Functional Genomics Approaches for Investigating Human Evolution

by
Tyler Fair

DISSERTATION
Submitted in partial satisfaction of the requirements for degree of
DOCTOR OF PHILOSOPHY

in

Biomedical Sciences

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

DocuSigned by:
Nadav Ahituv
77BA96E7DAE34F4...

Nadav Ahituv
Chair

DocuSigned by:
LUKE GILBERT

LUKE GILBERT

DocuSigned by:
Mercedes Paredes

Mercedes Paredes

DocuSigned by:
Alex Pollen
2102D531BF2D416...

Alex Pollen

Committee Members

Dedication

This work is dedicated to my parents, Vickie Grove, PhD, and David Fair, PhD. Thank you for support, inspiration, and love.

Acknowledgments

Here I can only attempt to acknowledge the many sources of support that I have had during my graduate training. No words will convey the full extent of my gratitude, appreciation, and privilege.

To Dr. Alex Pollen, thank you for being an exceptional advisor. You gave me the freedom to pursue my interests while remaining available without notice at any time. Discussing science, track and field, chocolate-covered almonds, or really anything with you is a joy. You are a role model and a friend. Joining your research group was one of the best decisions I have ever made.

To Dr. Nadav Ahituv, Dr. Luke Gilbert, and Dr. Mercedes Paredes, thank you for providing critical feedback and mentorship. Your support was invaluable throughout my graduate training as I encountered setbacks, obstacles, and failures. I am eager to discuss my future research with each of you. To Dr. Jonathan Weissman, thank you for being a fantastic rotation advisor who inspired my interest in functional genomics. To Dr. Mike Levine, thank you for teaching me a love for discovery.

To Irene Chen, Jiapei Chen, and David Shin, thank you for being exceptional friends over the last six years.

To Vickie, David, and Brittany, I would not have made it here without you.

Contributions

This dissertation includes several projects that I undertook during my graduate training. Chapter I features a review and perspective that I wrote with Dr. Alex Pollen. Chapter II features a study that I led with critical contributions from Dr. Bryan Pavlovic, Dr. Nathan Schaefer, and Dr. Alex Pollen that is currently under review. Chapter III features a study that I co-led with Dr. Richard She, with critical contributions from Dr. Nathan Schaefer, Reuben Saunders, Dr. Bryan Pavlovic, Dr. Jonathan Weissman, and Dr. Alex Pollen. The chapters included in this dissertation are reproduced from the following publications:

Chapter I

Fair, T. & Pollen, A. A. Genetic architecture of human brain evolution. *Curr. Opin. Neurobiol.* **80**, 102710 (2023)

Chapter III

She, R. *et al.* Comparative landscape of genetic dependencies in human and chimpanzee stem cells. *Cell* 186, 2977–2994.e23 (2023)

Functional Genomics Approaches for Investigating Human Evolution

Tyler Fair

Abstract

The enumeration of human-specific genetic variants by comparative studies of ape genomes paired with advances in programmable genetic editing enables the functional interrogation of our evolutionary history. Because it is difficult to predict how millions of genomic alterations within the hominid lineage contribute to differences in phenotypes, there is a critical need for high-throughput, systematic approaches to probe genetic variation. Here, we develop a quantitative genome-scale platform for identifying the phenotypic consequences of human-specific genetic variants at a cellular and molecular level. First, we investigate the ancestral function of structural variant-sized human-specific deletions (hDels) by performing CRISPR interference (CRISPRi) genetic screens in chimpanzee induced pluripotent stem (iPS) cells. We further characterize the epigenetic state of chromatin at hDels using Omni ATAC-seq and CUT&Tag and perform single-cell CRISPRi (Perturb-seq) to identify their *cis*- and *trans*-regulatory target genes. We discover hDels removing *cis*-regulatory elements controlling the expression of proliferation-modifying genes including *MBD3*, *MRPS14*, and *RPL26*, and identify the *cis*-regulatory target genes of 16 nonessential hDels intersecting Omni ATAC-seq, H3K4me1, and H3K27ac. Separately, we investigate the evolution of gene regulatory networks during recent hominid evolution by performing paired CRISPRi genetic screens in human and chimpanzee iPS cells. We identify 75 genes with species-specific effects on cellular proliferation, many of which interact in biological pathways including RNA metabolism, chromatin organization, cell cycle phase transition, and UFMylation. Together, these findings underscore the importance of deletions as a source of evolutionary innovation and suggest that the regulation of essential cellular processes has evolved along the human lineage.

Tables of Contents

List of Figures

Chapter I

Genetic architecture of human brain evolution

Tyler Fair[1,2,3] and Alex A. Pollen[1,3*]

Affiliations

[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

[*]Corresponding author: Alex A. Pollen, alex.pollen@ucsf.edu

Keywords

Human brain evolution; Evolutionary genetics; Population genetics; Cis-regulatory evolution; Gene duplication; Effect size distribution; MPRA; CRISPR screens; Single-cell RNA-Sequencing; Brain organoid models; Neural development

ABSTRACT

Comparative studies of hominids have long sought to identify mutational events that shaped the evolution of the human nervous system. However, functional genetic differences are outnumbered by millions of nearly neutral mutations, and the developmental mechanisms underlying human nervous system specializations are difficult to model and incompletely understood. Candidate-gene studies have attempted to map select human-specific genetic differences to neurodevelopmental functions, but it remains unclear how to contextualize the relative effects of genes that are investigated independently. Considering these limitations, we discuss scalable approaches for probing the functional contributions of human-specific genetic differences. We propose that a systems-level view will enable a more quantitative and integrative understanding of the genetic, molecular and cellular underpinnings of human nervous system evolution.

INTRODUCTION

Cognitive and behavioral adaptations in the human lineage have enabled symbolic thought, compositional language, long-term planning, and the intergenerational transfer of knowledge and technology. The brain, the center of the nervous system, supports higher-order cognitive functions and has changed remarkably in size, development, and connectivity in humans. Since divergence from a common ancestor with chimpanzees, evolutionary alterations in human brain development tripled the number of neurons produced prenatally[1], extended synaptic maturation through adolescence[2], and reorganized connectivity between higher-order association areas of the cerebral cortex[3]. Identifying the genetic, molecular and cellular basis for specializations of the human brain will provide insights into neurodevelopment, disease, and ultimately, what makes us human (Figure 1.1).

Little is known of the identity and number of functional genetic differences contributing to human brain evolution. Genetic studies of evolution are motivated by the expectation that certain mutations have detectable effects, or else, as suggested by Orr, there would be "little reason, after all, to ask non-trivial questions about the genes that underlie adaptation if one assumes that there are thousands of them, each with small and interchangeable effects on the phenotype"[4]. Indeed, the distribution of effect sizes among new advantageous mutations is predicted to be exponential[5], a prediction supported by experimental evidence in model systems[6,7]. For complex traits that vary among humans, including brain structure and activity, most of the heritability is driven by large numbers of small-effect loci, with each locus accounting for <0.1–1% of phenotypic variance[8,9]. However, because trait-mapping analyses require sequence diversity in human populations, they cannot evaluate fixed variants that have emerged since our divergence from a common ancestor with chimpanzees. In contrast to the omnigenicity of some complex traits[10], adaptive evolution in natural populations can involve a small number of large-effect mutations and many of smaller effect[4,6]. Regardless of the extent to which oligogenic or polygenic adaptation

influenced human evolution[7,11,12], most functional genetic differences are expected to fall along the tail of a distribution of effect sizes (Figure 1.2). Therefore, identifying large-effect mutations–should they exist–and defining the effect size distribution for a measurable phenotype are central challenges facing genetic studies of human evolution.

Characterizing functional genetic differences requires an understanding of the cellular and neurodevelopmental processes underlying human-specific changes in brain size, structure, and function. However, it is difficult to determine which processes, among many, underlie higher-level trait divergence. For example, cortical expansion, a distinctive feature of human brain evolution, is predicted to involve diverse cellular and developmental mechanisms including proliferation, differentiation, and migration[13]. Several classes of neural stem and progenitor cells, including neuroepithelial cells, radial glia, and intermediate progenitor cells control cortical neurogenesis, but their respective contributions to cortical expansion are unresolved[14,15]. Alterations in the regulation of G_1 phase[16,17], cleavage plane angle[18], glucose and glutamine metabolism[19,20], autocrine and paracrine signaling[21,22], and extrinsic factors[23] can influence the decision between self-renewing or neurogenic divisions, but measures of how these processes differ among primates are limited. Without clearly defined differences in the constituent parts of neurodevelopment, it is difficult to weigh the influence of a mutation on higher-level traits.

Here, we outline a systems-level view for studying human brain evolution. We discuss scalable techniques, including pooled reporter assays and CRISPR-based genetic screens, that enable unbiased detection and ranking of functional genetic differences altering either reporter gene expression or cellular phenotypes, respectively. We also highlight the need to systematically identify human-specific differences in molecular, cellular and developmental processes that together give rise to higher-level traits. As a conceptual framework, we envision defining the nodes and edges of a multi-level trait hierarchy tree[24] to establish an integrative understanding of

specializations of the human brain.

Prioritizing and Modeling Human-Specific Genetic Differences

Functional genetic differences are outnumbered by millions of nearly neutral mutations and are difficult to distinguish based on sequence. Due to the similarity between human and chimpanzee protein-coding genes, gene regulatory changes have long been predicted to underlie human-specific traits[25]. Moreover, mutations with few pleiotropic effects, such as those affecting *cis*-regulatory elements and duplicated genes, may be more permissible under natural selection[26–28]. Genetic studies of evolution in natural populations highlight the importance of *cis*-regulatory mutations[29]. However, the genetic substrates for evolutionary change–whether *cis*-regulatory or protein coding–can vary across traits[28,30], with instances of *cis*-regulatory mutations associated with morphological divergence[6] and protein-coding mutations associated with sensory transduction, immune, and reproductive adaptations[31,32]. Among classes of genomic alterations, loss-of-function single-nucleotide variants and indels, gene and noncoding deletions, and gene duplications are most likely to have phenotypic consequences[33]. To prioritize individual mutations for investigation, comparative genomics approaches have considered sequence conservation[34,35], substitution rates[36–39], copy number variation[40], transcriptomics and epigenomics[41,42], gene ontology, and other signatures of selection.

Characterization of select noncoding human-specific genetic differences illustrates the promise and remaining challenges for genetic studies of human brain evolution. Transient transgenic *lacZ* reporter assays in mouse embryos have identified variants that modify[42] or completely remove[34] *cis*-regulatory elements active in the developing forebrain. *HARE5*, a human accelerated regulatory enhancer identified using this technique, is capable of accelerating cell cycle and reducing G_1/G_2/M phase length compared to the chimpanzee sequence when driving expression of the Wnt signaling pathway receptor *FZD8*[43]. Several recent studies have precisely modeled

human-specific genetic differences in the orthologous endogenous locus in transgenic mice[44–46]. For example, substitutions in the human accelerated region *HACNS1* increase H3K27ac and expand the expression domain of *Gbx2* in limb chondrogenic mesenchyme compared to the chimpanzee and mouse sequences[46]. However, *HACNS1* does not have a detectable effect on limb morphology[46], highlighting the difficulty of linking even statistically exceptional genetic differences to organismal-level phenotypes. By contrast, replacement of the mouse *Cbln2* enhancer 2 with the corresponding Hominini-specific sequence, which lacks two putative SOX5-binding sites, transiently increases expression of *Cbln2* and promotes dendritic spine formation in prefrontal cortex[45]. These findings support the prediction that evolutionary alterations in *cis*-regulatory elements can influence molecular, cellular and developmental processes, but underscore limitations in experimental throughput and *in silico* models of phenotype.

Segmental duplications, which occur at an accelerated rate in apes[47], are sources of gene innovation[27,48], underscoring the potential impact of this mutational class on human evolution. Segmental duplications have primarily been studied by introducing synthetic expression constructs into animal models or by generating cerebral organoids from stem cells harboring loss-of-function mutations. Several human- or hominoid-specific duplicated genes expressed during cortical development, including *ARHGAP11B*[49], *CROCCP2*[50], *NOTCH2NL*[21,22], and *TBC1D3*[51], are capable of promoting radial glia or intermediate progenitor cell proliferation. Duplicated genes have also been linked to processes regulating neuronal function: the hominid-specific paralog *LRRC37B* reduces neuronal excitability in adult human cortical slices[52] and the human-specific paralog *SRGAP2C* delays dendritic spine maturation and leads to increased cortico–cortical connectivity and sensory discrimination when expressed in mice[53]. While these studies have provided insights into cellular and developmental roles for human-specific duplicated genes, attaining the physiological cell type-specific expression required to evaluate the impact of each

duplication on human evolution is constrained by the challenge of paralog-specific gene editing and the use of randomly integrating transgenes in cellular and animal models.

The human-specific genetic differences described here, both noncoding and coding, have been characterized using model systems with distinct advantages and limitations. Primary among these models, the mouse has enabled the association of select genetic differences with alterations in gene expression, neurodevelopment, and even behavior. More recently, cerebral organoids derived from genetically engineered cell lines have emerged as an *in vitro* model for comparative studies of cortical germinal zones in primate genetic background[54–58]. While both model systems support complex phenotyping strategies, they are limited in throughput. Consequently, nearly all human-specific genetic differences remain functionally uncharacterized. For the small number of genetic differences that have been linked to neurodevelopmental functions, the candidate-gene approach precludes systematic comparisons between genes. In the following section, we discuss scalable approaches for identifying and contextualizing functional human-specific genetic differences (Figure 1.3).

Scalably Identifying Functional Human-Specific Genetic Differences

The pooled reporter assay, commonly known as the massively parallel reporter assay (MPRA), is a high-throughput approach for screening genetic sequences for potential *cis*-regulatory activity[59]. MPRAs measure the capacity for a sequence to drive reporter gene expression by introducing the sequence upstream or within the 3' untranslated region of the reporter and detecting transcribed, linked barcodes or the sequence itself among cellular mRNAs[60]. MPRA libraries, episomal or integrating, commonly consist of sheared genomic DNA, captured regions of interest, or synthesized sequences. Recent MPRAs have identified hundreds of human-specific variants[61–63] and modern-human alleles[64] that drive differential reporter gene expression.

Together, these studies have found epistatic interactions between substitutions in predicted transcription factor binding sites, *cis*-regulatory effects largely independent of *trans*-regulatory environment, and organs enriched for genes linked to modern-human alleles with differential effects. These studies also demonstrate that variants typically have modest effects ($\log_2$ mRNA/pDNA < 1). However, MPRAs may not accurately reflect the quantitative impact that variants have on target gene expression due to technical limitations, including the influence of flanking sequence and construct design on reporter gene expression, recombination between variants and associated barcodes, and non-native genomic context[60,65]. Despite these limitations, MPRAs enable comparison of human and ancestral alleles and identification of variants affecting mRNA splicing, stability, and localization[66–69].

Targeted genomic engineering is critical for understanding the effect that human-specific genetic differences have on molecular and cellular processes. Gene-editing technologies, including CRISPR-Cas, enable sequences to be manipulated within their native genomic context for functional analysis. Adopted for genetic screening, CRISPR-Cas has been used to systematically identify protein-coding genes, noncoding RNAs, and *cis*-regulatory elements controlling a range of phenotypes including proliferation, drug or toxin sensitivity, reporter gene expression, morphology, subcellular localization, and transcriptional and epigenetic state[70]. Cas9, the first Cas endonuclease employed for programmable genetic editing, is directed by chimeric single-guide RNAs (sgRNAs) to complementary DNA for cleavage (CRISPRn), interference (CRISPRi), or activation (CRISPRa)[71]. Several candidate-gene studies have used CRISPRn to compare modern- and archaic-human alleles in cerebral organoids[20,72,73], but high-throughput application of CRISPR-Cas for screening human-specific genetic differences is lacking.

Genome-scale CRISPR-based genetic screens provide the opportunity to functionally evaluate and compare genetic differences that may have shaped human nervous system evolution. For

example, libraries of sgRNAs targeting human-specific genetic differences can support screening for cell-intrinsic regulators of proliferation in pluripotent stem cell (PSC)-derived neuroepithelial cells, a cell type predicted to contribute to cortical expansion[14]. In this experiment, high-throughput sequencing-based quantification of sgRNA abundance before and after a period of competitive growth would enable the identification of functional genomic sequences. However, many human-specific features, including differences in neuronal morphology, plasticity, myelination, and connectivity may not directly influence proliferation and will require additional phenotyping strategies. For example, an imaging-based CRISPR screen in excitatory projection neurons could identify human-specific genetic differences influencing dendritic arborization and spine density. The use of reporter constructs designed to measure complex cellular processes, including synaptogenesis, metabolism, and signaling will enable screening additional phenotypes that diverged during human evolution.

Recent advances coupling CRISPR-based genetic screens with single-cell RNA-sequencing will be useful for identifying human-specific genetic differences with transcriptional consequences. Single-cell CRISPR screens have been applied to map *cis*-regulatory elements to target genes[74], dissect both *cis*- and *trans*-regulatory effects for noncoding variants associated with neurodegenerative dementias[75] and blood traits[76], and identify transcription factors controlling cell state and fate determination in mosaic cerebral organoids[77]. With libraries of sgRNAs targeting human-specific genetic differences, this approach could be used to discover variants altering cell-cycle kinetics and the balance between self-renewing or neurogenic divisions in radial glia[16]. Single-cell CRISPR screens also allow for direct comparison of evolutionary mutations previously modeled independently. For example, generating mosaic cerebral organoids from cells harboring sgRNAs targeting genes implicated in human neurodevelopmental processes–including radial glia proliferation (*ARHGAP11B*, *NOTCH2NL*, *TBC1D3*, *TKTL1*) and synaptogenesis (*FOXP2*,

NOVA1, *SRGAP2C*)–could enable quantitative comparison of cell type and cell state abundance across perturbations and identification of shared mechanisms of action.

The application of CRISPR-based genetic screens to probe human-specific genetic differences is subject to several limitations. Current CRISPR-Cas systems do not allow for precise reconstruction of ancestral alleles at scale. Consequently, it is difficult to determine using loss-of-function or gain-of-function screens whether a single-nucleotide variant or amino acid substitution is causal. Structural variants are also challenging to screen, as dual-sgRNA programmed deletions occur at low frequencies in sgRNA-expressing cells[78]. Moreover, Cas9-induced double-strand break toxicity and mutational heterogeneity can obscure the identification of small-effect variants[79]. For all CRISPR-Cas modalities, the specificity of sgRNA sequences must be carefully considered during sgRNA library design to mitigate off-target activity[80]. When coupled with single-cell technologies, CRISPR screens are constrained by the cost of scRNA-seq and are powered to detect changes in gene expression primarily for highly-expressed genes[74]. Finally, forward genetic screens are currently unable to directly evaluate the effect that genetic differences have on higher-level human-specific traits. Despite these limitations, we anticipate that CRISPR screens will enable the assignment of molecular and cellular functions to human-specific genetic differences at scale.

Cataloging Divergent Molecular, Cellular, and Neurodevelopmental Processes

Establishing a catalog of molecular, cellular, and neurodevelopmental processes that have been modified in the human lineage is essential for informing mechanistic studies of human brain evolution (Figure 1.4). Several modes of action for human-specific genetic differences linked to neurodevelopment have been proposed, including Wnt signaling[43], Notch signaling[21,22], glutaminolysis[49], and histone methylation[51], but the extent to which these processes differ among

primates remains unresolved. Because biological pathways can diverge without inducing a concomitant change in phenotype[81,82], it is important to obtain measurements at multiple levels of the trait hierarchy tree to distinguish genetic differences driving evolutionary change from those buffered by epistatic interactions.

Comparative studies of cell cycle regulation in primate neural progenitor cells reveal a promising connection between select human-specific genetic differences and neurodevelopmental processes. Motivated by the observation that cell cycle duration is lengthened in rhesus macaques compared to mice during neurogenesis[83], *in vitro* time-lapse imaging of human, chimpanzee, and orangutan cerebral organoids has revealed a human-specific increase in the duration of metaphase in neuroepithelial cells[54]. Introduction of modern human-specific amino acid substitutions in two genes associated with the mitotic spindle and kinetochore (*KIF18a* and *KNL1*) in mice recapitulates the prolongation of metaphase[73], suggesting that these genetic differences may be causal. Time-lapse imaging of human and rhesus macaque neural progenitor cells has also revealed a cell-intrinsic increase in progenitor cell clonal output; the progenitor cell expansion phase lasts approximately twice as long in human cells compared to rhesus macaque cells[84]. While this finding does not resolve the respective contributions of neuroepithelial cells, radial glia, and intermediate progenitor cells to human cortical expansion, it hints that human-specific features of neurodevelopment may be preserved in a genetically tractable *in vitro* environment.

Within the trait hierarchy tree, molecular features are currently most amenable to systematic characterization. By providing high-dimensional measurements across cell types, comparative single-cell transcriptomics and epigenomic profiling in primate primary tissue and cerebral organoids have revealed neurodevelopmental gene-regulatory programs specific to humans[55,58]. Pooled derivation of midbrain neurons from 215 human individuals[85] highlights the potential to

efficiently scale these approaches across primates to define cell-intrinsic molecular divergence. Similarly, the use of tetraploid composite cells containing human and chimpanzee genomes in the same nucleus enables disentangling *cis*- from *trans*-regulatory effects[86,87], and recent advances in artificial neural networks support linking *cis*-regulatory divergence to human-specific genetic differences[63,88,89]. Despite considerable progress mapping human-specific molecular features, many molecular differences do not have detectable functional consequences, and establishing a connection to higher-level traits is limited by current model systems and phenotyping strategies.

Recent advances in lineage tracing, spatial genomics, cellular reporters, and gene-editing technologies will enable systematically measuring cellular phenotypes across primates. To evaluate the clonal output of neural stem and progenitor cells, combining single-cell transcriptomics with cellular barcoding[90] will allow for constructing and comparing lineage trees between human and nonhuman primate cerebral organoids. Supported by algorithms for cellular image analysis[91], spatial transcriptomics and proteomics[92] will be useful for identifying changes in the morphology and organization of molecularly defined cell types within primate primary tissue. Cellular reporters for signaling[93], synaptogenesis[94], and metabolism[95] in primate cell models and xenografts[96] will permit measuring dynamic processes predicted to have diverged in the human lineage. Additionally, profiling stimulus-dependent responses in primate cell models[97,98] subject to genetic or non-genetic perturbations may unmask species differences absent in unperturbed cells[99]. Together, systematic comparative phenotyping will lay the foundation for an integrative understanding of the genetic, molecular and cellular basis for human brain evolution.

CONCLUSION

Candidate-gene studies have investigated select human-specific genetic differences in animal models and cerebral organoids, revealing striking connections between genotype and gene

expression, neurodevelopment, and even behavior. But reliance on complex models precludes comprehensive characterization of the millions of genetic differences that have accumulated in the human lineage. While strategies for prioritization support reverse genetic approaches, mutations with functional consequences on cellular and developmental processes can be indistinguishable from nearly neutral mutations based on genomic and epigenomic features. Here, we propose empirically defining an effect size distribution for human-specific genetic differences using forward genetic screens and systemically characterizing divergent molecular, cellular, and developmental processes in primate cell models and primary tissue. We envision defining the nodes and edges of a multi-level trait hierarchy tree to link the genetic architecture of a phenotype of interest to less experimentally tractable higher-level traits. This conceptual framework provides and informs multiple entry points for studies of human brain evolution.

ACKNOWLEDGEMENTS

The authors thank Jenelle Wallace, Nathan Schaefer, Sara Nolbrant, Riley Mangan, Craig Lowe, members of the Pollen lab for helpful discussions, and Helena Pinheiro for illustrations. We acknowledge the following funding sources: Ruth L. Kirschstein National Research Service Predoctoral Fellowship Award F31 HG011569-01A, 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.

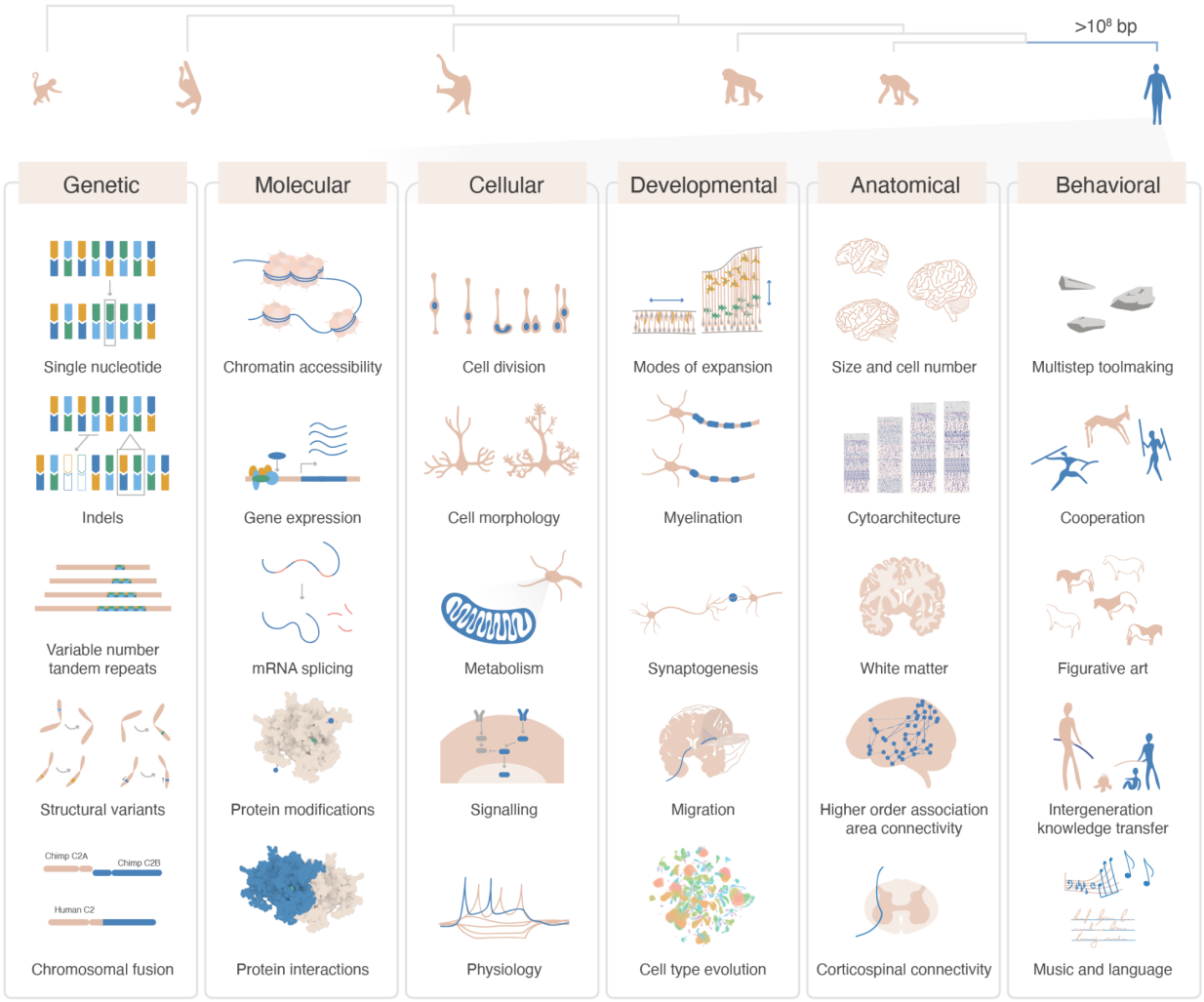


Figure 1.1. Linking human-specific genetic differences to human brain evolution. Functional genetic differences give rise to higher-level trait evolution by acting through molecular and cellular mechanisms. Phylogenetic tree (top) shows evolutionary relationships of humans and other primates, denoting approximate sequence divergence on the human branch. Illustrations (bottom) depict classes of genetic, molecular, cellular, developmental, and anatomical features influencing neural circuit function and human-specific behaviors. Although centered on the evolution of the human brain, this concept generalizes across organs and organisms.

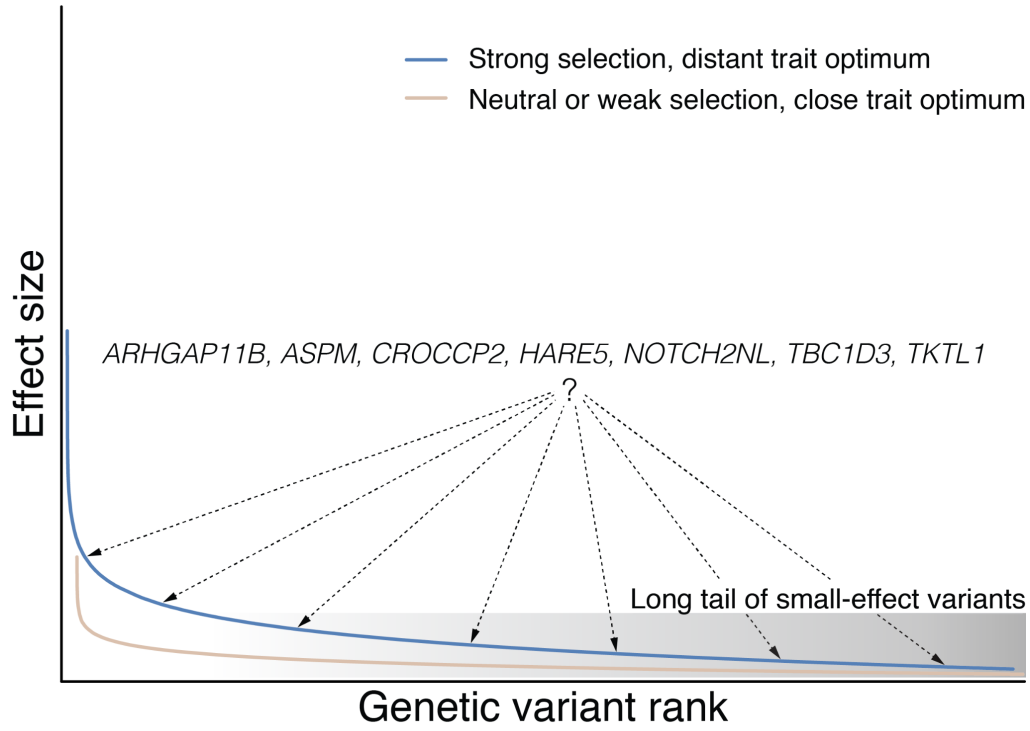


Figure 1.2. Predicting effect size distributions for adaptive human-specific genetic differences under multiple selective regimes.
The genetic architecture of human brain evolution, the distribution of genetic effects associated with cognitive and behavioral alterations in the environment of evolutionary adaptedness since divergence from a common ancestor with chimpanzees, is undetermined. Traits evolving under adaptive selection toward a distant optimum may be explained by a distribution enriched for strong-effect variants (blue) relative to traits with less selective pressure or a close optimum (beige)[5,24]. Under both selective regimes, most mutations will fall along the tail of an effect size distribution. Although select mutations have been implicated in human neurodevelopment–listed alphabetically–their placement along the distribution for a phenotype of interest is undetermined (arrows), and their effect size is likely influenced by genetic background[100], complicating experimental studies.

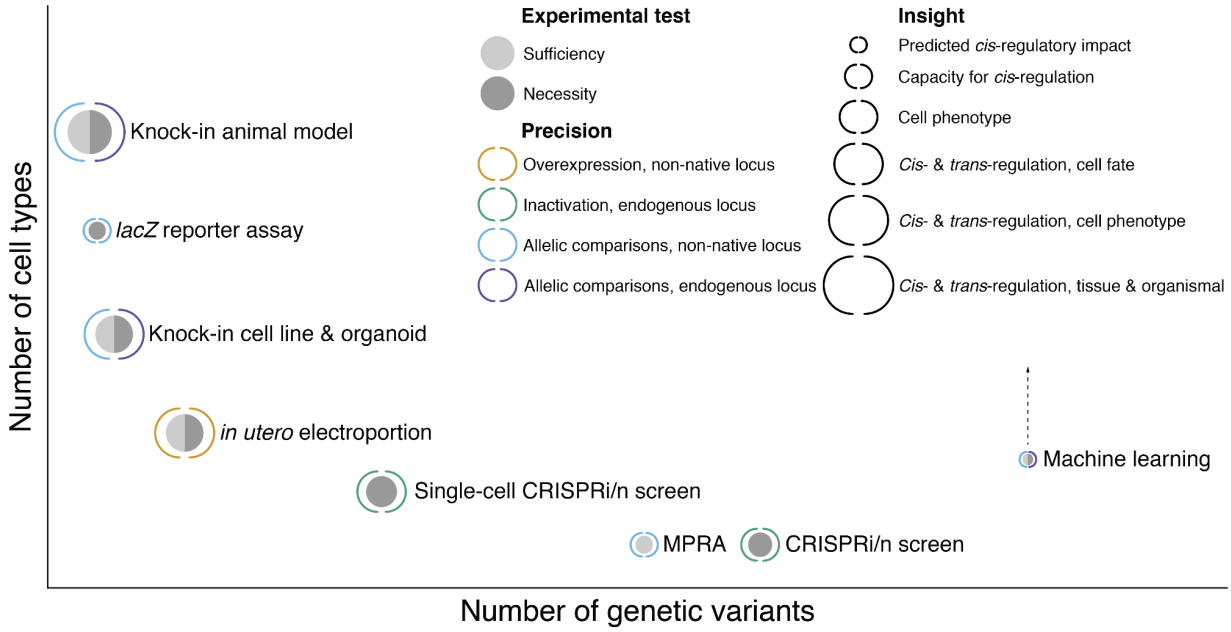


Figure 1.3. Interrogating human-specific genetic differences.
Techniques for interrogating the functional consequences of human-specific genetic differences vary along several dimensions, including: (1) the number of genetic differences assayed (*X*-axis), (2) the number of cell types assayed (*Y*-axis), (3) whether the technique evaluates the sufficiency (light gray) or necessity (dark gray) of a sequence, (4) whether the technique involves overexpression in a non-native genomic context (orange edge), inactivation in an endogenous locus (green edge), allelic comparisons in a non-native genomic context (light blue edge), or allelic comparisons in an endogenous locus (purple edge), and (5) insight, from predicted *cis*-regulatory effects (smallest oval) to behavioral effects in animal models (largest oval).

