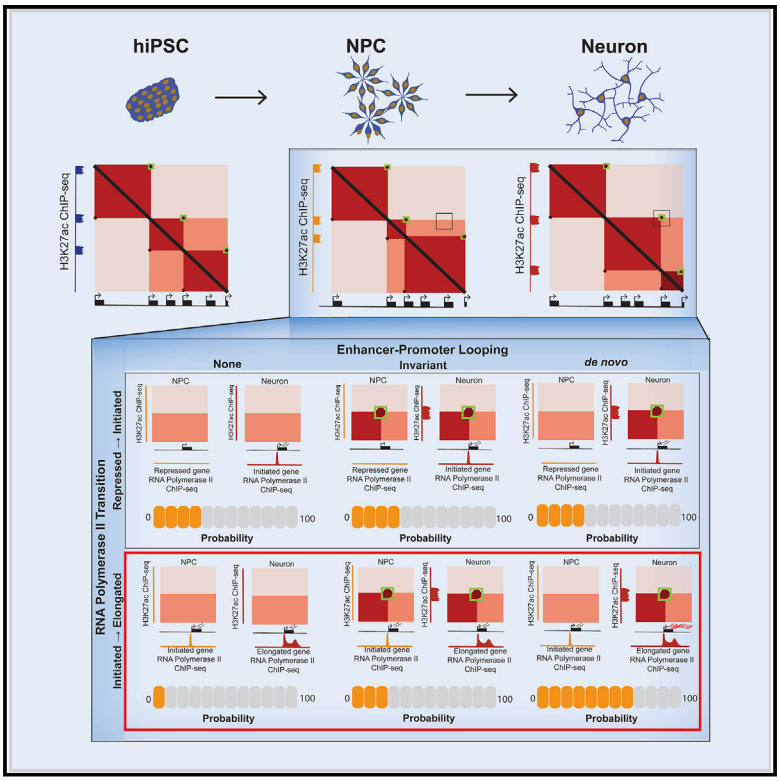
Cell Genomics

Resource

Cell-type-specific loops linked to RNA polymerase II elongation in human neural differentiation

Graphical abstract



Highlights

- Thousands of loops are lost and gained upon human iPSC differentiation to NPCs/neurons
- Elongated genes most often anchor cell-type-specific promoter-enhancer loops
- Initiated genes exhibit equal probability of looping versus not looping
- Loops anchoring elongated genes are severely disrupted by RNA polymerase II depletion

Authors

Katelyn R. Titus, Zoltan Simandi, Harshini Chandrashekar, Dominik Paquet, Jennifer E. Phillips-Cremins

Correspondence

jcremins@seas.upenn.edu

In brief

Titus et al. examine the structure-function relationship between chromatin looping and RNA polymerase II-mediated gene expression during human induced pluripotent stem cell differentiation to neural progenitors and post-mitotic neurons. They find that cell-type-specific promoter-enhancer loops strongly correlate with and may be functionally linked to RNA polymerase II elongation.

Titus et al., 2024, Cell Genomics *4*, 100606
August 14, 2024
https://doi.org/10.1016/j.xgen.2024.100606

CellPress

Cell Genomics

CellPress
OPEN ACCESS

Resource

Cell-type-specific loops linked to RNA polymerase II elongation in human neural differentiation

Katelyn R. Titus,[1,3,4,6] Zoltan Simandi,[1,3,4,6] Harshini Chandrashekar,[1,3,4,6] Dominik Paquet,[2,5] and Jennifer E. Phillips-Cremins[1,3,4,7,*]

[1]Department of Bioengineering, School of Engineering and Applied Science, University of Pennsylvania, Philadelphia, PA, USA
[2]Institute for Stroke and Dementia Research (ISD), University Hospital, LMU Munich, Munich, Germany
[3]Epigenetics Institute, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
[4]Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
[5]Munich Cluster for Systems Neurology (SyNergy), Munich, Germany
[6]These authors contributed equally
[7]Lead contact
*Correspondence: jcremins@seas.upenn.edu
https://doi.org/10.1016/j.xgen.2024.100606

SUMMARY

DNA is folded into higher-order structures that shape and are shaped by genome function. The role of long-range loops in the establishment of new gene expression patterns during cell fate transitions remains poorly understood. Here, we investigate the link between cell-specific loops and RNA polymerase II (RNA Pol II) during neural lineage commitment. We find thousands of loops decommissioned or gained *de novo* upon differentiation of human induced pluripotent stem cells (hiPSCs) to neural progenitor cells (NPCs) and post-mitotic neurons. During hiPSC-to-NPC and NPC-to-neuron transitions, genes changing from RNA Pol II initiation to elongation are >4-fold more likely to anchor cell-specific loops than repressed genes. Elongated genes exhibit significant mRNA upregulation when connected in cell-specific promoter-enhancer loops but not invariant promoter-enhancer loops or promoter-promoter loops or when unlooped. Genes transitioning from repression to RNA Pol II initiation exhibit a slight mRNA increase independent of loop status. Our data link cell-specific loops and robust RNA Pol II-mediated elongation during neural cell fate transitions.

INTRODUCTION

Mammalian development requires the precise spatiotemporal regulation of gene expression in defined cell types. Non-coding *cis*-regulatory elements (CREs) known as enhancers regulate cell-type-specific patterns of gene expression in metazoans.[1–3] CREs can be separated by kilobases (kb) up to megabases (Mb) from their target genes[4,5]; therefore, understanding how chromatin loops are involved in governing enhancer-promoter (E-P) communication is critical toward our knowledge of gene expression regulation.

There are three broad categories of long-range looping interactions: promoter-promoter (P-P) loops, E-P loops, and structural loops with no direct connection to promoters. In the case of structural loops, it is established that they are anchored by the architectural proteins CTCF and cohesin.[6,7] During interphase in steady-state mammalian cells, the cohesin complex extrudes DNA through its ring in an ATP-dependent manner to create transient "loops in the making."[8–11] Cohesin-mediated loop extrusion stalls at CTCF-occupied motifs oriented toward each other in a convergent orientation.[10,12,13] Extrusion stalling manifests empirically as focal "dot-like" structures representing loop anchors in ensemble Hi-C heatmaps.[14] Short-term degradation of CTCF or the Rad21 subunit of cohesin ablates structural loops.[15–17] By contrast, only a subset of long-range E-P or P-P loops are anchored by CTCF and sensitive to its degradation.[16] The zinc-finger-containing transcription factor Yin Yang 1 and subunits of the Mediator complex have recently been linked to E-P loop maintenance, but results vary depending on the cell type and methodologies used for molecular perturbation.[16–23] Overall, the mechanisms governing E-P and P-P loops and how they are functionally linked to transcription remain important open questions.

Enhancers and promoters are enriched for genetic motifs encoding binding sites for transcription factors. Through the process of protein-protein interactions, motif-bound transcription factors at CREs and transcription start sites (TSSs) can recruit co-factors and RNA polymerase II (RNA Pol II).[24] Transcription establishment requires assembly of the pre-initiation complex at the gene promoter, Cdk7-mediated C-terminal phosphorylation of RNA Pol II-serine 5 (Ser5P), and RNA Pol II initiation.[25–27] RNA Pol II transcribes promoter-proximal RNA until it pauses approximately 20–60 bp downstream of the TSS via the direct binding to negative elongation factor (NELF) and

DRB-sensitivity-inducing factor (DSIF).[28,29] Paused RNA Pol II is released into productive elongation via the CDK9 catalytic domain of positive transcription elongation factor (P-TEFb), which phosphorylates serine 2 on the C-terminal domain of RNA Pol II as well as NELF and DSIF.[25–27] RNA Pol II-mediated enhancer RNAs have also been detected at CREs in mammalian cells.[30] Two recent studies demonstrated that short-term RNA Pol II degradation can disrupt a subset of E-P loops in steady-state human cell lines,[31,32] suggesting a direct link between transcription and loop maintenance. However, the link between cell-type-specific E-P and P-P loops and the establishment of new RNA Pol II initiation and elongation patterns remains an important unanswered question.

Here, we set out to understand the relationship between loops and RNA Pol II during the establishment of new transcriptional programs during permanent changes in cell fate. We create genome-wide reference maps of long-range chromatin loops, gene expression, and RNA Pol II occupancy during neural lineage commitment in the transitions from human induced pluripotent stem cells (hiPSCs) to neural progenitor cells (NPCs) and NPCs to post-mitotic neurons. We uncover a strong link among cell-type-specific E-P loops gained *de novo* during differentiation, RNA Pol II elongation, and a robust increase of gene expression during human neural lineage commitment. We also demonstrate that loops anchored by elongated genes are particularly sensitive to short-term RNA Pol II degradation, whereas loops anchoring initiated genes bound by CTCF are protected from RNA Pol II perturbation. Our work sheds new light on the genome's dynamic and context-dependent structure-function relationship by linking chromatin loops to transcription elongation during human neural lineage commitment.

RESULTS

Rewiring of chromatin loops in a human iPSC model of early neural lineage commitment

We set out to conduct a genome-wide analysis of long-range chromatin loops in an *in vitro* human neurodevelopmental model consisting of three cellular stages: (1) undifferentiated hiPSCs (day 0); (2) iPSC-derived NPCs (days *in vitro* [DIV]35), and (3) post-mitotic cortical neurons (neurons, DIV65) (Figure 1A). We implemented a well-established, multi-stage protocol for growth-factor-mediated neuronal differentiation of iPSCs in monolayer tissue culture (STAR Methods).[33,34] After confirming the absence of karyotypic abnormalities, we used immunofluorescence and microscopy to evaluate the efficiency of differentiation (STAR Methods). We observed that hiPSCs exhibit homogeneous expression of pluripotency markers SSEA4 and OCT3/4 (also known as POU5F1) (Figure 1B). We further observed that NPCs homogeneously express forebrain progenitor markers FOXG1 and NESTIN and that post-mitotic neurons showed homogeneous expression of the pan-neural marker MAP2 and layer V-specific marker CTIP2 (Figure 1B). Thus, our hiPSCs, NPCs, and neurons homogeneously display the expected morphology and protein markers characteristic of the stage of neural lineage commitment.

We assayed higher-order chromatin folding genome-wide by creating Hi-C maps in hiPSCs, NPCs, and neurons. We acquired 396 (hiPSCs), 474 (NPCs), and 426 (neurons) million unique valid pairs for two biological replicates per condition, achieving a read depth sufficient for high-resolution chromatin loop calling (STAR Methods). We merged replicates to create genome-wide ensemble interaction frequency maps for all chromosomes at 10 kb resolution. To identify dot-like structures representative of *bona fide* loops in ensemble Hi-C data,[7,35–37] we employed statistical methods developed by our lab and others. Dots are characterized in Hi-C maps as a punctate group of adjacent pixels with significantly higher contact frequency compared to the surrounding local topologically associating domain (TAD) and sub-TAD structure. We identified 17,178 loops in hiPSCs, 12,827 loops in NPCs, and 14,752 loops in neurons (Table S1; STAR Methods). We verified that our loop calls were robust across a sweep of parameters[37] and verified by visual inspection of Hi-C heatmaps (Figure 1C).

We previously developed a statistical method, 3DeFDR, to identify loops with an invariant or cell-type-specific interaction frequency between two biological conditions.[36] To quantify changes in looping genome wide, we used 3DeFDR to compare hiPSC to NPC conditions and NPC to neuron conditions (Figures 1C–1E). We found 9,145 hiPSC-specific loops, 5,348 NPC-specific loops, and 6,800 invariant loops for the hiPSC-to-NPC transition (Figures 1D and 1E, top; Table S2). We further identified 2,311 NPC-specific loops, 4,231 neuron-specific loops, and 11,304 invariant loops for the NPC-to-neuron transition (Figures 1D and 1E, bottom; Table S3). Our cell-type-specific and invariant loop calls were confirmed by visual inspection of Hi-C heatmaps. At the *ADCY2* locus, we found loops that were decommissioned (prototypical hiPSC-specific loops) and gained *de novo* (prototypical NPC-specific and neuron-specific loops) upon differentiation (Figure 1C). Aggregate peak analysis plots of average interaction frequency validated our invariant and cell-type-specific loop calls across both hiPSC-to-NPC and NPC-to-neuron transitions (Figure 1E; STAR Methods). Together, these results demonstrate that differentiation of hiPSCs into NPCs and neurons results in substantial rewiring of loops genome wide.

Classifying genes into repressed, initiated, and elongated RNA Pol II occupancy

We next set out to classify genes at isoform resolution by their signature patterns of RNA Pol II occupancy. It is well established that repressed genes will exhibit minimal to no RNA Pol II signal, initiated genes will exhibit a strong peak-like signal at the TSS and a minimal signal at the gene body, and elongated genes will exhibit an RNA Pol II peak at the TSS as well as a domain-like signal across the gene body.[27,38] We generated RNA Pol II chromatin immunoprecipitation sequencing (ChIP-seq) libraries in hiPSCs, NPCs, and neurons (STAR Methods). To stratify genes into repressed, initiated, and elongated categories, first, we merged genes with the same TSSs and transcription end sites (TESs) into transcriptional units, thus circumventing redundancies due to shared TSSs. We next quantified (1) the maximum RNA Pol II signal in a window from −2 kb to the TSS, (2) the mean RNA Pol II signal across the gene body normalized by gene length, and (3) the mean signal within a window at the gene's 3′ end that

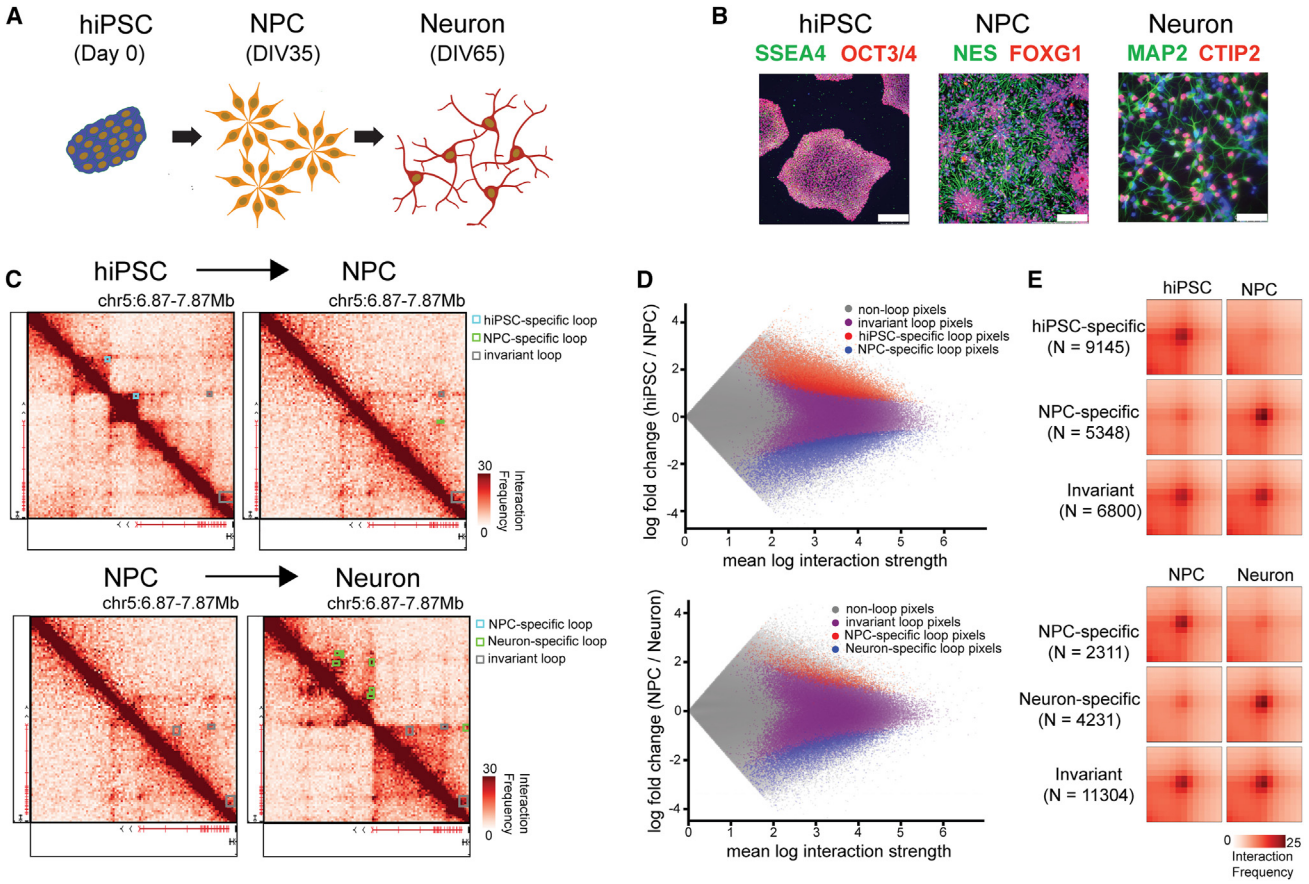


Figure 1. Differentiation of hiPSCs into NPCs and neurons leads to genome-wide rewiring of long-range chromatin loops

(A) A cartoon schematic of the differentiation from human induced pluripotent stem cells (hiPSCs) to neural progenitor cells (NPCs) and neurons. DIV represents the days of *in vitro* differentiation at which the samples were collected.

(B) Immunostaining of hiPSCs, NPCs, and neurons for cell-type-specific markers. Scale bar, 250 μm.

(C) Hi-C heatmaps at the *ADCY2* locus with annotated invariant and cell-type-specific loops between hiPSCs and NPCs (top row) and between NPCs and neurons (bottom row). *ADCY2* is highlighted in red on the gene track below the heatmap.

(D) MA plots of cell-type-specific loops (STAR Methods) between hiPSCs and NPCs (top row) and between NPCs and neurons (bottom row). Non-loops, gray. Invariant loops, purple. Cell-type-specific loops, red and blue.

(E) Aggregate peak analysis of differential and invariant loops across hiPSC and NPC conditions (top three rows) and NPC and neuron conditions (bottom three rows).

is proportional to the size of the transcriptional unit normalized by base pairs (Figures 2A and 2B; detailed in the STAR Methods). We verified that the computational strategy is effective in parsing repressed genes devoid of a RNA Pol II signal, initiated genes with an RNA Pol II peak at the promoter and depleted of a signal along the gene body, and elongated genes with a high RNA Pol II signal at the TSS and in a domain-like occupancy pattern along the gene body (Figure 2C; Table S4). Thus, we have classified repressed, initiated, and elongated transcriptional units across all three cell types.

