Mapping *cis*- and *trans*-regulatory target genes of human-specific deletions

Tyler Fair[1,2,3], Bryan J. Pavlovic[1,3], Nathan K. Schaefer[1,3], Alex A. Pollen[1,3]

[1]Eli and Edythe Broad Center of Regeneration Medicine and Stem Cell Research, University of California, San Francisco, San Francisco, CA, USA.
[2]Biomedical Sciences Graduate Program, University of California, San Francisco, San Francisco, CA, USA.
[3]Department of Neurology, University of California, San Francisco, San Francisco, CA, USA.

Abstract

Deletion of functional sequence is predicted to represent a fundamental mechanism of molecular evolution[1,2]. Comparative genetic studies of primates[2,3] have identified thousands of human-specific deletions (hDels), and the *cis*-regulatory potential of short (≤31 base pairs) hDels has been assessed using reporter assays[4]. However, how structural variant-sized (≥50 base pairs) hDels influence molecular and cellular processes in their native genomic contexts remains unexplored. Here, we design genome-scale libraries of single-guide RNAs targeting 7.2 megabases of sequence in 6,358 hDels and present a systematic CRISPR interference (CRISPRi) screening approach to identify hDels that modify cellular proliferation in chimpanzee pluripotent stem cells. By intersecting hDels with chromatin state features and performing single-cell CRISPRi (Perturb-seq) to identify their *cis*- and *trans*-regulatory target genes, we discovered 19 hDels controlling gene expression. We highlight two hDels, hDel_2247 and hDel_585, with tissue-specific activity in the liver and brain, respectively. Our findings reveal a molecular and cellular role for sequences lost in the human lineage and establish a framework for functionally interrogating human-specific genetic variants.

Introduction

Millions of single-nucleotide and structural variants (SVs)—deletions, duplications, insertions, and inversions ≥50 base pairs (bp) in length—have accumulated in the human lineage since divergence from nonhuman primates. Contained within this genetic variation are alterations to functional sequences that distinguish humans from nonhuman primates. However, the overwhelming majority of variants are predicted to be selectively neutral[5]. Among polymorphic variants, deletions are enriched for driving splicing and expression quantitative trait loci[6] and noncoding deletions predicted to be deleterious exhibit levels of purifying selection comparable to loss-of-function coding alleles[7]. Noncoding deletions removing *cis*-regulatory elements may also underlie instances of adaptive evolution[8]. We reasoned that inactivating human-specific deletions (hDels) using tiling CRISPRi-based genetic screens in chimpanzee cells would enable systematic genome-scale interrogation of the effect of this class of genomic alterations on cellular proliferation and gene expression.

Results

We focused on 7,278 SV-sized (≥50 bp) hDels previously identified through comparison of long-read great ape genomes[3] (Methods). hDels span 12.7 megabases (Mb) of the chimpanzee reference genome (panTro6), are a median of 626 bp (range 50 to 262,923 bp), and primarily intersect noncoding regions (52.4% of hDel base pairs are intronic, 47.4% are intergenic, and 0.2% are exonic). Compared with matched genomic sequences[9], hDels are enriched for repeat elements ($p < 10^{-3}$, 63.2% of hDel base pairs are repetitive) and intergenic regions ($p < 10^{-3}$) and depleted from introns ($p = 7 \times 10^{-3}$) and exons ($p < 10^{-3}$). To characterize the epigenetic state of chromatin at hDels present in the chimpanzee reference genome, we performed Omni ATAC-seq[10] (Extended Data Fig. 1) in chimpanzee induced pluripotent stem (iPS) cells from four individuals (C3624K, C3651, C8861, Pt5-C) and profiled H3K4me1^{ab8895}, H3K4me3^{ab8580}, H3K27ac^{ab4729}, and H3K27me3^{9733S} histone modifications using CUT&Tag[11] (Extended Data Fig. 2). Although hDels are depleted from Tn5-accessible and H3K4me1-, H3K4me3-, H3K27ac-, and H3K27me3-modified chromatin ($p < 10^{-3}$), we identified 290 hDels intersecting at least one of these epigenetic features, revealing sequences lost in the human lineage harboring candidate *cis*-regulatory elements.

To evaluate the functions of hDels in their native genomic contexts using a CRISPRi-based genetic screening approach, we first introduced dCas9-KRAB into the *CLYBL* safe harbor locus[12] in iPS cells from two male chimpanzees (Extended Data Fig. 3a, C3624K, Pt5-C). We then designed a library of sgRNAs tiling across all hDels and separately, a library targeting hDels intersecting epigenetic features associated with *cis*-regulatory elements.

To probe the effect of hDels as a class of human-specific SVs on a quantitative cellular phenotype, we designed a library of 170,904 sgRNAs tiling across 7.2 Mb of sequence within 6,358 hDels (Extended Data Fig. 3b, hDel-v1). We considered hDels independent of evolutionary conservation or epigenetic state, as these features may not be predictive of all classes of *cis*-regulatory elements[13]. To tile across all uniquely targetable hDels, sgRNAs were assigned to 50-bp genomic windows and the sgRNA with the highest predicted activity per window was selected for inclusion in hDel-v1 (Fig. 1a, median 52 bp between sgRNAs, median 14 sgRNAs per hDel). We reasoned that iPS cells would be a useful model for studying hDels because of their transcriptionally permissive chromatin structure[14] and sensitivity to

proliferation-modifying perturbations. We transduced chimpanzee CRISPRi iPS cells (C3624K, Pt5-C) with the lentiviral hDel-v1 sgRNA library, selected for sgRNA-expressing cells with puromycin, cultured cells for 10 days, and quantified sgRNA enrichment and depletion by high-throughput sequencing (Fig. 1a).

Analysis of technical and biological replicates revealed that sgRNAs were highly correlated between replicates of the same cell line (Fig. 1d and Extended Data Fig. 3, Pearson's r = 0.78 to 0.88) and between cell lines (r = 0.69). As expected, sgRNAs targeting the promoters of essential or proliferation-suppressor genes were depleted or enriched, respectively (Fig. 1d). Using DESeq2[15] to model sgRNA counts from hDel-v1 and compute sgRNA false discovery rates (FDRs), we identified 1,851 hDel-targeting sgRNAs modifying cellular proliferation (FDR < 0.01, Fig. 1c). Because proximal sgRNAs mediate variable CRISPRi activity, we assigned hDel-targeting sgRNAs to overlapping 500-bp genomic windows (250-bp step size) and combined sgRNAs into hDel FDRs using alpha-robust rank aggregation (α-RRA). Using this approach, we found that hDel_6304, a 382 bp deletion located within the first intron of the nucleosome remodeling and deacetylase (NuRD) complex subunit *MBD3*, increases cellular proliferation upon CRISPRi. (Fig. 1d). Genome-wide, we identified 16 hDels modifying cellular proliferation (FDR < 0.1, Fig. 1e).

To refine the identification of proliferation-modifying hDels, we designed a second tiling library (hDel-v2) with reduced spacing between proximal sgRNAs. Because reduced spacing supports the discovery of hDels that may not have been detected using hDel-v1, we included hundreds of hDels with a single proliferation-modifying sgRNA. We omitted a genomic window strategy for sgRNA selection to maximize tiling density, resulting in a library of 78,270 sgRNAs targeting 558 hDels (Fig. 2a, median 7 bp between sgRNAs, median 119 sgRNAs per hDel). As with hDel-v1, we transduced chimpanzee CRISPRi iPS cells (C3624K) with the lentiviral hDel-v2 sgRNA library, selected and cultured sgRNA-expressing cells, and quantified sgRNA enrichment and depletion by high-throughput sequencing.

hDel-targeting sgRNAs screened in both hDel-v1 and hDel-v2 were highly correlated (r = 0.73, Fig. 2c), as were technical replicates (Extended Data Fig. 4, r = 0.82). We discovered 38 hDels modifying cellular proliferation (FDR < 0.1, Fig. 2c), including 14 hDels intersecting Omni ATAC-seq, H3K4me1, H3K4me3, H3K27ac, or H3K27me3, epigenetic features associated with *cis*-regulatory elements (Fig. 2d). For example, hDel_7051, a 2,602 bp deletion of a long interspersed nuclear element (LINE) element within an intronic region of the SWI/SNF family chromatin remodeler *ATRX*, intersects Omni ATAC-seq, H3K4me1, and H3K27ac, and reduces cellular proliferation upon CRISPRi (Fig. 2e). Together, hDel-v1 and hDel-v2 identify cellular phenotypes for select hDels and indicate that despite the predicted phenotypic importance of SV-sized noncoding deletions[7], hDels as a class of human-specific SVs are largely dispensable for proliferation.

We next sought to map the *cis*- and *trans*-regulatory targets of proliferation-modifying hDels using single-cell CRISPRi. To facilitate hDel-gene mapping, we designed a compact library of sgRNAs (hDel-v3) targeting hDels identified using hDel-v2 (122 sgRNAs targeting 19 hDels, FDR < 0.05). As positive and negative controls, we included transcription start site (TSS)-targeting sgRNAs, putative *cis*-regulatory element-targeting sgRNAs[16], and non-targeting sgRNAs (Extended Data Fig. 5a). We transduced chimpanzee CRISPRi iPS cells (C3624K) with the lentiviral hDel-v3 sgRNA library, selected

and cultured sgRNA-expressing cells for 7 days, and performed single-cell RNA sequencing (Fig. 3a, Direct-capture Perturb-seq). After filtering, we recovered 16,810 sgRNA-expressing cells (Extended Data Fig. 5b-d, median 15,366 UMIs per cell, median 151 cells per sgRNA).

To identify the transcriptional consequences of hDel-v3 sgRNAs, we summed gene expression counts across cells containing each sgRNA (+sgRNA 'pseudobulk') and negative-control sgRNAs (+negative-control sgRNA 'pseudobulk') and performed a likelihood-ratio test using DESeq2. TSS-targeting sgRNAs mediated efficient target gene repression (Fig. 3b, 52.9 to 93.4%, median 85.0%, FDR < 0.1), demonstrating the performance of CRISPRi and the accuracy of sgRNA-cell assignments. As expected, we observed an enrichment of significant sgRNA-gene pairs for hDel-targeting, but not non-targeting, sgRNAs (Fig. 3c). We identified 4 hDel-gene pairs within 100 kb (FDR < 0.1), including hDel_6012-*RPL26*, hDel_349-*MRPS14*, and hDel_6304-*MBD3*.

Targeting of hDel_6012 (Fig. 3d) reduced the expression of the 60S ribosomal protein L26 *RPL26* (Fig. 3e, 33.0 to 49.8%, median 41.3%, FDR < 0.1), and we observed a highly correlated relationship between the effect of hDel_6012-targeting sgRNAs on cellular proliferation and *RPL26* expression (Fig. 3f, r = 0.97). hDel_6012 does not intersect epigenetic features associated with *cis*-regulatory elements (Fig. 3d), highlighting the value of evaluating hDels independent of chromatin profiling. Genome-wide, we identified 636 differentially expressed genes upon hDel_6012 CRISPRi, including 35 genes encoding S and L ribosomal proteins, 32 of which (91.4%) were up-regulated (Fig. 3g, FDR < 0.1). As expected, sgRNAs targeting hDel_6012 elicited correlated genome-wide transcriptional responses (Extended Data Fig. 6a, r = 0.57 to 0.68, FDR < 0.1).

sgRNAs targeting hDel_349 (Extended Data Fig. 6b) reduced the expression of the mitochondrial ribosomal 28S subunit *MRPS14* (Extended Data Fig. 6c, 60.3 to 88.3%, median 74.7%, FDR < 0.1), and we observed a highly correlated relationship between the effect of hDel_349-targeting sgRNAs on cellular proliferation and *MRPS14* expression (Extended Data Fig. 6d, r = 0.73). Genome-wide, we identified 11 differentially expressed genes upon hDel_349 CRISPRi, all of which correspond to mitochondrial or nuclear-mitochondrial transcripts (Extended Data Fig. 6e, FDR < 0.1), indicating that mitochondrial dysfunction underlies reduced proliferation. While we cannot exclude the possibility that hDel_349-targeting sgRNAs interfere with transcription at the *MRPS14* promoter, the sgRNAs reducing *MRPS14* expression are located -302 to -658 bp relative to the TSS, outside of the optimal range of CRISPRi [17], and sgRNAs targeting *MRPS14* in human K562 cells[17] do not exhibit proliferation-modifying effects at similar distances (Extended Data Fig. 6f), suggesting that the observed effects are specific to hDel_349. *MRPS14* variants in humans are associated with muscle hypotonia, cognitive delay, and midface retrusion[18], providing evidence that loss of a *MRPS14 cis*-regulatory element may alter the development of multiple tissues.

Targeting of hDel_6304 (Fig. 3h) reduced the expression of the NuRD complex subunit *MBD3* (Fig. 3i, 35.5 to 51.7%, median 47.2%, FDR < 0.1). We observed an inverse relationship between the effect of hDel_6304-targeting sgRNAs on cellular proliferation and *MBD3* expression (Fig. 3j, r = -0.94), consistent with depletion of MBD3–NuRD inhibiting the differentiation of highly proliferative pluripotent cells[19]. Genome-wide, we identified 83 differentially expressed genes upon hDel_6304 CRISPRi, including transcription factors controlling meso-endoderm differentiation (*EOMES*, *GATA6*, *LHX1*,

MIXL1), all of which were down-regulated (Fig. 3k, FDR < 0.1). hDel_6304 is also accessible in chimpanzee neural progenitor cells[20], suggesting that loss of hDel_6304 in the human lineage may contribute to increased neural stem and progenitor cell proliferation[21] by delaying terminal differentiation.

We next examined *RPL26*, *MRPS14*, and *MBD3* expression in human and chimpanzee cells. To quantify *cis*-regulatory divergence, we compared the expression of human and chimpanzee alleles in an identical *trans* environment. As expected, human alleles drove reduced *MRPS14* expression compared to chimpanzee alleles in human-chimpanzee allotetraploid iPS cells (*cis* contribution 13.1%, FDR < 0.05), explaining cross-species differences in gene expression (14.3%, FDR < 0.005)[22]. Similarly, *MBD3* expression is reduced in human, compared to chimpanzee, iPS cells (31.7%, FDR < 0.001)[23], and human alleles drove reduced *MBD3* expression (*cis* contribution 34.7%, FDR < 0.15)[22]. Surprisingly, *RPL26* expression from human alleles is increased compared to chimpanzee alleles (*cis* contribution 21.6%, FDR < 0.05)[22], indicating that additional *RPL26 cis*-regulatory alterations occurred in the human and chimpanzee lineages. Together, these findings suggest that hDel_349 and hDel_6304 remove *cis*-regulatory elements contributing to reduced *MRPS14* and *MBD3* expression in human cells, with *trans*-regulatory target genes controlling mitochondrial function and differentiation, respectively.

We also investigated potential sources of sgRNA off-target activity for proliferation-modifying hDels lacking *cis*-regulatory target genes. Because hDels are enriched for repeat elements, we focused on nucleotide homopolymers in hDel-targeting sgRNAs. Grouping hDel-v2 sgRNAs by the presence and position of N_4 homopolymers (AAAA, GGGG, CCCC) revealed that sgRNAs containing guanine (G_4) and cytosine (C_4) homopolymers exhibited pervasive off-target activity (Extended Data Fig. 7a,b). Off-target activity was dependent on the position of the homopolymer within the sgRNA, with the greatest toxicity observed for sgRNAs containing G_4 or C_4 near the center of the spacer sequence: sgRNAs containing G_4 at spacer position 9 were 5.0-fold more likely to have significant effects on cellular proliferation compared to all hDel-v2 sgRNAs (Extended Data Fig. 7c,d, 28.6%, FDR < 0.05), while sgRNAs containing C_4 at spacer position 11 were 2.4-fold more likely to have proliferation-modifying effects (Extended Data Fig. 7e,f, 13.7%, FDR < 0.05). Consequently, we excluded G_4/C_4-containing sgRNAs from hDel-v2 and hDel-v3 analysis. In hDel-v3, we did not observe correlated genome-wide transcriptional responses for homopolymer-containing sgRNAs, revealing no single transcriptional basis for proliferation-modifying effects. We also identified G_4-associated toxicity for sgRNAs tiling across *GATA1* and *MYC* in human K562 cells[24] (Extended Data Fig. 7g,h), providing evidence that sgRNA nucleotide homopolymers are an unappreciated source of off-target activity in CRISPRi-based genetic screens.

As hDels may control gene expression without directly modifying cellular proliferation, we focused on nonessential hDels intersecting epigenetic features associated with *cis*-regulatory elements. We designed a library of sgRNAs (hDel-v4) targeting hDels intersecting Omni ATAC-seq, H3K4me1, or H3K27ac (Fig. 4a, 888 sgRNAs targeting 163 hDels), including putative *cis*-regulatory element-targeting sgRNAs[16] and non-targeting sgRNAs as positive and negative controls, respectively (Extended Data Fig. 8a). To facilitate hDel-gene mapping using a larger library of sgRNAs, we transduced chimpanzee CRISPRi iPS cells (C3624K) with the lentiviral hDel-v4 sgRNA library at a high

multiplicity of infection, selected and cultured sgRNA-expressing cells for 7 days, and performed single-cell RNA sequencing (Fig. 4a, Direct-capture Perturb-seq). After filtering, we recovered 18,571 sgRNA-expressing cells, detecting multiple sgRNAs per cell (Fig. 4b, Extended Data Fig. 8b-d, median 7 sgRNAs per cell, median 26,989 UMIs per cell, median 195 cells per sgRNA).

