Cell type-specific *cis*-regulatory divergence in gene expression and chromatin accessibility revealed by human-chimpanzee hybrid cells

Ban Wang[1,*], Alexander L. Starr[1,*], Hunter B. Fraser[1,2]

*Indicates equal contribution, these authors may interchange their names on their CV or resume

1. Department of Biology, Stanford University, Stanford, CA, USA

2. Correspondence: hbfraser@stanford.edu

Abstract:

Although gene expression divergence has long been postulated to be the primary driver of human evolution, identifying the genes and genetic variants underlying uniquely human traits has proven to be quite challenging. Theory suggests that cell type-specific *cis*-regulatory variants may fuel evolutionary adaptation due to the specificity of their effects. These variants can precisely tune the expression of a single gene in a single cell type, avoiding the potentially deleterious consequences of *trans*-acting changes and non-cell type-specific changes that can impact many genes and cell types, respectively. It has recently become possible to quantify human-specific *cis*-acting regulatory divergence by measuring allele-specific expression in human-chimpanzee hybrid cells—the product of fusing induced pluripotent stem (iPS) cells of each species *in vitro*. However, these *cis*-regulatory changes have only been explored in a limited number of cell types. Here, we quantify human-chimpanzee *cis*-regulatory divergence in gene expression and chromatin accessibility across six cell types, enabling the identification of highly cell type-specific *cis*-regulatory changes. We find that cell type-specific genes and regulatory elements evolve faster than those shared across cell types, suggesting an important role for genes with cell type-

specific expression in human evolution. Furthermore, we identify several instances of lineage-specific natural selection that may have played key roles in specific cell types, such as coordinated changes in the *cis*-regulation of dozens of genes involved in neuronal firing in motor neurons. Finally, using novel metrics and a machine learning model, we identify genetic variants that likely alter chromatin accessibility and transcription factor binding, leading to neuron-specific changes in the expression of the neurodevelopmentally important genes *FABP7* and *GAD1*. Overall, our results demonstrate that integrative analysis of *cis*-regulatory divergence in chromatin accessibility and gene expression across cell types is a promising approach to identify the specific genes and genetic variants that make us human.

Introduction:

In the past few million years, humans have evolved a multitude of unique phenotypes (Shave et al., 2019; Vanderhaeghen & Polleux, 2023). For example, our cardiovascular system has evolved to enable extended periods of physical exertion and the unique aspects of our nervous system enable human language and toolmaking (Shave et al., 2019; Vanderhaeghen & Polleux, 2023). Previous research suggests that much of human adaptation may be caused by changes in gene expression (Fraser, 2013; King & Wilson, 1975; Reilly & Noonan, 2016; Romero et al., 2012). To catalog these changes, studies have compared gene expression in post-mortem tissues of humans and our closest living relatives, chimpanzees (Blake et al., 2020; Kelley & Gilad, 2020; Ma et al., 2022). Although thousands of differentially expressed genes have been identified in post-mortem samples, it is generally not possible to disentangle the effects of genetic differences from the effects of confounding factors such as differences in diet, environment, cell type abundances, age, post-mortem interval, etc. In addition, for many traits the relevant gene expression differences may be specific to early development, but it is impossible to study fetal development *in vivo* in non-human great apes due to both ethical and technical

49 difficulties. To circumvent these issues, several groups have used great ape iPS cells to study differences
50 in gene expression in cell types present in early development (Benito-Kwiecinski et al., 2021; Field et al.,
51 2019; Kanton et al., 2019; Pavlovic et al., 2018). While the use of iPS cells addresses many of the
52 confounding factors present in post-mortem comparisons, they also introduce new issues such as
53 interspecies differences in iPS cell differentiation kinetics, efficiency, and maturation. Overall, it remains
54 tremendously challenging to identify human-specific changes in gene expression, which limits our ability
55 to link expression differences to either phenotypic differences or natural selection in the human lineage.
56
57 One particularly powerful means of studying the evolution of gene expression is through the
58 measurement of allele specific expression (ASE) in hybrids between two species (Combs et al., 2018;
59 Fraser, 2011; Hu et al., 2022; Mack & Nachman, 2017; Wittkopp & Kalay, 2012). This approach has the
60 advantage of eliminating many confounding factors inherent to interspecies comparisons, including
61 differences in cell type composition, environmental factors, developmental stage, and response to
62 differentiation protocols. Because the trans-acting environments of the two alleles in a hybrid are
63 identical, ASE has the additional benefit of reflecting only *cis*-regulatory changes, which are thought to
64 be less pleiotropic and more likely to drive evolutionary adaptation than broader *trans*-acting changes
65 (Agoglia et al., 2021; Gokhman et al., 2021; Prud'homme et al., 2007; Wittkopp & Kalay, 2012).
66 Furthermore, ASE enables the use of powerful methods that can detect lineage-specific natural selection
67 and, as a result, contribute to our understanding of the selective pressures that have shaped the
68 evolution of a wide variety of species (Fraser, 2011). Until recently, it has not been possible to
69 disentangle *cis*- and *trans*-acting changes fixed in the human lineage since humans cannot hybridize with
70 any other species. However, the development of human-chimpanzee hybrid iPS cells via *in vitro* cell
71 fusion enables measurement of ASE in a wide variety of tissues and developmental contexts (Agoglia et
72 al., 2021; Gokhman et al., 2021; Song et al., 2021). This provides an effective platform to investigate

73 general principles of hominid gene expression evolution, detect lineage-specific selection, and identify
74 candidate gene expression changes underlying human-specific traits.
75
76 While gene expression divergence between humans and chimpanzees is well-studied, there has been
77 less focus on epigenetic differences, many of which are likely to underlie divergent gene expression
78 (García-Pérez et al., 2021; Kozlenkov et al., 2020; Netherlands Brain Bank et al., 2016; Trizzino et al.,
79 2017). Furthermore, these studies, regardless of whether they utilize post-mortem tissues or cell lines,
80 are subject to the same confounding factors mentioned above. Analogous to ASE, one can use the assay
81 for transposase accessible chromatin using sequencing (ATAC-seq) in interspecies hybrids to measure
82 allele-specific chromatin accessibility (ASCA) (Buenrostro et al., 2013; Corces et al., 2017; Liang et al.,
83 2021; S. Zhang et al., 2020). As with ASE, ASCA is unaffected by many confounders inherent to between-
84 species comparisons and only measures *cis*-regulatory divergence. Perhaps most importantly, ASCA can
85 implicate specific regulatory elements that likely underlie gene expression differences. These regulatory
86 elements can then be more closely studied to identify the likely causal genetic variants and the
87 molecular mechanisms by which those variants alter gene expression.
88
89 Here, we generated RNA-seq and ATAC-seq data from six cell types, derived from human-chimpanzee
90 hybrid iPS cells, and quantified ASE and ASCA. Using this dataset, we identified thousands of genes and
91 *cis*-regulatory elements showing cell type-specific ASE and ASCA. We found that cell type-specific genes
92 and *cis*-regulatory elements are more likely to have divergent expression and accessibility than their
93 more broadly expressed/accessible counterparts. Furthermore, we provide evidence for polygenic
94 selection on the expression level of genes associated with physiologically relevant gene sets including
95 sodium channels and syntaxin-binding proteins in motor neurons. Finally, we use newly developed
96 metrics and machine learning algorithms to link cell type-specific differences in chromatin accessibility

97 and gene expression and identify putative causal mutations underlying these differences. Using this
98 pipeline we identified motor neuron-specific increases in promoter chromatin accessibility and gene
99 expression for *FABP7*, which plays a key role in neurodevelopment but is not well-studied in neurons. In
100 addition, we focus on a human-accelerated region (HAR) near the promoter of *GAD1*. While this region
101 is accessible in all cell types, both the accessibility of the HAR and the expression of *GAD1* are only
102 chimpanzee-biased in motor neurons. Analysis of scRNA-seq from human and chimpanzee brain
103 organoids showed that increased expression of *GAD1* also occurs in ventral forebrain inhibitory neurons.
104 Overall, this study provides insight into the evolution of gene expression in hominids as well as a
105 resource that will inform functional genomic dissection of human-specific traits.
106
107 **Results**
108
109 **Cis-regulatory divergence of gene expression in six cell types is largely cell type-specific or shared**
110 **across all cell types**
111
112 To measure genome-wide cis-regulatory divergence in gene expression, we performed RNA-seq on six
113 cell types derived from human-chimpanzee hybrid iPS cells (Fig. 1a). The cell types profiled were from six
114 diverse developmental lineages including the motor neuron (MN), cardiomyocyte (CM), hepatocyte
115 progenitor (HP), pancreatic progenitor (PP), skeletal myocyte (SKM), and retinal pigment epithelium
116 (RPE) lineages. These represent all three germ layers and a variety of organs (Fig. 1a). It is worth noting
117 that these differentiations do not necessarily lead to a pure population of cells, but rather a population
118 of cells with different levels of maturity along a particular developmental lineage. For example, the SKM
119 population likely contains fully differentiated muscle fibers as well as a small population of proliferating
120 satellite cells; for clarity we refer to this as the SKM cell type. As these different cell types are not shared

121 between tissues, we use cell type-specific and tissue-specific interchangeably throughout the
122 manuscript.
123
124 Two independently generated hybrid lines were differentiated for each cell type and at least two
125 biological replicates per hybrid line per cell type were collected (see Methods). Each cell type was
126 sequenced to an average depth of 134 million paired-end reads (Supp Fig. 1). We used a computational
127 pipeline to quantify ASE adapted from the pipeline introduced by Agoglia et al (Agoglia et al., 2021).
128 Briefly, we computed ASE by mapping reads to both the human and chimpanzee genomes, correcting
129 for mapping bias, and assigning reads to the human or chimpanzee genome if a read contained one or
130 more human-chimpanzee single nucleotide differences (see Methods).
131
132 As expected, the samples clustered predominantly by cell type (Fig. 1b-c). Within four of the six cell type
133 clusters, individual samples clustered by line rather than species of origin, potentially indicating line to
134 line variability in differentiation (Fig. 1b). This highlights the importance of measuring ASE which, by
135 definition, is measured within each line and so is robust to variability between lines. Indeed, when
136 performing PCA within cell types using allelic counts (i.e. counting reads from the human allele and
137 chimpanzee allele separately), human and chimpanzee species differences were clearly separated by
138 principal component (PC) 1 or PC2 in each cell type (Fig. 1d, Supp Fig. 2). To assess the success of our
139 differentiations, we examined each cell type for the expression of known marker genes in our RNA-seq
140 data (Fig. 1e, Supp Fig. 3). All cell types express canonical marker genes and do not express pluripotency
141 markers, indicating that the differentiations were successful (Fig. 1e, Supp Fig 3, Additional file 1)
142 (Burridge et al., 2014; Chal et al., 2016; Korytnikov & Nostro, 2016; Mallanna & Duncan, 2013; Maury et
143 al., 2015; Sharma et al., 2019).
144

145 Because our hybrid cells were grown concurrently with their human and chimpanzee diploid "parental"
146 cells, we performed an additional check for purity of the hybrid lines by quantifying genome-wide ASE.
147 We noticed that among our 25 RNA-seq samples, the two PP hybrid2 samples had a slight bias towards
148 higher expression from the chimpanzee alleles across all chromosomes. This is likely due to a small
149 fraction of contaminating chimpanzee cells in these samples. We corrected for this by reducing the
150 chimpanzee allele counts such that the number of reads assigned to the human and chimpanzee alleles
151 was equal. By simulating contamination of a hybrid sample with chimpanzee cells, we found that this
152 correction was conservative and that the log fold-change estimates were largely unaffected by
153 contamination after this correction (Methods, Supp Fig. 4).
154
155 We next investigated which genes were differentially expressed between the human and chimpanzee
156 allele in each cell type (Fig. 2a). We identified thousands of genes showing significantly biased ASE in
157 each cell type at a false discovery rate (FDR) cutoff of 0.05 (Fig. 2b). We detected a comparable number
158 of ASE genes in all cell types except SKM. As a result, we repeated all subsequent analyses both including
159 and excluding SKM and obtained qualitatively similar results regardless of whether SKM was included.
160
161 While a considerable number of genes had significant ASE in all cell types, many more genes only had
162 significant ASE in a single cell type, suggesting cell type-specific *cis*-regulatory divergence (Fig. 2b). A
163 notable family of developmentally important genes that exemplifies differences in ASE across cell types
164 is the neurotrophins and their receptors (Fig. 2c) (Caporali & Emanueli, 2009; Huang & Reichardt, 2001).
165 For example, *NTRK3*, which plays a key role in the development of the nervous system, is only
166 differentially expressed in RPE and MN but is chimpanzee-biased in RPE and human-biased in MN (Supp
167 Fig. 5) (Ichim et al., 2012; Naito et al., 2017). In addition, the gene coding for its primary ligand (*NTF3*) is
168 differentially expressed in a variety of cell types yet is human-biased in all cell types except MN in which

169 it is chimpanzee-biased (Supp Fig. 5). *NTRK1* differential expression is similarly tissue-specific as it is
170 strongly chimpanzee-biased in MN, but human-biased in CM and SKM (Supp Fig. 5). These results
171 indicate that the regulatory landscape of these genes has undergone many complex *cis*-regulatory
172 changes as the human and chimpanzee lineages have diverged.
173
174 To further investigate the relationship between tissue-specificity and ASE, we asked whether genes with
175 variable expression across tissues are more likely to show ASE. Using a standard definition of cell type-
176 specific genes—those with detectable expression in only one cell type in our study—we found that cell
177 type-specific genes were typically enriched for ASE in the one cell type where they are expressed (Supp
178 Fig. 6) (GTEx Consortium, 2017; Jain & Tuteja, 2019). However, other cell type-specific expression
179 patterns such as uniquely low expression in a particular cell type may also indicate an important dosage-
180 sensitive function in that cell type. We therefore focused on a broader definition of cell type-specificity
181 in which genes that are differentially expressed between one cell type and all others in our study (FDR <
182 0.05 for each pairwise comparison) are considered cell type-specific for that cell type. We found that
183 this more inclusive definition, which identified many more cell type-specific genes, showed an even
184 more significant ASE enrichment than the narrower definition (Fig. 2d). This result is not sensitive to the
185 choice of FDR cutoff (Supp Fig. 7) nor driven solely by a subgroup of highly expressed genes (Supp Fig. 8).
186 This trend is also robust to separating samples into two groups and using one to define cell type-specific
187 genes and the other to identify differentially expressed genes. This controls for spurious relationships
188 that can result when the same data are used to define two different quantities which are then
189 compared (Supp Fig. 9) (Fraser, 2019).
190
191 This enrichment (Fig. 2d) suggests that tissue-specific genes may have less constraint and/or more
192 frequent positive selection on their expression. We reasoned that if the trend was solely driven by

193 constraint, then controlling for constraint—even if imperfectly—would be expected to reduce the
194 strength of the relationship. To investigate this, we binned genes by their variance in ASE across a large
195 cohort of human samples which we have previously shown acts as a reasonable proxy for evolutionary
196 constraint on gene expression (Castel et al., 2020; Starr et al., 2023). Across cell types, we generally
197 observe significant enrichments in each bin and little difference in enrichment between bins, suggesting
198 that differences in constraint on expression of cell type-specific vs. ubiquitously expressed genes are not
199 solely responsible for our observations (Supp Fig. 10). Furthermore, we observe even stronger
200 enrichments using an alternative constraint metric, the probability of haploinsufficiency score (pHI)
201 likely due to the larger number of genes for which pHI can be calculated (Supp Fig. 11) (Collins et al.,
202 2022). Overall, our analysis suggests that differences in constraint are unlikely to fully explain these
203 trends, suggesting a potential role for positive selection.
204
205 **Lineage-specific selection has acted on tissue-specific gene expression divergence**
206
207 Next, we sought to use our RNA-seq data to identify instances of lineage-specific selection. In the
208 absence of positive selection, one would expect that an approximately equal number of genes in a
209 pathway would have human-biased vs. chimpanzee-biased ASE. Significant deviation from this
210 expectation (as determined by the binomial test) rejects the null hypothesis of neutral evolution, instead
211 providing evidence of lineage-specific selection on this pathway. Using our previously published
212 modification of this test that incorporates a tissue-specific measure of constraint on gene expression, we
213 detected several signals of lineage-specific selection, some of which were cell type-specific (Starr et al.,
214 2023, Additional file 2). Notably, the four most significant enrichments were specific to motor neurons
215 and cardiomyocytes and are highly relevant to those cell types (Fig. 2e; Additional file 2). In
216 cardiomyocytes, the top pathway was “PPAR signaling pathway” which plays a key role in the regulation

217 of heart morphology and lipid metabolism (Montaigne et al., 2021). For example, *NR1H3* (also known as
218 *LXRA*) is strongly upregulated in human cardiomyocytes as well as all other cell types (Supp Fig. 12a).
219 Furthermore, this upregulation appears to have occurred in the human lineage based on data from non-
220 hybrid cardiomyocytes as well as adult hearts (Supp Fig. 12b) (Blake et al., 2020; Pavlovic et al., 2018).
221 Hybrid cells are essential in determining that the human-specific upregulation of *NR1H3* in
222 cardiomyocytes has a strong genetic component as *NR1H3* expression is very responsive to diet and
223 other environmental factors (Wang & Tontonoz, 2018).
224
225 In motor neurons, multiple categories showed a strong bias toward higher chimpanzee expression
226 including "sodium ion transmembrane transport" and "syntaxin binding". The genes in these categories
227 are of fundamental importance to the function of motor neurons as sodium ion transporters control
228 excitability and syntaxin binding proteins control the release of neurotransmitters from synaptic vesicles
229 (Brose et al., 2019; Meisler et al., 2021). Interestingly, several key genes in these sets appear to have
230 human-chimpanzee differences in expression that extend beyond motor neurons to other neuronal
231 types. For example, *SCN1B, SCN2B,* and *SYT2* have chimpanzee-biased ASE in our MN data. In contrast,
232 these genes have been observed to have lower expression in human cortical glutamatergic neurons
233 compared to neurons from chimpanzees and rhesus macaques (Supp Fig. 13) (Kozlenkov et al., 2020).
234 We note that several other genes in these genes sets are not differentially expressed between humans
235 and chimpanzees in this cortical neuron dataset, emphasizing the importance of studying individual
236 neuron types (Kozlenkov et al., 2020). Overall, the strong bias in gene expression of sodium ion
237 transporters and syntaxin binding proteins we observe suggests lineage-specific selection that may have
238 altered the electrophysiological properties of human motor neurons.
239
240 **Patterns of allele-specific chromatin accessibility reveal divergent *cis*-regulatory elements**

241

242 While ASE provides insight into what gene expression changes might underlie phenotypic differences
243 between humans and chimpanzees, in the absence of additional data it is very difficult to prioritize
244 which specific mutations might cause expression divergence. To begin to fill this gap, we generated
245 ATAC-seq data from five of the six cell types (all except RPE), with each cell type sequenced to an
246 average depth of 184 million paired-end reads (Supp Fig. 14). ATAC-seq uses a hyperactive Tn5
247 transposase to cleave DNA that is not bound by nucleosomes to enrich for accessible chromatin (Fig. 3a),
248 a hallmark of active *cis*-regulatory elements (CREs) (Buenrostro et al., 2013; Corces et al., 2017). We
249 estimated ASCA for individual open chromatin peaks by mapping reads to both species' reference
250 genomes, correcting for mapping bias, generating a unified list of peaks across all samples, and then
251 counting reads supporting each allele in each peak (see Methods). After extensive filtering, we were left
252 with 73,360 peaks across the five cell types with many fewer retained in SKM due to lower sequencing
253 depth (Supp Fig. 15). As expected, most genes had only a few peaks assigned to them (Supp Fig. 16). Cell
254 types clustered well using PC1 and PC2 of the ATAC-seq data, except for the HP sample which clustered
255 closely with PP samples (Fig. 3b, Supp Fig. 17). However, performing PCA on just the HP and PP samples
256 clearly separates the two cell types (Supp Fig. 18). Within each cell type, species differences were clearly
257 separated by PC1 or PC2 (Supp Fig. 19). As an example of ASCA, the accessibility of the promoter of *CTSF*
258 was strongly chimpanzee-biased, mirroring the chimpanzee-biased ASE for this gene (Fig. 3c).