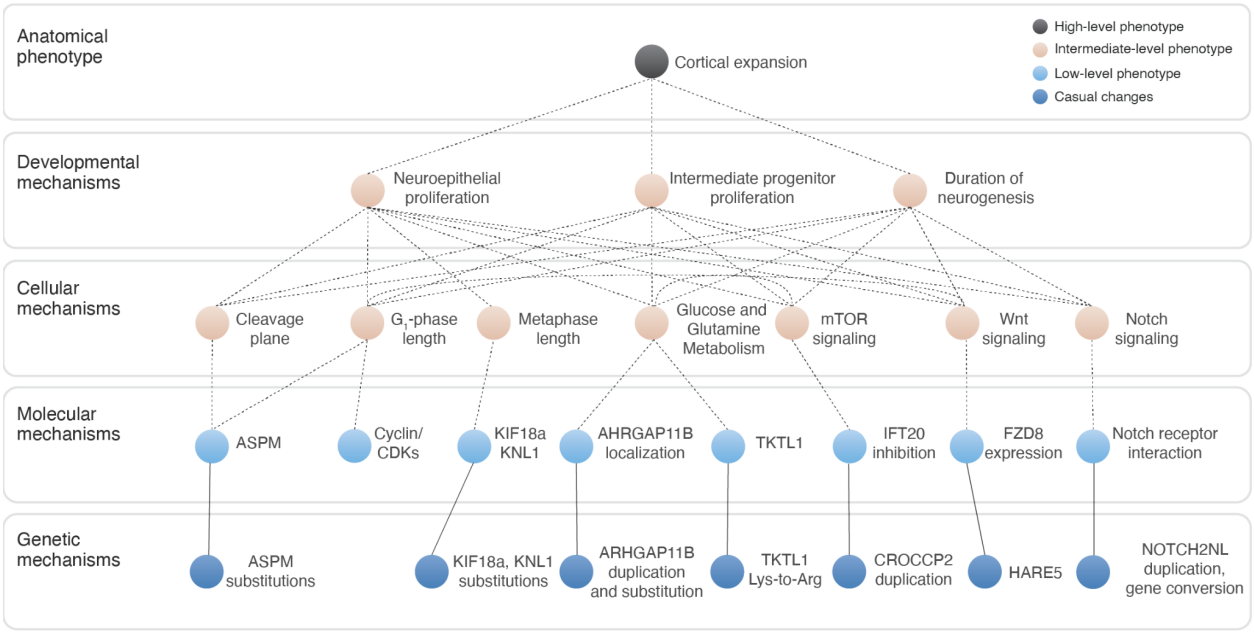


Figure 1.4. Envisioning a multi-level trait hierarchy tree for human cortical expansion. A multi-level trait hierarchy tree[24] is a conceptual framework for connecting functional genetic differences through molecular, cellular, and developmental mechanisms to higher-level traits. In this example, genetic differences associated with cortical expansion are arranged on the bottom row, with solid edges connecting these differences to molecular mechanisms, and dashed lines linking molecular mechanisms to putative cellular and developmental mechanisms. This framework generalizes across anatomical and behavioral traits and provides multiple entry points for identifying mechanisms underlying human evolution.

REFERENCES

1. Herculano-Houzel, S. The remarkable, yet not extraordinary, human brain as a scaled-up primate brain and its associated cost. *Proc. Natl. Acad. Sci. U. S. A.* **109 Suppl 1**, 10661–10668 (2012).
2. Petanjek, Z. *et al.* Extraordinary neoteny of synaptic spines in the human prefrontal cortex. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 13281–13286 (2011).
3. Buckner, R. L. & Krienen, F. M. The evolution of distributed association networks in the human brain. *Trends Cogn. Sci.* **17**, 648–665 (2013).
4. Orr, H. A. The genetic theory of adaptation: a brief history. *Nat. Rev. Genet.* **6**, 119–127 (2005).
5. Orr, H. A. The distribution of fitness effects among beneficial mutations. *Genetics* **163**, 1519–1526 (2003).
6. Peichel, C. L. & Marques, D. A. The genetic and molecular architecture of phenotypic diversity in sticklebacks. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **372**, (2017).
7. Dittmar, E. L., Oakley, C. G., Conner, J. K., Gould, B. A. & Schemske, D. W. Factors influencing the effect size distribution of adaptive substitutions. *Proc. Biol. Sci.* **283**, (2016).
8. Thompson, P. M. *et al.* ENIGMA and global neuroscience: A decade of large-scale studies of the brain in health and disease across more than 40 countries. *Transl. Psychiatry* **10**, 100 (2020).
9. Weiner, D. J. *et al.* Polygenic architecture of rare coding variation across 394,783 exomes. *Nature* (2023) doi:10.1038/s41586-022-05684-z.
10. Liu, X., Li, Y. I. & Pritchard, J. K. Trans Effects on Gene Expression Can Drive Omnigenic Inheritance. *Cell* **177**, 1022-1034.e6 (2019).
11. Mathieson, I. Human adaptation over the past 40,000 years. *Curr. Opin. Genet. Dev.* **62**, 97–104 (2020).
12. Pritchard, J. K., Pickrell, J. K. & Coop, G. The genetics of human adaptation: hard sweeps,

soft sweeps, and polygenic adaptation. *Curr. Biol.* **20**, R208-15 (2010).

13. Geschwind, D. H. & Rakic, P. Cortical evolution: judge the brain by its cover. *Neuron* **80**, 633–647 (2013).
14. Rakic, P. A small step for the cell, a giant leap for mankind: a hypothesis of neocortical expansion during evolution. *Trends Neurosci.* **18**, 383–388 (1995).
15. Kriegstein, A., Noctor, S. & Martínez-Cerdeño, V. Patterns of neural stem and progenitor cell division may underlie evolutionary cortical expansion. *Nat. Rev. Neurosci.* **7**, 883–890 (2006).
16. Dehay, C. & Kennedy, H. Cell-cycle control and cortical development. *Nat. Rev. Neurosci.* **8**, 438–450 (2007).
17. Taverna, E., Götz, M. & Huttner, W. B. The cell biology of neurogenesis: toward an understanding of the development and evolution of the neocortex. *Annu. Rev. Cell Dev. Biol.* **30**, 465–502 (2014).
18. Vaid, S. & Huttner, W. B. Progenitor-Based Cell Biological Aspects of Neocortex Development and Evolution. *Front Cell Dev Biol* **10**, 892922 (2022).
19. Namba, T. *et al.* Human-Specific ARHGAP11B Acts in Mitochondria to Expand Neocortical Progenitors by Glutaminolysis. *Neuron* **105**, 867-881.e9 (2020).
20. Pinson, A. *et al.* Human TKTL1 implies greater neurogenesis in frontal neocortex of modern humans than Neanderthals. *Science* **377**, eabl6422 (2022).
21. Suzuki, I. K. *et al.* Human-Specific NOTCH2NL Genes Expand Cortical Neurogenesis through Delta/Notch Regulation. *Cell* **173**, 1370-1384.e16 (2018).
22. Fiddes, I. T. *et al.* Human-Specific NOTCH2NL Genes Affect Notch Signaling and Cortical Neurogenesis. *Cell* **173**, 1356-1369.e22 (2018).
23. Lehtinen, M. K. *et al.* The Cerebrospinal Fluid Provides a Proliferative Niche for Neural Progenitor Cells. *Neuron* **69**, 893–905 (2011).
24. Barghi, N., Hermisson, J. & Schlötterer, C. Polygenic adaptation: a unifying framework to understand positive selection. *Nat. Rev. Genet.* **21**, 769–781 (2020).

25. King, M. C. & Wilson, A. C. Evolution at two levels in humans and chimpanzees. *Science* **188**, 107–116 (1975).
26. Carroll, S. B. Evo–devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* **134**, 25–36 (2008).
27. Ohno, S. *Evolution by Gene Duplication*. (Springer Berlin Heidelberg, 1970).
28. Hoekstra, H. E. & Coyne, J. A. The locus of evolution: evo devo and the genetics of adaptation. *Evolution* **61**, 995–1016 (2007).
29. Wray, G. A. The evolutionary significance of cis-regulatory mutations. *Nat. Rev. Genet.* **8**, 206–216 (2007).
30. Halligan, D. L. *et al.* Contributions of protein-coding and regulatory change to adaptive molecular evolution in murid rodents. *PLoS Genet.* **9**, e1003995 (2013).
31. Julius, D. & Nathans, J. Signaling by sensory receptors. *Cold Spring Harb. Perspect. Biol.* **4**, a005991 (2012).
32. Haygood, R., Babbitt, C. C., Fedrigo, O. & Wray, G. A. Contrasts between adaptive coding and noncoding changes during human evolution. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 7853–7857 (2010).
33. Abel, H. J. *et al.* Mapping and characterization of structural variation in 17,795 human genomes. *Nature* **583**, 83–89 (2020).
34. McLean, C. Y. *et al.* Human-specific loss of regulatory DNA and the evolution of human-specific traits. *Nature* **471**, 216–219 (2011).
35. Kronenberg, Z. N. *et al.* High-resolution comparative analysis of great ape genomes. *Science* **360**, eaar6343 (2018).
36. Pollard, K. S. *et al.* An RNA gene expressed during cortical development evolved rapidly in humans. *Nature* **443**, 167–172 (2006).
37. Prabhakar, S., Noonan, J. P., Pääbo, S. & Rubin, E. M. Accelerated evolution of conserved noncoding sequences in humans. *Science* **314**, 786 (2006).

38. Nielsen, R. R. *et al.* A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biol.* **3**, e170–e170 (2005).
39. Mangan, R. J. *et al.* Adaptive sequence divergence forged new neurodevelopmental enhancers in humans. *Cell* **185**, 4587-4603.e23 (2022).
40. Dennis, M. Y. *et al.* The evolution and population diversity of human-specific segmental duplications. *Nature ecology & evolution* **1**, 69 (2017).
41. Prescott, S. L. *et al.* Enhancer Divergence and cis-Regulatory Evolution in the Human and Chimp Neural Crest. *Cell* **163**, 68–83 (2015).
42. Reilly, S. K. *et al.* Evolutionary genomics. Evolutionary changes in promoter and enhancer activity during human corticogenesis. *Science* **347**, 1155–1159 (2015).
43. Boyd, J. L. *et al.* Human-chimpanzee differences in a FZD8 enhancer alter cell-cycle dynamics in the developing neocortex. *Curr. Biol.* **25**, 772–779 (2015).
44. Aldea, D. *et al.* Repeated mutation of a developmental enhancer contributed to human thermoregulatory evolution. *Proc. Natl. Acad. Sci. U. S. A.* **118**, (2021).
45. Shibata, M. *et al.* Hominini-specific regulation of CBLN2 increases prefrontal spinogenesis. *Nature* **598**, 489–494 (2021).
46. Dutrow, E. V. *et al.* Modeling uniquely human gene regulatory function via targeted humanization of the mouse genome. *Nat. Commun.* **13**, 304 (2022).
47. Marques-Bonet, T. *et al.* A burst of segmental duplications in the genome of the African great ape ancestor. *Nature* **457**, 877–881 (2009).
48. Shew, C. J. *et al.* Diverse Molecular Mechanisms Contribute to Differential Expression of Human Duplicated Genes. *Mol. Biol. Evol.* **38**, 3060–3077 (2021).
49. Heide, M. & Huttner, W. B. Human-Specific Genes, Cortical Progenitor Cells, and Microcephaly. *Cells* **10**, (2021).
50. Van Heurck, R. *et al.* CROCCP2 acts as a human-specific modifier of cilia dynamics and mTOR signaling to promote expansion of cortical progenitors. *Neuron* (2022)

doi:10.1016/j.neuron.2022.10.018.

51. Hou, Q.-Q., Xiao, Q., Sun, X.-Y., Ju, X.-C. & Luo, Z.-G. TBC1D3 promotes neural progenitor proliferation by suppressing the histone methyltransferase G9a. *Sci Adv* **7**, (2021).
52. Libé-Philippot, B. *et al.* LRRC37B is a species-specific regulator of voltage-gated channels and excitability in human cortical neurons. *bioRxiv* 2022.12.21.521423 (2022) doi:10.1101/2022.12.21.521423.
53. Schmidt, E. R. E. *et al.* A human-specific modifier of cortical connectivity and circuit function. *Nature* **599**, 640–644 (2021).
54. Mora-Bermúdez, F. *et al.* Differences and similarities between human and chimpanzee neural progenitors during cerebral cortex development. *Elife* **5**, 166 (2016).
55. Kanton, S. *et al.* Organoid single-cell genomic atlas uncovers human-specific features of brain development. *Nature* **574**, 418–422 (2019).
56. Fischer, J. *et al.* Human-specific ARHGAP11B ensures human-like basal progenitor levels in hominid cerebral organoids. *EMBO Rep.* e54728 (2022).
57. Mostajo-Radji, M. A., Schmitz, M. T., Montoya, S. T. & Pollen, A. A. Reverse engineering human brain evolution using organoid models. *Brain Res.* **1729**, 146582 (2020).
58. Pollen, A. A. *et al.* Establishing Cerebral Organoids as Models of Human-Specific Brain Evolution. *Cell* **176**, 743-756.e17 (2019).
59. Patwardhan, R. P. *et al.* Massively parallel functional dissection of mammalian enhancers in vivo. *Nat. Biotechnol.* **30**, 265–270 (2012).
60. 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).
61. Girskis, K. M. *et al.* Rewiring of human neurodevelopmental gene regulatory programs by human accelerated regions. *Neuron* **109**, 3239-3251.e7 (2021).
62. Uebbing, S. *et al.* Massively parallel discovery of human-specific substitutions that alter enhancer activity. *Proc. Natl. Acad. Sci. U. S. A.* **118**, e2007049118 (2021).

63. Whalen, S. *et al.* Machine learning dissection of human accelerated regions in primate neurodevelopment. *Neuron* (2023) doi:10.1016/j.neuron.2022.12.026.

64. Weiss, C. V. *et al.* The cis-regulatory effects of modern human-specific variants. *Elife* **10**, e63713 (2021).

65. Lindhorst, D. & Halfon, M. S. Reporter gene assays and chromatin-level assays define substantially non-overlapping sets of enhancer sequences. *bioRxiv* 2022.04.21.489091 (2022) doi:10.1101/2022.04.21.489091.

66. Adamson, S. I., Zhan, L. & Graveley, B. R. Vex-seq: high-throughput identification of the impact of genetic variation on pre-mRNA splicing efficiency. *Genome Biol.* **19**, 71 (2018).

67. Griesemer, D. *et al.* Genome-wide functional screen of 3'UTR variants uncovers causal variants for human disease and evolution. *Cell* **184**, 5247-5260.e19 (2021).

68. Sample, P. J. *et al.* Human 5' UTR design and variant effect prediction from a massively parallel translation assay. *Nat. Biotechnol.* **37**, 803–809 (2019).

69. Hansen, T. J., Fong, S., Capra, J. A. & Hodges, E. Human gene regulatory evolution is driven by the divergence of regulatory element function in both cis and trans. *bioRxiv* 2023.02.14.528376 (2023) doi:10.1101/2023.02.14.528376.

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

71. Knott, G. J. & Doudna, J. A. CRISPR-Cas guides the future of genetic engineering. *Science* **361**, 866–869 (2018).

72. Trujillo, C. A. *et al.* Reintroduction of the archaic variant of NOVA1 in cortical organoids alters neurodevelopment. *Science* **371**, (2021).

73. Mora-Bermúdez, F. *et al.* Longer metaphase and fewer chromosome segregation errors in modern human than Neanderthal brain development. *Science Advances* **8**, eabn7702 (2022).

74. Gasperini, M. *et al.* A Genome-wide Framework for Mapping Gene Regulation via Cellular

Genetic Screens. *Cell* **176**, 377-390.e19 (2019).

75. Cooper, Y. A. *et al.* Functional regulatory variants implicate distinct transcriptional networks in dementia. *Science* **377**, eabi8654 (2022).
76. 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.
77. Fleck, J. S. *et al.* Inferring and perturbing cell fate regulomes in human brain organoids. *Nature* 1–8 (2022).
78. 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).
79. Álvarez, M. M., Biayna, J. & Supek, F. TP53-dependent toxicity of CRISPR/Cas9 cuts is differential across genomic loci and can confound genetic screening. *Nat. Commun.* **13**, 4520 (2022).
80. Tycko, J. *et al.* Mitigation of off-target toxicity in CRISPR-Cas9 screens for essential non-coding elements. *Nat. Commun.* **10**, 4063 (2019).
81. True, J. R. & Haag, E. S. Developmental system drift and flexibility in evolutionary trajectories. *Evolution and Development* vol. 3 109–119 Preprint at https://doi.org/10.1046/j.1525-142x.2001.003002109.x (2001).
82. Wang, S. H., Hsiao, C. J., Khan, Z. & Pritchard, J. K. Post-translational buffering leads to convergent protein expression levels between primates. *Genome Biol.* **19**, 83 (2018).
83. Kornack, D. R. & Rakic, P. Changes in cell-cycle kinetics during the development and evolution of primate neocortex. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 1242–1246 (1998).
84. Otani, T., Marchetto, M. C., Gage, F. H., Simons, B. D. & Livesey, F. J. 2D and 3D Stem Cell Models of Primate Cortical Development Identify Species-Specific Differences in Progenitor

Behavior Contributing to Brain Size. *Cell Stem Cell* **18**, 467–480 (2016).

85. Jerber, J. *et al.* Population-scale single-cell RNA-seq profiling across dopaminergic neuron differentiation. *Nat. Genet.* **53**, 304–312 (2021).
86. 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).
87. Pavlovic, B. J., Fox, D., Schaefer, N. K. & Pollen, A. A. Rethinking nomenclature for interspecies cell fusions. *Nat. Rev. Genet.* **23**, 315–320 (2022).
88. Kelley, D. R. Cross-species regulatory sequence activity prediction. *PLoS Comput. Biol.* **16**, e1008050 (2020).
89. Avsec, Ž. *et al.* Base-resolution models of transcription-factor binding reveal soft motif syntax. *Nat. Genet.* **53**, 354–366 (2021).
90. Delgado, R. N. *et al.* Individual human cortical progenitors can produce excitatory and inhibitory neurons. *Nature* vol. 601 397–403 Preprint at https://doi.org/10.1038/s41586-021-04230-7 (2022).
91. Moen, E. *et al.* Deep learning for cellular image analysis. *Nat. Methods* **16**, 1233–1246 (2019).
92. Fang, R. *et al.* Conservation and divergence of cortical cell organization in human and mouse revealed by MERFISH. *Science* **377**, 56–62 (2022).
93. Qian, Y., Celiker, O. T., Wang, Z., Guner-Ataman, B. & Boyden, E. S. Temporally multiplexed imaging of dynamic signaling networks in living cells. *bioRxiv* 2022.08.22.504781 (2022) doi:10.1101/2022.08.22.504781.
94. Berryer, M. H. *et al.* An automated high-content synaptic phenotyping platform in human neurons and astrocytes reveals a role for BET proteins in synapse assembly. *bioRxiv* 2022.05.17.492322 (2022) doi:10.1101/2022.05.17.492322.
95. Chandris, P., Giannouli, C. C. & Panayotou, G. Imaging Approaches for the Study of Metabolism in Real Time Using Genetically Encoded Reporters. *Front Cell Dev Biol* **9**,

725114 (2021).

96. Marchetto, M. C. *et al.* Species-specific maturation profiles of human, chimpanzee and bonobo neural cells. *Elife* **8**, 1748 (2019).
97. Ward, M. C., Banovich, N. E., Sarkar, A., Stephens, M. & Gilad, Y. Dynamic effects of genetic variation on gene expression revealed following hypoxic stress in cardiomyocytes. *Elife* **10**, e57345 (2021).
98. Pizzollo, J. *et al.* Comparative Serum Challenges Show Divergent Patterns of Gene Expression and Open Chromatin in Human and Chimpanzee. *Genome Biol. Evol.* **10**, 826–839 (2018).
99. She, R. *et al.* Comparative landscape of genetic dependencies in human and chimpanzee stem cells. *bioRxiv* 2023.03.19.533346 (2023) doi:10.1101/2023.03.19.533346.
100. Tabbaa, M., Knoll, A. & Levitt, P. Mouse population genetics phenocopies heterogeneity of human Chd8 haploinsufficiency. *Neuron* (2023) doi:10.1016/j.neuron.2023.01.009.

Chapter II

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]

Affiliations

[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

[*]Corresponding author: Alex A. Pollen, alex.pollen@ucsf.edu

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] (Figure S2.1) 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] (Figure S2.2). 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 (Figure S3.1a, 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 (Figure S3.1b, 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 (Figure 2.1a, 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 (Figure 2.1a).

Analysis of technical and biological replicates revealed that sgRNAs were highly correlated between replicates of the same cell line (Figure 2.1d and Figure S2.3, 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 (Figure 2.1d). 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, Figure 2.1c). 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. (Figure 2.1d). Genome-wide, we identified 16 hDels modifying cellular proliferation (FDR < 0.1, Figure 2.1e).

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 (Figure 2.2a, 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 (Figure S2.4, r = 0.82). We discovered 38 hDels modifying cellular proliferation (FDR < 0.1, Figure 2.2c), including 14 hDels intersecting Omni ATAC-seq, H3K4me1, H3K4me3, H3K27ac, or H3K27me3, epigenetic features associated with *cis*-

regulatory elements (Figure 2.2d). 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 (Figure 2.2e). 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 (Figure S2.5a). 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 (Figure 2.3a, Direct-capture Perturb-seq). After filtering, we recovered 16,810 sgRNA-expressing cells (Figure S2.5b-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 (Figure 2.3b, 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 (Figure 2.3c). 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 (Figure 2.3d) reduced the expression of the 60S ribosomal protein L26 *RPL26* (Figure 2.3e, 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 (Figure 2.3f, r = 0.97). hDel_6012 does not intersect epigenetic features associated with *cis*-regulatory elements (Figure 2.3d), 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 (Figure 2.3g, FDR < 0.1). As expected, sgRNAs targeting hDel_6012 elicited correlated genome-wide transcriptional responses (Figure S2.6a, r = 0.57 to 0.68, FDR < 0.1).

sgRNAs targeting hDel_349 (Figure S2.6b) reduced the expression of the mitochondrial ribosomal 28S subunit *MRPS14* (Figure S2.6c, 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 (Figure S2.6d, 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 (Figure S2.6e, 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 (Figure S2.6f), 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 (Figure 2.3h) reduced the expression of the NuRD complex subunit *MBD3* (Figure 2.3i, 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 (Figure 2.3j, 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 (Figure 2.3k, 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 (Figure S2.7a,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 (Figure S2.7c,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 (Figure S2.7e,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] (Figure S2.7g,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 (Figure 2.4a, 888 sgRNAs targeting 163 hDels), including putative *cis*-regulatory element-targeting sgRNAs[16] and non-targeting sgRNAs as positive and negative

controls, respectively (Figure S2.8a). 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 (Figure 2.4a, Direct-capture Perturb-seq). After filtering, we recovered 18,571 sgRNA-expressing cells, detecting multiple sgRNAs per cell (Figure 2.4b, Figure S2.8b-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 (Figure 2.4c). We identified 16 hDel-gene pairs within 100 kb (Figure 2.4d, Figure S2.8e, 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 (Figure 2.4h), sgRNAs targeting hDel_585 reduced the expression of *HADHA* (Figure 2.4i, 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 (Figure S2.9, PFC) revealed that hDel_585 is accessible in radial glia, intermediate progenitor cells, and excitatory and inhibitory neurons (Figure 2.4j). 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 (Figure 2.4k), 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] (Figure 2.4l-o). Additionally, sgRNAs targeting hDel_3779 reduced the expression of the RAC1 effector *FAM49B* (Figure 2.4l, 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* (Figure 2.4m, 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] (Figure 2.4o, 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 (Figure 2.4o, 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.

ACKNOWLEDGEMENTS

We thank Nadav Ahituv, Jingwen Ding, Luke Gilbert, Max Haeussler, Craig Lowe, Aaron McKenna, Katie Pollard, Joseph Replogle, Demian Sainz, 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.

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 (Figure S2.7; 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 (10^{th} percentile), 5,105 UMIs (10^{th} 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 (10^{th} percentile), 2,744 UMIs (10^{th} 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.

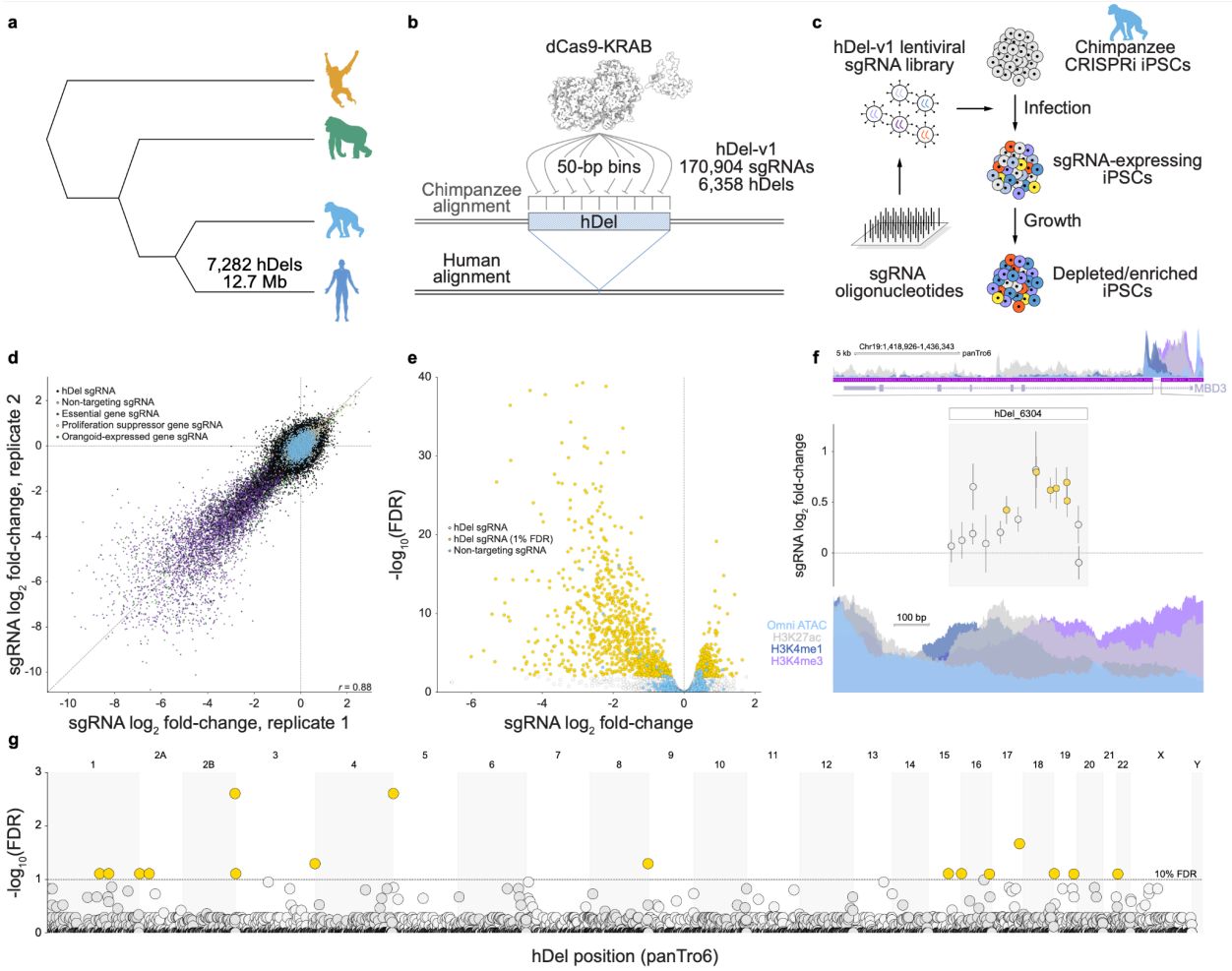


Figure 2.1. Genome-scale tiling CRISPRi-based genetic screens identify hDels modifying cellular proliferation.

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).

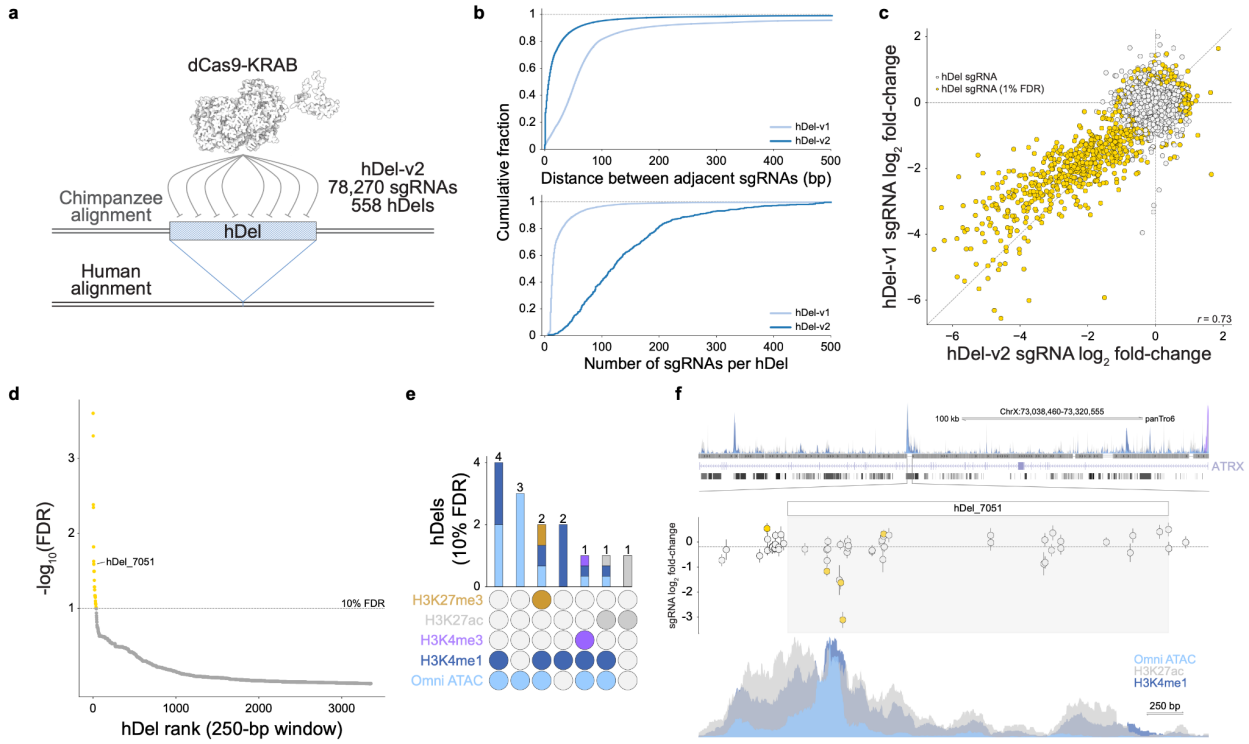


Figure 2.2. High-density tiling CRISPRi screen refines the boundaries of functional sequences within proliferation-modifying hDels.
a, High-density CRISPRi-based tiling of hDels (hDel-v2).
b, Distribution of the distance between adjacent hDel-targeting sgRNAs (top) and the number of sgRNAs per hDel (bottom) 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.

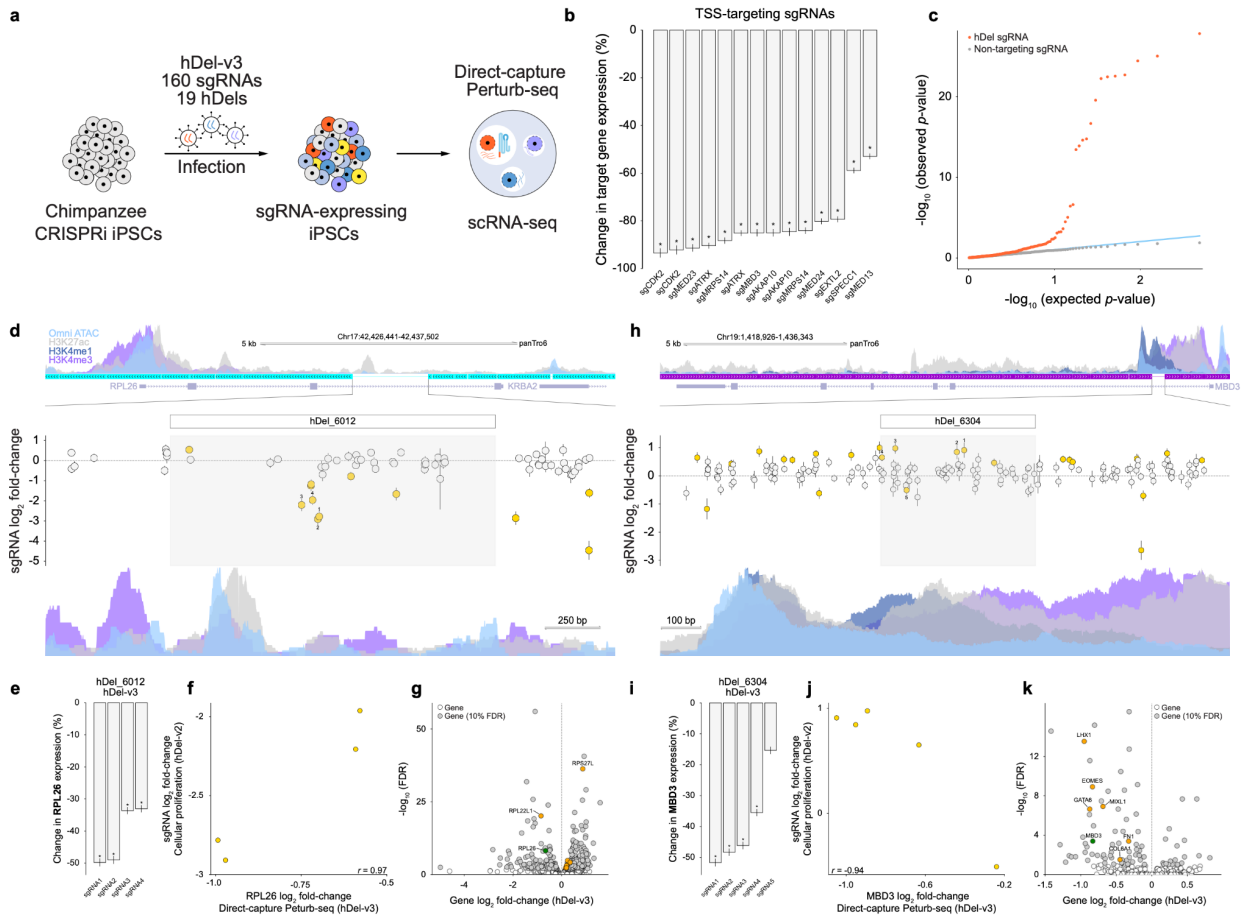


Figure 2.3. Mapping *cis*- and *trans*-regulatory target genes of proliferation-modifying hDels.

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).

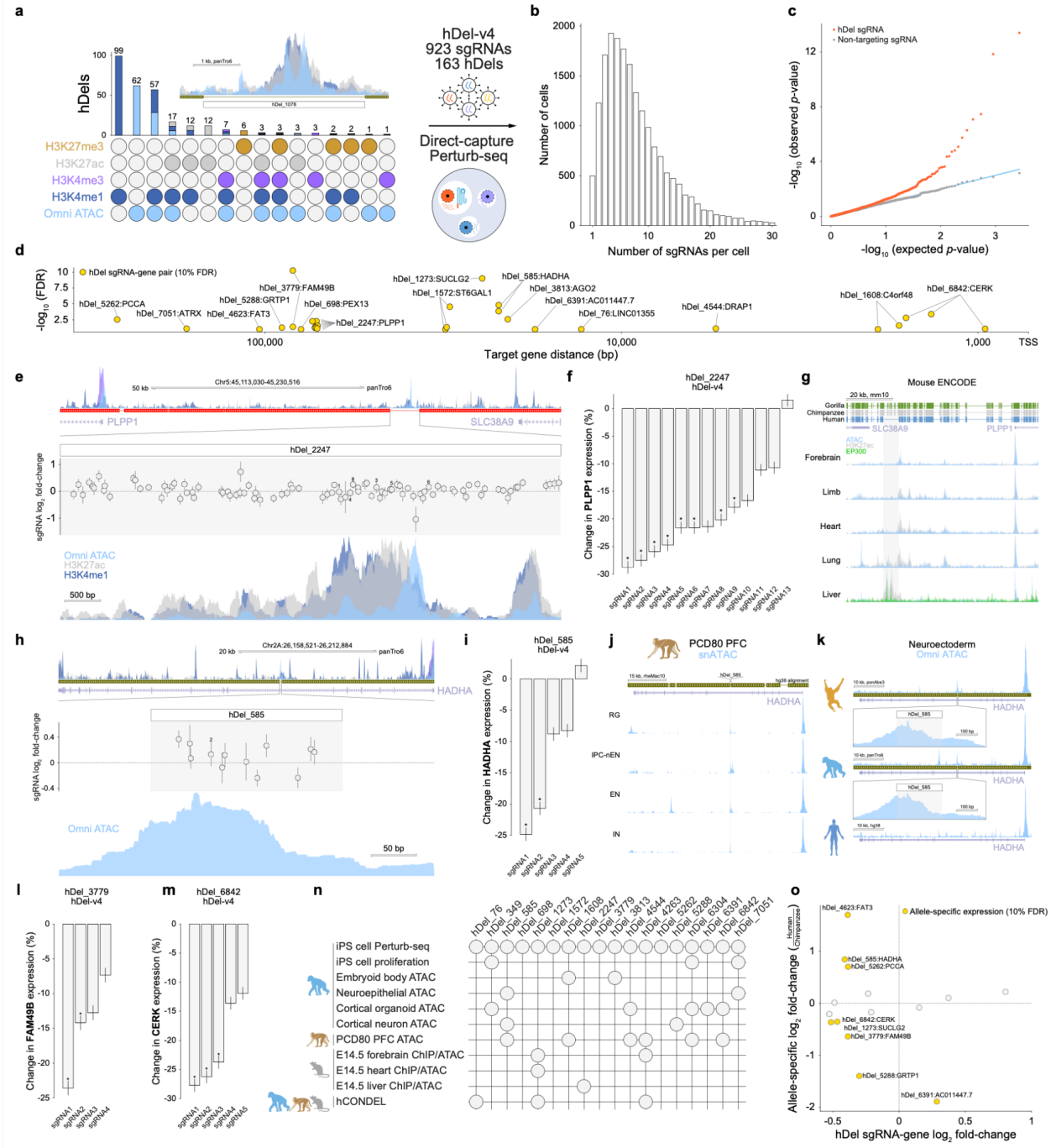


Figure 2.4. Nonessential hDels harbor *cis*-regulatory elements.
a, Upset plot of Omni ATAC-seq, H3K4me1, H3K4me3, H3K27ac, and H3K27me3 intersecting hDels and high multiplicity of infection single-cell CRISPRi screening approach (hDel-v4). Inset: hDel_1078.
b, Distribution of the number of sgRNAs per 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).