To identify the transcriptional consequences of hDel-v4 sgRNAs, we summed gene expression counts across cells containing each sgRNA (+sgRNA 'pseudobulk') and all other cells (-sgRNA 'pseudobulk') and performed a likelihood-ratio test using DESeq2. As expected, sgRNAs targeting a putative inhibitor of DNA binding *ID1 cis*-regulatory element[16] reduced the expression of *ID1* (23.6 to 37.2%, FDR < 0.1). As with hDel-v3, we observed an enrichment of significant sgRNA-gene pairs for hDel-targeting, but not non-targeting, sgRNAs (Fig. 4c). We identified 16 hDel-gene pairs within 100 kb (Fig. 4d, Extended Data Fig. 8e, FDR < 0.1), including hDel_2247-*PLPP1* and hDel_585-*HADHA*.

hDel_2247, a 6.8 kb intergenic deletion located between the phospholipid phosphatase *PLPP1* and the lysosomal amino acid transporter *SLC38A9*, intersects Omni ATAC-seq, H3K4me1, H3K27ac (Fig. 4e). Although hDel_2247 did not modify cellular proliferation (Fig. 4f), sgRNAs targeting hDel_2247 reduced the expression of *PLPP1* (Fig. 4f, 17.9 to 28.8%, median 23.2%, FDR < 0.1). *PLPP1* is a ubiquitously expressed phosphatase that hydrolyzes extracellular lipid phosphates including lysophosphatidic acid (LPA)[25]. To evaluate the *in vivo* activity of hDel_2247, we examined epigenetic features at the orthologous position in the mouse genome across a range of tissue types. We found that chimpanzee sequences within hDel_2247 conserved to mice intersect ATAC-seq, p300, and H3K27ac in the liver, but not other tissues, from embryonic day 11.5 (E11.5) to birth (P0, Fig. 4g, Fig. SX). Genetic disruption of *PLPP1* increases LPA in plasma and liver *in vivo*[26] and in primary hepatocytes *in vitro*[27], suggesting that increased extracellular levels of LPA may have systemic effects, including on neural stem and progenitor cell differentiation[28]. Together, these findings indicate that the loss of hDel_2247 in the human lineage removed a conserved *cis*-regulatory element regulating *PLPP1* with tissue-specific activity in the liver.

We next focused on hDel_585, a 207 bp intronic deletion located within the alpha subunit of the mitochondrial trifunctional protein *HADHA*. Although hDel_585 did not modify cellular proliferation (Fig. 4h), sgRNAs targeting hDel_585 reduced the expression of *HADHA* (Fig. 4i, 20.7 to 24.8%, FDR < 0.1). Because sequences within hDel_585 are not conserved to mice, we examined chromatin accessibility in additional nonhuman primates as a measure of *cis*-regulatory activity during development. Analysis of single-nucleus ATAC sequencing (snATAC-seq) of post-conception day 80 (PCD80) rhesus macaque prefrontal cortex (Extended Data Fig. 9, PFC) revealed that hDel_585 is accessible in radial glia, intermediate progenitor cells, and excitatory and inhibitory neurons (Fig. 4j). As deletions can increase *cis*-regulatory activity by removing transcriptional repressors or creating transcription factor binding sites [4], we performed Omni ATAC-seq of human, chimpanzee, and orangutan neuroepithelial cells to examine evolutionary differences in chromatin accessibility at the *HADHA* gene body. We did not observe accessibility at the boundaries of hDel_585 in human neuroepithelial cells (Fig. 4k), consistent with the loss of *cis*-regulatory sequence. These results provide evidence that the loss of hDel_585 in the human lineage removed a *cis*-regulatory element active in the forebrain regulating *HADHA*.

Several hDels linked to *cis*-regulatory target genes removed evolutionarily conserved sequences[2]. hDel_1273, linked to the GTP-specific beta subunit of succinyl-CoA synthetase *SUCLG2* (30.1%, FDR < 0.1), is partially conserved to platypus, and is predicted to be active in the developing mouse heart, limb, midbrain[29] (Fig. 4l-o). Additionally, sgRNAs targeting hDel_3779 reduced the expression of the RAC1 effector *FAM49B* (Fig. 4l, 14.2 to 23.6%, FDR < 0.1), a regulator of mitochondrial fission and cytoskeletal remodeling[30,31]. Finally, hDel_6842 intersects a mouse ENCODE proximal enhancer-like signature and is linked to the ceramide kinase *CERK* (Fig. 4m, 23.7 to 27.7%, median 26.2%, FDR < 0.1). CERK converts ceramide to ceramide 1-phosphate (C1P), a sphingolipid metabolite hydrolyzed by *PLPP1*[25], raising the possibility of epistasis between hDel_2247 and hDel_6842.

We also examined the relationship between hDels and human-chimpanzee *cis*-regulatory divergence. For *CERK*, *FAM49B*, *GRTP1*, and *SUCLG2*, human alleles drove reduced expression compared to chimpanzee alleles in allotetraploid iPS cells[22] (Fig. 4o, FDR < 0.1), consistent with the human-specific loss of *cis*-regulatory sequence. However, in certain cases, hDels removing *cis*-regulatory elements did not account for cross-species *cis*-regulatory divergence. For example, *ATRX* and *PLPP1* were expressed at similar levels from human and chimpanzee alleles, and *HADHA* was expressed at higher levels from human alleles (Fig. 4o, FDR < 0.1), suggesting that additional *cis*-regulatory alterations in the human and chimpanzee lineages attenuate or even reverse the *cis*-regulatory effects of select hDels, consistent with compensatory evolution within human-accelerated regions[32].

Discussion

We established a systematic approach for evaluating how sequences lost in the human lineage modify cellular proliferation and gene expression by performing CRISPRi-based genetic screens in chimpanzee cells. Using libraries of sgRNAs tiling across 7.2 Mb of sequence within 6,358 hDels, we assessed the effects of a class of human-specific SVs on a quantitative cellular phenotype. Although largely dispensable for proliferation, we identified hDels removing *cis*-regulatory elements controlling the expression of proliferation-modifying genes including *MBD3*, *MRPS14*, and *RPL26*, and discovered the *cis*-regulatory target genes of 16 nonessential hDels intersecting Omni ATAC-seq, H3K4me1, and H3K27ac. Several hDels linked to *cis*-regulatory target genes remove conserved sequences, and in certain cases, hDels account for human-chimpanzee *cis*-regulatory divergence, underscoring the importance of deletions as a source of evolutionary innovation[1].

Our study provides a framework for applying CRISPRi-based forward genetic screens to characterize human-specific genetic variants in their native genomic contexts at scale[33]. Because it is difficult to predict which variants are functional, scalable approaches, such as genetic screens in cellular models, are required to systematically probe the millions of base pair alterations that have accumulated in the human lineage. In contrast to pooled reporter assays, which measure the capacity for variants to drive reporter gene expression in synthetic constructs [34], CRISPRi enables linkages to cellular phenotypes and genome-wide transcriptional responses. Recent CRISPRi-based studies have linked single-nucleotide polymorphisms identified by GWAS to target genes in *cis* and *trans* in disease-relevant cell types [35,36], but the functions of human-specific variants as a class of genomic alterations are not restricted to a single tissue or cell type. Pluripotent stem cells provide a useful model for screening variants with unknown tissue specificities, as tissue-specific *cis*-regulatory elements may also be active

in undifferentiated cells due to their transcriptionally permissive chromatin structure[14]. Additional CRISPRi-based mapping in cell types associated with morphological evolution[37,38] will facilitate interrogation of *cis*-regulatory divergence in the human lineage.

The adoption of CRISPRi-based screens for probing human-specific variants is subject to several design considerations. For modeling certain developmental differences, such as the expansion of neural stem and progenitor cell populations during human neurogenesis[21], proliferation represents a scalable and quantitative cellular phenotype, but does not directly reveal *cis*-regulatory element-gene linkages. Alternatively, single-cell CRISPRi enables high-dimensional molecular phenotyping[39], but may require variant selection using genetic, epigenetic, or transcriptomic features. Independent of phenotyping strategy, human-specific SVs frequently contain repeat-rich sequences[3], necessitating careful consideration of mismatched off-target sites [40] and G_4/C_4 nucleotide homopolymers during the design of sgRNA libraries. Among CRISPR-Cas modalities, CRISPRi facilitates fine-mapping functional sequences within SVs, while nuclease-active Cas9 enables the reconstruction of derived and ancestral alleles using sgRNA pairs, but at efficiencies that may be incompatible with pooled screening[41].

In summary, the CRISPRi-based characterization of hDels presented here illuminates the loss of *cis*-regulatory elements in the human lineage and provides an approach, complemented by base and prime editors[42,43], for assigning molecular and cellular functions to all classes of human-specific genetic variants.

Methods

Sequencing

All Illumina sequencing was performed at the UCSF Center for Advanced Technology.

Cell lines and cell culture

Chimpanzee iPS cells from two healthy male donors (C3624K and Pt5-C) were cultured in v3.1 medium (65 ng/ml FGF2-G3, 2 ng/ml TGFβ1, 0.5 ng/ml NRG1, 20 µg/ml insulin, 20 µg/ml transferrin, 20 ng/ml sodium selenite, 200 µg/ml ascorbic acid 2-phosphate, 2.5 mg/ml bovine serum albumin (BSA), 30 ng/ml heparin, 15 µM adenosine, guanosine, cytidine, uridine nucleosides, 6 µM thymidine nucleosides in Dulbecco's Modified Eagle's Medium (DMEM)/Ham's F-12 (Corning, 10-092-CM)) supplemented with 100 IU/ml penicillin, and 100 µg/ml streptomycin at 37°C and 5% CO_2. At ≥80% confluency, cultures were dissociated with phosphate-buffered saline (PBS) supplemented with 0.5 mM ethylenediaminetetraacetic acid (EDTA), resuspended in v3.1 medium supplemented with 2 µM thiazovivin (MedChemExpress, HY-13257), and seeded on Matrigel-coated cell culture plates (Corning, 354230).

CRISPRi iPS cell line generation

To establish chimpanzee iPS cell lines expressing dCas9-KRAB (*ZNF10/KOX1*), 1×10^6 iPS cells were seeded at a density of 100,000 cells/cm^2 and transfected with 3 µg pC13N-dCas9-BFP-KRAB (Addgene, 127968), 0.375 µg pZT-C13-L1 (Addgene, 62196), and 0.375 µg pZT-C13-R1 (Addgene, 62197) using 10 µl Lipofectamine Stem (Invitrogen, STEM00001)[44]. After 7 days, BFP-positive iPS cells were isolated by single-cell fluorescence-activated cell sorting. Transgene integration at the *CLYBL* locus was verified by PCR. To assess CRISPRi-mediated transcriptional repression, iPS cells were transduced with non-targeting or *SEL1L*-targeting sgRNAs[45], selected with 1.5 µg/ml puromycin (Gibco, A1113803), and processed for quantitative reverse transcription PCR.

Lentivirus production and titration

Lenti-X HEK293T cells (Takara Bio, 632180) were maintained in DMEM/Ham's F-12 supplemented with 10% fetal bovine serum (FBS), 100 IU/ml penicillin, and 100 µg/ml streptomycin at 37°C and 5% CO_2. To generate lentivirus, 150 mm cell culture dishes were coated with 10 µg/ml poly-D-lysine (Sigma-Aldrich, P7405) and Lenti-X HEK293T cells were seeded at a density of 85,000 cells/cm^2. The following day, medium was replaced and Lenti-X HEK293T cells were transfected with 23.1 µg hDel sgRNA library transfer plasmid, 7.6 µg pMD2.G (Addgene, 12259), and 13.9 µg psPAX2 (Addgene, 12260) using 125 µl Mirus *Trans*IT-293 (Mirus, MIR 2700) in Opti-MEM (Gibco, 31985062). Lentivirus-containing supernatant was harvested two days post-transfection, filtered through a 0.45 µm PVDF membrane (Millipore, SLHV033RS), and concentrated using precipitation solution (Alstem, VC100). To determine functional lentiviral titer, iPS cells were transduced in a dilution series and fluorophore-positive populations were quantified using a flow cytometer three days post-transduction.

Human-specific deletions

hDel sequences (Table S11.1[3]) were extracted from panTro6 contigs and aligned to the assembled panTro6 reference genome. hDel coordinates were compared to the UCSC hg38-panTro6 net alignment using BEDTools (v. 2.27.1) and discrepancies between hDels, hCONDELs[2], and the UCSC hg38-panTro6 net alignment were resolved by reciprocal BLAT.

hDel enrichment analysis

Intersections between hDels and genomic features were tested for significance using a resampling approach. After discarding all hDel sequences mapped to unplaced contigs and alternate haplotypes, 1000 matched null sets of genomic features were sampled from panTro6 (excluding unplaced contigs and alternate haplotypes) using bootRanges[46]. Regions were sampled using a block length of 500 kb, excluding approximately 51 Mb of blacklisted sequence (excludeOption='drop', withinChrom=FALSE). This blacklist was generated by combining runs of at least 1 kb with less than 100% 50-mer mappability identified using GenMap[47] and nuclear mitochondrial insertions (NUMTs) inferred as described previously[48].

After sampling matched null features, intersection between each sampled set of features and the feature set of interest (introns, exons, repeat elements, Tn-5 accessible regions, and pA-Tn5-accessible regions) was assessed using BEDTools[49]. After counting the number of base pairs in each intersection, *p*-values for enrichment or depletion of intersection were computed as the number of null feature sets for which intersection with the feature set of interest was greater (for enrichment testing) or lower (for depletion testing) than the observed intersection between hDels and features of interest.

Introns, exons, and intergenic regions were extracted from the chimpanzee reference genome (panTro6) annotated with the Comparative Annotation Toolkit[50,51]. Repeat regions were identified using the UCSC panTro6 RepeatMasker annotations. All features were merged using BEDTools before performing intersections.

hDel sgRNA library design

For all hDel sgRNA libraries (hDel-v1, hDel-v2, hDel-v3, and hDel-v4), candidate hDel-targeting sgRNAs were identified and scored for predicted off-target activity against mismatched target sites in the chimpanzee reference genome (panTro6) using FlashFry[52] (v. 1.15; --maxMismatch=3; --scoringMetrics JostandSantos,dangerous,minot). Candidate sgRNAs (GN_{19}) with >1 perfect-match target site, CRISPRi specificity score < 0.2, maximal predicted CRISPRi activity at any off-target site > 0.8, or TTTT sequences were excluded from all libraries.

To maximize coverage of hDel sequences in the hDel-v1 library, sgRNAs were grouped into non-overlapping 50-bp bins corresponding to the genomic location of their target sequence. Candidate sgRNAs were then scored for predicted on-target activity using DeepHF, and the sgRNA with the highest DeepHF score in each 50-bp bin was selected for inclusion in the library (n = 3,121 hDels). For hDels targeted by fewer than 10 sgRNAs after filtering and binning, sgRNAs were ranked by their DeepHF scores and a sequentially increasing number of sgRNAs were selected per 50-bp bin until all hDels were targeted by at least 10 sgRNAs (n = 1,531 hDels). For the remaining hDels targeted by fewer than 10 sgRNAs, sgRNAs targeting human-conserved sequence flanking either side of each hDel

were filtered, binned, and ranked as described above until each hDel ± 250 bp was targeted by at least 5 sgRNAs (*n* = 1,706 hDels). In total, 170,904 sgRNAs tiling 6,358 hDels were included in hDel-v1 (median distance between sgRNAs: 52 bp; median number of sgRNAs per hDel: 14). Non-targeting sgRNAs (*n* = 3,000 sgRNAs) were generated by scoring random $GN_{19}NGG$ sequences against panTro6 and filtering for 0 perfect-match target sites. As protein-coding controls, sgRNAs targeting the promoters of essential genes (*n* = 8,068 sgRNAs targeting 2,017 genes), proliferation-suppressor genes (*n* = 1,692 sgRNAs targeting 423 genes), and chimpanzee organoid-expressed genes (*n* = 15,189 sgRNAs targeting 5,063 genes) were selected from hCRISPRi-v2 after off-target scoring against panTro6. The complete hDel-v1 sgRNA library contains 198,718 sgRNAs. All hDel-v1 sgRNAs were screened in a single pool.

Candidate hDel *cis*-regulatory elements were screened at higher tiling density in the hDel-v2 sgRNA library (*n* = 558 hDels, see hDel CRISPRi screening analysis). For the 50 hDels with the greatest number of sgRNAs passing off-target filters described above, all sgRNAs within ± 500 bp of MAGeCK- and iAnalyzer-identified hDel genomic windows and DESeq2-identified singleton sgRNAs were selected for inclusion in the library. For all other hDels, all hDel-targeting sgRNAs, as well as sgRNAs targeting human-conserved sequence flanking either side of each hDel (± 500 bp), were included. In total, 78,270 sgRNAs tiling 558 hDels were included in hDel-v2 (median distance between sgRNAs: 7 bp; median number of sgRNAs per hDel: 119). Non-targeting sgRNAs (*n* = 2,000 sgRNAs) and sgRNAs targeting the promoters of essential genes (*n* = 600 sgRNAs targeting 98 genes) and proliferation-suppressor genes (*n* = 394 sgRNAs targeting 60 genes) were selected from hDel-v1 and CEV-v1, respectively. The complete hDel-v2 sgRNA library contains 81,264 sgRNAs. All hDel-v2 sgRNAs were screened in a single pool.