259

260 As a first step in analyzing the ATAC-seq data, we intersected the peaks we identified with the genomic
261 annotations of chromatin states. These fifteen categories, predicted across many tissues and cell types
262 by the chromHMM model (Vu & Ernst, 2022), include terms such as "TSS" (transcription start site) and
263 "enhancer". We then plotted the median of the absolute human-chimpanzee ASCA log fold-changes for
264 each chromatin state and cell type (Fig 3d). The TSS and promoter annotations were the least divergent

265 in their accessibility, whereas regions of heterochromatin were the most divergent (Fig. 3d). To explore
266 the relationship between interspecies differences in ASCA and ASE, we assigned peaks to the nearest
267 TSS and computed the Pearson correlation between the allelic log fold-change of chromatin accessibility
268 and expression within each cell type and chromatin state. As expected, TSS and promoter annotations
269 showed the strongest correlation between ASCA and ASE, and correlations were stronger when
270 including only differentially accessible peaks (Fig. 3e, Methods). Intriguingly however, ASCA of regions
271 annotated as heterochromatin, polycomb repressed, or quiescent were as strongly correlated with ASE
272 as elements identified as enhancers or DNase hypersensitivity sites (Fig. 3e). Notably this result is robust
273 to how peaks and chromHMM annotations were intersected (Supp Fig. 20a), as well as to removal of all
274 peaks even slightly overlapping TSS or promoter-related annotations (Supp Fig. 20b). It should be noted
275 that chromHMM annotations such as heterochromatin do not imply that a region is constitutively
276 heterochromatic, but instead reflect the most common chromatin state across a large compendium of
277 cell types (Vu & Ernst, 2022). Indeed, the fact that we are focusing on ATAC-seq peaks indicates that the
278 chromatin in these regions is accessible in at least one cell type in our study. This result suggests that
279 CREs that are heterochromatic in some cell types may be more prone to large changes in accessibility
280 during evolution (Fig 3d), with significant impacts on the cis-regulation of nearby genes (Fig 3e).
281 Throughout, we refer to CREs assigned to the chromHMM annotation heterochromatin as
282 "heterochromatin CREs".
283
284 To further investigate the intriguing relationship between heterochromatin ASCA and ASE, we asked
285 whether TSS proximity is an important factor. First, we removed all CREs annotated as promoters and
286 recomputed correlations separately for CREs within 30 kilobases of a TSS (proximal CREs) and those
287 greater than 30 kb away (distal CREs). Interestingly, differences in accessibility in proximal
288 heterochromatin CREs have weak correlations with ASE compared to proximal enhancers (Supp Fig.

289 21a). However, differences in accessibility in distal heterochromatin CREs were roughly as strongly
290 correlated with ASE as differences in accessibility in enhancer regions (Supp Fig. 21b).
291
292 Next, we partitioned genes into cell type-specifically and ubiquitously expressed and recomputed the
293 correlations. While the results for ubiquitously expressed genes mirrored our initial finding of a
294 relatively strong relationship between accessibility in heterochromatin CREs and gene expression,
295 differences in accessibility in heterochromatin CREs were less correlated with ASE for cell type-specific
296 genes than differences in accessibility in enhancers (Supp Fig. 22). Altogether, our analysis suggests that
297 large changes in accessibility of distal heterochromatin CREs may be particularly important in the cis-
298 regulatory evolution of more ubiquitously expressed genes.
299
300 Next, we investigated whether the analog of the relationship between cell type-specific gene expression
301 and ASE (Fig. 2d) holds for chromatin accessibility. Since the number of called peaks is largely dependent
302 on sequencing depth (Supp Fig. 14, 23a), we performed down-sampling to equalize power to detect
303 peaks across cell types (Supp Fig. 23b, Methods). We then called peaks on the down-sampled data and
304 identified peaks as cell type-specific if they were called as peaks in only one cell type. In agreement with
305 the gene expression data, we observed that cell type-specific peaks are enriched for ASCA across all cell
306 types and this enrichment generally holds when using varying $\log_2$ fold-change or p-value cutoffs (Supp
307 Figs. 24-25). Analogous to our analysis of gene expression, we also applied a broader definition of cell
308 type-specificity to the ATAC data, in which a peak was considered specific to a cell type if that peak had
309 an absolute $\log_2$ fold-change greater than a chosen threshold (e.g. 0.5 or 1) across all pairwise
310 comparisons with other cell types. We observe strong enrichment for ASCA in cell type-specific peaks
311 using this definition except when using the most stringent cutoffs due to the very low number of peaks
312 meeting these criteria (Supp Figs. 26-27, Methods). Notably, we observe the same enrichments when

313 controlling for a recently published metric for constraint on non-coding elements that compares the
314 observed vs. expected number of human polymorphisms (S. Chen et al., 2022), suggesting that
315 differences in evolutionary constraint may not be solely responsible for the observed trends (Supp Fig.
316 28).
317
318 Finally, it is possible that CREs and genes that are less conserved will have more SNPs, and therefore
319 more power to call ASCA and ASE, leading to systematically biased estimates. There is a weak positive
320 correlation between the number of SNPs and the $-\log_{10}$(FDR) for ASE and a weak negative or no
321 correlation for ASCA (Supp Fig. 29). Similarly, we observe a weak relationship between the number of
322 SNPs in CREs or genes and absolute log fold-change estimates (Supp Fig. 30). Although the relationship
323 between the number of SNPs and ASE/ASCA is weak, we confirmed that cell type-specific genes and
324 peaks are still strongly enriched for ASE and ASCA when stratifying by number of SNPs (Supp Fig. 31-32).
325 Overall, our analysis suggests that the result that more cell type-specific genes and CREs are more
326 evolutionarily diverged is robust to a variety of possible confounders.
327
328 We next explored the relationship between cell type-specific ASCA and ASE. To do this, we developed a
329 novel metric called differential expression enrichment (dEE) to quantify how specific the log fold-change
330 is to a particular cell type or tissue. Our method is based on expression enrichment (EE) (Yu et al., 2006),
331 a metric that measures how specific gene expression is to a certain cell type/tissue. dEE estimates how
332 cell type-specific ASE is for a gene (Supp Fig. 33, Methods) and, analogously, dCAE (differential
333 chromatin accessibility enrichment) measures how cell type-specific ASCA is for a *cis*-regulatory element
334 (Supp Fig. 34, Methods). dEE and dCAE are high in a cell type if there is a high absolute log fold-change in
335 that cell type and much lower absolute log-fold changes or log fold-changes in the opposite direction in
336 the other cell types. For example, dEE would be close to one for a gene in a cell type if the gene had

337 strongly human-biased ASE in that cell type and very weakly human-biased or chimpanzee-biased ASE in
338 the other cell types. On the other hand, if a gene did not have any strong allelic bias, that gene would
339 have dEE close to zero. dEE is conceptually related to a metric we have previously introduced, diffASE
340 (Combs et al., 2018; Hu et al., 2022; York et al., 2018), and generalizes diffASE to an arbitrary number of
341 cell types and any assay that produces log fold-changes. Using these metrics, we identified 154 instances
342 in which a gene with cell type-specific ASE (i.e., high dEE) had a peak with cell type-specific ASCA in the
343 same cell type (i.e., high dCAE, see Additional file 3 for the full list). Of these, 95 showed ASCA and ASE in
344 the same direction which is more than expected by chance (77 expected; $p < 0.005$, binomial test).
345 These results suggest that tissue-specific *cis*-regulatory divergence in chromatin accessibility may often
346 impact tissue-specific gene expression, though this divergence is neither necessary nor sufficient to do
347 so.
348
349 **Identifying candidate causal *cis*-regulatory variants by integrating ASE and ASCA across cell types**
350
351 As high dEE and dCAE in a given cell type might be indicative of a causal link between the change in
352 chromatin accessibility and the change in expression, we focused on the 95 peak-gene pairs with
353 matching direction and used two different strategies to identify examples to investigate in detail. First,
354 we prioritized genes known to play important roles in development. For example, we found that the
355 promoter of *FABP7* has human-biased ASCA specifically in motor neurons (Fig. 4a-b) and that the *FABP7*
356 gene has human-biased ASE in motor neurons (Fig 4c). *FABP7* is used as a marker of glial cells and neural
357 progenitor cells (NPCs) and plays a key role in NPC proliferation and astrocyte function (Arai et al., 2005;
358 De Rosa et al., 2012; Ebrahimi et al., 2016; Watanabe et al., 2007). Using previously published single-
359 nucleus RNA-seq data from humans, chimpanzees, and rhesus macaques, we confirmed that *FABP7*
360 shows a human-derived up-regulation in several neuronal subtypes but not glial cells (Supp Fig. 35) (Ma

361 et al., 2022). To investigate the genetic basis of this cell type-specific divergence, we leveraged a
362 machine learning model, Sei (K. M. Chen et al., 2022), to nominate potentially causal variants in the
363 promoter of *FABP7* (see Methods). Sei is a deep neural network that takes DNA sequence as input and
364 predicts the probability that the sequence has a particular epigenetic state in a variety of cell types and
365 tissues (Fig. 4d) (K. M. Chen et al., 2022). While single-base substitutions differing between human and
366 chimpanzee had only minor impacts on predicted *cis*-regulatory activity, "chimpanizing" the human
367 sequence of the *FABP7* promoter at two small indels (by deleting one base at chr6: 122,779,291 and
368 inserting three bases at chr6: 122,779,115) was sufficient to make the Sei predictions for the
369 chimpanized human sequence closely match the predictions for the complete chimpanzee sequence
370 (Fig. 4e-g). Making only one of these changes had substantial but weaker effects in both cases,
371 suggesting that both mutations might be functionally important (Fig. 4f,g). The 1-base insertion in the
372 human lineage introduces a binding site for the neuronally expressed transcription factors GLIS2/3,
373 suggesting a potential molecular mechanism (Calderari et al., 2018; Castro-Mondragon et al., 2022; Ke
374 et al., 2015).
375
376 As another approach to ranking the 95 peaks, we searched for peaks containing human-chimpanzee
377 sequence differences in otherwise highly conserved genomic positions, since these could reflect changes
378 in selective pressure. Using PhyloP scores for 241 placental mammals (Sullivan et al., 2023) to assess
379 conservation, one of the top-ranked peaks was a putative enhancer six kilobases away from the TSS of
380 *GAD1*, which plays a key role in the synthesis of the neurotransmitter GABA (Feldblum et al., 1993).
381 Notably, part of this peak has been classified as a human accelerated region (HAR) (Girskis et al., 2021;
382 Hubisz & Pollard, 2014; Pollard et al., 2006)—a short sequence that is highly conserved in mammals yet
383 contains an unusual number of human-specific mutations. Both the accessibility in the peak and *GAD1*
384 expression are only chimpanzee-biased in motor neurons (Fig. 4a-c, Supp Fig. 36). Applying Sei to

385 estimate the predicted effect of every variant in this region, we found that a single-nucleotide
386 substitution within the HAR (chr2: 170,823,193) has by far the largest predicted cis-regulatory effect and
387 most closely matches the differences in Sei predictions between the full human and chimpanzee
388 haplotypes (Fig. 4d-g). Interestingly, this mutation is predicted to disrupt a binding site for several basic
389 helix-loop-helix transcription factors that play essential roles in neuronal differentiation such as Ascl1
390 (Supp Fig. 37) (Castro-Mondragon et al., 2022; Mizuguchi et al., 2006; Yang et al., 2017).
391
392 As *GAD1* is only highly expressed in GABAergic neurons (and was therefore lowly expressed in the cell
393 types studied here, Supp Fig. 38a), we investigated whether this reduced expression of human *GAD1*
394 also occurs in cortical organoids which contain GABAergic neurons together with other cell types in
395 which *GAD1* is not highly expressed. We analyzed our previously published data from human-
396 chimpanzee hybrid cortical organoids (Agoglia et al., 2021) and found that the expression of *GAD1* from
397 the chimpanzee allele spikes higher than that of the human allele around day 50 of hybrid cortical
398 organoid differentiation before dropping in expression over time to match the human expression level
399 (Supp Fig. 38b). Because ASE in the hybrid cells controls for any potential interspecies differences in
400 differentiation kinetics or cell type composition, this difference must be the result of *cis*-regulatory
401 divergence between humans and chimpanzees. This expression difference is also more pronounced in
402 comparisons of human and chimpanzee parental cortical organoids, with a higher absolute log fold-
403 change at day 50, day 100, and day 150, only returning to equal expression at day 200 (Supp Fig. 38c).
404 While this could be due to differences in cell type proportion between human and chimpanzee
405 organoids, it might also be due to a reinforcing *trans*-acting effect.
406
407 To test whether this difference in expression also occurs specifically during GABAergic neuron
408 differentiation we examined *GAD1* expression in single cell RNA-seq data from human and chimpanzee

409 cortical organoids (Kanton et al., 2019). Consistent with our cortical organoid results, we observed a
410 peak in *GAD1* expression in less mature chimpanzee GABAergic neurons that is absent in the
411 corresponding part of the trajectory in human neurons (Supp Fig. 39). Notably, a similar trend holds
412 regardless of which GABAergic sub-trajectory (i.e., equivalent to GABAergic neurons from the caudal,
413 lateral, or medial ganglionic eminences) is examined suggesting this difference is not unique to a
414 particular type of GABAergic neuron (Supp Fig. 39). Finally, we examined the accessibility of the putative
415 *GAD1* enhancer more closely. Consistent with a potential role for this enhancer in the spike in *GAD1*
416 expression during development, the accessibility of this enhancer mirrors the expression of *GAD1* in
417 human cortical and striatal organoids (Supp Fig. 40) with high accessibility between day 50 and day 100
418 before decreasing somewhat near day 150 (Trevino et al., 2020). Overall, our results demonstrate how
419 the combination of RNA-seq, ATAC-seq, and machine learning models can nominate variants that may
420 be responsible for cell type-specific changes in gene expression and chromatin accessibility.
421
422 **Discussion**
423
424 In this study, we quantified human-chimpanzee *cis*-regulatory divergence in gene expression and
425 chromatin accessibility in six different cell types representing diverse developmental lineages. Across the
426 thousands of genes with ASE, we found that most *cis*-regulatory divergence is specific to one or a few
427 cell types. Furthermore, we found that divergent *cis*-regulation is linked to tissue-specificity, with tissue-
428 specific genes being enriched for ASE and tissue-specific regulatory elements being enriched for ASCA.
429 As this result was largely unchanged when stratifying by evolutionary constraint, our results suggest that
430 changes in the expression of genes with more cell type-specific expression patterns may be less
431 deleterious than changes in more broadly expressed genes, supporting the idea that cell-type specific
432 divergence may be less pleiotropic (Wittkopp & Kalay, 2012). Overall, this suggests that broad changes

433 in expression in cell type-specifically expressed genes may be an important substrate for evolution, but
434 it remains unclear whether positive selection or lower constraint plays a larger role in driving the faster
435 evolution of more cell type-specifically expressed genes. Future work will be required to more precisely
436 quantify the relative roles of positive selection and evolutionary constraint in driving changes in gene
437 expression.
438
439 We also identified several sets of genes evolving under lineage-specific selection that may have played a
440 role in establishing unique facets of human physiology and behavior. Most interestingly, we found
441 evidence for selection on sodium ion transporters and syntaxin binding proteins that may alter the
442 electrophysiological properties of motor neurons, and potentially other types of neurons as well (Brose
443 et al., 2019; Meisler et al., 2021). The complexity of the molecular machinery regulating neuronal
444 excitability and synaptic vesicle release make it difficult to say what the effects of these gene expression
445 changes are on the excitability of motor neurons without electrophysiology data from human and great
446 ape neurons coupled with perturbation of candidate genes. However, given the divergence in
447 locomotion and motor skills between humans and chimpanzees, one could speculate that these changes
448 may have had some role in the evolution of motor control and learning in humans.
449
450 In this work, we developed two metrics—dEE and dCAE—to quantify the degree of cell type-specific
451 differential expression and accessibility. These metrics are largely analogous to widely used metrics that
452 quantify tissue- or cell type-specific expression level and applicable to any comparison of log fold-
453 changes across conditions. They markedly improved our ability to identify matching cell type-specific
454 ASE and ASCA and led to the identification of 95 peak-gene pairs that had highly cell type-specific
455 concordant changes in accessibility and expression.
456