We next assessed CTCF occupancy patterns at the promoters of repressed, initiated, and elongated genes. Using CTCF ChIP-seq in hiPSCs, NPCs, and neurons, we intersected the 2 kb segment upstream of all classified TSSs with CTCF peaks (Figure 2C; Table S5). Across the three cellular stages, we observed that less than 20% of promoters for repressed transcriptional units are occupied by CTCF. Promoters of initiated and elongated transcriptional units exhibit increases in CTCF occupancy compared to repressed genes. Our observations suggest that promoter occupancy of the architectural protein CTCF increases during the transition from repressed to initiated but no substantial increase in genome-wide promoter occupancy occurs during the transition from initiation to elongation. Thus, it is unlikely that CTCF binding at the promoter alone could distinguish between initiated and elongated genes.

We sought to verify the elongated gene group by measuring their mRNA levels with bulk RNA-seq. We further stratified non-redundant, unique transcriptional units by those that were elongated only in hiPSCs, only in NPCs, or only in neurons (Figure 2D). We found that hiPSC-specific elongated transcriptional units exhibited significantly higher mRNA levels in hiPSCs compared to NPCs and neurons. Similarly, we found that NPC-specific elongated transcriptional units had significantly higher expression in NPCs compared to hiPSCs and neurons.

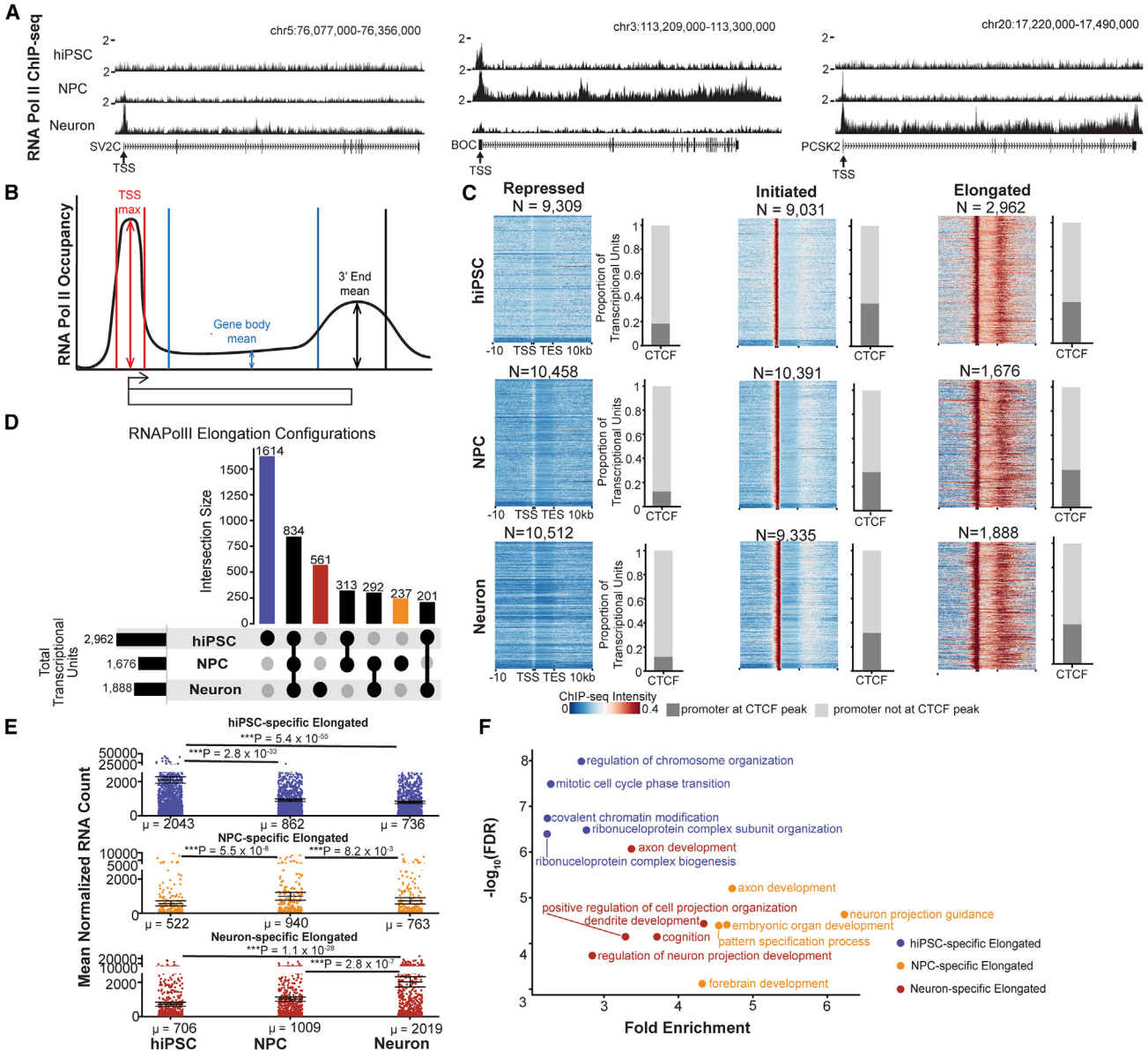


Figure 2. Categorization of transcriptional units by RNA Pol II occupancy into repressed, initiated, and elongated

(A) Examples of transcriptional units exhibiting RNA Pol II occupancy indicative of initiation in neurons (left), elongation in NPCs (middle), and elongation in neurons (right).

(B) Schematic representing the windows used to categorize transcriptional units into repressed, initiated, and elongated by RNA Pol II occupancy patterns (STAR Methods).

(C) RNA Pol II ChIP-seq heatmaps for transcriptional units categorized as repressed (left), initiated (middle), and elongated (right) in each cell type. Each row represents an individual transcriptional unit. Adjacent to the heatmaps are stacked bar plots illustrating the relative proportions of promoters co-localized with CTCF peaks.

(D) Genes classified as elongated in hiPSCs, NPCs, and neurons. Horizontal bars, the number of unique transcriptional units exhibiting cell-type-specific elongation. hiPSC-only (blue), NPC-only (orange), and neuron (red) elongated transcriptional units.

(E) mRNA levels from RNA-seq for transcriptional units classified into blue, red, and orange in (D). Points represent the mean normalized RNA count from three biological replicates. Horizontal lines represent the mean across all points. *p* values, two-tailed Mann-Whitney U (MWU) test with significance based on $\alpha = 0.01$.

(F) Gene Ontology analysis for transcriptional units classified into blue, red, and orange in (D).

Neuron-specific elongated transcriptional units exhibited significantly higher expression in neurons compared to hiPSCs and NPCs (Figure 2E). All three groups of cell-type-specific elongated genes showed the expected ontology for each cellular state (Figure 2F), thus verifying our approach for stratifying elongated genes by RNA Pol II occupancy.

CellPress
OPEN ACCESS

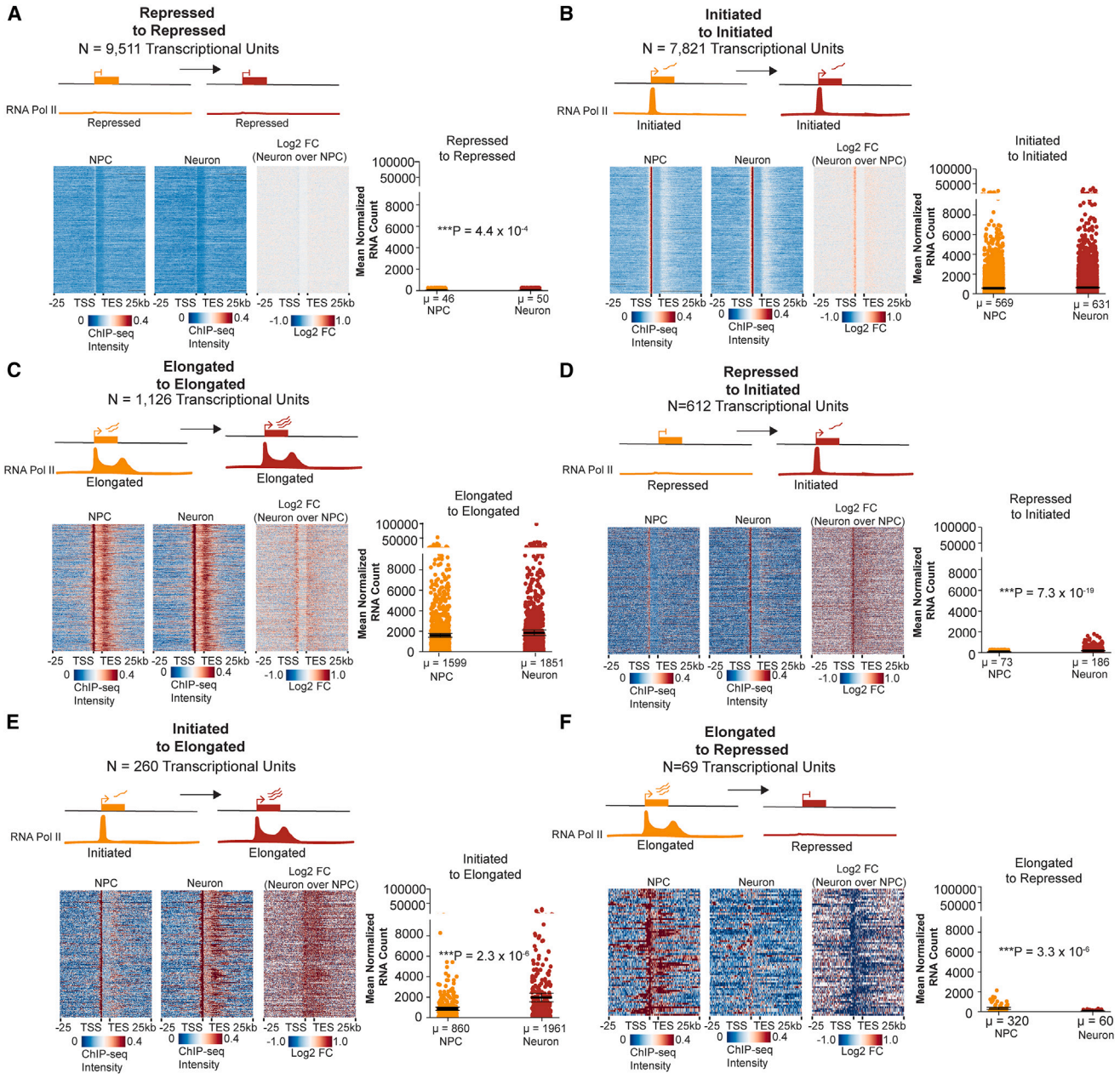


Figure 3. Dynamic RNA Pol II occupancy patterns during the cell fate transition from human NPCs to post-mitotic neurons

RNA Pol II ChIP-seq heatmaps and mRNA levels from RNA-seq for transcriptional units categorized during the transition from NPCs to post-mitotic neurons, including (A) repressed (NPCs) to repressed (neurons), (B) initiated (NPCs) to initiated (neurons), (C) elongated (NPCs) to elongated (neurons), (D) repressed (NPCs) to initiated (neurons), (E) initiated (NPCs) to elongated (neurons), and (F) elongated (NPCs) to repressed (neurons). Each row represents an individual transcriptional unit. Points in the mRNA plots on the right in each image represent the mean normalized RNA count from three biological replicates. Horizontal lines represent the mean across all points. *p* values, two-tailed MWU test with significance based on $\alpha = 0.01$.

Dynamic RNA Pol II occupancy during hiPSC-to-NPC and NPC-to-neuron cell fate transitions

Terminal differentiation of proliferating NPCs to post-mitotic neurons represents a unique cell fate transition in which chromatin is no longer subjected to the cell cycle. Focusing first on the NPC-to-neuron transition, we characterized genes into categories indicative of dynamic RNA Pol II occupancy during cell fate transitions, including (1) constitutively repressed in both NPCs and neurons, (2) constitutively initiated in both NPCs and neurons, (3) constitutively elongated in both NPCs and neurons, (4) repressed (NPCs) to initiated (neurons), (5) initiated (NPCs) to elongated (neurons), and (6) elongated (NPCs) to repressed (neurons) (Figure 3). We confirmed that the transcriptional units constitutively repressed in the NPC-to-neuron transition

displayed minimal RNA Pol II occupancy and negligible mRNA levels (Figure 3A). We also confirmed that transcriptional units constitutively initiated and constitutively elongated in NPC-to-neuron transition displayed similar RNA Pol II occupancy and mRNA levels between these cell types (Figures 3B and 3C). Genes that transition from repressed to initiated during the NPC-to-neuron transition showed increased RNA Pol II occupancy at promoter regions in neurons and a moderate but significant increase in mRNA levels (Figure 3D). Moreover, genes that transition from initiated to elongated gained a pronounced domain-like pattern of RNA Pol II signaling along the gene body in neurons along with a significant upregulation in mRNA levels (Figure 3E). We also identified a group of genes displaying occupancy patterns consistent with RNA Pol II decommissioning from elongation in NPCs to minimal RNA Pol II occupancy in neurons, and these genes exhibit the expected downregulation in mRNA levels (Figure 3F). Such patterns were not restricted to the NPC-to-neuron transition, as the same RNA Pol II occupancy and gene expression patterns for all six gene classes were similarly identified in the hiPSC-to-NPC transition (Figure S1). Taken together, these data uncover gene classes with distinct dynamic RNA Pol II occupancy patterns during hiPSC-to-NPC and NPC-to-neuron cell fate transitions.

Genes transitioning from initiated to elongated during neural differentiation are strongly enriched for gained cell-type-specific looping interactions

To understand the link between loops and RNA Pol II occupancy, we stratified promoters of unique transcriptional units into those that are anchoring any NPC-specific (class I, blue), neuron-specific (class 2, green), mixed (class 3, yellow), or cell-type-invariant (class 4, dark gray) chromatin loops and those not engaged in any looping interactions (class 5, light gray) (Figures 4A and S2; STAR Methods). At baseline, 29.5% of constitutively repressed transcriptional units connect in cell-type-specific or invariant loops, whereas 79.2% of initiated-to-elongated transcriptional units connect in cell-type-specific or invariant loops during the NPC-to-neuron transition (Figures 4B, first versus fifth barplot, and S2). Transcriptional units transitioning from repressed NPCs to initiated neurons exhibit equal probability of looping versus not looping to distal enhancers (Figure 4B, fourth barplot). By contrast, transcriptional units transitioning from initiated to elongated are 3.5-fold more likely to be engaged in cell-type-specific loops compared to transcriptional units that are constitutively repressed (Figure 4B, fifth barplot). Our data indicate that most transcriptional units with RNA Pol II occupancy patterns indicative of elongation are anchored by invariant or cell-type-specific loops, whereas RNA Pol II initiation occurs with similar likelihood in transcriptional units that loop or do not loop.

A noteworthy observation is that both initiated and elongated genes are enriched for cell-type-specific chromatin loops, which arise *de novo* during neural differentiation. We observed that 22.3% of transcriptional units transitioning from repressed-to-initiated RNA Pol II occupancy (Figure 4B, fourth barplot) and 33% of transcriptional units transitioning from initiated-to-elongated RNA Pol II occupancy (Figure 4B, fifth barplot) gain neuron-specific loops that form *de novo* during the NPC-to-neuron transition (e.g., green class 2 neuron-specific loops). By contrast, only 7.5% of transcriptional units constitutively repressed and 12.7% of transcriptional units constitutively initiated in the NPC-to-neuron transition engage in neuron-specific class 2 loops (Figure 4B, first and second barplots). Example neuron-specific loops gained in the NPC-to-neuron transition can be observed at the *SV2C* gene classified as transitioning from repressed to initiated (Figure 4C) and the *PCSK2* gene classified as transitioning from initiated to elongated (Figure 4D). We observed similar trends during the iPSC-to-NPC cell fate transition (Figures S3A–S3E and S4). Our data suggest that transcriptional units are 3- to 4-fold more likely to engage in cell-type-specific chromatin loops when they are initiated or elongated compared to repressed during hiPSC-to-NPC and NPC-to-neuron cell fate transitions.

Genes transitioning from elongated to repressed during neural differentiation are strongly enriched for decommissioned cell-type-specific looping interactions

We also examined the relationship between loops and genes transitioning from elongated to repressed during NPC-to-neuron differentiation. Transcriptional units that are elongated in NPCs and lose RNA Pol II occupancy in neurons are strongly enriched in loops that are decommissioned in neurons (e.g., blue class 1 NPC-specific loops) (Figure 4B). Specifically, we observe that 28.9% of decommissioned transcriptional units also lose loops, which is 6.5-fold higher than the 4.3% of constitutively repressed transcriptional units that engage in class 1 NPC-specific loops and 3.8-fold higher than the 7.5% of constitutively elongated transcriptional units that engage in class 1 NPC-specific loops (Figure 4B, sixth barplot, blue). Loop decommissioning is exemplified at the *BOC* gene classified with an elongated-to-repressed RNA Pol II occupancy pattern (Figure 4E). Taken together, we observe that the establishment of new cell-type-specific loops correlates with transitions into initiated and elongated RNA Pol II, and the decommissioning of loops correlates with the loss of RNA Pol II occupancy during neural differentiation.

Genes transitioning from repressed to initiated exhibit slight increases in mRNA levels independent of loop status

We next set out to distinguish between cell-type-specific E-P and P-P loops for their relationship to RNA Pol II and mRNA levels. We stratified the NPC-specific, neuron-specific, and invariant loops in Figure 4A and the iPSC-specific, NPC-specific, and invariant loops in Figure S3 into the subset connecting E-P loops (Figures 5A–5C, S5, S6, and S7A–S7C) and those connecting P-P loops (Figures S8A–S8C, S5, S6, and S9A–S9C). We integrated E-P and P-P loops with genes classified by changes in RNA Pol II occupancy and examined mRNA levels for both hiPSC-to-NPC (Figures S6, S7, and S9) and NPC-to-neuron (Figures 5, S5, and S8) transitions (STAR Methods).