Two additional sgRNA libraries were designed for single-cell CRISPRi screening to facilitate mapping hDel *cis*-regulatory element-gene pairs. For hDel-v3, all essential sgRNAs targeting α-RRA-identifed 250-bp hDel genomic windows (*n* = 122 sgRNAs targeting 19 hDels, see hDel CRISPRi screening analysis) were selected for inclusion in the library. sgRNAs targeting the promoters hDel-proximal genes (*n* = 18 sgRNAs targeting 12 genes), core-control non-targeting sgRNAs (Replogle *Cell* 2022) (*n* = 10 sgRNAs), and sgRNAs targeting putative *cis*-regulatory elements (Gasperini *Cell* 2019) (*n* = 10 sgRNAs targeting 5 *cis*-regulatory elements) were also included. The complete hDel-v3 sgRNA library contains 160 sgRNAs.

For hDel-v4, nonessential sgRNAs targeting hDels marked by chromatin state features associated with *cis*-regulatory elements including Omni ATAC-seq, H3K4me1, and H3K27ac (see Omni ATAC-seq and CUT&Tag) were selected for inclusion in the library. sgRNAs passing off-target filters described above and lacking AAAA, TTTT, GGGG, or CCCC sequences were scored for predicted on-target activity using DeepHF, and the 5 sgRNAs with the highest DeepHF scores targeting each hDel Tn5- or pA-Tn5-accessible region were included (*n* = 888 sgRNAs targeting 163 hDels). Core-control non-targeting sgRNAs (*n* = 25 sgRNAs), and sgRNAs targeting putative *cis*-regulatory elements (*n* = 10 sgRNAs targeting 5 *cis*-regulatory elements) were also included. The complete hDel-v4 sgRNA library contains 923 sgRNAs.

hDel sgRNA library cloning

Oligonucleotide pools were designed with flanking PCR adapter sequences and restriction sites (*BstXI*, *BlpI*), synthesized by Agilent Technologies, and cloned into the sgRNA expression vector pCRISPRia-

v2 (Addgene, 84832; hDel-v1, hDel-v2) or pJR101 (Addgene, 187241; hDel-v3, hDel-v4) as described previously[45]. Briefly, oligonucleotide pools were amplified by 8 to 10 cycles of PCR using NEBNext Ultra II Q5 Master Mix (New England Biolabs, M0544X), digested with *BstXI* and *BlpI*, size-selected by polyacrylamide gel electrophoresis, ligated into *BstXI*- and *BlpI*-digested pCRISPRia-v2 or pJR101, and introduced into MegaX DH10B $T1^R$ cells by electroporation (Invitrogen, C640003; Bio-Rad, 1652660).

hDel CRISPRi screening

Chimpanzee CRISPRi iPS cells (C3624K or Pt5-C, hDel-v1; C3624K, hDel-v2) were dissociated with Accutase (Innovative Cell Technologies, AT104-500), resuspended in v3.1 medium supplemented with 2 μM thiazovivin and 5 μg/ml polybrene (Mirus, MIR 6620), transduced with the hDel-v1 or hDel-v2 lentiviral sgRNA library at a target infection rate of 25%, and plated at a density of 85,000 cells/cm^2 in Matrigel-coated 5-layer cell culture flasks (Corning, 353144). Two days post-transduction, cells were dissociated with Accutase, resuspended in v3.1 medium supplemented with 2 μM thiazovivin and 1.5 μg/ml puromycin, and plated at a density of 100,000 cells/cm^2. Four days post-transduction, 200 M cells were harvested (t_0) and 300 M cells were resuspended in v3.1 medium supplemented with 1.5 μg/ml puromycin and plated (≥1000× sgRNA library representation). Selection efficiency was assessed using a flow cytometer (≥70% BFP+). Every two days, cells were dissociated with Accutase, resuspended in v3.1 medium supplemented with 2 μM thiazovivin, and plated. Technical replicates were maintained separately for the duration of the screen. After 10 days of growth, 200 M cells from each technical replicate were harvested (t_{final}). Genomic DNA was isolated from pelleted cells by column purification (Macherey-Nagel, 740950.50), and the sgRNA expression cassette was amplified by 22 cycles of PCR using NEBNext Ultra II Q5 Master Mix and primers containing Illumina P5/P7 termini and sample-specific TruSeq indices. Each sample was distributed into individual 100 μl reactions in 96-well plates, each containing 10 μg genomic DNA. Following amplification, reactions from each sample were pooled and a 100 μl aliquot was purified by double-sided SPRI selection (0.65×, 1×). Purified libraries were quantified using Agilent Bioanalyzer, pooled at equimolar concentrations, and sequenced on Illumina HiSeq 4000 using a custom sequencing primer (SE50; oCRISPRi_seq V5).

hDel CRISPRi screening analysis

Single-end sequencing reads were aligned to hDel-v1 or hDel-v2 and counted using MAGeCK[53] (v. 0.5.9.4; count). For hDel-v1 (C3624K, Pt5-C), sgRNAs were assigned to overlapping 500-bp hDel genomic windows (250-bp step size) depending on the genomic location of their target sequence. hDel genomic windows targeted by fewer than 5 sgRNAs were excluded from analysis. hDel genomic windows and sgRNA counts were then used as input for analysis by MAGeCK alpha-robust rank aggregation (α-RRA) (test --paired --gene-test-fdr-threshold 0.05 --norm-method control --gene-lfc-method alphamean), iAnalyzeR (combining sgRNA *Z*-scores for each genomic window using Stouffer's method followed by the Benjamini-Hochberg procedure), or DESeq2 (design=~individual+time). The union of the hDel 500-bp genomic windows and hDel-targeting sgRNAs identified using these approaches (n = 313 genomic windows MAGeCK α-RRA 10% FDR, n = 147 genomic windows iAnalyzeR 5% FDR, n = 87 genomic windows MAGeCK α-RRA 10% FDR and iAnalyzeR 5% FDR, n = 202 sgRNAs DESeq2 1% FDR) were included in the hDel-v2 sgRNA library (n = 558 hDels). For the

hDel-v1 Manhattan plot in Fig. 1, sgRNA adjusted *p*-values from DESeq2 were combined into FDRs corresponding to each 500-bp hDel genomic window using α-RRA (v. 0.5.9; --control).

For hDel-v2 (C3624K), sgRNA counts were used as input for differential analysis by DESeq2 (design=~time). sgRNAs containing GGGG sequences following spacer position 5 or CCCC sequences between spacer positions 10 and 12 were excluded from analysis due to pervasive off-target effects (Extended Data Fig.7; fraction of significantly enriched or depleted sgRNAs at least twofold greater than all hDel-v2 sgRNAs). Enriched sgRNAs ($\log_2$ fold-change ≥1) initially present at low abundance (≤5th percentile) were also excluded from analysis (n = 159 sgRNAs). hDel-targeting sgRNAs were assigned to nonoverlapping 250-bp hDel genomic windows, and sgRNA adjusted *p*-values from DESeq2 were combined into FDRs corresponding to each hDel genomic window using α-RRA (--control).

Single-cell hDel CRISPRi screening

For hDel-v3, chimpanzee CRISPRi C3624K iPS cells were dissociated with Accutase (Innovative Cell Technologies, AT104-500), resuspended in v3.1 medium supplemented with 2 μM thiazovivin and 5 μg/ml polybrene (Mirus, MIR 6620), transduced with the hDel-v3 lentiviral sgRNA library at a target infection rate of 10%, and plated at a density of 100,000 cells/cm^2. The following day, v3.1 medium was replaced. Two days post transduction, v3.1 medium was replaced and supplemented with 1.5 μg/ml puromycin. Selection continued until six days post transduction. Seven days post transduction, iPS cells were dissociated with Accutase and resuspended in 1× PBS supplemented with 0.04% BSA for single-cell RNA sequencing (Direct-capture Perturb-seq). iPS cells were partitioned into Gel Beads-in-emulsion (GEMs) across five wells using the 10x Genomics Chromium Controller and cDNA libraries from polyadenylated mRNAs and Feature Barcode-compatible sgRNAs were generated by following the 10x Genomics Chromium Next GEM Single Cell 3' Reagent Kits v3.1 (Dual Index) User Guide (CG000316 Rev D). cDNA libraries from mRNAs and sgRNAs were quantified using Agilent Bioanalyzer, pooled at a 4:1 molar ratio, and sequenced on Illumina NovaSeq 6000 (28x10x10x90).

For hDel-v4, chimpanzee CRISPRi C3624K iPS cells were transduced with the hDel-v4 lentiviral sgRNA library at a target infection rate of >95%.

Single-cell CRISPRi screening analysis

A mismatch map for hDel-v3 and hDel-v4 was generated using kITE[54] and indexed using kallisto[55] (v. 0.48.0; kb ref --workflow kite). Paired-end sequencing reads from the gene expression and CRISPR screening libraries were then pseudoaligned and error-collapsed using kallisto | bustools (v. 0.42.0; kb count --workflow kite:10xFB --filter bustools). The chimpanzee reference genome (panTro6) annotated with the Comparative Annotation Toolkit[50,51] was used as the reference transcriptome.

To assign sgRNAs to cells, a two-component Poisson-Gaussian mixture model[56] was fit for each sgRNA using the $\log_2$-transformed UMIs in the bustools-filtered cell by sgRNA matrix. Assignments were made when the posterior probability of a cell belonging to the second component of the mixture model was >0.5.

For hDel-v3, cells with fewer than 2,049 genes detected (10th percentile), 5,105 UMIs (10th percentile), and greater than 15% mitochondrial UMIs were filtered from the dataset using Scanpy (v.

1.9.1). After intersecting Cell Barcodes in the gene expression and CRISPR screening UMI matrices, 16,810 cells were retained for analysis (median genes detected per cell: 4,564; median gene UMIs per cell: 15,366; median cells per sgRNA: 151; median UMIs per sgRNA per cell: 1,597).

For hDel-v4, cells with fewer than 1,106 genes detected (10th percentile), 2,744 UMIs (10th percentile), and less than 1% or greater than 15% mitochondrial UMIs were filtered from the dataset. After intersecting Cell Barcodes in the gene expression and CRISPR screening UMI matrices, 18,571 cells were retained for analysis (median genes detected per cell: 6,143; median gene UMIs per cell: 26,989; median sgRNAs per cell: 7; median cells per sgRNA: 195; median UMIs per sgRNA per cell: 495).

To perform differential gene expression testing for hDel-v3, the unnormalized gene expression UMIs were summed across cells containing each sgRNA (+sgRNA 'pseudobulk') and cells containing non-targeting sgRNAs (+non-targeting sgRNA 'pseudobulk') using ADPBulk (https://github.com/noamteyssier/adpbulk) and a likelihood-ratio test was performed individually for each sgRNA using DESeq2 controlling for GEM well (design=~GEM+sgRNA; test='LRT', reduced=~GEM).

To perform differential gene expression testing for hDel-v4, the unnormalized gene expression UMIs were summed across cells containing each sgRNA (+sgRNA 'pseudobulk') and all other cells (-sgRNA 'pseudobulk') using ADPBulk and a likelihood-ratio test was performed individually for each sgRNA using DESeq2 controlling for GEM well (design=~GEM+sgRNA; test='LRT', reduced=~GEM). Cells containing any sgRNA targeting the same hDel were excluded from the -sgRNA 'pseudobulk'.

For hDel-v3 and hDel-v4, genes detected in fewer than 2,000 cells were filtered from the dataset prior to differential gene expression testing.

To identify differentially expressed genes for hDel-targeting sgRNAs, sgRNA-gene pairs were *Z*-score normalized and a *p*-value was calculated from the survival function of a normal distribution for each gene within 100 kb of any hDel. For each sgRNA-gene pair, the $\log_2$ fold-change was divided by the standard error of the $\log_2$ fold-change prior to *Z*-score normalization and a gene-wise null distribution was created from all sgRNA-gene pairs separated by ≥100 kb. Gene-wise *p*-values were concatenated, and the Benjamini-Hochberg procedure was applied to sgRNA-gene pairs separated by ≤100 kb.

Omni ATAC-seq

Chimpanzee iPS cells from four healthy donors (C3624K dCas9-KRAB, Pt5-C dCas9-KRAB, C8861, and C3651) were rinsed with DMEM/Ham's F-12 and dissociated with PBS supplemented with 0.5 mM EDTA. Cells were washed and resuspended in cold PBS supplemented with 0.04% BSA. Following counting, 100,000 cells were resuspended in 100 µl cold lysis buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.01% digitonin, 0.1% Tween-20, 0.1% NP40, and 1% BSA) by pipetting three times and incubated on ice for 3 min. Following lysis, 1 ml cold wash buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.1% Tween-20, and 1% BSA) was added and nuclei were centrifuged at 500×*g* for 5 min at 4°C. Supernatant was removed using P1000, P200, and P20 pipettes, and nuclei were resuspended in 50 µl transposition mix (25 µl 2× transposition buffer (Active Motif), 10 µl transposase (Active Motif), 2 µl 10× PBS, 0.5 µl 0.5% digitonin, 0.5 µl 10% Tween-20, and 12 µl dH_2O) by pipetting three times. Transposition reactions were incubated for 30 min at 37°C while shaking at 1,000 rpm and cleaned using DNA Clean & Concentrator-5 (Zymo Research, D4003). Transposed libraries were amplified in 50 µl PCRs (0.5 µl Q5 DNA polymerase, 10 µl 5× Q5 reaction buffer, 1 µl 10

mM dNTP mix, 2.5 µl 25 µM i7 index primer and 2.5 µl 25 µM i5 index primer[57]) using the following cycling parameters: 72°C for 5 min; 98°C for 30 s; 10 cycles of 98°C for 10 s, 63°C for 30 s, and 72°C for 1 min. Amplified transposed libraries were purified by SPRI selection (1.2×), quantified using Agilent Bioanalyzer, and pooled at equimolar concentrations. Purified libraries were sequenced on Illumina HiSeq 4000 (PE100).

Neuroepithelial cell differentiation and Omni ATAC-seq

Chimpanzee (C3624K dCas9-KRAB, C8861), human (H20961), and orangutan (Jos-3C1) iPS cells were seeded at a density of 100,000 cells/cm^2 in v4 medium (v3.1 medium with 5 mg/ml BSA) containing CEPT[58]. The following day (Day 1), medium was replaced with DMEM/Ham's F-12 differentiation medium (DMEM/Ham's F-12 containing 1% polyvinyl alcohol, 100 µg/ml 2-Phospho-L-ascorbic acid trisodium salt, 20 ng/ml sodium selenite, 20 µg/ml holo-transferrin, 20 µg/ml insulin, and 0.1 mg/ml Primocin) supplemented with 0.5 µM LDN193189, 10 µM SB431542, and 0.1 µM Wnt-C59. Medium was replaced on Day 3. On Day 5, medium was replaced with DMEM/Ham's F-12 differentiation medium supplemented with 0.5 µM LDN193189, 10 µM SB431542, and CEPT. On Day 7, cells were dissociated with Accutase and seeded at a density of 800,000 cells/cm^2 onto polyethylenimine (0.1%) and Matrigel-coated plates in DMEM/Ham's F-12 differentiation medium supplemented with a modified 6F[59] formulation (1 µM SB431542, 0.1 µM K02288, 0.1 µM AKTiVIII, 0.075 µM MK2006, 0.1 µM LDN193189, 0.5 µM CHIR99021, 0.2 µM NVP-TNKS656, 25 ng/ml SHH), and CEPT. On Day 10, medium was replaced with DMEM/Ham's F-12 differentiation medium supplemented with 25 ng/ml SHH, 50 ng/ml FGF8, and CEPT. On Day 12, medium was replaced with DMEM/Ham's F-12 differentiation medium supplemented with 25 ng/ml SHH, and 50 ng/ml FGF8. On Day 15, medium was replaced and cells were rinsed with DMEM/Ham's F-12 and dissociated with Accutase. Omni ATAC-seq was performed as described above. Purified libraries were sequenced on Illumina Novaseq X (PE150).

Omni ATAC-seq analysis

Paired-end sequencing reads were trimmed using Cutadapt[60] (v. 3.4; -q 20 --minimum-length 20) and aligned to panTro6 using Bowtie 2[61] (v. 2.2.5; --very-sensitive -X 2000 -k 10). SAM files were converted to BAM format while discarding alignments with MAPQ < 15, sorted by position or read name, and indexed using SAMtools[62] (v. 1.10, -q 15). Tn5-accessible regions were called following PCR duplicate removal using Genrich (https://github.com/jsh58/Genrich) (v. 0.6.1; -j -r -e chrM -q 0.05). Tn5-accessible regions were intersected using BEDTools multiinter (v. 2.27.1) and only regions common to all four iPS cell lines were retained for analysis (n = 94,003). BAM files were converted to bedGraph format using deepTools[63] (v. 3.5.1; --binSize 1 --ignoreDuplicates) and visualized using SparK (https://github.com/harbourlab/SparK) (v. 2.6.2). To identify intersections between Omni ATAC-seq and hDels, at least half of a Tn5-accessible region was required to intersect a hDel.