457 One such example is a human-derived increase in *FABP7* expression in several types of human neurons.
458 As *FABP7* is not highly expressed in adult mouse neurons, the functional consequences of its higher
459 expression in human neurons are difficult to predict (Yao et al., 2021). *FABP7* plays a role in the uptake
460 of the fatty acid Docosahexaenoic Acid (DHA), an important component of neuronal membranes (Akbar
461 et al., 2005; Choi et al., 2021). DHA promotes neuronal survival through phosphatidylserine
462 accumulation, so it is possible that the human-specific *FABP7* expression increases neuronal DHA uptake
463 leading to reduced apoptosis in human neurons during development and ultimately contributing to a
464 larger number of neurons in humans (Akbar et al., 2005; Choi et al., 2021).
465
466 In addition, we identified a highly conserved developmentally dynamic enhancer near *GAD1* that may
467 have partially lost activity in the human lineage resulting in a decrease in *GAD1* expression early in
468 GABAergic neuron development. By integrating with the deep learning model Sei (K. M. Chen et al.,
469 2022), we identified a variant that may account for the chimpanzee-biased ASCA in this region .
470 Interestingly, the ASE of *GAD1* was coupled with a relatively small (though significant) magnitude of
471 ASCA. This could potentially reflect divergence in transcription factor binding that leaves a “footprint”
472 resulting in subtle ASCA (Vierstra et al., 2020). Overall, our data suggest that this enhancer has lost
473 activity in the human lineage, potentially altering the expression pattern of *GAD1* during neuronal
474 development. *GAD1* is the rate-limiting enzyme for GABA synthesis so GABA levels are likely responsive
475 to changes in *GAD1* expression (Feldblum et al., 1993). GABA release has complex context-specific
476 effects on neurodevelopment, making it difficult to speculate as to what the phenotypic effects of
477 reduced GABA synthesis during human neurodevelopment might be (Ben-Ari et al., 2012). However, the
478 high conservation of this *cis*-regulatory element in placental mammals implies that its human-specific
479 disruption is likely to have important neurodevelopmental effects. Careful perturbation of this enhancer
480 and *GAD1* expression in mouse models will be required to explore this further.

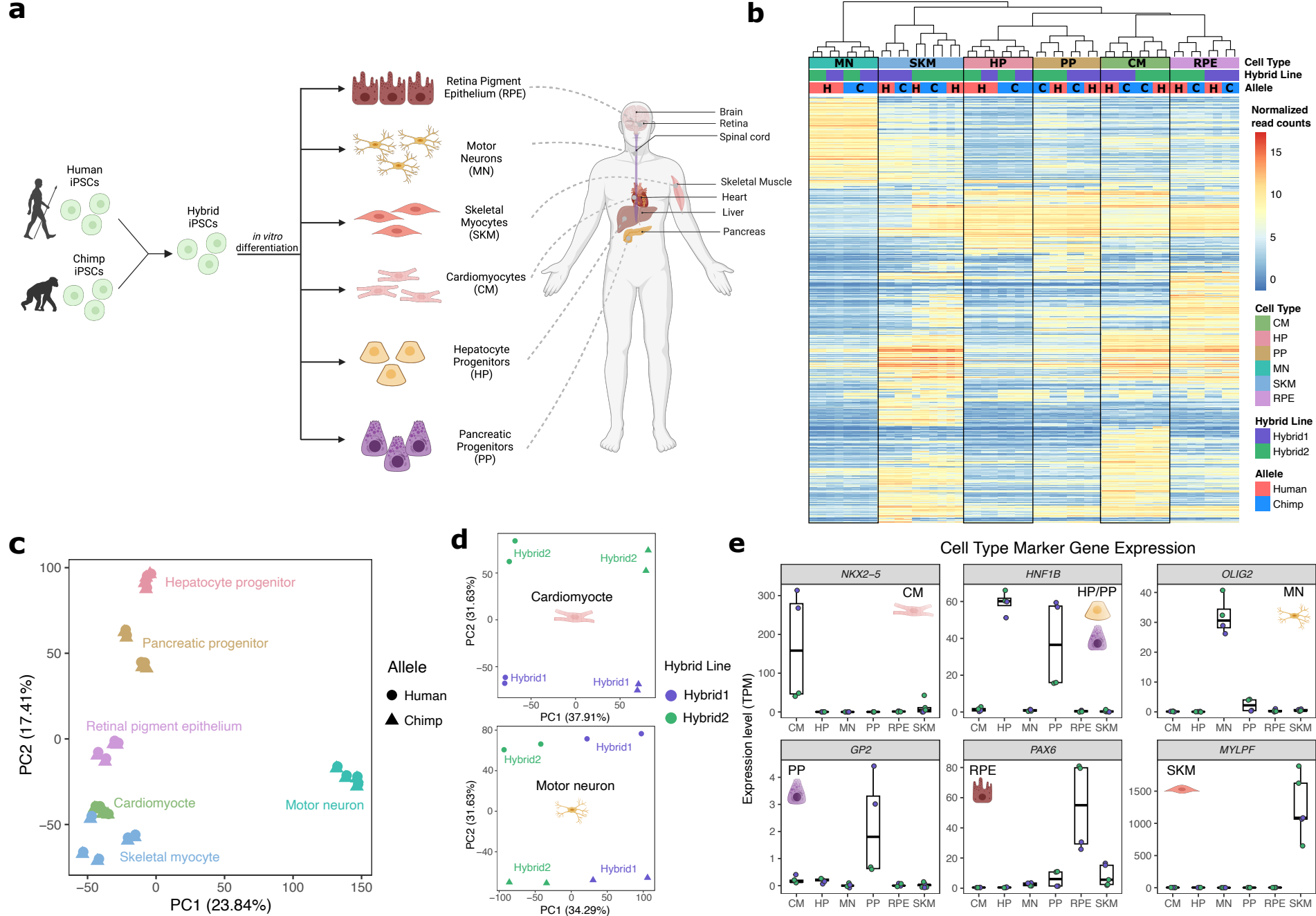
481

482 In addition to following up on our findings on *GAD1* and *FABP7*, there are other exciting future
483 directions for this work. First, additional bulk assays such as those that measure methylation, chromatin
484 conformation, and translation rate could lead to a better understanding of what molecular features
485 ultimately lead to cell type-specific changes in gene expression. Furthermore, the use of deep single cell
486 profiling of hybrid lines derived from iPSCs from multiple individuals of each species during
487 differentiation could enable the identification of many more highly context-specific changes in gene
488 expression and chromatin accessibility such as the differences in *GAD1* we highlighted here. Finally,
489 integration with data from massively parallel reporter assays and deep learning models will help us link
490 specific variants to the molecular differences we identified in this study.

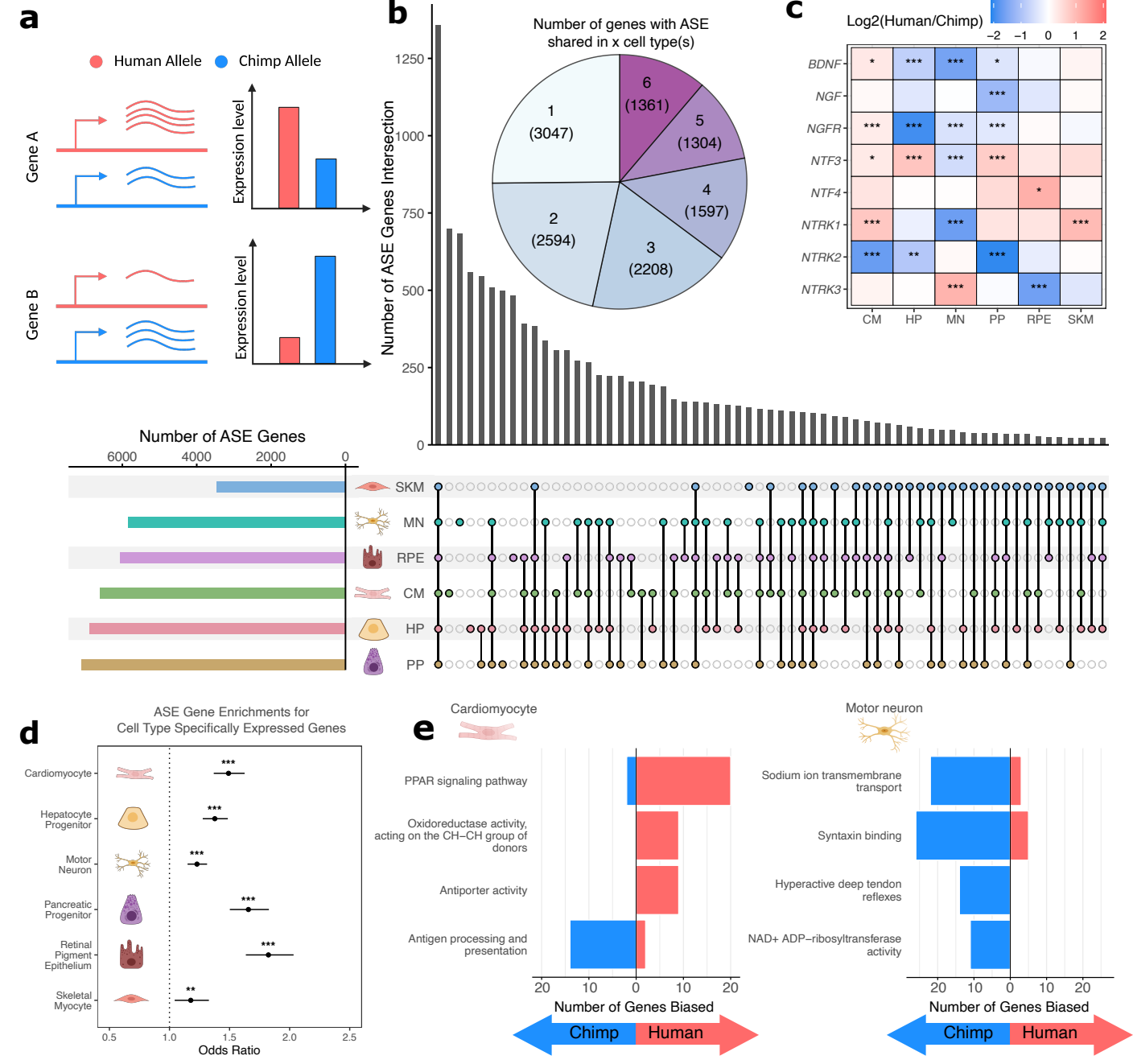
491

492 Overall, our study provides foundational data, insight, and computational tools that will improve our
493 understanding of cell type-specific *cis*-regulatory evolution and the role it has played in the
494 establishment of human-specific traits.

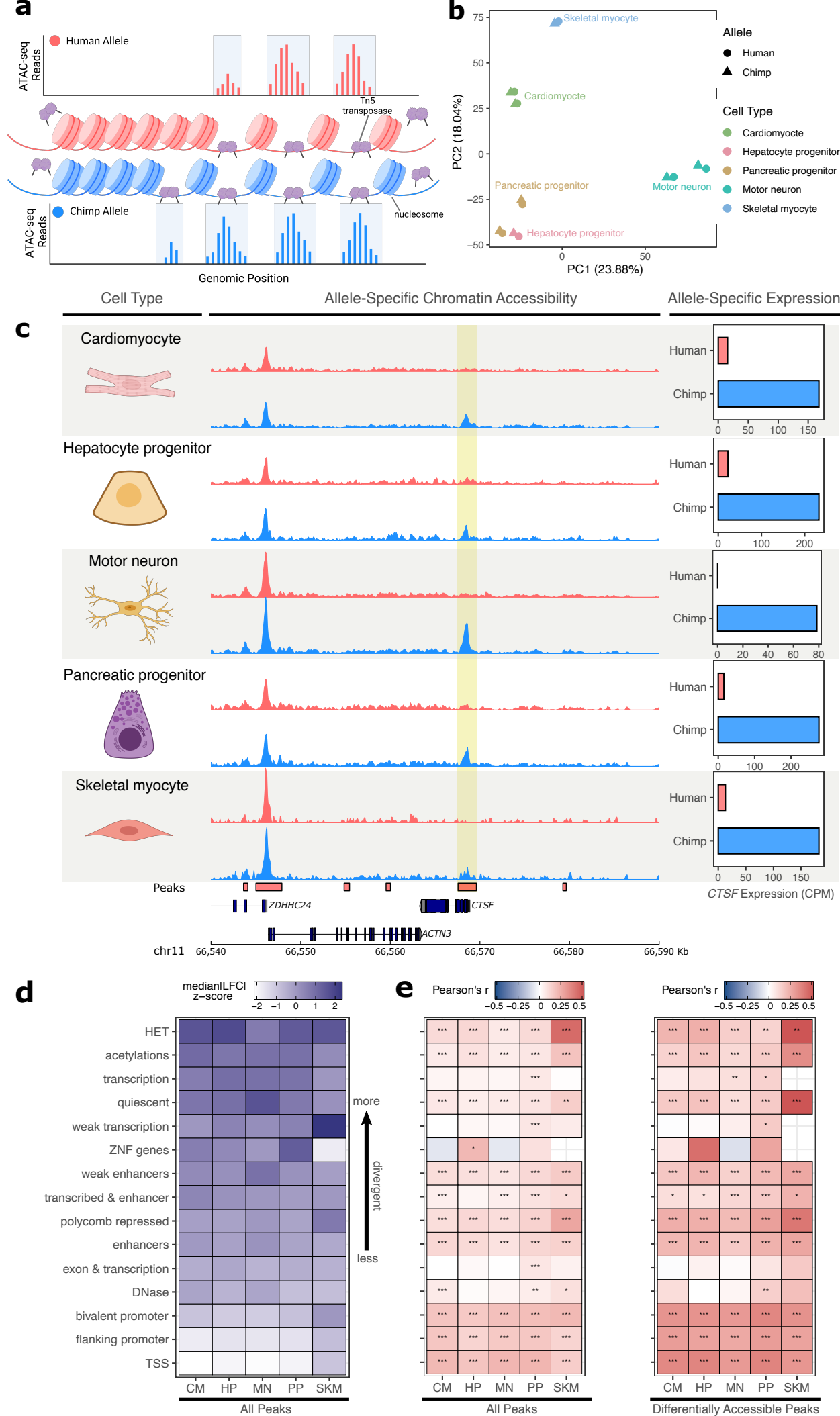
495


a
Human iPSCs
Chimp iPSCs
Hybrid iPSCs
in vitro differentiation
Retina Pigment Epithelium (RPE)
Motor Neurons (MN)
Skeletal Myocytes (SKM)
Cardiomyocytes (CM)
Hepatocyte Progenitors (HP)
Pancreatic Progenitors (PP)
Brain
Retina
Spinal cord
Skeletal Muscle
Heart
Liver
Pancreas
b
MN
SKM
HP
PP
CM
RPE
Cell Type
Hybrid Line
Allele
Normalized read counts
15
10
5
0
Cell Type
CM
HP
PP
MN
SKM
RPE
Hybrid Line
Hybrid1
Hybrid2
Allele
Human
Chimp
c
Hepatocyte progenitor
Pancreatic progenitor
Retinal pigment epithelium
Cardiomyocte
Motor neuron
Skeletal myocyte
PC2 (17.41%)
PC1 (23.84%)
Allele
Human
Chimp
d
Cardiomyocte
Motor neuron
Hybrid1
Hybrid2
PC2 (31.63%)
PC1 (37.91%)
PC1 (34.29%)
Hybrid Line
Hybrid1
Hybrid2
e
Cell Type Marker Gene Expression
NKX2–5
HNF1B
OLIG2
GP2
PAX6
MYLPF
CM
HP/PP
MN
PP
RPE
SKM
Expression level (TPM)
CM HP MN PP RPE SKM

