**Title: Disentangling cell-intrinsic and extrinsic factors underlying gene expression evolution**

Authors: Alexander L. Starr[1] †, Toshiya Nishimura[2,3,4] †, Kyomi J. Igarashi[5], Chihiro Funamoto[4], Hiromitsu Nakauchi[2,6], Hunter B. Fraser[1]*

Affiliations:

1. Department of Biology, Stanford University, Stanford, CA 94305, USA
2. Institute for Stem Cell Biology and Regenerative Medicine, Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305, USA.
3. WPI Premium Research Institute for Human Metaverse Medicine (WPI-PRIMe), Osaka University, Osaka, 565-0871, Japan (current address for T.N.)
4. Division of Stem Cell and Organoid Medicine, Department of Genome Biology, Graduate School of Medicine, Osaka University, Osaka, 565-0871, Japan
5. Department of Genetics, Stanford University School of Medicine, Stanford, CA, 94305, USA
6. Division of Stem Cell Therapy, Distinguished Professor Unit, The Institute of Medical Science, The University of Tokyo, Minato-ku, Tokyo 108-8639, Japan.

†These authors contributed equally to this work

*Indicates corresponding author: hbfraser@stanford.edu

**Abstract:**

Chimeras have played a foundational role in biology, for example by enabling the classification of developmental processes into those driven intrinsically by individual cells versus those driven

extrinsically by their extracellular environment. Here, we extend this framework to decompose evolutionary divergence in gene expression and other quantitative traits into cell-intrinsic, extrinsic, and intrinsic-extrinsic interaction components. Applying this framework to reciprocal rat-mouse chimeras, we found that the majority of gene expression divergence is attributable to cell-intrinsic factors, though extrinsic factors also play an integral role. For example, a rat-like extracellular environment extrinsically up-regulates the expression of a key transcriptional regulator of the endoplasmic reticulum (ER) stress response in some but not all cell types, which in turn strongly predicts extrinsic up-regulation of its target genes and of the ER stress response pathway as a whole. This effect is also seen at the protein level, suggesting propagation through multiple regulatory levels. We also demonstrate that our framework is applicable to a cellular trait, neuronal differentiation, and estimated the intrinsic and extrinsic contributions to its divergence. Finally, we show that imprinted genes are dramatically mis-expressed in species-mismatched environments, suggesting that mismatch between rapidly evolving intrinsic and extrinsic mechanisms controlling gene imprinting may contribute to barriers to interspecies chimerism. Overall, our conceptual framework opens new avenues to investigate the mechanistic basis of evolutionary divergence in gene expression and other quantitative traits in any multicellular organism.

**Main text:**

## Introduction

Interspecies chimeras—formed by grafting donor cells from one species into the host embryo of another—have enabled seminal discoveries in vertebrate development dating back over a century(*1–5*). Chimeras have been instrumental in determining whether effects of a specific gene or developmental process are cell-intrinsic (also referred to as cell-autonomous) or cell-

extrinsic (non-cell-autonomous)(*6*, *7*). In general, a process is said to be cell-intrinsic if the cells responsible for executing that process behave in a similar way regardless of the genetic background of surrounding cells. For example, quail donor cranial neural crest cells destined to form the beak follow their faster cell-intrinsic developmental timeline and execute their own spatial and temporal programs for gene expression when they are transplanted into a slower-developing duck host embryo(*4*, *8*, *9*). On the other hand, a process is cell-extrinsic when dependent on the genotype of other cells; duck host epidermal cells in the same quail-duck chimeras are extrinsically induced by the quail donor neural crest to alter their gene expression to make quail-like feathers and a conical egg tooth(*4*, *8*, *9*). Although this categorization of developmental processes into extrinsically vs. intrinsically controlled has been immensely fruitful, there is no general framework to quantitatively deconvolve genome-wide measurements into cell-extrinsic and intrinsic components.

In contrast, the quantitative dissection of two fundamentally different molecular mechanisms to divergence in gene expression, a key driver of evolutionary adaptation and phenotypic novelty, has been extensively explored(*10*, *11*). These mechanisms are known as *cis* and *trans*: *cis* refers to regulatory elements such as promoters and enhancers that affect nearby genes on the same chromosome, whereas *trans*-acting factors involve diffusible molecules such as transcription factors (TFs) or noncoding RNAs that can regulate genes throughout the genome(*11*). A landmark study used *Drosophila* interspecies hybrids to disentangle the *cis*- and *trans*-regulatory contributions to gene expression divergence(*11*). Since then, genome-wide *cis*- and *trans*-regulatory variation has been estimated for species across the tree of life, finding that the majority of variation within most species is caused by *trans*-acting mechanisms(*12*–*15*). While *cis*-acting divergence is always cell-intrinsic, both intrinsic and extrinsic factors can lead to *trans*-acting divergence, yet their relative contributions to phenotypic evolution remain unknown.

78 Here, we unify these concepts and introduce a framework to disentangle the contributions of
79 intrinsic factors, extrinsic factors, and their interactions to evolutionary divergence in gene
80 expression and other quantitative traits.
81
82 **Results**
83
84 ***Quantifying extrinsic and intrinsic divergence with interspecies chimeras***
85
86 A chimera is an amalgamation of two sets of cells, donor and host. We define host cells as the
87 cells derived from the injected blastocyst; when host cells make up a large majority, they are in
88 species-matched environments similar to wildtype organisms (Fig 1A). Donor cells are the
89 progeny of the cells that were injected into the blastocyst; when these are a small minority, they
90 are in species-mismatched environments determined by the host (Fig 1B-C). Throughout, we
91 use donor cells/cells in species-mismatched environments interchangeably and host cells/cells
92 in species-matched environments interchangeably.
93
94 To illustrate the concept of extrinsic and intrinsic divergence in gene expression, we can
95 consider an orthologous gene that is more highly expressed in mice than in rats. If mouse cells
96 express the gene at the same level regardless of their extracellular environment, and likewise
97 for rat cells, then we infer that the expression difference is determined by the cell of origin rather
98 than its extracellular environment, suggesting cell-intrinsic divergence (Fig 1B). However, if
99 donor and host cells express this gene at equal levels within each chimera, then we infer that
100 this divergence is determined by the extracellular environment of the cells in that animal,
101 suggesting extrinsic divergence in gene expression caused by divergence in the extracellular
102 environment (Fig 1C). In practice, we can measure the intrinsic and extrinsic divergence in gene
103 expression using single-cell RNA-seq (scRNA-seq) of reciprocal interspecies chimeras (Fig 1D).

104

105 To explore this concept empirically, we first generated reciprocal rat-like and mouse-like
106 chimeras via blastocyst injection (Fig 1E, Methods). Importantly, we only used tissues with less
107 than 10% donor contribution so that the extrinsic environment was primarily composed of host
108 cells. As an example, we observed clear populations of host and donor cells (marked by
109 TdTomato) in the forebrain of both mouse-like and rat-like chimeras (Fig 1F-G). Consistent with
110 previous work, different brain regions showed distinct patterns, as donor cells were rarer and
111 more evenly distributed in the ganglionic eminence compared to the developing neocortex in
112 mouse-like chimeras (Fig 1G)(*16*).

113

114 Next, we enriched for donor cells from stage-matched embryos (E13.5 in mouse, E15.5 in rat),
115 then generated and analyzed scRNA-seq data from the forebrain and connective tissue/spinal
116 cord of both mouse-like (i.e. made up primarily of mouse host cells) and rat-like chimeras (Fig
117 1E, Figs. S1-S2, Methods). After removing low-quality cells, identifying cell types and filtering
118 those with fewer than 10 cells in any sample type (host mouse, host rat, donor mouse, and
119 donor rat), we retained 4,720 cells distributed across 11 cell types. As an example, we can
120 consider the expression of *Efnb3* in forebrain glutamatergic neurons (abbreviated as
121 brain.glut.neu). In species-matched environments, we observed two-fold higher expression in
122 mouse than in rat (Fig 2A, left; the bars to the left of the dashed line are the same data as the
123 bars for species-matched environments to the right of the dashed line and are included twice for
124 clarity). However, in both chimeras, *Efnb3* expression in the donor cells mirrored that of the host
125 species (Fig 2A, middle/right; see Fig. S3A for the per-cell normalized count distribution). This
126 suggests that the divergence in the expression of *Efnb3* in forebrain glutamatergic neurons was
127 driven almost entirely by extrinsic factors. On the other hand, the expression of *Lsm6* was much
128 higher in mouse cells regardless of their extracellular environment, suggesting purely intrinsic
129 divergence between species (Fig 2B, Fig. S3B).

However, most genes do not fit neatly into either category. For example, in both mouse and rat cells, *Plcxd2* had higher expression in a mouse-like environment than in a rat-like environment (Fig 2C, Fig. S3C). However, it also has higher expression in rat cells than in mouse cells in both chimeras, consistent with intrinsically higher expression in rat cells. These two opposing differences cancel out, resulting in nearly identical expression in host cells (Fig 2C) (see Figs. S4-9 and Supplementary Text 1 for discussion of how developmentally dynamic gene expression and differences in the spatial distribution of donor cells between cell types might affect our results). While most genes primarily exhibit some combination of extrinsic and intrinsic divergence, some genes have evidence for a complex interaction between the two forms of divergence. For example, *Pou3f2* is down-regulated in species-mismatched environments in both chimeras, which cannot be explained by any simple combination of intrinsic and extrinsic divergence, suggesting an interaction causing *Pou3f2* in mouse cells to respond differently to the extracellular environment than it does in rat cells (Fig 2D, Fig. S3D).

***A quantitative framework to disentangle extrinsic and intrinsic components of divergence***

Next, we developed a quantitative framework to estimate the extrinsic, intrinsic, and interaction components of gene expression divergence for each gene (see Methods). To summarize these results genome-wide, we visualize these values on a scatter plot so that genes with purely intrinsic divergence are on the x-axis and those with purely extrinsic divergence are on the y-axis (Fig 2F-G, Figs. S10-11, Supplementary Table 1). Across all cell types, we consistently observed that the genome-wide intrinsic component of gene expression divergence is larger than the extrinsic component, which is in turn larger than the interaction component (Fig 2G and Figs. S10-11). We estimate that 21-28% of gene expression divergence is cell-extrinsic across

156 cell types, suggesting that a substantial fraction of the *trans*-regulatory divergence between
157 species is due to differences in the extracellular environment.
158
159 We found that intrinsic divergence is substantially more correlated across cell types than
160 extrinsic divergence (Fig. S12A), consistent with the idea that *cis*-regulatory variation is also
161 less tissue-specific than *trans*(*17–19*). Interestingly, neural and mesenchymal cell types
162 clustered separately in terms of intrinsic divergence, but for extrinsic divergence they instead
163 clustered by location within the embryo, resulting in spinal neurons clustering with other cells
164 from more caudal parts of the embryo such as mesenchyme (Fig. S12B). This suggests that
165 when specifically comparing extrinsically-driven divergence, the similar extracellular
166 environment shared between both neuronal and non-neuronal caudal cell types could be the
167 dominant factor, overriding the tendency for similar cell types such as neurons to cluster
168 together. In addition, we evaluated whether several gene-level variables were correlated with
169 proportion intrinsic vs proportion extrinsic divergence. The best predictor of proportion intrinsic
170 vs proportion extrinsic divergence for individual genes was each gene's expression difference
171 between mouse and rat: all genes with >4-fold difference had primarily intrinsic divergence (Fig.
172 S12C), consistent with the finding that *cis*-regulatory divergence (which is cell-intrinsic)
173 contributes more to large changes in gene expression(*20*). Other factors like expression level
174 and tissue-specificity had little to no predictive power (Fig. S12D).
175
176 ***Extrinsic and intrinsic divergence can propagate through transcriptional networks***
177
178 To investigate whether extrinsic divergence converges on particular pathways, we tested
179 whether extrinsically-driven genes were enriched for different gene sets, identifying dozens of
180 enrichments (Supplementary Table 2). For example, we found that genes involved in the ER
181 stress response, such as *Hsp90b1*, had >2-fold enrichment for up-regulation in a rat-like

182 environment in 10 out of 11 cell types, Fig 3A-B, Fig. S13A-B)(*21*). Conversely, we found that
183 genes encoding histone methyltransferases, such as *Setbp1*, were highly enriched for up-
184 regulation in a mouse-like environment in neural but not connective tissue (Fig. S14A-D)(*22*).
185 Finally, many pathways were strongly enriched for extrinsic divergence in only one cell type,
186 such as positive regulation of programmed cell death which was specifically up-regulated by
187 forebrain GABAergic neurons in a rat-like environment (Fig. S15A-D).
188
189 We next hypothesized that intrinsic and extrinsic regulation may propagate through
190 transcriptional regulatory networks; for example, extrinsic regulation of a TF may correlate with
191 extrinsic regulation of its target genes. In our analysis of ER stress response genes (Fig 3A) we
192 noticed that divergence in the expression of *Xbp1*, a key transcriptional regulator of ER stress
193 response genes(*23*), was almost entirely extrinsic in some cell types such as mesenchyme (Fig
194 3D, Fig. S16A). To determine whether extrinsically driven changes in *Xbp1* expression might
195 drive the extrinsic divergence of its target genes, we compared the degree of extrinsic up-
196 regulation in the rat-like environment for *Xbp1* target genes (defined as genes up-regulated by
197 *Xbp1* overexpression)(*24*) with that of *Xbp1* itself in each cell type(*25*). Remarkably, we
198 observed a strong correlation between these two quantities, where cell types with extrinsic up-
199 regulation of *Xbp1* in the rat-like environment showed a similar pattern of extrinsic up-regulation
200 for its target genes (Spearman's rho = 0.70, p = 0.025, Fig 3E). Moreover, we observed a
201 similarly strong correlation between *Xbp1* extrinsic divergence and the extrinsic divergence of
202 ER stress response-associated genes (Spearman's rho = 0.73, p = 0.014, Fig. S17A). This
203 suggests that extrinsic regulation of *Xbp1* may lead to extrinsic regulation of its target genes
204 (Fig 3F), though we cannot rule out a potential role for other unobserved factors such as post-
205 transcriptional regulation(*23*).
206

207 Interestingly, although *Xbp1* had a similar magnitude of up-regulation in both rat host spinal
208 glutamatergic neurons and rat host mesenchyme compared to their mouse host counterparts
209 (Fig 3D, G), this was almost entirely due to intrinsic divergence in the spinal neurons (Fig 3G,
210 Fig. S16B). Analogous to the pattern for extrinsic divergence (Fig 3E), we found that cell types
211 with intrinsic up-regulation of *Xbp1* in rat cells showed a similar pattern of intrinsic up-regulation
212 of its target genes (Spearman's rho = 0.79, p = 0.006, Fig 3H), although in this case *Xbp1*
213 intrinsic regulation was not as strongly predictive of intrinsic up-regulation of ER stress response
214 genes more broadly (Spearman's rho = 0.37, p = 0.29, Fig. S17B). This suggests that
215 intrinsically driven divergence in *Xbp1* expression may propagate to its target genes (Fig 3I) in
216 much the same way as extrinsic divergence. More generally, we tested whether the known
217 targets of many different TFs were enriched for unidirectional extrinsic divergence. Notably, the
218 targets of several well-known regulators of ER stress including *Atf3*, *Atf6*, and *Jun* were also
219 enriched for extrinsic up-regulation in a rat-like environment, similar to *Xbp1* (Fig. 3J)(*26–28*).
220 The extrinsic regulation of these TFs was not correlated with that of *Xbp1* (Fig. S18) suggesting
221 divergence in distinct extrinsic factors that converge on TFs regulating the ER stress response.
222 Overall, we observed 155 enriched TFs at a 10% FDR, compared to just one enriched TF when
223 performing an identical procedure with intrinsic regulation (Fig 3K, Supplementary Table 3).
224
225 The single TF with targets enriched for intrinsic divergence was *Nfe2l1* (which encodes the
226 protein Nrf1), a well-established regulator of proteasomal genes (Fig. 3K)(*29*). For example,
227 *Pomp*, a target of Nrf1 involved in the formation and activation of the proteasome, was
228 consistently more highly expressed in rat cells than mouse cells regardless of the extracellular
229 environment (Fig. S19A-F)(*29*). In contrast, the targets of other proteasome TFs such as *Nfe2l2*
230 (encoding the protein Nrf2) and *Stat3* were enriched for extrinsic up-regulation in mouse-like
231 and rat-like environments, respectively (Fig. 3J)(*30*, *31*). This heterogeneity in the mode
232 (intrinsic vs extrinsic) and direction (rat vs. mouse up-regulation) of divergence in proteasomal