CUT&Tag

Concanavalin A-coated beads (Epicypher, 21-1401) were washed twice and resuspended in binding buffer (20 mM HEPES pH 7.5, 10 mM KCl, 1 mM $CaCl_2$, and 1 mM $MnCl_2$). Chimpanzee iPS cells from two healthy donors (C3624K dCas9-KRAB and C8861) were rinsed with DMEM/Ham's F-12 and dissociated with PBS supplemented with 0.5 mM EDTA. Following counting, 175,000 cells were washed twice in 1 ml wash buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM spermidine, and 1

tablet cOmplete EDTA-free protease inhibitor cocktail) and incubated with 11 µl concanavalin A-coated beads on an end-over-end rotator for 10 min. The cell and bead mixture was placed on a magnetic stand and unbound supernatant was discarded. Bead-bound cells were resuspended in 50 µl primary antibody buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM spermidine, 1 tablet cOmplete EDTA-free protease inhibitor cocktail, 0.05% digitonin, 2 mM EDTA, and 0.1% BSA) and 0.5 µl primary antibody (1:100 dilution; H3K4me1 (Abcam, ab8895, lot GR3426435-2), H3K4me3 (Abcam, ab8580, lot GR3425199-1), H3K27ac (Abcam, ab4729, lot GR3442878-1), or H3K27me3 (Cell Signaling Technology, 9733S, lot 19)) was added. Bead-bound cells and primary antibody were mixed by pipetting and placed on a nutator overnight at 4°C. The following day, the primary antibody solution was placed on a magnetic stand and supernatant was discarded. Bead-bound cells were resuspended in 50 µl secondary antibody buffer (goat anti-rabbit IgG (Epicypher, 13-0047) diluted 1:100 in 20 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM spermidine, 1 tablet cOmplete EDTA-free protease inhibitor cocktail, and 0.05% digitonin) and placed on a nutator at room temperature for 60 min. Bead-bound cells were washed three times in 1 ml digitonin-wash buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM spermidine, 1 tablet cOmplete EDTA-free protease inhibitor cocktail, and 0.05% digitonin), resuspended in 50 µl digitonin-300 buffer (20 mM HEPES pH 7.5, 300 mM NaCl, 0.5 mM spermidine, 1 tablet cOmplete EDTA-free protease inhibitor cocktail, and 0.05% digitonin) containing pAG-Tn5 (Epicypher, 15-1017), and placed on a nutator at room temperature for 60 min. Following pA-Tn5 binding, bead-bound cells were washed three times in 1 ml digitonin-300 buffer, resuspended in 125 µl transposition buffer (10 mM $MgCl_2$ in digitonin-300 buffer), and incubated at 37°C for 60 min. Following transposition, 4.2 µl 0.5 M EDTA, 1.25 µl 10% SDS, and 1.1 µl 20 mg/ml proteinase K (Invitrogen, AM2546) were added and bead-bound cells were incubated at 55°C for 60 min. Bead-bound cells were placed on a magnetic stand and DNA-containing supernatant was cleaned using ChIP DNA Clean & Concentrator (Zymo Research, D5201). Transposed libraries were amplified in 50 µl PCRs (25 µl NEBNext HiFi 2× PCR master mix, 2 µl 10 µM i7 index primer and 2 µl 10 µM i5 index primer[57]) using the following cycling parameters: 58°C for 5 min; 72°C for 5 min; 98°C for 45 s; 13 cycles of 98°C for 15 s, 60°C for 10 s; and 72°C for 1 min. Amplified transposed libraries were purified by SPRI selection (1.3×), quantified using Agilent Bioanalyzer, and pooled at equimolar concentrations. Purified libraries were sequenced on Illumina NovaSeq 6000 (PE150).

CUT&Tag analysis

Paired-end sequencing reads were trimmed using Cutadapt (-q 20 --minimum-length 20) and aligned to panTro6 using Bowtie 2 (--very-sensitive -X 700 -k 10). SAM files were converted to BAM format while discarding alignments with MAPQ < 15, sorted by position or read name, and indexed using SAMtools (-q 15). pA-Tn5-accessible regions were called following PCR duplicate removal using Genrich (-j -r -e chrM -q 0.05). For each primary antibody, pA-Tn5-accessible regions were intersected using BEDTools intersect and only regions common to both iPS cell lines were retained for analysis. BAM files were converted to bedGraph format using deepTools (--binSize 1 --ignoreDuplicates) and visualized using SparK. To identify intersections between CUT&Tag and hDels, at least half of a pA-Tn5-accessible region was required to intersect a hDel.

Mouse ENCODE analysis

Mouse ENCODE ATAC-seq and H3K27ac and p300 ChIP-seq datasets from E14.5 forebrain, limb, heart, lung, and liver tissues were downloaded from (https://www.encodeproject.org/). For all datasets,

the bigWig file “fold-change over control, isogenic replicates 1,2” was used. bigWig files were converted to bedGraph format using UCSC bigWigToBedGraph and visualized using SparK.

snATAC-seq of PCD80 rhesus macaque prefrontal cortex

Prefrontal cortex (PFC) was microdissected from post-conception day 80 (PCD80) rhesus macaque cortex[64]. Tissue was enzymatically dissociated in papain (Worthington, LK003176) containing DNase I for 30 min at 37°C and gently triturated to form a single-cell suspension. Cells were washed and resuspended in cold PBS supplemented with 0.04% BSA. Following counting, 100,000 cells were resuspended in 100 µl cold lysis buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.01% digitonin, 0.1% Tween-20, 0.1% NP40, and 1% BSA) by pipetting three times and incubated on ice for 3 min. Following lysis, 1 ml cold wash buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.1% Tween-20, and 1% BSA) was added and nuclei were centrifuged at 500×*g* for 5 min at 4 °C. Nuclei were partitioned into GEMs using the 10x Genomics Chromium Controller and transposed libraries were generated by following the 10x Genomics Chromium Chromium Next GEM Single Cell ATAC Reagent Kits v1.1 User Guide (CG000209 Rev G). Amplified transposed libraries were quantified using Agilent Bioanalyzer and sequenced on Illumina NovaSeq 6000 (51x12x24x51).

snATAC-seq analysis

Paired-end sequencing reads were aligned to rheMac10 (BSgenome.Mmulatta.UCSC.rheMac10, TxDb.Mmulatta.UCSC.rheMac10.refGene) using Cell Ranger ATAC (v. 2.1.0) and processed with ArchR[65] (v. 1.0.2). Cells with fewer than 3,000 fragments or a transcription start site enrichment score less than 20 were filtered from the dataset (n = 5,166 cells retained, median 9,142 fragments per cell, median 31.8 TSS enrichment score). Cellular subsets were identified by iterative latent semantic indexing dimensionality reduction followed by graph-based clustering using ArchR. Fragments were then summed across cells within each cluster using pycisTopic[66] (v. 1.0.3) and visualized using SparK.

Data Availability and Code Availability

Paired-end sequencing reads (FASTQ) are deposited on SRA under BioProject PRJNA100279.

Notebooks implementing analyses are available at https://github.com/tdfair.

Acknowledgements

We thank Nadav Ahituv, Irene Chen, Jiapei Chen, Jingwen Ding, Clif Duhn, Luke Gilbert, Max Haeussler, Craig Lowe, Aaron McKenna, Katie Pollard, Joseph Replogle, Demian Sainz, Matthew Schmitz, David Shin, Noam Teyssier, and members of the Pollen group for valuable discussions. This work was supported by the following funding sources: Ruth L. Kirschstein National Research Service Predoctoral Fellowship Award F31 HG011569-01A1 (TF), Weill Neurohub Fellowship (NKS), National Institutes of Health DP2MH122400-01, Schmidt Futures Foundation, Shurl and Kay Curci Foundation Innovative Genomics Institute Award. AAP is a New York Stem Cell Foundation Robertson Investigator.

Author Contributions

T.F. performed experiments and analyzed data. B.J.P performed neuroepithelial cell differentiations. N.K.S performed enrichment analyses. T.F. and A.A.P conceived the study and wrote the manuscript with input from all authors.

References:

1. Olson, M. V. When less is more: gene loss as an engine of evolutionary change. *Am. J. Hum. Genet.* **64**, 18–23 (1999).
2. McLean, C. Y. *et al.* Human-specific loss of regulatory DNA and the evolution of human-specific traits. *Nature* **471**, 216–219 (2011).
3. Kronenberg, Z. N. *et al.* High-resolution comparative analysis of great ape genomes. *Science* **360**, eaar6343 (2018).
4. Xue, J. R. *et al.* The functional and evolutionary impacts of human-specific deletions in conserved elements. *Science* **380**, eabn2253 (2023).
5. Kimura, M. The neutral theory of molecular evolution: a review of recent evidence. *Jpn. J. Genet.* **66**, 367–386 (1991).
6. Ebert, P. *et al.* Haplotype-resolved diverse human genomes and integrated analysis of structural variation. *Science* **372**, (2021).
7. Abel, H. J. *et al.* Mapping and characterization of structural variation in 17,795 human genomes. *Nature* **583**, 83–89 (2020).
8. Chan, Y. F. *et al.* Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. *Science* **327**, 302–305 (2010).
9. Davis, E. S. *et al.* matchRanges: generating null hypothesis genomic ranges via covariate-matched sampling. *Bioinformatics* **39**, (2023).
10. Corces, M. R. *et al.* An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* **14**, 959–962 (2017).
11. Kaya-Okur, H. S. *et al.* CUT&Tag for efficient epigenomic profiling of small samples and single

cells. *Nat. Commun.* **10**, 1930 (2019).

12. Cerbini, T. *et al.* Transcription activator-like effector nuclease (TALEN)-mediated CLYBL targeting enables enhanced transgene expression and one-step generation of dual reporter human induced pluripotent stem cell (iPSC) and neural stem cell (NSC) lines. *PLoS One* **10**, e0116032 (2015).
13. Mannion, B. J. *et al.* Uncovering Hidden Enhancers Through Unbiased In Vivo Testing. *bioRxiv* 2022.05.29.493901 (2022) doi:10.1101/2022.05.29.493901.
14. Percharde, M., Bulut-Karslioglu, A. & Ramalho-Santos, M. Hypertranscription in Development, Stem Cells, and Regeneration. *Dev. Cell* **40**, 9–21 (2017).
15. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
16. Gasperini, M. *et al.* A Genome-wide Framework for Mapping Gene Regulation via Cellular Genetic Screens. *Cell* **176**, 377–390.e19 (2019).
17. Nuñez, J. K. *et al.* Genome-wide programmable transcriptional memory by CRISPR-based epigenome editing. *Cell* **184**, 2503–2519.e17 (2021).
18. Jackson, C. B. *et al.* A variant in MRPS14 (uS14m) causes perinatal hypertrophic cardiomyopathy with neonatal lactic acidosis, growth retardation, dysmorphic features and neurological involvement. *Hum. Mol. Genet.* **28**, 639–649 (2019).
19. Kaji, K. *et al.* The NuRD component Mbd3 is required for pluripotency of embryonic stem cells. *Nat. Cell Biol.* **8**, 285–292 (2006).
20. Kanton, S. *et al.* Organoid single-cell genomic atlas uncovers human-specific features of brain development. *Nature* **574**, 418–422 (2019).
21. Geschwind, D. H. & Rakic, P. Cortical evolution: judge the brain by its cover. *Neuron* **80**, 633–647 (2013).
22. Song, J. H. T. *et al.* Genetic studies of human-chimpanzee divergence using stem cell fusions. *Proc. Natl. Acad. Sci. U. S. A.* **118**, (2021).

23. Gallego Romero, I. *et al.* A panel of induced pluripotent stem cells from chimpanzees: a resource for comparative functional genomics. *Elife* **4**, e07103 (2015).
24. Fulco, C. P. *et al.* Systematic mapping of functional enhancer-promoter connections with CRISPR interference. *Science* **354**, 769–773 (2016).
25. Tang, X., Benesch, M. G. K. & Brindley, D. N. Lipid phosphate phosphatases and their roles in mammalian physiology and pathology. *J. Lipid Res.* **56**, 2048–2060 (2015).
26. Tomsig, J. L. *et al.* Lipid phosphate phosphohydrolase type 1 (LPP1) degrades extracellular lysophosphatidic acid in vivo. *Biochem. J* **419**, 611–618 (2009).
27. Taddeo, E. P. *et al.* Lysophosphatidic acid counteracts glucagon-induced hepatocyte glucose production via STAT3. *Sci. Rep.* **7**, 127 (2017).
28. Medelnik, J.-P. *et al.* Signaling-Dependent Control of Apical Membrane Size and Self-Renewal in Rosette-Stage Human Neuroepithelial Stem Cells. *Stem Cell Reports* **10**, 1751–1765 (2018).
29. He, Y. *et al.* Improved regulatory element prediction based on tissue-specific local epigenomic signatures. *Proceedings of the National Academy of Sciences* **114**, E1633–E1640 (2017).
30. Chattaragada, M. S. *et al.* FAM49B, a novel regulator of mitochondrial function and integrity that suppresses tumor metastasis. *Oncogene* **37**, 697–709 (2018).
31. Fort, L. *et al.* Fam49/CYRI interacts with Rac1 and locally suppresses protrusions. *Nat. Cell Biol.* **20**, 1159–1171 (2018).
32. Whalen, S. *et al.* Machine learning dissection of human accelerated regions in primate neurodevelopment. *Neuron* (2023) doi:10.1016/j.neuron.2022.12.026.
33. Fair, T. & Pollen, A. A. Genetic architecture of human brain evolution. *Curr. Opin. Neurobiol.* **80**, 102710 (2023).
34. Klein, J. C. *et al.* A systematic evaluation of the design and context dependencies of massively parallel reporter assays. *Nat. Methods* **17**, 1083–1091 (2020).
35. Cooper, Y. A. *et al.* Functional regulatory variants implicate distinct transcriptional networks in dementia. *Science* **377**, eabi8654 (2022).

36. Morris, J. A. *et al.* Discovery of target genes and pathways of blood trait loci using pooled CRISPR screens and single cell RNA sequencing. *bioRxiv* 2021.04.07.438882 (2021) doi:10.1101/2021.04.07.438882.

37. Prescott, S. L. *et al.* Enhancer Divergence and cis-Regulatory Evolution in the Human and Chimp Neural Crest. *Cell* **163**, 68–83 (2015).

38. Pollen, A. A. *et al.* Establishing Cerebral Organoids as Models of Human-Specific Brain Evolution. *Cell* **176**, 743–756.e17 (2019).

39. Przybyla, L. & Gilbert, L. A. A new era in functional genomics screens. *Nat. Rev. Genet.* **23**, 89–103 (2022).

40. Tycko, J. *et al.* Mitigation of off-target toxicity in CRISPR-Cas9 screens for essential non-coding elements. *Nat. Commun.* **10**, 4063 (2019).

41. Gasperini, M. *et al.* CRISPR/Cas9-Mediated Scanning for Regulatory Elements Required for HPRT1 Expression via Thousands of Large, Programmed Genomic Deletions. *The American Journal of Human Genetics* vol. 101 192–205 Preprint at https://doi.org/10.1016/j.ajhg.2017.06.010 (2017).

42. Anzalone, A. V., Koblan, L. W. & Liu, D. R. Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. *Nat. Biotechnol.* **38**, 824–844 (2020).

43. Erwood, S. *et al.* Saturation variant interpretation using CRISPR prime editing. *Nat. Biotechnol.* **40**, 885–895 (2022).

44. She, R. *et al.* Comparative landscape of genetic dependencies in human and chimpanzee stem cells. *bioRxiv* (2023) doi:10.1101/2023.03.19.533346.

45. Gilbert, L. A. *et al.* Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell* **159**, 647–661 (2014).

46. Mu, W. *et al.* bootRanges: flexible generation of null sets of genomic ranges for hypothesis testing. *Bioinformatics* **39**, (2023).

47. Pockrandt, C., Alzamel, M., Iliopoulos, C. S. & Reinert, K. GenMap: ultra-fast computation of

genome mappability. *Bioinformatics* **36**, 3687–3692 (2020).