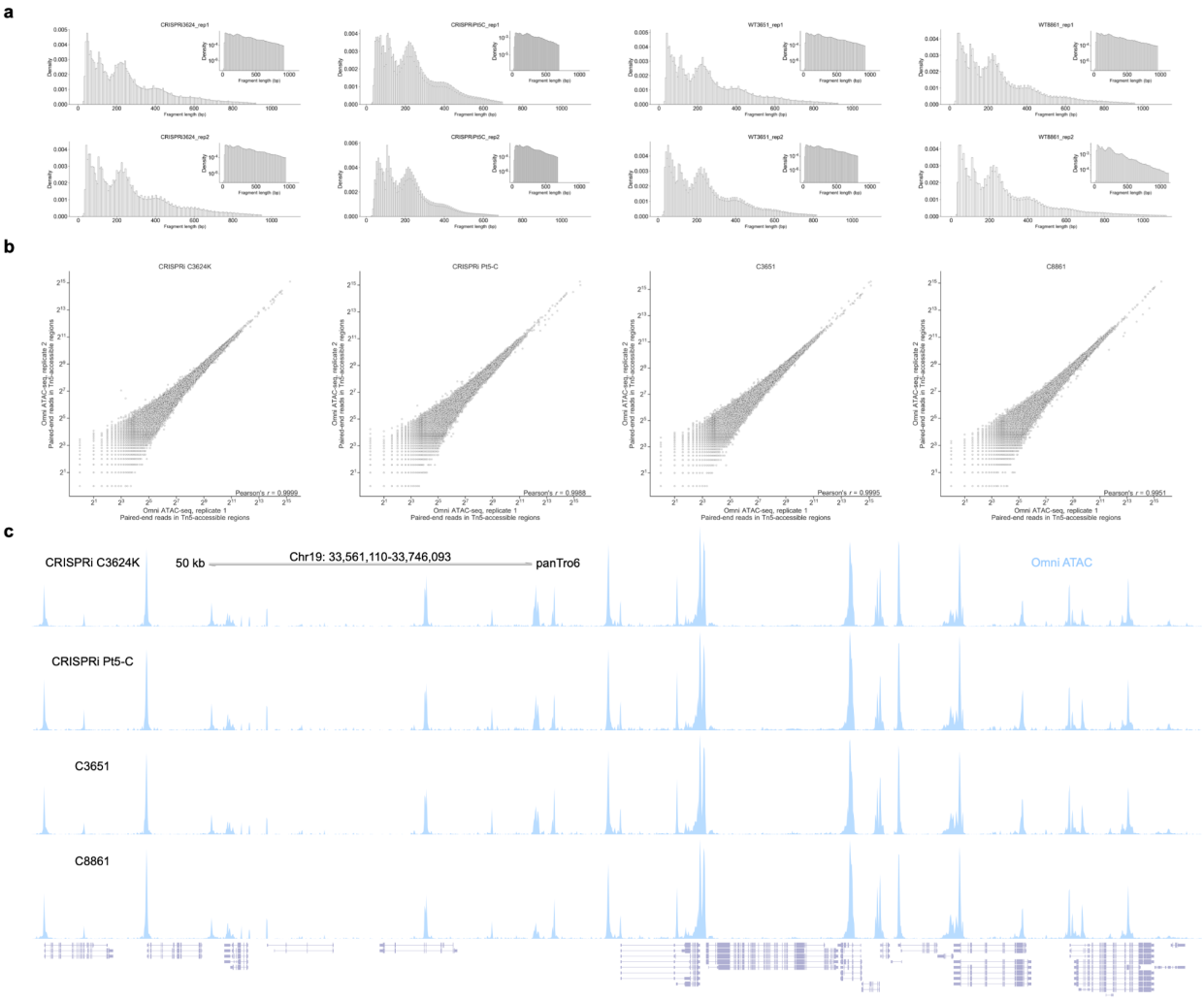


Figure S2.1. Omni ATAC-seq of chimpanzee iPS cells.
a, Distribution of Omni ATAC-seq fragment sizes. Inset: log-transformed histogram.
b, Correlation of reads within Tn5-accessible regions for Omni ATAC-seq technical replicates. Each point represents a Tn5-accessible region in any of the four iPS cells lines (5% FDR).
c, Omni ATAC-seq across a 184 kb region of the chimpanzee reference genome (panTro6).

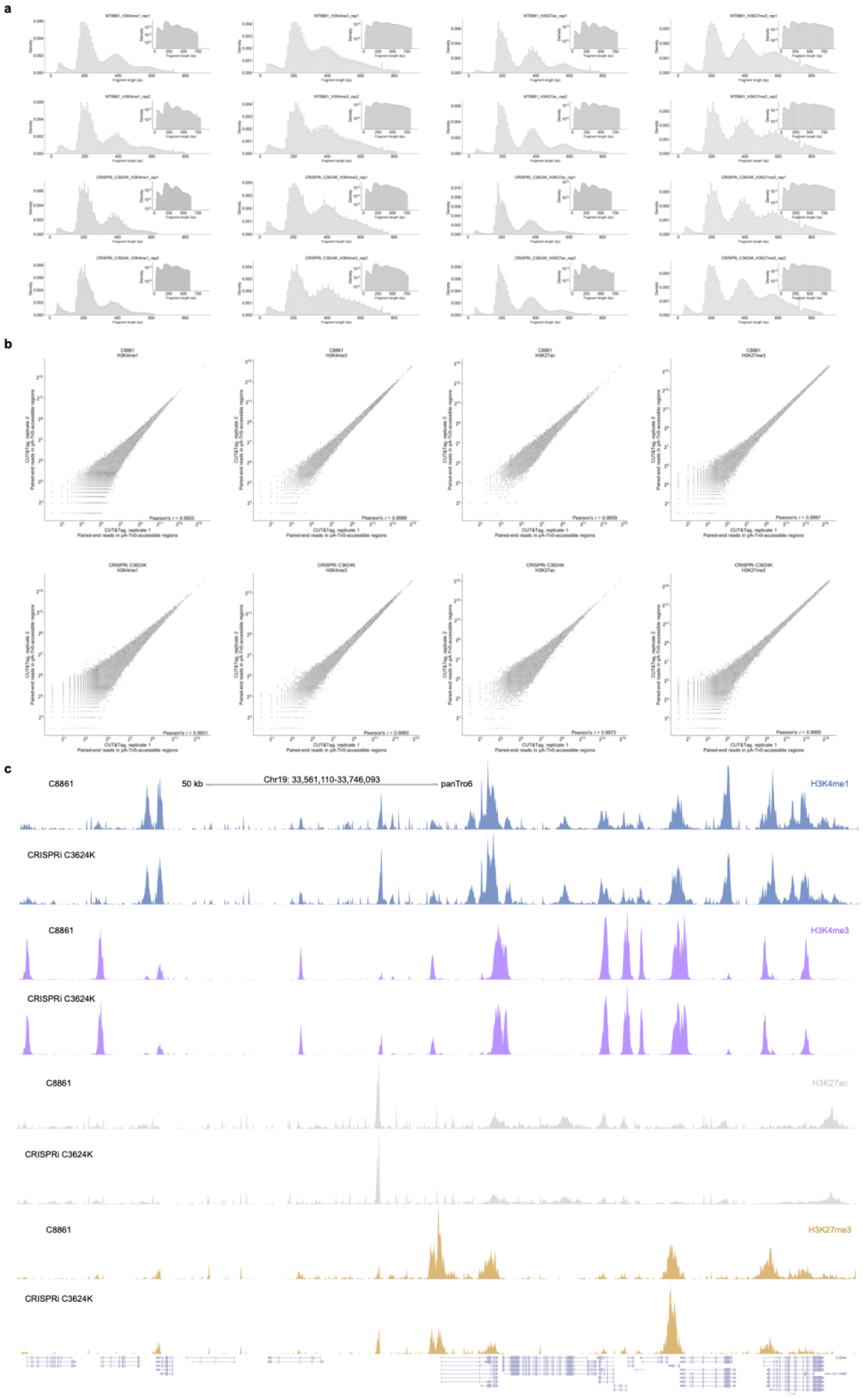


Figure S2.2. Profiling histone modifications with CUT&Tag of chimpanzee iPS cells.
a, Distribution of CUT&Tag fragment sizes for H3K4me1^{ab8895}, H3K4me3^{ab8580}, H3K27ac^{ab4729}, and H3K27me3^{9733S}. Inset: log-transformed histogram.

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).

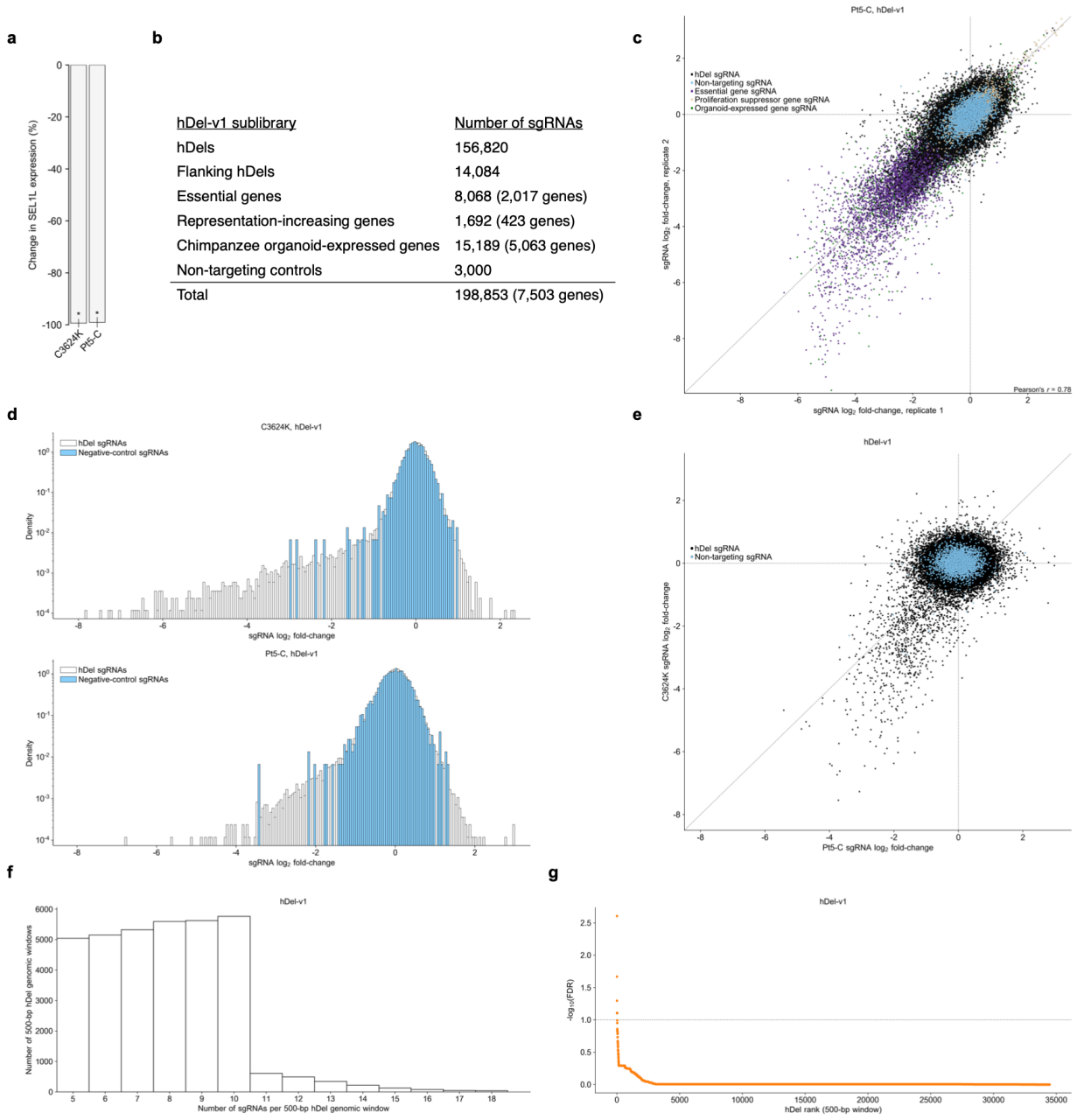


Figure S2.3. hDel-v1 CRISPRi-based genetic screens.
a, Change in *SEL1L* expression for chimpanzee CRISPRi iPS cells (C3624K, Pt5-C) harboring *SEL1L*-targeting sgRNAs compared to non-targeting sgRNAs as assessed by RT-qPCR.
b, hDel-v1 sublibraries.
c, Scatterplot of sgRNA $\log_2$ fold-change for hDel-v1 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.

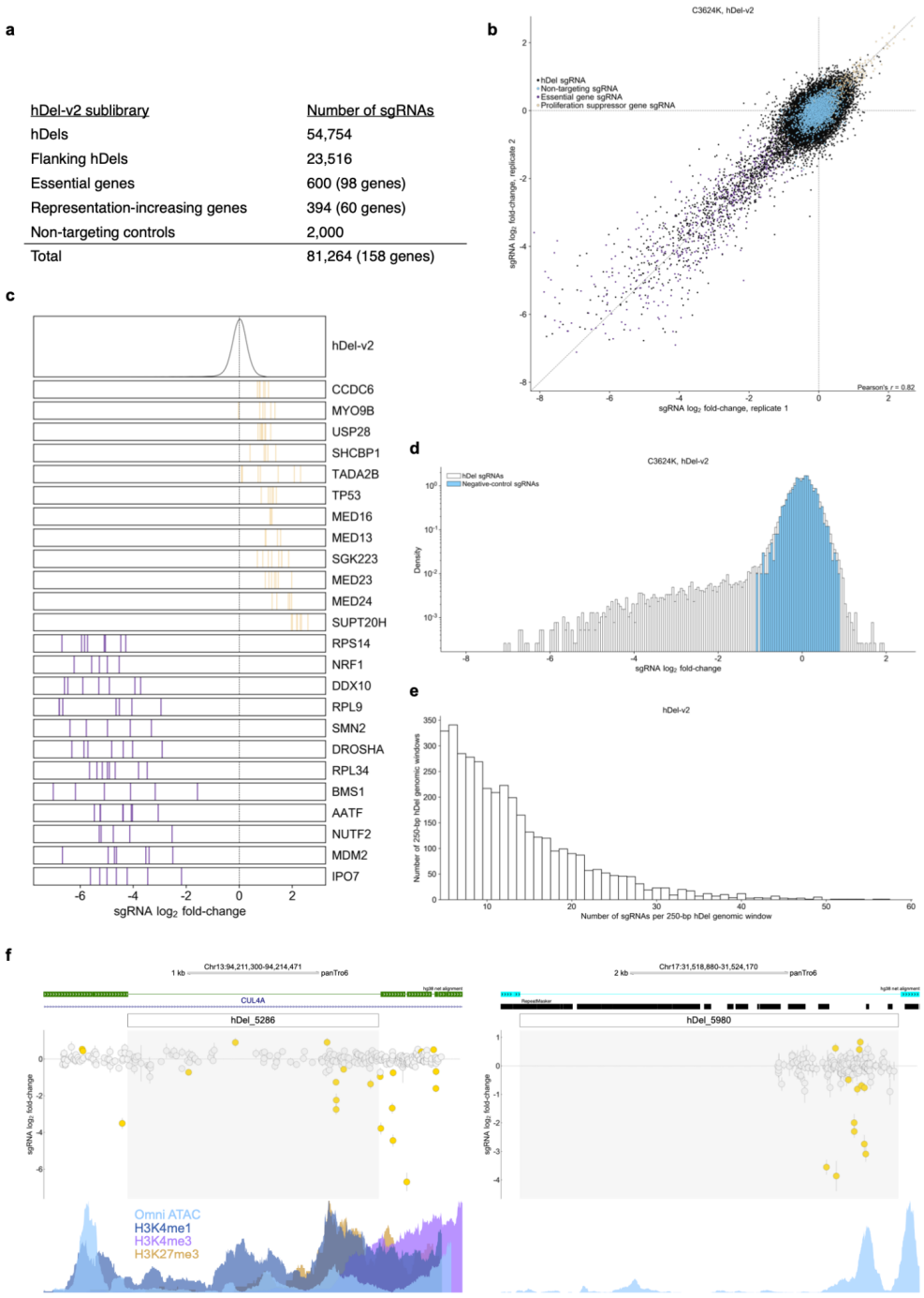


Figure S2.4. hDel-v2 CRISPRi-based genetic screen.
a, hDel-v2 sublibraries.
b, Scatterplot of sgRNA $\log_2$ fold-change for hDel-v2 technical replicates in C3624K.
c, Rug plot of sgRNA $\log_2$ fold-change for the 12 most enriched (khaki) and depleted genes (purple) as ranked by average sgRNA $\log_2$ 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

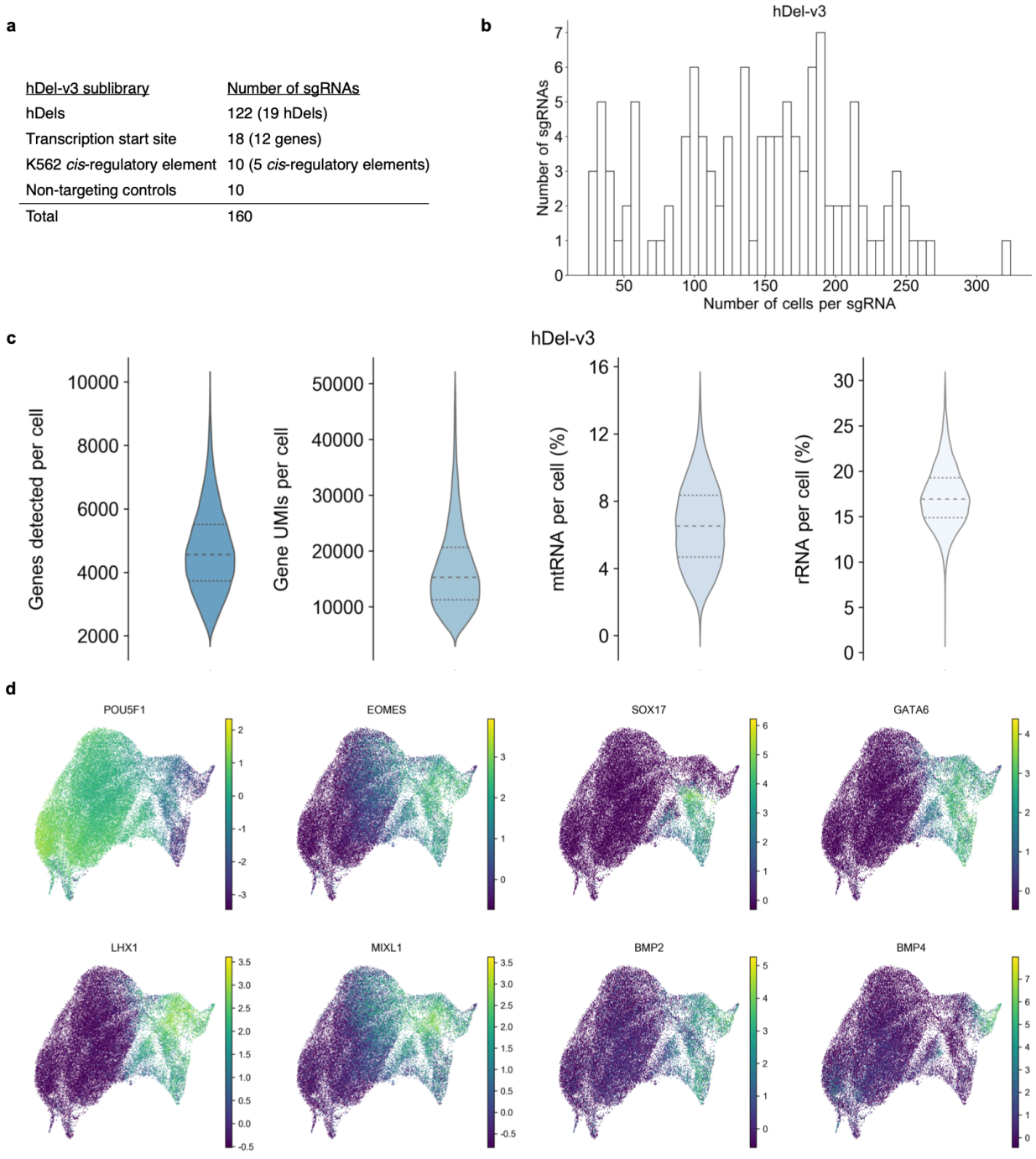


Figure S2.5. hDel-v3 Direct-capture Perturb-seq.
a, hDel-v3 sublibraries.
b, Distribution of the number of cells per sgRNA.
c, Distribution of the number of genes detected (far left), gene UMIs (center left), percent mtRNA (center right), and percent rRNA (far right) per cell.
d, UMAP projections colored by the normalized, log-transformed, and scaled expression of the indicated genes.

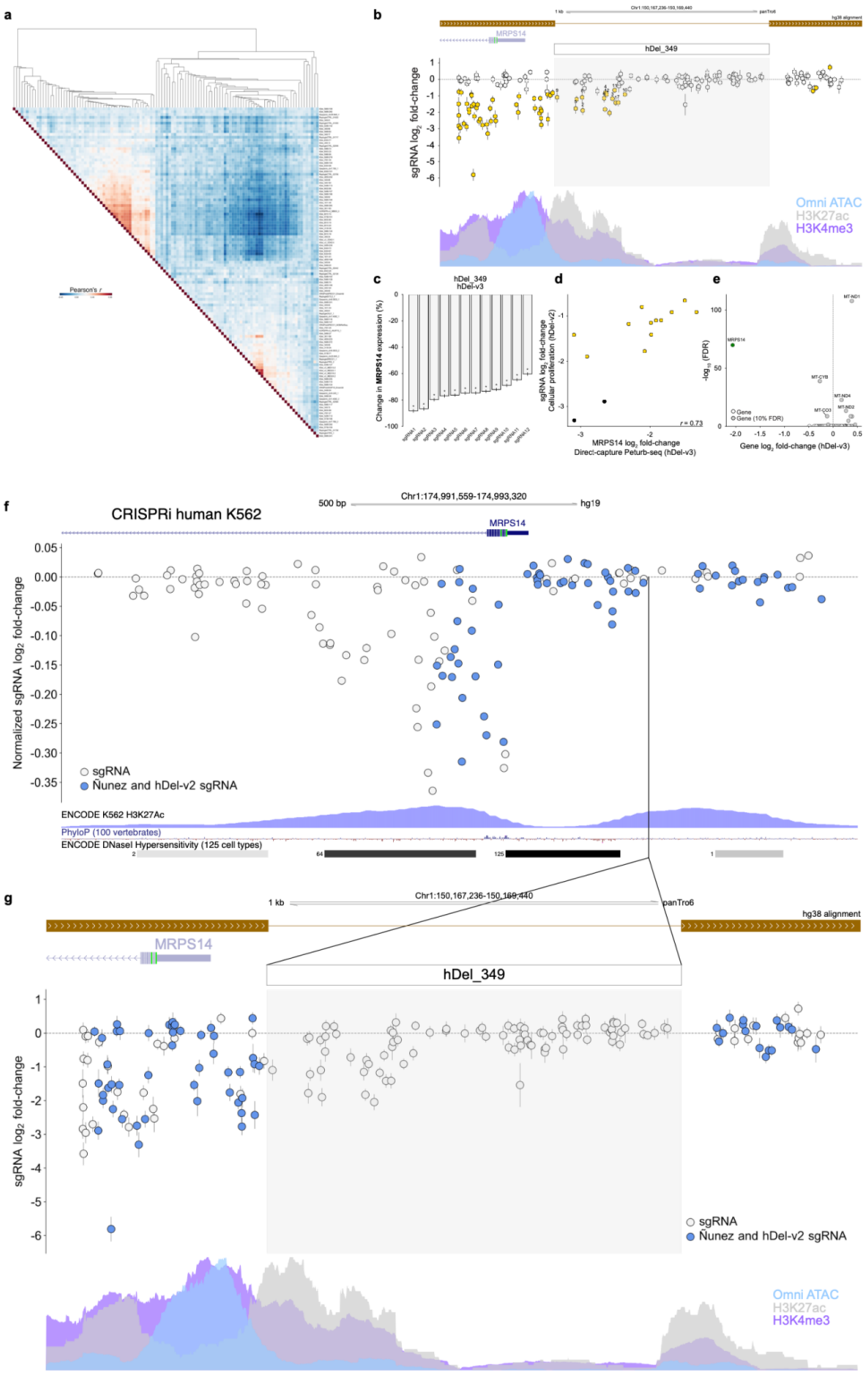


Figure S2.6. hDel-v3 *trans* differential expression and hDel_349.
a, hDel-v3 sublibraries.
b, Heatmap of Pearson correlations of gene $\log_2$ fold-change and hierarchical clustering for differentially expressed genes. The union of any gene differentially expressed (n = 2,864 genes, FDR < 0.1) in a 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.

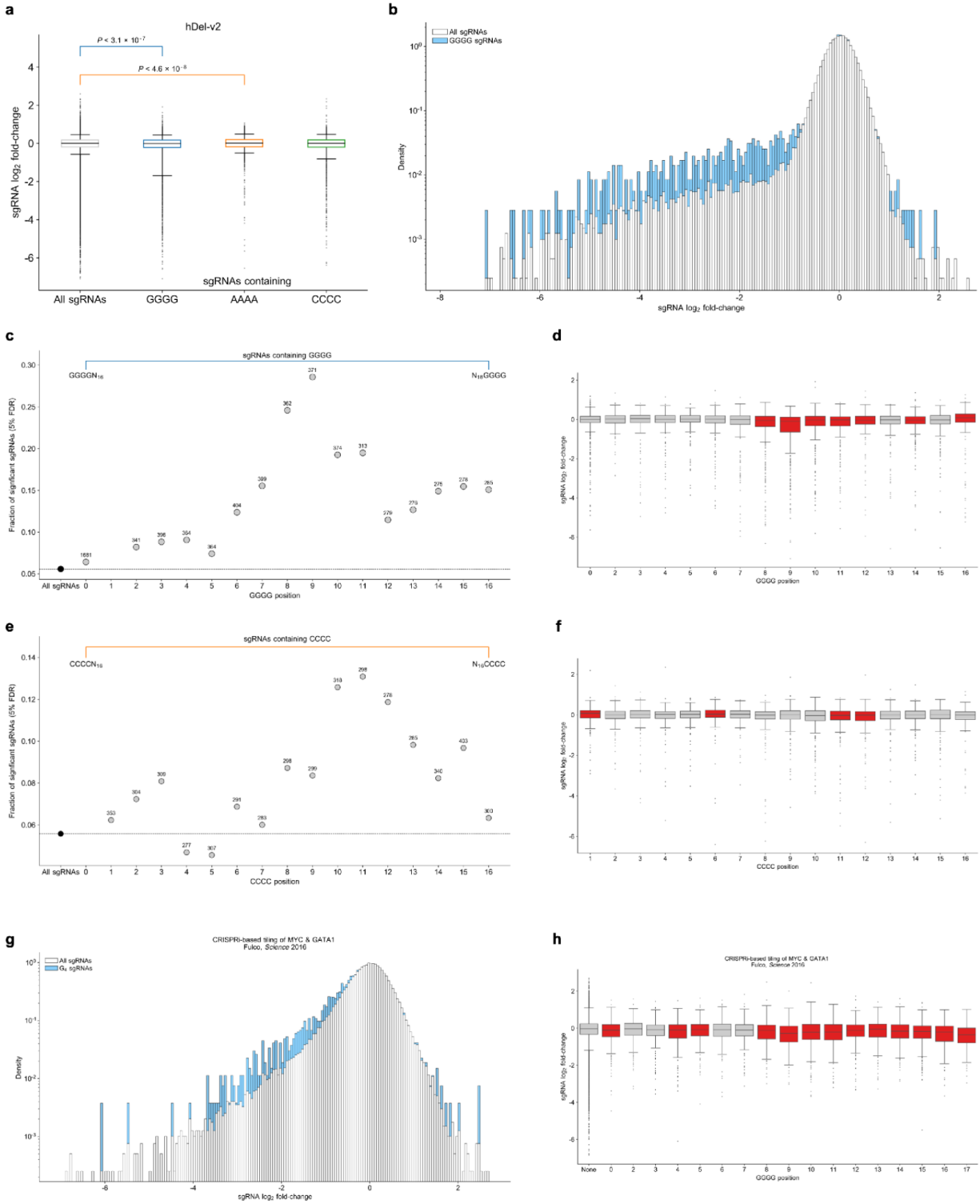


Figure S2.7. sgRNA nucleotide homopolymer-associated toxicity.
a, Boxplots of sgRNA $\log_2$ fold-change for all hDel-v2 sgRNAs (far left), sgRNAs containing G_4 nucleotide homopolymers (center left), sgRNAs containing A_4 nucleotide homopolymers (center right), and sgRNAs containing C_4 nucleotide homopolymers (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).

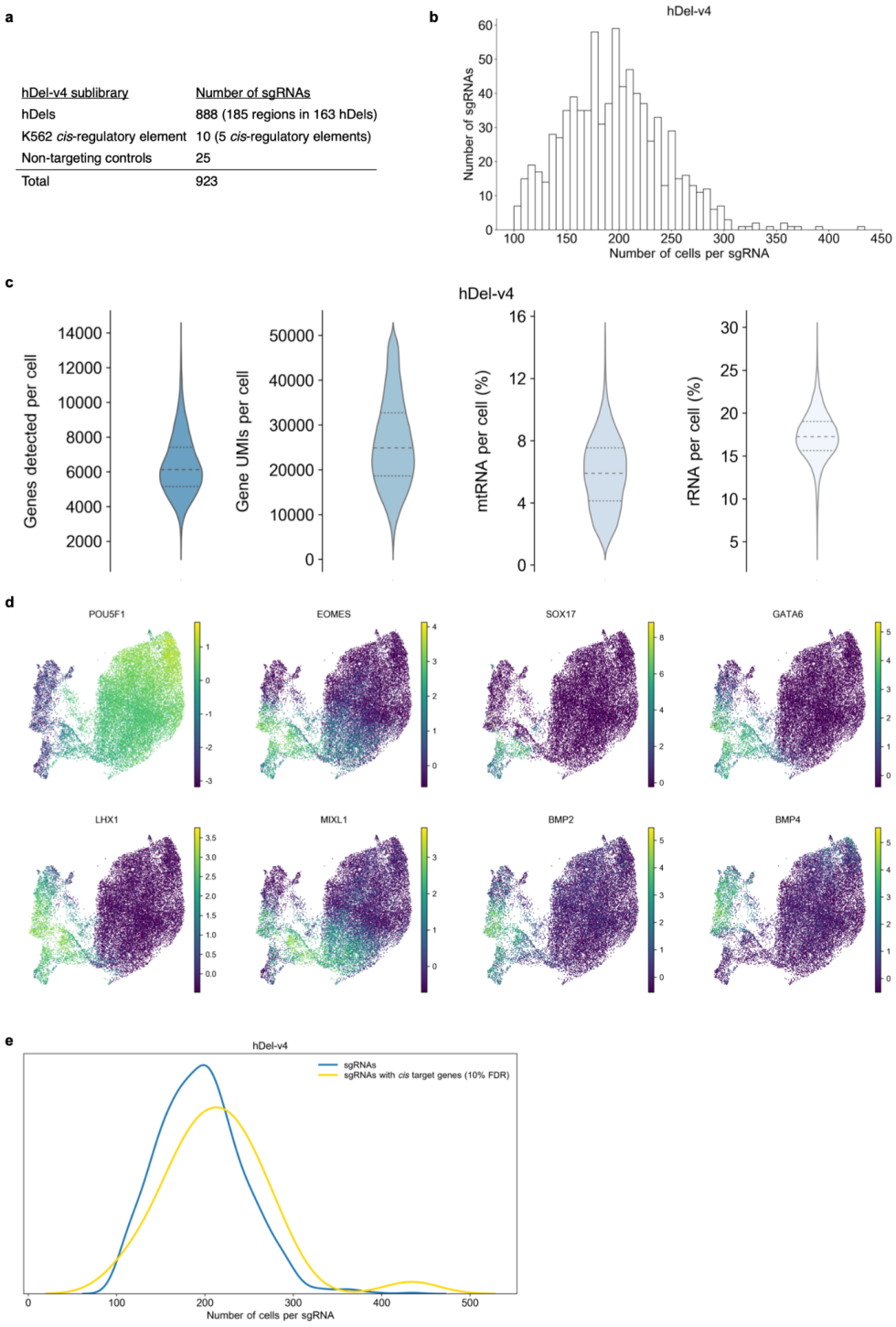


Figure S2.8. hDel-v4 Direct-capture Perturb-seq.
a, hDel-v4 sublibraries.
b, Distribution of the number of cells per sgRNA.
c, Distribution of the number of genes detected (far left), gene UMIs (center left), percent mtRNA (center right), and percent rRNA (far right) per cell.
d, UMAP projections colored 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).