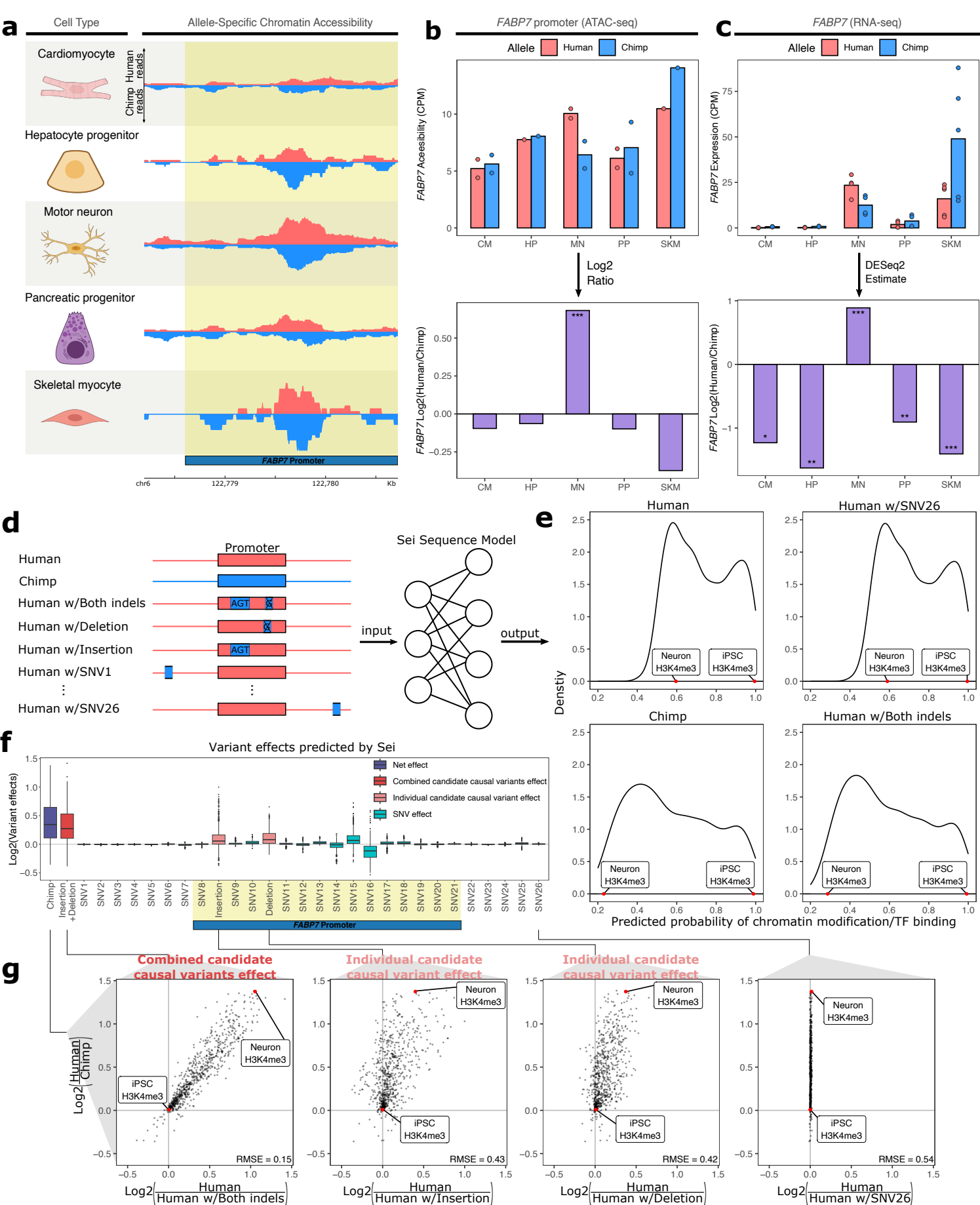
496 **Figure 1: Allele-specific expression across diverse human-chimpanzee hybrid cell types. a)** Six cell types
497 were differentiated from human-chimpanzee hybrid induced pluripotent stem cells. These six cell types
498 represent diverse body systems, including motor neurons for the central nervous system, retinal
499 pigment epithelium for eye, skeletal myocytes for skeletal muscle, cardiomyocytes for the heart,
500 hepatocyte progenitors for the liver, and pancreatic progenitors for the pancreas. **b)** Heatmap showing
501 the result of hierarchical clustering performed on genes with highly variable normalized allele counts. **c)**
502 Result of running PCA on normalized allelic counts for all samples and cell types. **d)** Result of PCA
503 performed on normalized allele counts for each individual cell type separately. Cardiomyocytes and
504 motor neurons are shown here. **e)** Expression of marker genes for each cell type.


a
Human Allele
Chimp Allele
Gene A
Gene B
Expression level
Expression level
b
Number of genes with ASE shared in x cell type(s)
1 (3047)
2 (2594)
3 (2208)
4 (1597)
5 (1304)
6 (1361)
Number of ASE Genes Intersection
1250
1000
750
500
250
0
Number of ASE Genes
6000
4000
2000
0
SKM
MN
RPE
CM
HP
PP
c
Log2(Human/Chimp)
−2 −1 0 1 2
BDNF
NGF
NGFR
NTF3
NTF4
NTRK1
NTRK2
NTRK3
CM HP MN PP RPE SKM
d
ASE Gene Enrichments for Cell Type Specifically Expressed Genes
Cardiomyocyte
Hepatocyte Progenitor
Motor Neuron
Pancreatic Progenitor
Retinal Pigment Epithelium
Skeletal Myocyte
0.5 1.0 1.5 2.0 2.5
Odds Ratio
e
Cardiomyocyte
PPAR signaling pathway
Oxidoreductase activity, acting on the CH−CH group of donors
Antiporter activity
Antigen processing and presentation
20 10 0 10 20
Number of Genes Biased
Chimp
Human
Motor neuron
Sodium ion transmembrane transport
Syntaxin binding
Hyperactive deep tendon reflexes
NAD+ ADP−ribosyltransferase activity
20 10 0 10 20
Number of Genes Biased
Chimp
Human

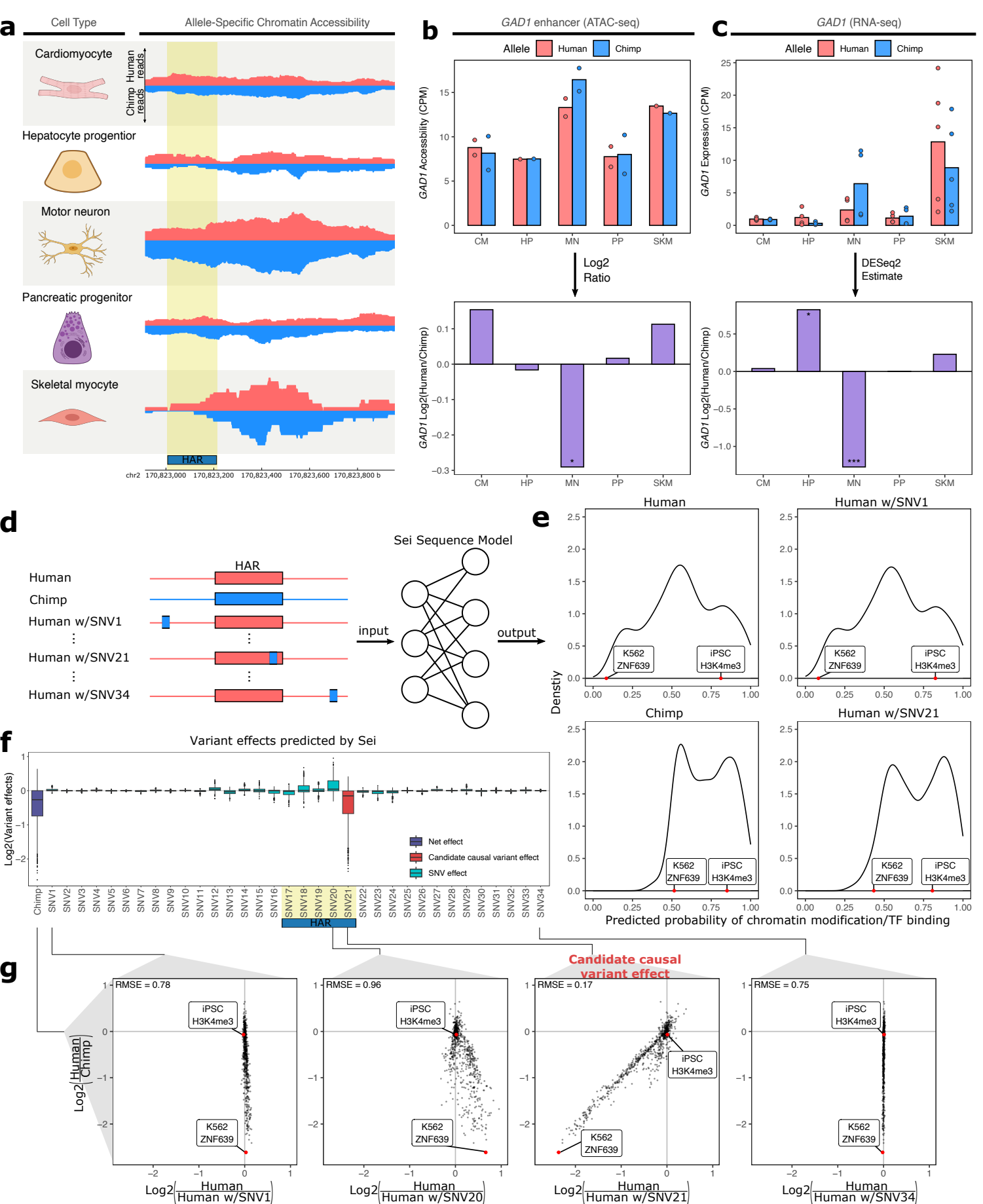
505 **Figure 2: Human-chimpanzee ASE is largely cell type-specific. a**) Outline of measurement of allele-
506 specific expression. Reads from the human and chimpanzee alleles are counted and differences in read
507 counts identified. **b)** Thousands of genes with ASE were identified for each cell type, with many genes
508 only showing ASE in a single cell type (Conway et al., 2017). **c)** The neurotrophins and their receptors as
509 examples of genes showing cell type-specific ASE patterns. DESeq2 estimate of $\log_2$ fold-change
510 (human/chimpanzee) are shown in the heatmap and significance is indicated by asterisks where ***
511 indicates FDR < 0.005, ** indicates FDR < 0.01, and * indicates FDR < 0.05. Zero asterisks (i.e. a blank
512 box) indicates FDR > 0.05. **d)** Plot showing that genes with ASE are enriched for genes showing cell type-
513 specific expression patterns across all cell types. Asterisks indicate p-values rather than FDR using the
514 same system as in 2c. **e)** Top gene sets with evidence for lineage-specific selection in cardiomyocytes
515 and motor neurons are shown. The length of the bars indicates the number of genes in a category with
516 biased expression in each cell type.
517


a
Human Allele
ATAC-seq Reads
Tn5 transposase
nucleosome
Chimp Allele
ATAC-seq Reads
Genomic Position
b
Skeletal myocyte
Cardiomyocte
Pancreatic progenitor
Motor neuron
Hepatocyte progenitor
PC2 (18.04%)
PC1 (23.88%)
Allele
Human
Chimp
Cell Type
Cardiomyocte
Hepatocyte progenitor
Pancreatic progenitor
Motor neuron
Skeletal myocyte
Cell Type
Allele-Specific Chromatin Accessibility
Allele-Specific Expression
c
Cardiomyocyte
Hepatocyte progenitor
Motor neuron
Pancreatic progenitor
Skeletal myocyte
Human
Chimp
Peaks
ZDHHC24
CTSF
ACTN3
chr11
66,540
66,550
66,560
66,570
66,580
66,590 Kb
CTSF Expression (CPM)
d
medianlLFCl z-score
HET
acetylations
transcription
quiescent
weak transcription
ZNF genes
weak enhancers
transcribed & enhancer
polycomb repressed
enhancers
exon & transcription
DNase
bivalent promoter
flanking promoter
TSS
more
divergent
less
CM
HP
MN
PP
SKM
All Peaks
e
Pearson's r
All Peaks
Differentially Accessible Peaks

518 **Figure 3: Allele-specific chromatin accessibility across diverse human-chimpanzee hybrid cell types. a)**
519 Schematic outlining the ATAC-seq protocol. A hyperactive transposase cleaves accessible DNA and adds
520 adapters enabling measurement of chromatin accessibility. **b)** PCA on normalized allelic counts from
521 ATAC-seq. **c)** ASCA in the promoter of *CTSF,* and ASE for the *CTSF* gene. **d)** Differences in ASCA were
522 quantified and plotted separately based on chromHMM annotation. The order is based on the median
523 of z-score transformed absolute log fold-change between human and chimpanzee across all cell types,
524 with higher z-scores indicating greater divergence in accessibility. **e)** Pearson correlation between ASE
525 and ASCA for all cell types with all peaks (left) or only differentially accessible peaks (right, defined as
526 peaks with nominal binomial p-value less than 0.05). Pearson's r values are shown in the heatmap and
527 significance is indicated by asterisks where *** indicates $p < 0.005$, ** indicates $p < 0.01$, and * indicates
528 $p < 0.05$. Zero asterisks (i.e. a blank box) indicates $p > 0.05$.
529


a
Cell Type
Allele-Specific Chromatin Accessibility
Cardiomyocyte
Human reads
Chimp reads
Hepatocyte progenitor
Motor neuron
Pancreatic progenitor
Skeletal myocyte
FABP7 Promoter
chr6
122,779
122,780
Kb
b
FABP7 promoter (ATAC-seq)
Allele
Human
Chimp
FABP7 Aceesibility (CPM)
CM
HP
MN
PP
SKM
Log2 Ratio
FABP7 Log2(Human/Chimp)
c
FABP7 (RNA-seq)
FABP7 Expression (CPM)
DESeq2 Estimate
d
Promoter
Human
Chimp
Human w/Both indels
Human w/Deletion
Human w/Insertion
Human w/SNV1
Human w/SNV26
AGT
input
Sei Sequence Model
output
e
Human
Human w/SNV26
Chimp
Human w/Both indels
Neuron H3K4me3
iPSC H3K4me3
Denstiy
Predicted probability of chromatin modification/TF binding
f
Variant effects predicted by Sei
Net effect
Combined candidate causal variants effect
Individual candidate causal variant effect
SNV effect
Log2(Variant effects)
Chimp
Insertion +Deletion
Insertion
Deletion
FABP7 Promoter
g
Combined candidate causal variants effect
Individual candidate causal variant effect
Individual candidate causal variant effect
Log2(Human/Chimp)
Log2(Human/Human w/Both indels)
Log2(Human/Human w/Insertion)
Log2(Human/Human w/Deletion)
Log2(Human/Human w/SNV26)
RMSE = 0.15
RMSE = 0.43
RMSE = 0.42
RMSE = 0.54

530 **Figure 4: Motor neuron-specific human-biased ASE and ASCA for *FABP7* and the promoter of *FABP7*. a)**
531 Allelic ATAC-seq tracks are shown in the peak containing the annotated *FABP7* promoter (highlighted in
532 yellow). **b)** The top panel shows allelic CPM of the *FABP7* promoter across cell types and the bottom
533 panel shows the log fold-change across cell types. **c)** The top panel shows allelic CPM of the *FABP7* gene
534 across cell types and the bottom panel shows the log fold-change across cell types. **d)** Outline of the
535 process for variant effect prediction with Sei for *FABP7*. All sequences input to Sei were centered at the
536 *FABP7* promoter. The human sequence, chimpanzee sequence, and partially "chimpanized" human
537 sequences (modified by systematically switching the human allele to the chimpanzee allele separately
538 for each human-chimpanzee difference) were fed into Sei to predict the effects of these variants on
539 chromatin state. **e)** Histogram of the probabilities of various chromatin states and transcription factor
540 binding predicted by Sei was plotted for the human sequence, the chimpanzee sequence, and the
541 human sequence with one human-chimpanzee difference swapped to match the chimpanzee sequence.
542 The human sequence with SNV26 changed to the chimpanzee allele and the human sequence switched
543 at both indels are shown as examples. A histone modification (H3K4me3) predicted in two cell types was
544 labeled to illustrate how the predictions depend on both input sequence and cell type. **f)** Plot of the
545 predicted effects of all single nucleotide differences and indels between the human and chimpanzee
546 genomes in the Sei input window (the *FABP7* promoter is highlighted in yellow). **g)** Scatterplots showing
547 the correlation of the effects of both indels (left panel), each individual indel (middle two panels), and a
548 representative SNV (right panel) on Sei predictions with the difference in Sei predictions between the
549 human and chimpanzee sequences. The root mean square error (RMSE) was computed and shown in
550 each figure.
551


a
Cell Type
Allele-Specific Chromatin Accessibility
Cardiomyocyte
Human reads
Chimp reads
Hepatocyte progenitor
Motor neuron
Pancreatic progenitor
Skeletal myocyte
HAR
chr2 170,823,000 170,823,200 170,823,400 170,823,600 170,823,800 b
b
GAD1 enhancer (ATAC-seq)
Allele Human Chimp
GAD1 Accessibility (CPM)
CM HP MN PP SKM
Log2 Ratio
GAD1 Log2(Human/Chimp)
c
GAD1 (RNA-seq)
GAD1 Expression (CPM)
DESeq2 Estimate
d
HAR
Human
Chimp
Human w/SNV1
Human w/SNV21
Human w/SNV34
input
Sei Sequence Model
output
e
Human
Human w/SNV1
Chimp
Human w/SNV21
Denstiy
K562 ZNF639
iPSC H3K4me3
Predicted probability of chromatin modification/TF binding
f
Variant effects predicted by Sei
Log2(Variant effects)
Net effect
Candidate causal variant effect
SNV effect
Chimp SNV1 SNV2 SNV3 SNV4 SNV5 SNV6 SNV7 SNV8 SNV9 SNV10 SNV11 SNV12 SNV13 SNV14 SNV15 SNV16 SNV17 SNV18 SNV19 SNV20 SNV21 SNV22 SNV23 SNV24 SNV25 SNV26 SNV27 SNV28 SNV29 SNV30 SNV31 SNV32 SNV33 SNV34
HAR
g
Candidate causal variant effect
RMSE = 0.78
RMSE = 0.96
RMSE = 0.17
RMSE = 0.75
iPSC H3K4me3
K562 ZNF639
Log2(Human/Chimp)
Log2(Human/Human w/SNV1)
Log2(Human/Human w/SNV20)
Log2(Human/Human w/SNV21)
Log2(Human/Human w/SNV34)

552 **Figure 5: Motor neuron-specific chimpanzee-biased ASE for *GAD1* and ASCA for a HAR near the *GAD1***
553 **TSS. a)** Allelic ATAC-seq tracks are shown for the peak near the *GAD1* TSS that contains a HAR
554 (highlighted in yellow). **b)** The top panel shows allelic CPM of the CRE near the *GAD1* TSS across cell
555 types and the bottom panel shows the log fold-change across cell types. **c)** The top panel shows allelic
556 CPM of the *GAD1* gene across cell types and the bottom panel shows the log fold-change across cell
557 types. **d)** Outline of the process for variant effect prediction with Sei for *GAD1*. All input sequences to Sei
558 were centered at the HAR. The human sequence, the chimpanzee sequence, and modified sequences
559 with the human sequence altered at each substitution to match the chimpanzee sequence were fed into
560 the Sei sequence model to predict the effects of these variants on the chromatin state. **e)** Histogram of
561 the probabilities of various chromatin states and transcription factor binding predicted by Sei was
562 plotted for the human sequence, the chimpanzee sequence, and two examples in which the human
563 sequence with only one SNV "chimpanized" (human w/SNV) was input to Sei. The histogram of the
564 probability of the sequence having a particular epigenomic annotation (predicted by Sei) was plotted for
565 human, chimpanzee, human w/SNV1 changed to match the chimpanzee sequence, and human w/SNV21
566 changed to match the chimpanzee sequence. Two epigenomic annotations were labeled as examples
567 that show the different values output by Sei with these two different sequence inputs. **f)** Plot of the
568 predicted effects of all single nucleotide differences between human and chimpanzee in the Sei input
569 window centered at the HAR (highlighted in yellow). Positions were switched to the chimpanzee allele
570 individually. **g)** Scatterplots showing the correlation of the effects of four SNVs on Sei predictions with
571 the difference in Sei predictions between the human and chimpanzee sequences. The root mean square
572 error (RMSE) was computed and shown in each figure.