48. Lareau, C. A. *et al.* Massively parallel single-cell mitochondrial DNA genotyping and chromatin profiling. *Nat. Biotechnol.* **39**, 451–461 (2021).
49. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
50. Fiddes, I. T. *et al.* Comparative Annotation Toolkit (CAT)-simultaneous clade and personal genome annotation. *Genome Res.* **28**, 1029–1038 (2018).
51. Mao, Y. *et al.* A high-quality bonobo genome refines the analysis of hominid evolution. *Nature* **594**, 77–81 (2021).
52. McKenna, A. & Shendure, J. FlashFry: a fast and flexible tool for large-scale CRISPR target design. *BMC Biol.* **16**, 74 (2018).
53. Li, W. *et al.* MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol.* **15**, 554 (2014).
54. Sina Booeshaghi, A., Min, K. H. (joseph), Gehring, J. & Pachter, L. Quantifying orthogonal barcodes for sequence census assays. *bioRxiv* 2022.10.09.511501 (2022) doi:10.1101/2022.10.09.511501.
55. Melsted, P. *et al.* Modular, efficient and constant-memory single-cell RNA-seq preprocessing. *Nat. Biotechnol.* **39**, 813–818 (2021).
56. Replogle, J. M. *et al.* Combinatorial single-cell CRISPR screens by direct guide RNA capture and targeted sequencing. *Nat. Biotechnol.* **38**, 954–961 (2020).
57. Buenrostro, J. D. *et al.* Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486–490 (2015).
58. Chen, Y. *et al.* A Versatile Polypharmacology Platform Promotes Cytoprotection and Viability of Human Pluripotent and Differentiated Cells. *bioRxiv* 815761 (2019).
59. Varga, B. V. *et al.* Signal requirement for cortical potential of transplantable human neuroepithelial stem cells. *Nat. Commun.* **13**, 2844 (2022).

60. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12 (2011).

61. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

62. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

63. Ramírez, F., Dündar, F., Diehl, S., Grüning, B. A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* **42**, W187–91 (2014).

64. Schmitz, M. T. *et al.* The development and evolution of inhibitory neurons in primate cerebrum. *Nature* **603**, 871–877 (2022).

65. Granja, J. M. *et al.* ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat. Genet.* **53**, 403–411 (2021).

66. Bravo González-Blas, C. *et al.* SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks. *Nat. Methods* (2023) doi:10.1038/s41592-023-01938-4.

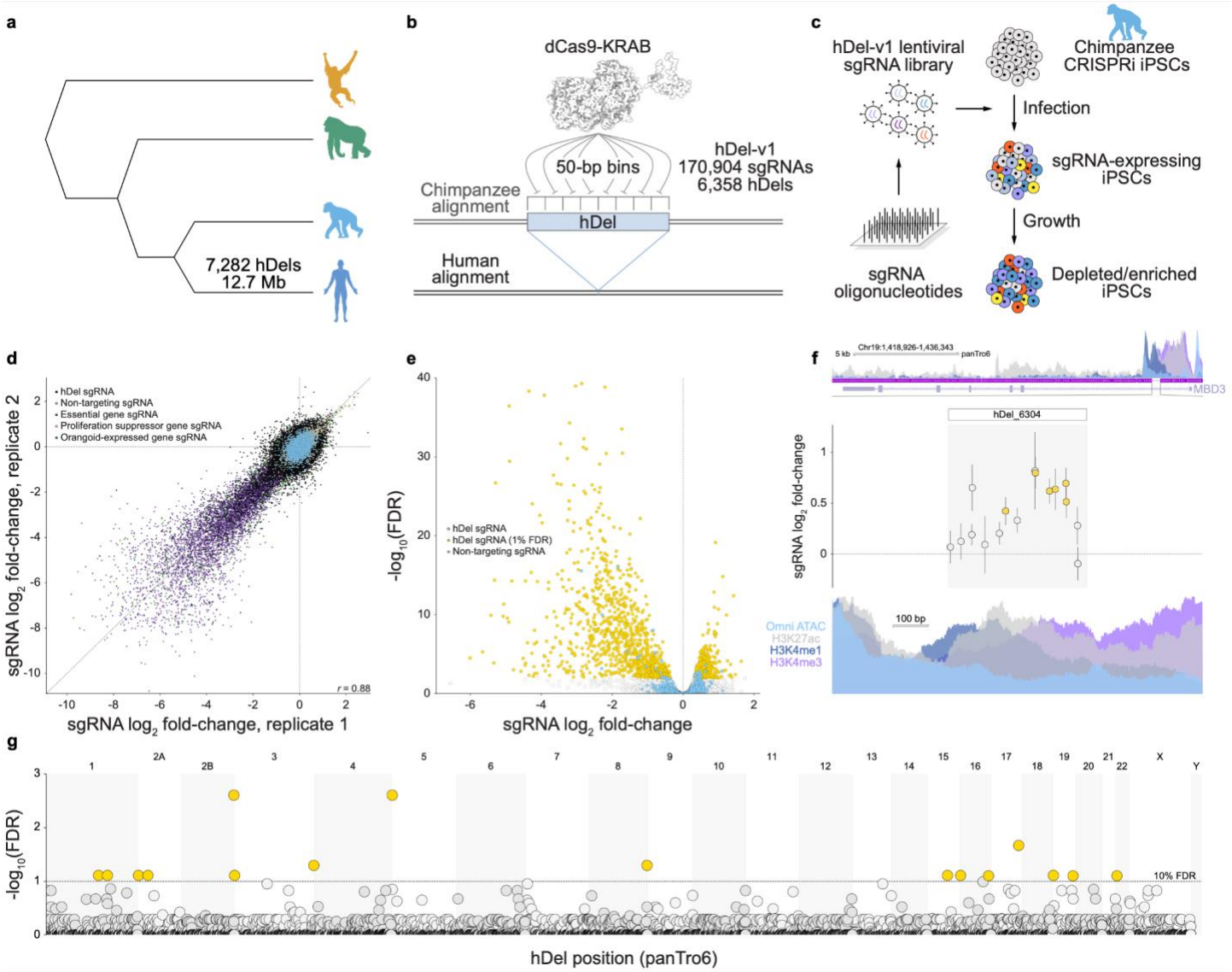


Figure 1. Genome-scale tiling CRISPRi-based genetic screens identify hDels modifying cellular proliferation

a, Great ape cladogram. The number of deletions assigned to the human lineage[3] and the number of base pairs removed are labeled.
b, CRISPRi-based tiling of hDels (hDel-v1). hDel-v1 sgRNAs were assigned to and selected from 50-bp genomic windows.
c, hDel-v1 screening approach in chimpanzee iPS cells.
d, Scatterplot of sgRNA $\log_2$ fold-change for hDel-v1 technical replicates in C3624K.
e, Volcano plot of hDel-targeting and non-targeting sgRNA $\log_2$ fold-change and DESeq2 Benjamini-Hochberg-adjusted *p*-value.
f, hDel_6304-targeting sgRNA $\log_2$ fold-change (gold, FDR < 0.05) and *MBD3* Omni ATAC-seq, H3K4me1, H3K4me3, and H3K27ac in C3624K.
g, Manhattan plot of hDel position in the chimpanzee reference genome (panTro6) and α-RRA Benjamini-Hochberg-adjusted *p*-value for 500-bp hDel genomic windows (gold, FDR < 0.1).

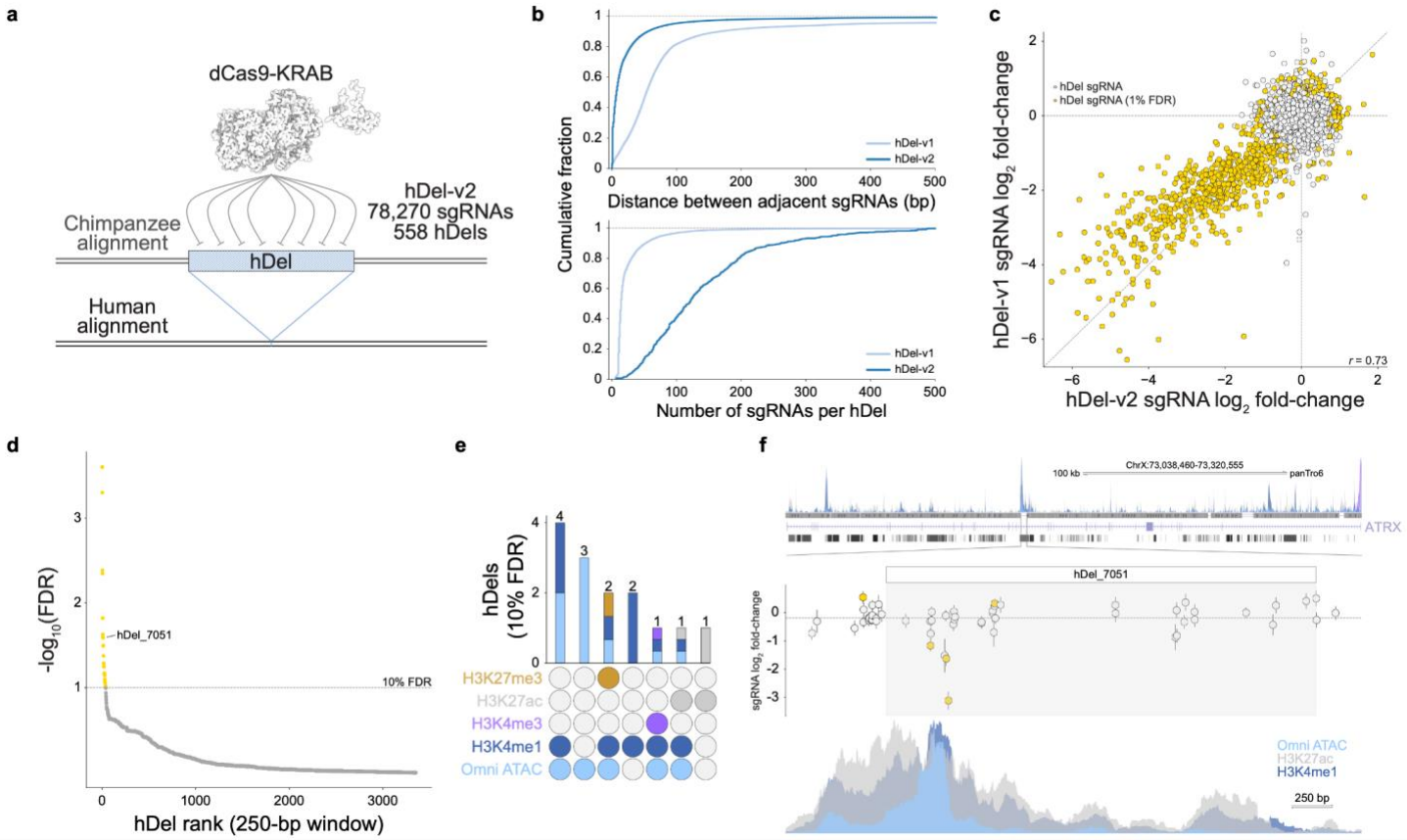


Figure 2. High-density tiling CRISPRi screen refines the boundaries of functional sequences within proliferation-modifying hDels

a, High-density CRISPRi-based tiling of hDels (hDel-v2).
b, Distribution of the distance between adjacent hDel-targeting sgRNAs (top) and the number of sgRNAs per hDel (bottom) for hDel-v1 and hDel-v2.
c, Scatterplot of sgRNA $\log_2$ fold-change for hDel-targeting sgRNAs screened in hDel-v1 and hDel-v2 (n = 18,148 sgRNAs).
d, Identification of proliferation-modifying hDels. 250-bp hDel genomic windows are ranked by α-RRA Benjamini-Hochberg-adjusted p-value (gold, FDR < 0.1).
e, Upset plot of Omni-ATAC seq, H3K4me1, H3K4me3, H3K27ac, and H3K27me3 intersecting hDels (FDR < 0.1) in C3624K.
f, hDel_7051-targeting sgRNA $\log_2$ fold-change (gold, FDR < 0.05) and *ATRX* Omni-ATAC seq, H3K4me1, H3K4me3, and H3K27ac in C3624K.

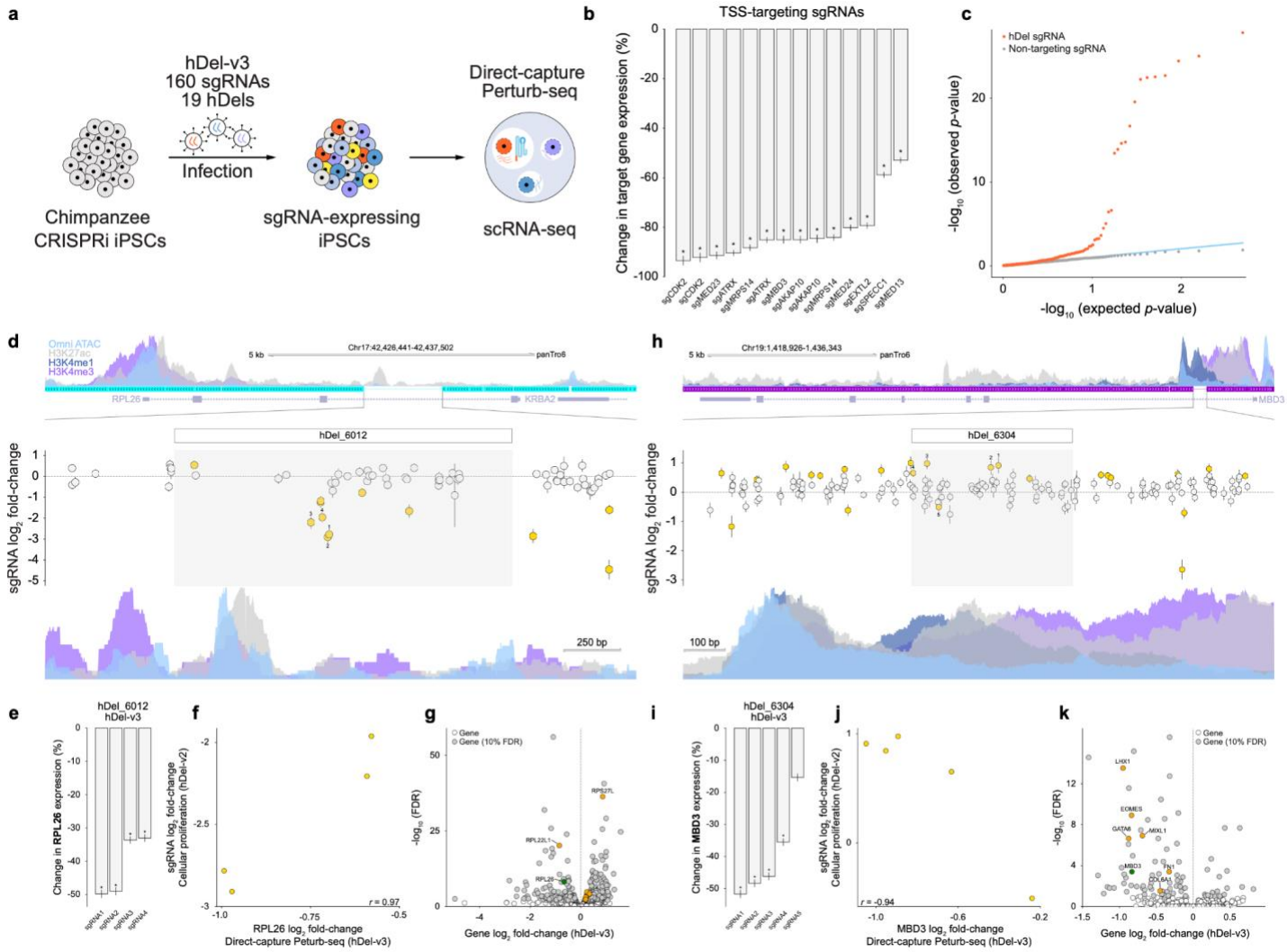


Figure 3. Mapping *cis*- and *trans*-regulatory target genes of proliferation-modifying hDels

a, Single-cell CRISPRi screening approach (hDel-v3).

b, Differential gene expression for cells ('pseudobulk') harboring the indicated transcription start site-targeting sgRNA (*FDR < 0.1).

c, Distributions of observed and expected (uniform) *p*-values for *cis* differential expression for hDel-targeting sgRNA-gene pairs (orange) and non-targeting sgRNA-gene pairs (gray, downsampled). Blue line: observed *p*-value = expected *p*-value.

d, hDel_6012-targeting sgRNA $\log_2$ fold-change (hDel-v2; gold, FDR < 0.05) and *RPL26* Omni-ATAC seq, H3K4me1, H3K4me3, and H3K27ac in C3624K.

e, Differential *RPL26* expression for cells harboring the indicated hDel_6012-targeting sgRNA (*FDR < 0.1).

f, Scatterplot of hDel_6012-targeting sgRNA $\log_2$ fold-change (cellular proliferation, hDel-v2) and *RPL26* $\log_2$ fold-change (gene expression, hDel-v3).

g, *trans* differential expression for cells harboring hDel_6012-targeting sgRNAs (gene $\log_2$ fold-change and Benjamini-Hochberg-adjusted *p*-value). Green: *cis* target gene; orange: ribosomal genes (FDR < 0.1).

h, hDel_6304-targeting sgRNA $\log_2$ fold-change (hDel-v2; gold, FDR < 0.05) and *MBD3* Omni-ATAC seq, H3K4me1, H3K4me3, and H3K27ac in C3624K.

i, Differential *MBD3* expression for cells harboring the indicated hDel_6304-targeting sgRNA (*FDR < 0.1).
j, Scatterplot of hDel_6304-targeting sgRNA $\log_2$ fold-change (cellular proliferation, hDel-v2) and *MBD3* $\log_2$ fold-change (gene expression, hDel-v3).
k, *trans* differential expression for cells harboring hDel_6304-targeting sgRNAs (gene $\log_2$ fold-change and Benjamini-Hochberg-adjusted *p*-value). Green: *cis* target gene; orange: meso-endoderm (FDR < 0.1).