233 TFs contrasted with the four ER stress response TFs mentioned above, whose targets were all
234 enriched for extrinsic up-regulation in a rat-like environment. Interestingly, the conflicting modes
235 and directions of divergence of these proteasomal TFs did not correlate strongly with the
236 divergence of their target genes (Fig. S20A-C), suggesting that heterogeneity in the regulatory
237 divergence of multiple TFs targeting the same genes might lead to more complex patterns than
238 the relatively simple propagation seen for *Xbp1* and its targets.
239
240 To investigate the consequences of having multiple proteasomal TFs with differing modes and
241 directions of expression divergence, we explored how proteasomal subunits were regulated
242 across cell types. For example, in rat cells, the proteasomal subunit *Psmb1* had higher
243 expression in a rat-like environment in forebrain GABAergic neurons, but higher expression in a
244 mouse-like environment in spinal glutamatergic neurons and chondrocytes (Fig 4A-C, Fig.
245 S21A-C). Interestingly, many proteasomal genes showed a similar pattern (Fig 4D-F). In
246 forebrain GABAergic neurons, expression was up-regulated both intrinsically in rat cells and
247 extrinsically by the rat-like environment. However, in spinal glutamatergic neurons, the intrinsic
248 effect was preserved but the extrinsic rat-like environment had the opposite effect, down-
249 regulating expression. Lastly, for chondrocytes, both effects were flipped compared to forebrain
250 GABAergic neurons: both the intrinsic and extrinsic effects of rat were repressive. Based on this
251 finding, we looked for other gene sets that showed similarly variable enrichments across cell
252 types and identified context-specific patterns of intrinsic and extrinsic components for
253 spliceosomal genes as well as nuclear-encoded subunits of mitochondrial respiratory complex I
254 (Figs. S22-23). This suggests that highly coordinated context-specificity, integrating both
255 extrinsic and intrinsic signals, has evolved to regulate multiple protein complexes, perhaps to
256 rapidly modulate their abundances in response to both extrinsic and intrinsic cues.
257
258 ***Extrinsic divergence in the protein expression of two ER stress response genes***

259

260 To test whether the extrinsically-driven gene expression divergence we observed translates to
261 the protein level, we used immunofluorescence (IF) to quantify extrinsic, intrinsic, and
262 interaction components of divergence in protein levels for two proteins: *Jun* and *Hspa5*(*32*). We
263 prioritized these genes because they are involved in the ER stress response, show strong
264 extrinsic divergence across cell types (especially in GABAergic forebrain neurons and
265 progenitors, Fig 5A-B, Fig. S24A-D), and are sufficiently highly expressed to be detected with
266 IF. In the ganglionic eminences (which contain both GABAergic progenitors and recently born
267 GABAergic neurons) of two mouse-like and two rat-like chimeras, we observed extrinsically-
268 driven up-regulation in rat-like environments for both proteins, closely matching the estimates
269 from our scRNA-seq data (Fig 5C-E). This suggests that extrinsic divergence in mRNA
270 abundances can translate to the protein level, and also that our estimates of extrinsic
271 divergence for these genes are unlikely to be inflated by experimental variability.

272

273 ***Extrinsic and intrinsic components of neural progenitor differentiation kinetics***

274

275 As discussed above, our framework is applicable to most quantitative traits. To demonstrate
276 this, we explored the application of this framework to estimate extrinsic, intrinsic, and interaction
277 components to a cellular-level quantitative trait: the ratio of post-mitotic neurons to neural
278 progenitors, which reflects the timing of neurogenesis. If the timing of neurogenesis is
279 intrinsically controlled, we would expect a higher neuron to progenitor ratio in mouse regardless
280 of whether the cells were in a mouse-like or rat-like environment (Fig. 5F, top). On the other
281 hand, if this trait is extrinsically controlled, we would expect similar ratios for both mouse and rat
282 cells in the same extracellular environment (Fig. 5F, bottom). Strikingly, across three different
283 branches of neurogenesis (forebrain GABAergic, forebrain glutamatergic, and spinal), we
284 consistently found that the ratio of neurons to progenitors was much higher in mouse host cells

285 than in rat donor cells, consistent with an intrinsic component for the faster development of
286 mouse neurons and faster global maturation rate of mouse embryos (Fig. S25A-C). However,
287 we also found this ratio to be similar for mouse donor cells and rat host cells, consistent with
288 previous work showing an effect of the rat maternal environment on the timing of neurogenesis
289 (Fig. S25A-C)(*7*). Overall, we estimated an average contribution of 44% intrinsic, 23% extrinsic,
290 and 33% interaction for divergence in the timing of mouse and rat neuronal differentiation (Fig
291 5G).
292
293 ***Imprinted genes are strongly mis-expressed in species-mismatched environments***
294
295 During our analysis, we noticed that several imprinted genes had strikingly large interaction
296 components. For example, while the imprinted gene *Grb10* was highly expressed in both mouse
297 and rat chondrocytes when in species-matched environments, its expression was nearly zero in
298 both mouse and rat chondrocytes in species-mismatched environments (Fig. 6A, Fig.
299 S26A)(*33*). In GABAergic progenitors *Grb10* expression in mouse cells was unaffected by the
300 extracellular environment, but was undetectable in rat cells in a mouse-like environment (Fig.
301 6B, Fig. S26B). In contrast, *Grb10* expression was unaltered in species-mismatched
302 environments in forebrain GABAergic neurons (Fig. 6C, Fig. S26C). This stark divergence was
303 not restricted to *Grb10* as, for example, *Igf2* is more highly expressed in rat chondrocytes than
304 mouse chondrocytes in species-matched environments, but drops close to zero when rat
305 chondrocytes are in a mouse-like environment (Fig. S27A, D)(*34*).
306
307 To explore whether this pattern holds for imprinted genes more generally, we tested whether
308 previously reported imprinted genes were enriched for interactions between extrinsic and
309 intrinsic divergence. Remarkably, we observed that imprinted genes were strongly enriched for
310 large interaction components ($p < 0.001$ in both connective and nervous tissue, Fig. 6D). Given

311 that genes expressed from the maternal allele (maternally-expressed) and expressed from the
312 paternal allele (paternally-expressed) often have opposing effects on growth, we split imprinted
313 genes based on the expressed allele(*35*). We found that paternally-expressed genes were
314 strongly enriched for higher expression in species-mismatched environments ($p < 0.01$ in both
315 brain and connective tissue), whereas maternally-expressed genes were enriched for higher
316 expression in species-matched environments only in connective tissue ($p < 0.01$ in connective
317 tissue, $p > 0.5$ in brain tissue), suggesting that maternally-expressed and paternally-expressed
318 genes are often affected by species-mismatched environments in opposite ways (Fig. 6E). The
319 pattern for paternally expressed genes is exemplified by *Zdbf2*, which is expressed
320 approximately 2-fold higher in species-mismatched environments for both mouse cells and rat
321 cells across connective and nervous cell types **(**Fig. S27B-C, E-F**)**(*36*).
322
323 To determine if this pattern of disrupted imprinting generalized to other types of cells and
324 chimeras, we sought to test for this effect in other chimera scRNA-seq data sets. Because these
325 studies do not have reciprocal chimeras, we could not disentangle the extrinsic and interaction
326 components, but we could test the simpler prediction that imprinted genes should be highly
327 sensitive to the extracellular environment. In one previous study, rat-mouse chimeras were
328 generated in which the hearts were almost entirely derived from rat donor cells, and their gene
329 expression was compared to wildtype rat heart cells from the same embryonic stage(*37*).
330 Despite the data being independently generated using different host and donor strains, tissues,
331 and developmental stages, we observe that that *Grb10* and *Igf2* were highly expressed in
332 wildtype rat but were nearly undetectable in rat cells present in a mouse-like environment
333 across all 11 heart data cell types tested (Fig. 6F, Fig. S28A-B). More generally, imprinted
334 genes were strongly enriched for differential expression between rat cells in a rat-like
335 environment vs. rat cells in a mouse-like environment ($p < 0.05$ in all cell types except
336 macrophages where $p = 0.11$, Fig. 6G).

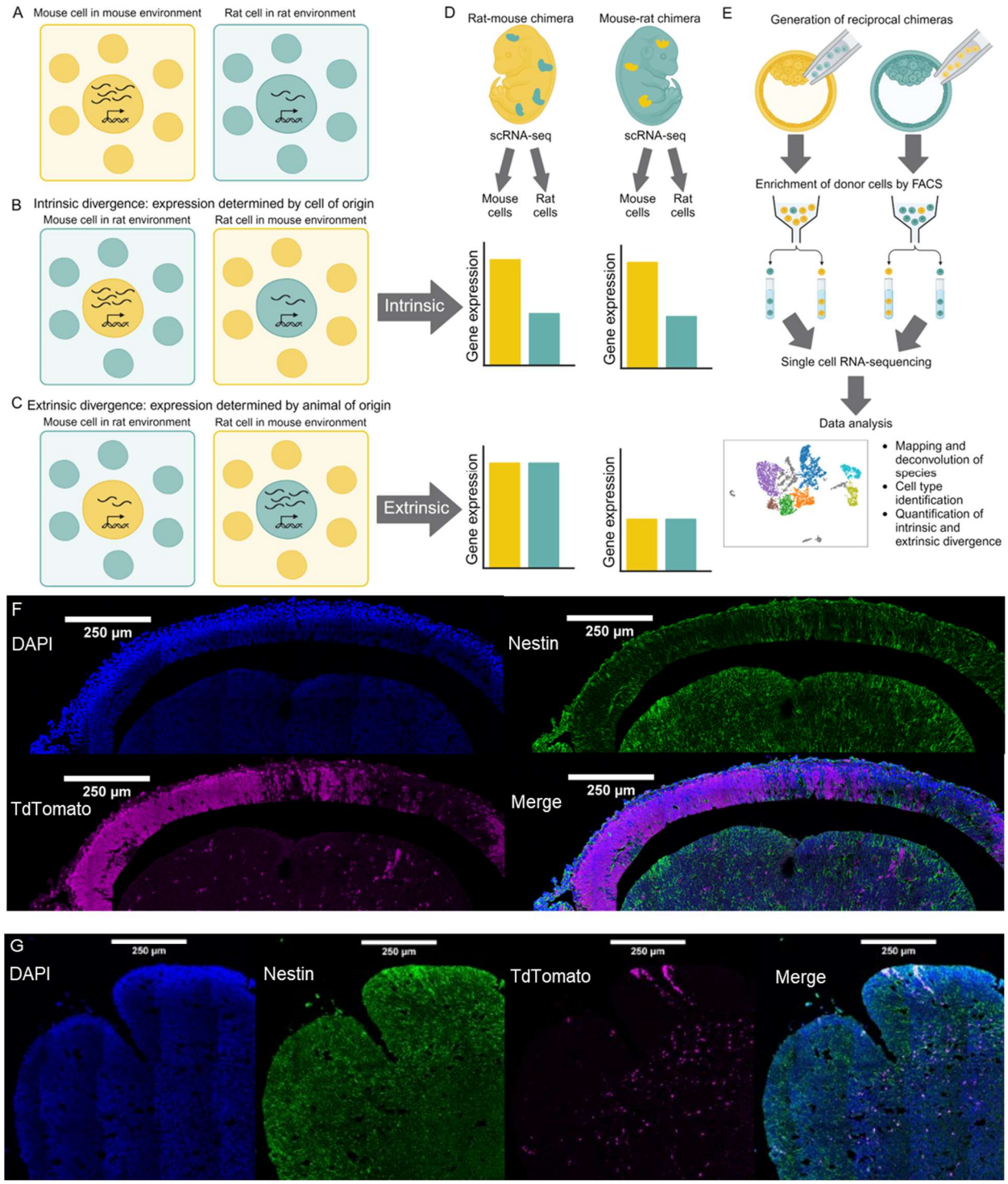
One possible explanation for disrupted imprinting in donor cells is that imprinting could already be disrupted in the ESCs from which these cells are derived. To test this, we examined a published bulk RNA-seq dataset of adult parathyroid gland cells where the same mouse ESCs were injected into both rat and mouse blastocysts and compared to wildtype mice with no blastocyst injection(*38*). If imprinting disruption is due to the species-mismatched environment rather than any technical factor related to cell lines or chimera generation, then we would expect imprinted genes to be mis-expressed in rat-mouse chimeras but not mouse-mouse chimeras. To evaluate this, we hierarchically clustered all samples based on the expression of all expressed imprinted genes. We found that samples from species-mismatched environments clearly clustered together, whereas samples from wildtype mice and donor mouse cells in a species-matched environment were intermixed in a separate cluster (Fig. 6G). For example, *Grb10* and *Igf2* were strongly mis-expressed in a species-mismatched environment (Fig. S29A). Overall, these results suggest that a species-mismatched environment, not blastocyst injection of ESCs more generally, underlies mis-expression of imprinted genes and that mis-expression can persist into adulthood.