Our combined analyses revealed several key points. First, for transcriptional units transitioning from repressed to initiated RNA Pol II occupancy patterns, we observed a small increase in mRNA levels regardless of whether the transcriptional unit engaged in cell-type-specific E-P loops, invariant E-P loops, or no loops at all (Figures 5A and S7A). Further, we observed slight

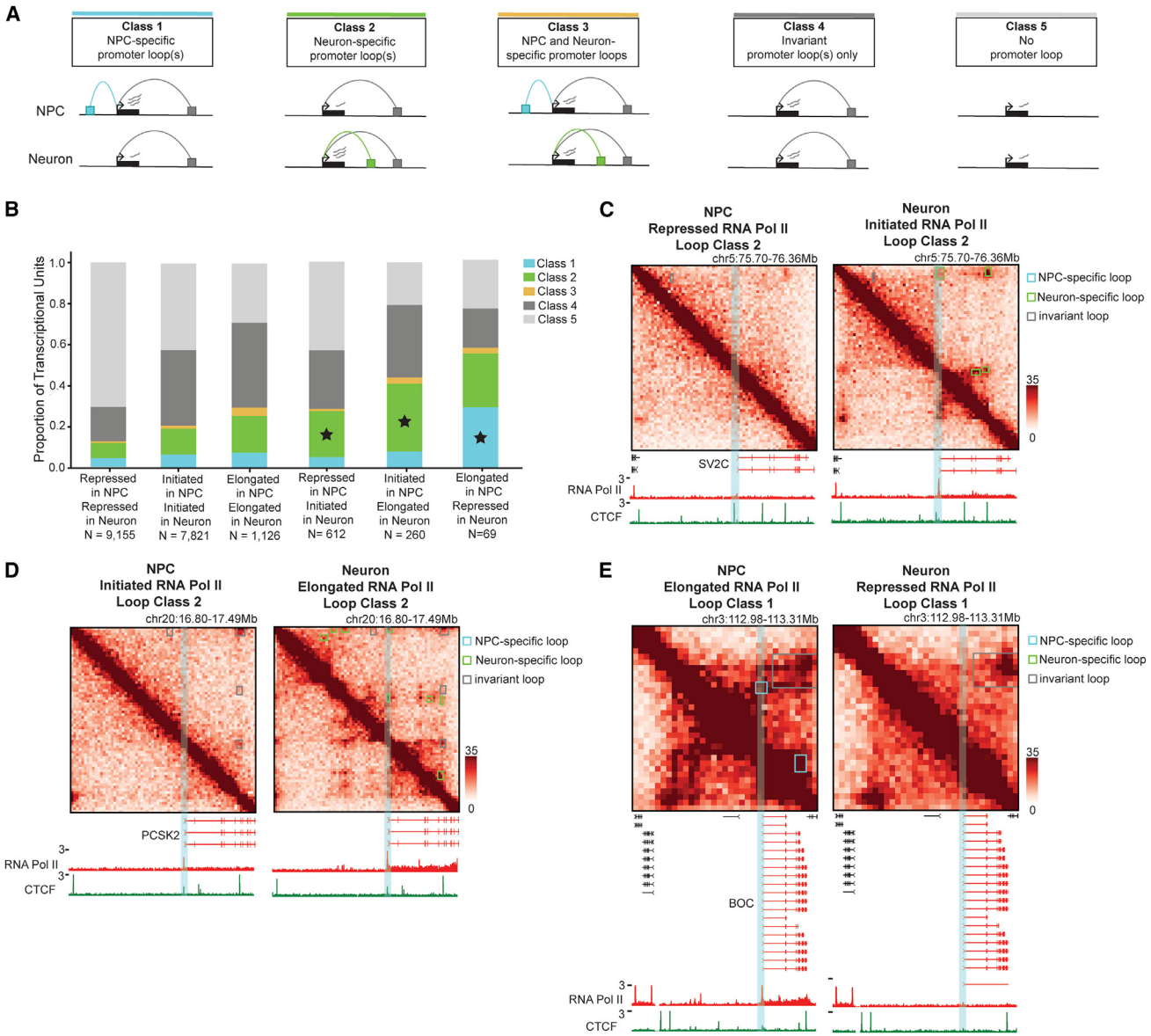


Figure 4. Genes transitioning from initiated to elongated are more likely to form cell-type-specific loops than constitutively repressed and initiated genes

(A) Schematic depicting promoters of unique transcriptional units classified as anchoring NPC-specific loops (class 1, blue), neuron-specific loops (class 2, green), mixed loops (class 3, yellow), cell-type-invariant loops (class 4, dark gray), or not looping (class 5, light gray).

(B) Proportion of genes engaged in the five looping categories from (A). Black stars (★) indicate loop classifications of interest illustrated in (C)–(E).

(C–E) Hi-C heatmaps during NPC-to-neuron differentiation: (C) the *SV2C* gene illustrating RNA Pol II occupancy characteristic of the repressed-to-initiated cell fate transition, (D) the *PCSK2* gene illustrating RNA Pol II occupancy characteristic of initiated-to-elongated genes, and (E) the *BOC* gene illustrating RNA Pol II occupancy characteristic of elongated-to-repressed transition. Tracks below Hi-C heatmaps show RNA Pol II and CTCF ChIP-seq data.

increases in mRNA levels whether the transcriptional units engaged in E-P or P-P loops (Figures 5A, S7A, S8A, and S9A). Trends were similar in both the hiPSC-to-NPC and NPC-to-neuron transitions (Figures 5 and S7–S9). These findings suggest that the recruitment of RNA Pol II to the TSS in the repressed-to-initiated transition during neural differentiation can be independent of looping status.

Genes transitioning from initiated to elongated exhibit robust mRNA upregulation primarily when connected in cell-type-specific E-P loops

We next examined genes transitioning from initiated to elongated RNA Pol II occupancy. We observed a striking upregulation of mRNA levels for those connected in cell-type-specific E-P loops (Figures 5B and S7B, second column). Similar patterns

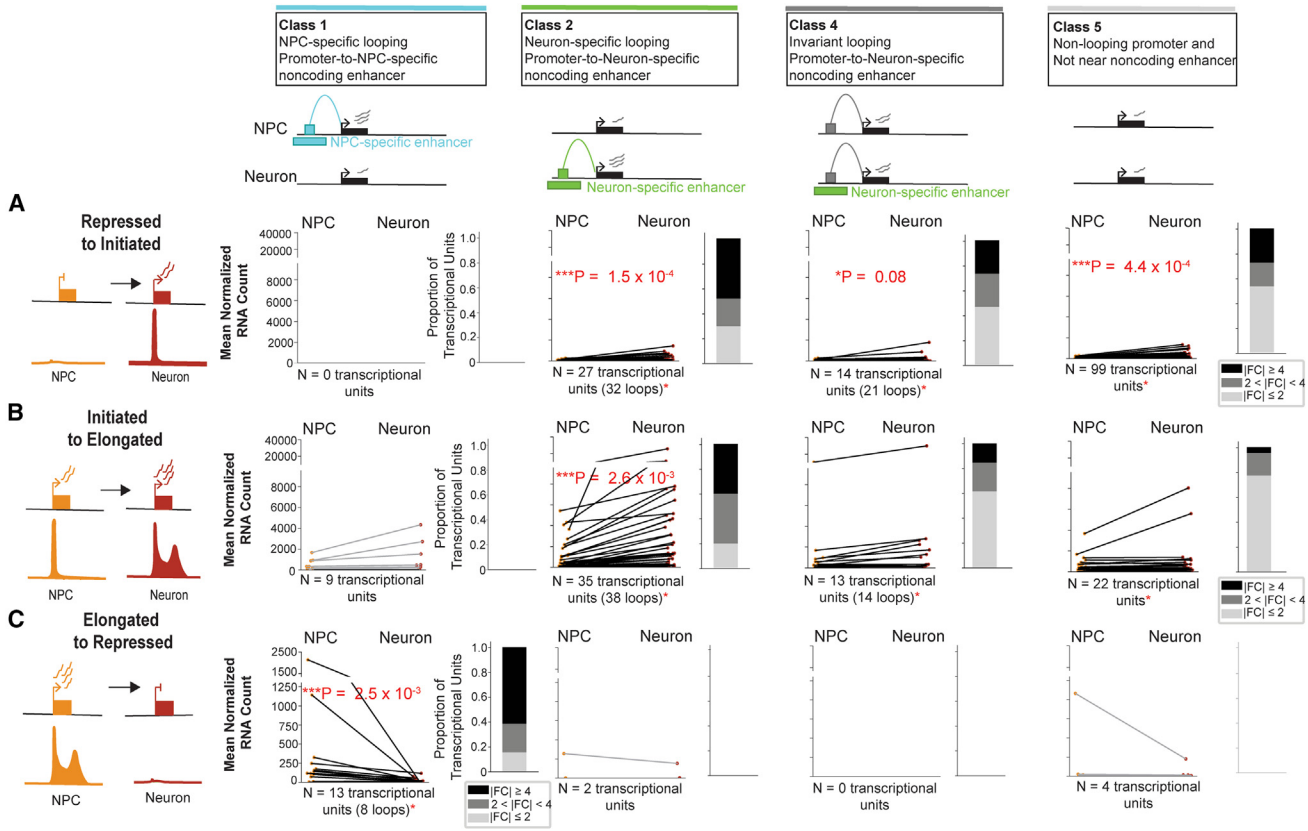


Figure 5. Upregulation of mRNA levels is highest for elongated genes engaged in *de novo* cell-type-specific enhancer-promoter loops during NPC-to-neuron differentiation

(A–C) mRNA levels calculated from RNA-seq for transcriptional units stratified as (A) repressed to initiated, (B) initiated to elongated, and (C) elongated to repressed during NPC-to-neuron differentiation. These three transcript classes are stratified into those that (column 1) anchor NPC-specific enhancer-promoter loops decommissioned in differentiation (class 1, blue), (column 2) anchor neuron-specific enhancer-promoter loops gained *de novo* in differentiation (class 2, green), (column 3) anchor cell-type-invariant loops connecting promoters to neuron-specific promoters (class 4, dark gray), and (column 4) do not loop and lack enhancers within a 80 kb vicinity of the TSS (class 5, light gray). Each point is the mean normalized RNA count across three replicates. A two-tailed MWU test was used to compute *p* values. An asterisk (*) reaches significance under α = 0.1 and triple asterisks (***) reach significance under α = 0.01. Stacked bar plots represent the proportion of transcriptional units that exhibit absolute fold change expression less than 2, greater than 2 and less than 4, and greater than 4 between NPC and neuron cellular states. Plots and *p* values are shown for conditions with sample sizes sufficient to assess trends.

of mRNA upregulation did not occur for initiated-to-elongated genes when connected in P-P loops, invariant E-P loops, or no loops (Figures 5B and S7B, fourth column, and S8B and S9B, second column). Observations were similar in both the hiPSC-to-NPC and NPC-to-neuron transitions (Figures 5 and S7–S9). Together, these data indicate that although genes transitioning from repressed to elongated exhibit slight increases in mRNA levels independent of looping status, the transition from initiation to elongation correlates with significantly higher levels of steady-state mRNA levels when the transcriptional units are engaged in cell-type-specific E-P loops during neural differentiation.

Decommissioning from elongation to repression can involve the breaking of cell-type-specific E-P loops and can also occur when genes do not engage in loops

We also investigated genes transitioning from elongated to repressed RNA Pol II occupancy. We observed downregulation of mRNA levels whether the transcriptional units engaged in decommissioned cell-type-specific E-P loops or did not loop (Figures 5C and S7C, first barplot). We did not find consistent decreases in mRNA levels when decommissioned genes engaged in invariant E-P loops or P-P loops (Figures S8C and S9C). Similar trends were observed in the iPSC-to-NPC and NPC-to-neuron transitions (Figures 5 and S7–S9). These results demonstrate that the decommissioning from productive elongation can involve the breaking of cell-type-specific E-P loops and also occur when the genes do not engage in loops.

Cell-type-specific E-P looping and RNA Pol II signal at enhancers correlate with mRNA upregulation during neural differentiation

To better understand the correlation between cell-type-specific E-P looping and mRNA levels in the NPC-to-neuron transition, we stratified cell-type-specific enhancers into those with an enhancer RNA Pol II signal and those without (Figures S10A–S10C; STAR Methods). For repressed-to-initiated genes, we

found that 38% and 36% of neuron-specific enhancers exhibit RNA Pol II occupancy when anchoring the base of *de novo* E-P loops and invariant E-P loops, respectively (Figure S10A). For elongated-to-repressed genes, we observed similar patterns (Figure S10C). By contrast, for initiated-to-elongated genes, we found that 68% and 62% of neuron-specific enhancers exhibit RNA Pol II occupancy when anchoring the base of *de novo* E-P loops and invariant E-P loops, respectively (Figure S10B). Overall, we found a strong enrichment of RNA Pol II at enhancers anchoring the base of *de novo* E-P loops anchoring genes that transition from initiation to elongation during neural differentiation.

We also assessed how the presence of RNA Pol II at enhancers affects mRNA levels at the base of cell-type-specific E-P loops (Figures S10D–S10F). For initiated-to-elongated genes, we found a robust increase in mRNA levels when *de novo* promoter-enhancer loops are anchored by enhancers with RNA Pol II (Figure S10E, column 3). The same gene expression patterns were not observed at the same loops when the enhancers did not have RNA Pol II signal (Figure S10E, column 4). Our results suggest that the presence of RNA Pol II at looped enhancers strongly correlates with a robust upregulation of mRNA levels at elongated genes.

Short-term RNA Pol II perturbation disrupts loops anchored by elongated gene promoters

Finally, we sought to functionally test our observation of a strong correlation among RNA Pol II elongation, mRNA levels, and cell-type-specific loops between enhancers and promoters of elongated genes. We re-analyzed published Micro-C data from the DLD-1 cell line in which an auxin-inducible degron was used to degrade RNA Pol II.[32] Using published RNA Pol II ChIP-seq[39] and CTCF CUT&Tag[32] data in wild-type DLD-1 cells, we stratified transcriptional units into those with or without CTCF promoter occupancy and exhibiting repressed, initiated, and elongated RNA Pol II occupancy patterns (Figures 6A–6C; Table S6). Similar to our NPC and neuron model systems, we found that 6% of repressed, 46% of initiated, and 64% of elongated transcriptional units genome wide have CTCF bound at the promoters in DLD-1 cells (Figure S11). We identified loops genome wide (N = 24,358) in Micro-C data from control DLD-1 cells without auxin treatment (STAR Methods). Focusing on transcriptional units with promoters engaged in loops, we found that 27% of repressed transcriptional units, 68% of initiated transcriptional units, and 76% of elongated transcriptional units have at least one CTCF peak at the promoter (Figures S11 and 6A–6C). Transcriptional units with CTCF at the promoter are more likely to be engaged in promoter loops than those without CTCF (Figure S11). Promoter loops anchoring repressed transcriptional units exhibit negligible to slight changes in interaction frequency upon lost RNA Pol II occupancy (Figure 6A). Promoter loops anchoring initiated transcriptional units only exhibit a noticeable decrease in interaction frequency when CTCF is not bound to the promoter (Figure 6B). In contrast, we observed a striking decrease in interaction frequency of promoter loops anchoring elongated genes upon degradation of RNA Pol II occupancy after auxin treatment (Figure 6C). Elongated transcriptional units are susceptible to loop disruption upon RNA Pol II degradation independent of CTCF status (Figure 6C).

To validate our observations in an independent model system, we also re-analyzed published Micro-C data from the JM8.N4 mouse embryonic stem cell (ESC) line chemically treated with triptolide (TRP) and flavopiridol (FLV).[40] TRP inhibits transcription initiation by inducing proteasome-dependent degradation of RNA Pol II.[41] FLV inhibits the phosphorylation of P-TEFb, therefore preventing productive elongation.[42] We used RNA Pol II[40] and CTCF ChIP-seq[43] data in wild-type JM8.N4 mouse ESCs to stratify transcriptional units into those with or without CTCF promoter occupancy and exhibiting repressed, initiated, and elongated RNA Pol II occupancy patterns (Figures 6D–6F; Table S6). We identified 12,417 loops genome wide in Micro-C data generated in wild-type JM8.N4 mouse ESCs (STAR Methods). Promoter loops anchoring repressed transcriptional units exhibit negligible to slight changes in interaction frequency upon chemical treatment with either FLV or TRP, respectively (Figure 6D). Promoter loops anchoring initiated transcriptional units are not changed upon TRP treatment and only slightly weakened upon FLV treatment (Figure 6E). By contrast, we observed a marked weakening of promoter loops anchoring elongated transcriptional units upon FLV treatment, especially when CTCF is bound to the promoter (Figure 6F). These results further reinforce that RNA Pol II is necessary for the maintenance of promoter-enhancer loops anchoring elongated genes.

DISCUSSION

Understanding how chromatin loops interplay with RNA Pol II is important toward understanding the principles governing gene expression regulation in development. Here, we use a model of human iPSC differentiation to NPCs and neurons and create genome-wide kb-resolution maps of higher-order chromatin folding, gene expression, and RNA Pol II occupancy. We uncover an unexpected strong correlation between cell-type-specific E-P loops and genes with RNA Pol II occupancy patterns indicative of an initiation-to-elongation transition during differentiation. Elongated genes connected in E-P loops gained *de novo* during differentiation display robust upregulation of mRNA. By contrast, elongated genes anchoring P-P loops or invariant E-P loops or that are unlooped exhibit modest to negligible changes in mRNA levels. Moreover, we observe that genes with RNA Pol II patterns indicative of repressed-to-initiated and elongated-to-repressed transitions show changes in mRNA levels whether they are engaged in loops or not looping. Together, our data suggest that robust increases in transcription elongation strongly correlate with *de novo* E-P loops, whereas transcription repression, initiation, or decommissioning can occur whether the genes engage in loops or not (Figure 7A).