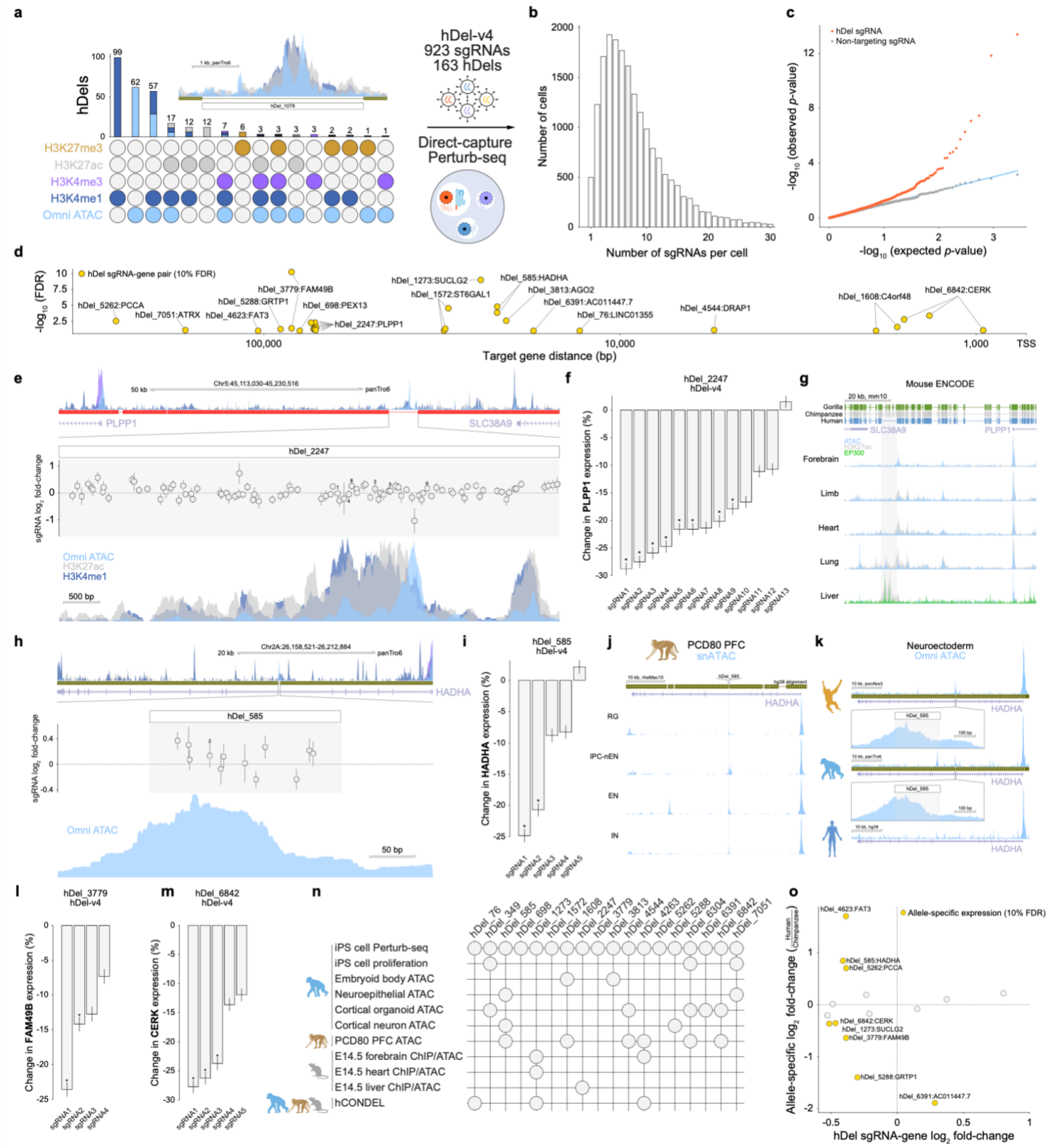


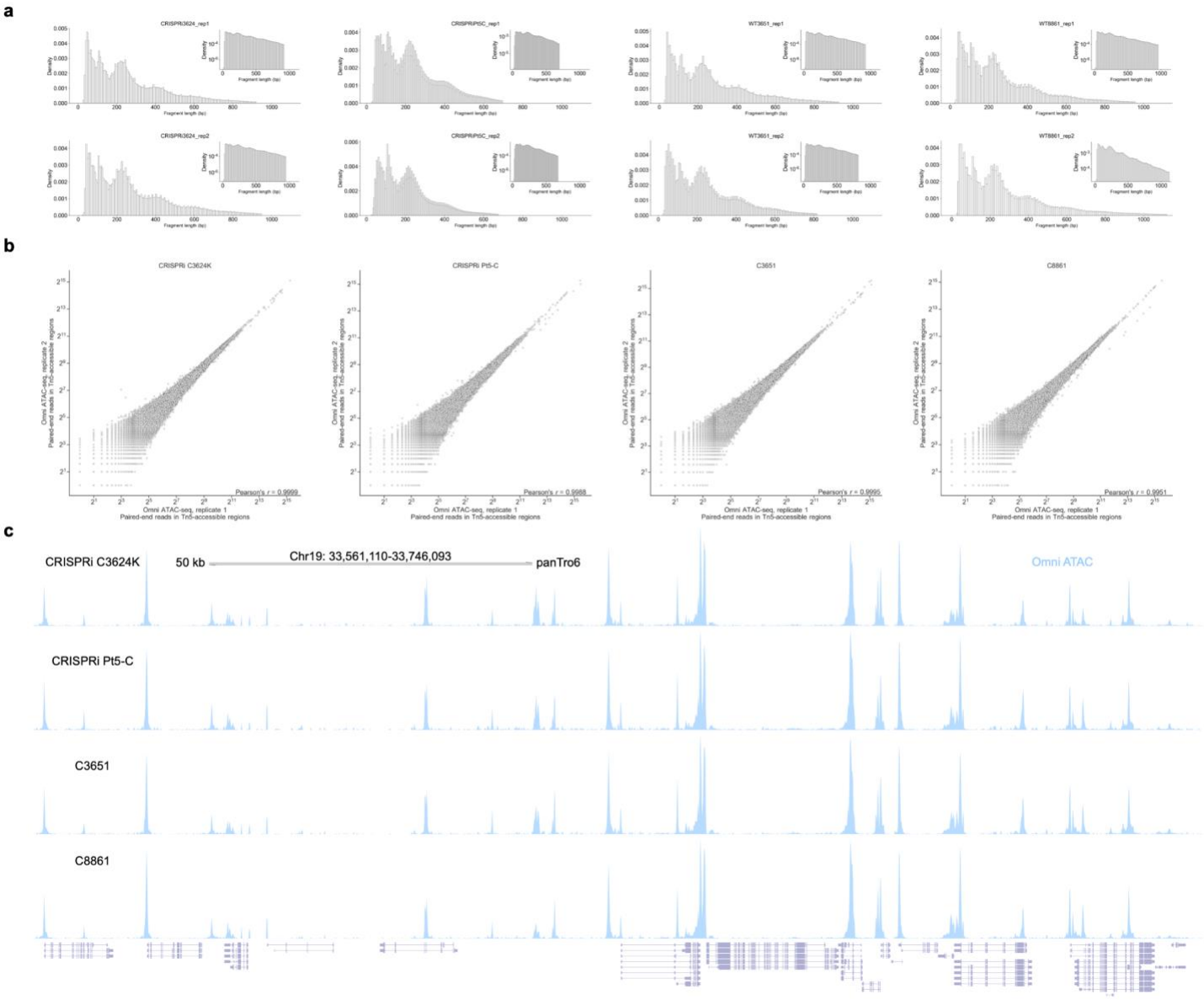
Figure 4. Nonessential hDels harbor *cis*-regulatory elements

a, Upset plot of Omni ATAC-seq, H3K4me1, H3K4me3, H3K27ac, and H3K27me3 intersecting hDels and high multiplicity of infection single-cell CRISPRi screening approach (hDel-v4). Inset: hDel_1078.

b, Distribution of the number of sgRNAs per cell (hDel-v4).

c, Distributions of observed and expected (uniform) *p*-values for *cis* differential expression for hDel-targeting sgRNA-gene pairs (orange) and non-targeting sgRNA-gene pairs (gray, downsampled). Blue line: observed *p*-value = expected *p*-value.
d, Distance between hDel-targeting sgRNA and TSS for corresponding *cis* target gene (FDR < 0.1).
e, hDel_2247-targeting sgRNA $\log_2$ fold-change (hDel-v1) and Omni ATAC-seq, H3K4me1, H3K4me3, and H3K27ac in C3624K.
f, Differential *PLPP1* expression for cells ('pseudobulk') harboring the indicated hDel_2247-targeting sgRNA (*FDR < 0.1).
g, Mouse ENCODE ATAC-seq, p300, and H3K27ac in E14.5 forebrain, limb, heart, lung, and liver tissues. Shaded region: hDel_2247 orthologous sequence.
h, hDel_585-targeting sgRNA $\log_2$ fold-change (hDel-v1) and Omni ATAC-seq, H3K4me1, H3K4me3, and H3K27ac in C3624K.
i, Differential *HADHA* expression for cells harboring the indicated hDel_585-targeting sgRNA (*FDR < 0.1).
j, snATAC-seq of PCD80 rhesus macaque prefrontal cortex at the *HADHA* gene body. RG: radial glia, IPC-nEN: intermediate progenitor cell-newborn excitatory neuron, EN: excitatory neuron, IN: inhibitory neuron. Shaded region: hDel_585 orthologous sequence.
k, Omni ATAC-seq in iPS cell-derived neuroepithelial cells from orangutan (top), chimpanzee (middle), and human (bottom) at the *HADHA* gene body. Shaded region: hDel_585 orthologous sequence.
l, Differential *FAM49B* expression for cells harboring the indicated hDel_3779-targeting sgRNA (*FDR < 0.1).
m, Differential *CERK* expression for cells harboring the indicated hDel_6842-targeting sgRNA (*FDR < 0.1).
n, Intersection of hDels with identified *cis* target genes in chimpanzee iPS cells with ATAC-seq and ChIP-seq from chimpanzee, rhesus macaque, and mouse, and hCONDELs.
o, Scatterplot of allele-specific gene $\log_2$ fold-change in human-chimpanzee allotetraploid iPS cells[22] and hDel-targeting sgRNA-gene $\log_2$ fold-change (hDel-v4).

Extended Data

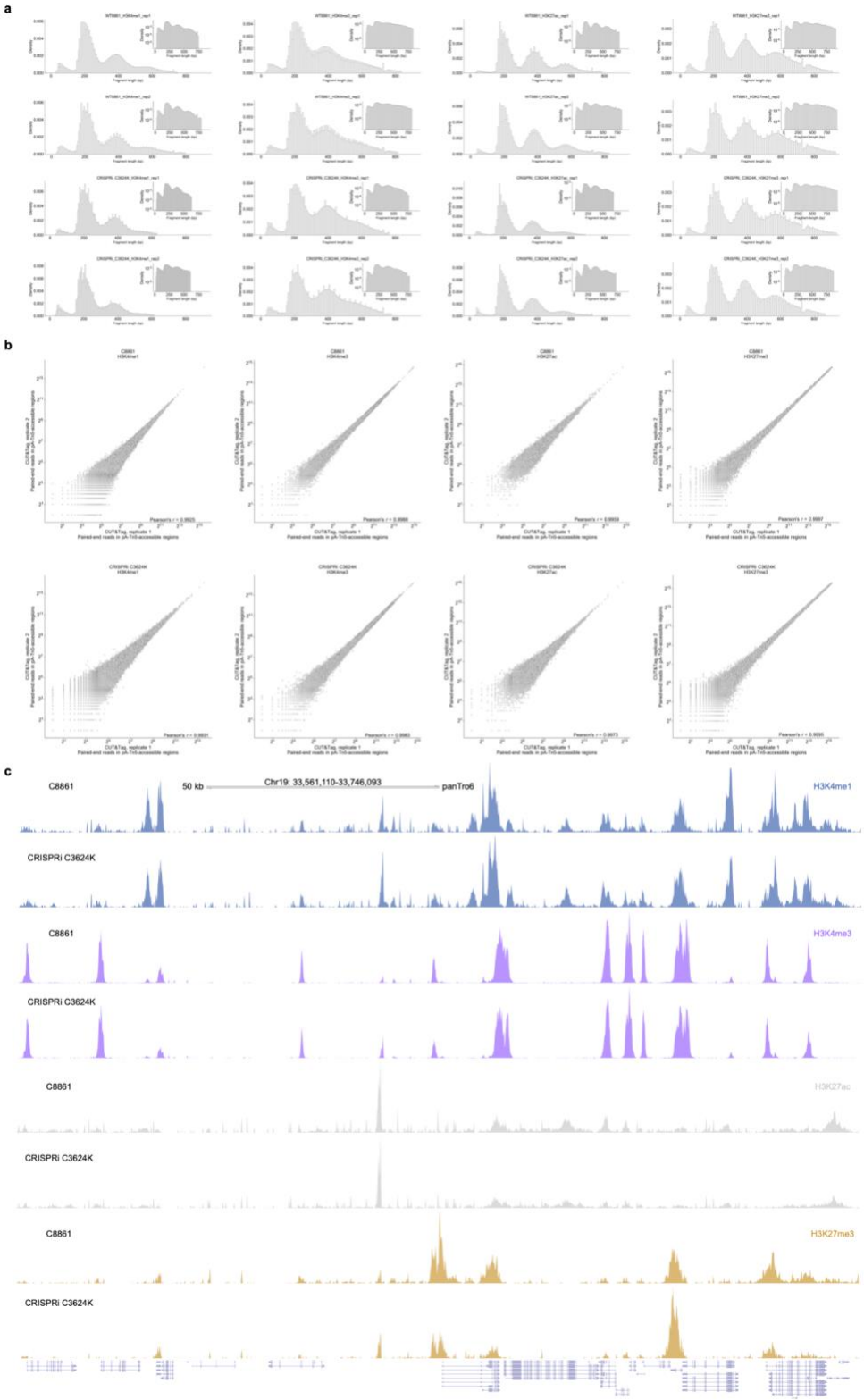


Extended Data Fig. 1 Omni ATAC-seq of chimpanzee iPS cells.

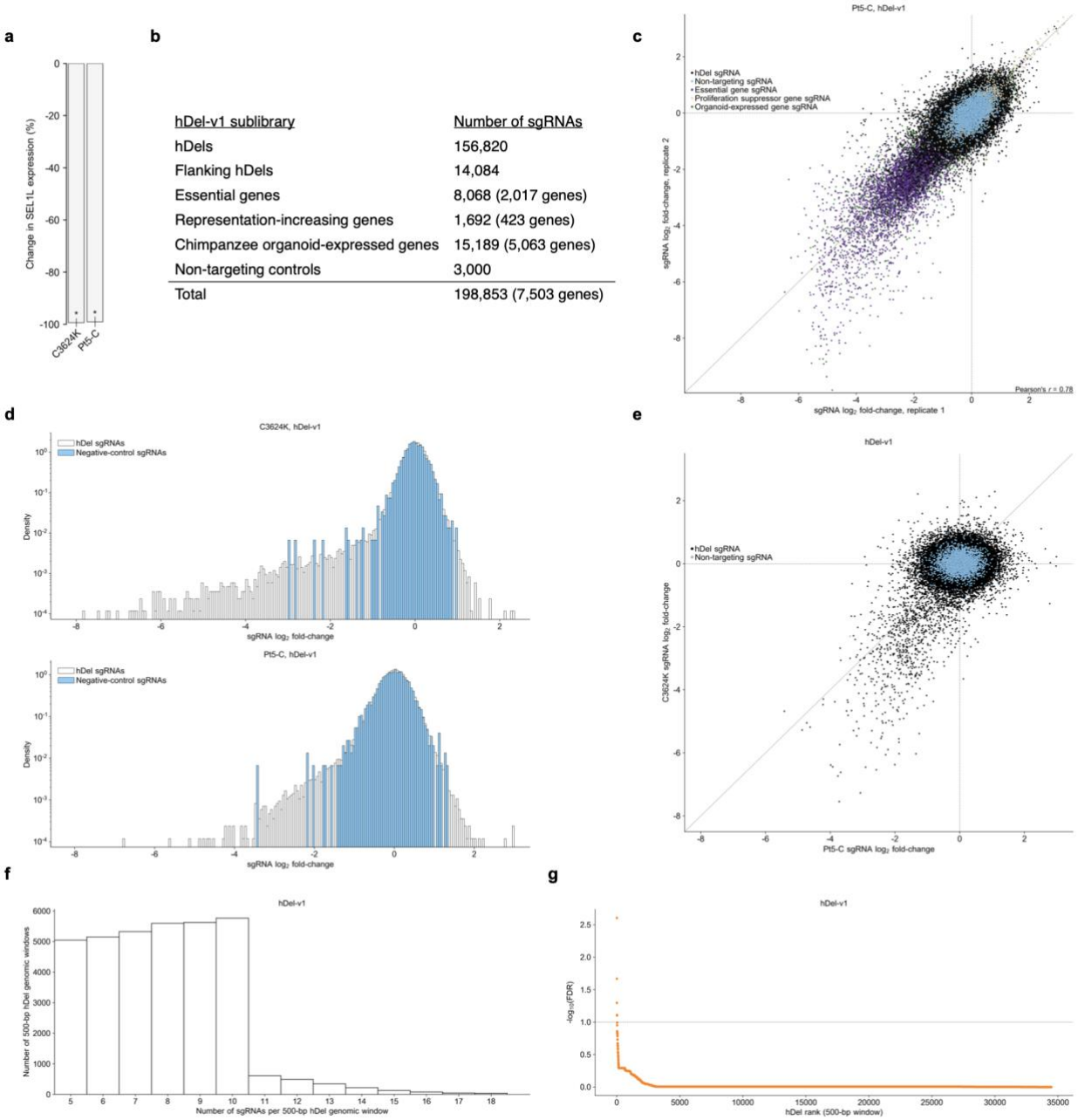
a, Distribution of Omni ATAC-seq fragment sizes. Inset: log-transformed histogram.

b, Correlation of reads within Tn5-accessible regions for Omni ATAC-seq technical replicates. Each point represents a Tn5-accessible region in any of the four iPS cells lines (5% FDR).

c, Omni ATAC-seq across a 184 kb region of the chimpanzee reference genome (panTro6).