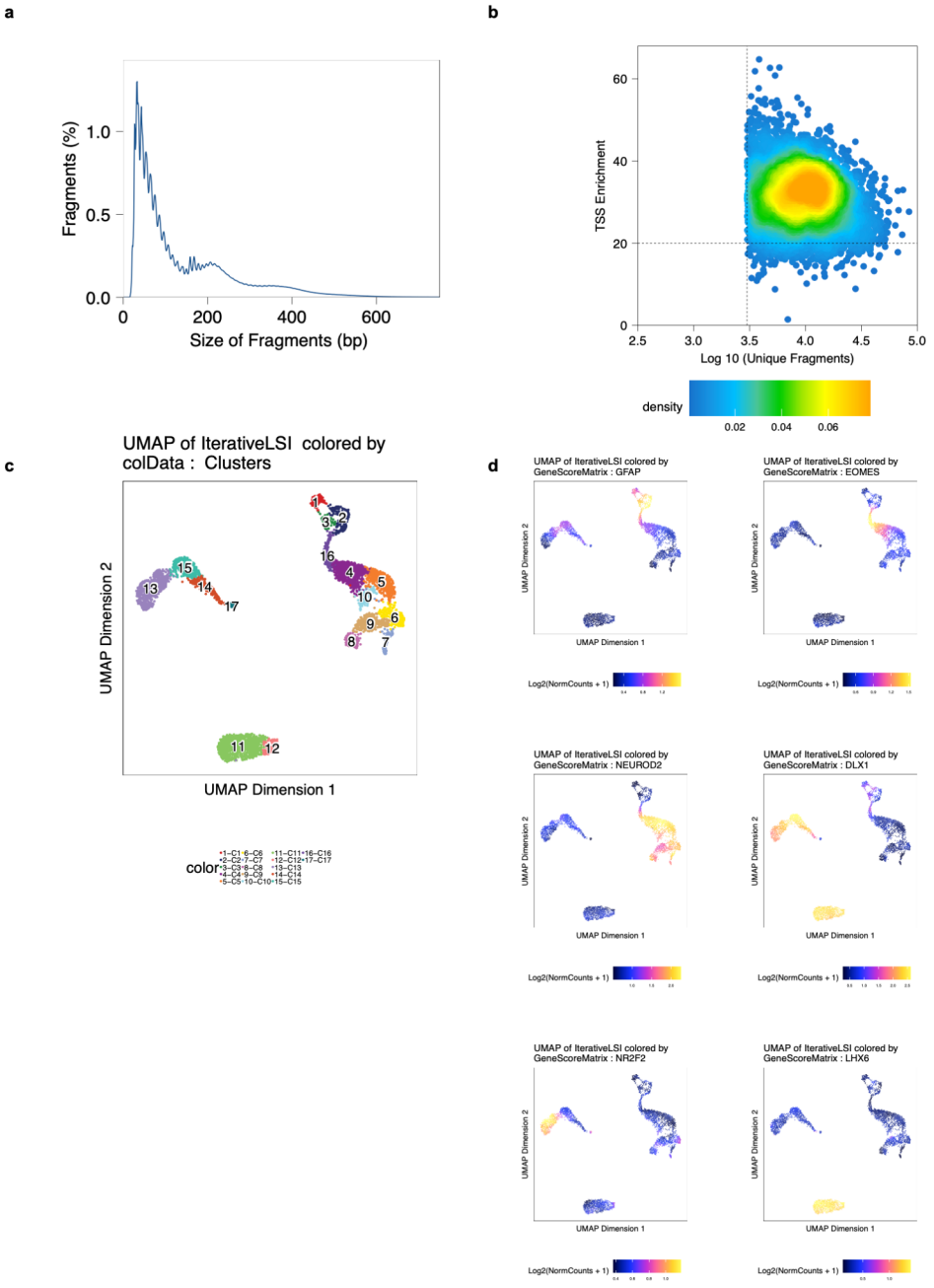


Figure S2.9 snATAC-seq of PCD80 rhesus macaque prefrontal cortex.
a, Distribution of snATAC-seq fragment sizes.
b, Density plot of the number of fragments and ArchR TSS enrichment score per cell.
c, UMAP projection colored by clusters (*n* = 17), including radial glia (RG, 1, 2, 3), intermediate progenitor 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.

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.

Chapter III

Comparative landscape of genetic dependencies in human and chimpanzee stem cells

Richard She[1,9], Tyler Fair[2,3,9], Nathan K. Schaefer[2,7], Reuben A. Saunders[1,4], Bryan J. Pavlovic[2,7], Jonathan S. Weissman[1,5,6,8,*], and Alex A. Pollen[2,7,10,*]

Affiliations

[1]Whitehead Institute for Biomedical Research, Cambridge, MA, USA

[2]Eli and Edythe Broad Center of Regeneration Medicine and Stem Cell Research, University of California, San Francisco, San Francisco, CA, USA

[3]Biomedical Sciences Graduate Program, University of California, San Francisco, San Francisco, CA, USA

[4]Department of Cellular and Molecular Pharmacology, University of California at San Francisco, San Francisco, CA, USA

[5]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA

[6]Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, MA, USA

[7]Department of Neurology, University of California, San Francisco, San Francisco, CA, USA

[8]David H. Koch Institute for Integrative Cancer Research, Massachusetts Institute Technology, Cambridge, MA 02142, USA

[9]These authors contributed equally

[10]Lead contact

*Correspondence: weissman@wi.mit.edu (J.S.W.), alex.pollen@ucsf.edu (A.A.P.)

https://doi.org/10.1016/j.cell.2023.05.043

ABSTRACT

Comparative studies of great apes provide a window into our evolutionary past, but the extent and identity of cellular differences that emerged during hominin evolution remain largely unexplored. We established a comparative loss-of-function approach to evaluate whether human cells exhibit distinct genetic dependencies. By performing genome-wide CRISPR interference screens in human and chimpanzee pluripotent stem cells, we identified 75 genes with species-specific effects on cellular proliferation. These genes comprised coherent processes, including cell-cycle progression and lysosomal signaling, which we determined to be human-derived by comparison with orangutan cells. Human-specific robustness to *CDK2* and *CCNE1* depletion persisted in neural progenitor cells and cerebral organoids, supporting the G1-phase length hypothesis as a potential evolutionary mechanism in human brain expansion. Our findings demonstrate that evolutionary changes in human cells reshaped the landscape of essential genes and establish a platform for systematically uncovering latent cellular and molecular differences between species.

INTRODUCTION

Comparative studies of humans and chimpanzees, our closest extant relatives, have long sought to define the evolutionary origins of unique human features. Within 7 million years, humans evolved numerous specializations, from bipedalism to the 3-fold expansion of the cerebral cortex.[1,2] Many of these novel human traits emerge from changes in cell behavior during development. These changes in cell behavior may in turn reflect underlying differences in molecular circuits modified during the short timescale of hominin evolution. However, we currently lack a framework for systematically identifying which molecular pathways play divergent roles in conserved developmental cell types.

Current approaches for studying the molecular basis of human evolution include reconstructing candidate mutations at specific loci in model organisms, but only a handful of mutations in non-coding regulatory regions and coding genes has been examined in detail. Among conserved non-coding elements with unexpected changes in the human lineage, specific loci have been linked to gene expression changes in distal limbs,[3] increased sweat gland number,[4] and increased neural proliferation.[5] Among coding changes, two human-specific coding mutations in *FOXP2* have been proposed to contribute to human language capabilities, based on functional studies in mouse models and human genetics,[6,7] and three modern human-specific mutations in *KIF18A* and *KNL1* prolong metaphase and reduce segregation errors in neural progenitor cells (NPCs).[8] In addition, recent duplications and subsequent modifications of *ARHGAP11B* and *NOTCH2NL* have been implicated in the expansion of the human cortex,[9–12] supporting predictions that human-specific mutations may influence proliferation of NPCs during development.[13,14] Nonetheless, connecting individual candidate mutations to evolved human traits remains challenging because most mutations are neutral or low effect size, analyses are low throughput, and we lack a detailed understanding of the divergence in cellular and developmental phenotypes that ultimately give rise to species differences.

In parallel, high-throughput genomics-based approaches have described gene regulatory changes that may contribute to species differences. Because ape primary tissue is largely inaccessible during early development, recent studies have employed stem cell-derived models as an experimentally tractable system for comparative analyses of species differences during development. Thousands of cell-type-specific gene expression differences have been identified in pluripotent stem cells (PSCs),[15,16] cardiomyocytes,[17] endoderm,[18] neural crest,[19] and cortical neurons.[20–22] However, these gene expression differences comprise a mixture of neutral changes, causal changes, and indirect downstream consequences, and genes that mediate species differences may have conserved expression. Therefore, it can be difficult to ascertain which molecular changes, among hundreds or thousands, drive differences in cellular physiology.

The history of developmental genetics provides a template for linking the function of individual genes to organismal phenotypes. Early mutagenesis screens in *Drosophila melanogaster* identified genes critical for body axis patterning.[23,24] Many of these genes belonged to highly conserved cell signaling pathways that also coordinate development in vertebrates, such as Wnt,[25,26] Hedgehog,[27] and BMP.[28] More recent efforts in organismal screening involve several international consortia that have generated large collections of knockout mice to investigate more complex vertebrate phenotypes.[29–31] The success of these genetic approaches has resulted in the functional annotation of many of the genes that guide mammalian development. However, despite the conservation of many core developmental principles from fruit flies to mice and to humans, these shared molecular functions do not account for how our species evolved to be different.

To apply functional genomics approaches to questions of species divergence, we harnessed recent advances in CRISPR-based technologies that have enabled genome-scale perturbation

screens across thousands of human cell lines.[32–34] These efforts have mapped landscapes of genetic dependencies with an enrichment of essential genes in coherent pathways that typically cluster by cell type of origin.[35,36] Extending this approach to studies of comparative evolution could reveal genes or cellular processes with divergent functional roles in homologous cell types. Illuminating the extent and identity of recently evolved genetic dependencies would complement individual candidate gene approaches, descriptive comparative genomics analyses, and single-species loss-of-function studies. However, whether genetic dependencies diverged in closely related hominin species and how this knowledge could reveal previously unappreciated differences in cellular physiology remain unexplored.

To evaluate the extent of conservation and divergence in genetic dependencies between human and chimpanzee, we established a comparative loss-of-function screening approach in PSCs. PSCs are a model for the earliest stages of development, capturing features of the inner cell mass of the blastocyst, including the capacity to differentiate into all germ layers at a stage that precedes species differences in developmental timing and cell-type composition. The state of pluripotency is well conserved between human and chimpanzee PSCs at the level of the transcriptome, epigenome, and cell fate potential,[15] providing a homologous cell type for species comparison. In addition, PSCs have greater levels of open chromatin and gene expression than somatic cells,[37] enabling large-scale study of gene function for genes later expressed in diverse cell types. As PSCs are poised to self-renew or differentiate into all germ layers based on environmental cues, we reasoned that changes in proliferation in PSCs could provide a sensitive measure for species-specific responses to a wide range of genetic perturbations.

Performing genetic screens using an *in vitro* model confers several advantages that could support isolation of molecular and cellular species differences. First, the ability to grow large numbers of PSCs enables a pooled library approach with multiple redundant library elements targeting each

gene. Second, laboratory cell culture provides a well-defined and highly controlled environment, which minimizes extrinsic sources of variation. Lastly, the scalability of pooled screening allows for retesting of each cellular phenotype in PSCs derived from multiple individuals of each species to account for individual variation within a species. Thus, we conducted genome-wide CRISPR interference (CRISPRi) screens in human and chimpanzee PSCs. Despite high levels of conservation, our screens revealed that genetic dependencies can diverge in short evolutionary timescales, that species differences are organized into coherent pathways and protein complexes, and that human-specific changes have evolved in gene networks promoting G1/S progression in PSCs and NPCs. In addition to these specific insights, our study establishes a broadly applicable experimental approach for uncovering latent molecular differences between closely related species.

RESULTS

Genome-wide CRISPRi screening in human and chimpanzee stem cells

To enable comparative CRISPR-based genetic screening, we engineered CRISPRi machinery[32] at the *CLYBL* safe harbor locus[38] in two human and two chimpanzee PSC lines (Figure 3.1A). For the two human individuals, we chose two well-characterized cell lines: WTC11[39–41] an induced pluripotent stem cell (iPSC) line; and H1,[42] an embryonic stem cell (ESC) line. For the chimpanzee individuals, we chose two iPSC lines used in previous studies: C3649 and Pt5-C.[15,43]

To identify genes that modify cellular growth and survival, we infected each cell line with the genome-wide lentiviral hCRISPRi-v2 sgRNA library[44] (5 sgRNAs/gene), selected for sgRNA-expressing cells with puromycin, cultured cells for 10 days, and quantified sgRNA enrichment and depletion by high-throughput sequencing. While hCRISPRi-v2 was designed to target the human genome, 77.4% of sgRNAs perfectly matched targets in the chimpanzee reference genome (panTro6); sgRNAs with mismatches were not considered for analyses of species differences

(Figure S3.1A). Across all four screens, we observed robust depletion of sgRNAs targeting common essential genes and enrichment of sgRNAs targeting proliferation suppressor genes. Analysis of technical and biological replicates revealed strong sgRNA correlations for replicates of the same cell line (Pearson's r = 0.80–0.97, Figure 3.1B) and for different cell lines within species (r = 0.69–0.83). In addition, all four genetic screens sensitively and precisely distinguished Dependency Map (DepMap) common essential and non-essential genes,[45,46] recalling 82.9%–92.6% of common essential genes at 95% precision (Figure 3.1C).

We next sought to identify genes with species-specific effects on cellular proliferation. To do so, we utilized MAGeCK[47] and developed a bootstrapping-based method that accounted for both the number of significantly enriched or depleted sgRNAs targeting a gene and the magnitude of sgRNA $\log_2$ fold change (Figure S3.1B). While the large majority of essential genes were shared between species (Figure 3.1B), we identified 583 candidate species-specific essential genes and 202 candidate species-specific proliferation suppressor genes (Figures 3.1D and S3.1C). Importantly, applying this approach to sets composed of one individual of each species identified far fewer candidate genes (n = 3–12 genes, Figure 3.1B), highlighting the influence of species on gene essentiality. As an additional quality control, we confirmed that sgRNAs targeting regulators of apoptosis such as *BAK1* and *BAD* were enriched in all four PSC screens (Figure S3.1D). We unexpectedly discovered that sgRNAs targeting *TP53* were enriched only in the two human cell lines, suggesting that the two chimpanzee PSC lines used in primary screening were p53-unresponsive. We reasoned that a fraction of the candidate species differences might thus be attributable to p53 status and would require additional testing with p53-responsive chimpanzee cell lines (Figures S3.1E and S3.1F). In conclusion, our approach shows that genome-wide CRISPRi screening can be applied to closely related species to establish a comparative essential-ome and to nominate candidate species differences.

Human and chimpanzee can be distinguished by genetic dependencies

We next sought to validate candidate species differences across multiple independently derived human and chimpanzee PSCs to distinguish species differences from those driven by individual variation,[48–50] adaptation to cell culture,[51] or by somatic cell reprogramming.[52] We engineered new CRISPRi stem cell lines from four human (H20961B, H21792A, H23555A, and H28126B) and four chimpanzee (C3624K, C8861G, C40280L, and C40290F) individuals. To minimize technical variation, we selected high-quality cell lines with normal karyotypes that were reprogrammed with identical protocols and maintained in identical media. Cell lines from both species were previously shown to differentiate into all three germ layers via teratoma formation and embryoid body assays, functionally validating pluripotency.[15] Finally, the human and chimpanzee lines were previously shown to share comparable pluripotency scores with strongly overlapping patterns of H3K27me3 and H3K27ac at pluripotency genes[15] and similar transcriptional trajectories of differentiation,[18] suggesting that the cell lines were in a comparable state of pluripotency. We used bulk RNA-seq to verify that the CRISPRi-engineered cell lines exhibited comparably high expression levels of canonical pluripotency markers *OCT4*, *SOX2*, and *NANOG*, compared with the original source lines (Figures S3.2A and S3.2B). No significant expression differences were observed for these genes in our human- and chimpanzee-engineered cell lines ($p = 0.42$, 0.98, and 0.94, two-tailed t test). We analyzed copy-number variation using CaSpER[53] and genome sequencing coverage to rule out the presence of large duplications or deletions (Figure S3.2C). In addition, we assessed CRISPRi cell lines for p53 responsiveness by measuring sensitivity to nutlin-3a, a small molecule MDM2 inhibitor that induces p53-dependent autophagy and apoptosis.[54,55] All new human and chimpanzee lines plus WTC11 and H1 were MDM2/p53-responsive, while C3649 and Pt5-C, the chimpanzee lines used for genome-scale screening, were non-responsive to MDM2/p53 perturbations (Figures S3.1E and S3.1F).

To enable secondary screening, we designed a comparative essential validation (CEV-v1) library

consisting of 7,847 sgRNAs targeting the transcriptional start sites of 963 genes from our genome-scale datasets (8 sgRNAs per gene) and 1,845 negative-control sgRNAs (Figures 3.2A and S3.2D–S3.2G). Due to the scalability of pooled screening, we targeted an inclusive set of genes with significant or suggestive differences between species in the primary screens as well as gene families with notable evolutionary histories.[56,57] To reduce human-specific bias, we required every sgRNA in CEV-v1 to perfectly match target sites in the human (hg38) and chimpanzee (panTro6) reference genomes.[58] In total, we performed 16 CRISPRi screens, using the CEV-v1 sgRNA library (Figure 3.2A). The validation screens were performed in the four newly constructed CRIPSRi PSC lines of each species. In addition, we retested three of the four PSC lines used for genome-scale screening (Pearson's r = 0.76–0.89) and performed biological replicate screens in separate laboratories for five cell lines (three human lines and two chimpanzee lines, r = 0.70–0.92). Notably, hierarchical clustering of cell lines by similarity of sgRNA profiles separated all the human (including ESCs and iPSCs) from all the chimpanzee individuals (Figure 3.2B). Decomposition of each cell line's sgRNA profile by principal-component analysis also grouped individuals by species, with the main axes of variation relating to shared changes in sgRNA representation over time (PC1) and species-specific changes (PC2) (Figure 3.2C). Together, our findings show that stem cells from humans and chimpanzees can be distinguished by their responses to genetic perturbations.

Molecular nature of core species-specific genetic dependencies

We next sought to identify genes underlying the differences between human and chimpanzee sgRNA profiles. We used DESeq2[59] to model sgRNA counts from all human and chimpanzee CEV-v1 screens and contrasted the species terms to detect sgRNAs with divergent effects on cellular proliferation. We identified 1,133 sgRNAs with evidence for differences between species (1% false discovery rate [FDR], |chimpanzee−human $\log_2$ fold change| ≥ 0.5), while negative-control sgRNAs were tightly distributed around zero (Figure 3.2D). Using α-RRA[47] to combine

sgRNA p values, we found 75 genes with robust species-specific effects on cellular proliferation at a 1% FDR (Figures 3.2E–3.2G and S3.3). This reduction in the number of genes from the CEV-v1 library was the result of a combination of factors: (1) the use of more stringent statistical thresholds, (2) a fraction of genes that replicated in validation screens of the original cell lines but not in the additional cell lines, and (3) the exclusion of genes whose effects on cellular proliferation depended on *TP53* status. Together, these findings reveal a stringent set of species-specific genetic dependencies that emerged in recent human and chimpanzee evolution.

We first explored whether species-specific genetic dependencies could relate to changes in the coding sequence or regulation of the target genes themselves that might suggest divergent species-specific activities of these genes. Several genes in the set exhibited unexpected coding sequence changes. For example, *ASPM*, which causes microcephaly when mutated, contains protein domains with signatures of positive selection in the human lineage[60–63] and was essential in human but not chimpanzee PSCs. Similarly, KATNA1, which physically interacts with ASPM to promote microtubule disassembly at mitotic spindle poles,[64] contains a nearly fixed modern human-specific mutation that is distinct from the Neanderthal and chimpanzee allele[65] and acted as a suppressor of proliferation in chimpanzee but not human cells. However, these examples were exceptions, and signatures of adaptive selection, as well as overall non-synonymous substitutions, were depleted among the set of 75 species-specific genetic dependencies, compared with the genome-wide distribution ($p < 0.01$ and $p < 10{-}6$, respectively, Kolmogorov-Smirnov test, Figures S3.4A and S3.4B).[66] Several genes with divergent genetic dependencies also displayed quantitative gene expression changes. For example, *MTCH2*, a gene involved in mitochondrial metabolism and apoptosis,[67] displayed significantly higher expression in human PSCs (fold change = 1.28, FDR < 10−3)15 and was specifically essential in human PSCs. In contrast, *ACAT2*, a gene involved in lipid metabolism, exhibited significantly higher expression in chimpanzee PSCs (fold change = 2.73, FDR < 10−6)[15] but was also specifically essential in

human cells. Despite these examples, the 75 gene set was also depleted for species differences in gene expression ($p = 0.035$, Figure S3.4C, two-tailed t test). Together, these analyses suggest that coding or regulatory changes in CRISPRi target genes themselves do not account for the majority of species-specific growth differences we observed and that a multitude of indirect effects driven by genetic background may underlie divergent dependencies.[68]

We next asked whether species-specific genetic dependencies involved groups of genes known to interact, a pattern that could suggest divergent requirements for conserved pathways. As essentiality phenotypes are typically shared among genes within known functional modules,[69,70] such coherence could also provide an additional test of internal consistency. Indeed, functionally related genes emerged with consistent patterns of depletion or enrichment within each species. Analysis using the STRING database[71] revealed an enrichment for protein–protein interactions ($p < 10-6$, p value calculated by string-db.org) and components of several well-established biological processes (Figure 3.3A). For example, we observed that all five core components of the UFMylation pathway (*UFM1*, *UFL1*, *UFC1*, *UBA5*, and *DDRGK1*) were essential only in human PSCs (Figure S3.4D). By contrast, all four subunits of the MOZ histone acetyltransferase complex (*KAT6A*, *BRPF1*, *ING5*, and *EAF6*) acted as proliferation suppressors in chimpanzee PSCs (Figure S3.5D). Accessory proteins to the vacuolar-type ATPase (*ATP6AP1* and *ATP6AP2*) and the highest ranking DepMap co-dependent gene *WDR7*[36] were specifically essential in human PSCs, whereas core subunits were essential in both species (Figure 3.3B). Strikingly, human PSCs were robust to depletion of cell-cycle regulators cyclin-dependent kinase 2 (*CDK2*), its activating partner *cyclin E1* (*CCNE1*), and cyclin-dependent kinase 4 (*CDK4*). For all three genes, we observed at least six sgRNAs that were essential across six chimpanzee individuals but non-essential across six human individuals (Figure 3.3C).

The consistent depletion of many sgRNAs targeting the same gene and multiple genes involved

in the same biological process indicates that off-target activity is unlikely to explain proliferation differences between species. In principle, species-specific differences could also result from differential effectiveness of CRISPRi-mediated transcriptional repression (e.g., due to histone occupancy or transcriptional start-site variability). To evaluate this possibility, we measured the efficacy of sgRNA-mediated repression for several candidate genes (*CDK2*, *CCNE1*, and *RBL1*). In all cases, measurements of transcript abundance by qRT-PCR revealed >90% knockdown in both species (Figure 3.3D), suggesting that proliferation differences were not driven by incomplete knockdown efficiency in one species. In summary, these results highlight the ability of our screening approach to isolate biologically meaningful networks of genes that mediate species differences in cell behavior when perturbed, in contrast to gene expression profiling, which often reveals a complex mixture of both direct and indirect effects.

Human-specific sensitivity to perturbations to lysosomal V-ATPase

While both our primary screen and validation screens measured growth and survival, changes in proliferation can reflect a wide range of cellular phenotypes, from differentiation to growth factor signaling. We next investigated the human-specific sensitivity to loss of *ATP6AP1* and *ATP6AP2*. ATP6AP1 and ATP6AP2 are accessory proteins to the lysosomal V-type ATPase. As the main proton pump responsible for maintaining the pH gradient of the lysosome, non-duplicated core subunits of the V-ATPase were essential in both species, as expected (Figure 3.3B). Cryoelectron microscopy of the V-ATPase complex has implicated ATP6AP1 in the assembly of the V0 complex of the V-ATPase.[72] In addition, ATP6AP1 comprises a transmembrane helix and an extensive luminal domain that bears extensive structural homology to lysosomal-associated membrane proteins (LAMPs) and forms extensive contacts with ATP6AP2. Staining with LysoTracker red and LysoSensor green in *ATP6AP1*-depleted cells revealed no significant defects in maintenance of lysosomal pH (Figure S3.4E). These results are consistent with the core function of V-ATPase being strictly essential. However, loss of *ATP6AP1* has also been

implicated in major cellular signaling pathways that are mediated by the lysosome.[73–75] Thus, we performed a western blot to measure phosphorylation of ribosomal protein RPS6 (pS6), a well-established downstream metric of mTORC1 activity. Depletion of *ATP6AP1* or *ATP6AP2* resulted in diminished pS6 signal in both species. However, pS6 was selectively abolished in human cells depleted for *ATP6AP1* (Figures 3.3E and S3.4F). These data thus link the human-specific growth defect of *ATP6AP1* sgRNAs observed in our pooled screens to an increased reliance of human cells on ATP6AP1-mediated mTORC1 signaling.

Human PSCs are robust to depletion of CDK2 and cyclin E

We next investigated whether chimpanzee-specific sensitivity to loss of several cell-cycle genes could be linked to changes in cell-cycle progression, following repression of these genes (Figure 3.4A). To do so, we measured the proportion of cells in G1, S-phase and G2/M via incorporation of the thymidine analog 5-ethynyl-2′-deoxyuridine (EdU) and Hoechst, a DNA-binding dye. Consistent with the early mammalian embryo, PSCs undergo rapid cell-cycle progression with a shortened G1 phase, compared with somatic cells.[76] For wild-type cells, only ∼10% of cells were classified in G1 phase (Figure 3.4B). However, the absolute fraction of G1 cells was influenced by environmental factors such as confluence and nutrient availability (Figure S3.5A). Therefore, we measured the effect of CRISPRi-mediated gene repression in an internally controlled co-culture, with wild-type cells (GFP−) and sgRNA-expressing cells (GFP+) mixed within the same well. Knockdown of *CDK2* or *cyclin E1* in chimpanzee PSCs led to a roughly 2-fold accumulation of cells in G1 (Figure 3.4C; $p < 10-3$ for both, two-tailed t test), consistent with the well-established role of cyclin E1-CDK2 in regulating the G1/S transition. By contrast, knockdown of *CDK2* in human PSCs had no effect on cell-cycle progression, and knockdown of *cyclin E1* produced only a limited accumulation of G1 cells. We confirmed that these differences were not mediated by incomplete sgRNA-mediated knockdown in human PSCs (Figure 3.3D). In addition, differences in cell viability were minor, and no species differences in G1 proportions were observed in the

absence of cell-cycle perturbations (Figures S3.5B and S3.5C). Lastly, we confirmed that the sgCDK2-mediated increase in G1 length could be observed in chimpanzee cells via live imaging with a FUCCI reporter (Figure S3.5D and S3.5E).[77] Thus, our data suggest that human PSCs are less dependent on cyclin E1-CDK2 for G1/S-phase transition than chimpanzee PSCs.

Cyclin E1-CDK2 is a central regulator of the G1/S cell-cycle transition[78] and is commonly essential[35] and frequently dysregulated across human cancer cell lines.[79] In contrast, Cdk2−/− knockout mice are fully viable and develop normally, albeit with reduced body size.[80,81] Subsequent studies showed that cell-cycle progression could be rescued in the absence of cyclin E1-CDK2 by cyclin A-CDK1 and cyclin E-CDK1 activity.[82,83] Therefore, we reasoned that human cells might compensate for the loss of cyclin E1-CDK2 via stronger cyclin A2-CDK1 activity, as cyclin homologs cyclin E2 and cyclin A1 are not expressed in PSCs. Consistent with this model, *CDK1* was more highly expressed in human PSCs (FDR < 10−2), while *CDK2* and *cyclin E1* were more highly expressed in chimpanzee PSCs (FDR < 10−3) (Figures 3.4D and S3.5F). As a functional test, we overexpressed *CDK1* in chimpanzee PSCs in conjunction with sgRNA-mediated repression of *CDK2* or *cyclin E1* and quantified the progression of cells through G1 phase. We found that 2.5-fold overexpression of *CDK1* was sufficient for rescuing the sensitivity of chimpanzee PSCs to *CDK2* or *cyclin E1* depletion and accelerated progression through G1/S-phase transition (Figures 3.4E and 3.4F).

Next, we extended our co-culture studies to additional cell-cycle regulators with known interactions with cyclin E-CDK2. Given the dependence of chimpanzee PSCs on cyclin E-CDK2, we reasoned that repression of an inhibitor of this cyclin-CDK complex might have species-specific effects on cell-cycle progression. We first investigated the consequences of repressing retinoblastoma-like 1 (*RBL1*/p107), a tumor suppressor homologous to retinoblastoma protein (*RB*). RBL1, like RB, represses cell cycle via inhibition of E2F transcription factors.[84–86] However,

E2F is de-repressed in rapidly dividing stem cells, compared with other cell types, owing to the need for rapid cell cycling.[87] Indeed, repression of *RB* did not elicit a growth effect in either species (Figure S3.5G), consistent with E2F de-repression and prior studies of PSCs.[88,89] In contrast, RBL1 possesses an ability unique among RB family proteins to directly inhibit the kinase activities of cyclin A/E-CDK2.[90] Repression of *RBL1* resulted in faster growth and a reduction in the fraction of cells in G1 in both species (Figure 3.4C). However, *RBL1* effects were larger for chimpanzee cells ($p < 0.01$, two-tailed t test). Given the accumulation of chimpanzee PSCs in G1 upon repression of *cyclin E* or *CDK2* (Figure 3.4C), these results further support a model in which cyclin E-CDK2 exerts greater control over G1/S transition in chimpanzee, compared with human PSCs.

In addition, we examined the chimpanzee-specific sensitivity to depletion of *FAM122A*, a cell-cycle regulator that acts upstream of *CDK1/2* via interactions with *CHK1* and *PP2A* (Figure 3.4A). FAM122A acts as an inhibitor of phosphatase PP2A-B55α,[91] which in turn acts in opposition to CDK1 and CDK2 by dephosphorylating key substrates such as WEE1 and CDC25[92] (Figure 3.4A). We observed that loss of *FAM122A* phenocopied loss of *CDK2* and led to accumulation of G1 cells in chimpanzee but not human PSCs (Figure 3.4C; $p < 0.05$, two-tailed t test). This species difference appeared to be independent of WEE1 accumulation, as *FAM122A*-depleted chimpanzee cells exhibited increased sensitivity to the WEE1 inhibitor adavosertib (Figure 3.4G),[93] suggesting that the effects of *FAM122A* loss may depend on other PP2A targets such as CDC25A. Moreover, we observed that *FAM122A* depletion in PSCs of both species promoted resistance to the CHK1 inhibitor prexasertib (Figure 3.4G),[94] highlighting the conserved upstream interaction between FAM122A and CHK1.

Finally, we applied pharmacological approaches to examine species differences in enforcing cell-cycle checkpoint. Because rapid cell divisions render PSCs sensitive to replication stress and DNA damage,[95–98] we tested whether wild-type human and chimpanzee PSCs would respond

differently to CHK1 and WEE1 inhibition in the absence of genetic perturbations. To do so, we mixed GFP-tagged human cells with mCherry-tagged chimpanzee cells in a ~50/50 co-culture competition experiment. With no drug treatment, the proportion of human and chimpanzee cells remained unchanged. However, upon either prexasertib or adavosertib treatment, we observed substantially higher survival of chimpanzee cells compared with human cells (Figure 3.4H). Collectively, these genetic and pharmacological results suggest that chimpanzee PSCs may enforce a more robust S-phase and G2/M checkpoint, compared with human PSCs, with higher endogenous CDK1 levels providing a potential mechanism for human PSCs to overcome inactivation of *CDK2*, *FAM122A*, *CHK1*, or *WEE1* to promote cell-cycle re-entry.

Cell-cycle perturbations alter NPC expansion

We wondered whether the molecular differences that we observed between species in stem cells would also manifest in differentiated cell types. As differences in G1/S regulation have long been hypothesized as an evolutionary mechanism for changing brain size,[99–102] we investigated whether human-specific robustness to depletion of cell-cycle factors would persist in NPCs. Previous studies have established both the necessity and sufficiency of genes promoting G1/S transition for proliferative NPC divisions in animal model systems.[100–106] However, it is not known whether human NPCs possess recently evolved characteristics that imbue them with an enhanced ability to maintain proliferative divisions. We generated CRISPRi human and chimpanzee NPCs and assessed how depletion of *cyclin E1*, *CDK2*, *RBL1*, and *FAM122A* affected cell-cycle progression and self-renewal (Figures 3.5A–3.5D and S3.6A–S3.6C). In contrast to PSCs, NPCs undergo substantially slower progression through cell cycle, with ~50% of cells in G1 phase compared with ~10% in PSCs (Figure 3.5B). Nonetheless, knockdown of *CDK2* ($p < 0.05$, two-tailed t test) or *cyclin E1* ($p < 10-3$, two-tailed t test) caused an additional accumulation of chimpanzee, but not human, NPCs in G1 (Figure 3.5C). Meanwhile, *RBL1* knockdown reduced the fraction of G1 cells in both human and chimpanzee (Figure 3.5C).

Depletion of *FAM122A* resulted in G2/M accumulation in chimpanzee but not human NPCs (Figure 3.5D), implying a greater role for PP2A activation at G2 in chimpanzee NPCs compared with PSCs.

We further tested the species-specific responses to genetic perturbations in cerebral organoids. Quantifications of organoid size at Day 18 revealed that human organoid development was robust to depletion of *CDK2*, *CDK4*, or *CCNE1* but sensitive to depletion of *ATP6AP1* (Figures 3.6A, 3.6B, and S3.6D). However, chimpanzee organoids seeded with cells expressing sgRNAs targeting *CDK2*, *CDK4*, or *CCNE1* were substantially smaller. To examine whether these size differences arose from changes in growth at the earliest stages of patterning versus ongoing differences in neural progenitor proliferation, we measured G1-phase length in day 9 organoids. We confirmed that depletion of *CDK2* in chimpanzee organoids continued to impair proliferation and resulted in a larger fraction of cells in G1 phase, with no such effect observed in the corresponding human organoids (Figure 3.6C). These results suggest that the increased robustness of human NPCs to depletion of regulators of G1/S progression could potentially bias human cells toward proliferative divisions, as has been proposed by developmental models.

Evolutionary origin of molecular species differences

To determine the evolutionary origin of human- and chimpanzee-specific genetic dependencies, we extended our comparative studies to orangutan PSCs.[107] While humans and chimpanzees diverged roughly 7 million years ago,[108] orangutans diverged from other great apes 13–18 million years ago.[109] Thus, we could infer by maximum parsimony that any genetic dependencies shared between orangutans and chimpanzees, but not humans, were likely to have been present in the common ancestor and subsequently diverged in the human lineage. We performed three-way species comparisons across genes representing several biological processes with coherent species differences in our dataset, using sgRNAs with perfectly matched targets in all three

species. For two sgRNAs targeting *CDK2*, we observed a significant depletion of sgRNA-expressing cells over the course of 10 days in both chimpanzee and orangutan PSCs (Figures 3.7A and S3.7A). In contrast, no such depletion was observed in human PSCs. We further confirmed that the differences we observed were not due to differences in sgRNA activity, as knockdown efficiency exceeded 90% in all three species. In addition, we further observed human-specific robustness to repression of *CDK4* (Figure 3.7B) and *cyclin E1* (Figures S3.7B and S3.7C). Based on these data, we inferred that robustness to perturbations of the G1/S transition evolved along the human lineage, otherwise dependence on *CDK2*, *CDK4*, and *cyclin E1* would have had to evolve on two separate occasions in the chimpanzee and orangutan lineages with the same direction of effect for each gene. Next, we evaluated the human-specific sensitivity to repression of *ATP6AP1*. We observed that the *ATP6AP1* sensitivity was not shared by chimpanzee or orangutan PSCs, suggesting that altered responses to cellular metabolism, including the increased reliance on ATP6AP1 for mTORC1 signaling that we observed, also evolved along the human lineage (Figure 3.7C).

By contrast, repression of *KAT6A* promoted proliferation in chimpanzee PSCs but not in human or orangutan PSCs, arguing that this molecular feature was derived in chimpanzees (Figure 3.7D). Similarly, sensitivity to *UFL1* repression was common to human and orangutan PSCs but diverged in chimpanzee PSCs (Figure 3.7E). In sum, our data indicate that distinct genetic dependencies arose recently in both the human and chimpanzee lineages, highlighting the importance of experimentally defining the extent, identity, and phylogenetic origin of cellular and molecular differences derived in humans to inform our understanding of human evolution.