The link between RNA Pol II and looping has been debated, with conflicting results linked to low-resolution datasets or pharmacological agents with non-specific indirect effects.[22,31,32,40,44–47] Most recently, two studies employed a degron to achieve short-term RNA Pol II degradation and uncovered a previously underappreciated effect on E-P loop maintenance.[31,32] Here, we build on these works by creating genome-wide maps of cell-type-invariant and cell-type-specific chromatin loops during the transitions from hiPSCs to NPCs and NPCs to post-mitotic neurons. We observe a strong correlation between *de novo* cell-type-specific E-P

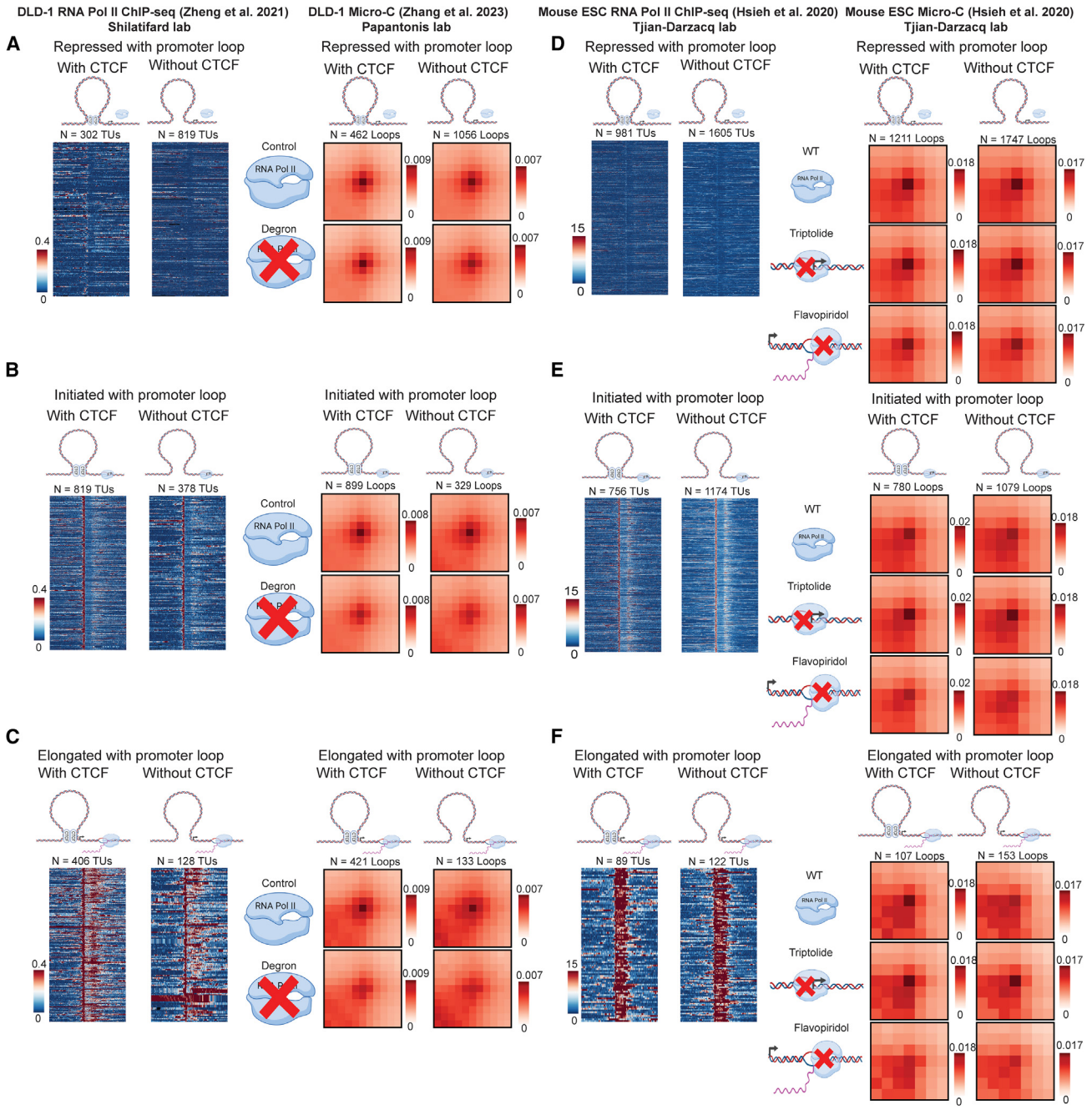


Figure 6. The interaction frequency of loops anchored by elongated genes is particularly susceptible to disruption upon RNA Pol II perturbation

(A–C) RNA Pol II ChIP-seq heatmaps for transcriptional units categorized as (A) repressed, (B) initiated, and (C) elongated in the wild-type DLD-1 cell line. (D–F) RNA Pol II ChIP-seq heatmaps for transcriptional units categorized as (D) repressed, (E) initiated, and (F) elongated in wild-type JM8.N4 mouse ESCs. Transcriptional units are further stratified by CTCF occupancy in a 2 kb window upstream of the TSS. The right column has aggregate peak analysis heatmaps of a Micro-C signal at loops called before and after short-term RNA Pol II perturbation. Some illustrations were created using Biorender.com.

chromatin loops, transcription elongation, and robust upregulation of mRNA levels. By contrast, we find that genes can transition from repressed-to-initiated RNA Pol II whether they engage in cell-type-specific loops or cell-type-invariant loops or do not loop to distal enhancers. We recognize that some E-P models are based on smaller datasets. Nevertheless, we believe that our approach offers a valuable trade-off, allowing us to explore a wider subtype of interactions and support a hypothesized role for E-P loops in the transition from RNA Pol II initiation to elongation. Therefore, we re-analyzed published Micro-C data from an established RNA Pol II

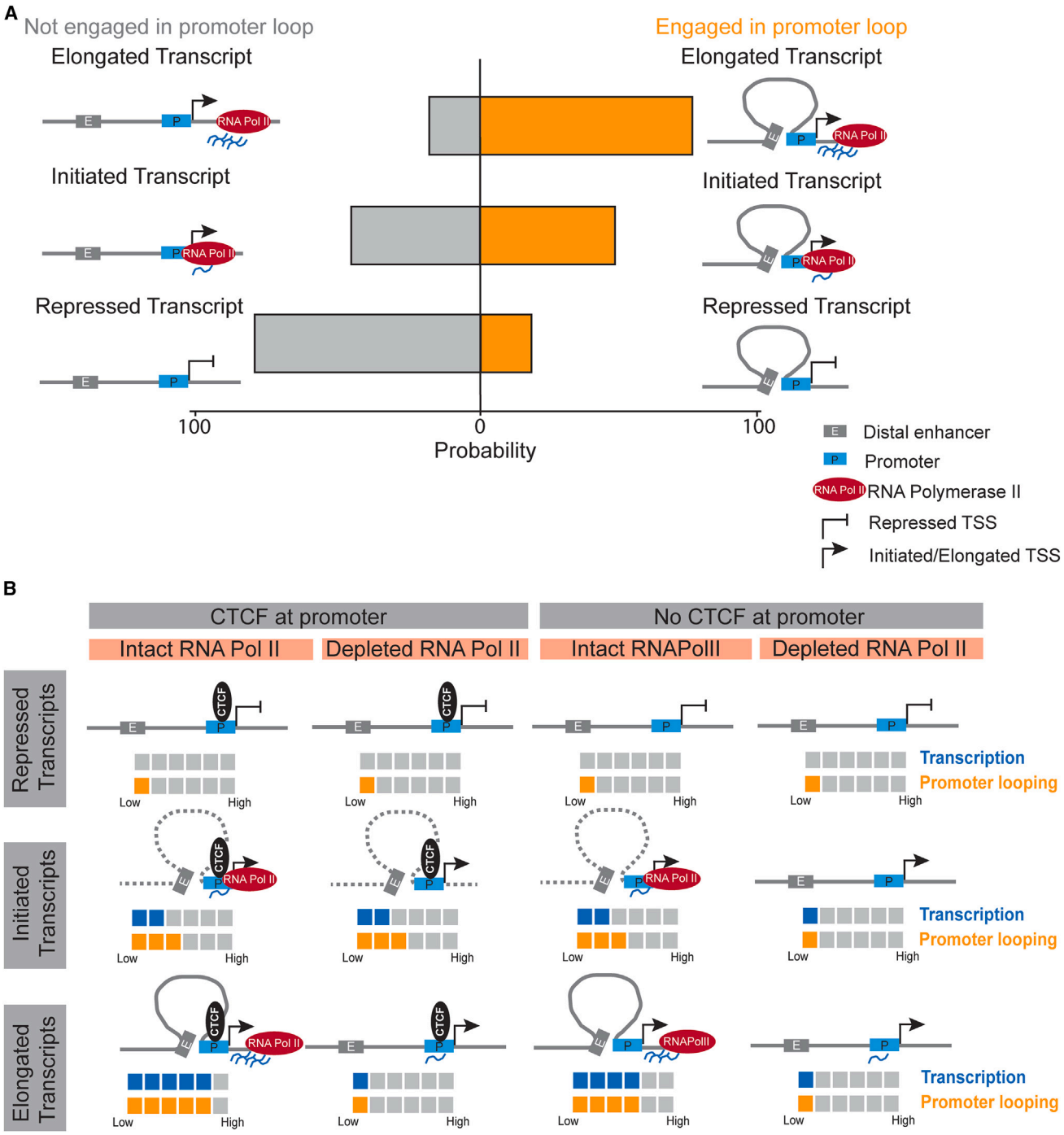


Figure 7. Schematic model of the link between RNA Pol II elongation, cell-type-specific E-P loops, and robust upregulation of gene expression during neural cell fate transitions

(A) The large majority of genes transitioning from initiated to elongated during neural differentiation connect in invariant or cell-type-specific looping interactions. Genes that remain repressed during differentiation generally do not loop. Genes transitioning from repressed to initiated during differentiation exhibit equal probability of looping or not looping.

(B) Models representing the sensitivity of loops formed by repressed, initiated, and elongated genes with or without a promoter bound CTCF to the short-term degradation of RNA Pol II.

degron system in a dividing mammalian cell line but taking steps to stratify gene isoforms among those that are repressed, elongated, and initiated. We demonstrate that loops anchored by initiated genes are slightly sensitive to RNA Pol II degradation but only when the promoter is CTCF independent. CTCF-bound initiated promoters anchor loops that are protected against disruption by RNA Pol II knockdown (Figure 7B). By contrast, the sensitivity of chromatin loops to RNA Pol II degradation is particularly strong when anchored by elongated promoters, regardless of CTCF occupancy (Figure 7B). While CTCF has been a major focus in the current study, it is important to point out the possible involvement of other structural proteins, such as cohesin in RNA Pol II pause and release.[48,49] Our observations reinforce that RNA Pol II causally contributes to the strength of loops anchored by elongated genes.

The models by which TSSs communicate with distal CREs to spatiotemporally regulate transcription remain hotly debated when CREs and TSSs are separated by kb to Mb of DNA. One leading model is the formation of stable long-range E-P loops in which promoter activity depends on sustained enhancer proximity via tethering.[19,50–53] A second leading "kiss-and-run" model involves the transient interaction of CREs with their target TSS.[54] Transient E-P contacts could, in principle, deposit persistent information on the TSSs to facilitate transcription, such as post-translational protein modifications, chromatin-modifying enzymes, or Pol II itself. E-P loops would be important for transcription, but ensemble Hi-C or DNA fluorescence *in situ* hybridization measurements of interaction frequency might not correlate with bulk mRNA measurements or even correlate temporally with burst size/frequency.[55,56] A third model, loop-independent long-range communication, involves the diffusion of biomolecules between the CRE and TSS.[56] Finally, a fourth "condensate" model involves the association of promoters and enhancers within foci representing concentrated subnuclear microenvironments of regulatory enzymes, proteins, and RNAs.[57] It has been proposed that Mediator, Brd4, and RNA Pol II can form foci—potentially through liquid-liquid phase-separation mechanisms.[58,59] Here, we take a step toward gathering data to build evidence toward testing these models. We observe that the large majority of elongated genes connect in invariant or cell-type-specific looping interactions, whereas repressed genes do not loop and initiated genes exhibit equal probability of looping or not looping (Figure 7A). Punctate focal dots represent loops detectable in ensemble Hi-C data because they are typically present in a large proportion of cells and unlikely to be transient. Therefore, our data are not consistent with transient kiss-and-run or loop-independent models for transcription elongation. Rather, our analyses suggest that genes with high levels of RNA Pol II elongation and robust upregulation of mRNA levels show genome folding patterns in ensemble Hi-C data more consistent with sustained E-P tethering in loops (model 1). While our analysis captures strong regulatory loops, weaker interactions may limit our understanding of the full regulatory landscape. Our data also cannot rule out condensates (model 4). Because condensates range from 200 to 500 nm in size, they, in principle, could create local environments of proximal access to similar biomolecules without direct contact.[57] Future causal studies should aim to dissect the causal role for loop extrusion and condensate formation in the loops observed connecting distal enhancers and promoters to transcription elongation.

The models governing long-range E-P communication remain an exciting open area, and further perturbative and single-cell imaging and genomics studies will garner further insight toward how each model alone or individually impacts gene regulation at the stages of initiation, pausing, and elongation. Altogether, our results highlight a link between transcriptional elongation and cell-type-specific E-P chromatin loops and emphasize the importance of future work dissecting the causal role for looping in gene expression regulation during human neural lineage commitment.

Limitations of the study

We aim to raise the limitations of our study. First, the E-P subclasses are small in size (Figures 5 and S7–S9). The limited size is due to the filtering process shown in the flowcharts (Figures S2 and S4–S6). Second, the RNA Pol II perturbation data we re-analyzed from human DLD-1 cells and mouse ESCs (Figure 6) examine a different biological context than our hiPSC-derived neuron model (Figures 1, 2, 3, 4, and 5). The published data study the role of RNA Pol II perturbation during the maintenance of gene expression in cell lines, whereas our model system reflects the establishment of new gene expression patterns during cell fate transitions. Third, although we found a correlation between RNA Pol II elongation and cell-type-specific promoter-enhancer loops, the causal role of loops in the establishment of new RNA Pol II elongation patterns remains unclear since we did not perform mechanistic perturbations in the hiPSC-derived neuron model system.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODELS AND SUBJECT DETAILS
  - Maintenance of human induced pluripotent stem cell (human iPSC) culture
  - hiPSC differentiation to neural progenitors (NPCs) and post-mitotic neurons
- METHOD DETAILS
  - Immunocytochemistry and microscopy
  - Total RNA-seq
  - Chromatin fixation for ChIP-seq and Hi-C
  - ChIP-seq
  - Hi-C
  - Library preparation (ChIP-seq, Hi-C)
  - hg38 RefSeq reference transcriptome
  - RNA-seq analysis
  - Hi-C pre-processing
  - Hi-C expect0065d modeling
  - Hi-C p Values
  - Hi-C multiple testing correction
  - Hi-C loop clustering
  - Hi-C differential loop calling
  - ChIP-seq analysis
  - RNAPolII occupancy analysis

- Intersection of RNAPolII occupancy profiles with CTCF peaks
- Gene ontology analysis
- Parsing RNAPolII transition classes
- Parsing looping classes
- Defining differential enhancer regions
- Classification of NPC-specific and neuron-specific loops using promoter positions and differential H3K27ac enhancer regions
- Stratification of promoter-promoter looping genes and non-looping genes
- Stratification of enhancers at cell-type specific enhancer-promoter loops by RNAPolII occupancy
- RNAPolII Occupancy Analysis of DLD-1 RNAPolII ChIP-seq
- Analysis of DLD-1 CTCF CUT&Tag
- Analysis of DLD-1 Micro-C
- mm10 RefSeq reference genome
- RNAPolII Occupancy Analysis of JM8.N4 mouse ES cell RNAPolII ChIP-seq
- Analysis of JM8.N4 mouse ES cell CTCF ChIP-seq
- Analysis of JM8.N4 mouse ES cell Micro-C

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xgen.2024.100606.

ACKNOWLEDGMENTS

This work was funded by the NIH NIMH (1DP1MH129957; J.E.P.-C.); the NIH NINDS (5-R01-NS114226; J.E.P.-C.); the 4D Nucleome Common Fund (1U01DK127405; J.E.P.-C.); an NSF CAREER Award (CBE-1943945; J.E.P.-C.); the NSF Emerging Frontiers Research Innovation (EFMA19-33400; J.E.P.-C.); CZI Neurodegenerative Disease Pairs Awards (2020-221479-5022 and DAF2022-250430; J.E.P.-C.); Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy within the framework of the Munich Cluster for Systems Neurology (EXC 2145 SyNergy – ID 390857198; D.P.); and donors of the ADR AD2019604S, a program of the BrightFocus Foundation (D.P.).