Extended Data Fig. 2 Profiling histone modifications with CUT&Tag of chimpanzee iPS cells.
a, Distribution of CUT&Tag fragment sizes for H3K4me1^{ab8895}, H3K4me3^{ab8580}, H3K27ac^{ab4729}, and H3K27me3^{9733S}. Inset: log-transformed histogram.
b, Correlation of reads within pA-Tn5-accessible regions for CUT&Tag technical replicates. Each point represents a pA-Tn5-accessible region in any of the two iPS cells lines (5% FDR).
c, CUT&Tag across a 184 kb region of the chimpanzee reference genome (panTro6).



hDel-v1 sublibrary	Number of sgRNAs
hDels	156,820
Flanking hDels	14,084
Essential genes	8,068 (2,017 genes)
Representation-increasing genes	1,692 (423 genes)
Chimpanzee organoid-expressed genes	15,189 (5,063 genes)
Non-targeting controls	3,000
Total	198,853 (7,503 genes)

Extended Data Fig. 3 hDel-v1 CRISPRi-based genetic screens.

a, Change in *SEL1L* expression for chimpanzee CRISPRi iPS cells (C3624K, Pt5-C) harboring *SEL1L*-targeting sgRNAs compared to non-targeting sgRNAs as assessed by RT-qPCR.

b, hDel-v1 sublibraries.

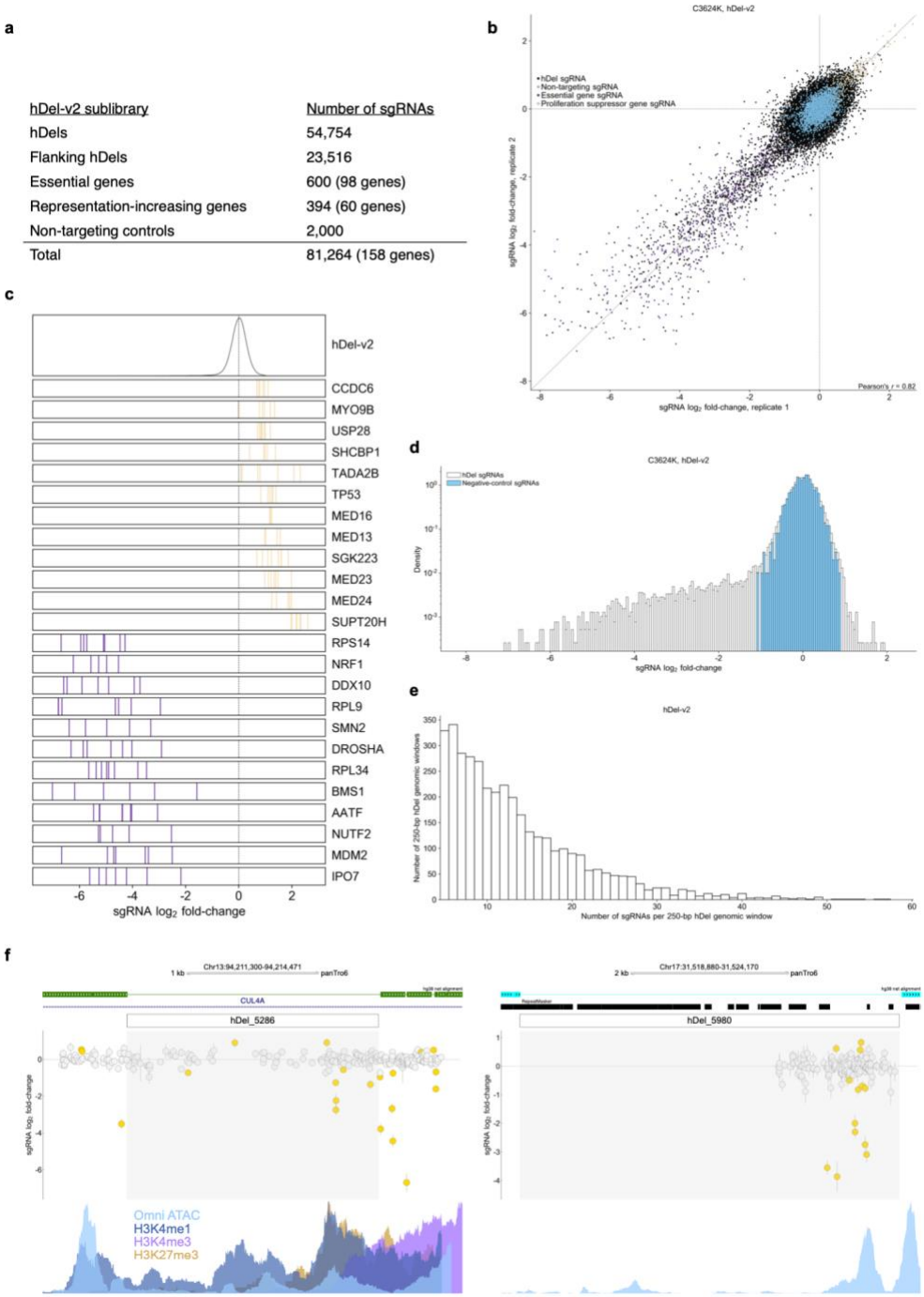
c, Scatterplot of sgRNA $\log_2$ fold-change for hDel-v1 technical replicates in Pt5-C.

d, Distribution of sgRNA $\log_2$ fold-change for hDel-targeting and non-targeting sgRNAs in C3624K (top) and Pt5-C (bottom).

e, Scatterplot of sgRNA $\log_2$ fold-change for hDel-targeting and non-targeting sgRNAs in C3624K and Pt5-C.

f, Distribution of the number of sgRNAs per 500-bp hDel genomic window.

g, 500-bp hDel genomic windows ranked by α-RRA Benjamini-Hochberg-adjusted *p*-value.



Extended Data Fig. 4 hDel-v2 CRISPRi-based genetic screen.

a, hDel-v2 sublibraries.

b, Scatterplot of sgRNA $\log_2$ fold-change for hDel-v2 technical replicates in C3624K.

c, Rug plot of sgRNA $\log_2$ fold-change for the 12 most enriched (khaki) and depleted genes (purple) as ranked by average sgRNA $\log_2$ fold-change. Each vertical line represents a transcription start site-targeting sgRNA.

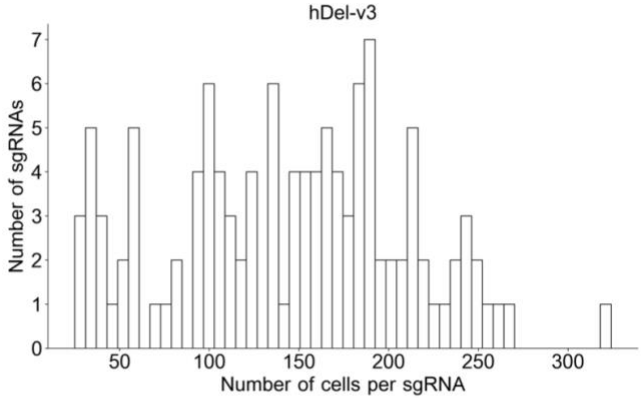
d, Distribution of sgRNA $\log_2$ fold-change for hDel-targeting and non-targeting sgRNAs in C3624K.

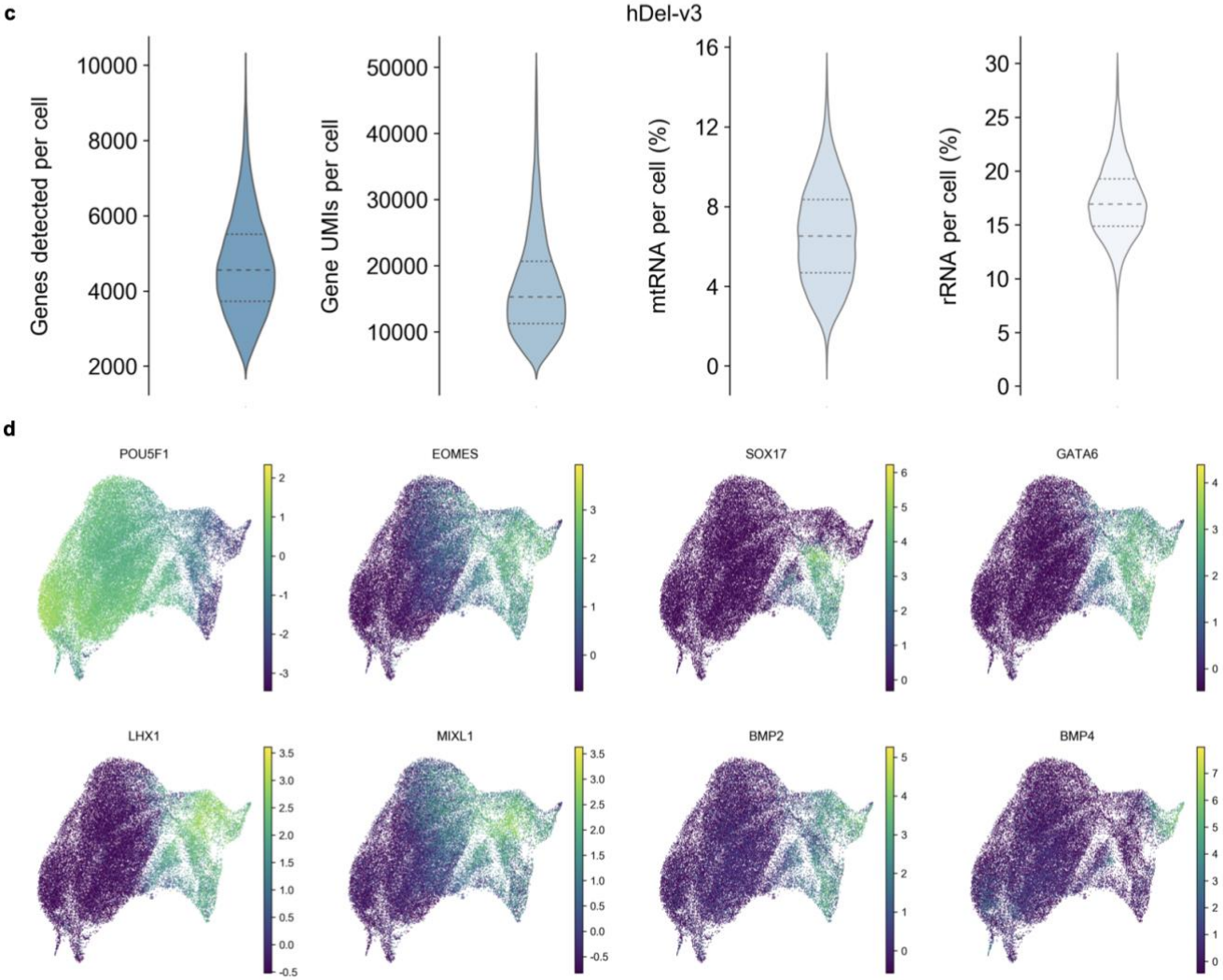
e, Distribution of the number of sgRNAs per 250-bp hDel genomic window.

f, 250-bp hDel genomic windows intersecting epigenetic features. hDel_5286 (left) and hDel_5980 (right).

a

hDel-v3 sublibrary	Number of sgRNAs
hDels	122 (19 hDels)
Transcription start site	18 (12 genes)
K562 *cis*-regulatory element	10 (5 *cis*-regulatory elements)
Non-targeting controls	10
Total	160





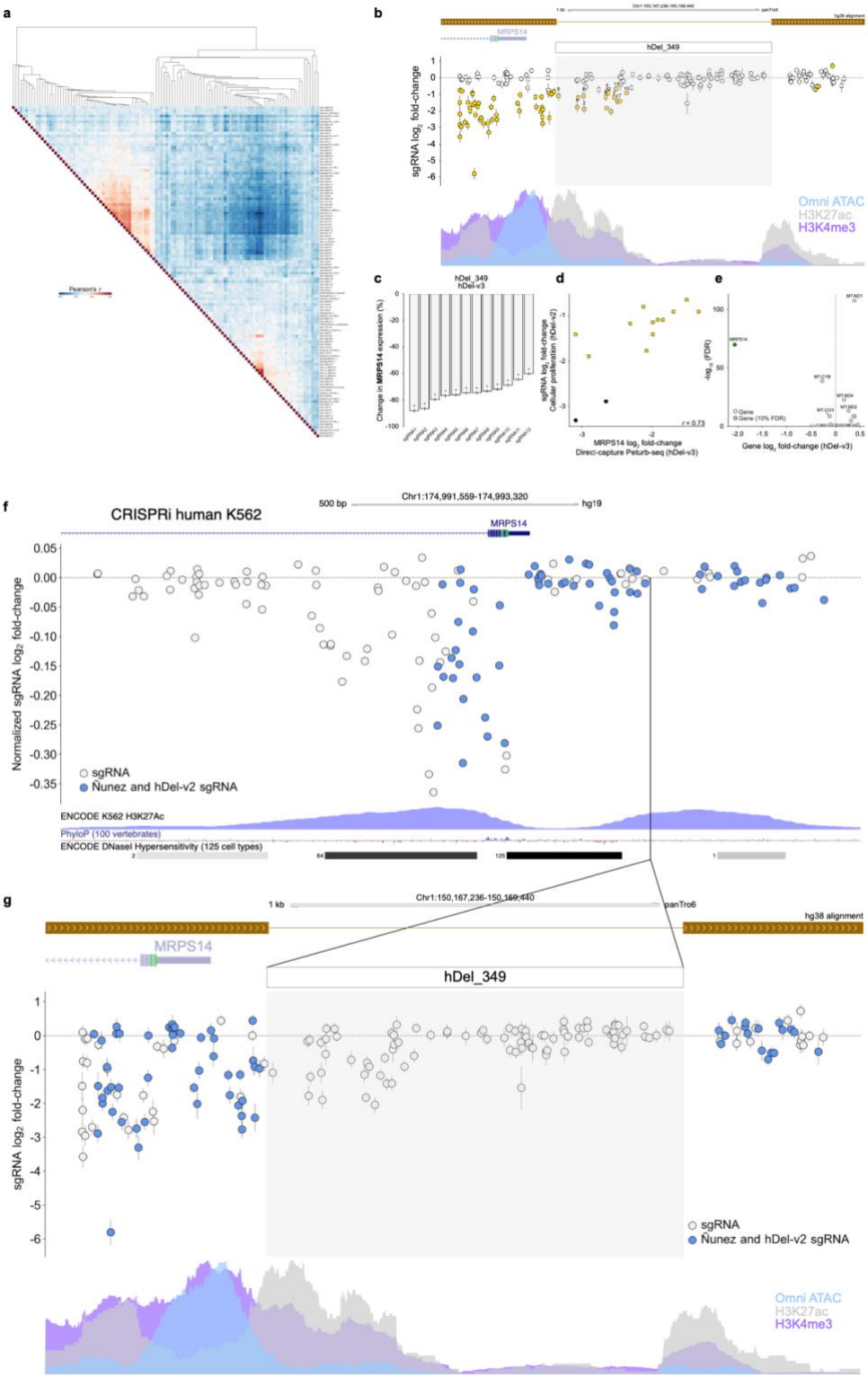
Extended Data Fig. 5 hDel-v3 Direct-capture Perturb-seq.

a, hDel-v3 sublibraries.

b, Distribution of the number of cells per sgRNA.

c, Distribution of the number of genes detected (far left), gene UMIs (center left), percent mtRNA (center right), and percent rRNA (far right) per cell.

d, UMAP projections colored by the normalized, log-transformed, and scaled expression of the indicated genes.