Finally, we investigated whether imprinting may also be disrupted in *ex vivo* human-macaque chimeric embryos(*39*, *40*). We found that many imprinted genes, such as *PLAGL1,* showed strong mis-regulation in human cells growing in a macaque-like embryo (Fig. 6I). Imprinted genes were enriched for differential expression between human cells vs. macaque cells in the same macaque-like environment, as well as between human cells in a human-like environment vs. in a macaque-like environment ($p < 0.01$ in each). These results suggest that imprinting disruption in chimeras does not depend on the maternal environment (which is not present for *ex vivo* chimeras) and occurs in both rodents and primates. Overall, our results suggest that the

362 interaction between extrinsic and intrinsic divergence results in widespread mis-expression of
363 imprinted genes in interspecies chimeras.
364
365 **Conclusion**
366
367 Building on foundational work in developmental and evolutionary biology, we have developed a
368 broadly applicable framework to decompose divergence in quantitative traits into cell-extrinsic
369 and cell-intrinsic components. Furthermore, our framework can provide insight into the genetic
370 basis of interspecies chimeric incompatibility(*41*). For example, we identified strong extrinsic
371 divergence in ER stress response genes such as *Hspa5,* which is vital for neuronal survival
372 during development(*42*), as well as extrinsic divergence in histone methyltransferases, which
373 play a key role in controlling the tempo of neuronal development(*43*). In addition, we identified
374 widespread mis-regulation of imprinted genes as a result of interactions between extrinsic and
375 intrinsic divergence. As many imprinted genes play key roles in growth and cell survival, our
376 results also suggest that modulating the expression of imprinted genes could be a promising
377 avenue to improve the growth and survival of cells in species-mismatched environments(*44*).
378
379 A recent study reported rat-mouse chimeras where rat ESCs were injected into host mice
380 blastocysts that were unable to generate a forebrain (*Hesx1*$^{-/-}$)(*16*). This study is challenging to
381 directly compare to ours since rat donor cells composed ~60% of the forebrain at the time of
382 scRNA-seq profiling, suggesting that the extracellular environment was neither mouse-like nor
383 rat-like. In addition, the study did not have reciprocal chimeras so could not directly quantify
384 intrinsic, extrinsic divergence, or their interactions. Perhaps in part because of these and other
385 technical factors, our results differ substantially from theirs. Specifically, they found the pace of
386 neurodevelopment (at a gross anatomical level) in rat donor cells to be extrinsically controlled,
387 while the overall transcriptome was intrinsically determined. In contrast, using reciprocal

chimeras and our quantitative framework we found both neurodevelopment (at a single-cell level) and the transcriptome to be subject to about twice as much intrinsic as extrinsic divergence, along with a substantial effect of interaction between these.

Our study opens the door to myriad avenues for further research. For example, sampling chimeras across a time course of development will allow us to disentangle the role of developmental time from intrinsic and extrinsic divergence to better understand the evolution of developmentally dynamic gene expression. In addition, our framework could be applied in conjunction with spatial transcriptomic measurements to further decompose extrinsic divergence into local and organ-wide components. Overall, the application of our framework to a wide range of quantitative traits will enable novel insights into the mechanistic basis of phenotypic evolution across the tree of multicellular life.


A
Mouse cell in mouse environment
Rat cell in rat environment
B
Intrinsic divergence: expression determined by cell of origin
Mouse cell in rat environment
Rat cell in mouse environment
Intrinsic
C
Extrinsic divergence: expression determined by animal of origin
Mouse cell in rat environment
Rat cell in mouse environment
Extrinsic
D
Rat-mouse chimera
Mouse-rat chimera
scRNA-seq
scRNA-seq
Mouse cells
Rat cells
Mouse cells
Rat cells
Gene expression
Gene expression
Gene expression
Gene expression
E
Generation of reciprocal chimeras
Enrichment of donor cells by FACS
Single cell RNA-sequencing
Data analysis
• Mapping and deconvolution of species
• Cell type identification
• Quantification of intrinsic and extrinsic divergence
F
DAPI
250 µm
Nestin
250 µm
TdTomato
250 µm
Merge
250 µm
G
DAPI
250 µm
Nestin
250 µm
TdTomato
250 µm
Merge
250 µm

401

402

**Fig. 1: Conceptual model of estimating extrinsic and intrinsic divergence using reciprocal chimeras. A)** Rat and mouse cells in species-matched environments. In wildtype animals, the divergence in expression of a single gene between species is determined by divergence intrinsic to mouse and rat cells, as well as divergence in the extracellular environment. In this example, expression is 2.5-fold higher in mouse cells compared to rat cells when both cells are in species-matched environments. However, it is unclear how much of this divergence is extrinsically driven or intrinsically driven. **B)** Example of intrinsic divergence. In interspecies chimeras, we can measure gene expression in mouse cells in a rat-like environment and in rat cells in a mouse-like environment. If the expression of the gene remains unchanged compared to species-matched environments, then the divergence in gene expression must be due to intrinsic mechanisms. **C)** Example of extrinsic divergence. If the expression of this gene in mouse cells in a rat-like environment matches the expression of the gene in rat cells in a rat-like environment and vice versa, then the divergence in gene expression must be due to extrinsic divergence in the extracellular environment. **D)** Outline of measuring intrinsic and extrinsic divergence genome-wide. To measure extrinsic and intrinsic divergence, we can generate mouse-like (a small proportion of rat cells and a large proportion of mouse cells) and rat-like (a small proportion of mouse cells and a large proportion of rat cells) reciprocal chimeras. We can then use scRNA-seq to measure gene expression in each cell type across the four different species-environment combinations. If the divergence in expression for a gene is similar in both rat-like and mouse-like environments, then it is driven by intrinsic divergence. On the other hand, if expression is similar when mouse and rat cells are in the same extracellular environment but different when comparing cells from two different environments, it is driven by extrinsic divergence. **E)** Outline of experimental procedure for this study. Reciprocal chimeras were generated via blastocyst injection and cells from mouse-like chimeras harvested at E13.5 and cells from rat-like chimeras harvested at a matched developmental stage (E15.5). Donor cells in species-mismatched environments were enriched

429 using fluorescence-activated cell sorting using TdTomato as a marker and single-cell RNA-
430 sequencing was performed. Gene expression measurements for each cell were then
431 computationally identified as mouse or rat, clustered, and then used to estimate extrinsic and
432 intrinsic divergence between species. **F)** Immunofluorescence image of the E13.5 forebrain of a
433 mouse-like chimera. Nuclei are stained with DAPI (blue) and Nestin (a neural stem cell marker)
434 is shown in green. TdTomato (magenta) marks donor rat cells in the mouse-like environment.
435 **G)** Immunofluorescence image of the E15.5 ganglionic eminence of a rat-like chimera. In this
436 case, TdTomato marks mouse cells in the rat-like environment.
437
438
439

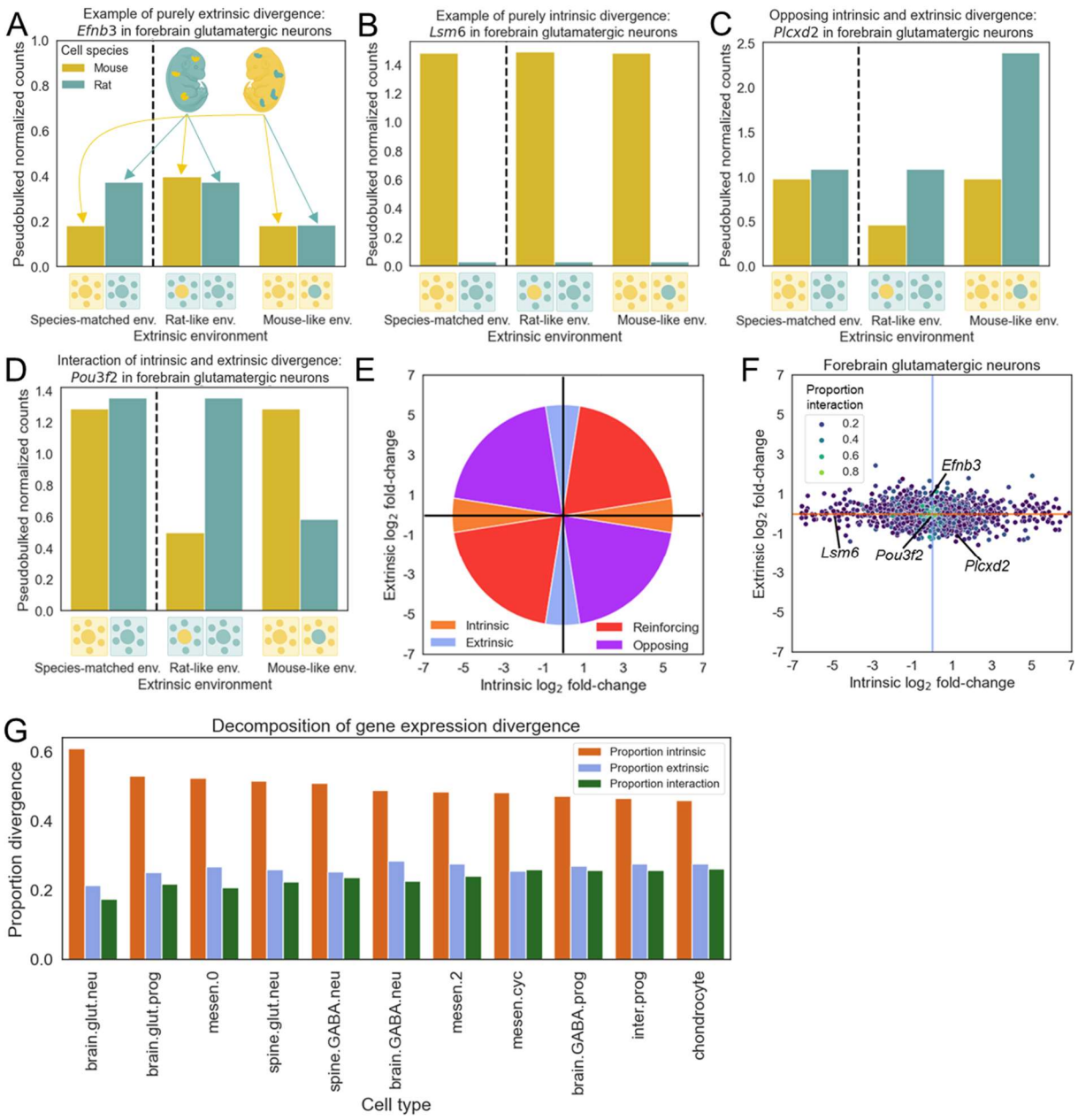
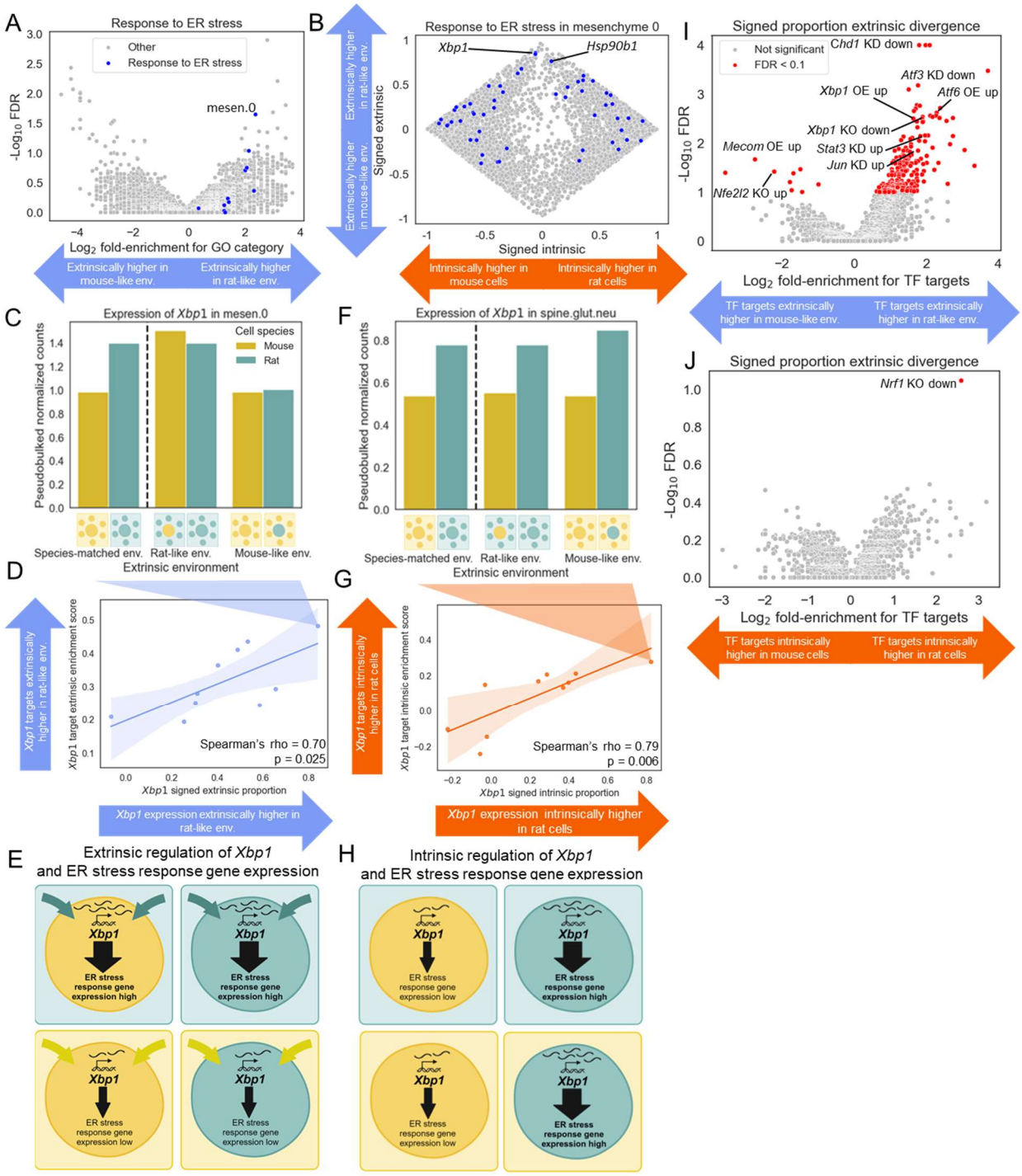



**Fig. 2: Decomposition of gene expression divergence into cell-intrinsic, cell-extrinsic, and interaction components. A)** Expression of *Efnb3* in forebrain glutamatergic neurons as an example of purely extrinsic divergence. On the left of the dashed line, the expression of *Efnb3* is shown in species-matched environments. To the right of the dashed line, expression in mouse cells and rat cells is first shown in a rat-like environment and then in a mouse-like environment. The data used for the bars in species-matched environments to the left of the dashed line is

447 from the same host cells as the data in species-matched environments to the right of the
448 dashed line and is included for clarity. **B)** Expression of *Lsm6* in forebrain glutamatergic neurons
449 as an example of intrinsic divergence. **C)** Expression of *Plcxd2* as an example of opposing
450 extrinsic and intrinsic divergence. **D)** Expression of *Pou3f2* in forebrain glutamatergic neurons.
451 **E)** Schematic outlining how each gene's location in Fig. 2F indicates intrinsic and extrinsic
452 contributions to its expression divergence. Purely intrinsic divergence lies along the x-axis,
453 purely extrinsic divergence along the y-axis, and a combination of both (which can be
454 reinforcing or opposing in direction) occurs closer to the four diagonals. **F)** Plot of intrinsic and
455 extrinsic divergence in forebrain glutamatergic neurons. Each point represents a single gene
456 and the example genes shown in Fig. 2A-D are indicated on the graph. Points are colored by
457 the proportion of gene expression divergence explained by interaction. **G)** Barplot showing the
458 mean proportion of gene expression divergence explained by intrinsic divergence, extrinsic
459 divergence, and their interaction across all genes that met filtering criteria in each cell type.