DISCUSSION

Loss-of-function screens have provided fundamental insights into the genes that regulate the development of model organisms. Here, we applied genetic screens to human and chimpanzee

PSCs to examine whether the requirements for essential genes could differ in closely related species. By performing paired genome-wide CRISPRi screens, we uncovered a landscape of divergent genetic dependencies. Despite human and chimpanzee PSCs being similar in their cellular morphology, response to *in vitro* differentiation protocols, and core set of essential genes, we identified 75 genes with divergent roles in controlling cellular proliferation. We observed that many of these genes were organized in coherent protein complexes and biochemical pathways. By contrast, existing state-of-the-art comparative approaches, including RNA-seq and chromatin state profiling,[20,22] have identified thousands of differentially expressed and accessible genes between humans and non-human primates but cannot directly evaluate the role of each gene in key cellular processes such as survival, proliferation, and differentiation. In addition, limited coherence among differentially expressed genes makes it difficult to pinpoint divergent pathways or protein complexes. Our data thus comprise a rich resource that interfaces with existing studies of gene regulation and chromatin states and provide a functional genomics guide for future candidate gene approaches.

How might the genetic dependencies we observed in PSCs relate to organismal differences that manifest during development? Intriguingly, one of the strongest observations that emerged from our unbiased genome-wide screening approach was human-specific robustness to depletion of cell-cycle factors, which persisted in NPCs. This finding aligns with long-standing hypotheses that changes in cell-cycle regulation could play a role in human-specific cortical expansion. The G1-phase length hypothesis proposes that factors lengthening G1 duration in NPCs increase the probability of differentiation toward non-proliferative neuronal fates, while factors reducing G1 length promote proliferative self-renewal of NPCs.[99,102,103] Indeed, loss of *CDK2* or *CDK4* in mouse NPCs prolongs G1 length and causes premature neuronal differentiation at the expense of self-renewal.[105] Conversely, exogenous overexpression of *CDK4* and *cyclin D1* in mouse and ferret reduces G1 length, promotes self-renewing divisions in basal progenitor cells, and results in

increased brain size and cortical area, while preserving a structurally normal, six-layered cortex.[106] In humans, mutations that promote cyclin D2 stability lead to megalencephaly.[110] These studies underscore the influence of inputs to the G1/S transition on brain expansion during development.[101] However, whether this developmental mechanism changed specifically in recent human evolution remained unexplored. Our demonstration that human NPCs are more likely than chimpanzee NPCs to continue cycling upon equivalent repression of *CDK2* or *cyclin E1* connects proposed developmental mechanisms to molecular changes that occurred in human evolution. Although physiological stressors do occur during development that influence the size of the neural progenitor pool,[21,111–113] it remains unknown how external environmental stimuli or the intrinsic tempo of differentiation might differ between humans and chimpanzees and interface with the changes in G1 regulation that we observed. Human-specific genetic dependencies could result from evolutionary changes in cell behavior or developmental systems drift, a process in which cell behaviors are conserved, but the underlying circuitry changes.[114,115] Regardless of the impact on cell behavior, the altered genetic dependencies or drug sensitivities represent recently evolved substrates for disease vulnerabilities and further evolutionary changes.[116] We expect that future studies connecting the response of human and chimpanzee NPCs to a wider range of genetic and physiological perturbations will provide further insights into the evolutionary mechanisms by which the proliferative capacity of NPCs has increased along the human lineage.

The endeavor to study the molecular basis of human evolution has been compared with searching for needles in a haystack, as human-specific genetic variants and gene expression changes are numerous and predominantly neutral.[2] By contrast, our finding that human and chimpanzee PSCs exhibit distinct genetic dependencies, even for genes that lack clear expression or protein-coding sequence divergence, provides a complementary approach for isolating recently evolved functional changes in human gene networks. This strategy mirrors the progression of cancer genetics research from sequencing and transcriptomics efforts such as TCGA[117] to functional

genetics-based efforts such as DepMap.[35,36] Moreover, while driver mutations can be identified in tumors based on their independent recurrence, human evolution has occurred only once, highlighting the added value of a functional genomics platform. We expect that loss-of-function profiling can be extended to cellular models of genetic variation and disease risk within humans to identify shared vulnerabilities and convergent pathway level differences. Lastly, our approach can be readily applied in differentiated cell types and interfaced with higher dimensional measurements of cell phenotypes,[118–120] opening the door to future efforts for understanding molecular control of species differences across stages of development.

Limitations of the study

While our study systematically maps species-specific genetic dependencies in PSCs, it is unclear which dependencies will persist across diverse differentiation trajectories and whether new cell-type-specific dependencies will arise. We applied a principled and systematic exploration of the changes in the genetic landscape that evolved during the short divergence time between humans and chimpanzees. However, to further connect human-specific robustness to depletion of cell-cycle regulators to physiological differences in brain development, we will need to understand how naturally occurring environmental factors can produce species differences in NPC cell-cycle properties and proliferative potential. In addition, the *in vitro* models of NPCs that we employed produce homologous cell types to those found *in vivo* but may exhibit differences in metabolism, spatial architecture, and maturation speed. Finally, our primary screen unexpectedly included two p53-non-responsive chimpanzee cell lines, which we accounted for by constructing four additional p53-responsive cell lines in the validation screen. The frequency of pro-survival adaptations in PSCs highlights the need for multiple representatives of each species and quality controls (such as the nutlin-3a assay) to distinguish species differences from line-to-line variation.

METHODS

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Alex A. Pollen (Alex.Pollen@ucsf.edu).

Materials availability

Materials used in this study will be provided upon request and available upon publication.

Experimental model and study participant details

Cell Lines

Human embryonic cell line H1 (WiCell), human induced pluripotent stem cell lines (20961B, 21792A, 23555A, 28126B, WTC11), chimpanzee induced pluripotent stem cell lines (3624K, C3649, 40280L, 40290F, 8861G, Pt5-C), Sumatran orangutan induced pluripotent stem cell line (Josephine, San Diego Zoo 11045-4593).

Media Formulations

mTESR1 was purchased from Stem Cell Technologies (cat. 85850) and supplemented with 100 units/ml penicillin, 100 µg/ml streptomycin, and 292 µg/ml L-glutamine (Gibco, cat. 10378016). StemFlex was purchased from Gibco (cat. A3349401) and supplemented with 100 units/ml penicillin, 100 µg/ml streptomycin, and 292 µg/ml L-glutamine. HEK293Ts were cultured in DMEM (ThermoFisher, cat. 11965118) and supplemented with 10% FBS (VWR, cat. 97068-085, lot 043K20), 100 units/ml penicillin, 100 µg/ml streptomycin, and 292 µg/ml L-glutamine. Neuronal differentiation media was prepared as described in,[121] with DMEM F/12 (ThermoFisher, cat. 21331020), CTS Neurobasal Medium (ThermoFisher, cat. A1371201), 1x N-2 supplement CTS

(ThermoFisher, cat. A1370701), 10 μM SB431542 (StemMACS, TGFβ inhibitor; Miltenyi, cat. 130-106-543), 100 ng/ml Noggin (recombinant human; Miltenyi, cat. 130-103-456), and 100 units/ml penicillin, 100 μg/ml streptomycin, and 292 μg/ml L-glutamine. Thiazovivin (Stem Cell Technologies, cat. 72252) was included at a concentration of 2 μM during passaging. The CEPT cocktail[122] consisting of 50 nM chroman 1 (Tocris Bioscience, cat. 7163), 5 μM emricasan (Selleck Chemicals, cat. S7775), polyamine supplement diluted 1:1000 (Sigma-Aldrich, cat. P8483), and 0.7 μM trans-ISRIB (Tocris Bioscience, cat. 5284) was included during single cell sorting and lipofection.

Cerebral organoids were seeded in Sasai media #1 (GMEM supplemented with 20% KSR, 1x Penicillin/Streptomycin/Glutamate, 0.11 mg/mL sodium pyruvate, 0.1 mM β-mercaptoethanol, 10 μM thiazovivin, 5 μM SB431542, 200 nM LDN-193189, and 1 μM Wnt-C59).[123] Wnt inhibitor Wnt-C59 was withdrawn after day 6.

Construction of CRISPRi cell lines

All wildtype cell lines tested negative for mycoplasma prior to the start of cell line engineering. The CRISPRi effector protein dCas9-KRAB (*KOX1*) was introduced into either the *CLYBL* or *AAVS1* safe harbor locus.[38,124] For the AAVS1 locus, cell lines were constructed via lipofection of three plasmids: 1) A modified version of pX458 (Addgene #48138), containing both Cas9 nuclease and a sgRNA targeting AAVS1. The sgRNA spacer was modified from the original plasmid by cutting with type IIS restriction endonuclease BbsI-HF. Complementary oligos containing the sgRNA spacer and proper overhangs were annealed and ligated with T4 ligase. 2) A modified version of the previously published Gen3-AAVS1 vector[125] containing an optimized dCas9-XTEN-KRAB-P2A-EGFP or mCherry[126] driven by the chicken beta-actin (CAG) promoter, flanked by homology arms to AAVS1. 3) pEF1-BCL-XL, a plasmid expressing BCL-XL, the anti-apoptotic isoform of BCL2L1, from the EF-1α promoter. For the *CLYBL* locus, four plasmids were

lipofected based on previously published methods[38]: 1) pZT-C13-L1 (Addgene #62196) 2) pZT-C13-R1 (Addgene # 62197), with plasmids 1 and 2 encoding TALENs that target the *CLYBL* locus. 3) dCas9-XTEN-KRAB-P2A-BFP (Addgene #127968) and 4) pEF-BCL-X. Lipofection was performed as follows: two days prior to transfection, cells were switched into mTESR1 media, which does not contain heparin, on a non-passaging day. One day prior to transfection, ~400,000 cells were plated into a Matrigel-coated (Corning, cat. 354230) 6-well plate with mTESR1 supplemented with 2 μM thiazovivin. On the day of transfection, a 3 μg mixture of plasmids 1-3 was made at a mass ratio of 5:5:1, added to a mixture of 96 μl Opti-MEM (Gibco, cat. 31985062) and 4 μl Lipofectamine Stem (ThermoFisher, cat. STEM00003), and incubated at room temperature for 10 minutes. Media was aspirated from the PSC plate and replaced with 2 ml Opti-MEM supplemented with CEPT. The lipid/plasmid DNA complexes were then added to plate and incubated for 4 hours, after which 2 ml of mTESR1 supplemented with CEPT was overlaid. 24 hours post-transfection, media was replaced with StemFlex supplemented with CEPT. 48 hours post-transfection, media was replaced with StemFlex supplemented with 2 μM thiazovivin, and cells were then passaged for 10-14 days to dilute out the transfected plasmids. Single-cell clones and one polyclonal population per cell line were then sorted on a Sony MA900. Expanded populations were cryopreserved in Bambanker preservation media (ThermoFisher, cat. 50999554). Six CRISPRi engineered PSC lines (3624K, 8861G, 21792A, 40280L, 40290F and Pt5-C) were functionally validated with an sgRNA targeting B2M, a gene encoding a non-essential cell surface protein. B2M levels were measured by staining with an APC anti-human β2-microglobulin antibody (BioLegend, cat. 316312).

Nutlin-3a pharmacological assay for testing TP53/MDM2 responsiveness

CRISPRi engineered cell lines were tested for TP53/MDM2 responsiveness based on sensitivity to Nutlin-3a, an active enantiomer of Nutlin-3, which is a small molecule MDM2 inhibitor that induces p53-dependent autophagy and apoptosis.[54,55] Cells were passaged with Versene (PBS-

EDTA) to avoid use of ROCK inhibitor, which promotes survival. 80-100% confluent wells were split 1 to 1 and plated densely into new wells with StemFlex supplemented with 10 μM Nutlin-3a. Consistent with previous reports, TP53/MDM2 responsive cell lines underwent apoptosis within 24 hours.

Lentivirus production, concentration, and titration

Lentivirus for CRISPRi screening was produced in HEK293T cells. HEK293Ts were seeded at a density of 80,000 cells/cm2 24 hr. prior to transfection in 15 cm dishes. Next, each dish was transduced with 20 μg sgRNA library, 6.75 μg standard lentivirus packaging vectors, and 81 μl Mirus transfection reagent (VWR, cat. 10767-122) in Opti-MEM. 24 hr. post-transfection, media was replaced and supplemented with 1X ViralBoost (Alstem, cat. VB100). Supernatant was collected at 48 hr. post-transfection and concentrated 1:10 with Lenti-X Concentrator (Takara Bio, cat. 631231). Concentrated lentivirus was titered in PSCs based on BFP expression 3 days post-infection using a flow cytometer.

Pooled genome-wide CRISPRi screening

CRISPRi PSCs expressing dCas9-KRAB were dissociated with Accutase (Innovative Cell Technologies, cat. AT104-500), resuspended in StemFlex supplemented with 2 μM thiazovivin and 5 μg/ml polybrene (Sigma-Aldrich, cat. TR-1003-G), transduced with the lentiviral hCRISPRi-v2 sgRNA library at a target infection rate of 25-40%, and plated in Matrigel-coated 5-layer cell culture flasks (Corning, cat. 353144) at a density of 65,000-80,000 cells/cm2. The following day, StemFlex medium was replaced. Two days after infection, cells were dissociated with Accutase, resuspended in StemFlex supplemented with 2 μM thiazovivin and 1.5 μg/ml puromycin (Goldbio, cat. P-600-100), and plated in 5-layer cell culture flasks. The following day, medium was replaced with StemFlex supplemented with 1.5 μg/ml puromycin. Four days after infection, 100 M cells were harvested for the initial time point (t0), while 250-300 M cells were resuspended in StemFlex

supplemented with 1.5 µg/ml puromycin and plated in 5-layer cell culture flasks (>1000x sgRNA library representation). Selection efficiency was assessed by flow cytometry (>70% BFP+). Every two days, cells were dissociated with Accutase, resuspended in StemFlex supplemented with 2 µM thiazovivin, and plated at a density of 80,000-100,000 cells/cm2. Technical replicates were cultured separately for the duration of the screen. After 10 days of growth, 150 M cells from each technical replicate were harvested for the final time point (tfinal). Genomic DNA was isolated from frozen cell pellets using the Macherey-Nagel NucleoSpin Blood XL kit (Macherey-Nagel). Isolated DNA was quantified using a NanoDrop (ThermoFisher) and the sgRNA expression cassette was amplified by 22 cycles of PCR using NEBNext Ultra II Q5 Master Mix (NEB) and primers containing Illumina P5/P7 termini and sample-specific TruSeq indices. Each sample was distributed into 150-200 individual 100 µl reactions in 96-well plates, each with 10 µg genomic DNA as input. Following amplification, reactions from each sample were pooled and a 100 µl aliquot was purified using AMPure XP beads (Beckman-Coulter) with a two-sided size selection. Purified libraries were quantified by Qubit (ThermoFisher) and sequenced on an Illumina HiSeq 4000 instrument (SE50, 5% PhiX) with a custom sequencing primer (oCRISPRi_seq V5).

CEV-v1 validation screening library design

To validate genome-wide screens, a Comparative Essential Validation (CEV-v1) sgRNA library consisting of 9,692 sgRNAs targeting 963 candidate species-specific essential or proliferation suppressor genes was constructed. hCRISPRi-v2 sgRNAs with perfect-match targets in panTro6 exhibiting significant depletion or enrichment in the genome-wide screens were retained in CEV-v1 (*n* = 3589 sgRNAs). In addition, new sgRNAs with perfect-match target sites in the human (hg38) and chimpanzee (panTro6) reference genomes were chosen based on their position relative to the FANTOM-annotated transcriptional start site,[44] on-target activity predicted by DeepHF,[127] and off-target potential predicted by a genome-wide search of mismatched target sites[58,128] in both reference genomes. Briefly, after performing off-target filtering (one perfect-

match target, CRISPRi specificity score > 0.20, maximal predicted activity at any off-target site < 0.80), candidate sgRNAs were categorized by their position relative to the FANTOM TSS and then ranked by their DeepHF score. A threshold DeepHF score was imposed by excluding sgRNAs with predicted activities less than one standard deviation below the mean of all candidate sgRNAs (minimum score: 0.4378). Eight sgRNAs were selected for each gene as well as 1845 non-targeting sgRNAs from hCRISPRi-v2. Oligonucleotide pools were designed with flanking PCR and restriction sites (BstXI, BlpI), synthesized by Agilent Technologies, and cloned into the sgRNA expression vector pCRISPRia-v2 (Addgene #84832) as described previously.[32]

Pooled validation CRISPRi screening

Validation screens were performed in conditions consistent with the genome-wide screens. Briefly, CRISPRi PSCs expressing dCas9-KRAB were dissociated with Accutase, resuspended in StemFlex supplemented with 2 μM thiazovivin and 5 μg/ml polybrene, transduced with the lentiviral CEV-v1 sgRNA library at a target infection rate of 25-40%, and plated in Matrigel-coated 3-layer cell culture flasks (Corning, cat. 353143) at a density of 65,000-80,000 cells/cm2. Cells were dissociated, plated, selected with puromycin, and grown on the same schedule as used for the genome-wide screens. Technical replicates were cultured separately for the duration of the screen and >1000x sgRNA library representation was maintained.

In addition to assessing the reproducibility of species-specific genetic dependencies across biological replicates (n = 5 human cell lines and n = 6 chimpanzee cell lines), we also assessed reproducibility between screens and across sites. 3 out of 4 PSC lines from the initial screen were re-tested with the CEV-v1 validation library. Several individual cell lines were screened twice (H20961B, H23555A, H28126B, C3624K, C8861G), with replicate screens were performed independently at the Whitehead Institute and UCSF. 4 out of the 5 lines retested at both UCSF and Whitehead were independently constructed CRISPRi cell lines from the same source iPSC

line, with CRISPRi machinery inserted at either the *AAVS1* or *CLYBL* locus with either GFP, mCherry, or BFP as fluorescent markers.

Quantitative RT-PCR

Triplicates of human (28126B), chimpanzee (40280L), and orangutan (11045-4593) PSCs were grown in a 6-well plate and infected with sgRNAs at an MOI of ~0.3. 48 hours post-infection, cells were expanded in 2 μg/ml puromycin and allowed to recover for 48 hours. At post-infection day 4, sgRNA-expressing cells were sorted based on BFP+ expression using a Sony MA900. sgRNA-expressing cells were isolated by FACS as the depletion of cells containing essential sgRNAs occurred more rapidly than the removal of sgRNA-negative cells through puromycin selection.

Cells were then allowed to recover for 48-96 hours, until they reached ~60-80% confluence on a 6-well plate and were harvested 6 to 8 days post-infection. For each biological replicate, RNA was extracted with a Direct-zol RNA miniprep kit (Zymo Research, cat. R2051). RNA was reverse transcribed with SuperScript IV VILO (ThermoFisher, cat. 11756050), and cDNA was amplified with the DyNAmo ColorFlash SYBR Green kit (ThermoFisher, cat. F416L). Primers for GAPDH were used as loading controls and no-RT controls were performed to control for genomic DNA contamination. Amplifications were performed in duplicate and quantified on a QuantStudio Flex 7 Real-Time PCR system in 96-well plates. Final data points reported are averages of the two duplicate qRT-PCR amplifications.

RNA-seq library prep

Wild-type human and chimpanzee cells were grown to 70% confluence and RNA was extracted by adding RNAse-free Trizol (ThermoFisher, cat. 15596026) to each pellet and processing with the Zymo Research Direct-zol RNA miniprep kit (Zymo Research, cat. R2050). RNA-seq was performed using the Illumina TruSeq Stranded Total RNA kit (Illumina, cat. 20020599) according

to the manufacturer's instructions, with the exception of the final PCR step for which only 10 cycles were used to prevent overamplification. The final pooled library was sequenced on an Illumina HiSeq 2500 (SE50).

Western Blot

Human (28126B) and chimpanzee (40280L) PSCs were transduced with lentiviral constructs containing sgRNAs targeting *ATP6AP1* or *ATP6AP2*. One additional human (21792A) and chimpanzee (8861G) was also separately infected with an sgRNA targeting *ATP6AP1*. 48 hours post-infection, sgRNA-expressing cells were selected using 1.5 µg/ml puromycin. Cells were then recovered in normal growth media from 4 to 6 days post-infection. Cells were harvested at 6 days post-infection, along with separate wells of three wild-type human (H1, 21792A, 28126B) and three wild-type chimpanzee (3624K, 40280L, 8861G) cell lines. Day 6 was selected as a time point for harvesting because it is roughly the earliest time point at which a pure population of sgRNA expressing cells can be produced for western blots. In addition, growth for additional days can lead to larger differences in cell viability between species. Cells from each well of a 6-well plate were lysed in ~250 µl ice-cold RIPA buffer supplemented with protease inhibitor (ThermoFisher, cat. A32965). After 30 minutes of incubation in lysis buffer at 4°C, cells were centrifuged at 16,000 x g for 5 minutes at 4°C. Supernatant was collected and snap frozen in liquid nitrogen and stored at -80°C.

Protein concentrations in each lysate were quantified using a Bradford BCA kit (ThermoFisher, cat. PI23227) Lysate was normalized to 1 µg/µl in RIPA buffer. 30 µl of lysate was added to 10 µl of NuPage Sample Buffer (4x), heated to 70°C on a PCR thermocycler, and loaded onto a Bolt 4-12% polyacrylamide gel (ThermoFisher, NW04122BOX). The gel was run for 45 minutes at 165 V in MOPS buffer. Protein was then transferred onto a nitrocellulose membrane (BioRad, cat. 1704270) with a Bio-Rad Trans-Blot Turbo (BioRad, cat. 1704150). The membrane was blocked

with Intercept (PBS) Blocking Buffer (LI-COR, cat. 927-90003) for 1 hour at RT. Membrane was incubated overnight at 4°C with anti-pS6 primary antibody at a 1:1,000 dilution. Membrane was washed 3x with TBST and incubated with secondary antibody at 1:15,000 dilution. Membrane was washed 3x with TBST and imaged on a LI-COR Odyssey CLX. Afterwards, antibodies were stripped from the membrane with NewBlot IR Stripping Buffer (LICOR, cat. 928-40028). Membrane was reblotted with anti-GAPDH antibody at 1:1,000 dilution and incubated overnight at 4°C. Membrane was washed 3x with TBST and incubated with secondary antibody at 1:15,000 dilution. Membrane was washed 3x with TBST and imaged on a LI-COR Odyssey CLX.

Cell death and viability analysis

Cell viability and cell death assays were performed on chimpanzee PSC line 40280L. Cells were infected with sgRNAs targeting cell cycle genes *CDK2*, *CDK4*, *CCNE1*, *RBL1*, or a non-targeting control sgRNA. Cells were infected in triplicate with sgRNAs at 0.3 MOI and selected with addition of 2 μg/ml puromycin from 2 to 4 days post-infection. Cells were recovered from day 4 to day 6 post-infection in StemFlex supplemented with 2 μM thiazovivin to match conditions from the growth screens. Cell viability was measured on a Countess 3 automated cell counter and a Chemometec Nucleocounter NC-202.

Cell cycle EdU staining

Human (28126B) and chimpanzee (40280L) PSCs were infected with pCRISPRia-v2 (Addgene Cat# 84832) containing: 1) a targeting sgRNA or 2) a non-targeting sgRNA or 3) a combination of an sgRNA and a *CDK1* overexpression ORF expressed from the EF-1α promoter typically used to express GFP and the puromycin resistance cassette. The *CDK1* ORF was expressed downstream of sfGFP-P2A and replaced the puromycin resistance cassette. The BFP marker from the original pCRISPRia-v2 construct was replaced with sfGFP to enable use of the BFP channel for Hoechst staining during the cell cycle assay. On the day of infection, cells were

transduced with lentivirus and 8 μg/ml polybrene. Lentivirus was titered to MOI ~ 1, such that GFP+ sgRNA expressing cells represented between 25% and 40% of the population at 48 hours post-infection. No puromycin selection was performed to maintain a mixed co-culture of sgRNA expressing cells and uninfected cells. Cells were grown in StemFlex media supplemented with 2 μM thiazovivin on the day of infection and withdrawn via replacement of fresh media 24 hours post-infection. Cells were then assayed at Day 6, after being freshly split at Day 5 into ROCKi containing media to prevent over-confluence and nutrient depleted media. This time point was chosen because it is one of the earliest time points at which cells have recovered from lentivirus infection and after sgRNA-mediated gene repression has been fully activated. In addition, growth for additional days could lead to larger differences in cell viability or growth rate between species.

Cell cycle phase measurements were performed with the Click-iT Plus EdU Alexa Fluor 647 Flow Cytometry Assay Kit (ThermoFisher, cat. C10635). 10 μM EdU was added directly to each well without a fresh media change and cells were incubated for roughly 1 hour, after which cells were harvested with Accutase. The cell pellet was washed once with 500 μl PBS supplemented with 1% BSA, pelleted again, and fixed with 4% paraformaldehyde for 15 minutes at RT, protected from light. Cells were washed and permeabilized for 15 minutes at RT. EdU detection was then accomplished via click chemistry of an Alexa Fluor 647 to an EdU antibody, which was incubated with cells for 30 minutes at RT. After 1 wash, 10 μg/ml of Hoechst 33342 (ThermoFisher, cat. H3570) was added and incubated for 15 minutes. Cells were then directly analyzed by flow cytometry. Data were analyzed with custom MATLAB scripts – after filtering for viable cells and doublets, G1, S, and G2/M gates were manually drawn and saved with the function impoly() for each sample. sgRNA+ populations were determined by GFP+ expression, and identical G1, S, and G2/M gates were used for sgRNA+ cells and sgRNA- cells within each sample.

Cell cycle drug treatments

For Figure 3.4G, human iPSC line H28126B was used and chimpanzee iPSC line C40280L was used. Cells were infected with an sgRNA targeting *FAM122A*, with 15-30% of cells infected. No puromycin selection was performed. At day 4 post-infection, cells were Accutase passaged into StemFlex supplemented with 2 μM thiazovivin and drug, with prexasertib (Chk1i) or adavosertib (Wee1i) added at 62 nM. At day 6, cells were replated and fresh drug was added to ensure removal of dead cells. At day 8, the fraction of sgRNA+ (BFP+) surviving cells were analyzed by flow cytometry.

For Figure 3.4H, human iPSC line H21792A and chimpanzee iPSC C40280L were co-cultured with a 50/50 initial seeding density. After one normal passage, cells were dissociated with Accutase and passaged in StemFlex supplemented with 2 μM thiazovivin and drug, with prexasertib or adavosertib added at 125 nM. 24hr after drug treatment, cells were replated and fresh drug was added to ensure removal of dead cells. 48 hours post drug treatment, the ratio of human cells (GFP+) chimpanzee cells (mCherry+) was analyzed by flow cytometry.

FUCCI reporter live imaging

An updated version of the FUCCI cell cycle reporter was cloned into a lentiviral expression vector with a universal chromatin opening element (UCOE), an EF-1α promoter driving Cdt1 (1-100) C-terminal fused to mCherry, and a woodchuck hepatitis virus post-transcriptional regulatory element (WPRE).[77] A four amino acid SGGS linker was used between mCherry and the Cdt1 fragment. This construct constitutes half of the FUCCI cell cycle reporter and has been reported to fluoresce red during G1 phase but is otherwise degraded by both SCFSkp2 and CUL4Ddb1 E3 ubiquitin ligases.

Human (21792A) and chimpanzee (3624K) cells were infected with the Cdt1 reporter. At day 5

post-infection, polyclonal populations of cells expressing the Cdt1 reporter were sorted based on mCherry expression on a Sony MA900. Reporter cells were then cross validated with the commercial CLICK-iT EdU cell cycle assay, with mCherry positive cells nearly all classified as G1 phase by the EdU assay (88%). Similarly, mCherry negative cells were nearly fully depleted from G1 phase (3%).

Next, Cdt1 reporter cells were infected with either an sgRNA targeting *CDK2* or a non-targeting sgRNA control, with 25-40% of cells expressing an sgRNA based on GFP expression. At day 4 post-infection, cells were seeded onto an Ibidi 4-well µ-Slide, with one well for each condition. On day 5 post-infection, cells were live imaged on a Nikon Ti automated inverted microscope with incubation enclosure for a period of 24.83 hours at an interval of 10 minutes across 38 stage positions.

Neural progenitor cell differentiation

Human (28126B) and chimpanzee (40280L) PSCs were differentiated into neural progenitor cells (NPCs) as described in[121] with the following modifications. Differentiation media was made without ventral and caudalization patterning factors (sonic hedgehog agonist and GSK3i CHIR99021). PSCs were maintained on Matrigel (Corning, cat. 354230) prior to day 0 plating onto Lam-111 coated plates. Cells were seeded at 20,000 cells/cm^2, as measured by Chemometec Nucleocounter NC-202, roughly twice the published density to ensure robust survival. NPCs were evaluated for purity at days 7-11 of differentiation with antibody staining against NPC markers Pax6 and Nestin. To do so, cells were dissociated with Accutase, pelleted and washed, then fixed and permeabilized with the BD Cytofix/Cytoperm kit (ThermoFisher, cat. BDB554714). 100 µl cells were stained with 2 µl human anti-Pax6-APC (Miltenyi, cat. 130-123-328) + 5 µl mouse anti-Nestin-PE (Biolegend, cat. 656805) and evaluated by flow cytometry. In addition, NPCs were plated on µ-Slide 4 Well chambers (Ibidi cat. 80426), stained with antibodies against Pax6 and

Nestin, and visualized by fluorescence microscopy on a RPI spinning disk confocal microscope.

Cerebral organoid differentiation

Human (H1) and Chimpanzee (8861G) cells were reverse transduced with lentiviral constructs containing sgRNAs targeting *CDK2*, *CDK4*, *CCNE1*, *RBL1*, *ATP6AP1*, *ASPM*, or a non-targeting sgRNA. An additional pair of human (23555A) and chimpanzee (40280L) organoids were transduced with lentiviral constructs containing sgRNAs targeting cell cycle factors *CDK2*, *CDK4*, *CCNE1*, *RBL1*, or a non-targeting sgRNA. 48 hours post-infection, cells expressing sgRNA were selected for via addition of 1.5 µg/ml puromycin. Cells were then recovered in normal growth media from 4 days post-infection to 6 days post-infection. At 6 days post-infection, each sgRNA condition was single cell dissociated with Accutase, counted on a Chemometec Nucleocounter NC-202, and seeded into an ultra low-attachment V-bottom 96 well plate (Prime Surface cat. MS-9096VZ). Four to twelve wells were seeded per condition with 10,000 cells seeded per well in 100 µl of Sasai Media #1.

Organoid formation was performed according to the protocol established by the Sasai lab[123] with the following modifications. On days 0-6, wnt inhibitor Wnt-C59 was added to promote patterning to telencephalon. 200 nM LDN-193189 was added to promote differentiation from days 0 through 18.

Cerebral organoid EdU staining

Human (H1, 20961B, 23555A) and Chimpanzee (8861G and 40280L) cells were reverse transduced with lentiviral constructs containing sgRNAs targeting *CDK2* or a non-targeting sgRNA. 48 hours post-infection, cells expressing sgRNA were selected for via addition of 1.5 µg/ml puromycin. Only one day of puromycin selection was performed to maintain a sub-population (20-30%) of non-sgRNA expressing cells to enable co-culture analysis. Cells were

then recovered in normal growth media from 3 days post-infection to 6 days post-infection. At 6 days post-infection, each sgRNA condition was dissociated to single cells with Accutase, counted on a Chemometec Nucleocounter NC-202, and seeded into an ultra low-attachment V-bottom 96 well plate (Prime Surface cat. MS-9096VZ). Twelve wells were seeded per replicate in each condition with 10,000 cells seeded in 100 µl of Sasai Media #1.

Organoids were collected on day 9 for cell cycle measurements to examine whether perturbations continue to influence cell cycle progression during NPC patterning and expansion. This time point was also chosen because several lines (20961B, 23555A, 8861G, and 40280L) exhibited significant silencing of the EF-1α promoter at day 18, which drives GFP expression within the sgRNA cassette. Some silencing was also apparent at day 9, with a bias for non-silenced cells to be actively proliferating, resulting in an apparent reduction of cells in G1 phase for the non-targeting sgRNA. Organoids were incubated in 10 µM EdU with a fresh media change and cells were incubated for roughly 1 hour, after which organoids were dissociated with a pre-warmed solution of Papain (Worthington Biochemical Corporation) mixed with DNase I according to manufacturer's instructions. For single-cell dissociation, organoid samples were cut into small pieces, and incubated with Papain for 30 minutes. Cells were then pelleted and then directly fixed in 4% paraformaldehyde for 15 minutes at room temperature, protected from light. Cell cycle phase measurements were then performed as described above with the Click-iT Plus EdU Alexa Fluor 647 Flow Cytometry Assay Kit (ThermoFisher, cat. C10635).

Orangutan CRISPRi growth comparison

CRISPRi machinery was engineered into orangutan PSCs[107] at the AAVS1 locus via the three plasmid lipofection method described above (see construction of CRISPRi cell lines). To account for mutations in the orangutan genome, the sgRNA for the Cas9 nuclease component was modified to perfectly match the orangutan AAVS1 locus. However, flanking regions were not

modified. sgRNAs targeting *CDK2*, *CDK4*, *CCNE1*, *ATP6AP1*, *KAT6A*, and *UFL1* were transduced into human (28126B), chimpanzee (40280L), and orangutan PSCs via lentivirus at MOI ~1. Cells were transduced in triplicate in 24-well plates and passaged every 2 days in StemFlex supplemented with 2 μM thiazovivin. At each passage, a portion of cells were quantified by flow cytometry on a BD LSRFortessa. The fraction of sgRNA+ expressing cells was determined based on the fraction of BFP+ cells. Measurements were collected until day 10 for *CDK2*, *CDK4*, *CCNE1*, and *ATP6AP1*. Measurements were collected until day 14 for *KAT6A* and *UFL1*. We chose to carry out 4 additional days of growth for *KAT6A* and *UFL1* because the growth differences between species were smaller for these genes compared to *CDK2*, *CDK4*, and *ATP6AP1*, and the additional depletion of sgRNA-expressing cells that occurred between days 10 to 14 enabled finer resolution of species differences. In parallel, cells were transduced with sgRNAs targeting *CDK2*, *CDK4*, *CCNE1*, *ATP6AP1*, *KAT6A*, and *UFL1* and expanded into a 6-well plate. 48h post-infection, cells were expanded in 2 μg/ml puromycin and allowed to recover for 48 hours. At day 5 post-infection, sgRNA+ cells were sorted based on BFP+ expression using a Sony MA900. RNA was extracted, cDNA was reverse transcribed, and qRT-PCR was used to quantify the degree of sgRNA-mediated depletion in biological triplicate (as described above in Quantitative RT-PCR).

Data analysis for pooled genome-wide CRISPRi screens

Sequencing data were aligned to hCRISPRi-v2 and quantified using the ScreenProcessing pipeline (https://github.com/mhorlbeck/ScreenProcessing).[44] sgRNA counts were then processed using MAGeCK[47] test (--norm-method control --remove-zero control --gene-test-fdr-threshold 0.10 --remove-zero-threshold 50 --gene-lfc-method alphamean) and separately using a custom analysis method inspired by MAGeCK. Briefly, sgRNA counts were normalized by the median ratio method. Mean-variance modeling was performed with non-targeting sgRNAs as the control group, and sample mean and variance values were used to parameterize a negative binomial

distribution (Figure S3.1B). *P*-values were then calculated for each sgRNA based on the tail probability of the negative binomial and *P*-value cut-offs were chosen such that 95% of non-targeting sgRNAs were not significant.