AUTHOR CONTRIBUTIONS

Conceptualization, K.R.T., H.C., Z.S., and J.E.P.-C.; methodology/visualization, K.R.T., H.C., Z.S., D.P., and J.E.P.-C.; investigation, K.R.T., H.C., Z.S., and J.E.P.-C.; funding, J.E.P.-C. and D.P.; administration, J.E.P.-C.; writing & editing, K.R.T., H.C., Z.S., and J.E.P.-C.; reagents, D.P.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: December 4, 2023
Revised: May 11, 2024
Accepted: June 17, 2024
Published: July 10, 2024

REFERENCES

1. Benoist, C., and Chambon, P. (1981). In vivo sequence requirements of the SV40 early promotor region. Nature *290*, 304–310. https://doi.org/10.1038/290304a0.
2. Levine, M. (2010). Transcriptional enhancers in animal development and evolution. Curr. Biol. *20*, R754–R763. https://doi.org/10.1016/j.cub.2010.06.070.
3. Long, H.K., Prescott, S.L., and Wysocka, J. (2016). Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. Cell *167*, 1170–1187. https://doi.org/10.1016/j.cell.2016.09.018.
4. Uslu, V.V., Petretich, M., Ruf, S., Langenfeld, K., Fonseca, N.A., Marioni, J.C., and Spitz, F. (2014). Long-range enhancers regulating Myc expression are required for normal facial morphogenesis. Nat. Genet. *46*, 753–758. https://doi.org/10.1038/ng.2971.
5. Tolhuis, B., Palstra, R.J., Splinter, E., Grosveld, F., and de Laat, W. (2002). Looping and interaction between hypersensitive sites in the active beta-globin locus. Mol. Cell *10*, 1453–1465.
6. Tang, Z., Luo, O.J., Li, X., Zheng, M., Zhu, J.J., Szalaj, P., Trzaskoma, P., Magalska, A., Wlodarczyk, J., Ruszczycki, B., et al. (2015). CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. Cell *163*, 1611–1627. https://doi.org/10.1016/j.cell.2015.11.024.
7. Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., and Aiden, E.L. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell *159*, 1665–1680. https://doi.org/10.1016/j.cell.2014.11.021.
8. Alipour, E., and Marko, J.F. (2012). Self-organization of domain structures by DNA-loop-extruding enzymes. Nucleic Acids Res. *40*, 11202–11212. https://doi.org/10.1093/nar/gks925.
9. Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N., and Mirny, L.A. (2016). Formation of Chromosomal Domains by Loop Extrusion. Cell Rep. *15*, 2038–2049. https://doi.org/10.1016/j.celrep.2016.04.085.
10. Sanborn, A.L., Rao, S.S.P., Huang, S.C., Durand, N.C., Huntley, M.H., Jewett, A.I., Bochkov, I.D., Chinnappan, D., Cutkosky, A., Li, J., et al. (2015). Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. Proc. Natl. Acad. Sci. USA *112*, E6456–E6465. https://doi.org/10.1073/pnas.1518552112.
11. Goloborodko, A., Marko, J.F., and Mirny, L.A. (2016). Chromosome Compaction by Active Loop Extrusion. Biophys. J. *110*, 2162–2168. https://doi.org/10.1016/j.bpj.2016.02.041.
12. de Wit, E., Vos, E.S.M., Holwerda, S.J.B., Valdes-Quezada, C., Verstegen, M.J.A.M., Teunissen, H., Splinter, E., Wijchers, P.J., Krijger, P.H.L., and de Laat, W. (2015). CTCF Binding Polarity Determines Chromatin Looping. Mol. Cell *60*, 676–684. https://doi.org/10.1016/j.molcel.2015.09.023.
13. Guo, Y., Xu, Q., Canzio, D., Shou, J., Li, J., Gorkin, D.U., Jung, I., Wu, H., Zhai, Y., Tang, Y., et al. (2015). CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. Cell *162*, 900–910. https://doi.org/10.1016/j.cell.2015.07.038.
14. Beagan, J.A., and Phillips-Cremins, J.E. (2020). On the existence and functionality of topologically associating domains. Nat. Genet. *52*, 8–16. https://doi.org/10.1038/s41588-019-0561-1.
15. Nora, E.P., Goloborodko, A., Valton, A.L., Gibcus, J.H., Uebersohn, A., Abdennur, N., Dekker, J., Mirny, L.A., and Bruneau, B.G. (2017). Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. Cell *169*, 930–944.e22. https://doi.org/10.1016/j.cell.2017.05.004.
16. Hsieh, T.H.S., Cattoglio, C., Slobodyanyuk, E., Hansen, A.S., Darzacq, X., and Tjian, R. (2022). Enhancer-promoter interactions and transcription are largely maintained upon acute loss of CTCF, cohesin, WAPL or YY1. Nat. Genet. *54*, 1919–1932. https://doi.org/10.1038/s41588-022-01223-8.
17. Rao, S.S.P., Huang, S.C., Glenn St Hilaire, B., Engreitz, J.M., Perez, E.M., Kieffer-Kwon, K.R., Sanborn, A.L., Johnstone, S.E., Bascom, G.D., Bochkov, I.D., et al. (2017). Cohesin Loss Eliminates All Loop Domains. Cell *171*, 305–320.e24. https://doi.org/10.1016/j.cell.2017.09.026.
18. Phillips-Cremins, J.E., Sauria, M.E.G., Sanyal, A., Gerasimova, T.I., Lajoie, B.R., Bell, J.S.K., Ong, C.T., Hookway, T.A., Guo, C., Sun, Y., et al. (2013). Architectural protein subclasses shape 3D organization of genomes during lineage commitment. Cell *153*, 1281–1295. https://doi.org/10.1016/j.cell.2013.04.053.
19. Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S., et al.

(2010). Mediator and cohesin connect gene expression and chromatin architecture. Nature *467*, 430–435. https://doi.org/10.1038/nature09380.

20. Beagan, J.A., Duong, M.T., Titus, K.R., Zhou, L., Cao, Z., Ma, J., Lachanski, C.V., Gillis, D.R., and Phillips-Cremins, J.E. (2017). YY1 and CTCF orchestrate a 3D chromatin looping switch during early neural lineage commitment. Genome Res. *27*, 1139–1152. https://doi.org/10.1101/gr.215160.116.
21. Weintraub, A.S., Li, C.H., Zamudio, A.V., Sigova, A.A., Hannett, N.M., Day, D.S., Abraham, B.J., Cohen, M.A., Nabet, B., Buckley, D.L., et al. (2017). YY1 Is a Structural Regulator of Enhancer-Promoter Loops. Cell *171*, 1573–1588.e28. https://doi.org/10.1016/j.cell.2017.11.008.
22. Ramasamy, S., Aljahani, A., Karpinska, M.A., Cao, T.B.N., Velychko, T., Cruz, J.N., Lidschreiber, M., and Oudelaar, A.M. (2023). The Mediator complex regulates enhancer-promoter interactions. Nat. Struct. Mol. Biol. *30*, 991–1000. https://doi.org/10.1038/s41594-023-01027-2.
23. El Khattabi, L., Zhao, H., Kalchschmidt, J., Young, N., Jung, S., Van Blerkom, P., Kieffer-Kwon, P., Kieffer-Kwon, K.R., Park, S., Wang, X., et al. (2019). A Pliable Mediator Acts as a Functional Rather Than an Architectural Bridge between Promoters and Enhancers. Cell *178*, 1145–1158.e20. https://doi.org/10.1016/j.cell.2019.07.011.
24. Levine, M., Cattoglio, C., and Tjian, R. (2014). Looping back to leap forward: transcription enters a new era. Cell *157*, 13–25. https://doi.org/10.1016/j.cell.2014.02.009.
25. Kwak, H., and Lis, J.T. (2013). Control of transcriptional elongation. Annu. Rev. Genet. *47*, 483–508. https://doi.org/10.1146/annurev-genet-110711-155440.
26. Chen, F.X., Smith, E.R., and Shilatifard, A. (2018). Born to run: control of transcription elongation by RNA polymerase II. Nat. Rev. Mol. Cell Biol. *19*, 464–478. https://doi.org/10.1038/s41580-018-0010-5.
27. Core, L., and Adelman, K. (2019). Promoter-proximal pausing of RNA polymerase II: a nexus of gene regulation. Genes Dev. *33*, 960–982. https://doi.org/10.1101/gad.325142.119.
28. Lis, J.T. (2019). A 50 year history of technologies that drove discovery in eukaryotic transcription regulation. Nat. Struct. Mol. Biol. *26*, 777–782. https://doi.org/10.1038/s41594-019-0288-9.
29. Roeder, R.G. (2019). 50+ years of eukaryotic transcription: an expanding universe of factors and mechanisms. Nat. Struct. Mol. Biol. *26*, 783–791. https://doi.org/10.1038/s41594-019-0287-x.
30. Kim, T.K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S., et al. (2010). Widespread transcription at neuronal activity-regulated enhancers. Nature *465*, 182–187. https://doi.org/10.1038/nature09033.
31. Barshad, G., Lewis, J.J., Chivu, A.G., Abuhashem, A., Krietenstein, N., Rice, E.J., Ma, Y., Wang, Z., Rando, O.J., Hadjantonakis, A.K., and Danko, C.G. (2023). RNA polymerase II dynamics shape enhancer-promoter interactions. Nat. Genet. *55*, 1370–1380. https://doi.org/10.1038/s41588-023-01442-7.
32. Zhang, S., Übelmesser, N., Barbieri, M., and Papantonis, A. (2023). Enhancer-promoter contact formation requires RNAPII and antagonizes loop extrusion. Nat. Genet. *55*, 832–840. https://doi.org/10.1038/s41588-023-01364-4.
33. Paquet, D., Kwart, D., Chen, A., Sproul, A., Jacob, S., Teo, S., Olsen, K.M., Gregg, A., Noggle, S., and Tessier-Lavigne, M. (2016). Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. Nature *533*, 125–129. https://doi.org/10.1038/nature17664.
34. Dannert, A., Klimmt, J., Cardoso Gonçalves, C., Crusius, D., and Paquet, D. (2023). Reproducible and scalable differentiation of highly pure cortical neurons from human induced pluripotent stem cells. STAR Protoc. *4*, 102266. https://doi.org/10.1016/j.xpro.2023.102266.
35. Gilgenast, T.G., and Phillips-Cremins, J.E. (2019). Systematic Evaluation of Statistical Methods for Identifying Looping Interactions in 5C Data. Cell Syst. *8*, 197–211.e13. https://doi.org/10.1016/j.cels.2019.02.006.
36. Fernandez, L.R., Gilgenast, T.G., and Phillips-Cremins, J.E. (2020). 3DeFDR: statistical methods for identifying cell type-specific looping interactions in 5C and Hi-C data. Genome Biol. *21*, 219. https://doi.org/10.1186/s13059-020-02061-9.
37. Emerson, D.J., Zhao, P.A., Cook, A.L., Barnett, R.J., Klein, K.N., Saulebekova, D., Ge, C., Zhou, L., Simandi, Z., Minsk, M.K., et al. (2022). Cohesin-mediated loop anchors confine the locations of human replication origins. Nature *606*, 812–819. https://doi.org/10.1038/s41586-022-04803-0.
38. Adelman, K., and Lis, J.T. (2012). Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. Nat. Rev. Genet. *13*, 720–731. https://doi.org/10.1038/nrg3293.
39. Zheng, B., Aoi, Y., Shah, A.P., Iwanaszko, M., Das, S., Rendleman, E.J., Zha, D., Khan, N., Smith, E.R., and Shilatifard, A. (2021). Acute perturbation strategies in interrogating RNA polymerase II elongation factor function in gene expression. Genes Dev. *35*, 273–285. https://doi.org/10.1101/gad.346106.120.
40. Hsieh, T.H.S., Cattoglio, C., Slobodyanyuk, E., Hansen, A.S., Rando, O.J., Tjian, R., and Darzacq, X. (2020). Resolving the 3D Landscape of Transcription-Linked Mammalian Chromatin Folding. Mol. Cell *78*, 539–553.e8. https://doi.org/10.1016/j.molcel.2020.03.002.
41. Wang, Y., Lu, J.-j., He, L., and Yu, Q. (2011). Triptolide (TPL) inhibits global transcription by inducing proteasome-dependent degradation of RNA polymerase II (Pol II). PLoS One *6*, e23993.
42. Chao, S.H., and Price, D.H. (2001). Flavopiridol inactivates P-TEFb and blocks most RNA polymerase II transcription in vivo. J. Biol. Chem. *276*, 31793–31799. https://doi.org/10.1074/jbc.M102306200.
43. Hansen, A.S., Pustova, I., Cattoglio, C., Tjian, R., and Darzacq, X. (2017). CTCF and cohesin regulate chromatin loop stability with distinct dynamics. Elife *6*, e25776.
44. Jiang, Y., Huang, J., Lun, K., Li, B., Zheng, H., Li, Y., Zhou, R., Duan, W., Wang, C., Feng, Y., et al. (2020). Genome-wide analyses of chromatin interactions after the loss of Pol I, Pol II, and Pol III. Genome Biol. *21*, 158. https://doi.org/10.1186/s13059-020-02067-3.
45. Heinz, S., Texari, L., Hayes, M.G.B., Urbanowski, M., Chang, M.W., Givarkes, N., Rialdi, A., White, K.M., Albrecht, R.A., Pache, L., et al. (2018). Transcription Elongation Can Affect Genome 3D Structure. Cell *174*, 1522–1536.e22. https://doi.org/10.1016/j.cell.2018.07.047.
46. Zhang, S., Ubelmesser, N., Josipovic, N., Forte, G., Slotman, J.A., Chiang, M., Gothe, H.J., Gusmao, E.G., Becker, C., Altmuller, J., et al. (2021). RNA polymerase II is required for spatial chromatin reorganization following exit from mitosis. Sci. Adv. *7*, eabg8205. https://doi.org/10.1126/sciadv.abg8205.
47. Mitchell, J.A., and Fraser, P. (2008). Transcription factories are nuclear subcompartments that remain in the absence of transcription. Genes Dev. *22*, 20–25. https://doi.org/10.1101/gad.454008.
48. Muniz, L., Nicolas, E., and Trouche, D. (2021). RNA polymerase II speed: a key player in controlling and adapting transcriptome composition. The EMBO journal *40*, e105740.
49. Tei, S., Sakata, T., Yoshimura, A., Natsume, T., Kanemaki, M.T., Bando, M., and Shirahige, K. (2024). Cohesin regulates promoter-proximal pausing of RNA Polymerase II by limiting recruitment of super elongation complex. bioRxiv. https://doi.org/10.1101/2024.03.15.584908.
50. Carter, D., Chakalova, L., Osborne, C.S., Dai, Y.F., and Fraser, P. (2002). Long-range chromatin regulatory interactions in vivo. Nat. Genet. *32*, 623–626. https://doi.org/10.1038/ng1051.
51. Andrey, G., Montavon, T., Mascrez, B., Gonzalez, F., Noordermeer, D., Leleu, M., Trono, D., Spitz, F., and Duboule, D. (2013). A switch between topological domains underlies HoxD genes collinearity in mouse limbs. Science *340*, 1234167. https://doi.org/10.1126/science.1234167.
52. Mifsud, B., Tavares-Cadete, F., Young, A.N., Sugar, R., Schoenfelder, S., Ferreira, L., Wingett, S.W., Andrews, S., Grey, W., Ewels, P.A., et al. (2015). Mapping long-range promoter contacts in human cells with

high-resolution capture Hi-C. Nat. Genet. *47*, 598–606. https://doi.org/10.1038/ng.3286.