460


A
Response to ER stress
Other
Response to ER stress
mesen.0
-Log10 FDR
Log2 fold-enrichment for GO category
Extrinsically higher in mouse-like env.
Extrinsically higher in rat-like env.
B
Response to ER stress in mesenchyme 0
Xbp1
Hsp90b1
Signed extrinsic
Signed intrinsic
Extrinsically higher in rat-like env.
Extrinsically higher in mouse-like env.
Intrinsically higher in mouse cells
Intrinsically higher in rat cells
I
Signed proportion extrinsic divergence
Not significant
FDR < 0.1
Chd1 KD down
Atf3 KD down
Xbp1 OE up
Atf6 OE up
Xbp1 KO down
Mecom OE up
Stat3 KD up
Jun KD up
Nfe2l2 KO up
-Log10 FDR
Log2 fold-enrichment for TF targets
TF targets extrinsically higher in mouse-like env.
TF targets extrinsically higher in rat-like env.
C
Expression of Xbp1 in mesen.0
Cell species
Mouse
Rat
Pseudobulked normalized counts
Species-matched env.
Rat-like env.
Mouse-like env.
Extrinsic environment
F
Expression of Xbp1 in spine.glut.neu
Pseudobulked normalized counts
Species-matched env.
Rat-like env.
Mouse-like env.
Extrinsic environment
J
Signed proportion extrinsic divergence
Nrf1 KO down
-Log10 FDR
Log2 fold-enrichment for TF targets
TF targets intrinsically higher in mouse cells
TF targets intrinsically higher in rat cells
D
Xbp1 targets extrinsically higher in rat-like env.
Xbp1 target extrinsic enrichment score
Spearman's rho = 0.70
p = 0.025
Xbp1 signed extrinsic proportion
Xbp1 expression extrinsically higher in rat-like env.
G
Xbp1 targets intrinsically higher in rat cells
Xbp1 target intrinsic enrichment score
Spearman's rho = 0.79
p = 0.006
Xbp1 signed intrinsic proportion
Xbp1 expression intrinsically higher in rat cells
E
Extrinsic regulation of Xbp1 and ER stress response gene expression
Xbp1
ER stress response gene expression high
ER stress response gene expression low
H
Intrinsic regulation of Xbp1 and ER stress response gene expression
Xbp1
ER stress response gene expression low
ER stress response gene expression high

461

462

463 **Fig. 3: Extrinsic and intrinsic divergence in the expression of *Xbp1* and its target genes.**
464 **A)** Enrichment of endoplasmic reticulum stress response genes for positive extrinsic divergence
465 across cell types. Each point is a GO biological process category in a cell type and the points
466 corresponding to the Response to ER stress GO category are colored blue. The x-axis shows
467 the $\log_2$ fold-enrichment and the y-axis shows the $-\log_{10}$ false discovery rate. **B)** Scatterplot
468 showing signed proportion intrinsic divergence (x-axis) and signed proportion extrinsic
469 divergence (y-axis) for all genes passing our filtering criteria for mesenchymal cluster 0. ER
470 stress response genes are shown in blue and all other genes are shown in grey. **C)** Expression
471 of *Xbp1* in mesenchymal cluster 0. **D)** Correlation between signed extrinsic proportion
472 divergence for *Xbp1* (x-axis) and the enrichment of its target genes for signed extrinsic
473 proportion divergence (y-axis). Each point is a cell type and cell types with extrinsically driven
474 increased expression of *Xbp1* in a rat-like environment have larger values on the x-axis. Cell
475 types with extrinsically driven increased expression of *Xbp1* target genes in a rat-like
476 environment have larger values on the y-axis. The line and shaded region represent the best fit
477 and 95% confidence interval of a linear model fit to the data. **E)** Conceptual model for *Xbp1*-
478 driven increased expression of ER stress response genes in a rat-like environment occurring in
479 some cell types. Due to extrinsic factor(s), *Xbp1* expression is increased in a rat-like
480 environment compared to a mouse-like environment. This then up-regulates *Xbp1* target genes,
481 many of which are ER stress response genes, in a rat-like environment. **F)** Expression of *Xbp1*
482 in spinal glutamatergic neurons. **G)** Correlation between signed intrinsic proportion divergence
483 for *Xbp1* (x-axis) and the enrichment of its target genes for signed intrinsic proportion
484 divergence (y-axis). Each point is a cell type and cell types with intrinsically driven increased
485 expression of *Xbp1* in rat cells have larger values on the x-axis. Cell types with intrinsically
486 driven increased expression of *Xbp1* target genes in rat cells have larger values on the y-axis.
487 The line and shaded region represent the best fit and 95% confidence interval of a linear model
488 fit to the data. **H)** Conceptual model for *Xbp1*-driven increased expression of ER stress

489 response genes in rat cells occurring in some cell types. Due to intrinsic factor(s), *Xbp1*
490 expression is increased in rat cells compared to mouse cells. This then up-regulates *Xbp1*
491 target genes, many of which are ER stress response genes, in rat cells compared to mouse
492 cells. **I)** Volcano plot of enrichment of TF target genes for signed extrinsic proportion
493 divergence. The y-axis shows the negative $\log_{10}$(FDR). The x-axis shows the $\log_2$ fold-
494 enrichment with target sets enriched for higher expression in rat-like environments on the right
495 and mouse-like environments on the left. KO stands for knockout, KD stands for knock down,
496 OE stands for overexpression, up implies increased expression after the experimental
497 manipulation, and down implies decreased expression after the experimental manipulation. **J)**
498 Volcano plot of enrichment of TF target genes for signed intrinsic proportion divergence. The y-
499 axis shows the negative $\log_{10}$(FDR). The x-axis shows the $\log_2$ fold-enrichment with target sets
500 enriched for higher expression in rat cells on the right and in mouse cells on the left.
501
502
503
504

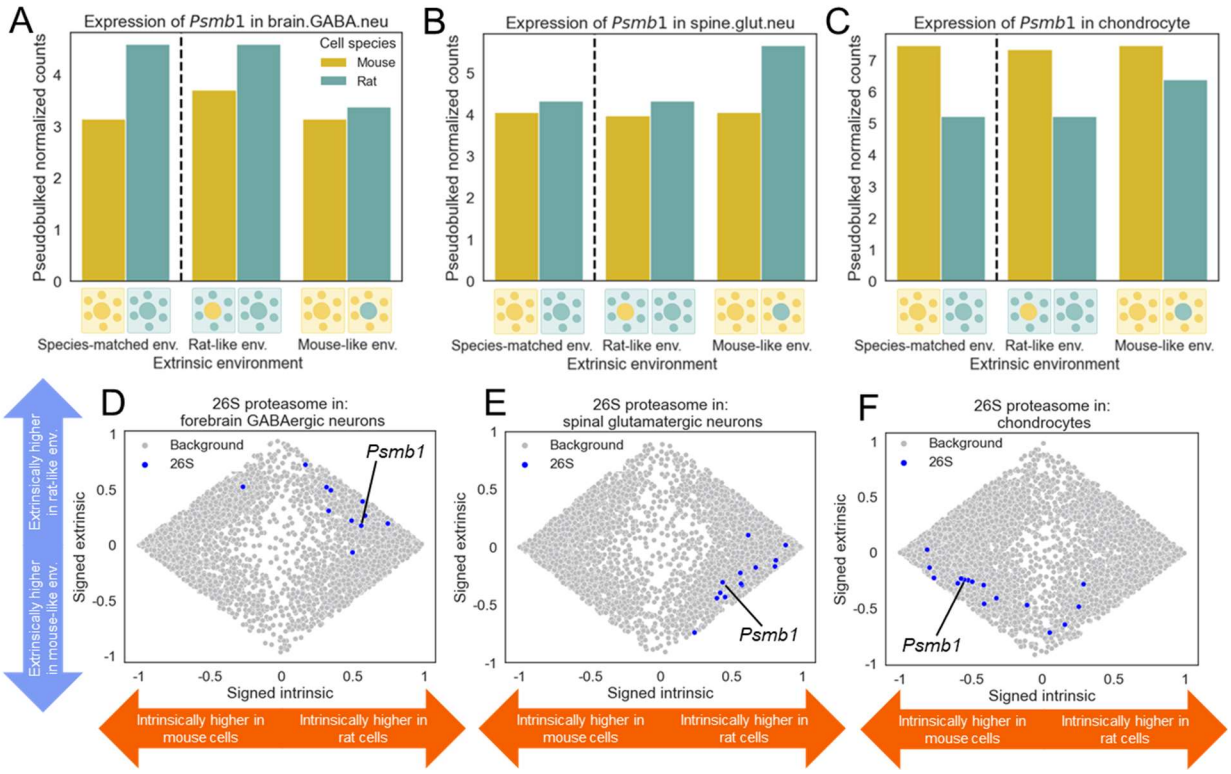
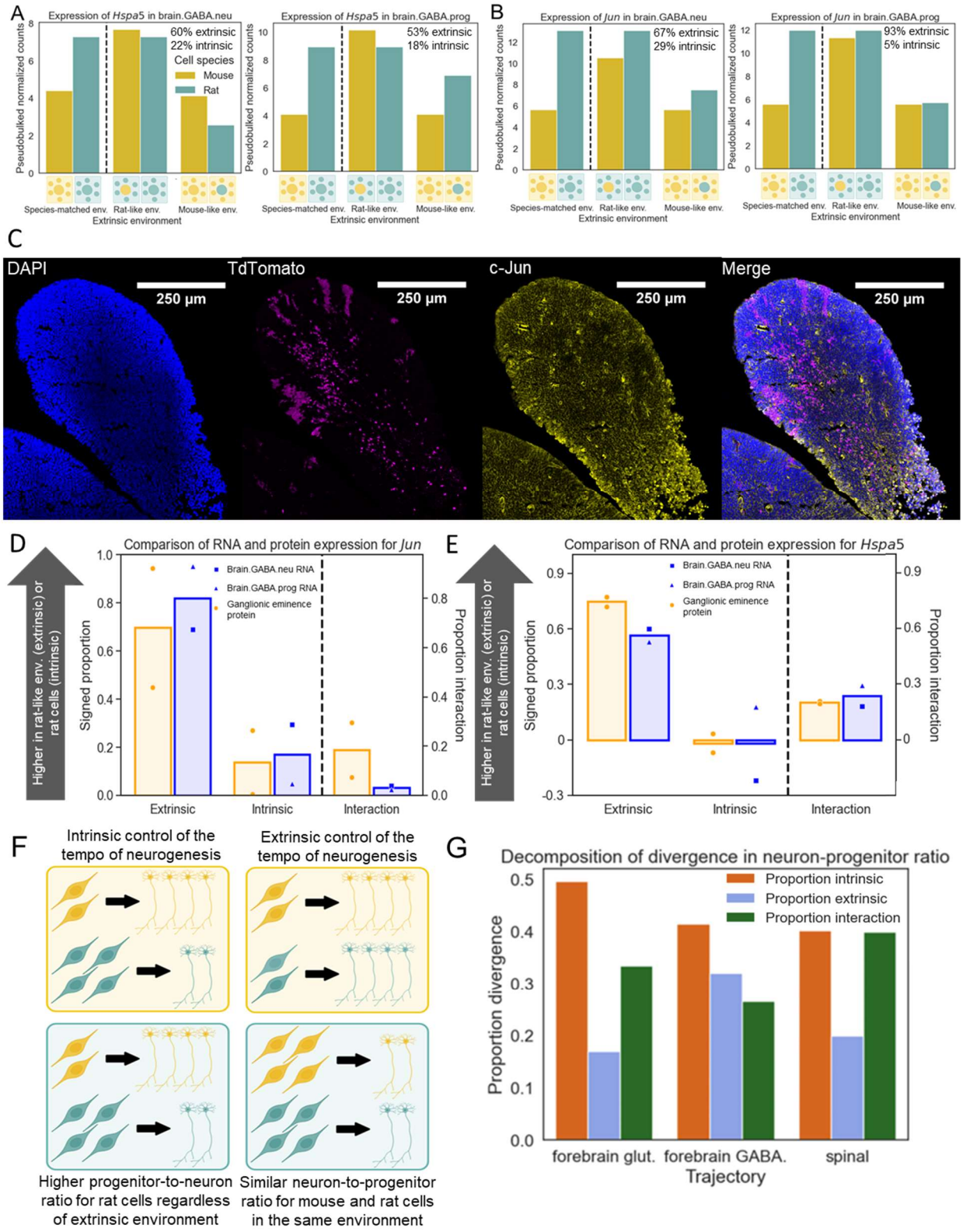
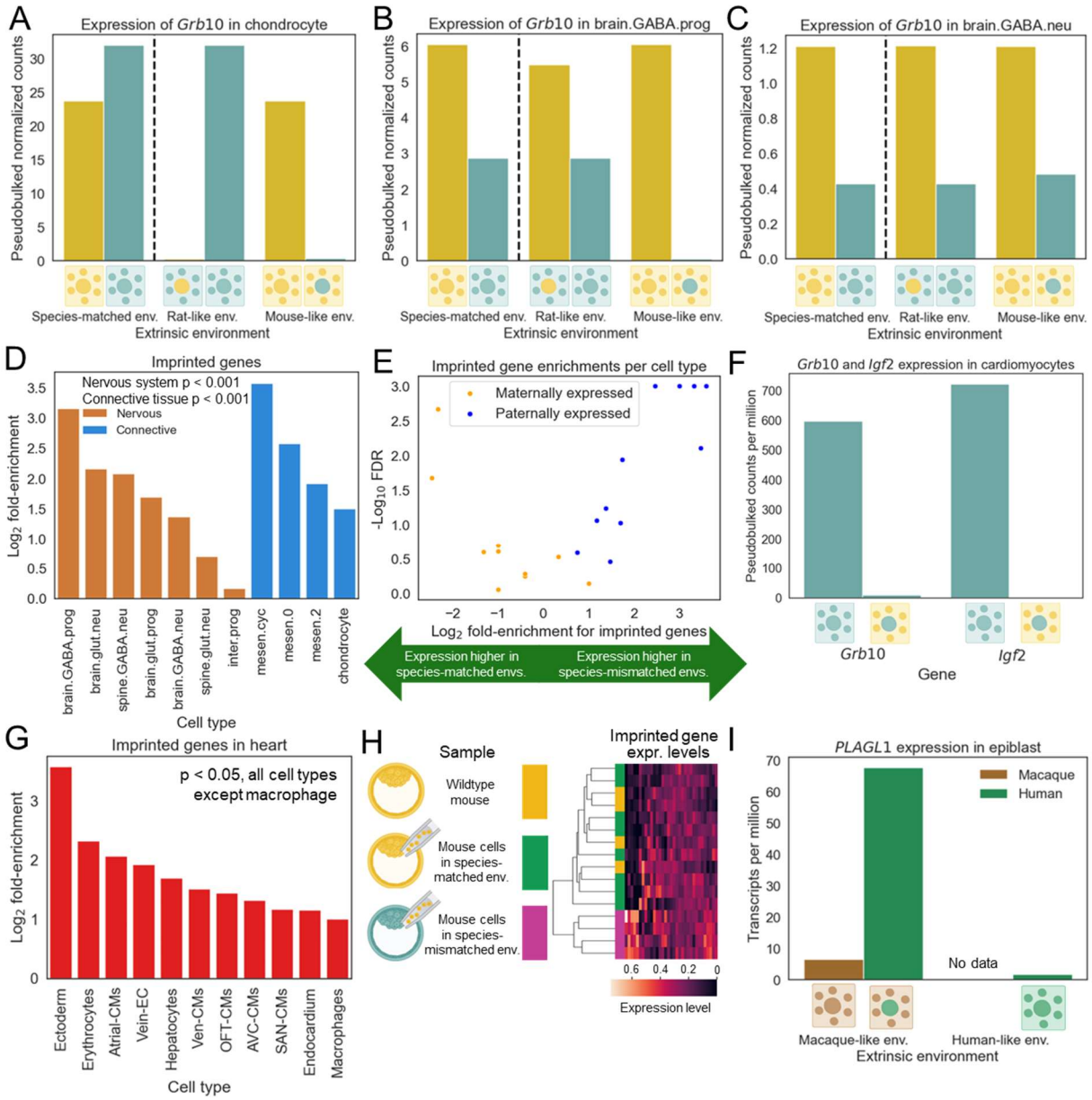



**Fig. 4: Cell-type specific intrinsic and extrinsic divergence of the expression of genes encoding proteasomal subunits across cell types. A)** Expression of *Psmb1* in forebrain GABAergic neurons. **B)** Same as in (A) but for spinal glutamatergic neurons. **C)** Same as in (A) but for chondrocytes. **D)** Scatterplot showing signed proportion intrinsic divergence (x-axis) and signed proportion extrinsic divergence (y-axis) for all genes passing our filtering criteria for forebrain GABAergic neurons. Genes coding for 26S proteasomal subunits are shown in blue and all other genes are shown in grey. **E)** Same as in (D) but for spinal glutamatergic neurons. **F)** Same as in (D) but for chondrocytes.