For each gene, sgRNAs were filtered by two criteria: 1) perfect alignment to both the human (hg38) and chimpanzee (panTro6) reference genomes, as determined by FlashFry,[58] and 2) significance according to the negative binomial distribution. Only sgRNAs passing both filters were retained for analysis, resulting in a variable number of sgRNAs per gene (0-5 sgRNAs). The remaining sgRNA counts were converted to log2 fold-change and averaged to produce a gene score. Significance testing for gene scores was performed by bootstrapping non-targeting sgRNAs, with groups of 1-5 random non-targeting sgRNAs assigned to each control gene to select candidates from the initial genome-wide screens.

Essential genes for each screen (Figure 3.1B) were determined by the degree of depletion among sgRNAs targeting the gene. To facilitate equal comparison among screens, the top 3000 most depleted genes in each screen were defined as essential (with mean sgRNA depletion greater than 4-fold for all such genes). In addition, genes with mean sgRNA depletion less than 2-fold were defined as non-essential. Each intersection set was then constructed based on two inclusion criteria: genes were required to be essential for each member of the set and non-essential for all non-members of the set.

For comparison of technical replicates (Figure 3.1B), identification of shared essential genes (Figure 3.1B), and assessment of screen performance using DepMap Public 21Q4 gene sets (Figure 3.1C), all hCRISPRi-v2 sgRNAs, including those with mismatched targets in the panTro6 reference genome, were included for analysis. For identification of candidate genes with species-specific effects on proliferation (Figure 3.1D), only sgRNAs with perfect-match targets in the

panTro6 reference genome (77.4%, *n* = 79417/102640) and transcriptional start sites targeted by at least three sgRNAs after excluding mismatched sgRNAs (86.7%, *n* = 17804/20528), were retained for analysis.

For screen analysis using MAGeCK, genes with false discovery rates (FDRs) less than 10% for both individuals within a species and FDRs greater than 25% in both individuals from the opposite species were considered as candidates for validation screening.

Copy number variation analysis

Chromosomal copy number variations (CNV) were inferred with the InferCNV R package (version 1.2.1), which predicts CNVs based on gene expression data. InferCNV was run in 'subclusters' analysis mode using 'random_trees' as the subclustering method with gene expression quantified for both species by alignment to the hg38 reference genome. Average gene expression across all six individuals in each species was used as the background column. The cut-off for the minimum average read count per gene among reference cells was set to 1, per software recommendation for bulk RNA-seq data. CNV prediction was performed with the 'i6' Hidden Markov Model, whose output CNV states were filtered with the included Bayesian mixture model with a threshold of 0.1 to find the most confident CNVs. All other options were set to their default values.

To check for copy number variation at a selected set of cell cycle-related genes, we analyzed whole-genome shotgun sequencing data. Genomic DNA from all libraries sequenced on an Illumina sequencer in 151 bp paired-end mode was provided for analysis courtesy of the laboratory of Gregory Wray and mapped to chimpanzee reference panTro6[129] using bwa-mem2 (https://ieeexplore.ieee.org/document/8820962) with default parameters. PicardTools (https://broadinstitute.github.io/picard) was used to add read group information and mark

duplicates, and baseline coverage histograms were generated using BEDTools genomecov,[130] from which the 5th, 50th, and 95th percentile of coverage for each library, both genome-wide and across chromosome X, were extracted. Gene-level features for all genes listed as cyclins, cyclin dependent kinases, and class III Cys-based CDC25 phosphatases in the HGNC database[131] were selected from a recent chimpanzee gene annotation[132] and the coverage at each base across the full length of each gene in the set for each library was counted and summed using samtools mpileup.[133] For this step, only primary alignments containing mapped reads not marked as duplicates, with minimum map quality of 20, were considered (samtools view -F1284 -q20).

CEV-v1 library selection criteria

Candidate genes for the CEV-v1 library were chosen based on the union of the bootstrapping method, MAGeCK analysis, and additional genes selected based on notable evolutionary history. The two methods were largely complementary. In total, we classified 339 candidate species-specific genetic dependencies with MAGeCK and 796 candidates with the more permissive bootstrapping method (57 candidate genes with “notable evolutionary history” were later added). By constructing the CEV-v1 validation library based on the union of these gene sets, we aimed to compile a permissive list of candidates to account for potential false negatives from the initial screens.

963 candidate species-specific essential or proliferation suppressor genes were selected for inclusion in the Comparative Essential Validation (CEV-v1) sgRNA library according to a series of criteria (Figure S3.3D).

For MAGeCK, genes with false discovery rates (FDRs) less than 10% for both individuals within a species and FDRs greater than 25% in both individuals from the opposite species were considered as candidates for validation screening (*n* = 418 genes).

Lastly, a total of 57 candidate genes in the CEV-v1 validation library were selected based on notable evolutionary history. Of these, 42/57 covered previously annotated human-specific gene duplication events[56,134] that were targetable by unique sgRNAs. These gene duplications included genes such as ARHGAP11A/B and NOTCH2/2NL, which have been implicated in the expansion of the human cortex.[9–12] The remaining 15/57 genes were approaching significance and were selected based on gene function, such as AGO1/2 and DROSHA, which we hypothesized to be involved in repression of transposable retroelements expressed in pluripotent stem cells.

Data analysis for CEV-v1 validation screens

Sequencing data were aligned to CEV-v1 and quantified using the ScreenProcessing pipeline and MAGeCK. A matrix containing sgRNA counts from all CEV-v1 screens (excluding C3649 and Pt5-C due to non-responsiveness to *MDM2*/*TP53* perturbations) was assembled and used as input for differential analysis by DESeq2. Briefly, each sample was annotated by species, individual, and timepoint and a design matrix was created to model the species-specific effect of time (t0 vs. tfinal) while controlling for individual effects (modeled as fixed effects) within each species. The human and chimpanzee species terms were then contrasted to extract a Benjamini–Hochberg-adjusted *P*-value and log2 fold-change for each sgRNA. sgRNA adjusted *P*-values were combined into gene FDRs using alpha-robust rank aggregation (α-RRA) from MAGeCK and the α threshold (to remove the effect of insignificant sgRNAs from the assessment of gene significance) was set according to the fraction of sgRNAs with an adjusted *P*-value < 0.01. For each gene, log2 fold-change was computed as the mean of the four sgRNAs with the largest absolute fold-change. To exclude genes with shared effects from being erroneously called as species-specific, any gene with an FDR in both the human and chimpanzee species terms less than the highest FDR for any gene with at least one sgRNA passing the α threshold in α-RRA was discarded. For each gene in the set of 75 genes with species-specific effects reported in

Figure 3.2, we required that three conditions be met in the chimpanzee–human contrast term: (1) gene FDR < 0.01, (2) at least three sgRNAs targeting the gene pass the α threshold in α-RRA, and (3) gene $\log_2$ fold-change difference ≥ 0.75 between species. We used the STRING database v11.5[71] to identify known and predicted protein–protein interactions among this set of 75 genes.

To quantify sources of variation in CEV-v1 screens, a matrix of sgRNA counts was assembled as described above and normalized using edgeR145 calcNormFactors. Normalized sgRNA counts were then prepared for linear modeling using variancePartition voomWithDreamWeights and a linear mixed model was fit using variancePartition fitExtractVarPartModel. The categorical variables species, individual, and timepoint were modeled as random effects. Because most individuals were only screened once, the individual term describes variation attributable to differences between independent screens as well as differences between individuals. For each gene (Figure S3.2B), gene-level estimates of variance were determined by computing the mean variance attributable to each variable for all sgRNAs targeting that gene.

Potential p53-dependent candidate genes were flagged according to two methods. First, each gene was cross-referenced to previous growth screens performed in eight p53 wild-type and six p53 mutant AML cell lines.[135] Second, the two p53-unresponsive chimpanzee cell lines were compared to the four p53-responsive chimpanzee cells with a two-tailed t-test.

Analysis of protein-coding and gene expression changes

To obtain coding sequences for homologous transcripts from human and chimpanzee reference genomes, we downloaded human protein and transcript sequences from Gencode release 36[136] and chimpanzee protein and transcript sequences from the Comparative Annotation Toolkit[137] annotation on reference version panTro6 produced as part of a recent study.[132] For each human transcript of each protein coding gene, we obtained the transcript sequence and its canonical

translation, and we then extracted the corresponding chimpanzee transcript and canonical translation by matching the Ensembl transcript ID to its chimpanzee counterpart ("source transcript" field in the chimpanzee gene annotation). For both the human and chimpanzee sequence of each transcript, we then compared the translated sequence at every possible start codon and frame to the canonical amino acid sequence, determining the start codon and frame that produced the canonical amino acid sequence to be "correct" and removing bases thus found to belong to the 5' or 3' UTR (upstream of the correct start codon or downstream of the correct stop codon).

With coding sequences for homologous transcripts, we then aligned the human and chimpanzee protein sequences using the pairwise2 module from Biopython[138] with a BLOSUM62 substitution matrix. We then deleted codons in transcripts corresponding to amino acids that aligned to a gap in the other amino acid sequence. Finally, we deleted stop codons from the ends of sequences. We then wrote out each pairwise alignment to a control file for PAML[139] and calculated relevant statistics, including dN, dS, N, and S, using PAML's implementation of the Yang and Nielson (yn00) algorithm.[140] Finally, to avoid undefined values, we set dS to 1/S where dS was zero and selected the median dN value and median dN/dS value per gene for analysis.

Distributions of dN and dN/dS were compared for the full set of genes, DepMap common essential genes, and validation screen hits by two-sided Kolmogorov-Smirnov test (ks.test in R).

RNA-seq analysis

Raw bulk RNA-seq reads from wild-type chimpanzee and human iPSCs were adapter-trimmed using cutadapt[141] (with option -b AGATCGGAAGAGCACACGTCTGAACTCCAGTCA) and then pseudo-aligned to species-specific transcriptomes using kallisto[142] with options --single -l 200 -s 20. Transcriptomes were extracted from species-specific gtf annotations using the gffread utility[143]

using the -w output option. Human transcripts were obtained from the Gencode[136] comprehensive gene annotation v36 (GTF), using genome assembly hg38,[144] and the chimpanzee annotation was obtained from a recent study that produced a hierarchical alignment of primate genome assemblies[132] and annotated the assemblies using the Comparative Annotation Toolkit.[137]

To ensure consistency of gene names across the annotations, we downloaded the set of gene aliases from the HUGO Gene Nomenclature Committee website (www.genenames.org)[131] and searched for gene names present in the chimpanzee but missing in the human annotation, mapped to aliases present in the human annotation. This led us to rename five genes in the non-human primate annotations (DEC1 to *BHLHE40*, *DUSP27* to *DUSP29*, *AC073585.2* to *FAM24B*, *LOR* to *LOXL2*, and *TNRC6C-AS1* to *TMC6*); we also renamed *CCNP* in the human annotation to *CNTD2*.

After counting transcript abundances using kallisto, we converted them to gene counts using the tximport command in the tximport R package[145] with the options type='kallisto' and countsFromAbundance='no'. We then created one human and one chimpanzee data set in DESeq2.[59] We extracted VST-transformed counts for plotting using the function vst with option blind=TRUE and ran the DESeq linear model fitting using the function DESeq with betaPrior=TRUE.

FUCCI reporter live analysis

Images were pre-processed in MATLAB to rescale intensity to fit the full dynamic range of the image. Due to variable illumination at different stage positions, manual adjustments to the intensity range were made to minimize background signal and to enable accurate segmentation of mCherry-Cdt1+ nuclei.

Image segmentation was performed using a CellProfiler[146] pipeline with the following steps: 1) IdentifyPrimaryObjects, 2) MeasureObjectSizeShape, 3) MeasureObjectIntensity, 4) ConvertObjectsToImage, 5) TrackObjects, 6) SaveImages, 7) ExportToSpreadsheet. All image segmentation was performed in the mCherry channel, though GFP fluorescence was measured for each segmented object to determine sgRNA status. Primary cell segmentation was performed with a typical diameter range of 8 to 60 pixels and a threshold correction factor of 3. Thresholding was performed with the global minimum cross-entropy method. Nuclei were then tracked from one time point to the next with the overlap method, with a maximum distance to consider matches of 40 pixels.

Further post-processing refinement of the nuclei tracking step was performed to match nuclei that had moved enough during the 10 minute imaging interval to have zero overlap with the previous timepoint, but likely represented the same cell based on spatial proximity. For nuclei that were unassigned based on the CellProfiler TrackObjects module, we calculated the Euclidean distance to all nuclei in the previous frame based on the centroids of each object. If the unassigned nucleus was within 25 pixels of a nucleus from the previous frame that did not already have a matched nucleus in the current frame, we then reassigned both nuclei to the same label. We did not consider the possibility of nuclei splitting into two daughter cells, as the mCherry-Cdt1 reporter fluorescence is restricted to G1 phase, and thus nuclei were tracked to a single object in each frame.

We measured the lifetime of each nucleus based on the number of frames in which the nucleus could be segmented and tracked within our time series. Nuclei present at the initial time point or the final time point were filtered, as its lifetime could extend for an unknown period beyond the start or end of the experiment. In addition, nuclei with lifetime shorter than 30 minutes or 6 hours were filtered out, as these nuclei likely represent extremes in overall reporter expression that may

confound accurate measurement of G1-phase length. Lastly, nuclei were filtered based on area, with a minimum area of 60 square pixels and a maximum area of 550 square pixels. GFP was quantified as the average GFP intensity within the segmented nucleus, averaged across the entire time course of the object's lifetime. sgRNA expressing cells were classified as cells with mean GFP intensity greater than 0.15 (on a 0 to 1 scale) and sgRNA non-expressing cells within the same well were classified as cells with intensity less than 0.02. The distributions of nuclei lifetimes were then visualized on violin plots for each cell line and sgRNA perturbation.

Cerebral organoid brightfield image analysis

Brightfield images were captured on a EVOS M5000 microscope with a 4x objective. Day 18 was selected as a measurement time point because after 18, organoids are typically transferred from individual wells of a 96 well plate to 6-well petri dishes, which precludes standardized imaging. For a small number of conditions (40280L sgCDK2 and sgCCNE1), a fraction of organoids disaggregated due to cell death before day 18 and size measurements were not collected.

Images were first manually pre-processed in ImageJ to maximize contrast between organoid edges and background debris. Contrast adjusted images were then quantified with a CellProfiler[146] pipeline with the following steps: 1) IdentifyPrimaryObjects, 2) MeasureObjectSizeShape, 3) ConvertObjectsToImage, 4) SaveImages, 5) ExportToSpreadsheet. Organoid segmentation was performed with a typical diameter range of 150 to 600 pixels and a threshold correction factor of 1. Thresholding was performed with the global minimum cross-entropy method. Organoid segmentation was manually checked for accuracy and absence of multiple objects.

Data and code availability

- Raw sequencing data are deposited on GEO accession number GSE212297.
- All code for the analyses performed on the CRISPRi screens is publicly available at https://github.com/tdfair.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGEMENTS

We thank Eva Okrent (www.evaokrent.com) for all illustrations contained in this manuscript. We thank Joseph Min for running the CaSpER CNV pipeline, Sofie Salama and Andrew Field for providing orangutan iPSCs, Martin Kampmann for providing pC13N-dCas9-BFP-KRAB plasmids, Gregory Wray for providing whole-genome shotgun sequencing data from PSC lines, and Brian DeVeale for helpful discussions and critical review of the manuscript. We acknowledge the following funding sources: Helen Hay Whitney Foundation Postdoctoral Fellowship (R.S.), Ruth L. Kirschstein National Research Service Predoctoral Fellowship Award F31 HG011569-01A1 (T.F.), Weill Neurohub Fellowship (N.K.S.), Fannie and John Hertz Fellowship (R.A.S.), NSF Graduate Research Fellowship (R.A.S.), National Institutes of Health DP2MH122400-01 (A.A.P.), Schmidt Futures Foundation (A.A.P.), and Shurl and Kay Curci Foundation Innovative Genomics Institute Award (A.A.P.). J.S.W. is a Howard Hughes Medical Institute Investigator. A.A.P. is a New York Stem Cell Foundation Robertson Investigator.

Author contributions

R.S., T.F., J.S.W., and A.A.P. conceived the study design, executed all experiments, and wrote the manuscript. N.K.S. processed RNA-seq data and performed dN/dS calculations. R.A.S. provided the H1 *CLYBL* CRISPRi ES cell line and *AAVS1* integration plasmids. B.J.P. provided PSC lines and guidance during the primary screens. J.S.W. and A.A.P. supervised all aspects of this work.

Declaration of interests

The authors declare no competing interests.

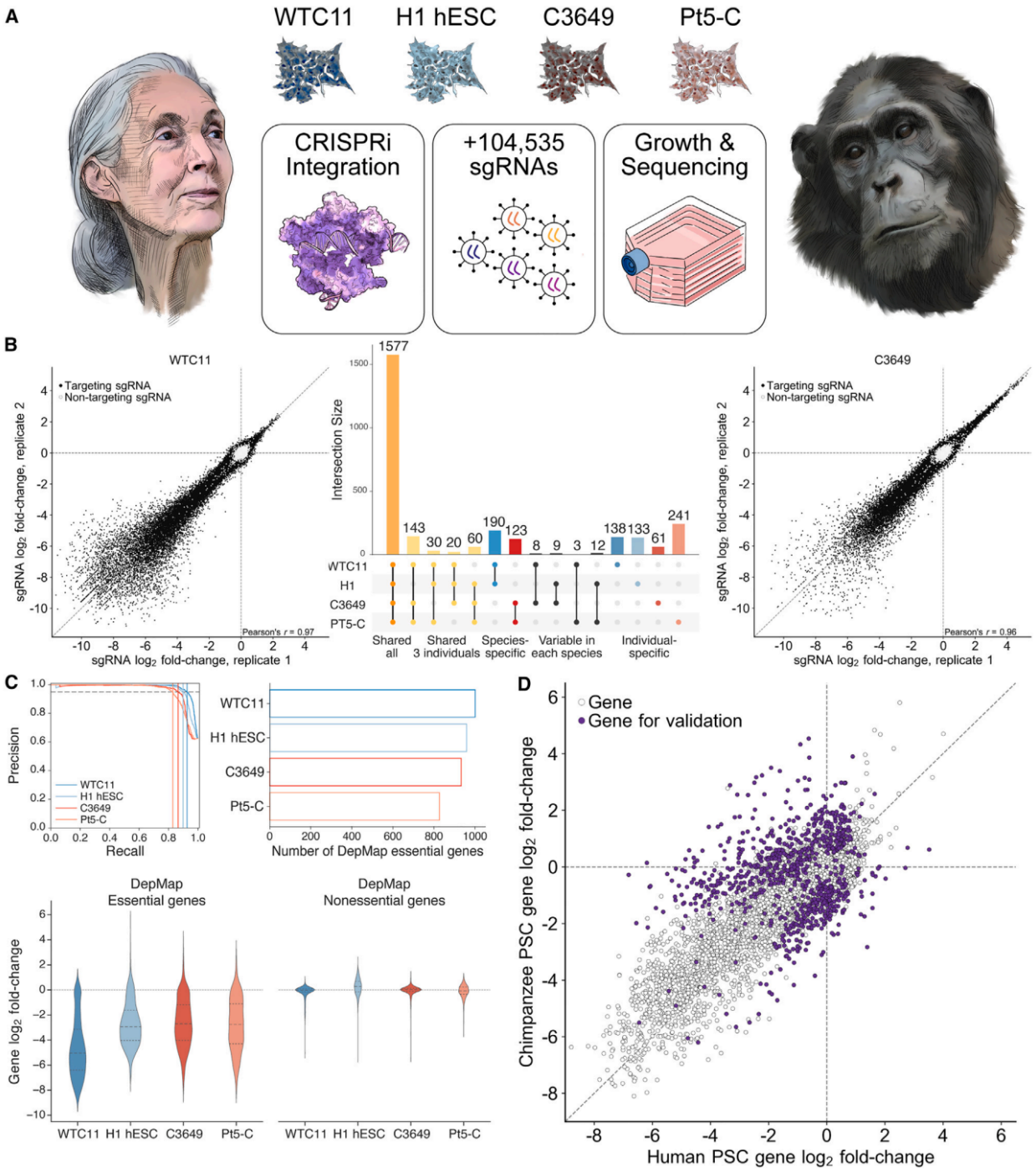


Figure 3.1. Genome-wide CRISPRi screens in human and chimpanzee stem cells identify candidate species-specific genetic dependencies.

(A) Schematic of CRISPRi screening approach, with original portraits of Jane Goodall and an adult female chimpanzee from Kibale Chimpanzee Project for artistic representation of human and chimpanzee species differences. Two human (WTC11 and H1) and two chimpanzee (C3649 and Pt5-C) PSC lines were engineered to express dCas9-KRAB, infected with the lentiviral hCRISPRi-v2 sgRNA library, and grown competitively for 10 days. Depleted and enriched sgRNAs were detected by high-throughput sequencing.

(B) Scatterplots of sgRNA $\log_2$ fold change for WTC11 and C3649 technical replicates and UpSet plot showing the intersection of essential genes across all four screens.
(C) Precision-recall analysis (top left) for each screen. Precision and recall were determined using DepMap essential and non-essential genes. The number of DepMap essential genes (top right) identified by MAGeCK (5% FDR, $\log_2$ fold change < -1.5). Distribution of $\log_2$ fold change for DepMap essential (bottom left) and non-essential (bottom right) genes.
(D) Species-level gene $\log_2$ fold change across genome-wide CRISPRi screens. Gene-level phenotypes were computed as the mean of the three sgRNAs with the largest absolute $\log_2$ fold change. sgRNAs lacking perfect-match targets in the chimpanzee genome were excluded from analysis.

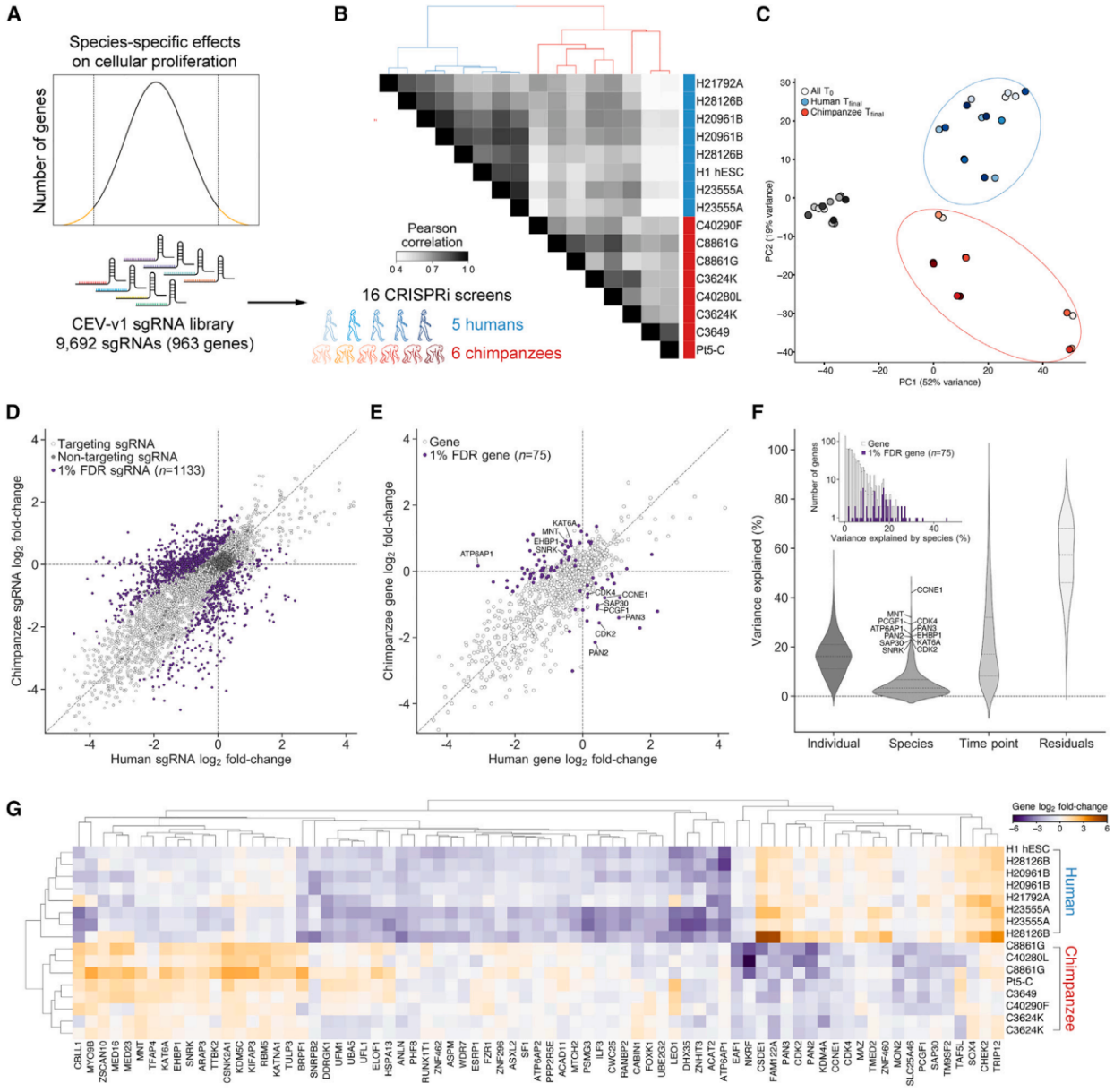


Figure 3.2. Species-specific genetic dependencies validated across five human and six chimpanzee individuals.

(A) Schematic of validation sgRNA library design and CRISPRi screening approach.

(B) Heatmap of Pearson correlations and hierarchical clustering for sgRNA profiles across 16 validation CRISPRi screens. Individuals listed twice are replicate screens performed in separate laboratories.

(C) Principal-component analysis of sgRNA counts at t_0 (black circle) and t_{final} (red and blue circles).

(D) Scatterplot of $\log_2$ fold change of sgRNA counts, modeled by DESeq2. 1,133 sgRNAs with significant species differences (FDR < 0.01) colored in purple and negative-control sgRNAs colored in dark gray.

(E) Species-level gene $\log_2$ fold change across validation CRISPRi screens. Gene-level phenotypes were computed as the mean of the four sgRNAs with the largest absolute $\log_2$ fold

change. The 12 genes with the greatest variance in sgRNA $\log_2$ fold change attributable to species are labeled.
(F) Dream-variancePartition analysis for quantifying sources of variation in sgRNA counts attributable to individual, species, and time point (t_0 vs. t_{final}).
(G) Heatmap of gene $\log_2$ fold change and hierarchical clustering for 75 genes with species-specific effects on cellular proliferation across validation CRISPRi screens (1% FDR).

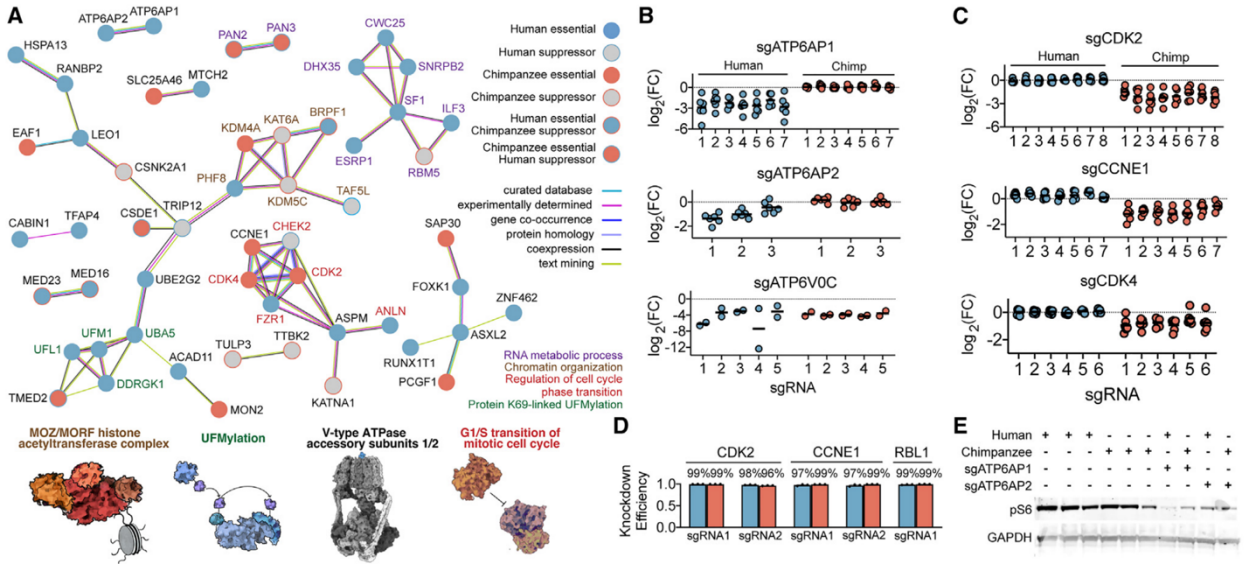


Figure 3.3. Core species-specific genetic dependencies.
(A) Species-specific genetic dependencies with STRING protein-protein associations. Illustrations of pathways and protein complexes with coherent species-specific effects.
(B) Strip plots of $\log_2$ fold change for sgRNAs targeting *ATP6AP1*, *ATP6AP2*, and *ATP6V0C*. Data derived from CEV-v1 validation screens for *ATP6AP1*, *ATP6AP2*, and only from primary genome-wide screen plotted for *ATP6V0C*.
(C) Strip plots of $\log_2$ fold change for sgRNAs targeting *CDK2*, *CCNE1*, and *CDK4*. Each circle represents the sgRNA $\log_2$ fold change for one sgRNA in one human (blue) or chimpanzee (red) individual. Each strip plot contains a variable number of columns, corresponding to the number of significant sgRNAs targeting each gene.
(D) qRT-PCR measurements of knockdown efficiency in human (28126B, blue, n = 3) and chimpanzee (40280L, red, n = 3) PSCs.
(E) Western blot for phospho-S6 (pS6) expression and GAPDH loading control for three wild-type human (H1, 21792A, and 28126B) and three wild-type chimpanzee (3624K, 40280L, and 8861G) cell lines, and cell lines depleted for *ATP6AP1* or *ATP6AP2* (28126B and 40280L).

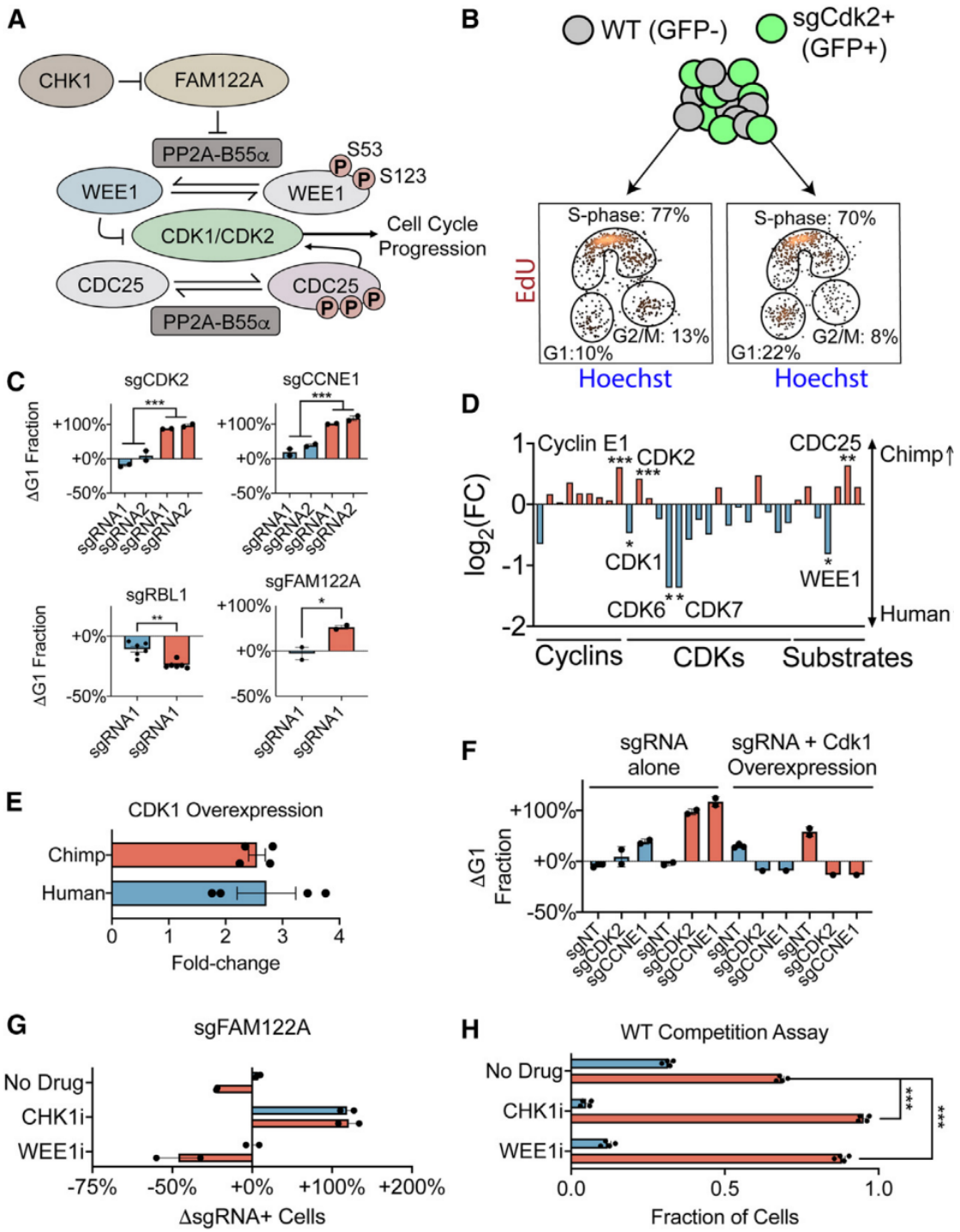


Figure 3.4. Divergent regulation of cell-cycle progression in human and chimpanzee cells. (A) Schematic for CDK1/CDK2 regulatory network. CDK1 and CDK2 phosphorylate key substrates WEE1 and CDC25, leading to degradation of WEE1 and activation of CDC25. Phosphatase PP2A dephosphorylates WEE1 and CDC25 at the same sites. CHK1 inhibits FAM122A, and FAM122A inhibits PP2A.

(B) Cell-cycle proportions in chimpanzee wild-type cells (GFP^-) and sgRNA-containing cells (GFP^+) grown in co-culture.

(C) Change in the fraction of human (28126B, blue) and chimpanzee (40280L, red) cells in G1 phase upon knockdown of CDK2 ($p < 10^{-3}$, n = 2), cyclin E1 ($p < 10^{-3}$, n = 2), RBL1 ($p < 10^{-2}$, n = 6), and FAM122A ($p < 0.05$, n = 2), calculated by two-tailed t test.
(D) Comparative gene expression data from human and chimpanzee PSCs for core cell-cycle regulators (*$p < 0.05$, **$p < 10^{-2}$, ***$p < 10^{-3}$, p values from[15]).
(E) qRT-PCR measurements of the degree of CDK1 overexpression (n = 2 with two qRT-PCR primer sets).
(F) Change in the fraction of cells in G1 phase upon overexpression of CDK1 in conjunction with CDK2 or cyclin E1 knockdown.
(G) Change in the fraction of FAM122A sgRNA-containing cells in the presence of no drug, Chk1 inhibitor prexasertib (CHK1i), or WEE1 inhibitor adavosertib (WEE1i) (n = 2, 2 days of drug treatment).
(H) Fraction of wild-type human (blue, 21792A) vs. wild-type chimpanzee (red, 40280L) cells grown in co-culture in the presence of no drug, CHK1i, or WEE1i (n = 4). For all bar charts, data are plotted as mean ± SEM with individual data points representing biological replicates.