53. Sanyal, A., Lajoie, B.R., Jain, G., and Dekker, J. (2012). The long-range interaction landscape of gene promoters. Nature *489*, 109–113. https://doi.org/10.1038/nature11279.
54. Brandao, H.B., Gabriele, M., and Hansen, A.S. (2021). Tracking and interpreting long-range chromatin interactions with super-resolution live-cell imaging. Curr. Opin. Cell Biol. *70*, 18–26. https://doi.org/10.1016/j.ceb.2020.11.002.
55. Alexander, J.M., Guan, J., Li, B., Maliskova, L., Song, M., Shen, Y., Huang, B., Lomvardas, S., and Weiner, O.D. (2019). Live-cell imaging reveals enhancer-dependent Sox2 transcription in the absence of enhancer proximity. Elife *8*, e41769. https://doi.org/10.7554/eLife.41769.
56. Benabdallah, N.S., Williamson, I., Illingworth, R.S., Kane, L., Boyle, S., Sengupta, D., Grimes, G.R., Therizols, P., and Bickmore, W.A. (2019). Decreased Enhancer-Promoter Proximity Accompanying Enhancer Activation. Mol. Cell *76*, 473–484.e7. https://doi.org/10.1016/j.molcel.2019.07.038.
57. Boija, A., Klein, I.A., Sabari, B.R., Dall'Agnese, A., Coffey, E.L., Zamudio, A.V., Li, C.H., Shrinivas, K., Manteiga, J.C., Hannett, N.M., et al. (2018). Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. Cell *175*, 1842–1855.e16. https://doi.org/10.1016/j.cell.2018.10.042.
58. Cho, W.K., Spille, J.H., Hecht, M., Lee, C., Li, C., Grube, V., and Cisse, I.I. (2018). Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. Science *361*, 412–415. https://doi.org/10.1126/science.aar4199.
59. Lu, H., Yu, D., Hansen, A.S., Ganguly, S., Liu, R., Heckert, A., Darzacq, X., and Zhou, Q. (2018). Phase-separation mechanism for C-terminal hyperphosphorylation of RNA polymerase II. Nature *558*, 318–323. https://doi.org/10.1038/s41586-018-0174-3.
60. Ramirez, F., Ryan, D.P., Gruning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dundar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. *44*, W160–W165. https://doi.org/10.1093/nar/gkw257.
61. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137. https://doi.org/10.1186/gb-2008-9-9-r137.
62. Langmead, B., Schatz, M.C., Lin, J., Pop, M., and Salzberg, S.L. (2009). Searching for SNPs with cloud computing. Genome Biol. *10*, R134. https://doi.org/10.1186/gb-2009-10-11-r134.
63. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842. https://doi.org/10.1093/bioinformatics/btq033.
64. Servant, N., Varoquaux, N., Lajoie, B.R., Viara, E., Chen, C.J., Vert, J.P., Heard, E., Dekker, J., and Barillot, E. (2015). HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. Genome Biol. *16*, 259. https://doi.org/10.1186/s13059-015-0831-x.
65. Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol. *34*, 525–527. https://doi.org/10.1038/nbt.3519.
66. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550. https://doi.org/10.1186/s13059-014-0550-8.
67. Soneson, C., Love, M.I., and Robinson, M.D. (2015). Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Res. *4*, 1521. https://doi.org/10.12688/f1000research.7563.2.
68. Liao, Y., Wang, J., Jaehnig, E.J., Shi, Z., and Zhang, B. (2019). WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. Nucleic Acids Res. *47*, W199–W205. https://doi.org/10.1093/nar/gkz401.
69. Marco-Sola, S., Sammeth, M., Guigó, R., and Ribeca, P. (2012). The GEM mapper: fast, accurate and versatile alignment by filtration. Nat. Methods *9*, 1185–1188. https://doi.org/10.1038/nmeth.2221.
70. Dominguez, C., Heras, J., and Pascual, V. (2017). IJ-OpenCV: Combining ImageJ and OpenCV for processing images in biomedicine. Comput. Biol. Med. *84*, 189–194. https://doi.org/10.1016/j.compbiomed.2017.03.027.
71. Roayaei Ardakany, A., Gezer, H.T., Lonardi, S., and Ay, F. (2020). Mustache: multi-scale detection of chromatin loops from Hi-C and Micro-C maps using scale-space representation. Genome Biol. *21*, 256. https://doi.org/10.1186/s13059-020-02167-0.
72. Abdennur, N., and Mirny, L.A. (2020). Cooler: scalable storage for Hi-C data and other genomically labeled arrays. Bioinformatics *36*, 311–316. https://doi.org/10.1093/bioinformatics/btz540.
73. Abdennur, N., Abraham, S., Fudenberg, G., Flyamer, I.M., Galitsyna, A.A., Goloborodko, A., Imakaev, M., Oksuz, B.A., and Venev, S.V. (2022). Cooltools: enabling high-resolution Hi-C analysis in Python. bioRxiv. https://doi.org/10.1101/2022.10.31.514564.
74. Zhang, H., Emerson, D.J., Gilgenast, T.G., Titus, K.R., Lan, Y., Huang, P., Zhang, D., Wang, H., Keller, C.A., Giardine, B., et al. (2019). Chromatin structure dynamics during the mitosis-to-G1 phase transition. Nature *576*, 158–162. https://doi.org/10.1038/s41586-019-1778-y.
75. Sun, J.H., Zhou, L., Emerson, D.J., Phyo, S.A., Titus, K.R., Gong, W., Gilgenast, T.G., Beagan, J.A., Davidson, B.L., Tassone, F., and Phillips-Cremins, J.E. (2018). Disease-Associated Short Tandem Repeats Co-localize with Chromatin Domain Boundaries. Cell *175*, 224–238.e15. https://doi.org/10.1016/j.cell.2018.08.005.
76. Norton, H.K., Emerson, D.J., Huang, H., Kim, J., Titus, K.R., Gu, S., Bassett, D.S., and Phillips-Cremins, J.E. (2018). Detecting hierarchical genome folding with network modularity. Nat. Methods *15*, 119–122. https://doi.org/10.1038/nmeth.4560.
77. Chandrashekar, H., Simandi, Z., Choi, H., Ryu, H.-S., Waldman, A.J., Nikish, A., Muppidi, S.S., Gong, W., Paquet, D., and Phillips-Cremins, J.E. (2024). A multi-looping chromatin signature predicts dysregulated gene expression in neurons with familial Alzheimer's disease mutations. bioRxiv. https://doi.org/10.1101/2024.02.27.582395.
78. Beagan, J.A., Gilgenast, T.G., Kim, J., Plona, Z., Norton, H.K., Hu, G., Hsu, S.C., Shields, E.J., Lyu, X., Apostolou, E., et al. (2016). Local Genome Topology Can Exhibit an Incompletely Rewired 3D-Folding State during Somatic Cell Reprogramming. Cell Stem Cell *18*, 611–624. https://doi.org/10.1016/j.stem.2016.04.004.
79. Phillips-Cremins, J.E., and Corces, V.G. (2013). Chromatin insulators: linking genome organization to cellular function. Mol. Cell *50*, 461–474. https://doi.org/10.1016/j.molcel.2013.04.018.
80. Beagan, J.A., Pastuzyn, E.D., Fernandez, L.R., Guo, M.H., Feng, K., Titus, K.R., Chandrashekar, H., Shepherd, J.D., and Phillips-Cremins, J.E. (2020). Three-dimensional genome restructuring across timescales of activity-induced neuronal gene expression. Nat. Neurosci. *23*, 707–717. https://doi.org/10.1038/s41593-020-0634-6.
81. Kim, J.H., Rege, M., Valeri, J., Dunagin, M.C., Metzger, A., Titus, K.R., Gilgenast, T.G., Gong, W., Beagan, J.A., Raj, A., and Phillips-Cremins, J.E. (2019). LADL: light-activated dynamic looping for endogenous gene expression control. Nat. Methods *16*, 633–639. https://doi.org/10.1038/s41592-019-0436-5.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
OCT4	Santa Cruz Biotechnology	Cat# SC9081, RRID:AB_2167703
SSEA4	R&D Systems	Cat# MAB1435, RRID:AB_357704
Nestin	R&D Systems	Cat# MAB1259, RRID:AB_2251304
FoxG1	Abcam	Cat# ab18259, RRID:AB_732415
CTIP2	Abcam	Cat# ab18465, RRID:AB_2064130
SATB2	Abcam	Cat# ab51502, RRID:AB_882455
Anti-Mouse IgG H&L Alexa Fluor 488	Abcam	Cat# ab150113, RRID:AB_2576208
Anti-Rat IgG H&L Alexa Fluor 647	Abcam	Cat# ab150159, RRID:AB_2566823
Goat Anti-Rabbit IgG H&L Alexa Fluor 594	Abcam	Cat# ab150080, RRID:AB_2650602
IgG from rabbit serum	Sigma-Aldrich	Cat# I8140, RRID:AB_1163661
CTCF	Millipore	Cat# 07–729, RRID:AB_441965
H3K27ac	Millipore	Cat# 07–449, RRID:AB_310624
YY1	Abcam	Cat# ab109237, RRID:AB_10890662
RNA pol II antibody (mAb) (Clone 4H8)	Active Motif	Cat# 39097, RRID:AB_2732926
Chemical, peptides, and recombinant proteins		
StemFlex culture media	Thermo Fisher Scientific	Cat # A3349401
Geltrex	Thermo Fisher Scientific	Cat # A1569601
Essential 6 media	Thermo Fisher Scientific	Cat # A1516401
Penicillin-streptomycin	Gibco	Cat# 15140122
Versene Solution	Gibco	Cat# 15040066
Accutase	Gibco	Cat# A1110501
Y-27632 (Dihydrochloride)	Stem Cell Technologies	Cat# 72304
Neurobasal media	Thermo Fisher Scientific	Cat# 21203-049
DMEM/F12 GlutaMax	Thermo Fisher Scientific	Cat# 10565-018
B27 supplement	Thermo Fisher Scientific	Cat# 17504-044
GlutaMax supplement	Thermo Fisher Scientific	Cat# 35050
Non-Essential Amino Acid	Thermo Fisher Scientific	Cat# 11140-050
N-2 supplement	Thermo Fisher Scientific	Cat# 17502048
Insulin	Sigma	Cat# I1882
2-mercapto-ethanol	Thermo Fisher Scientific	Cat# 21985-023
SB431542	Selleckchem	Cat# S1067
LDN-193189	Selleckchem	Cat# S2618
poly-L-ornithine	Sigma-Aldrich	Cat# P4957
Laminin	Thermo Fisher Scientific	Cat# 23017015
FGF-2	Stem Cell Technologies	Cat# 78003
Uridine	Sigma	Cat# U3750-1G
STEMdiff Neural Rosette Selection Reagent	Stem Cell Technologies	Cat# 5832
DAPT	Sigma-Aldrich	Cat# D5942
PBS	Corning	Cat# 21-040-CV
100% Ethanol	Decon Labs	Cat# 2716
EDTA, pH 8.0	Invitrogen	Cat# 15575020
Synth-a-Freeze	Gibco	Cat# A1254201
Formaldehyde solution	Sigma-Aldrich	Cat# F8775
Formaldehyde solution	Pierce	Cat# 28908
HEPES-KOH, pH 7.5	Boston BioProducts	Cat# BBH-75-K

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Igepal CA-630	Sigma-Aldrich	Cat# I8896
5M NaCl	Invitrogen	Cat# AM9760G
Protease Inhibitor	Sigma	Cat# P8340
SDS solution, 10%	Fisher Scientific	Cat# 15553027
TWEEN 20	Sigma	P9416-50ML
Sodium bicarbonate	Sigma-Aldrich	Cat# S5761
Sodium deoxycholate	Sigma-Aldrich	Cat# D6750
TE buffer, pH 8.0	Invitrogen	Cat# AM9858
Lithium chloride, ultra dry	Alfa Aesar	Cat# 1368406
Magnesium chloride solution	Sigma	Cat# M1028-100ML
Tris-HCl, pH 8.0	Invitrogen	Cat# 15568025
Triton X-100 solution	Sigma-Aldrich	Cat# 93443
TE buffer, pH 8.0	Invitrogen	Cat# AM9858
Nuclease-free water	Sigma-Aldrich	Cat# W4502
VECTASHIELD Antifade Mounting Medium	Vector Laboratories	Cat# H-1200
Protein A Agarose beads	Invitrogen	Cat# 15918014
Protein G Agarose beads	Invitrogen	Cat# 15920010
Agencourt AMPure XP beads	Beckman Coulter	Cat# A63881
RNaseA	Roche	Cat# 10109169001
Proteinase K	NEB	Cat# P8107S
PMSF solution	Sigma	Cat# 93482-50ML-F
Nuclease-free water	Sigma-Aldrich	Cat# W4502
Ultrapure Phenol/Chloroform/Isoamyl Alcohol	Fisher Scientific	Cat# BP1752I100
Critical commercial assays		
Human Karyotype Panel	NanoString	Cat# XT-CSO-KAR15-012
Mycoplasma detection kit	ATCC	Cat# 30-1012K
Direct-zol total RNA isolation kit	Zymo	Cat# R2061
Qubit RNA HS assay	Invitrogen	Cat# Q32852
RNA 6000 Pico reagent kit	Agilent Technologies	Cat# 5067-1511
TruSeq Stranded Total RNA Library Prep Human/Mouse/Rat (48 Samples) (gold)	Illumina	Cat# 20020598
TruSeq RNA Single Indexes Set A and B	Illumina	Cat# 20020492
Qubit dsDNA HS assay kit	Invitrogen	Cat# Q32851
Agilent DNA 1000 reagent kit	Agilent Technologies	Cat# 5067–1504
Arima-HiC kit	Arima Genomics	Cat# A510008
Kapa Library Quantification Kit	KAPA Biosystems	Cat# KK4835
NEBNext Ultra II DNA Library Prep Kit for Illumina	New England Biolabs	Cat# E7645S
NextSeq (150-cycle) v2.5 High Output kit	Illumina, Inc	Cat# 20024907
NextSeq 500 High Output v2 Kit (75 cycles) kit	Illumina, Inc	Cat# 20024906
Deposited data		
RNAPolII ChIP-seq in hiPSC, hiPSC-NPC, and NPC-Neuron	This study	GEO: GSE220103
CTCF ChIP-seq in hiPSC, hiPSC-NPC, and NPC-Neuron	This study	GEO: GSE220103
H3K27ac ChIP-seq in hiPSC, hiPSC-NPC, and NPC-Neuron	This study	GEO: GSE220103
Hi-C in hiPSC, hiPSC-NPC, and NPC-Neuron	This study	GEO: GSE220103
RNA-seq in hiPSC, hiPSC- NPC, and NPC-Neuron	This study	GEO: GSE220103
Experimental models: cell lines		
7889SA3.5	Paquet et al.[33]	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
ImageJ	NIH	https://imagej.nih.gov/ij/
Adobe Photoshop (v 23.2.2)	Adobe	https://www.adobe.com/
Leica LAS X	Leica	https://www.leica-microsystems.com/products/m icroscope-software/p/leica-las-x-ls/downloads/
deeptools (v 3.5.1)	Ramirez et al.[60]	https://deeptools.readthedocs.io/en/develop/
MACS2 (v 2.1.1.20160309)	Zhang et al.[61]	https://pypi.org/project/MACS2/
Bowtie (v 0.12.7)	Langmead et al.[62]	http://bowtie-bio.sourceforge.net/index.shtml
Bedtools (v 2.29.1)	Quinlan and Hall[63]	https://github.com/arq5x/bedtools2
HiC-Pro (v 2.7.7) and (v 3.1.0)	Servant et al.[64]	https://github.com/nservant/HiC-Pro
Kallisto (v 0.46.1)	Bray et al.[65]	https://pachterlab.github.io/kallisto/about
DESEQ2 (v 1.22.1)	Love et al.[66]	https://doi.org/10.1186/s13059-014-0550-8
tximport	Soneson et al.[67]	https://bioconductor.org/packages/release/bioc/html/tximport.html
WebGestalt (v 0.4.4)	Liao et al.[68]	https://github.com/bzhanglab/WebGestaltR
GEM-mappability (Binary pre-release 2)	Marco-Sola et al.[69]	https://sourceforge.net/projects/gemlibrary/files/gem-library/
3DeFDR-HiC (v 0.2.1)	Fernandez et al.[36]	https://bitbucket.org/creminslab/hic3defdr
pyBigWig (v 0.3.18)	Ramirez et al.[60]	https://github.com/deeptools/pyBigWig
opencv-python (v 4.5.4.58)	Dominguez et al.[70]	https://docs.opencv.org/4.x/d6/d00/tutorial_py_root.html
Mustache (v 1.3.1)	Roayaei et al.[71]	https://github.com/ay-lab/mustache
Cooler (v 0.9.2)	Abdennur et al.[72]	https://github.com/open2c/cooler
Cooltools (v 0.5.4)	Open2C[73]	https://github.com/open2c/cooltools

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources, reagents, or other materials should be directed to and will be addressed by the lead contact, Dr. Jennifer E. Phillips-Cremins (jcremins@seas.upenn.edu).

Materials availability

All unique reagents generated in this study are available from the lead contact with a completed Materials Transfer Agreement upon reasonable request.

Data and code availability

- Raw sequencing files and key intermediate files generated in this study have been deposited in the Gene Expression Omnibus GEO: GSE220103.
- All code is provided freely at the below zenodo link: https://doi.org/10.5281/zenodo.11106963. https://zenodo.org/records/11106963.

EXPERIMENTAL MODELS AND SUBJECT DETAILS

Maintenance of human induced pluripotent stem cell (human iPSC) culture

A subclone of the previously described 7889SA human induced pluripotent stem cell (hiPSC) line was used for each experiment. The line was established from AG07889 male fibroblasts (Coriell Institute for Medical Research, Camden, NJ) and previously characterized as described.[33] Upon arrival, the subclone (7889SA3.5) was expanded and a master stock of cells were frozen from the early passage. The same master stocks were thawed, cultured, and subsequently used for each experiment.

hiPSCs were cultured on 10 cm cell culture plastic dishes (Corning, #430167) and pre-coated with 6 mL Geltrex hESC-Qualified Reduced Growth Factor Basement Membrane Matrix (Thermo Fisher Scientific, #A1569601) for 1 h at 37°C hiPSCs were plated and maintained in StemFlex stem cell culture media (Thermo Fisher Scientific, #A3349401) supplemented with 1% (v/v) penicillin-streptomycin (Thermo Fisher Scientific, #15140122). Cultures were incubated in a humidified incubator at 37°C and 5% CO_2. To passage,

cells were detached from dishes by washing with 1xPBS and then incubating with 4 mL of Versene Solution (Thermo Fisher Scientific, 15040066) at 37°C for 3–5 min. Following inactivation by the addition of 10 mL of fresh cell culture media, cells were split in a 1:3 to 1:8 dilution and seeded onto Geltrex-coated plates, prepared freshly on the same day.

To verify the cellular state of the hiPSC clones, an early passage stock vial was characterized for (1) normal karyotype, (2) characteristic stem cell morphology, (3) expression of genes indicating pluripotent state, and (4) absence of mycoplasma. A human Karyotype Panel (NanoString, XT-CSO-KAR15-012) was used to assess karyotype from genomic DNA isolated from the master stock. using phenol:chloroform and ethanol precipitation, DNA was extracted from the cells and submitted for NanoString service provided by the Genomics Facility at The Wistar Institute, Philadelphia, PA. Immunofluorescence staining was used to characterize the presence of pluripotency markers, we used for OCT4 (Santa Cruz, SC9081, 1:100 dilution) and SSEA4 (R&D, MAB1435, 10 μg/mL) as described in detail under immunocytochemistry and microscopy. 10^5 hiPSCs were collected into the existing culture media and processed with a mycoplasma detection kit (ATCC, 30-1012K) according to manufacturer's protocol, to confirm the absence of mycoplasma contamination. Following the initial characterization of the master stock, cultures were continuously monitored for expected pluripotent cell morphology by daily visual assessment under the microscope and routine staining for OCT4 and SSEA4.

hiPSC differentiation to neural progenitors (NPCs) and post-mitotic neurons

NPCs and post-mitotic neurons were differentiated from hiPSCs using our recently described cortical neuron differentiation protocol with a few additional optimizations.[34] Briefly, 24 h prior]differentiation, StemFlex media was replaced with Essential 6 (E6) media (Thermo Fisher Scientific, A1516401) on hiPSCs growing on 10 cm cell culture dishes. Induction of neural precursor cell differentiation was initiated by detaching hiPSCs with 4mL Accutase (Thermo Fisher Scientific, #A1110501) followed by Accutase inactivation and single cell suspension with 16 mL of E6 media. Cells were centrifuged, resuspended in E6 media supplemented with ROCK inhibitor (Stem Cell Technologies, #72304) and seeded into Geltrex-coated 12-well tissue culture plates (Falcon, #353043), Medium was replaced to 2 mL neural induction (NI) medium supplemented with LDN-193189 (Selleckchem, #S2618, 1:10000) and SB431542 (Selleckchem, #S1067, 1:1000). The day of change to NI media was counted as day *in vitro* 0 (DIV0). Cells were maintained in NI media for 8 days (DIV8) where media was changed every day. To differentiate hiPSC derived neural precursor cells (hiPSC-NPCs), DIV8 cells were enzymatically dissociated using Accutase and pelleted cells were resuspended at 30 million cells per mL in NI medium that was supplemented with ROCK inhibitor. Cells were plated as 250-μL spots onto dry 6-well plates there were pre-coated with poly-L-ornithine (Sigma-Aldrich, #P4957) and laminin (Thermo Fisher Scientific, #23017015, 20x dilution), and 2 mL NI medium supplemented with ROCK inhibitor were slowly added to avoid detachment of the cells. Media was daily changed until DIV10. On DIV10 NI media was replaced with neural maintenance (NM) media. 20 ng mL FGF-2 (Stem Cell Technologies Inc., #78003) were added for 2 days as soon as neural rosettes were apparent in the cultures. When cells with neuronal morphology started to emerge from rosettes, rosettes were manually isolated under the microscope under sterile conditions and cultivated in NM media until DIV35 when NPC samples were harvested.

To further differentiate cells into terminally differentiated, post-mitotic neurons (NPC-Neurons), NPC rosettes were triturated into single cells using Accutase, followed by inactivation with fresh Neurobasal medium supplemented with B-27 serum-free supplement (Thermo Fisher Scientific, #17504044), 2 mM GlutaMax (Thermo Fisher Scientific, #35050061), 1% (v/v) Penicillin-Streptomycin (Thermo Fisher Scientific, #15140–122)). Cells were seeded at a ~200,000–500,000 hiPSC-NPCs per well density onto a 24-well poly-L-ornithine/laminin-coated plate. In the first 7 days after plating, 10 μM DAPT (Sigma-Aldrich, #D5942) and 5-FU was added to the media to eliminate proliferating cells and enhancing terminal differentiation. Neurons were harvested at DIV65. Homogeneous morphology was confirmed, and presence of neuronal markers were assessed as described in detail under immunocytochemistry and microscopy.

METHOD DETAILS

Many of the methods listed below have already been described extensively in our previous manuscripts.[37,74–77] We state the same or near similar methodological steps below to ensure reproducibility and clarity.