573 **Methods:**

574 **Generation of multiple human-chimpanzee hybrid cell types**

575 We used two previously described human-chimpanzee hybrid iPS cell lines (hybrid1 and hybrid2,
576 previously denoted Hy1-25 and Hy1-30 respectively) (Agoglia et al., 2021). Before differentiation, cells
577 were routinely cultured on matrigel in mTeSR1 or mTeSR Plus (Stem Cell Technologies cat #85850 or cat
578 #100-0276). Culture and *in vitro* differentiation of iPS cells into six cell types (motor neurons (MN),
579 cardiomyocytes (CM), hepatocyte progenitors (HP), pancreatic progenitors (PP), skeletal myocytes
580 (SKM), and retinal pigment epithelium (RPE)) was carried out by the Columbia Stem Cell Core Facility
581 using published protocols (Burridge et al., 2014; Chal et al., 2016; Korytnikov & Nostro, 2016; Mallanna
582 & Duncan, 2013; Maury et al., 2015; Sharma et al., 2019).

583

584 **Preparation of RNA-seq libraries**

585 All samples were cryopreserved in liquid nitrogen before RNA extraction (Milani et al., 2016). Cells were
586 gently thawed and then washed with PBS and cell pellets were collected via centrifugation at 1,000 RPM
587 for 5 minutes. Cell pellets were loosened by flicking the tube and an appropriate volume of Buffer RLT
588 based on the cell count were added following the RNeasy Mini Kit (Qiagen, 74104) protocol. Total RNA
589 extraction and on-column DNase digestion were performed using RNeasy Mini Kit (Qiagen, 74104) and
590 RNase-Free DNase Set (Qiagen, 79254). RNA quality was assessed using the Agilent Bioanalyzer RNA Pico
591 assay. Only samples with an RNA integrity number (RIN) greater than or equal to 7 were used to prepare
592 cDNA libraries. All RNA-seq libraries except three motor neuron libraries were prepared using the
593 TruSeq Stranded mRNA kit (Illumina, 20020594) and the TruSeq RNA CD Index Plate (Illumina,
594 20019792) from between 100 ng and 1 ug total RNA following the manufacturer's protocols. Due to low
595 yield of total RNA, three motor neuron libraries (2 hybrid2 and 1 hybrid1) were prepared using Illumina
596 Stranded mRNA Prep (Illumina, 20040532) and IDT for Illumina RNA UD Indexes Set A, Ligation (Illumina,

597 20040553). Notably, the four motor neuron libraries did not cluster by library preparation method. All
598 libraries were normalized, pooled at an equimolar ratio using Qubit measurements, and sequenced on
599 an Illumina HiSeq 4000 to generate 2x150bp paired-end reads.
600
601 **Identification of confident human-chimpanzee SNVs**
602 To identify a confident list of human-chimpanzee SNVs that could be used to quantify allele-specific
603 expression and chromatin accessibility, we first downloaded hg38-panTro6 MAF files from UCSC and
604 whole-genome sequencing data generated from the parental human and chimpanzee iPS cells (in the
605 form of bam files aligned to hg38 and panTro5, generously provided by the Gilad lab). We first
606 converted them back to fastq files and then mapped reads to panTro6 and hg38 (we mapped both
607 human and chimpanzee to both reference genomes) using bowtie2 with the flags —very-sensitive-local -
608 p 16 (Langmead & Salzberg, 2012). We then used a modified version of our previous approach to filter
609 out SNVs that could not be confidently identified as homozygous in the human and chimpanzee parental
610 lines (Agoglia et al., 2021). Briefly, we extracted SNVs and indels from both human and chimpanzee MAF
611 files, counted reads in the WGS data that supported the human, chimpanzee, or an alternative base at
612 that position, then filtered out any SNVs with < 2 reads or < 90% of reads supporting that species' base.
613 We then reformatted files, merged with indels for use in Hornet, and generated a modified bed file of
614 SNVs that includes the human and chimpanzee base at the SNV position (van de Geijn et al., 2015).
615
616 **Generation of allele-specific count tables**
617 An allele-specific expression pipeline adapted from Agoglia et al. and our updated high-confidence SNV
618 list was used. The whole pipeline was carried out twice independently using hg38 or panTro6 as the
619 reference genome. This approach was taken to eliminate genes showing strong mapping bias, defined
620 here as genes with an absolute difference in $\log_2$ fold-change between the panTro6-referenced and

621 hg38-referenced runs greater than one. All sequencing reads were trimmed with SeqPrep (adapters
622 specified by the manufacturer for the different library preparation kits) and mapped using STAR with
623 two passes and the following parameters: --outFilterMultimapNmax 1 (Dobin et al., 2013; John St. John,
624 n.d.). Uniquely aligned reads were deduplicated with Picard and Hornet (an implementation of WASP
625 which first removes reads overlapping indels) was used to correct for mapping bias (Broad Institute,
626 n.d.; van de Geijn et al., 2015). Reads were assigned to either the human allele if they contained one or
627 more human-chimpanzee single nucleotide differences that matched the human sequence and zero
628 positions that matched the chimpanzee sequence (and vice versa for assigning reads to the chimpanzee
629 allele) and counted per gene as previously described (Agoglia et al., 2021).
630

631 **Detection of aneuploidy on chromosome 20 and slight chimpanzee parental contamination in PP**
632 **hybrid2 samples**

633 In our quality control process, we plotted the $\log_2$ fold-change for each gene along every chromosome
634 and inspected the results. This revealed a clear bias toward the human allele on a part of chromosome
635 20 for hybrid2 samples, suggesting chromosome 20 aneuploidy which was also reported by Agoglia et al
636 (Agoglia et al., 2021). As a result, we excluded chromosome 20 from all downstream analyses. In
637 addition, we found that PP hybrid2 samples had a slight bias toward the chimpanzee allele across every
638 chromosome which was most likely due to a small fraction of contaminating chimpanzee cells in these
639 samples. Rather than removing these samples, we normalized the allele-specific count tables by
640 subtracting a small number of reads from the chimpanzee allele counts calculated based on the biased
641 ratio summarized from genome wide human and chimpanzee allele counts, to force a global log fold-
642 change (across all autosomes except chromosome 20) of zero between the human and chimpanzee
643 alleles. We applied this normalization to all other samples as well. To evaluate the success of this
644 strategy, we simulated the effects of chimpanzee iPSC contamination and our subsequent correction.

645 Specifically, using hybrid and chimpanzee parental iPSC RNA-seq from Agoglia et al (Agoglia et al., 2021)
646 from the same iPS cell lines as used in this study, we simulated chimpanzee iPSC contamination by
647 mixing chimpanzee RNA-seq data into hybrid2 data to reach a similar degree of chimpanzee bias level to
648 that observed in the PP hybrid2 samples. We then identified genes showing ASE (see "Identifying genes
649 with ASE") using the counts from the original hybrid samples, simulated contaminated samples, and
650 corrected simulated contaminated samples and compared the outputs (Supp Fig. 4).
651

652 **PCA and hierarchical clustering**

653 Allelic counts were normalized by DESeq2 rlog and principal components analysis (PCA) was performed
654 on rlog normalized allelic counts with default centering and scaling (Love et al., 2014). The top 1,000
655 variable genes with the highest variance of normalized allelic counts across all cell types were used to
656 compute Euclidean distance matrices. The R package pheatmap was used to do hierarchical clustering
657 and heatmap plotting.
658

659 **Identifying genes with ASE**

660 DESeq2 was used to measure allele-specific expression (ASE) in each cell type (Love et al., 2014). All
661 reads from chromosome 20 were removed (as mentioned above). Two replicates per hybrid line per cell
662 type (plus one additional replicate for SKM hybrid2 for a total of 3 samples) were used by DEseq2 with
663 model ~hybLine+Species to measure differential expression level. A likelihood ratio test (test="LRT",
664 betaPrior=FALSE) was used to compute p-values. P-values were then false discovery rate adjusted using
665 an implementation of the Benjamini-Hochberg correction in the R package qvalue (Benjamini &
666 Hochberg, 1995; Storey Lab, n.d.). Log fold-changes were shrunk as recommended by the DESeq2
667 pipeline (Zhu et al., 2019). Differentially expressed genes were defined as those with FDR < 0.05 when

668 aligned to hg38 and panTro6 as well as an absolute difference in log fold-change <= 1 when comparing
669 the results from the two alignments.
670
671 **Identifying cell type-specifically expressed genes**
672 For the more traditional definition of cell type-specific genes, we required transcripts per million (TPM)
673 < 1 for a gene in every cell type except one. In the cell type with TPM > 1, we varied how highly
674 expressed the gene had to be in that cell type (again using a TPM cutoff, varying between one and five)
675 to consider that gene to be specific to that cell type. A similar process to the one described in
676 "Identifying differentially expressed genes" was used to identify cell type-specific genes based on the
677 broader definition described in the main text. Rather than using allelic counts, total counts for each
678 sample (i.e. all uniquely mapping deduplicated reads regardless of their allelic origin) were computed by
679 summing all allelic and non-allelic counts**.** These counts were inputted to DESeq2 and the expression of
680 each gene was compared pairwise between all cell types (Love et al., 2014). Genes were defined as cell
681 type-specifically expressed in a cell type only if all pairwise comparisons between that cell type and
682 other cell types resulted in an FDR < 0.05 using both hg38 and panTro6 aligned counts. Due to the
683 markedly lower number of differentially expressed genes identified in SKM, results were computed both
684 including and excluding SKM. An analogous procedure was used to identify more broadly defined cell
685 type-specific peaks in the down-sampled ATAC-seq dataset. Peaks were defined as specific to a cell type
686 if the absolute log fold-change was greater than 0.5 across all pairwise comparisons with the other cell
687 types. We also tested an absolute log fold-change threshold of 1 to ensure that our results were not
688 sensitive to the choice of cutoff.
689
690 **Enrichment test for genes with cell type-specific expression patterns and genes showing ASE**

691 Odds ratios were calculated using the unconditional maximum likelihood estimate implemented in the R
692 package epitools function oddsratio.wald(), and 95% confidence intervals and p-values were calculated
693 using the normal approximation. A directly analogous procedure was performed to test for enrichment
694 of peaks with ASCA and cell type-specific peaks.
695
696 **Enrichment test stratified by expression level or evolutionary constraint**
697 Enrichment tests were carried out as in 'Enrichment test of cell type-specific expression patterns and
698 genes showing ASE' except that genes were split into five equal size bins depending on which factors
699 were used to stratify genes, and tests were done in each bin. When stratifying by expression level, genes
700 were ordered in ascending order based on expression level (TPM) and then split into five equally sized
701 bins where genes in the 0-20% bin are the most lowly expressed genes and genes in the 80-100% bin are
702 the most highly expressed genes. When stratifying by constraint metrics such as ASE variance (see
703 "Identification of lineage-specific selection on gene expression" for details on how ASE variance was
704 computed) or pHI, genes were ordered in ascending order based on ASE variance values and then split
705 into five equal size bins where the 0-20% bin contains genes with the lowest ASE variance (i.e. most
706 evolutionarily constrained) and the 80-100% bin contains genes with highest ASE variance (i.e. least
707 evolutionarily constrained). To stratify the ATAC data by constraint, we used the "QCed genomic
708 constraint by 1kb regions" computed by the gnomAD consortium (S. Chen et al., 2022). The gnomAD
709 consortium computed the ratio of the observed and expected number of human polymorphisms in 1 kb
710 regions tiling the human genome and then converted this metric to a z-score (see S. Chen et al., 2022 for
711 additional details). We further removed any regions that overlapped protein-coding exons from the
712 human gtf file using bedtools subtract (Quinlan & Hall, 2010). If a peak overlapped two or more 1
713 kilobase windows, it was assigned to the window with the highest constraint, mirroring the procedure
714 used by the gnomAD consortium to assign peaks to ENCODE regulatory elements (S. Chen et al., 2022).

715 Once the peaks were ranked by this metric, a procedure identical to that for the gene expression
716 constraint metrics outlined above was performed.
717
718 **Identification of lineage-specific selection on gene expression**
719 We developed a modified version of our previously published pipeline, which uses ASE values from
720 many individuals of a single species to estimate cis-regulatory constraint of each gene (Starr et al.,
721 2023). We restricted these ASE values to GTEx samples from the tissue(s) of origin for each cell type
722 (hepatocytes with liver, skeletal muscle with skeletal muscle, cardiomyocytes with the left ventricle of
723 the heart, and pancreatic progenitors with the pancreas). As RPE and MN did not have clear matching
724 tissues (e.g. GTEx does not include data from the eye), we compared RPE to all GTEx samples and MN to
725 all brain and peripheral nerve samples. We then used the Mann-Whitney U test to compare the human
726 population ASE distribution to the human-chimpanzee ASE distribution as previously described (Starr et
727 al., 2023). We employed the previously described signed ranking by Mann-Whitney p-value that
728 incorporates whether a gene has human or chimpanzee-biased ASE with GSEAPY and the binomial test
729 to identify instances of lineage-specific selection (Starr et al., 2023). Positive selection on a gene set is
730 only inferred if there is statistically significant human- or chimpanzee-biased ASE in that gene set (using
731 an FDR-corrected p-value from the binomial test). Due to the focus on tissue-specificity, we did not filter
732 redundant gene sets with GSEAPY FDR < 0.25 in multiple cell types (Starr et al., 2023; Subramanian et
733 al., 2005). To compute ASE variance, we used the same tissue-of-origin-matched data from GTEx and
734 computed the variance of the ASE ratios after filtering out samples with fewer than ten counts from the
735 reference or alternate allele (see Starr et al., 2023 for additional details). For example, if a sample had 11
736 reference counts and 2 alternate counts for a gene, that sample would be excluded for that gene.
737
738 **Preparation of ATAC-seq libraries**

739 We used the OmniATAC protocol with the only modification being the use of 25,000 cells instead of
740 50,000 since the fused iPS cells are tetraploid (Corces et al., 2017). All samples for ATAC-seq prep were
741 from the same vials used in RNA-seq library preparation except for the motor neuron libraries due to the
742 low yield of total RNA extracted from motor neurons. After library preparation and running samples on a
743 Bioanalyzer, we noticed a considerable number of fragments greater than 1000 bases in length. To
744 reduce these fragments, we size selected with Ampure beads using the protocol from the Kaestner lab
745 available here: https://www.med.upenn.edu/kaestnerlab/assets/user-content/documents/ATAC-
746 seq-Protocol-(Omni)-Kaestner-Lab.pdf.
747 After size selection and rerunning on the Bioanalzyer, we pooled the libraries together and sequenced
748 them to compute quality control metrics. We used the R package ChrAccR to compute TSS enrichment
749 scores (Mueller, Fabian, n.d.). We pooled all libraries with TSS enrichment score greater than 3.5. This
750 resulted in 2 CM libraries, 2 MN libraries, 2 PP libraries, 1 SKM library, and 1 HP library. After pooling,
751 libraries were sequenced on an Illumina Hiseq 4000 to produce 2x150 paired-end reads.
752
753 **Mapping the ATAC-seq data**
754 We trimmed reads using SeqPrep and then mapped them to the hg38 and panTro6 reference genomes
755 with bowtie2 in paired-end mode (John St. John, n.d.; Langmead & Salzberg, 2012). The following
756 parameters were used: -X 2000 --very-sensitive-local -p 16. After mapping, duplicates were removed via
757 Picard MarkDuplicates (Broad Institute, n.d.). We then removed multi-mapping reads with the command
758 samtools view -b -q 10 (Li et al., 2009). Due to the format of bowtie2's output, running Hornet on all
759 reads at once was excessively RAM intensive. Therefore, we split the bam files by chromosome and ran
760 Hornet on each of the chromosomes separately. We used the files of SNVs and indels generated as
761 described above as input to Hornet. After Hornet finished running, we used samtools merge to merge all

762 autosomes and sex chromosomes (we excluded the mitochondrial genome) to create a final bam file for
763 downstream analysis (Li et al., 2009).
764
765 **Peak calling and filtering**
766 As only one replicate was available for SKM and HP, we generated two pseudo-replicates by randomly
767 assigning reads to one of two files using Picard SplitSamByNumberOfReads (Broad Institute, n.d.). We
768 then called peaks on each file separately, as well as a merged file containing all the reads from a
769 particular cell type. For example, for MN, both replicates were pooled and peaks were called on that file
770 as well as the two replicates separately. Before peak calling, all bam files were converted to bed files.
771 We called peaks using MACS2 callpeak with the following arguments: -f BED -p 0.01 --nomodel --shift 75
772 --extsize 150 -B --SPMR --keep-dup all --call-summits (Y. Zhang et al., 2008, p. 2). We called peaks on
773 both the chimpanzee-referenced and human-referenced bam files. After peak calling, we sought to filter
774 peaks using a modified version of the ENCODE pipeline designed to eliminate peaks that lack a one-to-
775 one ortholog between humans and chimpanzees. The following pipeline was run on each cell type
776 separately. We first filtered peaks that were not called in both replicates as well as the pooled file using
777 code from the ENCODE pipeline based on bedtools and awk (Quinlan & Hall, 2010). We then used a
778 custom Python script to merge overlapping peaks and used UCSC LiftOver to lift the peaks from hg38 to
779 the panTro6 and back to hg38 as well as from panTro6 to hg38 (Kuhn et al., 2013). We then used
780 bedtools to intersect the resulting human referenced files and filtered out any peaks that did not have at
781 least 25% overlap with a peak in the other file (Quinlan & Hall, 2010). After filtering out peaks
782 overlapping ENCODE blacklisted regions and merging overlapping peaks again, we lifted the file that was
783 originally chimpanzee-referenced back to the chimpanzee genome (Amemiya et al., 2019). Finally, we
784 removed human-referenced peaks if their chimpanzee-referenced counterpart failed to lift over (Kuhn