A
Expression of Hspa5 in brain.GABA.neu
60% extrinsic
22% intrinsic
Cell species
Mouse
Rat
Expression of Hspa5 in brain.GABA.prog
53% extrinsic
18% intrinsic
Pseudobulked normalized counts
Species-matched env.
Rat-like env.
Mouse-like env.
Extrinsic environment
B
Expression of Jun in brain.GABA.neu
67% extrinsic
29% intrinsic
Expression of Jun in brain.GABA.prog
93% extrinsic
5% intrinsic
C
DAPI
TdTomato
c-Jun
Merge
250 µm
D
Comparison of RNA and protein expression for Jun
Higher in rat-like env. (extrinsic) or rat cells (intrinsic)
Brain.GABA.neu RNA
Brain.GABA.prog RNA
Ganglionic eminence protein
Signed proportion
Proportion interaction
Extrinsic
Intrinsic
Interaction
E
Comparison of RNA and protein expression for Hspa5
F
Intrinsic control of the tempo of neurogenesis
Extrinsic control of the tempo of neurogenesis
Higher progenitor-to-neuron ratio for rat cells regardless of extrinsic environment
Similar neuron-to-progenitor ratio for mouse and rat cells in the same environment
G
Decomposition of divergence in neuron-progenitor ratio
Proportion intrinsic
Proportion extrinsic
Proportion interaction
Proportion divergence
forebrain glut.
forebrain GABA.
spinal
Trajectory

515

516 **Fig. 5: Extrinsic divergence of *Jun* and *Hspa5* at the RNA and protein levels. A)**
517 Expression of *Hspa5* in forebrain GABAergic neurons (left) and progenitors (right). The
518 percentage of gene expression divergence explained by extrinsic and intrinsic factors is shown
519 in the upper right of each plot. **B)** The same as in (A) but showing the expression of *Jun*, which
520 codes for the protein c-Jun. **C)** Immunofluorescent staining for c-Jun in the medial ganglionic
521 eminence for a rat-like chimera. Donor mouse cells are marked with TdTomato and c-Jun
522 staining is shown in yellow. **D)** Comparison of intrinsic, extrinsic, and interaction divergence at
523 the RNA and protein level for *Jun*. To the left of the dashed line, the signed proportion
524 divergence is along the y-axis, with larger values indicating a greater proportion of either
525 intrinsic or extrinsic divergence leading to higher expression in rat. The extrinsic and intrinsic
526 divergence for protein levels across two chimera pairs is shown in blue and the divergence in
527 RNA expression in orange for GABAergic neurons (square marker) and neurons (triangle
528 marker). The bars indicate the average across replicates for protein and the two cell types for
529 RNA. To the right of the dashed line, the proportion of interaction divergence (which is non-
530 negative by definition) is shown with its own y-axis. **E)** Similar to (D) but showing *Hspa5*
531 divergence. **F)** Conceptual model for extrinsic or intrinsic divergence in the pace of
532 neurogenesis. The top outlines intrinsic divergence as mouse cells have a higher proportion of
533 neurons than progenitors regardless of whether cells are in a rat-like or mouse-like environment.
534 The bottom outlines extrinsic divergence in which the ratio of neurons to progenitors is similar
535 for mouse and rat cells in rat-like environment and similar for mouse and rat cells in a mouse-
536 like environment. **G)** Proportion intrinsic, extrinsic, and interaction divergence for neuron-
537 progenitor ratio in forebrain glutamatergic neurogenesis, forebrain GABAergic neurogenesis,
538 and spinal GABAergic and glutamatergic neurogenesis together.
539
540


A
Expression of Grb10 in chondrocyte
Pseudobulked normalized counts
Species-matched env.
Rat-like env.
Mouse-like env.
Extrinsic environment
B
Expression of Grb10 in brain.GABA.prog
C
Expression of Grb10 in brain.GABA.neu
D
Imprinted genes
Nervous system p < 0.001
Connective tissue p < 0.001
Nervous
Connective
Log2 fold-enrichment
brain.GABA.prog
brain.glut.neu
spine.GABA.neu
brain.glut.prog
brain.GABA.neu
spine.glut.neu
inter.prog
mesen.cyc
mesen.0
mesen.2
chondrocyte
Cell type
E
Imprinted gene enrichments per cell type
Maternally expressed
Paternally expressed
-Log10 FDR
Log2 fold-enrichment for imprinted genes
Expression higher in species-matched envs.
Expression higher in species-mismatched envs.
F
Grb10 and Igf2 expression in cardiomyocytes
Pseudobulked counts per million
Grb10
Igf2
Gene
G
Imprinted genes in heart
p < 0.05, all cell types except macrophage
Log2 fold-enrichment
Ectoderm
Erythrocytes
Atrial-CMs
Vein-EC
Hepatocytes
Ven-CMs
OFT-CMs
AVC-CMs
SAN-CMs
Endocardium
Macrophages
Cell type
H
Sample
Wildtype mouse
Mouse cells in species-matched env.
Mouse cells in species-mismatched env.
Imprinted gene expr. levels
Expression level
I
PLAGL1 expression in epiblast
Macaque
Human
Transcripts per million
No data
Macaque-like env.
Human-like env.
Extrinsic environment

541

542 **Fig. 6: Misexpression of imprinted genes in species-mismatched environments. A)**
543 Expression of the imprinted gene *Grb10* in chondrocytes. **B)** Expression of *Grb10* in forebrain
544 GABAergic progenitors. **C)** Expression of *Grb10* in forebrain GABAergic neurons. **D)**
545 Enrichment of imprinted genes for high absolute interaction divergence across cell types. The
546 bars represent the $\log_2$ fold-enrichment. **E)** Enrichment of maternally expressed and paternally
547 expressed genes for signed interaction divergence. Each point represents one of the eleven cell
548 types analyzed in this study. The y-axis shows the $-\log_{10}$(FDR) and the x-axis shows the $\log_2$
549 fold-enrichment. A negative enrichment score indicates higher expression in species-matched
550 environments, whereas a positive enrichment score indicates higher expression species-
551 mismatched environments. **F)** Expression of imprinted genes *Grb10* and *Igf2* in ventricular
552 cardiomyocytes. **G)** The same as in (D), but for cell types in the heart dataset and using the
553 absolute log fold-change between expression in a mouse-like environment and expression in a
554 rat-like environment to rank genes when computing enrichments. **H)** Heatmap of imprinted gene
555 expression across samples for adult parathyroid data. Samples were hierarchically clustered
556 using the Euclidean distance metric. **I)** Expression of the imprinted gene *PLAGL1* in macaque
557 and human epiblast. No data are shown for macaque cells in a human-like environment since
558 these have not been reported.

559 **Materials and methods**

560

561 **Animals**

562 Seven-week-old Wistar female rats and 10-week-old male rats were purchased from Charles
563 River Laboratory (Wilmington, MA; CRL: 003). Seven-week-old CD1 female mice and 10-week-
564 old male mice (CRL: 022) were purchased from Charles River Laboratory. Seven-week-old
565 C57BL/6 female mice and 10-week-old male mice were purchased from Jackson Laboratories
566 (Bar Harbor, ME: 000664). Littermates of the same sex were randomly assigned to experimental
567 groups. All rats and mice were housed in pathogen-free conditions with free access to food and
568 water. All animal protocols were approved by the Administrative Panel on Laboratory Animal Care
569 at Stanford University.

570 **ESC culture**

571 Undifferentiated tdTomato-labelled or GFP-labeled mouse ESCs (SUN106.2 and SGE2) were
572 maintained on mitomycin-c treated mouse embryonic fibroblasts (MEFs) in N2B27 medium(*45*)
573 containing 1000 U/ml LIF (Peprotech, Cranbury, NJ; 300-05), 1 μM MEK inhibitor PD0325901
574 (Tocris, Barton Ln, Abingdon, United Kingdom; 4192) and 3 μM GSK3 inhibitor CHIR99021
575 (Tocris: 4423). Undifferentiated tdTomato-labelled rat ESCs were maintained in N2B27 medium
576 containing 1 μM MEK inhibitor PD0325901, 3 μM CHIR99021, and 1000 U/ml of rat LIF as
577 described(*46*). Their pluripotency was confirmed by chimera generation assay. All PSC lines used
578 in this study were male lines.

579 **Embryo culture and manipulation**

580 Wild-type mouse embryos were prepared according to published protocols(*47*). In brief, morula-
581 stage embryos were obtained by uterus perfusion from superovulated CD1 mice at 2.5 days

582 postcoitum (dpc). Morula-stage embryos were cultured in KSOM-AA medium (CytoSpring,
583 Mountain View, CA; K0101) for 1 day and developed to blastocyst-stage embryos. Wild-type rat
584 blastocysts were obtained by uterus perfusion from female rats at 4.5 dpc and cultured in Rat
585 KSOM medium. For micromanipulation, mouse or rat ESCs were trypsinized and suspended in
586 mouse or rat ESC culture medium. A piezo-driven micromanipulator (Prime Tech, Tsuchiura,
587 Japan) was used to pierce the zona pellucida and trophectoderm under microscopy and 5–7
588 mouse or rat ESCs were introduced into blastocyst cavities near the inner cell mass. After
589 blastocyst injection, embryos were cultured for 1–2 hours. Mouse blastocysts were then
590 transferred into uteri of pseudopregnant recipient CD1 female mice at 2.5 dpc. Rat blastocysts
591 were then transferred into uteri of pseudopregnant recipient Wistar female rats at 3.5 dpc. Table
592 S1 shows results of the cell injection.

593 **Chimera dissection and cell preparation**

594 Embryonic day (E) 15.25 mouse-rat and E13.5 rat-mouse chimeras were dissected to harvest
595 forebrain and connective tissue. In brief, chimeric fetuses were collected from either E15.25 rat
596 or E13.5 mouse pregnant mothers and forebrain and connective tissues were harvested from
597 those chimeric fetuses. Both forebrains and connective tissues were dissociated in the solution
598 containing 2mg/ml Collagenase/Dispase (Roche, Basel, Switzerland; 10269638001) in Hanks'
599 Balanced Salt solution. After 30-60 min incubation at 37 degrees, 10% fetal bovine serum (FBS)
600 in PBS solution was added into the tissue solution to inactivate the enzymes. All the dissociated
601 cells were then filtered and used for the subsequent experiment.

602 **Flow cytometry**

603 Dissociated cells derived from forebrain and connective tissue were stained with APC-anti-mouse
604 CD45 antibody (Biolegend, San Diego, CA; 103112), PE-Cy7-anti rat CD45 (Biolegend: 202214).

605 Donor chimerism was analyzed by detecting CD45 negative, tdTomato or GFP-expressing cells
606 (detailed in Table S2). Both donor and host cells in the forebrain and connective tissue of chimeric
607 fetuses were sorted using a FACS Aria II (BD, Franklin Lakes, NJ).

608 **Single cell RNA-sequencing**

609 The chimeric fetuses from the CD1-Wistar chimeras that presented less than 10% donor
610 contribution to either forebrain or connective tissue were used for single cell RNA-sequencing.
611 The connective tissue and forebrain of three rat-like chimeras, three mouse-like chimeras, one
612 wildtype mouse, and one wildtype rat were used for scRNA-seq using 10x 3' v3.1 chemistry. The
613 connective tissue and forebrain were from different chimeras but pooled together before the cells
614 entered the 10x controller. Of the resulting eight libraries, the rat-like chimera in library MR1 had
615 high quality data for forebrain and connective tissue, the mouse-like chimera in library RM1 had
616 sufficiently high quality data for the neuronal lineage, the mouse-like chimera in library RM2 had
617 sufficiently high quality data for connective tissue, and both the wildtype mouse and rat libraries
618 had sufficiently high quality data for both tissue types.

619

620 **Histological analysis**

621 Tissues were fixed with 4% paraformaldehyde and embedded in Optimal Cutting Temperature
622 (O.C.T.) compound (Sakura Finetek, Tokyo, Japan). OCT-embedded sections were stained with
623 antibodies for fluorescence microscopic analysis. Each section was incubated with the primary
624 antibodies (Abs) for 24 hours at 4 degrees and with the secondary Abs for 1 hour at room
625 temperature (detailed in Table S3). Following a wash step, sections were mounted with Vibrance
626 Antifade Mounting Medium with DAPI (Vector LABORATORIES, Newark, CA; H-1800), and
627 observed under confocal laser scanning microscopy (Carl Zeiss, Oberkochen, Baden-
628 Württemberg, Germany).

629

630 **Alignment, species deconvolution, and clustering**

631

632 All fastq files were aligned to the mouse (GRCm39, from ENSEMBL v109), rat (mRatBN7.2,
633 from ENSEMBL v109), and a concatenated mouse and rat genome using CellRanger
634 v7.1.0(*48*). For mapping, we retained protein-coding genes, lncRNA, antisense transcripts,
635 snoRNA, snRNA, miRNA, and scaRNA. The resulting count matrices were read into scanpy
636 v1.9.5 and the proportion of aligned reads to each genome in the concatenated genome was
637 used to determine the species of origin of each cell(*49*). A cell was assigned to mouse if greater
638 than or equal to 70% of aligned reads aligned to the mouse genome, assigned to rat if greater
639 than or equal to 70% of aligned reads aligned to the rat genome, and discarded as a doublet
640 otherwise. For all subsequent steps, we only used the mouse-aligned-only counts for mouse
641 cells and the rat-aligned-only counts for rat cells. We then removed all genes on sex
642 chromosomes, restricted to mouse-rat one-to-one orthologs (defined using data from
643 ENSEMBL), and performed standard pre-processing steps to remove low quality cells(*48*). For
644 the latter, we performed standard filtering and removed cells with greater than 15% of counts
645 coming from mitochondrial reads and cells with n_genes_by_counts (a commonly used quality
646 control metric) greater than 7500. Next, we converted to the log normalized counts, identified
647 highly variable genes, computed principal components, used the Python implementation of
648 harmony v0.0.9 to integrate the mouse and rat cells, and then found nearest neighbors(*50*). We
649 then clustered the cells using the Leiden algorithm v0.10.2 with resolution equal to 0.1(*51*). This
650 first round of clustering split cells into neuronal, connective tissue, and hematopoietic lineages.
651 The neuronal and connective tissue lineage cells were then retained for further analysis (see
652 below).