Immunocytochemistry and microscopy

Stem cell markers (*OCT4* and *SSEA4*), neural precursor cell markers (*NESTIN, FOXG1*), and mature neuronal markers (*CTIP2, SATB2*) were assessed by immunofluorescence staining at Day 0, DIV35 (hiPSC-NPC), and DIV65 (NPC-Neuron). Briefly, cells were fixed in 4% paraformaldehyde (Pierce, cat# 28908), permeabilized in PBS/0.1% Triton X-100, and stained with primary antibodies anti-OCT4 (Santa Cruz, #SC9081, 1:100 dilution), anti-SSEA4 (R&D, #MAB1435, 10 μg/mL), anti-FoxG1 (Abcam, #ab18259), anti-Nestin (R&D System, #MAB1259), anti-SATB2 (Abcam, #ab51502) and anti-CTIP2 (Abcam, #ab18465). Secondary antibodies Goat Anti-Mouse IgG H&L Alexa Fluor 488 (Abcam, #ab150113), Goat Anti-Rat IgG H&L Alexa Fluor 647 (Abcam, #ab150159), and Goat Anti-Rabbit IgG H&L Alexa Fluor 594 (Abcam, #ab150080) were used for visualization. Cells were imaged on a Leica DMi8 microscope using software Leica LAS X and images were processed in ImageJ and Adobe Photoshop (Adobe). Image processing parameters were limited to brightness and contrast levels. These parameters were set equivalently across all images to obtain higher quality phase-contrast images and ensure that secondary antibody only controls did not show signal.

Total RNA-seq

Total RNA was extracted from 3 biological replicates of hiPSCs, NPCs, and neurons using the Direct-zol total RNA isolation kit (Zymo, #R2061) according to the manufacturer's protocol. The integrity of the RNA samples (RIN) were assessed using Agilent RNA 6000 Pico reagent kit on the Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA). 500 ng of isolated total RNA with a RIN value above 8 was used for ribosomal RNA depleted strand-specific RNA library preparation using the TruSeq Stranded Total RNA LT sample preparation kit with Ribo-Zero Gold (Illumina, #RS-122-2301). TruSeq RNA Single Indexes Set A and B were used and ligated onto cDNA (Illumina, #20020492) to enable multiplex sequencing. Library clean-up and size selection of ~300 bp fragments was performed by Agencourt AMPure XP beads (Beckman Coulter, #A63881) prior to PCR amplification and index incorporation for 15 cycles. Library quality and quantities were assessed using the Qubit high sensitivity DNA assay kit (Thermo Fisher Scientific, Q32851) and the Agilent DNA 1000 reagent kit (Agilent Technologies, 5067–1504) on the Agilent Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA). Sequencing of multiplexed, pooled samples were performed on Illumina NextSeq500 using a 150-cycle v2.5 High Output kit (Illumina, Cat# 20024907), generating 2 × 75bp, paired-end reads.

Chromatin fixation for ChIP-seq and Hi-C

Cells were fixed for downstream ChIP-seq and Hi-C assays as previously described.[20,36,37,74,75,78,79] Briefly, media was discarded, cells were washed with 6 mL of 1x PBS and 10 mL of freshly prepared fixation media (1% (v/v) formaldehyde in DMEM/F-12 (Thermo Fisher Scientific, #11320033)) was added to the cells growing on 10 cm cell culture dishes and incubated for 10 min at room temperature. Fixation media was diluted from 11% formaldehyde solution (50 mM HEPES-KOH (pH 7.5), 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 37% formaldehyde (Sigma, Cat# F8775). The fixation process was terminated by adding 125 mM glycine for 5 min at room temperature, followed by 15 min incubation at 4°C. Cross-linked cells were washed with pre-chilled PBS twice before flash freezing in liquid nitrogen. Pellets of cells were stored at −70°C.

ChIP-seq

ChIP-seq libraries were prepared as previously described,[20,36,37,74,75,78,79] with minor modifications. Briefly, crosslinked cell pellets (8–10 million cells) were lysed in pre-chilled cell lysis buffer (10 mM Tris pH 8.0, 10 mM NaCl, 0.2% (v/v) NP-40, 1 mM PMSF, 0.2% (v/v) Protease inhibitor cocktail (Sigma, #P8340)) on ice for 10 min. Using a Dounce homogenizer and 25–30 strokes with pestle A, cells were homogenized. To pellet and solubilize nuclei, lysates were centrifuged at 2,500 x g for 5 min at 4°C and pellets were resuspended in 500 μL nuclear lysis buffer (50 mM Tris pH 8.0, 10 mM EDTA, 1% (w/v) SDS, 1 mM PMSF, 0.2% (v/v) Protease inhibitor cocktail) on ice for 20 min before mixing with 300 μL IP dilution buffer (20 mM Tris pH 8.0, 2 mM EDTA, 150 mM NaCl, 1% (v/v) Triton X-100, 0.01% (w/v) SDS, 1 mM PMSF, 0.2% Protease inhibitor cocktail). Chromatin was sheared to ~200–600 bp DNA fragment size using Qsonica Q800R3 (Qsonica Sonicators, CT) with parameters of 100% amplitude and 30 s on/30 s off pulses. Fragment size was confirmed on 1% agarose gel from purified input samples. After sonication lysates were centrifuged at 16,000 x g and incubated for 2 h with pre-clearing solution consisting 50 μg of IgG (Sigma, #I8140), 175 μL of Protein A Agarose (Thermo Fisher Scientific, #15918014), and 175 μL of Protein G Agarose (Thermo Fisher Scientific, #15920010) in 3.7 mL of pre-chilled IP dilution buffer and 0.5 mL of nuclear lysis buffer by rotating at 10 rpm at 4°C. Chromatin immunoprecipitation was performed from supernatant using antibody-bound beads by rotating the samples at 10 rpm overnight (~12 h) at 4°C. The following antibodies were used for ChIP: anti-CTCF (Millipore, Cat# 07–729), anti-H3K27ac (Millipore, Cat# 07–449), anti-YY1 (Abcam, Cat# ab109237), anti-RNAPolII (Active Motif, Cat# 39097). 10 μg of antibody was used for each 8–10 million cell pellets.

Beads were washed once in IP wash buffer 1 (20 mM Tris pH 8.0, 2 mM EDTA, 50 mM NaCl, 1% (v/v) Triton X-100, 0.1% (w/v) SDS), twice in high-salt buffer (20 mM Tris pH 8.0, 2 mM EDTA, 500 mM NaCl, 1% (v/v) Triton X-100, 0.01% (w/v) SDS), once in IP wash buffer 2 (10 mM Tris pH 8.0, 1 mM EDTA, 250 mM lithium chloride, 1% (v/v) NP-40, 1% (w/v/) sodium deoxycholate), and twice in 1 x TE. All wash buffers were pre-chilled to ~4°C. Chromatin was eluted from beads in elution buffer (1% (w/v) SDS, 100 mM sodium bicarbonate), followed by RNA degradation by adding RNaseA (Roche, #10109169001) at 65°C for 1 h. Residual proteins were removed and reverse crosslinking of DNA was induced by adding 2.4 U of Proteinase K (NEB, #P8107S) to eluent and incubation at 65°C overnight. The ChIP DNA was purified and extracted using conventional phenol:chloroform extraction and ethanol precipitation methods. ChIP DNA was stored at −20°C until library preparation. 5 ng of purified ChIP-seq DNA was used for downstream library preparation as described under library preparation (ChIP-seq, Hi-C).

Hi-C

Using the Arima Hi-C kit (Arima Genomics, Inc., #A510008) according to the manufacturer's protocol, Hi-C was performed on ~2 million cross-linked cells per replicate per condition. The 5′-overhangs were filled in to label the digested ends with a biotinylated nucleotide following restriction digest of the chromatin with multiple enzymes. Spatially proximal digested ends of DNA were ligated, purified, and sheared to an average size of ~400 bp using a Covaris S220 sonicator. The sonicator was used at 140 W peak incident power, 10% duty factor, and 200 cycles per burst for 55 s. Size-selection and bead purification was performed to obtain 200-600 bp DNA fragments using DNA Purification Beads (AMPure XP Beads, Beckman Coulter, #A63881). Biotin-tagged ligation junctions were enriched after size-selection using Enrichment Beads provided in Arima-Hi-C kit. Streptavidin beads containing enriched DNA fragments were kept up to 3 days at −20°C, before proceeding with library preparation as described under library preparation (ChIP-seq, Hi-C).

Library preparation (ChIP-seq, Hi-C)

Sequencing libraries were prepared using the NEB Next Ultra II Library Prep Kit (NEB, #E7645S) following the manufacturer's protocol with minor modifications. For ChIP and Hi-C experiments, end-repair and dA-tailing of DNA was carried out according to the manufacturer's protocol. Adaptor-ligated Hi-C libraries were washed on streptavidin beads twice in 150 μL of wash buffer at 55°C and once in 100 μL of elution buffer (Arima Genomics, #A510008). Ligation products were eluted from streptavidin beads by boiling at 98°C for 10 min in 15 μL elution buffer. Size-selection of adaptor-ligated ChIP-seq libraries were performed using AgenCourt Ampure XP beads (Beckman Coulter, #A63881). Library amplification was carried out using NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB, #E7645S) according to the manufacturer's protocol. For ChIP-seq, <1 kb size DNA fragments were size-selected and amplified using 7–8 PCR cycles. After additional purification using AgenCourt Ampure XP beads (Beckman Coulter, #A63881), the quality of the individual libraries was assessed using Agilent Bioanalyzer High Sensitivity DNA Analysis Kits (Agilent, #5067–4626). DNA concentration was quantified using a Kapa Library Quantification Kit (KAPA Biosytems, #KK4835). Libraries were pooled and sequenced on an Illumina NextSeq 500 instrument using 75-cycle v2.5 High Output kit (Illumina, Cat# 20024906) for 75 bp single-end reads for ChIP-seq or 37 bp pair-end for Hi-C.

hg38 RefSeq reference transcriptome

The hg38 reference transcriptome was downloaded from the UCSC Table Browser on July 7th, 2021. NCBI RefSeq and RefSeq Curated were used as track and table options respectively (https://genome.ucsc.edu/cgi-bin/hgTables?hgsid=2124536004_l92s7VLx7DSbALl8aT7Yh5FE7R3X). Genes that (i) are in canonical chromosomes 1–22 and chromosome X, (ii) are not present in multiple chromosomes, (iii) have a gene length greater than 10 kb (iv) are not intersecting the GRCh38 blacklist (https://www.encodeproject.org/files/ENCFF356LFX/), centromeres, or telomeres were analyzed.

RNA-seq analysis

RNA-seq paired-end reads corresponding to the hg38 RefSeq reference transcriptome were pseudoaligned using kallisto quant with 100 bootstraps of gene quantification.[65] In R, stimulated counts were converted into DESeq2 format using the library ("tximportData") according to recommendations from the DESeq2 documentation.[66] Normalized counts were computed for each replicate and each cellular stage using DESeq2. Genes with the same transcription start sites (TSSs) and transcription end sites (TESs) were merged into transcriptional units and the counts were summed together. The final reference transcriptome had 27,030 transcriptional units.

Hi-C pre-processing

Paired end reads with a read length of 37 bp were aligned to the hg38 genome using bowtie2. Through the HiC-Pro software (v 2.7.7), the following default global parameters were used: –very-sensitive –L 30 –score-min L,-0.6,-0.2 –end-to-end–reorder and the following local parameters:–very-sensitive –L 20 –scoremin L,-0.6,-0.2 –end-to-end–reorder. Reads that were unmapped, non-uniquely mapped, and PCR duplicates were removed. Unique valid pairs were binned into uniform 10kb bins to generate *cis*-contact matrices and then merged across replicates.

To normalize for sequencing depth, samples were normalized using size factors conditioned on genomic distance from the diagonal. A size factor was computed for each sample by summing the counts of bin-bin pairs that are the same genomic distance and then dividing by the geometric mean. Then, the normalized matrices were rounded to the nearest whole number for downstream loop calling. Poorly mapped regions were removed from the normalized matrices based on an hg38 36-mer alignability track generated using the GEM-mappability alignment software. Pixels were masked if a 50kb window centered on that pixel had a mean mappability below 50%. An additional filter that removes high outliers was implemented to prevent balancing artifacts. Pixels that exhibited high fold changes (>4) relative to the median value of a neighborhood defined by a 5 × 5 footprint were removed. If a row had less than 35 non-zero pixels within 750 kb of the diagonal, it was masked. After filtering, the Knight-Ruiz matrix balancing algorithm was implemented and the final bias factors were retained.

Hi-C expect0065d modeling

With several modifications, an expected modeling strategy from our previous work and work of others was implemented as described in detail in our previously published works (https://bitbucket.org/creminslab/cremins_lab_loop_calling_pipeline_11_6_2021/src/initial/).[20,35–37,78–81]

The analysis was restricted to pixels with interaction distances within 10 Mb from Knight Ruiz balanced 10kb contact matrices. First, to account for the distance dependence of Hi-C counts, a one-dimensional expected model, D, was computed by determining the geometric mean of interaction counts for each of the first 1000 diagonals spanning 10 Mb (Equation 1):

$$D_d = (S_{a,b}) \quad \forall\ d\ such\ that\ 0 \leq d \leq 1000 \qquad \text{(Equation 1)}$$

where D_d is the expected value for interactions between pixels (a, b) separated by d bins and S is the balanced contact matrix with a pseudocount of 1 added to each count. Then to correct the one-dimensional expected model for local domain structures, each expected value $D_{i,j}$ was multiplied by five separate correction factors (Equations 2, 4, 6, 8, and 10). These correction factors were computed by summing all pixels (a, b) that fall within a geometric footprint centered on pixel i, j in the balanced matrix S and the

one-dimensional expected matrix D then finding the ratio between the two sums. The five geometric footprints are the donut footprint (Equations 2 and 3),

$$E_{i,j}^{DF} = D_{i,j} \times \frac{\sum_{(a,b) \in DF_{i,j}} S_{a,b}}{\sum_{(a,b) \in DF_{i,j}} D_{b-a}} \quad \text{(Equation 2)}$$

$$DF_{i,j} = \{(a,b) | (|a - i| \leq w) \wedge (|b - i| \leq w) \wedge (a \neq i) \wedge (b \neq j) \wedge ((|a - i| > p) \vee (|b - j| > p))\} \quad \text{(Equation 3)}$$

the lower left footprint (Equations 4 and 5).

$$E_{i,j}^{LLF} = D_{i,j} \times \frac{\sum_{(a,b) \in LLF_{i,j}} S_{a,b}}{\sum_{(a,b) \in LLF_{i,j}} D_{b-a}} \quad \text{(Equation 4)}$$

$$LLF_{i,j} = \{(a,b) \in DF_{i,j} | (a < i) \wedge (b < j)\} \quad \text{(Equation 5)}$$

vertical footprint (Equations 6 and 7).

$$E_{i,j}^{VF} = D_{I,j} \times \frac{\sum_{(a,b) \in VF_{i,j}} S_{a,b}}{\sum_{(a,b) \in VF_{i,j}} D_{b-a}} \quad \text{(Equation 6)}$$

$$VF_{i,j} = \{(a,b) | ((b = j - 1) \vee (b = j) \vee (b = j + 1)) \wedge (|a - i| > p) \wedge (|a - i| \leq w)\} \quad \text{(Equation 7)}$$

horizontal footprint (Equations 8 and 9).

$$E_{i,j}^{HF} = D_{i,j} \times \frac{\sum_{(a,b) \in HF_{i,j}} S_{a,b}}{\sum_{(a,b) \in HF_{i,j}} D_{b-a}} \quad \text{(Equation 8)}$$

$$HF_{i,j} = \{(a,b) \mid ((a = i - 1) \vee (a = i) \vee (a = i + 1)) \wedge (|b - j| > p) \wedge (|b - j| \leq w)\} \quad \text{(Equation 9)}$$

and the upper triangle footprint (Equations 10, 11, 12, 13, 14, 15, 16, and 17).

$$E_{i,j}^{UTF} = D_{i,j} \times \frac{\sum_{(a,b) \in UTF_{i,j}} S_{a,b}}{\sum_{(a,b) \in UTF_{i,j}} D_{b-a}} \quad \text{(Equation 10)}$$

$$UTF_{i,j} = \{(a,b) \in DF_{i,j} | b - a \geq j - i\} \quad \text{(Equation 11)}$$

respectively. All geometric footprints were parameterized by $p = 4$ and $w = 10$. An additional parameter was implemented where the minimum fraction of finite counts within a footprint must be greater than 0.2 for an expected value to be computed. The final expected value of pixel i,j was computed by finding $E_{i,j}^{UTF}$ when the interaction distance was less than 400kb. The final expected value was computed by finding the maximum expected value across the donut, lower left, vertical, and horizontal footprints when the interaction distance was greater than 400kb but less than 10 Mb (Equation 12).

$$E_{i,j} = \begin{cases} E_{i,j}^{UTF}, for\ b - a \leq 40 \\ max\left(E_{i,j}^{DF}, E_{i,j}^{LLF}, E_{i,j}^{HF}, E_{i,j}^{VF}\right), for\ 40 < b - a < 1000 \end{cases} \quad \text{(Equation 12)}$$

Hi-C p Values

To compute an integer biased expected value for comparison with the integer sequencing depth normalized count $X_{i,j}$, the final expected value $E_{i,j}$ and the bias vector c from Knight-Ruiz balancing was used. (Equation 13):

$$E_{i,j}^{biased} = E_{i,j} \times c_i \times c_j \quad \text{(Equation 13)}$$

A p-value $P_{i,j}$ was computed by testing the null hypothesis that $X_{i,j}$ was less than or equal to a Poisson-distributed random variable $X'_{i,j}$ with mean $E_{i,j}^{biased}$ (Equation 14):

$$P_{i,j} = P\left(X_{i,j} \leq X'_{i,j}\right); X'_{i,j} \sim Poisson\left(E_{i,j}^{biased}\right) \quad \text{(Equation 14)}$$

Hi-C multiple testing correction

To perform multiple testing correction, the lambda-chunking method from Aiden[7] was applied. First, we stratified pixels (i,j) by their biased expected values $E_{i,j}^{biased}$ using bins with a bin spacing $2^{1/3}$. After stratification by biased expected values, Benjamini-Hochberg false discovery rate control was performed on each chunk separately to obtain a q-value, $Q_{i,j}$. This q-value represents the significance of the maximum false discovery rate (FDR).

Hi-C loop clustering

After computing $Q_{i,j}$, clusters of nearby significant pixels were identified. First, an initial set of significant pixels was identified using: (1) a q-value $Q_{i,j} \leq 0.05$, a balanced contact value $S_{i,j} \geq 6$ and an observed over expected fold-change $FC \geq 1.5$. Clusters with fewer than three significant pixels were removed to reduce false positives. "Superclusters" composed of smaller clusters that were more likely to be true interactions, were found in the initial calls. Therefore, superclusters were refined by applying increasingly more stringent q-value thresholds on the order from 0.05 to 5e−6 FDR. If a cluster became smaller when the threshold was tightened, the smaller, more refined cluster was kept. If a cluster was lost entirely when the threshold was tightened, the cluster right before it was lost was kept. Finally, to avoid calling loops near the diagonal of the contact matrix, all clusters containing at least 1 pixel within 3 bins of diagonal were removed.