785 et al., 2013). As only a relatively small number of peaks failed to lift over (e.g. because the region was
786 split in the new genome), any peaks that failed to lift over were excluded.
787
788 **Annotating the peak lists**
789 To annotate the peaks, we used the list of TSSs defined by Horlbeck et al. to annotate peaks (Horlbeck et
790 al., 2016). We lifted over each TSS to hg38, expanded 1000 bases on either side of the midpoint of each
791 TSS to generate promoters, and merged any promoters that overlapped while retaining all unique gene
792 names associated with the promoter (Kuhn et al., 2013). We then used reciprocal LiftOver with panTro6
793 to filter out non-orthologous promoters and used bedtools intersect to link peaks to promoters and
794 expanded the peak to include the entirety of the promoter if necessary (Quinlan & Hall, 2010). Through
795 this process, we also compiled a list of non-promoter CREs (sometimes labeled as enhancers as
796 enhancers are thought to be the most common type of CRE). We used bedtools closest to link these
797 non-promoter CREs to the two closest protein coding genes (Quinlan & Hall, 2010). Notably, the gene
798 naming conventions differ for the Horlbeck et al. TSS list and the GTF file used for RNA-seq processing.
799 We altered all gene names in peaks to match those found in the GTF file. In some cases, the gene no
800 longer existed in the updated hg38 GTF in which case the gene name was replaced with NAN.
801
802 **Generating a unified peak list**
803 We next merged our cell type-specific peak list across all five cell types to create a unified peak list. To
804 do this, we iteratively intersected all the peaks with bedtools and then merged any overlapping peaks
805 (Quinlan & Hall, 2010). Finally, we added back any peaks that did not intersect a peak found in any other
806 cell types. We then took the chimpanzee and human-referenced versions of these peak lists and ran
807 them through the LiftOver-based non-homologous peak filtering pipeline described above to generate a
808 final file of all identified peaks as well as which cell type (s) they were called in (Kuhn et al., 2013). Then,

809 we reran the annotation pipeline described in 'Annotating the peak lists' on this new set of peaks. In
810 total, this process resulted in 251,669 ATAC peaks.
811
812 **Counting reads in peaks and further peak filtering**
813 First, we split the bam files into reads that we could confidently assign to the chimpanzee genome and
814 reads we could assign to the human genome. We used our bed file of high-confidence SNVs and
815 required at least one SNV matching the human genome as well as no SNVs matching the chimpanzee
816 genome for a read to be assigned to human (and vice versa for chimp). We then used a custom Python
817 script to reformat the peak list bed files as GTF files and used HTSeq to count reads in peaks using the
818 following parameters: -s no -m union -r pos (Anders et al., 2015). We only kept peaks if they had a mean
819 read count across replicates within a cell type of at least 25 from either allele. For example, if a peak has
820 an average of 27 reads from the human allele and an average of 10 reads from the chimpanzee allele in
821 MN, that peak would be kept in MN. On the other hand, if the same peak had an average of 24 reads
822 from the human allele and an average of 10 reads from the chimpanzee allele in CM, that peak would be
823 discarded for CM.
824
825 We next filtered the reads to remove peaks that might be differentially accessible but show evidence of
826 mapping bias or do not agree between replicates. To do this, we removed any peaks with an absolute
827 $\log_2$ fold-change greater than one in one replicate but with a fold-change of any magnitude in the
828 opposite direction in the other. This was not done for SKM or HP as we had only one replicate. We then
829 removed any peaks that had a log fold-change in opposite directions with an absolute difference greater
830 than 1 in at least one replicate when comparing the human-referenced and chimpanzee-referenced
831 counts. Finally, as described in section ''Detection of aneuploidy on chromosome 20 and slight
832 chimpanzee parental contamination in PP hybrid2 samples' for RNA-seq data analysis, we removed any

833 peaks on chr20 and took this as our final list of peaks for downstream analyses. Allelic counts were
834 normalized as described in the RNA-seq data analysis. We tested for allele-specific chromatin
835 accessibility (ASCA) using the binomial test applied to the normalized allelic counts (summed by species
836 within a cell type). We considered any peaks with a binomial p-value less than 0.05 to be nominally
837 differentially accessible. After this filtering, we retained 76,360 peaks for additional analysis. Using the
838 set of peaks that passed filtering for each cell type, we plotted the total number of promoter and non-
839 promoter peaks that passed filtering in each cell type as well as the number of peaks per gene.
840

841 **Down-sampling to identify cell type-specific ATAC-seq peaks**

842 As the number of peaks detected by ATAC-seq is generally a function of read depth and our read depth
843 varied widely across cell types, we restricted to one replicate (always hybrid1 if two replicates were
844 available) and down-sampled reads to match the SKM sample with lowest sequencing depth. We then
845 called peaks for cell types with a single ATAC replicate as described above.
846

847 **Allelic chromatin accessibility tracks**

848 Allelic bam files with reads originating from the human allele and the chimpanzee allele (respectively)
849 were obtained as described in 'Counting reads in peaks and further peak filtering'. Two replicates in CM,
850 MN, and PP were pooled by cell type. Bam files were converted into bigWig files using the python
851 package deepTools bamCoverage with options: --binSize 1 --normalizeUsing CPM --effectiveGenomeSize
852 2862010578 --ignoreForNormalization chr20 --extendReads (Ramírez et al., 2014). Tracks were
853 visualized and plotted using the python package pyGenomeTracks (Lopez-Delisle et al., 2021). When
854 comparing human and chimpanzee log fold-change track differences in each cell type, deepTools
855 bigwigCompare was used to compare between human bigWig and chimpanzee bigWig with options: --
856 pseudocount 1 --skipZeroOverZero --operation log2 -bs 1 (Ramírez et al., 2014).

857

858 **ChromHMM annotation and correlation with ASE**

859 A universal chromHMM annotation was obtained for each peak based on overlap with any of the 15
860 categories in chromHMM (excluding the blacklist category, for which peaks had already been removed)
861 (Ernst & Kellis, 2017; Vu & Ernst, 2022). Divergence was measured as the z-score transformed median of
862 the absolute $\log_2$ fold-change of human and chimpanzee normalized counts in each peak. Each peak was
863 assigned to the closest gene and then Pearson correlation was computed between the chromatin
864 accessibility $\log_2$ fold-change and the expression $\log_2$ fold-change for each peak and its nearest gene.
865 Pearson correlation was computed only on categories including at least 15 peaks. When showing results
866 for differentially accessible peaks, only peaks with binomial p-values less than 0.05 were kept and used
867 in computing the Pearson correlation. When assigning a unique chromHMM to each peak, the
868 chromHMM category that covered the largest portion of each peak was used. When filtering out
869 promoter-related annotations, peaks covering any promoter-related chromHMM categories ("TSS",
870 "flanking promoter" and "bivalent promoter") were filtered out and the analysis described above was
871 repeated. Of the 76,630 peaks retained, 76,221 overlapped at least one chromHMM annotation. We
872 recomputed these correlations (with promoters excluded) after splitting peaks into those less than 30
873 kilobases from the nearest TSS and peaks greater than or equal to 30 kilobases from the nearest TSS. We
874 also recomputed the correlations for cell type-specifically and ubiquitously expressed genes separately.

875

876 **Testing for influence of number of SNVs on p-values and log fold-change**

877 To assign SNVs to genes, we converted all exons in the human gtf file to bed format and intersected the
878 exons with our list of confident human-chimpanzee SNVs. We then used a custom python script to count
879 the number of unique SNVs assigned to each gene and plotted the relationship between DESeq2 p-value
880 and log fold-change estimate. For the ATAC peaks, we performed an identical procedure using the bed

881 file of peaks rather than a bed file of exons. We then plotted the number of SNVs in a peak with the
882 binomial test p-value and raw log fold-change estimate.
883
884 **Computation of differential expression enrichment (dEE) and differential chromatin accessibility**
885 **enrichment (dCAE)**
886 For each target cell type, taking CM as an example, the $\log_2$ fold-change for gene A was fixed as target
887 $\log_2$ fold-change, and the $\log_2$ fold-changes for gene A in the remaining cell types with an opposing sign
888 (compared to the target $\log_2$ fold-change) were set to zero. Then, the dEE value was calculated as the
889 proportion of the target $\log_2$ fold-change in the sum of the zeroed log fold-changes across all cell types.
890 For example, the dEE for gene A in CM would be abs (target LFC)/sum (abs(LFC after zeroing)). dEE
891 ranges from zero to one and low dEE value indicates differential expression with similar magnitude and
892 direction across cell types, and/or the gene does not have any strong allelic bias, whereas a high dEE
893 value indicates that this gene is only strongly differentially expressed (with the sign the $\log_2$ fold-change
894 has in that cell type) in a particular cell type. dCAE uses the same procedure as dEE except the table is
895 populated with the $\log_2$ fold-changes derived from chromatin accessibility measurements. dEE and dCAE
896 are sensitive to the inclusion or exclusion of cell types (by definition), so we excluded RPE when
897 integrative analysis combining results from dEE and dCAE was performed (to match the cell types for
898 which dCAE could be computed, Supp Fig. 33b). After restricting to genes defined as having significant
899 ASE or significant ASCA, we defined genes with dEE >= 0.75 in a particular cell type as showing cell type-
900 specific ASE and peaks with dCAE >= 0.75 in one cell type as showing cell type-specific ASCA. We used
901 bedtools intersect to intersect the peaks with our list of human-chimpanzee single nucleotide
902 differences and the 241-way placental mammal PhyloP scores (Quinlan & Hall, 2010; Sullivan et al.,
903 2023). We also checked whether peaks that contained human-chimpanzee differences in sites with high
904 PhyloP scores were in the list of HARs described in Girskis et al (Girskis et al., 2021).

905

906 **Predicting regulatory activity with single-variant resolution**

907 We used sequences in fasta format as input to the deep neural network model Sei (K. M. Chen et al.,
908 2022). Sei requires a 4096 base pair input sequence, so we put the center of our region of interest at the
909 center of the input window and expanded equally on either side to contain 4096 base pairs. The human
910 sequence was retrieved from hg38 and the corresponding chimpanzee sequence was retrieved from
911 panTro6. The effect size when comparing the probabilities of each sequence having a particular
912 chromatin state was computed as the log of the human sequence probability divided by the chimpanzee
913 sequence probability. Only annotations for which either the chimpanzee sequence or the human
914 sequence had a probability value greater than or equal to 0.5 were kept for downstream analysis. All
915 SNVs between human and chimpanzee in this input window were identified and ordered based on
916 coordinates. For each SNV position, the human sequence was changed to the chimpanzee allele at that
917 position to generate a new sequence that was input to Sei. The $\log_2$ fold-change for each chromatin
918 annotation was computed for each input sequence as described above and used as a measure of the
919 effect of this change on the sequence. Similarly, an indel can be introduced to modify the human
920 sequence and input to Sei. With indels, the center of the regions of interest (promoter or HAR) were
921 always at the center of the input window and the start or end of the sequence inputted to Sei could
922 possibly lose or gain base pairs. However, we found that for the small indels shown here this had
923 essentially no effect on the Sei output.

924

925 **Processing of publicly available datasets**

926 The data from Blake et al. and Pavlovic et al. were processed as previously described (Blake et al., 2020;
927 Pavlovic et al., 2018; Starr et al., 2023). For the Pavlovic et al. data, $\log_2$ fold-changes were computed in
928 DESeq2 with the scaled proportion of cardiomyocytes present in each sample (available in the

929 supplemental materials of Pavlovic et al.), sex, and whether cardiomyocytes were treated with T3 as
930 covariates (i.e. using the model ~sex+scaled_proportion_cardiomyocytes+T3_Treatment+species)
931 (Pavlovic et al., 2018). No covariates were included for Blake et al. as they had little impact on the data
932 (Blake et al., 2020). The $\log_2$ fold-changes and FDR corrected p-values were directly downloaded from
933 the supplemental materials of Kozlenkov et al. (Kozlenkov et al., 2020).
934
935 The processed data from Ma et al. (Ma et al., 2022) were downloaded from
936 http://resources.sestanlab.org/PFC/. We pseudobulked the data by cell type by summing counts within
937 each individual. We then separately input each pairwise comparison of two species (human to
938 chimpanzee or human to rhesus macaque) into DESeq2 with no covariates to test for differential
939 expression and compute $\log_2$ fold-changes.
940
941 The counts tables from Kanton et al.[12] were downloaded from
942 https://www.ebi.ac.uk/biostudies/arrayexpress/studies/E-MTAB-7552 and processed with SCANPY
943 (Kanton et al., 2019, p. 29409532; Wolf et al., 2018). The data were filtered by removing cells with
944 n_genes_by_counts > 2500 and >5% mitochondrial reads. We also removed cells with fewer than 200
945 unique genes and genes that had non-zero counts in fewer than 3 cells. After filtering, any chimpanzee
946 cells not falling in the category (defined by Kanton et al. (Kanton et al., 2019)) “ventral forebrain
947 progenitors and neurons” were eliminated and human cells not in the categories "ventral progenitors
948 and neurons 1", "ventral progenitors and neurons 2", or "ventral progenitors and neurons 3" were
949 similarly eliminated. We then merged the two counts tables, normalized/logarithmized the counts,
950 computed PCA, used harmony (Korsunsky et al., 2019) to integrate cells from different species (human
951 and chimpanzee), and found nearest neighbors with the harmonized principal components. We then ran

952 Leiden clustering with resolution = 0.5 to identify 7 subclusters (one of which appeared to be a technical
953 artifact with very low counts that was removed) (Traag et al., 2019).
954
955 We identified cell types and lineages using canonical marker genes (*MKI67* and *HES5* for progenitors,
956 *NKX2-1* and *LHX6* for the medial ganglionic eminence or MGE, *MEIS2* and *ZFHX3* for the lateral
957 ganglionic eminence or LGE, and *SCGN* and *NR2F1* for the caudal ganglionic eminence or CGE) (Su-Feher
958 et al., 2022). We then used the implementation of PAGA in SCANPY to compute pseudotime using the
959 first cell in the progenitor subcluster as the root (Wolf et al., 2019). We binned cells into five equal bins
960 along pseudotime and compared the expression of cells with non-zero counts for *GAD1* in each
961 pseudotime bin. Within each bin, we used a Wilcoxon test to test for higher expression of *GAD1* in
962 chimpanzee cells compared to human cells. We repeated the pseudotime analysis, binning, and
963 comparing of *GAD1* gene expression for each subtrajectory (MGE, LGE, and CGE).
964
965 **Description of Additional file 1:** This file contains an extended evaluation of the success of the
966 differentiations in generating the desired cell type and other cell types likely to be present in each
967 sample.
968
969 **Description of Additional file 2:** This file contains the results of running the test for selection described
970 by Starr et al (2023) on each cell type.
971
972 **Description of Additional file 3:** This file contains the peak-gene pairs that had concordant high dCAE
973 and high dEE in each cell type.
974
975 **Declarations:** The authors have nothing to declare.

Ethics approval and consent to participate: Not relevant to our study.

Consent for publication: Not relevant to our study.

Availability of data and materials: Raw and processed data generated by this study are publicly available through the Gene Expression Omnibus under accession GSE232949: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE232949. The snRNAseq data from Ma et al. are available here: http://resources.sestanlab.org/PFC/. The scRNAseq data from Kanton et al. are available here: https://www.ebi.ac.uk/biostudies/arrayexpress/studies/E-MTAB-7552. The bulk RNA-seq data from Blake et al. and Pavlovic et al. are available here: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE112356 and here: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE110471 respectively. The log fold-changes and associated statistics were used directly from the supplemental materials of Kozlenkov et al. The human-chimp pairwise alignment used to identify SNVs and indels is available here: https://hgdownload.soe.ucsc.edu/goldenPath/hg38/vsPanTro6/. The human and chimpanzee genomes used are available here: https://hgdownload.soe.ucsc.edu/goldenPath/hg38/bigZips/ and here: https://hgdownload.soe.ucsc.edu/goldenPath/panTro6/bigZips/ respectively. The 241-way PhyloP scores were downloaded from here: https://hgdownload.soe.ucsc.edu/goldenPath/hg38/cactus241way/. The “QCed genomic constraint by 1kb regions” from the gnomAD consortium are available here: https://gnomad.broadinstitute.org/downloads#v3. All scripts for performing analyses and making figures in this manuscript is publicly available at https://github.com/banwang27/multi-celltypes.

Competing interests: The authors have no competing interests to declare.

Funding: Funding for this work came from NIH R01HG012285. A.L.S was supported by an NDSEG fellowship.

Author contributions: HBF conceived of the study. ALS and BW performed all analysis, visualization, validation, and writing of software. ALS wrote the manuscript with input from BW and HBF. BW made all figures with input from ALS and HBF. All authors approved the publication of the manuscript.

Acknowledgements: The authors wish to acknowledge the Columbia Stem Cell Core for their hard work in differentiating these cells. We also acknowledge members of the Fraser lab past and present for helpful discussion. Figure 1a, 2a, and 3a were created with https://www.biorender.com/.