653

654 Several libraries did not contain a sufficient number of cells to be useful for our analysis.
655 Therefore, we used the rat-like chimera data from library MR1 for both neural and connective
656 tissue analyses, the mouse-like chimera data from library RM1 for neural analyses, and the
657 mouse-like chimera data from library RM2 for all subsequent analyses after the pseudobulking
658 step (see below).
659
660 Our main goal in clustering the cells was to identify cell types and remove cell subtypes that had
661 fewer than 10 cells in any of donor mouse cells, donor rat cells, host mouse cells, or host rat
662 cells in MR1 and RM1 (for neuronal lineages) or MR1 and RM2 (for connective tissue). We
663 repeated the harmony and Leiden-based procedure (with resolution equal to 0.5) on the neural
664 lineages and identified 18 subclusters. Due to a low number of cells in other cell types, we
665 restricted further analysis to forebrain glutamatergic progenitors (*Gli3* high, *Eomes*-, *Mki67*+),
666 forebrain GABAergic progenitors (*Gli3* low, *Eomes*-, *Mki67*+), intermediate progenitors
667 (*Eomes*+, *Mki67*+), forebrain glutamatergic neurons (*Slc17a6*+, *Tcf7l2*-, *Reln*-, *Hoxb8*-, *Mki67*-),
668 forebrain GABAergic neurons (*Gad1*+, *Hoxb8*-), spinal glutamatergic neurons (*Slc17a6*+,
669 *Hoxb8*+), and spinal GABAergic neurons (*Gad1*+, *Hoxb8*+)(*52–54*). As some forebrain
670 GABAergic subtypes had very few rat donor cells, we subclustered (Leiden resolution equal to
671 0.35) that cell type further, removed the subtypes with very few rat donor cells, and then merged
672 the remaining 3 cell subtypes which had similar cell type proportions across the four species-
673 environment combinations.
674
675 For the connective tissue, we reclustered the cells (Leiden resolution equal to 0.15) and
676 identified six clusters. We focused subsequent analysis on mesenchymal cells (*Postn* high,
677 *Prrx1* high) and chondrocytes (*Acan*+)(*55*). We subclustered both cell types (Leiden resolution
678 equal to 0.25 for mesenchymal cells and chondrocytes) and removed subtypes according to the
679 above criteria (i.e. at least 10 cells in each of the 4 species-environment combinations). After

680 this procedure, we retained one chondrocyte subcluster, two mesenchymal subclusters, and
681 cycling mesenchymal cells (*Mki67*+) for further analysis.
682
683 **Gene filtering, pseudobulking, and count normalization**
684
685 To avoid overly noisy estimates of intrinsic and extrinsic divergence, we removed genes that
686 were not expressed in greater than or equal to 20% of cells in at least one of the four species-
687 environment combinations. For each gene, we summed all counts across cells within a cell type
688 in each library for mouse and rat cells separately and removed genes that had fewer than ten
689 counts in all four species-environment combinations.
690
691 To normalize counts, we divided by total expression and multiplied by 10,000 for each
692 pseudobulked sample in each cell type separately. Rather than normalize the counts directly,
693 we first randomly sampled from the gene-count distribution within each sample in each cell type
694 so that the total number of counts was identical across all species-environment combinations.
695 We performed this procedure 100 times and took the mean normalized counts across all 100
696 samplings. This mean value was highly correlated with the actual normalized counts
697 (Spearman's rho > 0.999, $p < 10^{-300}$). This procedure was used instead of simply normalizing by
698 total counts because this can lead to biased estimation of log fold-changes when there are few
699 counts in one of the four species-environment combinations. For example, consider a scenario
700 in which donor mouse, host mouse, and host rat all have 1,000,000 total counts whereas donor
701 rat only has 100,000 counts. If a gene has 99 (100 after adding a pseudocount) counts in donor
702 mouse and host mouse and 0 (1 after adding a pseudocount) counts for donor rat and host rat
703 and we normalize to total counts and multiply by 10,000, we will have 1 normalized count in
704 donor mouse and host mouse, 0.01 normalized counts in host rat, and 0.1 normalized counts in
705 donor rat. If we then compute fold-changes, donor mouse divided by host rat would be 100 but

706 donor mouse divided by donor rat would be 10. This difference would propagate and (see
707 below) cause overestimation of the extrinsic or interaction component of gene expression and
708 underestimation of the intrinsic component.
709
710 From here, we will refer to normalized counts for mouse cells in a rat-like environment as DM
711 (donor mouse), counts for rat cells in a rat-like environment as HR (host rat), counts for mouse
712 cells in a mouse-like environment as HM (host mouse), and counts for rat cells in a mouse-like
713 environment as DR (donor rat).
714
715 **A framework to decompose gene expression divergence into extrinsic, intrinsic, and**
716 **interaction components**
717
718 Throughout, we use the $\log_2$ fold-change to measure gene expression divergence. We formulate
719 the total expression of a single gene in one of the species-environment combinations as the
720 product of an extrinsic component and an intrinsic component. For example, the expression of a
721 gene in a donor rat cell in a mouse-like environment ($D_R$) is equal to the intrinsic rat expression
722 ($I_R$) multiplied by the expression induced by the mouse-like environment ($E_M$). In this framework,
723 the normalized counts are:
724
725 $$H_M = I_M \times E_M$$
726 $$H_R = I_R \times E_R$$
727 $$D_M = I_M \times E_R$$
728 $$D_R = I_R \times E_M$$
729

730 Where $H_M$ is the gene expression for a host mouse cell in a mouse-like environment, $H_R$ is the
731 gene expression for a host rat cell in a rat-like environment, $D_M$ is the gene expression for a
732 donor mouse cell in a rat-like environment, and $D_R$ is the gene expression for a donor rat cell in
733 a mouse-like environment. Using these four values with three degrees of freedom, we can
734 obtain two independent estimates of the intrinsic divergence between mouse and rat ($I_1$ and $I_2$):
735

736 $$I_1 = log_2\left(\frac{H_R}{D_M}\right) = log_2\left(\frac{I_R \times E_R}{I_M \times E_R}\right) = log_2\left(\frac{I_R}{I_M}\right)$$

737 $$I_2 = log_2\left(\frac{D_R}{H_M}\right) = log_2\left(\frac{I_R \times E_M}{I_M \times E_M}\right) = log_2\left(\frac{I_R}{I_M}\right)$$

738

739 Similarly, we obtain two independent estimates of the extrinsic divergence between mouse and
740 rat ($E_1$ and $E_2$):

741 $$E_1 = log_2\left(\frac{D_M}{H_M}\right) = log_2\left(\frac{I_M \times E_R}{I_M \times E_M}\right) = log_2\left(\frac{E_R}{E_M}\right)$$

742 $$E_2 = log_2\left(\frac{H_R}{D_R}\right) = log_2\left(\frac{I_R \times E_R}{I_R \times E_M}\right) = log_2\left(\frac{E_R}{E_M}\right)$$

743

744 However, for some genes the effects of a rat-like environment on mouse cells will likely differ
745 from the effects of a rat-like environment on rat cells (and vice versa) due to interactions
746 between extrinsic and intrinsic divergence. In the absence of such an interaction, the two
747 estimates of intrinsic divergence should be equal and the two estimates of extrinsic divergence
748 should be equal, so the difference between these pairs of estimates is itself an estimate of the
749 interaction between extrinsic and intrinsic divergence. Therefore, to quantify this interaction, we
750 can compute:
751

752 $$I_1 - I_2 = log_2\left(\frac{H_R}{D_M}\right) - log_2\left(\frac{D_R}{H_M}\right) = log_2\left(\frac{H_R \times H_M}{D_M \times D_R}\right) = 2N$$

This value is identical to what we obtain starting with extrinsic divergence:

$$E_2 - E_1 = log_2\left(\frac{H_R}{D_R}\right) - log_2\left(\frac{D_M}{H_M}\right) = log_2\left(\frac{H_R \times H_M}{D_M \times D_R}\right) = 2N$$

Based on this framework, we obtain our final estimates of intrinsic, extrinsic, and interaction divergence by averaging the two estimates:

$$E = \frac{E_2 + E_1}{2}$$

$$I = \frac{I_1 + I_2}{2}$$

$$N = \frac{2N}{2}$$

To compute the proportion extrinsic, intrinsic, and interaction divergence, we take the magnitude of each quantity and divide it by the sum of the magnitudes of all three quantities:

$$I_p = \frac{|I|}{|I| + |E| + |N|}$$

$$E_p = \frac{|E|}{|I| + |E| + |N|}$$

$$N_p = \frac{|N|}{|I| + |E| + |N|}$$

Finally, the signed proportion extrinsic, intrinsic, and interaction divergence is then computed as the sign of the divergence multiplied by the proportion divergence for each quantity. We only

774 applied this framework to genes that had some evidence for divergence in expression level (i.e.
775 an absolute $\log_2$ fold-change greater than 0.5 in at least one of the four comparisons).
776
777 **Analysis of bulk RNA-seq data and developmentally dynamic gene expression**
778
779 To assess how developmentally dynamic gene expression might influence our results, we
780 focused on the E11.5, E13.5, and E15.5 (or their closest approximation e.g. E15 for rats)
781 timepoints collected from brain tissue in Cardoso-Moreira et al(*56*). For mice and rats
782 separately, we computed the Pearson correlation between expression level and developmental
783 stage(*57*). We then classified genes as having similar, opposite, or a shift in developmentally
784 dynamic gene expression that goes against the global difference in maturation rate between
785 mouse and rat (referred to as "temporal shift") according to the following criteria. We considered
786 a gene as having a similar change over development if the product of the correlation coefficients
787 for mice and rats was greater than 0.7 (increasing if both correlation coefficients were positive,
788 decreasing if both correlation coefficients were negative). Similarly, we classified genes as
789 opposite if the product of the correlation coefficients was less than -0.25. These different cutoffs
790 were chosen such that there were a sufficient number of genes for the analysis. As there were
791 many fewer genes with opposite expression trajectories over development, a less strict cutoff (-
792 0.25) was used. To classify genes as having a temporal shift, we restricted to the set of genes
793 with similar changes in expression over time and required that the absolute log fold-change
794 between mouse and rat at both E13.5 and E15.5 be less than 0.25, consistent with the logic
795 outlined in the supplementary text.
796
797 To further subdivide genes, we considered a gene as having higher expression in mouse if the
798 $\log_2$ fold-change of mouse expression divided by rat expression was greater than 0.5 and
799 considered a gene as having higher expression in rat if the $\log_2$ fold-change was less than -0.5.

800 Restricting to only the genes with similar gene expression trajectories in mouse and rat (i.e. the
801 product of the mouse and rat correlation coefficients was greater than 0.7), we classified genes
802 as increasing in expression over time if a gene had a positive correlation coefficient in mice and
803 rats and as decreasing over time if a gene had a negative correlation coefficient in both mice
804 and rats (for both decreasing and increasing genes, the product would be positive.
805
806 In general, to investigate the role of developmentally dynamic gene expression in our results
807 (see Supplementary Text 1 and Supplementary Figs. 4-9), we defined some gene set that we
808 would expect to be enriched for a target divergence (opposing extrinsic and intrinsic divergence,
809 reinforcing extrinsic and intrinsic divergence, mostly extrinsic divergence, or mostly interaction
810 divergence) under a model of purely intrinsic divergence combined with some form of
811 developmentally dynamic gene expression. We then used GSEAPY v1.0.6 preranked to test for
812 enrichment of that gene set at the top or bottom of the gene list when ranking by the target
813 (signed) divergence(*25*). Throughout the next section, we describe each gene set in terms of
814 what we would expect under the purely intrinsic model outlined above. We also always use
815 GSEAPY preranked with the following parameters: threads=4, permutation_num=1000,
816 format='png', seed=6, min_size = 10, max_size = 30000(*25*).
817
818 To define the first gene set in which we expect enrichment of both opposing and reinforcing
819 genes, we took the union of genes with similar expression trajectories in mice and rats and
820 genes that were classified as either higher in mouse or higher in rat. We then ranked genes by
821 the product of proportion intrinsic and proportion extrinsic and used GSEAPY preranked to test
822 for enrichment(*25*). To further investigate genes with opposing or reinforcing extrinsic and
823 intrinsic divergence, we ranked genes by the product of the signed extrinsic divergence and
824 signed intrinsic divergence so that genes with a large positive value for this metric are
825 reinforcing and genes with a large negative value are opposing. We then defined the gene set in

826 which we expect enrichment of opposing genes as the union of genes that either (1) were
827 increasing over time and had higher expression in mouse or (2) were decreasing over time and
828 had higher expression in rat. Similarly, for the gene set in which we expect enrichment of
829 reinforcing genes we took the union of genes that either (1) were decreasing over time and had
830 higher expression in mouse or (2) were increasing over time and had higher expression in rat.
831 We then ranked genes by the product of the signed intrinsic proportion and signed extrinsic
832 proportion (so that genes with a very low value are opposing and genes with a very high value
833 are reinforcing) and used GSEAPY preranked(*25*).
834
835 To investigate purely extrinsic genes, we defined the gene set in which we expect enrichment of
836 extrinsic genes as the temporal shift genes (defined above), sorted by extrinsic proportion, and
837 used GSEAPY preranked(*25*). Next, we split temporal shift genes into decreasing (expected to
838 be enriched for negative extrinsic divergence) and increasing (expected to be enriched for
839 positive extrinsic divergence) and input both gene sets, sorted by signed extrinsic proportion,
840 and used GSEAPY preranked(*25*).
841
842 Finally, to investigate interaction divergence genes we defined the gene set we expect to be
843 enriched for interaction divergence as the opposing genes (defined above), sorted by proportion
844 interaction, and used GSEAPY preranked(*25*). Next, we defined genes expected to have higher
845 expression in species-matched environments as genes with increasing expression in rat and
846 decreasing expression in mouse and genes expected to have higher expression in species-
847 mismatched environments as genes with increasing expression in mouse and decreasing
848 expression in rat. We then sorted the input gene list by the signed interaction proportion and
849 tested for enrichment using GSEAPY preranked(*25*).
850
851 **Data processing and analysis for predictors of intrinsic and extrinsic divergence**

852

853 We tested six variables for their correlation with [intrinsic proportion - extrinsic proportion]. In
854 general, if high quality data were not available for both rats and mice, we used data from the
855 human genome to avoid biasing our analysis. To measure gene expression divergence and
856 total expression in mice and rats, we used the $\log_2$ fold-change between and average
857 expression of mouse cells in a mouse environment and rat cells in a rat environment. To
858 measure evolutionary constraint, we used the metric for constraint on each gene published in
859 Zeng et al(*58*). As a proxy for regulatory complexity, we assigned candidate *cis*-regulatory
860 elements (CREs) from the ENCODE project to the nearest TSS of a protein-coding gene,
861 removed CREs greater than 100,000 bases from the nearest TSS of a protein coding gene, and
862 computed the total number of regulatory elements assigned to each gene(*59*). For cell type
863 specificity, we used total expression (as described above) for each of the eleven cell types in
864 our analysis as input to compute Tau, a well-established measure of tissue/cell type-specificity.
865 Tau is equal to one if a gene is completely specific to a single cell type or tissue, and zero if it is
866 equally expressed across all cell types or tissues(*60*). To compute tissue specificity in mice and
867 rats at E13.5 and E15.5 respectively, we averaged normalized counts across replicates for six
868 of the seven tissues from Cardoso-Moreira et al. (we excluded cerebellum/hindbrain due to its
869 transcriptional similarity to the rest of the brain) and computed Tau(*56*). Finally, we computed
870 the $\log_2$ fold-change between mouse and rat within each tissue (again using E13.5 and E15.5
871 respectively and excluding cerebellum/hindbrain) and used the variance of the $\log_2$ fold-change
872 across tissues as a measure of the tissue-specificity of gene expression divergence.
873