Hi-C differential loop calling

To call differential loops between conditions, 3DeFDR-HiC was used.[36] In brief, 3DeFDR-HiC performs a negative binomial likelihood ratio test for every pixel engaged in loops genome-wide. The negative binomial model was defined by several parameters. These parameters are (i) the average count per pixel across replicates for the two cell stages being compared (ii) a distance-dependent scaling factor and (iii) the Knight-Ruiz balancing bias factor, and (iv) an estimated distance dispersion per pixel across replicates for the two cell stages being compared. To obtain p-values against the null hypothesis that each looping pixel was not part of a differential loop a likelihood ratio test was performed. The Benjamini-Hochberg procedure was applied to correct these p-values an FDR threshold of 30%. Once differential pixels were identified, they were re-clustered back into their original clusters and compared for occupancy. If a cluster had more than 75% of its pixels as condition specific, then that cluster was classified as condition specific. Otherwise, the cluster was classified as invariant. To visually confirm differential loop calls, aggregate peak analysis was performed based on the mean Hi-C signal of a +/− 130 kb region centered on the differential loop calls.

ChIP-seq analysis

Using Bowtie, RNA Polymerase II (RNAPolII), CTCF, and H3K27ac ChIP-seq single-end reads were aligned to the hg38 genome. Reads with more than two alignments were removed. To account for sequencing depth differences, all immunoprecipitation libraries were downsampled to 24 million reads. Input libraries were also downsampled to 24 million reads. MACS2 callpeak was used to call peaks with a p-value cutoff of 1×10^{-4} for RNAPolII, CTCF, and H3K27ac, and a broad peak cutoff of 1×10^{-4} for RNAPolII and H3K27ac. Bedgraph pileups were also generated using the -B and –SPMR flags and then converted to bigwigs using UCSC bedGraphToBigWig. During preliminary analysis, it was noted the RNAPolII, CTCF, and H3K27ac ChIP-seq signals were not proportional across libraries. To address this, each pileup interval was divided by scalar size factors. To determine each size factor, the maximum global signal was found for each chromosome, followed by the mean of those maximums, and then mean was divided by 14.106 for the RNAPolII libraries, 11.339 for the CTCF libraries, and 9.965 for the H3K27ac libraries.

RNAPolII occupancy analysis

RNAPolII binding was profiled for each gene by computing the transcription start site (TSS) ChIP-seq pileup signal (S_{TSS}), the transcription end site (TES) ChIP-seq pileup signal (S_{TES}), and the gene body ChIP-seq pileup signal (S_{GB}). The total pileup signal spanning +/− 500 bins relative to the TSS was obtained using the values function from the pyBigWig python package (v0.3.18). Then, S_{TSS} was computed as the maximum ChIP-seq pileup signal within this region S_i (Equation 15):

$$S_{TSS} = \max_{i \in \{-500\ bp \ldots 500\ bp\}} (S_i) \quad \text{(Equation 15)}$$

To correct for enriched 3′ signal, the total pileup signal spanning +/− (gene length * 0.025) relative to the TES using the pyBigWig values function was also obtained. S_{TES} was computed as the mean ChIP-seq pileup signal within this region S_j (Equation 16):

$$S_{TES} = \operatorname{mean}_{j \in \{-(0.025 * gene\ length)\ bp \ldots +(0.025 * gene\ length)\ bp\}} (S_j) \quad \text{(Equation 16)}$$

Finally, the total pileup signal spanning +500 base-pairs from the TSS region to the start of the TES region was obtained. Before computing the gene body pileup signal, S_{GB}, the region was resized into 4,375 bins to normalize for gene length. This was accomplished using the cv2.resize function from the opencv-python package (v4.5.4.58). After resizing, the gene body signal, S_{GB}, was computed (Equation 17):

$$S_{GB} = \underset{k \in \{0\ bins \ldots 4375\ bins\}}{mean}(S_k) \quad \text{(Equation 17)}$$

Each gene was parsed into repressed, elongated, and initiated categories across each cellular stage. For a gene to be classified as repressed, four conditions must be met: $S_{TSS} < 0.75$ and $S_{GB} < 0.2$, $S_{TES} < 0.3$ and mean normalized gene expression less than 300. For a gene to be classified as initiated, three conditions must be met: $S_{TSS} > 0.75$ and $S_{GB} < 0.2$, and $S_{TES} < 0.3$. For a gene to be classified as elongated, two conditions must be met: $S_{TSS} > 0.75$ and $S_{GB} > 0.2$ 5. ChIP-seq pileup heatmaps were generated using deeptools.

Intersection of RNAPolII occupancy profiles with CTCF peaks

For each gene that exhibits each of the three RNAPolII occupancy profiles in each cellular stage, promoter regions (i.e., 2kb upstream of the TSS) were intersected with CTCF peaks identified in that cellular stage using bedtools intersect. For example, the promoters of genes that were classified as elongated in NPCs were intersected with CTCF peaks identified in NPCs, the promoters of genes that were classified as elongated in neuron were intersected with CTCF peaks identified in neuron, et cetera.

Gene ontology analysis

Using WebGestalt (http://www.webgestalt.org), gene ontology enrichment was performed. The following settings were used: Organism of interest = Homo Sapiens; Method of interest = Over-Representation Analysis; Functional Database = geneontology, Biological Process noRedundant; Select Reference Set, genome.

Parsing RNAPolII transition classes

For each cellular stage transition (i.e., hiPSC-to-NPC and NPC-to-neuron), each gene was parsed into nine RNAPolII transition classes based on their RNAPolII occupancy profiles. Genes were parsed as follows.

i. **Repressed-Repressed:** RNAPolII occupancy profile is classified as repressed in both prior and latter cellular stages.
ii. **Repressed-Initiated**: RNAPolII occupancy profile is classified as repressed in the prior cellular stage but initiated in the latter cellular stage.
iii. **Repressed**-Elongated: RNAPolII occupancy profile is classified as repressed in the prior cellular stage but elongated in the latter cellular stage.
iv. **Initiated-Initiated**: RNAPolII occupancy profile is classified as initiated in both prior and latter cellular stages.
v. **Initiated-Repressed**: RNAPolII occupancy profile is classified as initiated in the prior cellular stage but repressed in the latter cellular stage.
vi. **Initiated-Elongated**: RNAPolII occupancy profile is classified as initiated in the prior cellular stage but elongated in the latter cellular stage.
vii. **Elongated-Elongated**: RNAPolII occupancy profile is classified as elongated in both prior and latter cellular stages.
viii. **Elongated-Initiated**: RNAPolII occupancy profile is classified as elongated in the prior cellular stage but initiated in the latter cellular stage.
ix. **Elongated-Repressed**: RNAPolII occupancy profile is classified as elongated in the prior cellular stage but repressed in the latter cellular stage.

Parsing looping classes

For each RNAPolII transition class, promoters (i.e., 2kb upstream of the TSS) were intersected with both anchors of each differential and invariant loop calls (Figures 4 and S3).

For the hiPSC-to-NPC transition, a gene was classified as **Class 1** if the promoter intersected with one or more hiPSC-specific loops, **Class 2** if the promoter only intersected with one or more NPC-specific loops, **Class 3** if the promoter intersected with one or more hiPSC-specific loops and one or more NPC-specific loops, **Class 4** if the promoter intersected with only one or more invariant loops, and **Class 5** if the promoter did not intersect with any differential or invariant loops.

For the NPC-to-neuron transition, a gene was classified as **Class 1** if the promoter intersected with one or more NPC-specific loops, **Class 2** if the promoter only intersected with one or more neuron-specific loops, **Class 3** if the promoter intersected with one or more NPC-specific loops and one or more neuron-specific loops, **Class 4** if the promoter intersected with only one or more invariant loops, and **Class 5** if the promoter did not intersect with any differential or invariant loops.

Defining differential enhancer regions

For Figures 5 and S7, we filtered genes parsed by RNAPolII and looping status based on their proximity to differential enhancers. To identify differential enhancers, concatenated list of H3K27ac ChIP-seq peak calls identified in NPC and neurons were created. From

this list, any H3K27ac peaks that overlapped hg38 TSS±1kb, hg38 Refseq exons, 3′ UTRs and 5′ UTRs (downloaded from the UCSC Genome browser on October 25th 2022) were removed using bedtools. This resulted in H3K27ac peaks at only introns and intergenic regions. Finally, H3K27ac peaks were parsed into cell-type specific enhancers and invariant enhancers. First, the average bigwig signal across the peak interval was calculated using the pybigwig package (v0.3.13) in both conditions and then the fold change in H3K27ac-seq signal as [NPC stage/neuron stage] was calculated. A peak was assigned as an NPC-specific enhancer if it exhibited a fold change H3K27ac ChIP-seq signal [NPC stage/neuron stage] > 2.0. In addition, the NPC signal at the peak had to be above the 20th percentile signal threshold of all the enhancers. A peak was assigned as a neuron-specific enhancer peak using similar logic. The remaining enhancer peaks were classified as invariant enhancers.

The same method was applied to the H3K27ac ChIP-seq peak calls identified in hiPSC and NPCs. A peak was assigned as an hiPSC-specific enhancer if it exhibited a fold change H3K27ac ChIP-seq signal as [hiPSC stage/NPC stage] > 2.0 and if the hiPSC signal at the peak was above 20th percentile signal threshold of all the enhancers. A peak was assigned as an NPC-specific enhancer peak in the same manner with the conditions reversed.

Classification of NPC-specific and neuron-specific loops using promoter positions and differential H3K27ac enhancer regions

Loops formed by the genes present in the RNAPolII transition classes based on the cell-type specific H3K27ac enhancers from ChIP-seq peak calls were classified (see previous section). Loop anchors were defined as the region between start and end coordinates enclosing all the pixels in the looping cluster. Promoter regions were defined as 2kb upstream of the TSS of genes.

Promoter loops were classified as follows.

i. **Promoter–neuron-specific enhancer**: if one loop anchor colocalized with a promoter region and the other loop anchor colocalized with a neuron-specific enhancer regardless of the presence of other enhancers.
ii. **Promoter–NPC-specific enhancer**: if one loop anchor colocalized with a promoter region and the other loop anchor colocalized with a prior NPC-specific enhancer but not a neuron-specific enhancer.
iii. **Promoter–Cell type-invariant enhancer**: if one loop anchor colocalized with a promoter region and the other loop anchor colocalized with only invariant enhancers.
iv. **Promoter–Non-coding//non-enhancer**: if one loop anchor colocalized with a promoter region and the other loop anchor did not overlap with any enhancers.
v. **Promoter–Promoter:** if both loop anchors were colocalized with any promoter regions as defined above. This group is also mutually exclusive with the promoter-enhancer groups.

For the hiPSC-to-NPC transition, we classified promoter loops as follows.

i. **Promoter–NPC-specific enhancer**: if one anchor colocalized with promoter region and the other anchor colocalized with an NPC-specific enhancer regardless of the presence of other enhancers.
ii. **Promoter–iPSC-specific enhancer**: if one anchor colocalized with a promoter region and the other anchor colocalized with a prior hiPSC-specific enhancer but not an NPC-specific enhancer.
iii. **Promoter–Cell type-invariant enhancer**: if one anchor colocalized with a promoter region and the other anchor colocalized with only invariant enhancers.
iv. **Promoter–Non-coding//non-enhancer**: if one anchor colocalized with a promoter region and the other anchor did not colocalize with any enhancers.
v. **Promoter–Promoter:** if both anchors colocalized any promoter regions as defined above. This group is also mutually exclusive with the promoter-enhancer groups.

Stratification of promoter-promoter looping genes and non-looping genes

Promoter-promoter looping genes identified in the previous section and non-looping genes identified in the parsing looping classes section were further stratified based on the presence of close-range intronic and intergenic H3K27ac peaks (i.e., within ±80 kb of the TSS). This range was determined based on the resolution of the loop calling method. Genes that did not have any intronic and intergenic H3K72ac peaks within ±80 kb of the TSS were further characterized for gene expression in Figures 5 and S7–S9.

Stratification of enhancers at cell-type specific enhancer-promoter loops by RNAPolII occupancy

Using the values function from the pyBigWig python package, cell-type relevant RNAPolII ChIP-seq pileup signal at cell-type specific enhancers was obtained (i.e., RNAPolII ChIP-seq from neurons was intersected with neuron-specific enhancers). To normalize for enhancer length, each enhancer was binned into 1000 bins. This was accomplished using the cv2.resize function from the opencv-python package. Then, the mean signal was computed for each binned enhancer. For an enhancer to be classified as having RNAPolII signal, its mean value had to be greater than 0.2. Enhancer anchors of enhancer-promoter loops from Figure 5 were intersected with stratified enhancers using bedTools intersect to obtain proportions summarized in Figure S10A–C. For Figures S10D–S10F, A transcriptional unit was classified as being at an enhancer promoter loop with enhancer RNAPolII signal if at least one

enhancer at the enhancer anchor had RNAPolII signal. A transcriptional unit was classified as being at an enhancer-promoter loop without enhancer RNAPolII signal if none of the enhancers at the enhancer anchor had RNAPolII signal.

RNAPolII Occupancy Analysis of DLD-1 RNAPolII ChIP-seq

First, the RNAPolII ChIP-seq bigwig from[39] was lifted over from hg19 to hg38 using UCSC liftOver. Then, from the hg38 lifted bigwig, genes were classified as repressed, initiated, and elongated using the algorithm described in the RNAPolII occupancy analysis section. For a gene to be classified as repressed, three conditions must be met: $S_{TSS} < 0.75$ and $S_{GB} < 0.1$, $S_{TES} < 0.3$. For a gene to be classified as initiated, three conditions must be met: $S_{TSS} > 0.75$ and $S_{GB} < 0.1$, and $S_{TES} < 0.3$. For a gene to be classified as elongated, two conditions must be met: $S_{TSS} > 0.75$ and $S_{GB} > 0.2$.

Analysis of DLD-1 CTCF CUT&Tag

MACS2 bdgpeakcall was used to call CTCF peaks from the published bedgraphs of replicates 1 and 2 separately.[32] A p-value cutoff of 1×10^{-2} was used. Peak calls were then merged using bedtools merge. For each gene classified as repressed, initiated, or elongated, the promoter regions (i.e., 2kb upstream of the TSS) were intersected with the merged CTCF peaks using bedtools intersect.

Analysis of DLD-1 Micro-C

Using the provided mcool files from,[32] the Mustache package was used to identify loops in Micro-C from wild-type DLD-1 cells. The following parameters were used: -pt 0.05 and -r 5000. The loop epicenters that intersected the 2kb promoters of repressed, initiated, and elongated genes with and without CTCF were used in the paired aggregate peak analyses in Figure 6. Aggregate peak analyses were generated using cooltools. ChIP-seq pileup heatmaps of gene promoters that intersected loops called in the control condition in Figure 6 were generated using deeptools.

mm10 RefSeq reference genome

The mm10 reference transcriptome was downloaded from the UCSC Table Browser on February 2nd, 2024. NCBI RefSeq and RefSeq Curated were used as track and table options respectively (https://genome.ucsc.edu/cgi-bin/hgTables?hgsid=1345872709_tycZ8naeqyTXL51A2BV9FK8CsBk0). Genes that (i) are in canonical chromosomes 1–19 and chromosome X, and (ii) have a gene length greater than 2 kb were analyzed.

RNAPolII Occupancy Analysis of JM8.N4 mouse ES cell RNAPolII ChIP-seq

First, the RNAPolII ChIP-seq bigwig from[40] was used to classify genes as repressed, initiated, and elongated. This was accomplished using the algorithm described in the RNAPolII occupancy analysis section. For a gene to be classified as repressed, three conditions must be met: $S_{TSS} < 12$ and $S_{GB} < 9$, $S_{TES} < 12$. For a gene to be classified as initiated, three conditions must be met: $S_{TSS} > 45$ and $S_{GB} < 9$, and $S_{TES} < 12$. For a gene to be classified as elongated, two conditions must be met: $S_{TSS} > 45$ and $S_{GB} \geq 12$.

Analysis of JM8.N4 mouse ES cell CTCF ChIP-seq

Peak calls were obtained from.[43] For each gene classified as repressed, initiated, or elongated, the promoter regions (i.e., 2kb upstream of the TSS) were intersected with the CTCF peaks using bedtools intersect.

Analysis of JM8.N4 mouse ES cell Micro-C

For both the triptolide (TRP) and flavopiridol (FLV) Micro-C, only paired-end fastq files from replicate 2 were obtained and analyzed from GEO: GSE130275 to conserve computational resources. Both replicates were mapped to mm10 with an up-to-date version of HiC-Pro (v 3.1.0) using the same settings outlined in the Hi-C pre-processing section. Running the full HiC-Pro pipeline yielded 230,232,230 unique valid pairs for TRP replicate 2 and 220,970,201 unique valid pairs for FLV replicate 2. For the wild-type Micro-C, the wild-type allValidPairs file corresponding to $N = 2.6$ billion reads was downloaded from GEO: GSE130275. Due to dramatic sequencing depth differences, the wild-type, FLV, and TRP conditions were downsampled together. Trans valid pairs and chromosome Y were excluded from further analysis. Each chromosome was downsampled separately from other chromosomes, resulting in a total of 193,477,253 valid pairs genome-wide per condition. The hicpro2higlass.sh script from the HiC-Pro package (v 3.1.0) was used with default settings to construct three 10kb resolution.cool matrices from each downsampled allValidPairs file for aggregate peak analyses in Figure 6 and a 5kb resolution.cool matrix from the wild-type $N = 2.6$ billion reads allValidPairs file for loop calling. Each.cool matrix was ICE balanced using cooler balance with the following settings: –cis-only and –max-iters 500. Loops were called on the wild-type 5kb resolution.cool matrix using the Mustache package. The following parameters were used: -pt 0.05 and -r 5000. After loop calling, the 5kb bin coordinates of the loop epicenters were converted to 10kb bin coordinates for aggregate peak analyses in Figure 6. Only the loop epicenters that intersected the 2kb promoters of repressed, initiated, and elongated genes with and without CTCF were used. Aggregate peak analyses were generated using cooltools. ChIP-seq pileup heatmaps of genes whose promoters intersected loops called in the wild-type condition in Figure 6 were generated using deeptools.