1013 Agoglia, R. M., Sun, D., Birey, F., Yoon, S.-J., Miura, Y., Sabatini, K., Pașca, S. P., & Fraser, H.
1014 B. (2021). Primate cell fusion disentangles gene regulatory divergence in
1015 neurodevelopment. *Nature*, *592*(7854), 421–427. https://doi.org/10.1038/s41586-021-
1016 03343-3
1017 Akbar, M., Calderon, F., Wen, Z., & Kim, H.-Y. (2005). Docosahexaenoic acid: A positive
1018 modulator of Akt signaling in neuronal survival. *Proceedings of the National Academy of*
1019 *Sciences*, *102*(31), 10858–10863. https://doi.org/10.1073/pnas.0502903102
1020 Amemiya, H. M., Kundaje, A., & Boyle, A. P. (2019). The ENCODE Blacklist: Identification of
1021 Problematic Regions of the Genome. *Scientific Reports*, *9*(1), 9354.
1022 https://doi.org/10.1038/s41598-019-45839-z
1023 Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq—A Python framework to work with high-
1024 throughput sequencing data. *Bioinformatics (Oxford, England)*, *31*(2), 166–169.
1025 https://doi.org/10.1093/bioinformatics/btu638
1026 Arai, Y., Funatsu, N., Numayama-Tsuruta, K., Nomura, T., Nakamura, S., & Osumi, N. (2005).
1027 Role of Fabp7, a downstream gene of Pax6, in the maintenance of neuroepithelial cells
1028 during early embryonic development of the rat cortex. *The Journal of Neuroscience: The*
1029 *Official Journal of the Society for Neuroscience*, *25*(42), 9752–9761.
1030 https://doi.org/10.1523/JNEUROSCI.2512-05.2005
1031 Ben-Ari, Y., Khalilov, I., Kahle, K. T., & Cherubini, E. (2012). The GABA Excitatory/Inhibitory
1032 Shift in Brain Maturation and Neurological Disorders. *The Neuroscientist*, *18*(5), 467–
1033 486. https://doi.org/10.1177/1073858412438697
1034 Benito-Kwiecinski, S., Giandomenico, S. L., Sutcliffe, M., Riis, E. S., Freire-Pritchett, P., Kelava,
1035 I., Wunderlich, S., Martin, U., Wray, G. A., McDole, K., & Lancaster, M. A. (2021). An
1036 early cell shape transition drives evolutionary expansion of the human forebrain. *Cell*,
1037 *184*(8), 2084-2102.e19. https://doi.org/10.1016/j.cell.2021.02.050

1038 Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and
1039 Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B*
1040 *(Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x
1041 Blake, L. E., Roux, J., Hernando-Herraez, I., Banovich, N. E., Perez, R. G., Hsiao, C. J., Eres, I.,
1042 Cuevas, C., Marques-Bonet, T., & Gilad, Y. (2020). A comparison of gene expression
1043 and DNA methylation patterns across tissues and species. *Genome Research*, *30*(2),
1044 250–262. https://doi.org/10.1101/gr.254904.119
1045 Broad Institute. (n.d.). *Picard*.
1046 Brose, N., Brunger, A., Cafiso, D., Chapman, E. R., Diao, J., Hughson, F. M., Jackson, M. B.,
1047 Jahn, R., Lindau, M., Ma, C., Rizo, J., Shin, Y.-K., Söllner, T. H., Tamm, L., Yoon, T.-Y.,
1048 & Zhang, Y. (2019). Synaptic vesicle fusion: Today and beyond. *Nature Structural &*
1049 *Molecular Biology*, *26*(8), 663–668. https://doi.org/10.1038/s41594-019-0277-z
1050 Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y., & Greenleaf, W. J. (2013).
1051 Transposition of native chromatin for fast and sensitive epigenomic profiling of open
1052 chromatin, DNA-binding proteins and nucleosome position. *Nature Methods*, *10*(12),
1053 1213–1218. https://doi.org/10.1038/nmeth.2688
1054 Burridge, P. W., Matsa, E., Shukla, P., Lin, Z. C., Churko, J. M., Ebert, A. D., Lan, F., Diecke,
1055 S., Huber, B., Mordwinkin, N. M., Plews, J. R., Abilez, O. J., Cui, B., Gold, J. D., & Wu, J.
1056 C. (2014). Chemically defined generation of human cardiomyocytes. *Nature Methods*,
1057 *11*(8), 855–860. https://doi.org/10.1038/nmeth.2999
1058 Calderari, S., Ria, M., Gérard, C., Nogueira, T. C., Villate, O., Collins, S. C., Neil, H., Gervasi,
1059 N., Hue, C., Suarez-Zamorano, N., Prado, C., Cnop, M., Bihoreau, M.-T., Kaisaki, P. J.,
1060 Cazier, J.-B., Julier, C., Lathrop, M., Werner, M., Eizirik, D. L., & Gauguier, D. (2018).
1061 Molecular genetics of the transcription factor GLIS3 identifies its dual function in beta
1062 cells and neurons. *Genomics*, *110*(2), 98–111.
1063 https://doi.org/10.1016/j.ygeno.2017.09.001

1064 Caporali, A., & Emanueli, C. (2009). Cardiovascular actions of neurotrophins. *Physiological*
1065 *Reviews*, *89*(1), 279–308. https://doi.org/10.1152/physrev.00007.2008
1066 Castel, S. E., Aguet, F., Mohammadi, P., GTEx Consortium, Aguet, F., Anand, S., Ardlie, K. G.,
1067 Gabriel, S., Getz, G. A., Graubert, A., Hadley, K., Handsaker, R. E., Huang, K. H.,
1068 Kashin, S., Li, X., MacArthur, D. G., Meier, S. R., Nedzel, J. L., Nguyen, D. T., …
1069 Lappalainen, T. (2020). A vast resource of allelic expression data spanning human
1070 tissues. *Genome Biology*, *21*(1), 234. https://doi.org/10.1186/s13059-020-02122-z
1071 Castro-Mondragon, J. A., Riudavets-Puig, R., Rauluseviciute, I., Lemma, R. B., Turchi, L.,
1072 Blanc-Mathieu, R., Lucas, J., Boddie, P., Khan, A., Manosalva Pérez, N., Fornes, O.,
1073 Leung, T. Y., Aguirre, A., Hammal, F., Schmelter, D., Baranasic, D., Ballester, B.,
1074 Sandelin, A., Lenhard, B., … Mathelier, A. (2022). JASPAR 2022: The 9th release of the
1075 open-access database of transcription factor binding profiles. *Nucleic Acids Research*,
1076 *50*(D1), D165–D173. https://doi.org/10.1093/nar/gkab1113
1077 Chal, J., Al Tanoury, Z., Hestin, M., Gobert, B., Aivio, S., Hick, A., Cherrier, T., Nesmith, A. P.,
1078 Parker, K. K., & Pourquié, O. (2016). Generation of human muscle fibers and satellite-
1079 like cells from human pluripotent stem cells in vitro. *Nature Protocols*, *11*(10), 1833–
1080 1850. https://doi.org/10.1038/nprot.2016.110
1081 Chen, K. M., Wong, A. K., Troyanskaya, O. G., & Zhou, J. (2022). A sequence-based global
1082 map of regulatory activity for deciphering human genetics. *Nature Genetics*, *54*(7), 940–
1083 949. https://doi.org/10.1038/s41588-022-01102-2
1084 Chen, S., Francioli, L. C., Goodrich, J. K., Collins, R. L., Kanai, M., Wang, Q., Alföldi, J., Watts,
1085 N. A., Vittal, C., Gauthier, L. D., Poterba, T., Wilson, M. W., Tarasova, Y., Phu, W.,
1086 Yohannes, M. T., Koenig, Z., Farjoun, Y., Banks, E., Donnelly, S., … Karczewski, K. J.
1087 (2022). *A genome-wide mutational constraint map quantified from variation in 76,156*
1088 *human genomes* [Preprint]. Genetics. https://doi.org/10.1101/2022.03.20.485034

1089 Choi, W.-S., Xu, X., Goruk, S., Wang, Y., Patel, S., Chow, M., Field, C. J., & Godbout, R.
1090 (2021). FABP7 Facilitates Uptake of Docosahexaenoic Acid in Glioblastoma Neural
1091 Stem-like Cells. *Nutrients*, *13*(8), 2664. https://doi.org/10.3390/nu13082664
1092 Collins, R. L., Glessner, J. T., Porcu, E., Lepamets, M., Brandon, R., Lauricella, C., Han, L.,
1093 Morley, T., Niestroj, L.-M., Ulirsch, J., Everett, S., Howrigan, D. P., Boone, P. M., Fu, J.,
1094 Karczewski, K. J., Kellaris, G., Lowther, C., Lucente, D., Mohajeri, K., … Talkowski, M.
1095 E. (2022). A cross-disorder dosage sensitivity map of the human genome. *Cell*, *185*(16),
1096 3041-3055.e25. https://doi.org/10.1016/j.cell.2022.06.036
1097 Combs, P. A., Krupp, J. J., Khosla, N. M., Bua, D., Petrov, D. A., Levine, J. D., & Fraser, H. B.
1098 (2018). Tissue-Specific cis-Regulatory Divergence Implicates eloF in Inhibiting
1099 Interspecies Mating in Drosophila. *Current Biology*, *28*(24), 3969-3975.e3.
1100 https://doi.org/10.1016/j.cub.2018.10.036
1101 Conway, J. R., Lex, A., & Gehlenborg, N. (2017). UpSetR: An R package for the visualization of
1102 intersecting sets and their properties. *Bioinformatics*, *33*(18), 2938–2940.
1103 https://doi.org/10.1093/bioinformatics/btx364
1104 Corces, M. R., Trevino, A. E., Hamilton, E. G., Greenside, P. G., Sinnott-Armstrong, N. A.,
1105 Vesuna, S., Satpathy, A. T., Rubin, A. J., Montine, K. S., Wu, B., Kathiria, A., Cho, S.
1106 W., Mumbach, M. R., Carter, A. C., Kasowski, M., Orloff, L. A., Risca, V. I., Kundaje, A.,
1107 Khavari, P. A., … Chang, H. Y. (2017). An improved ATAC-seq protocol reduces
1108 background and enables interrogation of frozen tissues. *Nature Methods*, *14*(10), 959–
1109 962. https://doi.org/10.1038/nmeth.4396
1110 De Rosa, A., Pellegatta, S., Rossi, M., Tunici, P., Magnoni, L., Speranza, M. C., Malusa, F.,
1111 Miragliotta, V., Mori, E., Finocchiaro, G., & Bakker, A. (2012). A radial glia gene marker,
1112 fatty acid binding protein 7 (FABP7), is involved in proliferation and invasion of
1113 glioblastoma cells. *PloS One*, *7*(12), e52113.
1114 https://doi.org/10.1371/journal.pone.0052113

1115 Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson,
1116 M., & Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*
1117 *(Oxford, England)*, *29*(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635
1118 Ebrahimi, M., Yamamoto, Y., Sharifi, K., Kida, H., Kagawa, Y., Yasumoto, Y., Islam, A.,
1119 Miyazaki, H., Shimamoto, C., Maekawa, M., Mitsushima, D., Yoshikawa, T., & Owada,
1120 Y. (2016). Astrocyte-expressed FABP7 regulates dendritic morphology and excitatory
1121 synaptic function of cortical neurons. *Glia*, *64*(1), 48–62.
1122 https://doi.org/10.1002/glia.22902
1123 Ernst, J., & Kellis, M. (2017). Chromatin-state discovery and genome annotation with
1124 ChromHMM. *Nature Protocols*, *12*(12), 2478–2492.
1125 https://doi.org/10.1038/nprot.2017.124
1126 Feldblum, S., Erlander, M. G., & Tobin, A. J. (1993). Different distributions of GAD65 and
1127 GAD67 mRNAS suggest that the two glutamate decarboxylases play distinctive
1128 functional roles. *Journal of Neuroscience Research*, *34*(6), 689–706.
1129 https://doi.org/10.1002/jnr.490340612
1130 Field, A. R., Jacobs, F. M. J., Fiddes, I. T., Phillips, A. P. R., Reyes-Ortiz, A. M., LaMontagne,
1131 E., Whitehead, L., Meng, V., Rosenkrantz, J. L., Olsen, M., Hauessler, M., Katzman, S.,
1132 Salama, S. R., & Haussler, D. (2019). Structurally Conserved Primate LncRNAs Are
1133 Transiently Expressed during Human Cortical Differentiation and Influence Cell-Type-
1134 Specific Genes. *Stem Cell Reports*, *12*(2), 245–257.
1135 https://doi.org/10.1016/j.stemcr.2018.12.006
1136 Fraser, H. B. (2011). Genome-wide approaches to the study of adaptive gene expression
1137 evolution: Systematic studies of evolutionary adaptations involving gene expression will
1138 allow many fundamental questions in evolutionary biology to be addressed. *BioEssays*,
1139 *33*(6), 469–477. https://doi.org/10.1002/bies.201000094

1140 Fraser, H. B. (2013). Gene expression drives local adaptation in humans. *Genome Research*,
1141 *23*(7), 1089–1096. https://doi.org/10.1101/gr.152710.112
1142 Fraser, H. B. (2019). Improving Estimates of Compensatory cis–trans Regulatory Divergence.
1143 *Trends in Genetics*, *35*(1), 3–5. https://doi.org/10.1016/j.tig.2018.09.003
1144 García-Pérez, R., Esteller-Cucala, P., Mas, G., Lobón, I., Di Carlo, V., Riera, M., Kuhlwilm, M.,
1145 Navarro, A., Blancher, A., Di Croce, L., Gómez-Skarmeta, J. L., Juan, D., & Marquès-
1146 Bonet, T. (2021). Epigenomic profiling of primate lymphoblastoid cell lines reveals the
1147 evolutionary patterns of epigenetic activities in gene regulatory architectures. *Nature*
1148 *Communications*, *12*(1), 3116. https://doi.org/10.1038/s41467-021-23397-1
1149 Girskis, K. M., Stergachis, A. B., DeGennaro, E. M., Doan, R. N., Qian, X., Johnson, M. B.,
1150 Wang, P. P., Sejourne, G. M., Nagy, M. A., Pollina, E. A., Sousa, A. M. M., Shin, T.,
1151 Kenny, C. J., Scotellaro, J. L., Debo, B. M., Gonzalez, D. M., Rento, L. M., Yeh, R. C.,
1152 Song, J. H. T., … Walsh, C. A. (2021). Rewiring of human neurodevelopmental gene
1153 regulatory programs by human accelerated regions. *Neuron*, *109*(20), 3239-3251.e7.
1154 https://doi.org/10.1016/j.neuron.2021.08.005
1155 Gokhman, D., Agoglia, R. M., Kinnebrew, M., Gordon, W., Sun, D., Bajpai, V. K., Naqvi, S.,
1156 Chen, C., Chan, A., Chen, C., Petrov, D. A., Ahituv, N., Zhang, H., Mishina, Y.,
1157 Wysocka, J., Rohatgi, R., & Fraser, H. B. (2021). Human–chimpanzee fused cells reveal
1158 cis-regulatory divergence underlying skeletal evolution. *Nature Genetics*, *53*(4), 467–
1159 476. https://doi.org/10.1038/s41588-021-00804-3
1160 GTEx Consortium. (2017). Genetic effects on gene expression across human tissues. *Nature*,
1161 *550*(7675), 204–213. https://doi.org/10.1038/nature24277
1162 Horlbeck, M. A., Gilbert, L. A., Villalta, J. E., Adamson, B., Pak, R. A., Chen, Y., Fields, A. P.,
1163 Park, C. Y., Corn, J. E., Kampmann, M., & Weissman, J. S. (2016). Compact and highly
1164 active next-generation libraries for CRISPR-mediated gene repression and activation.
1165 *ELife*, *5*, e19760. https://doi.org/10.7554/eLife.19760

1166 Hu, C. K., York, R. A., Metz, H. C., Bedford, N. L., Fraser, H. B., & Hoekstra, H. E. (2022). Cis-
1167 Regulatory changes in locomotor genes are associated with the evolution of burrowing
1168 behavior. *Cell Reports*, *38*(7), 110360. https://doi.org/10.1016/j.celrep.2022.110360
1169 Huang, E. J., & Reichardt, L. F. (2001). Neurotrophins: Roles in neuronal development and
1170 function. *Annual Review of Neuroscience*, *24*, 677–736.
1171 https://doi.org/10.1146/annurev.neuro.24.1.677
1172 Hubisz, M. J., & Pollard, K. S. (2014). Exploring the genesis and functions of Human
1173 Accelerated Regions sheds light on their role in human evolution. *Current Opinion in*
1174 *Genetics & Development*, *29*, 15–21. https://doi.org/10.1016/j.gde.2014.07.005
1175 Ichim, G., Tauszig-Delamasure, S., & Mehlen, P. (2012). Neurotrophins and cell death.
1176 *Experimental Cell Research*, *318*(11), 1221–1228.
1177 https://doi.org/10.1016/j.yexcr.2012.03.006
1178 Jain, A., & Tuteja, G. (2019). TissueEnrich: Tissue-specific gene enrichment analysis.
1179 *Bioinformatics (Oxford, England)*, *35*(11), 1966–1967.
1180 https://doi.org/10.1093/bioinformatics/bty890
1181 John St. John. (n.d.). *SeqPrep*.
1182 Kanton, S., Boyle, M. J., He, Z., Santel, M., Weigert, A., Sanchís-Calleja, F., Guijarro, P., Sidow,
1183 L., Fleck, J. S., Han, D., Qian, Z., Heide, M., Huttner, W. B., Khaitovich, P., Pääbo, S.,
1184 Treutlein, B., & Camp, J. G. (2019). Organoid single-cell genomic atlas uncovers human-
1185 specific features of brain development. *Nature*, *574*(7778), 418–422.
1186 https://doi.org/10.1038/s41586-019-1654-9
1187 Ke, K., Song, Y., Shen, J., Niu, M., Zhang, H., Yuan, D., Ni, H., Zhang, Y., Liu, X., Dai, A., &
1188 Cao, M. (2015). Up-Regulation of Glis2 Involves in Neuronal Apoptosis After
1189 Intracerebral Hemorrhage in Adult Rats. *Cellular and Molecular Neurobiology*, *35*(3),
1190 345–354. https://doi.org/10.1007/s10571-014-0130-1