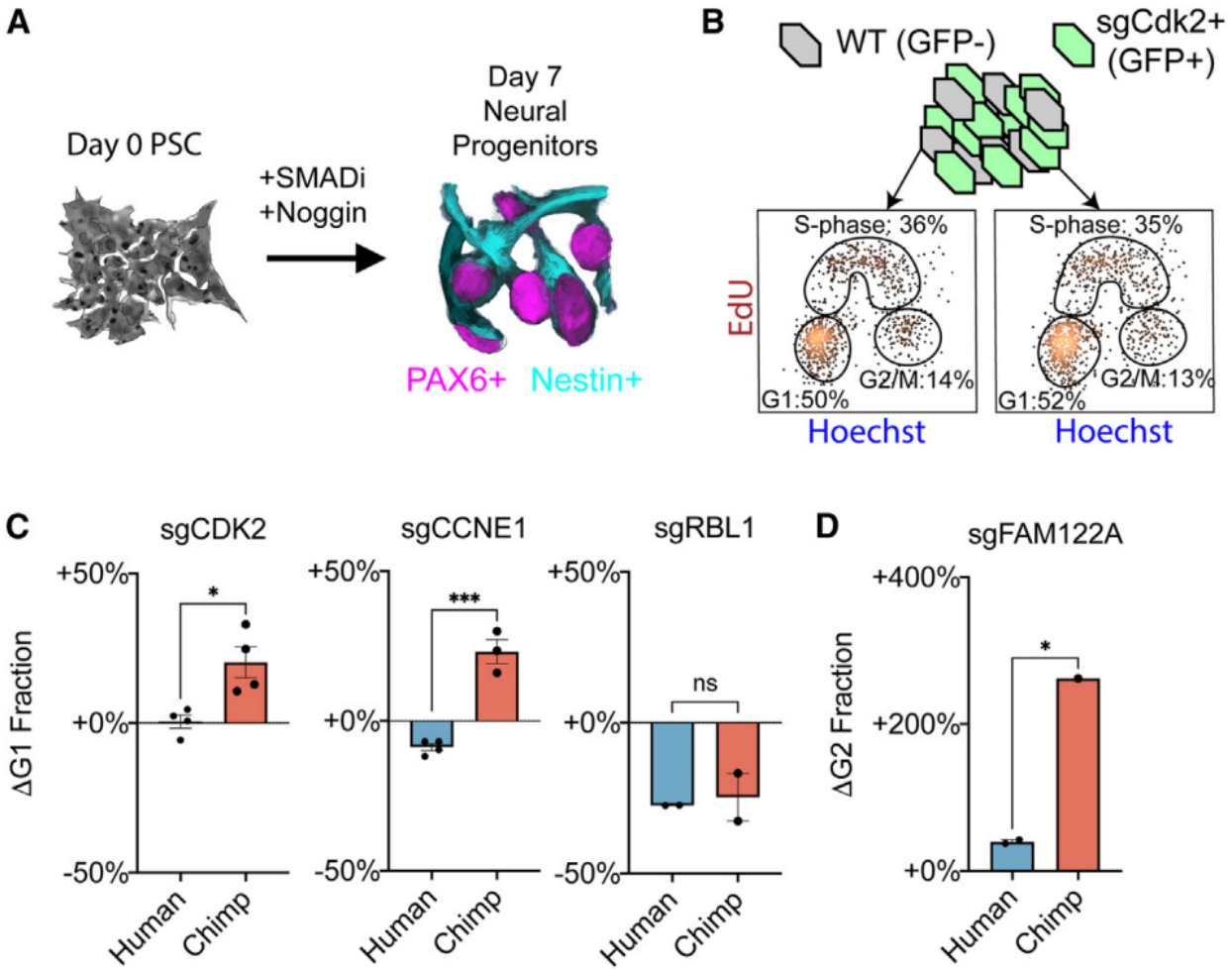


Figure 3.5. Human-specific robustness to cell-cycle perturbations persists in neural progenitor cells.

(A) Schematic for differentiation of PSCs into neural progenitor cells (NPCs).

(B) Cell-cycle proportions in human wild-type neural progenitor cells (GFP^-) and sgRNA-containing cells (GFP^+) grown in co-culture.

(C) Change in the fraction of human (28126B, blue) and chimpanzee (40280L, red) NPCs in G1 phase upon depletion of *CDK2* ($p < 0.05$, n = 4), *CCNE1* ($p < 10^{-3}$, n = 4 and n = 3), or *RBL1* (n.s., n = 2), calculated by two-tailed t test.

(D) Change in the fraction of NPCs in G2 phase upon depletion of *FAM122A* ($p < 0.05$, two-tailed t test, n = 2 and n = 1). Bar charts in (C) and (D) plotted as mean ± SEM, with individual data points representing biological replicates.

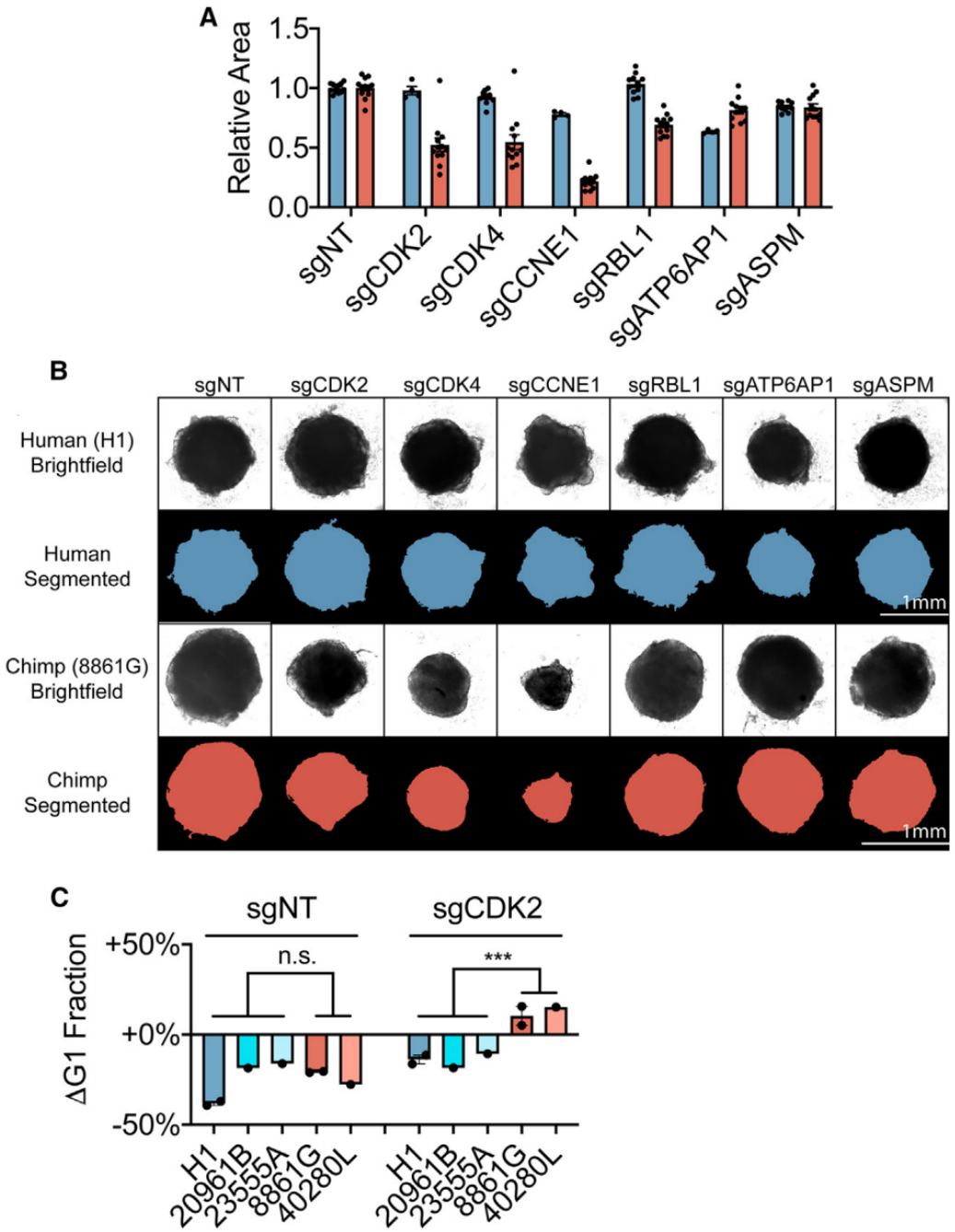


Figure 3.6. Human robustness to cell-cycle perturbations in cerebral organoids.
(A) Organoid size measurements for human (H1) and chimpanzee (8861G) cerebral organoids, measured on day 18 by bright-field microscopy (n = 4–12). Bar charts plotted as mean ± SEM, with each individual data point representing an independent organoid.
(B) Representative images of organoids from each sgRNA condition and corresponding image segmentation.
(C) Cell-cycle measurements for day 9 human (H1, 20961B, and 23555A; n = 2, n = 1, and n = 1; blue) and chimpanzee (8861G and 40280L; n = 2 and n = 1; red) organoids. Change in the fraction of cells in G1 phase in cells expressing an sgRNA targeting *CDK2* ($p < 0.001$, two-tailed t test) or a non-targeting sgRNA ($p = 0.56$, two-tailed t test).

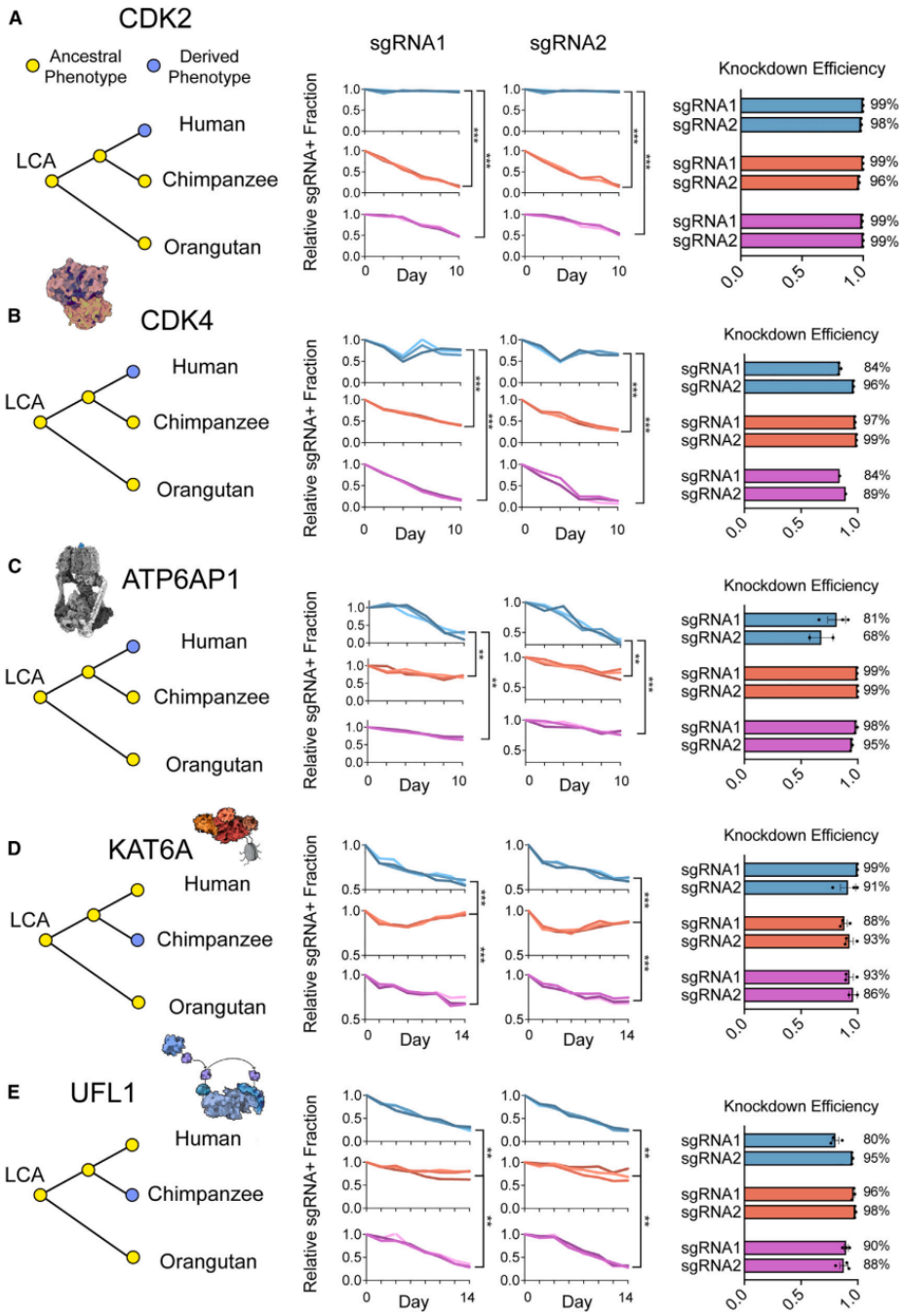


Figure 3.7. Orangutan PSCs suggest evolutionary origin of species-specific genetic dependencies.

(A) Change in the relative fraction of *CDK2* sgRNA-containing cells over time in human (blue, 28126B), chimpanzee (red, 40280L), and orangutan PSCs (purple, n = 3 for each species) (**p < 10^{-2}, ***p < 10^{-3}; p values calculated by two-tailed t test on final time point). qRT-PCR measurements of sgRNA knockdown efficiency for each sgRNA in all three species (n = 1–3). (B–E) Relative sgRNA fraction over time and qRT-PCR measurements for sgRNAs targeting (B) *CDK4*, (C) *ATP6AP1*, (D) *KAT6A*, and (E) *UFL1*. Individual data points represent biological replicates and qRT-PCR bar charts plotted as mean ± SEM.

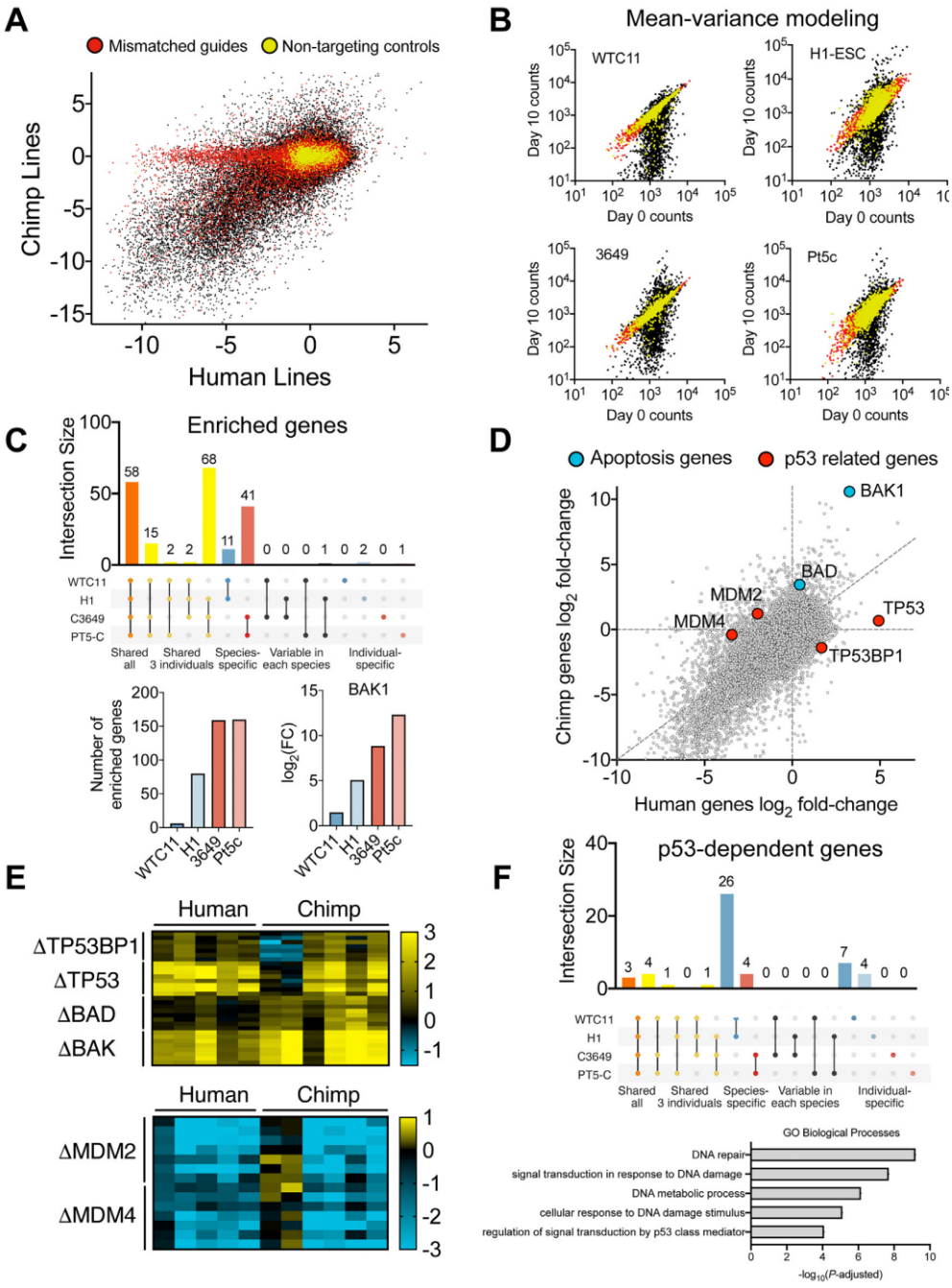
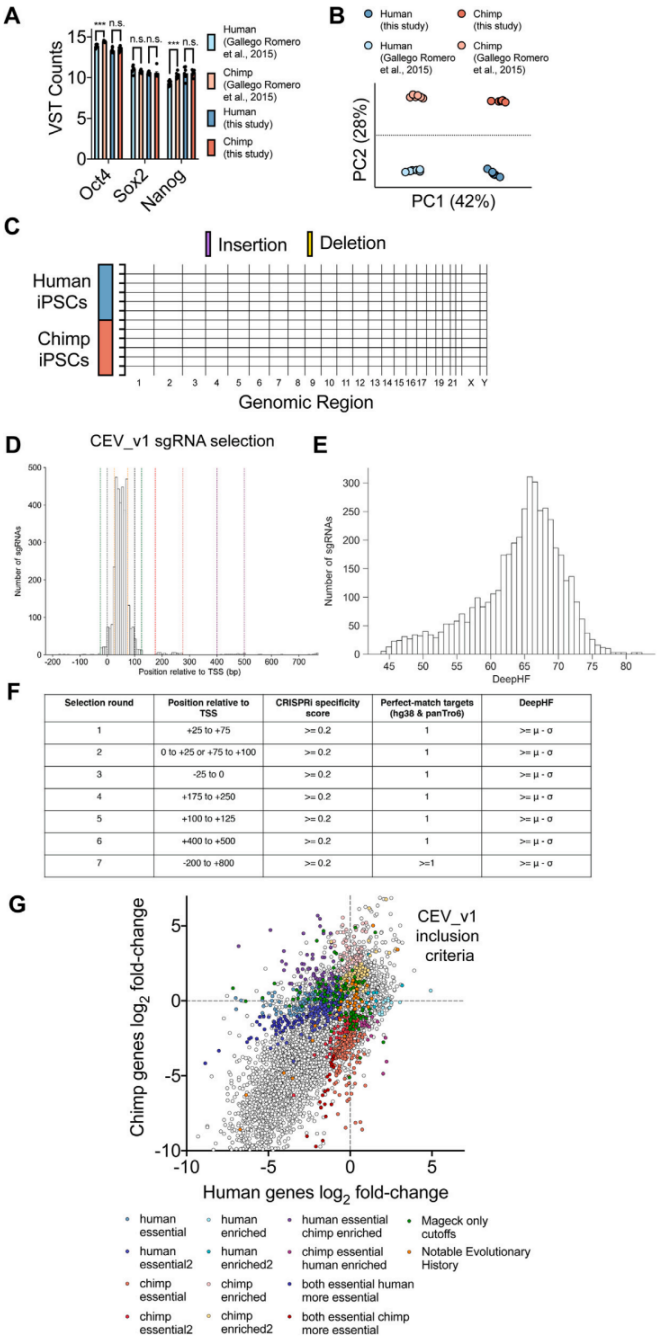


Figure S3.1. Genome-wide CRISPRi screens in human and chimpanzee PSCs, related to Figure 3.1.

(A) Log_2 fold change of sgRNA counts from genome-wide CRISPRi screens using the hCRISPRi-v2 sgRNA library, averaged across two human and two chimpanzee cell lines. sgRNAs containing mismatches to the chimpanzee genome are colored in red and non-targeting sgRNAs are colored in yellow. A substantial number of mismatched sgRNAs targeting essential genes are depleted in human PSCs but not in chimpanzee PSCs.

(B) Depletion or enrichment of sgRNA counts at growth day 10 compared to growth day 0. Non-targeting sgRNAs are colored in yellow, and sgRNAs characterized as non-significant by mean-variance modeling of a negative binomial distribution are colored in red.

(C) UpSet plot showing the intersection of enriched genes across all four screens. Total number of enriched genes with sgRNA enrichment score greater than 2-fold and FDR < 0.05 for each individual screen. Average $\log_2$ fold change for *BAK1* sgRNAs for each individual screen.
(D) Averaged species-level gene $\log_2$ fold change for apoptosis related genes and p53 related genes.
(E) Heatmap displaying $\log_2$ fold change of sgRNA counts across five human and six chimpanzee PSCs, with columns 1, 6, and 7 showing primary genome-wide screening data and remaining columns showing data from secondary validation screening. Columns 6 and 7 (Pt5-C and C3649) represent the two chimpanzee PSCs that exhibit TP53 mutant phenotypes.
(F) UpSet plot showing the intersection of candidate p53-dependent essential genes across all four screens. Gene ontology (GO) enrichment terms for the set of 127 candidate p53-dependent genes.



Selection round	Position relative to TSS	CRISPRi specificity score	Perfect-match targets (hg38 & panTro6)	DeepHF
1	+25 to +75	>= 0.2	1	>= μ - σ
2	0 to +25 or +75 to +100	>= 0.2	1	>= μ - σ
3	-25 to 0	>= 0.2	1	>= μ - σ
4	+175 to +250	>= 0.2	1	>= μ - σ
5	+100 to +125	>= 0.2	1	>= μ - σ
6	+400 to +500	>= 0.2	1	>= μ - σ
7	-200 to +800	>= 0.2	>=1	>= μ - σ

Figure S3.2. CEV-v1 validation screens in human and chimpanzee PSCs, related to Figure 2.

(A) Bulk RNA-seq VST-transformed counts for key pluripotency markers *OCT4*, *SOX2*, and *NANOG* across CRISPRi-engineered PSCs from this study and original source lines.

(B) Principal-component analysis of bulk RNA-seq transcriptomes, with PC1 capturing batch effects between two separate sequencing experiments and PC2 capturing human and chimpanzee species differences.

(C) CaSpER analysis of chromosomal copy-number variations from bulk RNA-seq data across all newly engineered CRISPRi PSC lines as aligned to the human hg38 reference genome. For the chimpanzee genome, chromosome 2 refers to 2a and 2b.

(D) Distribution of positions relative to the FANTOM-annotated TSS for sgRNAs in CEV-v1. Vertical colored lines indicate the selection round in which sgRNAs were chosen.
(E) Distribution of DeepHF on-target predictions for sgRNAs in CEV-v1.
(F) Selection criteria for CEV-v1 sgRNA library.
(G) Inclusion criteria for 963 genes selected as candidate species-specific genetic dependencies.

Figure S3.3. 75 genes with robust species-specific effects on cellular proliferation, related to Figure 2.
Strip plots of $\log_2$ fold change in sgRNA enrichment or depletion for 75 genes with species-specific effects on cellular proliferation (FDR < 1%), calculated by a-RRA, colored by individual.

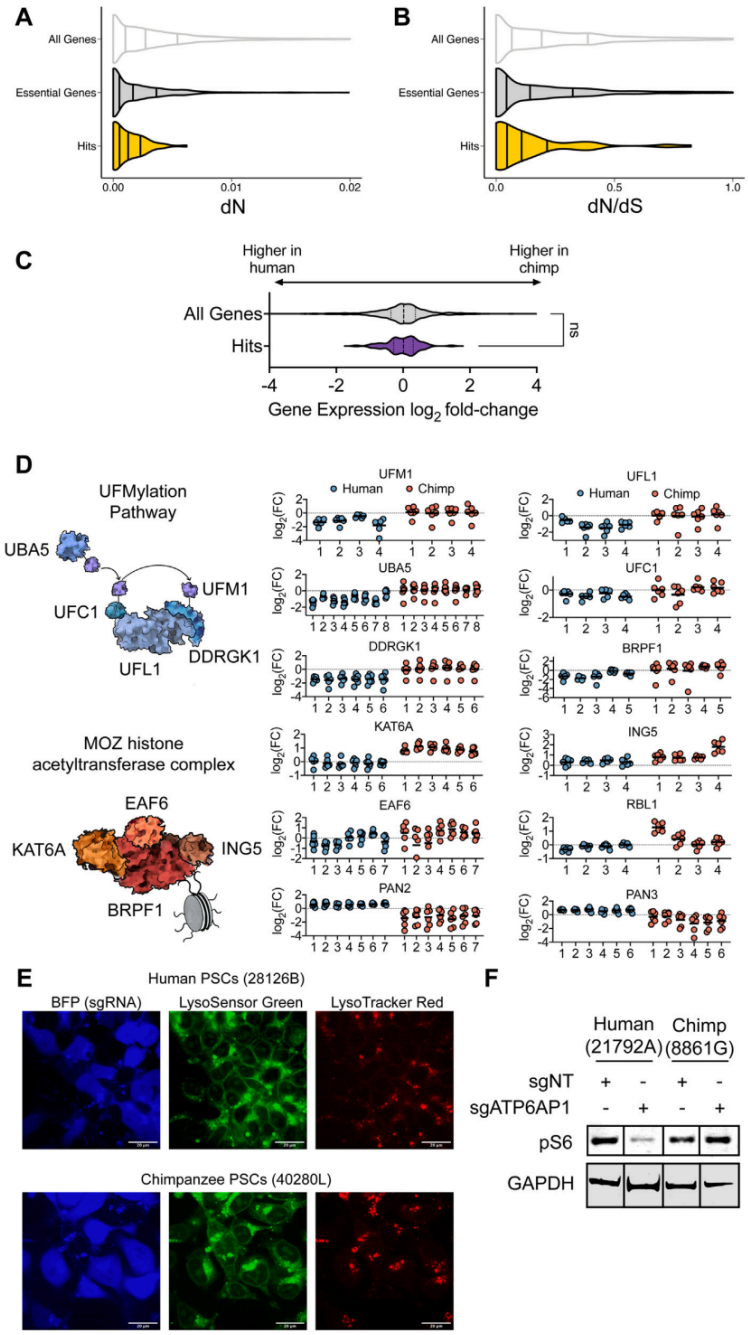


Figure S3.4. Species-specific genetic dependencies interact in biological processes and complexes, related to Figure 3.

(A and B) (A) dN, the rate of non-synonymous substitutions in a gene and (B) dN/dS values, the ratio between the rates of non-synonymous and synonymous substitutions, for 75 validated differential-essentiality genes from this study compared to all genes or essential genes.

(C) Comparative gene expression levels between human and chimpanzee PSCs for 75 validated differential-essentiality genes from this study vs. all genes expressed in PSCs.

(D) sgRNA depletion or enrichment for all active sgRNAs targeting members of the UFMylation pathway, MOZ histone acetylation complex, *RBL1*, and the PAN2/3 complex. Each circle represents the sgRNA $\log_2$ fold change for one sgRNA in one human (blue) or chimpanzee (red)

individual. Each strip plot contains a variable number of columns, corresponding to the number of significant sgRNAs targeting each gene. Genes with only one significant sgRNA (*ING5* and *RBL1*) are scored as less significant compared to genes with multiple significant sgRNAs and require validation of on-target effects.
(E) Human PSCs (28126B) and chimpanzee PSCs (40280L) with wild-type cells (BFP^-) co-cultured with cells expressing sgATP6AP1 (BFP^+) stained with LysoSensor Green and LysoTracker red. Scale bars, 20 mm.
(F) Western blot for phospho-S6 (pS6) expression and GAPDH loading control for one additional human (21792A) and one chimpanzee (8861G) cell line depleted for *ATP6AP1* with a non-targeting sgRNA control.

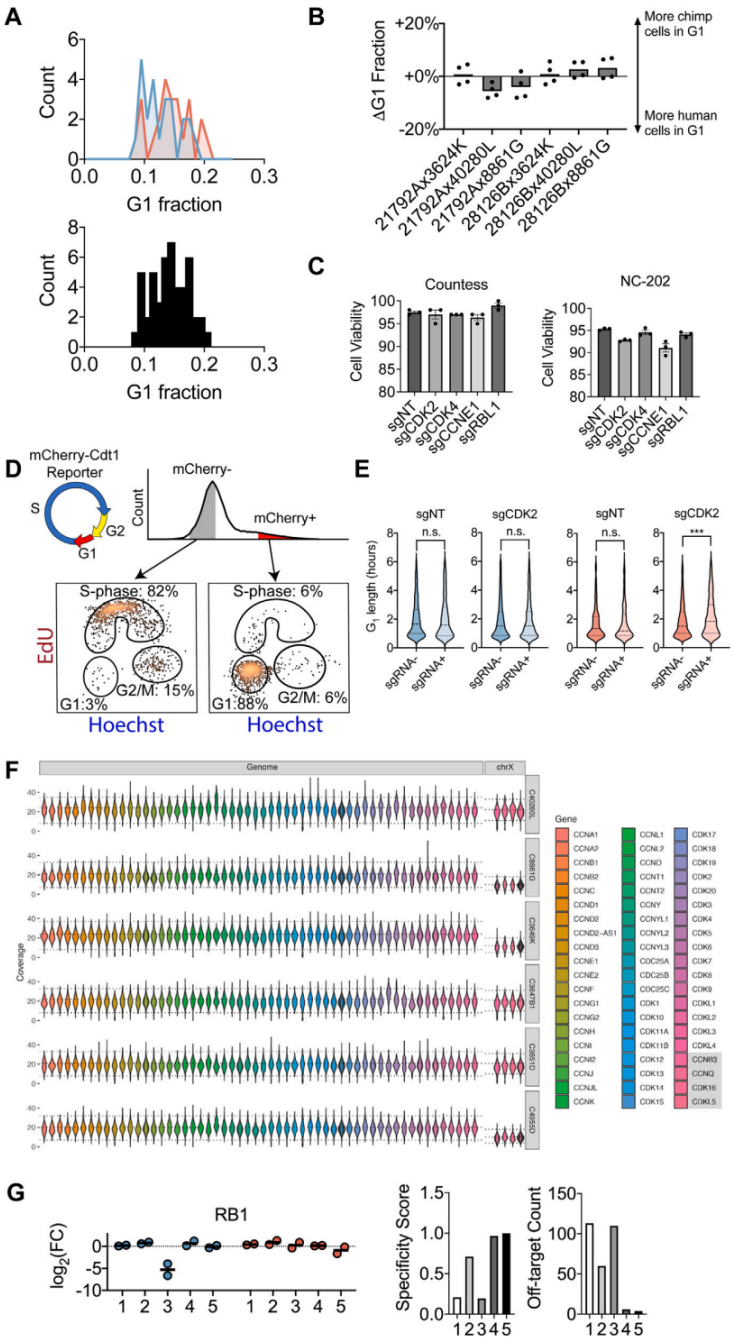


Figure S3.5. Measurements of species differences in G1 phase length, related to Figure 4.
(A) Absolute fraction of human (21792A and 28126B, blue) and chimpanzee (3624K, 40280L, and 8861G, red) PSCs in G1 phase, as measured by EdU incorporation and DNA content. Histogram of G1 fraction for both human and chimpanzee cells (black).
(B) Relative fraction of cells in G1 phase for six pairs of human and chimpanzee PSCs co-cultured in the same well (n = 4). (C) Cell viability measurements for chimpanzee PSCs (40280L) expressing sgRNAs targeting cell-cycle regulators (n = 3, SEM).
(D) FUCCI reporter cell line cross validation with cell-cycle proportion measurements via EdU incorporation and DNA content.
(E) Quantification of G1 phase length by live imaging of human (21792A) and chimpanzee (3624K) PSCs infected with either an sgRNA targeting CDK2 or a nontargeting sgRNA.

(F) Whole-genome shotgun sequence coverage at all genes in the HUGO Gene Nomenclature Committee gene groups cyclins, cyclin-dependent kinases, and class III Cys-based CDC25 phosphatases (https://www.genenames.org/). Each violin represents the coverage at each base across the entire body for each gene. The horizontal lines correspond to the 5th, 50th, and 95th percentiles of baseline coverage across the entire genome (‘‘Genome’’ panel) or the X chromosome (‘‘chrX’’ panel). Four genes in these sets located on chromosome X (CDKL5, CDK16, CCNB3, and CCNQ) are shown separately to account for different baseline coverage; these gene names are outlined in a gray box in the legend. The top three rows correspond to chimpanzee PSCs from individuals used in the present study (C40280L, C8861G, and C3649), while the bottom three rows correspond to similarly reprogrammed chimpanzee individuals.
(G) Strip plot of $\log_2$ fold change for sgRNAs targeting *RB1* with data derived from only from primary genome-wide screen. Computationally predicted specificity score and off-target counts for each of the five sgRNAs targeting *RB1*.