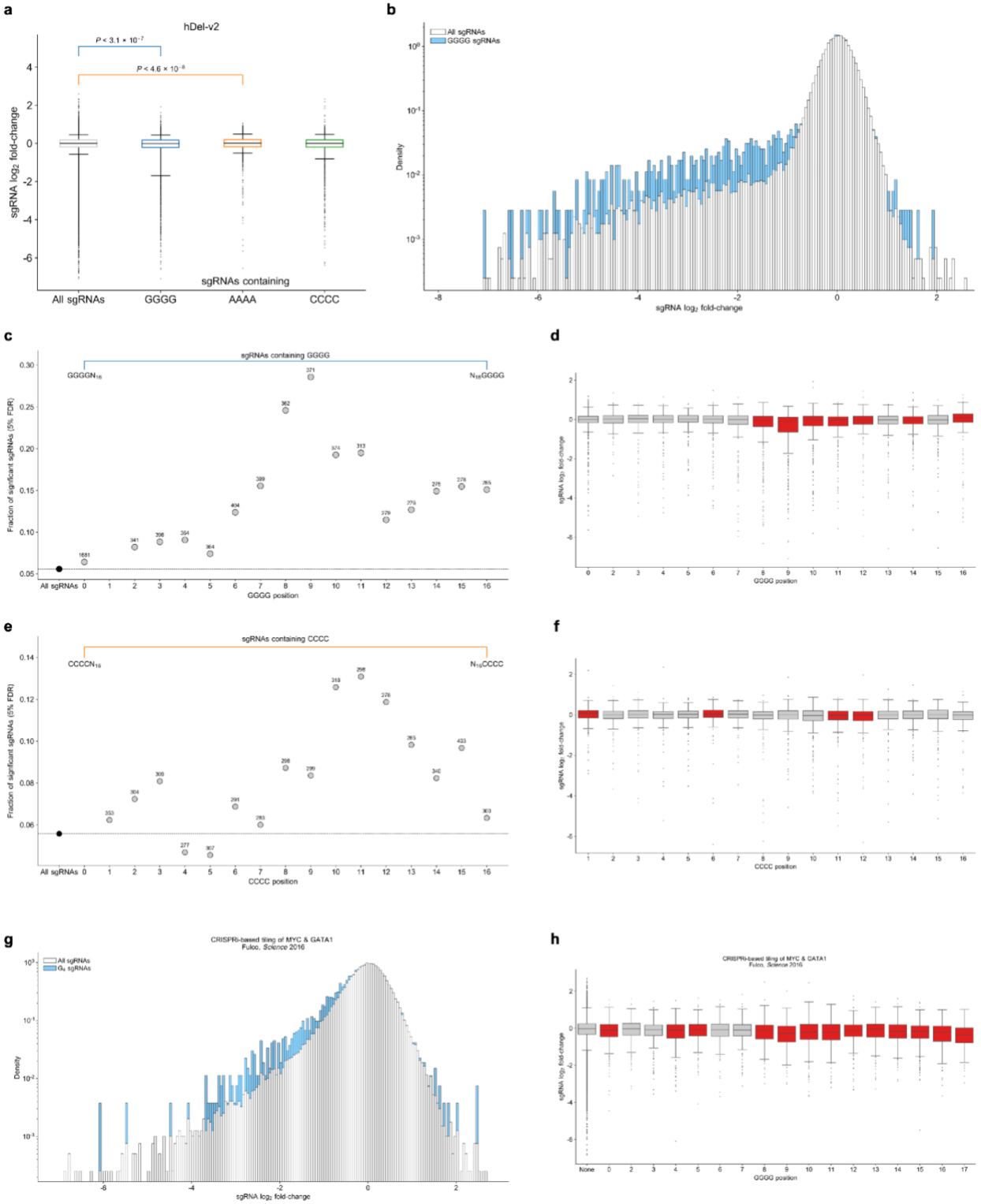
Extended Data Fig. 6 hDel-v3 *trans* differential expression and hDel_349.

a, hDel-v3 sublibraries.

b, Heatmap of Pearson correlations of gene $\log_2$ fold-change and hierarchical clustering for differentially expressed genes. The union of any gene differentially expressed (n = 2,864 genes, FDR < 0.1) in a sgRNA 'pseudobulk' (rows) is used for Pearson correlations and clustering.

c, hDel_349-targeting sgRNA $\log_2$ fold-change (hDel-v2; gold, FDR < 0.05) and *MRPS14* Omni-ATAC seq, H3K4me3, and H3K27ac in C3624K.

d, Differential *MRPS14* expression for cells harboring the indicated hDel_349-targeting sgRNA (*FDR < 0.1).
e, Scatterplot of hDel_349-targeting sgRNA $\log_2$ fold-change (cellular proliferation, hDel-v2) and *MRPS14* $\log_2$ fold-change (gene expression, hDel-v3).
f, *MRPS14* TSS-targeting sgRNA $\log_2$ fold-change in human K562 cells[17] and ENCODE K562 H3K27ac, PhyloP, and ENCODE DNaseI hypersensitivity. sgRNAs in blue are in the Nuñez et al. and hDel-v2 sgRNA libraries.
g, hDel_349-targeting sgRNA $\log_2$ fold-change (hDel-v2) and *MRPS14* Omni-ATAC seq, H3K4me3, and H3K27ac in chimpanzee iPS cells (C3624K). sgRNAs in blue are in the Nuñez et al. and hDel-v2 sgRNA libraries.

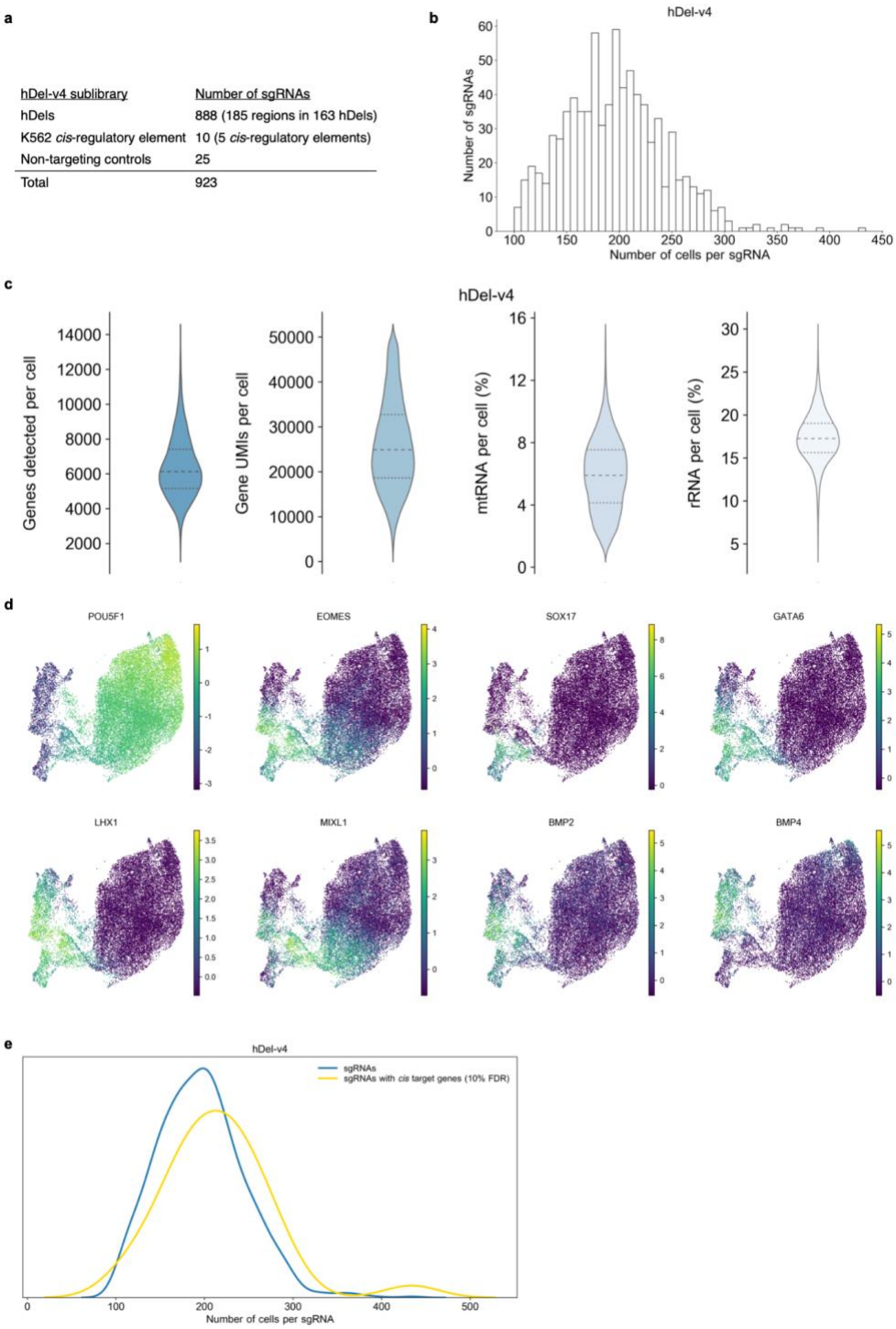


Extended Data Fig.7 sgRNA nucleotide homopolymer-associated toxicity.

a, Boxplots of sgRNA $\log_2$ fold-change for all hDel-v2 sgRNAs (far left), sgRNAs containing G_4 nucleotide homopolymers (center left), sgRNAs containing A_4 nucleotide homopolymers (center right), and sgRNAs containing C_4 nucleotide homopolymers (far right). *p*-values were obtained by Mann–Whitney *U* test. Boxes extend from the first quartile to the third quartile. Whiskers extend from boxes by 1.5× the interquartile range.

b, Distribution of sgRNA $\log_2$ fold-change for all hDel-v2 sgRNAs (white) and sgRNAs containing G_4 nucleotide homopolymers (blue).

c, Fraction of sgRNAs significantly modifying cellular proliferation (FDR < 0.05) for all hDel-v2 sgRNAs (black) and G_4-containing sgRNAs (grey). Position 0 corresponds to the most PAM-distal position in the sgRNA spacer sequence ($G_4N_{16}NGG$). The number of sgRNAs with G_4 nucleotide homopolymers at the indicated position is labeled.
d, Boxplots of sgRNA $\log_2$ fold-change for sgRNAs containing G_4 nucleotide homopolymers at the indicated position (red, Mann–Whitney *U* test *p*-value < 0.05).
e, Fraction of sgRNAs significantly modifying cellular proliferation (FDR < 0.05) for all hDel-v2 sgRNAs (black) and C_4-containing sgRNAs (grey). The number of sgRNAs with C_4 nucleotide homopolymers at the indicated position is labeled.
f, Boxplots of sgRNA $\log_2$ fold-change for sgRNAs containing C_4 nucleotide homopolymers at the indicated position (red, Mann–Whitney *U* test *p*-value < 0.05).
g, Distribution of sgRNA $\log_2$ fold-change for all *MYC*- and *GATA1*-targeting sgRNAs[24] (white) and sgRNAs containing G_4 nucleotide homopolymers (blue).
h, Boxplots of sgRNA $\log_2$ fold-change for *MYC*- and *GATA1*-targeting sgRNAs containing G_4 nucleotide homopolymers at the indicated position (red, Mann–Whitney *U* test *p*-value < 0.05).



Extended Data Fig. 8 hDel-v4 Direct-capture Perturb-seq.

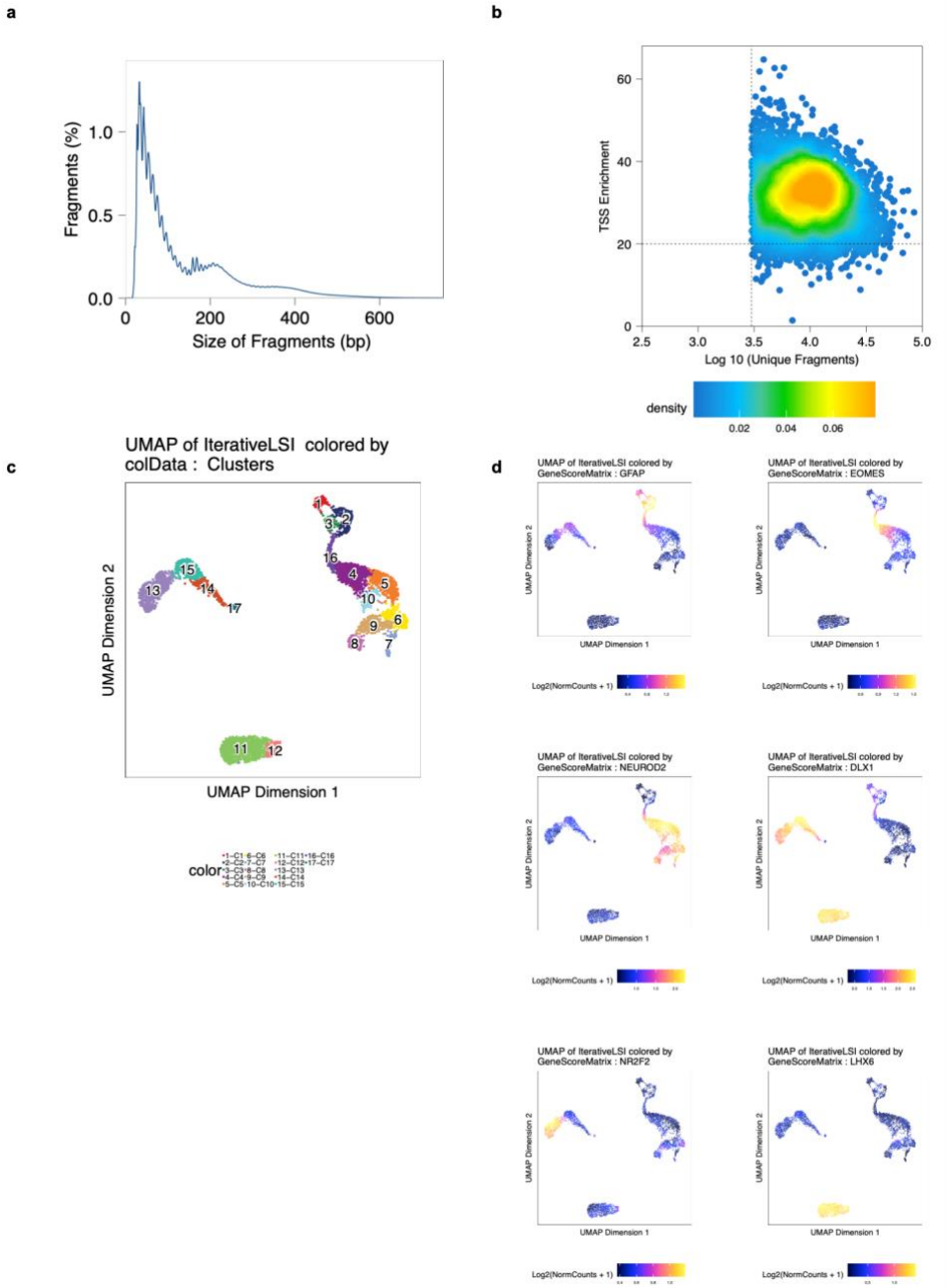
a, hDel-v4 sublibraries.

b, Distribution of the number of cells per sgRNA.

c, Distribution of the number of genes detected (far left), gene UMIs (center left), percent mtRNA (center right), and percent rRNA (far right) per cell.

d, UMAP projections colored by the normalized, log-transformed, and scaled expression of the indicated genes.

e, Distribution of the number of cells per sgRNA for all sgRNAs (blue) and sgRNAs with *cis* target genes (gold, FDR < 0.1).



Extended Data Fig. 9 snATAC-seq of PCD80 rhesus macaque prefrontal cortex.

a, Distribution of snATAC-seq fragment sizes.

b, Density plot of the number of fragments and ArchR TSS enrichment score per cell.

c, UMAP projection colored by clusters (n = 17), including radial glia (RG, 1, 2, 3), intermediate progenitor cells-newborn excitatory neurons (IPC-nEN, 4, 16), excitatory neurons (EN, 5, 6, 7, 8, 9, 10), consisting of deep layer EN (8) and upper layer EN (6,9), and inhibitory neurons (IN, 11, 12, 13, 14, 15, 17), consisting of MGE-derived IN (11, 12), CGE-derived IN (13), LGE-derived IN (14, 15).

d, UMAP projection colored by the ArchR gene score of the indicated genes.

Supplementary Information

Table S1. Human-specific deletion coordinates in the chimpanzee reference genome (panTro6)

Table S2. Tn5-accessible regions (Omni ATAC-seq) in chimpanzee iPS cells

Table S3. pA-Tn5-accessible regions (CUT&Tag) in chimpanzee iPS cells

Table S4. hDel-v1 sgRNA library and counts

Table S5. hDel-v1 α-RRA

Table S6. hDel-v2 sgRNA library and counts table

Table S7. hDel-v2 α-RRA

Table S8. hDel-v3 sgRNA library

Table S9. hDel-v3 *cis* sgRNA-gene pairs

Table S10. hDel-v3 *trans* sgRNA-gene pairs

Table S11. hDel-v4 sgRNA library

Table S12. hDel-v4 *cis* sgRNA-gene pairs

Table S13. Tn5-accessible regions (Omni ATAC-seq) in human, chimpanzee, and orangutan neuroepithelial cells