874 To merge these six variables into one matrix we converted mouse gene names to human gene
875 names using the orthologs from the Ensembl database and joined all the above metrics
876 together, removing any genes for which one or more of these metrics could not be

877 computed(*48*). We then calculated the Spearman correlation between [intrinsic proportion -
878 extrinsic proportion] and each of these variables(*57*).
879
880 **Enrichment analysis**
881
882 For the enrichment analysis, we either performed analysis on each cell type separately or
883 averaged expression in connective tissue, the neural lineage (i.e. including spinal neurons with
884 the forebrain cells), or across all eleven cell types. In the first two cases, we required that a
885 gene have passed our filtering criteria (see above) in at least three cell types and in the latter
886 case we required a gene to have passed the filtering criteria in at least five cell types.
887 Throughout, we always use GSEAPY preranked with parameters: threads=4,
888 permutation_num=1000, format='png', seed=6, min_size = 10, max_size = 300(*25*). To identify
889 enriched gene sets in extrinsically driven genes, we sorted the input gene list by the signed
890 extrinsic proportion, converted mouse gene symbols to human gene symbols using orthologs
891 defined by Ensembl, and used GSEAPY preranked(*25*, *48*). We tested gene sets from the gene
892 ontology (GO) biological process (2023 version), GO cellular component (2023 version), GO
893 molecular function (2023 version), MGI Level 4 Mammalian Phenotypes (2021 version), and the
894 CORUM protein complex database(*61*–*63*). In all cases, we used the versions directly
895 supported by GSEAPY. After running on all cell types, we retained all terms across the five
896 ontologies with FDR < 0.25.
897
898 To identify shared and cell type-specific enrichments, we counted the number of times each
899 term occurred in this final merged file. To further investigate shared terms, we visualized the
900 signed extrinsic and signed intrinsic proportion as shown in Fig. 3G-H. The $\log_2$ fold-enrichment
901 shown in many subFigs., is defined as the $\log_2$ fold-change for the number of genes in the gene
902 set with signed proportion extrinsic above the rank cutoff determined by GSEAPY preranked

903 and the number of genes below an identical cutoff (or vice versa if the enrichment was for
904 negative proportion extrinsic divergence). For example, consider an enrichment for positive
905 extrinsic proportion with 500 total genes above the GSEAPY cutoff, 15 of which are imprinted
906 genes. We would then count all imprinted genes in the 500 genes with the most negative signed
907 extrinsic proportion. If, continuing the example, 0 genes of those 500 were in the gene set, the
908 final $\log_2$ fold-change would be:
909

910 $$log_2\left(\frac{15+1}{0+1}\right) = 4$$

911

912 **Transcription factor target enrichment analysis**

913

914 For the transcription factor (TF) target enrichment analysis, we focused on the averaged signed
915 extrinsic and intrinsic proportion across all cell types to reduce noise, but also performed the
916 analysis on each cell type separately. The analysis was performed as described above, but
917 using the TF Perturbations Followed By Expression database in GSEAPY/MSigDB(*25*). We
918 then computed the Spearman correlation between the extrinsic TF target enrichment score for
919 genes with increased expression following *Xbp1* overexpression (from GSEAPY as described
920 above) and the signed proportion extrinsic for *Xbp1.* With regard to the significant enrichment
921 for targets *Nfe2l1* (which codes for the protein Nrf1) for positive intrinsic divergence, the gene is
922 often erroneously referred to as *NRF1* in the database of TF targets and so the significant
923 enrichment for *NRF1* we observe actually refers to *Nfe2l1* (and is referred to as *Nfe2l1*)
924 throughout the text. We repeated this analysis using the TF target enrichment score and signed
925 proportion extrinsic proportion for *Nfe2l2* and *Stat3*. We computed a similar correlation using
926 *Xbp1* and *Nrf1* intrinsic signed proportion and the intrinsic TF target enrichment score for *Xbp1*
927 and *Nrf1* target genes respectively. For *Xbp1*, we also computed the correlation between the

928 signed extrinsic or intrinsic proportion and the enrichment score for the Response to
929 Endoplasmic Reticulum Stress (Response to ER stress) Gene Ontology category(*63*).
930 Importantly, a strong correlation between signed extrinsic divergence in *Xbp1* expression and
931 enrichment for it targets or ER stress genes does not necessarily imply a strong correlation
932 between signed intrinsic divergence in *Xbp1* expression and the corresponding enrichments.
933 For example, if *cis*-acting mutations altered the expression of many *Xbp1* target genes, this
934 might lead to weak to no correlation for signed intrinsic divergence even if there were a strong
935 correlation for signed extrinsic divergence.
936
937 **Quantification and analysis of immunofluorescence (IF) images**
938
939 All quantification of fluorescence intensity was performed in Python and all IF images shown in
940 this work were made in ImageJ(*64*). We paired chimeras that were generated, fixed, processed,
941 stained, and imaged in the same batch with the same reagents. For all regions of interest,
942 selection, tuning, and parameter selection was performed using only the DAPI and TdTomato
943 channels before measuring fluorescence intensity for c-Jun or Hspa5, ensuring that we were
944 effectively blind to the outcome of the analysis and that the results were not biased.
945
946 To quantify fluorescence intensity, we read in the images and drew a rectangular region of
947 interest containing only the medial and lateral ganglionic eminences for chimera pair 1, and only
948 the medial ganglionic eminence for chimera 3 (as the lateral ganglionic eminence had no mouse
949 cells in the rat-like chimera). We then used a Gaussian filter to blur the DAPI and TdTomato
950 channels and masked the images into DAPI positive, DAPI negative, TdTomato positive, and
951 TdTomato negative. We tuned the Gaussian blue sigma parameter and cutoff for the masking
952 by hand until the DAPI mask included all of the area in the region of interest except clearly
953 visible larger regions without DAPI+ nuclei in the region of interest. For the TdTomato signal, we

954 ensured that the mask closely matched the original TdTomato signal in the region of interest.
955 The sigma and masking cutoff values selected in this process were used for all results
956 presented in the text.
957
958 We then considered all DAPI+ TdTomato+ pixels to be from donor cells (species-mismatched
959 environment) and all DAPI+ TdT- pixels to be from host cells (species-matched environment),
960 restricted to pixels with non-zero c-Jun or Hspa5 intensity values, and computed the average
961 fluorescent intensity in each set of pixels. For within chimera comparisons, we computed the
962 $\log_2$ fold-change between the average donor signal and the average host signal as was done for
963 the RNA-seq measurements. We also subtracted the average background signal from a
964 rectangular region with no nuclei from the values computed above before computing $\log_2$ fold-
965 changes between chimeras. As the sigma parameter and cutoff for the masking were chosen by
966 hand, we tested all possible combinations of three different masking cutoffs (the originally
967 selected cutoff and two other similar values) for DAPI and TdTomato for the rat-like and mouse-
968 like chimeras separately (81 combinations in total) and obtained highly similar log fold-changes
969 (the largest standard error for any log fold-change was 0.0046). These $\log_2$ fold-changes were
970 then used to compute the extrinsic, intrinsic, and interaction proportions for protein expression
971 divergence as described above for the gene expression divergence measured through scRNA-
972 seq.
973
974 To produce the images in the publication the images were first read into ImageJ. Next, channels
975 were assigned colors and the brightness and contrast tuned to visualize the fluorescent signal.
976 Scale bars were then added by synchronizing the windows, drawing horizontal line, and adding
977 a scale bar to each image with the Scale Bar tool. These images were then exported as .png
978 objects.
979

980 **Analysis of neuron-progenitor ratios**

981

982 For the forebrain glutamatergic cells, forebrain GABAergic cells, and the combined spinal cord
983 glutamatergic and GABAergic cells (combined due to the low number of progenitors we detect),
984 we computed the ratio of post-mitotic cells to progenitors (excluding intermediate progenitors) in
985 all four species-environment combinations. We then used these values to compute $\log_2$ fold-
986 changes and the intrinsic, extrinsic, and interaction proportions in the same way we did for the
987 normalized counts from the scRNA-seq data.

988

989 **Analysis of imprinted gene expression**

990

991 The list of imprinted genes for mice and humans was downloaded from
992 https://www.geneimprint.com/site/home through the "gene by species" tab. To test whether
993 imprinted genes were enriched for high absolute interaction divergence, we created an
994 imprinted gene set consisting of all "confident" and "predicted" imprinted genes in mice from the
995 above resource. Only 2 genes were predicted and not confident, so our results are unaltered by
996 restricting only to confident imprinted genes. We next performed enrichment analysis as
997 described above, including averaging across cell types within nervous and connective tissue,
998 with the following modification. Initially, we simply ranked genes by the absolute interaction
999 divergence and observed very strong enrichment. However, due to the GSEA running sum
1000 statistic being initially developed for signed metrics (e.g. log fold-change) we reasoned that this
1001 procedure could lead to spuriously strong enrichments(*25*). To address this, we rank
1002 transformed the list such that the genes with very large absolute interaction divergence have
1003 large positive values and genes with absolute interaction divergence near zero had negative
1004 values with equally large magnitude. For example, if there were 5,000 input genes, the gene
1005 with the highest absolute interaction divergence would have a value of 2,500, the next highest

1006 2,499, etc. The gene with the lowest absolute interaction divergence would have a value of -
1007 2,500, the next lowest -2,499, etc. Using this strategy, we still observe strong enrichment (we
1008 report the p-values obtained using this strategy in the text and Fig. 6D). For the signed
1009 enrichment analysis, we ranked genes by the signed interaction divergence and used GSEAPY
1010 preranked with the imprinted gene set.
1011
1012 To analyze the data from chimeric rodent hearts, we downloaded the counts matrices and
1013 associated metadata (including the species of origin for each cell) from GEO:
1014 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE236400(*37*). The cell type and other
1015 metadata were shared by the original authors of the study.
1016
1017 We then pseudobulked counts by summing across all cells within a cell type for chimera 1,
1018 chimera 2, and the wildtype rat sample separately, removed genes with fewer than twenty
1019 counts in all samples, and computed counts per million (CPM). We then took the mean of the
1020 two chimera replicates and computed the $\log_2$ fold-change of rat-like environment CPM divided
1021 by mouse-like environment CPM. This value was then used in place of interaction divergence
1022 for the enrichment analyses described above.
1023
1024 To analyze the data from adult parathyroid glands, we downloaded files containing TPM
1025 estimates for each mouse transcript from GEO:
1026 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE232600. We then collapsed TPM
1027 estimates to the gene level by summing the TPM for each transcript assigned to individual
1028 genes. As each sample was generated from a very small number of cells (approximately 20),
1029 expression estimates for many genes were noisy so we used the median TPM across replicates
1030 per condition as our estimate of the expression level of that gene. We removed genes with
1031 median TPM less than 5 in all conditions (species-mismatched environment, species-matched

1032 environment, and wildtype mice). We restricted to imprinted genes as described above,
1033 normalized the expression of each gene, and applied hierarchical clustering with the default
1034 parameters in seaborn clustermap. The log fold-change was computed by comparing the
1035 median TPMs. All imprinted genes with absolute $\log_2$ fold-change between expression for
1036 species-mismatched and species-matched environment samples greater than 1 were plotted in
1037 Fig. S29A.

1038

1039 To analyze the data from human and macaque epiblast, we downloaded the transcripts per
1040 million (TPM) matrices for human-macaque chimeric embryos and wildtype human embryos
1041 from GEO: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE155381 and
1042 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE109555 respectively. We restricted to
1043 cells annotated as epiblast for the human-macaque chimeric data and used Ensembl to identify
1044 one-to-one orthologs between crab-eating macaques and humans. For the wildtype human
1045 embryos, the raw count matrices and cell type annotations were unavailable. Therefore, we
1046 restricted to cells with greater than one TPM for *NANOG*, *SOX2*, or *POU5F1* (also known as
1047 *OCT4*) as putative epiblast cells. We then averaged TPM across all cells in each species-
1048 environment category, filtered genes that were expressed below 1 TPM in all categories, and
1049 computed the $\log_2$ fold-change of human in a macaque-like environment divided by macaque in
1050 a macaque-like environment and human in a macaque-like environment divided by human in a
1051 human-like environment after adding one pseudo-TPM to all values. We then performed the
1052 enrichment analysis using absolute $\log_2$ fold-change as described for mouse and rat heart data.
1053 Although the comparison of human data across studies undoubtedly introduces batch effects, it
1054 is very unlikely that those batch effects would specifically increase the absolute $\log_2$ fold-change
1055 of imprinted genes as a whole.

1056

## References and Notes


1. H. Spemann, H. Mangold, Induction of embryonic primordia by implantation of organizers from a different species. 1923. *Int. J. Dev. Biol.* **45**, 13–38 (2001).

2. N. M. Le Douarin, The ontogeny of the neural crest in avian embryo chimaeras. *Nature* **286**, 663–669 (1980).

3. Twitty, Victor C., CORRELATED GENETIC AND EMBRYOLOGICAL EXPERIMENTS ON TRITURUS. I and II. *The Journal of Experimental Zoology* **74** (1936).

4. R. A. Schneider, Neural crest and the origin of species-specific pattern. *genesis* **56**, e23219 (2018).

5. Le Douarin N, McLaren A., *Chimeras in Developmental Biology* (Academic Press, 1984).

6. F. Suchy, H. Nakauchi, Interspecies chimeras. *Curr. Opin. Genet. Dev.* **52**, 36–41 (2018).

7. B. K. Stepien, R. Naumann, A. Holtz, J. Helppi, W. B. Huttner, S. Vaid, Lengthening Neurogenic Period during Neocortical Development Causes a Hallmark of Neocortex Expansion. *Curr. Biol.* **30**, 4227-4237.e5 (2020).

8. B. F. Eames, R. A. Schneider, Quail-duck chimeras reveal spatiotemporal plasticity in molecular and histogenic programs of cranial feather development. *Dev. Camb. Engl.* **132**, 1499–1509 (2005).

9. R. A. Schneider, J. A. Helms, The Cellular and Molecular Origins of Beak Morphology. *Science* **299**, 565–568 (2003).

10. P. J. Wittkopp, G. Kalay, Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nat. Rev. Genet.* **13**, 59–69 (2012).

11. P. J. Wittkopp, B. K. Haerum, A. G. Clark, Evolutionary changes in cis and trans gene regulation. *Nature* **430**, 85–88 (2004).

12. J. D. Coolon, C. J. McManus, K. R. Stevenson, B. R. Graveley, P. J. Wittkopp, Tempo and mode of regulatory evolution in *Drosophila*. *Genome Res.* **24**, 797–808 (2014).

13. X. Liu, Y. I. Li, J. K. Pritchard, Trans Effects on Gene Expression Can Drive Omnigenic Inheritance. *Cell* **177**, 1022-1034.e6 (2019).

14. F. W. Albert, J. S. Bloom, J. Siegel, L. Day, L. Kruglyak, Genetics of trans-regulatory variation in gene expression. *eLife* **7**, e35471 (2018).