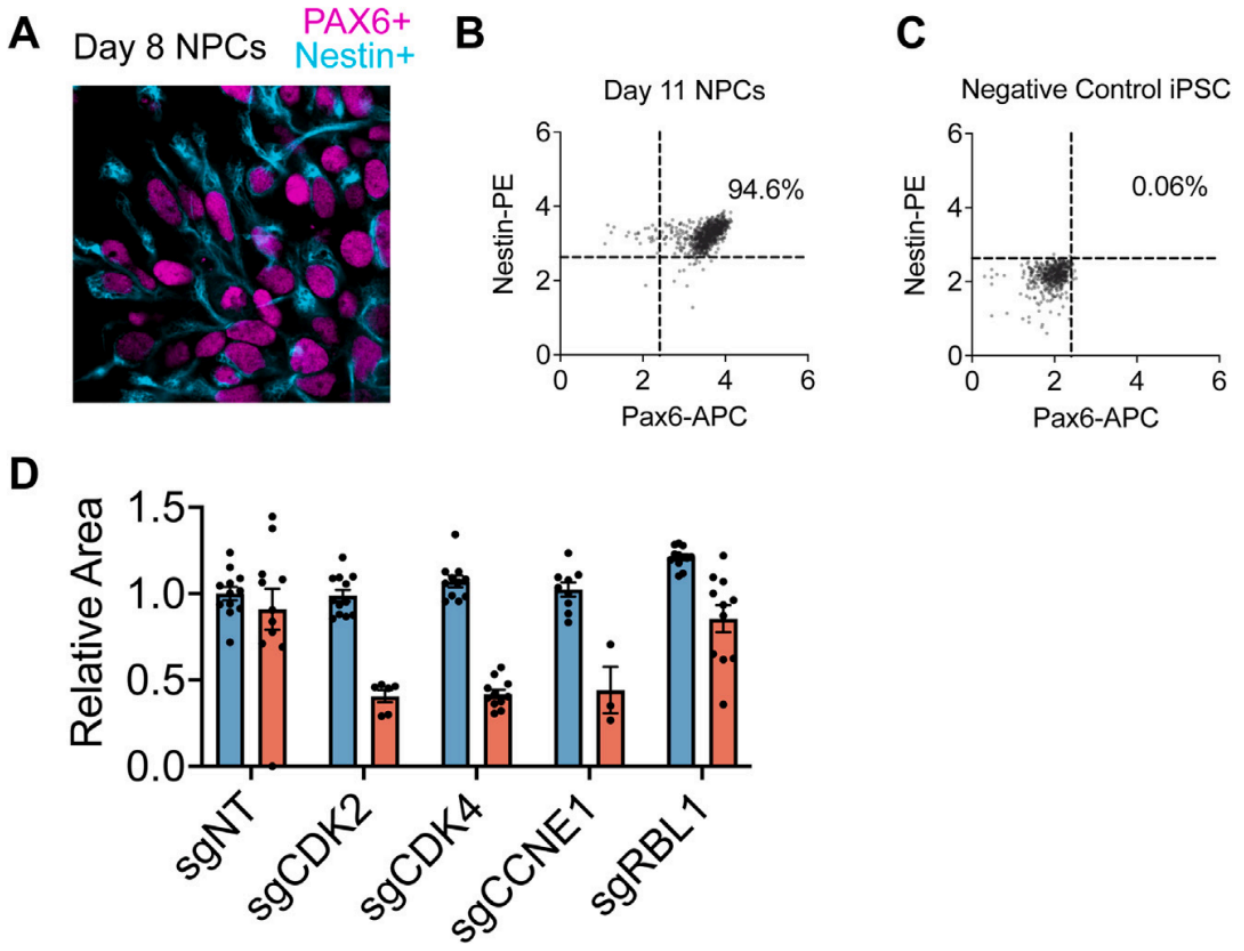


Figure S3.6. Derivation of human and chimpanzee NPCs, related to Figure 5.
(A) Chimpanzee neural progenitor cells (40280L) stained for Pax6 and Nestin, visualized by confocal microscopy.
(B) Chimpanzee neural progenitor cells (40280L) stained for Pax6 and Nestin, quantified by flow cytometry.
(C) Negative-control PSCs stained for Pax6 and Nestin, quantified by flow cytometry.
(D) Organoid size measurements for human (23555A, blue) and chimpanzee (40280L, red) cerebral organoids, measured on day 18 by bright-field microscopy (N = 3–12). Bar charts plotted as mean ± SEM, with each individual data point representing an independent organoid.

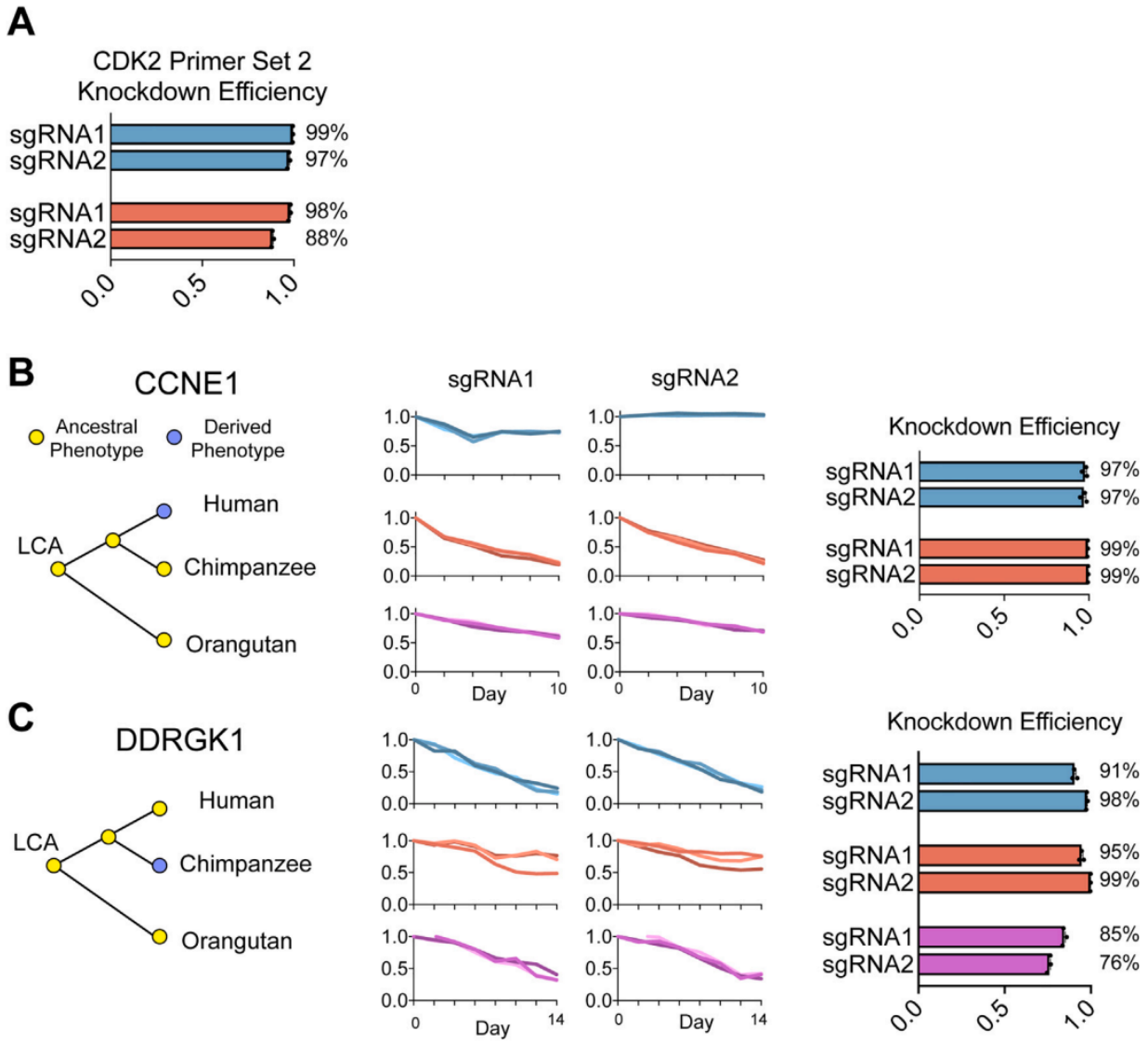


Figure S3.7. Tri-species comparison of human, chimpanzee, and orangutan PSCs expressing sgRNAs targeting CCNE1 and DDRGK1, related to Figure 7.

(A) qRT-PCR measurements of sgRNA knockdown efficiency for sgCDK2, measured with an alternative primer set.

(B and C) Change in the relative fraction of *CCNE1* (B) and *DDRGK1* (C) sgRNA-containing cells over time in human (28126B, blue), chimpanzee (40280L, red), and orangutan (purple) PSCs. qRT-PCR measurements of sgRNA knockdown efficiency for each sgRNA.

REFERENCES

1. Muller, M. N., Wrangham, R. W. & Pilbeam, D. R. *Chimpanzees and Human Evolution*. (Harvard University Press, 2017).
2. Varki, A. & Altheide, T. K. Comparing the human and chimpanzee genomes: searching for needles in a haystack. *Genome Res.* **15**, 1746–1758 (2005).
3. Dutrow, E. V. *et al.* Modeling uniquely human gene regulatory function via targeted humanization of the mouse genome. *Nat. Commun.* **13**, 304 (2022).
4. Aldea, D. *et al.* Repeated mutation of a developmental enhancer contributed to human thermoregulatory evolution. *Proc. Natl. Acad. Sci. U. S. A.* **118**, (2021).
5. Boyd, J. L. *et al.* Human-chimpanzee differences in a FZD8 enhancer alter cell-cycle dynamics in the developing neocortex. *Curr. Biol.* **25**, 772–779 (2015).
6. Enard, W. *et al.* Molecular evolution of FOXP2, a gene involved in speech and language. *Nature* **418**, 869–872 (2002).
7. Lai, C. S., Fisher, S. E., Hurst, J. A., Vargha-Khadem, F. & Monaco, A. P. A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature* **413**, 519–523 (2001).
8. Mora-Bermúdez, F. *et al.* Longer metaphase and fewer chromosome segregation errors in modern human than Neandertal brain development. *bioRxiv* 2022.03.30.486431 (2022) doi:10.1101/2022.03.30.486431.
9. Fiddes, I. T. *et al.* Human-Specific NOTCH2NL Genes Affect Notch Signaling and Cortical Neurogenesis. *Cell* **173**, 1356–1369.e22 (2018).
10. Florio, M., Namba, T., Pääbo, S., Hiller, M. & Huttner, W. B. A single splice site mutation in human-specific *ARHGAP11B* causes basal progenitor amplification. *Sci Adv* **2**, e1601941 (2016).
11. Heide, M. *et al.* Human-specific *ARHGAP11B* increases size and folding of primate neocortex in the fetal marmoset. *Science* **369**, 546–550 (2020).

12. Suzuki, I. K. *et al.* Human-Specific NOTCH2NL Genes Expand Cortical Neurogenesis through Delta/Notch Regulation. *Cell* **173**, 1370–1384.e16 (2018).
13. Kriegstein, A., Noctor, S. & Martínez-Cerdeño, V. Patterns of neural stem and progenitor cell division may underlie evolutionary cortical expansion. *Nat. Rev. Neurosci.* **7**, 883–890 (2006).
14. Rakic, P. A small step for the cell, a giant leap for mankind: a hypothesis of neocortical expansion during evolution. *Trends Neurosci.* **18**, 383–388 (1995).
15. Gallego Romero, I. *et al.* A panel of induced pluripotent stem cells from chimpanzees: a resource for comparative functional genomics. *Elife* **4**, e07103 (2015).
16. Marchetto, M. C. N. *et al.* Differential L1 regulation in pluripotent stem cells of humans and apes. *Nature* **503**, 525–529 (2013).
17. Pavlovic, B. J., Blake, L. E., Roux, J., Chavarria, C. & Gilad, Y. A Comparative Assessment of Human and Chimpanzee iPSC-derived Cardiomyocytes with Primary Heart Tissues. *Sci. Rep.* **8**, 15312 (2018).
18. Blake, L. E. *et al.* A comparative study of endoderm differentiation in humans and chimpanzees. *Genome Biol.* **19**, 162 (2018).
19. Prescott, S. L. *et al.* Enhancer divergence and cis-regulatory evolution in the human and chimp neural crest. *Cell* **163**, 68–83 (2015).
20. Kanton, S. *et al.* Organoid single-cell genomic atlas uncovers human-specific features of brain development. *Nature* **574**, 418–422 (2019).
21. Marchetto, M. C. *et al.* Species-specific maturation profiles of human, chimpanzee and bonobo neural cells. *Elife* **8**, (2019).
22. Pollen, A. A. *et al.* Establishing Cerebral Organoids as Models of Human-Specific Brain Evolution. *Cell* **176**, 743–756.e17 (2019).
23. Nüsslein-Volhard, C. & Wieschaus, E. Mutations affecting segment number and polarity in Drosophila. *Nature* **287**, 795–801 (1980).

24. Wieschaus, E. & Nüsslein-Volhard, C. The Heidelberg Screen for Pattern Mutants of Drosophila: A Personal Account. *Annu. Rev. Cell Dev. Biol.* **32**, 1–46 (2016).

25. Miller, J. R. The Wnts. *Genome Biol.* **3**, REVIEWS3001 (2002).

26. Peifer, M. & Wieschaus, E. The segment polarity gene armadillo encodes a functionally modular protein that is the Drosophila homolog of human plakoglobin. *Cell* **63**, 1167–1176 (1990).

27. Hooper, J. E. & Scott, M. P. The Drosophila patched gene encodes a putative membrane protein required for segmental patterning. *Cell* **59**, 751–765 (1989).

28. Costa, M., Wilson, E. T. & Wieschaus, E. A putative cell signal encoded by the folded gastrulation gene coordinates cell shape changes during Drosophila gastrulation. *Cell* **76**, 1075–1089 (1994).

29. Dickinson, M. E. *et al.* High-throughput discovery of novel developmental phenotypes. *Nature* **537**, 508–514 (2016).

30. de Angelis, M. H. *et al.* Analysis of mammalian gene function through broad-based phenotypic screens across a consortium of mouse clinics. *Nat. Genet.* **47**, 969–978 (2015).

31. White, J. K. *et al.* Genome-wide generation and systematic phenotyping of knockout mice reveals new roles for many genes. *Cell* **154**, 452–464 (2013).

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

33. Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods* **11**, 783–784 (2014).

34. Wang, T., Wei, J. J., Sabatini, D. M. & Lander, E. S. Genetic screens in human cells using the CRISPR-Cas9 system. *Science* **343**, 80–84 (2014).

35. Pacini, C. *et al.* Integrated cross-study datasets of genetic dependencies in cancer. *Nat. Commun.* **12**, 1661 (2021).

36. Tsherniak, A. *et al.* Defining a Cancer Dependency Map. *Cell* **170**, 564–576.e16 (2017).

37. Gaspar-Maia, A., Alajem, A., Meshorer, E. & Ramalho-Santos, M. Open chromatin in pluripotency and reprogramming. *Nat. Rev. Mol. Cell Biol.* **12**, 36–47 (2011).

38. 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).

39. Liu, S. J. *et al.* CRISPRi-based genome-scale identification of functional long noncoding RNA loci in human cells. *Science* **355**, (2017).

40. Tian, R. *et al.* CRISPR Interference-Based Platform for Multimodal Genetic Screens in Human iPSC-Derived Neurons. *Neuron* **104**, 239–255.e12 (2019).

41. Kreitzer, F. R. *et al.* A robust method to derive functional neural crest cells from human pluripotent stem cells. *Am. J. Stem Cells* **2**, 119–131 (2013).

42. Thomson, J. A. *et al.* Embryonic stem cell lines derived from human blastocysts. *Science* **282**, 1145–1147 (1998).

43. Whalen, S. *et al.* Machine learning dissection of human accelerated regions in primate neurodevelopment. *Neuron* **111**, 857–873.e8 (2023).

44. Horlbeck, M. A. *et al.* Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation. *Elife* **5**, (2016).

45. Blomen, V. A. *et al.* Gene essentiality and synthetic lethality in haploid human cells. *Science* **350**, 1092–1096 (2015).

46. Hart, T., Brown, K. R., Sircoulomb, F., Rottapel, R. & Moffat, J. Measuring error rates in genomic perturbation screens: gold standards for human functional genomics. *Mol. Syst. Biol.* **10**, 733 (2014).

47. Li, W. *et al.* MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol.* **15**, 554 (2014).

48. Cahan, P. & Daley, G. Q. Origins and implications of pluripotent stem cell variability and

heterogeneity. *Nat. Rev. Mol. Cell Biol.* **14**, 357–368 (2013).

49. Kilpinen, H. *et al.* Common genetic variation drives molecular heterogeneity in human iPSCs. *Nature* **546**, 370–375 (2017).
50. Prado-Martinez, J. *et al.* Great ape genetic diversity and population history. *Nature* **499**, 471–475 (2013).
51. Baker, D. E. C. *et al.* Adaptation to culture of human embryonic stem cells and oncogenesis in vivo. *Nat. Biotechnol.* **25**, 207–215 (2007).
52. Merkle, F. T. *et al.* Human pluripotent stem cells recurrently acquire and expand dominant negative P53 mutations. *Nature* **545**, 229–233 (2017).
53. Serin Harmanci, A., Harmanci, A. O. & Zhou, X. CaSpER identifies and visualizes CNV events by integrative analysis of single-cell or bulk RNA-sequencing data. *Nat. Commun.* **11**, 89 (2020).
54. Setoguchi, K., TeSlaa, T., Koehler, C. M. & Teitell, M. A. P53 Regulates Rapid Apoptosis in Human Pluripotent Stem Cells. *J. Mol. Biol.* **428**, 1465–1475 (2016).
55. Vassilev, L. T. *et al.* In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. *Science* **303**, 844–848 (2004).
56. Dennis, M. Y. & Eichler, E. E. Human adaptation and evolution by segmental duplication. *Curr. Opin. Genet. Dev.* **41**, 44–52 (2016).
57. Pontis, J. *et al.* Hominoid-Specific Transposable Elements and KZFPs Facilitate Human Embryonic Genome Activation and Control Transcription in Naive Human ESCs. *Cell Stem Cell* **24**, 724–735.e5 (2019).
58. McKenna, A. & Shendure, J. FlashFry: a fast and flexible tool for large-scale CRISPR target design. *BMC Biol.* **16**, 74 (2018).
59. 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).
60. Evans, P. D. *et al.* Adaptive evolution of ASPM, a major determinant of cerebral cortical size

in humans. *Hum. Mol. Genet.* **13**, 489–494 (2004).

61. Kouprina, N. *et al.* Accelerated evolution of the ASPM gene controlling brain size begins prior to human brain expansion. *PLoS Biol.* **2**, E126 (2004).
62. Montgomery, S. H., Capellini, I., Venditti, C., Barton, R. A. & Mundy, N. I. Adaptive evolution of four microcephaly genes and the evolution of brain size in anthropoid primates. *Mol. Biol. Evol.* **28**, 625–638 (2011).
63. Zhang, J. Evolution of the human ASPM gene, a major determinant of brain size. *Genetics* **165**, 2063–2070 (2003).
64. Jiang, K. *et al.* Microtubule minus-end regulation at spindle poles by an ASPM-katanin complex. *Nat. Cell Biol.* **19**, 480–492 (2017).
65. Kuhlwilm, M. & Boeckx, C. A catalog of single nucleotide changes distinguishing modern humans from archaic hominins. *Sci. Rep.* **9**, 8463 (2019).
66. Dumas, G., Malesys, S. & Bourgeron, T. Systematic detection of brain protein-coding genes under positive selection during primate evolution and their roles in cognition. *Genome Res.* **31**, 484–496 (2021).
67. Gross, A. BCL-2 family proteins as regulators of mitochondria metabolism. *Biochim. Biophys. Acta* **1857**, 1243–1246 (2016).
68. Liu, X., Li, Y. I. & Pritchard, J. K. Trans Effects on Gene Expression Can Drive Omnigenic Inheritance. *Cell* **177**, 1022–1034.e6 (2019).
69. Kim, E. *et al.* Dynamic rewiring of biological activity across genotype and lineage revealed by context-dependent functional interactions. *Genome Biol.* **23**, 140 (2022).
70. Wainberg, M. *et al.* A genome-wide atlas of co-essential modules assigns function to uncharacterized genes. *Nat. Genet.* **53**, 638–649 (2021).
71. Szklarczyk, D. *et al.* The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* **49**, D605–D612 (2021).

72. Wang, L., Wu, D., Robinson, C. V., Wu, H. & Fu, T.-M. Structures of a Complete Human V-ATPase Reveal Mechanisms of Its Assembly. *Mol. Cell* **80**, 501–511.e3 (2020).

73. Jewell, J. L., Russell, R. C. & Guan, K.-L. Amino acid signalling upstream of mTOR. *Nat. Rev. Mol. Cell Biol.* **14**, 133–139 (2013).

74. Zhou, J. *et al.* mTOR supports long-term self-renewal and suppresses mesoderm and endoderm activities of human embryonic stem cells. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 7840–7845 (2009).

75. Zoncu, R. *et al.* mTORC1 senses lysosomal amino acids through an inside-out mechanism that requires the vacuolar H(+)-ATPase. *Science* **334**, 678–683 (2011).

76. Becker, K. A. *et al.* Self-renewal of human embryonic stem cells is supported by a shortened G1 cell cycle phase. *J. Cell. Physiol.* **209**, 883–893 (2006).

77. Sakaue-Sawano, A. *et al.* Genetically Encoded Tools for Optical Dissection of the Mammalian Cell Cycle. *Mol. Cell* **68**, 626–640.e5 (2017).

78. Hochegger, H., Takeda, S. & Hunt, T. Cyclin-dependent kinases and cell-cycle transitions: does one fit all? *Nat. Rev. Mol. Cell Biol.* **9**, 910–916 (2008).

79. Hwang, H. C. & Clurman, B. E. Cyclin E in normal and neoplastic cell cycles. *Oncogene* **24**, 2776–2786 (2005).

80. Berthet, C., Aleem, E., Coppola, V., Tessarollo, L. & Kaldis, P. Cdk2 knockout mice are viable. *Curr. Biol.* **13**, 1775–1785 (2003).

81. Ortega, S. *et al.* Cyclin-dependent kinase 2 is essential for meiosis but not for mitotic cell division in mice. *Nat. Genet.* **35**, 25–31 (2003).

82. Aleem, E., Kiyokawa, H. & Kaldis, P. Cdc2-cyclin E complexes regulate the G1/S phase transition. *Nat. Cell Biol.* **7**, 831–836 (2005).

83. Satyanarayana, A. & Kaldis, P. Mammalian cell-cycle regulation: several Cdks, numerous cyclins and diverse compensatory mechanisms. *Oncogene* **28**, 2925–2939 (2009).

84. Rubin, S. M., Sage, J. & Skotheim, J. M. Integrating Old and New Paradigms of G1/S Control.

Mol. Cell **80**, 183–192 (2020).

85. Shirodkar, S. *et al.* The transcription factor E2F interacts with the retinoblastoma product and a p107-cyclin A complex in a cell cycle-regulated manner. *Cell* **68**, 157–166 (1992).
86. Zatulovskiy, E., Zhang, S., Berenson, D. F., Topacio, B. R. & Skotheim, J. M. Cell growth dilutes the cell cycle inhibitor Rb to trigger cell division. *Science* **369**, 466–471 (2020).
87. Liu, L., Michowski, W., Kolodziejczyk, A. & Sicinski, P. The cell cycle in stem cell proliferation, pluripotency and differentiation. *Nat. Cell Biol.* **21**, 1060–1067 (2019).
88. Steenpass, L. Generation of two H1 hESC sublines carrying a heterozygous and homozygous knock-out of RB1. *Stem Cell Res.* **25**, 270–273 (2017).
89. Döpper, H. *et al.* Biallelic and monoallelic deletion of the RB1 promoter in six isogenic clonal H9 hESC lines. *Stem Cell Res.* **45**, 101779 (2020).
90. Zhu, L., Harlow, E. & Dynlacht, B. D. p107 uses a p21CIP1-related domain to bind cyclin/cdk2 and regulate interactions with E2F. *Genes Dev.* **9**, 1740–1752 (1995).
91. Fan, L. *et al.* FAM122A, a new endogenous inhibitor of protein phosphatase 2A. *Oncotarget* **7**, 63887–63900 (2016).
92. Mochida, S., Ikeo, S., Gannon, J. & Hunt, T. Regulated activity of PP2A-B55 delta is crucial for controlling entry into and exit from mitosis in Xenopus egg extracts. *EMBO J.* **28**, 2777–2785 (2009).
93. Hirai, H. *et al.* Small-molecule inhibition of Wee1 kinase by MK-1775 selectively sensitizes p53-deficient tumor cells to DNA-damaging agents. *Mol. Cancer Ther.* **8**, 2992–3000 (2009).
94. King, C. *et al.* LY2606368 Causes Replication Catastrophe and Antitumor Effects through CHK1-Dependent Mechanisms. *Mol. Cancer Ther.* **14**, 2004–2013 (2015).
95. Aladjem, M. I. *et al.* ES cells do not activate p53-dependent stress responses and undergo p53-independent apoptosis in response to DNA damage. *Curr. Biol.* **8**, 145–155 (1998).
96. Chuykin, I. A., Lianguzova, M. S., Pospelova, T. V. & Pospelov, V. A. Activation of DNA damage response signaling in mouse embryonic stem cells. *Cell Cycle* **7**, 2922–2928 (2008).

97. Turinetto, V. *et al.* High basal γH2AX levels sustain self-renewal of mouse embryonic and induced pluripotent stem cells. *Stem Cells* **30**, 1414–1423 (2012).

98. Ahuja, A. K. *et al.* A short G1 phase imposes constitutive replication stress and fork remodelling in mouse embryonic stem cells. *Nat. Commun.* **7**, 10660 (2016).

99. Dehay, C. & Kennedy, H. Cell-cycle control and cortical development. *Nat. Rev. Neurosci.* **8**, 438–450 (2007).

100. Lukaszewicz, A. *et al.* G1 phase regulation, area-specific cell cycle control, and cytoarchitectonics in the primate cortex. *Neuron* **47**, 353–364 (2005).

101. Pilaz, L.-J. *et al.* Forced G1-phase reduction alters mode of division, neuron number, and laminar phenotype in the cerebral cortex. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 21924–21929 (2009).

102. Taverna, E., Götz, M. & Huttner, W. B. The cell biology of neurogenesis: toward an understanding of the development and evolution of the neocortex. *Annu. Rev. Cell Dev. Biol.* **30**, 465–502 (2014).

103. Calegari, F., Haubensak, W., Haffner, C. & Huttner, W. B. Selective lengthening of the cell cycle in the neurogenic subpopulation of neural progenitor cells during mouse brain development. *J. Neurosci.* **25**, 6533–6538 (2005).

104. Lange, C., Huttner, W. B. & Calegari, F. Cdk4/cyclinD1 overexpression in neural stem cells shortens G1, delays neurogenesis, and promotes the generation and expansion of basal progenitors. *Cell Stem Cell* **5**, 320–331 (2009).

105. Lim, S. & Kaldis, P. Loss of Cdk2 and Cdk4 induces a switch from proliferation to differentiation in neural stem cells. *Stem Cells* **30**, 1509–1520 (2012).

106. Nonaka-Kinoshita, M. *et al.* Regulation of cerebral cortex size and folding by expansion of basal progenitors. *EMBO J.* **32**, 1817–1828 (2013).

107. Field, A. R. *et al.* Structurally Conserved Primate LncRNAs Are Transiently Expressed during Human Cortical Differentiation and Influence Cell-Type-Specific Genes. *Stem Cell Reports*

12, 245–257 (2019).

108. Langergraber, K. E. *et al.* Generation times in wild chimpanzees and gorillas suggest earlier divergence times in great ape and human evolution. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 15716–15721 (2012).
109. Glazko, G. V. & Nei, M. Estimation of divergence times for major lineages of primate species. *Mol. Biol. Evol.* **20**, 424–434 (2003).
110. Mirzaa, G. *et al.* De novo CCND2 mutations leading to stabilization of cyclin D2 cause megalencephaly-polymicrogyria-polydactyly-hydrocephalus syndrome. *Nat. Genet.* **46**, 510–515 (2014).
111. Pirozzi, F., Nelson, B. & Mirzaa, G. From microcephaly to megalencephaly: determinants of brain size. *Dialogues Clin. Neurosci.* **20**, 267–282 (2018).
112. Wang, M. *et al.* Increased Neural Progenitor Proliferation in a hiPSC Model of Autism Induces Replication Stress-Associated Genome Instability. *Cell Stem Cell* **26**, 221–233.e6 (2020).
113. Yoshida, H. *et al.* Apaf1 is required for mitochondrial pathways of apoptosis and brain development. *Cell* **94**, 739–750 (1998).
114. Halfon, M. S. Perspectives on Gene Regulatory Network Evolution. *Trends Genet.* **33**, 436–447 (2017).
115. True, J. R. & Haag, E. S. Developmental system drift and flexibility in evolutionary trajectories. *Evol. Dev.* **3**, 109–119 (2001).
116. Farkas, Z. *et al.* Gene loss and compensatory evolution promotes the emergence of morphological novelties in budding yeast. *Nat Ecol Evol* **6**, 763–773 (2022).
117. Cancer Genome Atlas Research Network *et al.* The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.* **45**, 1113–1120 (2013).
118. Adamson, B. *et al.* A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. *Cell* **167**, 1867–1882.e21 (2016).

119. Dixit, A. *et al.* Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell* **167**, 1853–1866.e17 (2016).

120. Feldman, D. *et al.* Optical Pooled Screens in Human Cells. *Cell* **179**, 787–799.e17 (2019).

121. Nolbrant, S., Heuer, A., Parmar, M. & Kirkeby, A. Generation of high-purity human ventral midbrain dopaminergic progenitors for in vitro maturation and intracerebral transplantation. *Nat. Protoc.* **12**, 1962–1979 (2017).

122. Chen, Y. *et al.* A versatile polypharmacology platform promotes cytoprotection and viability of human pluripotent and differentiated cells. *Nat. Methods* **18**, 528–541 (2021).

123. Kadoshima, T. *et al.* Self-organization of axial polarity, inside-out layer pattern, and species-specific progenitor dynamics in human ES cell-derived neocortex. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 20284–20289 (2013).

124. Smith, J. R. *et al.* Robust, persistent transgene expression in human embryonic stem cells is achieved with AAVS1-targeted integration. *Stem Cells* **26**, 496–504 (2008).

125. Mandegar, M. A. *et al.* CRISPR Interference Efficiently Induces Specific and Reversible Gene Silencing in Human iPSCs. *Cell Stem Cell* **18**, 541–553 (2016).

126. Replogle, J. M. *et al.* Maximizing CRISPRi efficacy and accessibility with dual-sgRNA libraries and optimal effectors. *Elife* **11**, (2022).

127. Wang, D. *et al.* Optimized CRISPR guide RNA design for two high-fidelity Cas9 variants by deep learning. *Nat. Commun.* **10**, 4284 (2019).

128. Jost, M. *et al.* Titrating gene expression using libraries of systematically attenuated CRISPR guide RNAs. *Nat. Biotechnol.* **38**, 355–364 (2020).

129. Kronenberg, Z. N. *et al.* High-resolution comparative analysis of great ape genomes. *Science* **360**, (2018).

130. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

131. Tweedie, S. *et al.* Genenames.org: the HGNC and VGNC resources in 2021. *Nucleic Acids*

Res. **49**, D939–D946 (2021).

132. Mao, Y. *et al.* A high-quality bonobo genome refines the analysis of hominid evolution. *Nature* **594**, 77–81 (2021).
133. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
134. Dougherty, M. L. *et al.* Transcriptional fates of human-specific segmental duplications in brain. *Genome Res.* **28**, 1566–1576 (2018).
135. Wang, T. *et al.* Gene Essentiality Profiling Reveals Gene Networks and Synthetic Lethal Interactions with Oncogenic Ras. *Cell* **168**, 890–903.e15 (2017).
136. Frankish, A. *et al.* GENCODE 2021. *Nucleic Acids Res.* **49**, D916–D923 (2021).
137. Fiddes, I. T. *et al.* Comparative Annotation Toolkit (CAT)-simultaneous clade and personal genome annotation. *Genome Res.* **28**, 1029–1038 (2018).
138. Cock, P. J. A. *et al.* Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**, 1422–1423 (2009).
139. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
140. Yang, Z. & Nielsen, R. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol. Biol. Evol.* **17**, 32–43 (2000).
141. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12 (2011).
142. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).
143. Pertea, G. & Pertea, M. GFF Utilities: GffRead and GffCompare. *F1000Res.* **9**, (2020).
144. Schneider, V. A. *et al.* Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. *Genome Res.* **27**, 849–864 (2017).

145. Soneson, C., Love, M. I. & Robinson, M. D. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res.* **4**, 1521 (2015).

146. Stirling, D. R. *et al.* CellProfiler 4: improvements in speed, utility and usability. *BMC Bioinformatics* **22**, 433 (2021).

Chapter IV

Conclusions and Outlook

Genetic studies of human evolution provide insights into the origins of humans[1], the processes that underlie recent human adaptations[2], and the genetic variants that are specific to the human lineage[3]. Distinguishing neutral from non-neutral human-specific genetic variants using analytical and experimental approaches is central to the future of these investigations. However, advantageous variants are rare[4,5], many adaptive processes have a polygenic basis[6], and variants in *cis*-regulatory elements[7], a favored evolutionary substrate[8], may have detectable effects only within precise spatiotemporal domains[9,10], complicating the assignment of variants to functions. Advances in long-read sequencing of primate genomes, deep learning, nonhuman primate cellular models, and CRISPR-Cas systems are critical for addressing these challenges.

Whole-genome sequencing across primates[11], including Telomere-to-Telomere (T2T) great ape genomes, enables cataloging the millions of genetic variants have accumulated in the human lineage since human, chimpanzee, and bonobo species diverged from a common ancestor[12]. The improved contiguity of T2T great ape genomes will refine human-specific single-nucleotide and structural variant identification and the annotation of repeat-rich sequences. As nearly 90% of primate-specific candidate *cis*-regulatory elements intersect transposable elements[13], high-resolution annotation of these features will be essential for characterizing lineage-specific *cis*-regulatory evolution. For variants that are unobserved in human populations, identifying evolutionarily constrained sequences across primates provides a complementary approach for predicting variants that are likely to be disease-causing in humans[11].

The revolution in functional genomics that has occurred since the repurposing of CRISPR-Cas systems for programmable genetic and epigenetic editing[14,15] is paramount to the interrogation of human-specific genetic variants identified by genome sequencing. In Chapter II, I describe a

CRISPRi-based approach for investigating the ancestral function of human-specific deletions using nonhuman primate cellular models and primary cells. I consider this work to be a very early step towards the functional interrogation of variants lost, gained, inverted, duplicated, or otherwise altered in the human lineage in their native genomic contexts. These efforts will be greatly aided by the improved annotation of the epigenetic state of chromatin across cell types and tissues in nonhuman primates, the establishment of experimental systems that capture human and nonhuman primate cells differentiating in the presence of CRISPR-based genetic perturbations in a common environment, and the mutagenesis of human-specific single-nucleotide variants by base-editing screens[16]. Assigning cellular and molecular functions to all classes of human-specific variants may provide the foundation for a multi-level trait hierarchy tree[17] unifying future investigations.

REFERENCES

1. Nielsen, R. *et al.* Tracing the peopling of the world through genomics. *Nature* **541**, 302–310 (2017).
2. Fan, S., Hansen, M. E. B., Lo, Y. & Tishkoff, S. A. Going global by adapting local: A review of recent human adaptation. *Science* **354**, 54–59 (2016).
3. Kronenberg, Z. N. *et al.* High-resolution comparative analysis of great ape genomes. *Science* **360**, (2018).
4. Eyre-Walker, A. & Keightley, P. D. The distribution of fitness effects of new mutations. *Nat. Rev. Genet.* **8**, 610–618 (2007).
5. Orr, H. A. Fitness and its role in evolutionary genetics. *Nat. Rev. Genet.* **10**, 531–539 (2009).
6. Barghi, N., Hermisson, J. & Schlötterer, C. Polygenic adaptation: a unifying framework to understand positive selection. *Nat. Rev. Genet.* **21**, 769–781 (2020).
7. King, M. C. & Wilson, A. C. Evolution at two levels in humans and chimpanzees. *Science* **188**, 107–116 (1975).
8. Carroll, S. B. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* **134**, 25–36 (2008).
9. Levine, M. Transcriptional enhancers in animal development and evolution. *Curr. Biol.* **20**, R754–63 (2010).
10. Long, H. K., Prescott, S. L. & Wysocka, J. Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. *Cell* **167**, 1170–1187 (2016).
11. Gao, H. *et al.* The landscape of tolerated genetic variation in humans and primates. *Science* **380**, eabn8153 (2023).
12. Chimpanzee Sequencing and Analysis Consortium. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437**, 69–87 (2005).
13. Andrews, G. *et al.* Mammalian evolution of human cis-regulatory elements and transcription

factor binding sites. *Science* **380**, eabn7930 (2023).

14. Knott, G. J. & Doudna, J. A. CRISPR-Cas guides the future of genetic engineering. *Science* **361**, 866–869 (2018).
15. 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).
16. Schmidt, R. *et al.* Base-editing mutagenesis maps alleles to tune human T cell functions. *Nature* (2023) doi:10.1038/s41586-023-06835-6.
17. Fair, T. & Pollen, A. A. Genetic architecture of human brain evolution. *Curr. Opin. Neurobiol.* **80**, 102710 (2023).

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

DocuSigned by:
Tyler Fair
AF25821C7D724DE...
Author Signature

11/30/2023
Date