Kelley, J. L., & Gilad, Y. (2020). Effective study design for comparative functional genomics. *Nature Reviews Genetics*, *21*(7), 385–386. https://doi.org/10.1038/s41576-020-0242-z

King, M.-C., & Wilson, A. C. (1975). Evolution at Two Levels in Humans and Chimpanzees: Their macromolecules are so alike that regulatory mutations may account for their biological differences. *Science*, *188*(4184), 107–116. https://doi.org/10.1126/science.1090005

Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P., & Raychaudhuri, S. (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. *Nature Methods*, *16*(12), 1289–1296. https://doi.org/10.1038/s41592-019-0619-0

Korytnikov, R., & Nostro, M. C. (2016). Generation of polyhormonal and multipotent pancreatic progenitor lineages from human pluripotent stem cells. *Methods*, *101*, 56–64. https://doi.org/10.1016/j.ymeth.2015.10.017

Kozlenkov, A., Vermunt, M. W., Apontes, P., Li, J., Hao, K., Sherwood, C. C., Hof, P. R., Ely, J. J., Wegner, M., Mukamel, E. A., Creyghton, M. P., Koonin, E. V., & Dracheva, S. (2020). Evolution of regulatory signatures in primate cortical neurons at cell-type resolution. *Proceedings of the National Academy of Sciences*, *117*(45), 28422–28432. https://doi.org/10.1073/pnas.2011884117

Kuhn, R. M., Haussler, D., & Kent, W. J. (2013). The UCSC genome browser and associated tools. *Briefings in Bioinformatics*, *14*(2), 144–161. https://doi.org/10.1093/bib/bbs038

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

1217 Liang, D., Elwell, A. L., Aygün, N., Krupa, O., Wolter, J. M., Kyere, F. A., Lafferty, M. J., Cheek,
1218 K. E., Courtney, K. P., Yusupova, M., Garrett, M. E., Ashley-Koch, A., Crawford, G. E.,
1219 Love, M. I., de la Torre-Ubieta, L., Geschwind, D. H., & Stein, J. L. (2021). Cell-type-
1220 specific effects of genetic variation on chromatin accessibility during human neuronal
1221 differentiation. *Nature Neuroscience*, *24*(7), 941–953. https://doi.org/10.1038/s41593-
1222 021-00858-w
1223 Lopez-Delisle, L., Rabbani, L., Wolff, J., Bhardwaj, V., Backofen, R., Grüning, B., Ramírez, F., &
1224 Manke, T. (2021). pyGenomeTracks: Reproducible plots for multivariate genomic
1225 datasets. *Bioinformatics (Oxford, England)*, *37*(3), 422–423.
1226 https://doi.org/10.1093/bioinformatics/btaa692
1227 Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and
1228 dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550.
1229 https://doi.org/10.1186/s13059-014-0550-8
1230 Ma, S., Skarica, M., Li, Q., Xu, C., Risgaard, R. D., Tebbenkamp, A. T. N., Mato-Blanco, X.,
1231 Kovner, R., Krsnik, Ž., de Martin, X., Luria, V., Martí-Pérez, X., Liang, D., Karger, A.,
1232 Schmidt, D. K., Gomez-Sanchez, Z., Qi, C., Gobeske, K. T., Pochareddy, S., … Sestan,
1233 N. (2022). Molecular and cellular evolution of the primate dorsolateral prefrontal cortex.
1234 *Science*, eabo7257. https://doi.org/10.1126/science.abo7257
1235 Mack, K. L., & Nachman, M. W. (2017). Gene Regulation and Speciation. *Trends in Genetics*,
1236 *33*(1), 68–80. https://doi.org/10.1016/j.tig.2016.11.003
1237 Mallanna, S. K., & Duncan, S. A. (2013). Differentiation of hepatocytes from pluripotent stem
1238 cells. *Current Protocols in Stem Cell Biology*, *26*, 1G.4.1-1G.4.13.
1239 https://doi.org/10.1002/9780470151808.sc01g04s26
1240 Maury, Y., Côme, J., Piskorowski, R. A., Salah-Mohellibi, N., Chevaleyre, V., Peschanski, M.,
1241 Martinat, C., & Nedelec, S. (2015). Combinatorial analysis of developmental cues

efficiently converts human pluripotent stem cells into multiple neuronal subtypes. *Nature Biotechnology*, *33*(1), 89–96. https://doi.org/10.1038/nbt.3049

Meisler, M. H., Hill, S. F., & Yu, W. (2021). Sodium channelopathies in neurodevelopmental disorders. *Nature Reviews Neuroscience*, *22*(3), 152–166. https://doi.org/10.1038/s41583-020-00418-4

Milani, P., Escalante-Chong, R., Shelley, B. C., Patel-Murray, N. L., Xin, X., Adam, M., Mandefro, B., Sareen, D., Svendsen, C. N., & Fraenkel, E. (2016). Cell freezing protocol suitable for ATAC-Seq on motor neurons derived from human induced pluripotent stem cells. *Scientific Reports*, *6*, 25474. https://doi.org/10.1038/srep25474

Mizuguchi, R., Kriks, S., Cordes, R., Gossler, A., Ma, Q., & Goulding, M. (2006). Ascl1 and Gsh1/2 control inhibitory and excitatory cell fate in spinal sensory interneurons. *Nature Neuroscience*, *9*(6), 770–778. https://doi.org/10.1038/nn1706

Montaigne, D., Butruille, L., & Staels, B. (2021). PPAR control of metabolism and cardiovascular functions. *Nature Reviews Cardiology*, *18*(12), 809–823. https://doi.org/10.1038/s41569-021-00569-6

Mueller, Fabian. (n.d.). *ChrAccR*.

Naito, Y., Lee, A. K., & Takahashi, H. (2017). Emerging roles of the neurotrophin receptor TrkC in synapse organization. *Neuroscience Research*, *116*, 10–17. https://doi.org/10.1016/j.neures.2016.09.009

Netherlands Brain Bank, Vermunt, M. W., Tan, S. C., Castelijns, B., Geeven, G., Reinink, P., de Bruijn, E., Kondova, I., Persengiev, S., Bontrop, R., Cuppen, E., de Laat, W., & Creyghton, M. P. (2016). Epigenomic annotation of gene regulatory alterations during evolution of the primate brain. *Nature Neuroscience*, *19*(3), 494–503. https://doi.org/10.1038/nn.4229

Pavlovic, B. J., Blake, L. E., Roux, J., Chavarria, C., & Gilad, Y. (2018). A Comparative Assessment of Human and Chimpanzee iPSC-derived Cardiomyocytes with Primary

Heart Tissues. *Scientific Reports*, *8*(1), 15312. https://doi.org/10.1038/s41598-018-33478-9

Pollard, K. S., Salama, S. R., Lambert, N., Lambot, M.-A., Coppens, S., Pedersen, J. S., Katzman, S., King, B., Onodera, C., Siepel, A., Kern, A. D., Dehay, C., Igel, H., Ares, M., Vanderhaeghen, P., & Haussler, D. (2006). An RNA gene expressed during cortical development evolved rapidly in humans. *Nature*, *443*(7108), 167–172. https://doi.org/10.1038/nature05113

Prud'homme, B., Gompel, N., & Carroll, S. B. (2007). Emerging principles of regulatory evolution. *Proceedings of the National Academy of Sciences*, *104*(suppl_1), 8605–8612. https://doi.org/10.1073/pnas.0700488104

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics (Oxford, England)*, *26*(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033

Ramírez, F., Dündar, F., Diehl, S., Grüning, B. A., & Manke, T. (2014). deepTools: A flexible platform for exploring deep-sequencing data. *Nucleic Acids Research*, *42*(Web Server issue), W187-191. https://doi.org/10.1093/nar/gku365

Reilly, S. K., & Noonan, J. P. (2016). Evolution of Gene Regulation in Humans. *Annual Review of Genomics and Human Genetics*, *17*, 45–67. https://doi.org/10.1146/annurev-genom-090314-045935

Romero, I. G., Ruvinsky, I., & Gilad, Y. (2012). Comparative studies of gene expression and the evolution of gene regulation. *Nature Reviews Genetics*, *13*(7), 505–516. https://doi.org/10.1038/nrg3229

Sharma, R., Khristov, V., Rising, A., Jha, B. S., Dejene, R., Hotaling, N., Li, Y., Stoddard, J., Stankewicz, C., Wan, Q., Zhang, C., Campos, M. M., Miyagishima, K. J., McGaughey, D., Villasmil, R., Mattapallil, M., Stanzel, B., Qian, H., Wong, W., … Bharti, K. (2019). Clinical-grade stem cell-derived retinal pigment epithelium patch rescues retinal

degeneration in rodents and pigs. *Science Translational Medicine*, *11*(475), eaat5580. https://doi.org/10.1126/scitranslmed.aat5580

Shave, R. E., Lieberman, D. E., Drane, A. L., Brown, M. G., Batterham, A. M., Worthington, S., Atencia, R., Feltrer, Y., Neary, J., Weiner, R. B., Wasfy, M. M., & Baggish, A. L. (2019). Selection of endurance capabilities and the trade-off between pressure and volume in the evolution of the human heart. *Proceedings of the National Academy of Sciences*, *116*(40), 19905–19910. https://doi.org/10.1073/pnas.1906902116

Song, J. H. T., Grant, R. L., Behrens, V. C., Kučka, M., Roberts Kingman, G. A., Soltys, V., Chan, Y. F., & Kingsley, D. M. (2021). Genetic studies of human-chimpanzee divergence using stem cell fusions. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(51), e2117557118. https://doi.org/10.1073/pnas.2117557118

Starr, A. L., Gokhman, D., & Fraser, H. B. (2023). Accounting for cis-regulatory constraint prioritizes genes likely to affect species-specific traits. *Genome Biology*, *24*(1), 11. https://doi.org/10.1186/s13059-023-02846-8

Storey Lab. (n.d.). *Qvalue*.

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., & Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, *102*(43), 15545–15550. https://doi.org/10.1073/pnas.0506580102

Su-Feher, L., Rubin, A. N., Silberberg, S. N., Catta-Preta, R., Lim, K. J., Ypsilanti, A. R., Zdilar, I., McGinnis, C. S., McKinsey, G. L., Rubino, T. E., Hawrylycz, M. J., Thompson, C., Gartner, Z. J., Puelles, L., Zeng, H., Rubenstein, J. L. R., & Nord, A. S. (2022). Single cell enhancer activity distinguishes GABAergic and cholinergic lineages in embryonic mouse basal ganglia. *Proceedings of the National Academy of Sciences*, *119*(15), e2108760119. https://doi.org/10.1073/pnas.2108760119

Sullivan, P. F., Meadows, J. R. S., Gazal, S., Phan, B. N., Li, X., Genereux, D. P., Dong, M. X., Bianchi, M., Andrews, G., Sakthikumar, S., Nordin, J., Roy, A., Christmas, M. J., Marinescu, V. D., Wallerman, O., Xue, J. R., Li, Y., Yao, S., Sun, Q., … Lindblad-Toh, K. (2023). *Leveraging Base Pair Mammalian Constraint to Understand Genetic Variation and Human Disease* [Preprint]. Genomics. https://doi.org/10.1101/2023.03.10.531987

Traag, V. A., Waltman, L., & van Eck, N. J. (2019). From Louvain to Leiden: Guaranteeing well-connected communities. *Scientific Reports*, *9*(1), 5233. https://doi.org/10.1038/s41598-019-41695-z

Trevino, A. E., Sinnott-Armstrong, N., Andersen, J., Yoon, S.-J., Huber, N., Pritchard, J. K., Chang, H. Y., Greenleaf, W. J., & Pașca, S. P. (2020). Chromatin accessibility dynamics in a model of human forebrain development. *Science (New York, N.Y.)*, *367*(6476), eaay1645. https://doi.org/10.1126/science.aay1645

Trizzino, M., Park, Y., Holsbach-Beltrame, M., Aracena, K., Mika, K., Caliskan, M., Perry, G. H., Lynch, V. J., & Brown, C. D. (2017). Transposable elements are the primary source of novelty in primate gene regulation. *Genome Research*, *27*(10), 1623–1633. https://doi.org/10.1101/gr.218149.116

van de Geijn, B., McVicker, G., Gilad, Y., & Pritchard, J. K. (2015). WASP: Allele-specific software for robust molecular quantitative trait locus discovery. *Nature Methods*, *12*(11), 1061–1063. https://doi.org/10.1038/nmeth.3582

Vanderhaeghen, P., & Polleux, F. (2023). Developmental mechanisms underlying the evolution of human cortical circuits. *Nature Reviews Neuroscience*, *24*(4), 213–232. https://doi.org/10.1038/s41583-023-00675-z

Vierstra, J., Lazar, J., Sandstrom, R., Halow, J., Lee, K., Bates, D., Diegel, M., Dunn, D., Neri, F., Haugen, E., Rynes, E., Reynolds, A., Nelson, J., Johnson, A., Frerker, M., Buckley, M., Kaul, R., Meuleman, W., & Stamatoyannopoulos, J. A. (2020). Global reference

mapping of human transcription factor footprints. *Nature*, *583*(7818), 729–736. https://doi.org/10.1038/s41586-020-2528-x

Vu, H., & Ernst, J. (2022). Universal annotation of the human genome through integration of over a thousand epigenomic datasets. *Genome Biology*, *23*(1), 9. https://doi.org/10.1186/s13059-021-02572-z

Wang, B., & Tontonoz, P. (2018). Liver X receptors in lipid signalling and membrane homeostasis. *Nature Reviews Endocrinology*, *14*(8), 452–463. https://doi.org/10.1038/s41574-018-0037-x

Watanabe, A., Toyota, T., Owada, Y., Hayashi, T., Iwayama, Y., Matsumata, M., Ishitsuka, Y., Nakaya, A., Maekawa, M., Ohnishi, T., Arai, R., Sakurai, K., Yamada, K., Kondo, H., Hashimoto, K., Osumi, N., & Yoshikawa, T. (2007). Fabp7 Maps to a Quantitative Trait Locus for a Schizophrenia Endophenotype. *PLoS Biology*, *5*(11), e297. https://doi.org/10.1371/journal.pbio.0050297

Wittkopp, P. J., & Kalay, G. (2012). Cis-regulatory elements: Molecular mechanisms and evolutionary processes underlying divergence. *Nature Reviews Genetics*, *13*(1), 59–69. https://doi.org/10.1038/nrg3095

Wolf, F. A., Angerer, P., & Theis, F. J. (2018). SCANPY: Large-scale single-cell gene expression data analysis. *Genome Biology*, *19*(1), 15. https://doi.org/10.1186/s13059-017-1382-0

Wolf, F. A., Hamey, F. K., Plass, M., Solana, J., Dahlin, J. S., Göttgens, B., Rajewsky, N., Simon, L., & Theis, F. J. (2019). PAGA: Graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biology*, *20*(1), 59. https://doi.org/10.1186/s13059-019-1663-x

Yang, N., Chanda, S., Marro, S., Ng, Y.-H., Janas, J. A., Haag, D., Ang, C. E., Tang, Y., Flores, Q., Mall, M., Wapinski, O., Li, M., Ahlenius, H., Rubenstein, J. L., Chang, H. Y., Buylla, A. A., Südhof, T. C., & Wernig, M. (2017). Generation of pure GABAergic neurons by

transcription factor programming. *Nature Methods*, *14*(6), 621–628. https://doi.org/10.1038/nmeth.4291

Yao, Z., van Velthoven, C. T. J., Nguyen, T. N., Goldy, J., Sedeno-Cortes, A. E., Baftizadeh, F., Bertagnolli, D., Casper, T., Chiang, M., Crichton, K., Ding, S.-L., Fong, O., Garren, E., Glandon, A., Gouwens, N. W., Gray, J., Graybuck, L. T., Hawrylycz, M. J., Hirschstein, D., … Zeng, H. (2021). A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation. *Cell*, *184*(12), 3222-3241.e26. https://doi.org/10.1016/j.cell.2021.04.021

York, R. A., Patil, C., Abdilleh, K., Johnson, Z. V., Conte, M. A., Genner, M. J., McGrath, P. T., Fraser, H. B., Fernald, R. D., & Streelman, J. T. (2018). Behavior-dependent *cis* regulation reveals genes and pathways associated with bower building in cichlid fishes. *Proceedings of the National Academy of Sciences*, *115*(47). https://doi.org/10.1073/pnas.1810140115

Yu, X., Lin, J., Zack, D. J., & Qian, J. (2006). Computational analysis of tissue-specific combinatorial gene regulation: Predicting interaction between transcription factors in human tissues. *Nucleic Acids Research*, *34*(17), 4925–4936. https://doi.org/10.1093/nar/gkl595

Zhang, S., Zhang, H., Zhou, Y., Qiao, M., Zhao, S., Kozlova, A., Shi, J., Sanders, A. R., Wang, G., Luo, K., Sengupta, S., West, S., Qian, S., Streit, M., Avramopoulos, D., Cowan, C. A., Chen, M., Pang, Z. P., Gejman, P. V., … Duan, J. (2020). Allele-specific open chromatin in human iPSC neurons elucidates functional disease variants. *Science*, *369*(6503), 561–565. https://doi.org/10.1126/science.aay3983

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Brown, M., Li, W., & Liu, X. S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biology*, *9*(9), R137. https://doi.org/10.1186/gb-2008-9-9-r137

Zhu, A., Ibrahim, J. G., & Love, M. I. (2019). Heavy-tailed prior distributions for sequence count data: Removing the noise and preserving large differences. *Bioinformatics*, *35*(12), 2084–2092. https://doi.org/10.1093/bioinformatics/bty895