15. Q. Wang, Y. Jia, Y. Wang, Z. Jiang, X. Zhou, Z. Zhang, C. Nie, J. Li, N. Yang, L. Qu, Evolution of cis- and trans-regulatory divergence in the chicken genome between two contrasting breeds analyzed using three tissue types at one-day-old. *BMC Genomics* **20**, 933 (2019).

16. J. Huang, B. He, X. Yang, X. Long, Y. Wei, Y. Gao, Y. Fang, W. Ying, Z. Wang, C. Li, Y. Zhou, S. Li, L. Shi, F. Guo, H. Zhou, H. Yang, J. Wu, “Interspecies blastocyst complementation generates

functional rat cell-derived forebrain tissues in mice" (preprint, Developmental Biology, 2023); https://doi.org/10.1101/2023.04.13.536774.

17. S. Kim-Hellmuth, F. Aguet, M. Oliva, M. Muñoz-Aguirre, S. Kasela, V. Wucher, S. E. Castel, A. R. Hamel, A. Viñuela, A. L. Roberts, S. Mangul, X. Wen, G. Wang, A. N. Barbeira, D. Garrido-Martín, B. B. Nadel, Y. Zou, R. Bonazzola, J. Quan, A. Brown, A. Martinez-Perez, J. M. Soria, GTEx Consortium, G. Getz, E. T. Dermitzakis, K. S. Small, M. Stephens, H. S. Xi, H. K. Im, R. Guigó, A. V. Segrè, B. E. Stranger, K. G. Ardlie, T. Lappalainen, F. Aguet, S. Anand, K. G. Ardlie, S. Gabriel, G. A. Getz, A. Graubert, K. Hadley, R. E. Handsaker, K. H. Huang, S. Kashin, X. Li, D. G. MacArthur, S. R. Meier, J. L. Nedzel, D. T. Nguyen, A. V. Segrè, E. Todres, B. Balliu, A. N. Barbeira, A. Battle, R. Bonazzola, A. Brown, C. D. Brown, S. E. Castel, D. F. Conrad, D. J. Cotter, N. Cox, S. Das, O. M. De Goede, E. T. Dermitzakis, J. Einson, B. E. Engelhardt, E. Eskin, T. Y. Eulalio, N. M. Ferraro, E. D. Flynn, L. Fresard, E. R. Gamazon, D. Garrido-Martín, N. R. Gay, M. J. Gloudemans, R. Guigó, A. R. Hame, Y. He, P. J. Hoffman, F. Hormozdiari, L. Hou, H. K. Im, B. Jo, S. Kasela, M. Kellis, S. Kim-Hellmuth, A. Kwong, T. Lappalainen, X. Li, Y. Liang, S. Mangul, P. Mohammadi, S. B. Montgomery, M. Muñoz-Aguirre, D. C. Nachun, A. B. Nobel, M. Oliva, Y. Park, Y. Park, P. Parsana, A. S. Rao, F. Reverter, J. M. Rouhana, C. Sabatti, A. Saha, M. Stephens, B. E. Stranger, B. J. Strober, N. A. Teran, A. Viñuela, G. Wang, X. Wen, F. Wright, V. Wucher, Y. Zou, P. G. Ferreira, G. Li, M. Melé, E. Yeger-Lotem, M. E. Barcus, D. Bradbury, T. Krubit, J. A. McLean, L. Qi, K. Robinson, N. V. Roche, A. M. Smith, L. Sobin, D. E. Tabor, A. Undale, J. Bridge, L. E. Brigham, B. A. Foster, B. M. Gillard, R. Hasz, M. Hunter, C. Johns, M. Johnson, E. Karasik, G. Kopen, W. F. Leinweber, A. McDonald, M. T. Moser, K. Myer, K. D. Ramsey, B. Roe, S. Shad, J. A. Thomas, G. Walters, M. Washington, J. Wheeler, S. D. Jewell, D. C. Rohrer, D. R. Valley, D. A. Davis, D. C. Mash, P. A. Branton, L. K. Barker, H. M. Gardiner, M. Mosavel, L. A. Siminoff, P. Flicek, M. Haeussler, T. Juettemann, W. J. Kent, C. M. Lee, C. C. Powell, K. R. Rosenbloom, M. Ruffier, D. Sheppard, K. Taylor, S. J. Trevanion, D. R. Zerbino, N. S. Abell, J. Akey, L. Chen, K. Demanelis, J. A. Doherty, A. P. Feinberg, K. D. Hansen, P. F. Hickey, F. Jasmine, L. Jiang, R. Kaul, M. G. Kibriya, J. B. Li, Q. Li, S. Lin, S. E. Linder, B. L. Pierce, L. F. Rizzardi, A. D. Skol, K. S. Smith, M. Snyder, J. Stamatoyannopoulos, H. Tang, M. Wang, L. J. Carithers, P. Guan, S. E. Koester, A. R. Little, H. M. Moore, C. R. Nierras, A. K. Rao, J. B. Vaught, S. Volpi, Cell type–specific genetic regulation of gene expression across human tissues. *Science* **369**, eaaz8528 (2020).

18. The GTEx Consortium, F. Aguet, S. Anand, K. G. Ardlie, S. Gabriel, G. A. Getz, A. Graubert, K. Hadley, R. E. Handsaker, K. H. Huang, S. Kashin, X. Li, D. G. MacArthur, S. R. Meier, J. L. Nedzel, D. T. Nguyen, A. V. Segrè, E. Todres, B. Balliu, A. N. Barbeira, A. Battle, R. Bonazzola, A. Brown, C. D. Brown, S. E. Castel, D. F. Conrad, D. J. Cotter, N. Cox, S. Das, O. M. De Goede, E. T. Dermitzakis, J. Einson, B. E. Engelhardt, E. Eskin, T. Y. Eulalio, N. M. Ferraro, E. D. Flynn, L. Fresard, E. R. Gamazon, D. Garrido-Martín, N. R. Gay, M. J. Gloudemans, R. Guigó, A. R. Hame, Y. He, P. J. Hoffman, F. Hormozdiari, L. Hou, H. K. Im, B. Jo, S. Kasela, M. Kellis, S. Kim-Hellmuth, A. Kwong, T. Lappalainen, X. Li, Y. Liang, S. Mangul, P. Mohammadi, S. B. Montgomery, M. Muñoz-Aguirre, D. C. Nachun, A. B. Nobel, M. Oliva, Y. Park, Y. Park, P. Parsana, A. S. Rao, F. Reverter, J. M. Rouhana, C. Sabatti, A. Saha, M. Stephens, B. E. Stranger, B. J. Strober, N. A. Teran, A. Viñuela, G. Wang, X. Wen, F. Wright, V. Wucher, Y. Zou, P. G. Ferreira, G. Li, M. Melé, E. Yeger-Lotem, M. E. Barcus, D. Bradbury, T. Krubit, J. A. McLean, L. Qi, K. Robinson, N. V. Roche, A. M. Smith, L. Sobin, D. E. Tabor, A. Undale, J. Bridge, L. E. Brigham, B. A. Foster, B. M. Gillard, R. Hasz, M. Hunter, C. Johns, M. Johnson, E. Karasik, G. Kopen, W. F. Leinweber, A. McDonald, M. T. Moser, K. Myer, K. D. Ramsey, B. Roe, S. Shad, J. A. Thomas, G.

Walters, M. Washington, J. Wheeler, S. D. Jewell, D. C. Rohrer, D. R. Valley, D. A. Davis, D. C. Mash, P. A. Branton, L. K. Barker, H. M. Gardiner, M. Mosavel, L. A. Siminoff, P. Flicek, M. Haeussler, T. Juettemann, W. J. Kent, C. M. Lee, C. C. Powell, K. R. Rosenbloom, M. Ruffier, D. Sheppard, K. Taylor, S. J. Trevanion, D. R. Zerbino, N. S. Abell, J. Akey, L. Chen, K. Demanelis, J. A. Doherty, A. P. Feinberg, K. D. Hansen, P. F. Hickey, F. Jasmine, L. Jiang, R. Kaul, M. G. Kibriya, J. B. Li, Q. Li, S. Lin, S. E. Linder, B. L. Pierce, L. F. Rizzardi, A. D. Skol, K. S. Smith, M. Snyder, J. Stamatoyannopoulos, H. Tang, M. Wang, L. J. Carithers, P. Guan, S. E. Koester, A. R. Little, H. M. Moore, C. R. Nierras, A. K. Rao, J. B. Vaught, S. Volpi, The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).

19. GTEx Consortium, Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).

20. B. P. H. Metzger, F. Duveau, D. C. Yuan, S. Tryban, B. Yang, P. J. Wittkopp, Contrasting Frequencies and Effects of *cis* - and *trans* -Regulatory Mutations Affecting Gene Expression. *Mol. Biol. Evol.* **33**, 1131–1146 (2016).

21. A. Hoter, M. El-Sabban, H. Naim, The HSP90 Family: Structure, Regulation, Function, and Implications in Health and Disease. *Int. J. Mol. Sci.* **19**, 2560 (2018).

22. Y. Zhu, D. Sun, M. Jakovcevski, Y. Jiang, Epigenetic mechanism of SETDB1 in brain: implications for neuropsychiatric disorders. *Transl. Psychiatry* **10**, 115 (2020).

23. S.-M. Park, T.-I. Kang, J.-S. So, Roles of XBP1s in Transcriptional Regulation of Target Genes. *Biomedicines* **9**, 791 (2021).

24. M. F. Gregor, E. S. Misch, L. Yang, S. Hummasti, K. E. Inouye, A.-H. Lee, B. Bierie, G. S. Hotamisligil, The role of adipocyte XBP1 in metabolic regulation during lactation. *Cell Rep.* **3**, 1430–1439 (2013).

25. A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, J. P. Mesirov, Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci.* **102**, 15545–15550 (2005).

26. H.-C. Ku, C.-F. Cheng, Master Regulator Activating Transcription Factor 3 (ATF3) in Metabolic Homeostasis and Cancer. *Front. Endocrinol.* **11**, 556 (2020).

27. Y. Adachi, K. Yamamoto, T. Okada, H. Yoshida, A. Harada, K. Mori, ATF6 is a transcription factor specializing in the regulation of quality control proteins in the endoplasmic reticulum. *Cell Struct. Funct.* **33**, 75–89 (2008).

28. W. Yu, B. Wang, L. Zhou, G. Xu, Endoplasmic Reticulum Stress-Mediated p62 Downregulation Inhibits Apoptosis via c-Jun Upregulation. *Biomol. Ther.* **29**, 195–204 (2021).

29. Z. Sha, A. L. Goldberg, Proteasome-Mediated Processing of Nrf1 Is Essential for Coordinate Induction of All Proteasome Subunits and p97. *Curr. Biol.* **24**, 1573–1583 (2014).

30. J. R. Vangala, S. Dudem, N. Jain, S. V. Kalivendi, Regulation of PSMB5 Protein and β Subunits of Mammalian Proteasome by Constitutively Activated Signal Transducer and Activator of Transcription 3 (STAT3). *J. Biol. Chem.* **289**, 12612–12622 (2014).

31. F. He, X. Ru, T. Wen, NRF2, a Transcription Factor for Stress Response and Beyond. *Int. J. Mol. Sci.* **21**, 4777 (2020).

32. J. Wang, J. Lee, D. Liem, P. Ping, HSPA5 Gene encoding Hsp70 chaperone BiP in the endoplasmic reticulum. *Gene* **618**, 14–23 (2017).

33. A. M. Juan, Y. H. Foong, J. L. Thorvaldsen, Y. Lan, N. A. Leu, J. G. Rurik, L. Li, C. Krapp, C. L. Rosier, J. A. Epstein, M. S. Bartolomei, Tissue-specific Grb10/Ddc insulator drives allelic architecture for cardiac development. *Mol. Cell* **82**, 3613-3631.e7 (2022).

34. N. Giannoukakis, C. Deal, J. Paquette, C. G. Goodyer, C. Polychronakos, Parental genomic imprinting of the human IGF2 gene. *Nat. Genet.* **4**, 98–101 (1993).

35. G. E. Moore, M. Ishida, C. Demetriou, L. Al-Olabi, L. J. Leon, A. C. Thomas, S. Abu-Amero, J. M. Frost, J. L. Stafford, Y. Chaoqun, A. J. Duncan, R. Baigel, M. Brimioulle, I. Iglesias-Platas, S. Apostolidou, R. Aggarwal, J. C. Whittaker, A. Syngelaki, K. H. Nicolaides, L. Regan, D. Monk, P. Stanier, The role and interaction of imprinted genes in human fetal growth. *Philos. Trans. R. Soc. B Biol. Sci.* **370**, 20140074 (2015).

36. J. Glaser, J. Iranzo, M. Borensztein, M. Marinucci, A. Gualtieri, C. Jouhanneau, A. Teissandier, C. Gaston-Massuet, D. Bourc'his, The imprinted Zdbf2 gene finely tunes control of feeding and growth in neonates. *eLife* **11**, e65641 (2022).

37. G. Coppiello, P. Barlabé, M. Moya-Jódar, G. Abizanda, C. Pogontke, C. Barreda, E. Iglesias, J. Linares, E. Arellano-Viera, E. Larequi, P. San Martín-Úriz, X. Carvajal-Vergara, B. Pelacho, M. M. Mazo, J. M. Pérez-Pomares, A. Ruiz-Villalba, A. Ullate-Agote, F. Prósper, X. L. Aranguren, Generation of heart and vascular system in rodents by blastocyst complementation. *Dev. Cell* **58**, 2881-2895.e7 (2023).

38. M. Kano, N. Mizuno, H. Sato, T. Kimura, R. Hirochika, Y. Iwasaki, N. Inoshita, H. Nagano, M. Kasai, H. Yamamoto, T. Yamaguchi, H. Suga, H. Masaki, E. Mizutani, H. Nakauchi, Functional calcium-responsive parathyroid glands generated using single-step blastocyst complementation. *Proc. Natl. Acad. Sci.* **120**, e2216564120 (2023).

39. F. Zhou, R. Wang, P. Yuan, Y. Ren, Y. Mao, R. Li, Y. Lian, J. Li, L. Wen, L. Yan, J. Qiao, F. Tang, Reconstituting the transcriptome and DNA methylome landscapes of human implantation. *Nature* **572**, 660–664 (2019).

40. T. Tan, J. Wu, C. Si, S. Dai, Y. Zhang, N. Sun, E. Zhang, H. Shao, W. Si, P. Yang, H. Wang, Z. Chen, R. Zhu, Y. Kang, R. Hernandez-Benitez, L. Martinez Martinez, E. Nuñez Delicado, W. T. Berggren, M. Schwarz, Z. Ai, T. Li, H. Deng, C. Rodriguez Esteban, W. Ji, Y. Niu, J. C. Izpisua Belmonte, Chimeric contribution of human extended pluripotent stem cells to monkey embryos ex vivo. *Cell* **184**, 2020-2032.e14 (2021).

41. H. Masaki, H. Nakauchi, Interspecies chimeras for human stem cell research. *Development* **144**, 2544–2547 (2017).

42. B. Ramani, I. V. L. Rose, A. Pan, R. Tian, K. Ma, J. J. Palop, M. Kampmann, "Scalable, cell type-selective, AAV-based *in vivo* CRISPR screening in the mouse brain" (preprint, Genomics, 2023); https://doi.org/10.1101/2023.06.13.544831.

43. G. Ciceri, A. Baggiolini, H. S. Cho, M. Kshirsagar, S. Benito-Kwiecinski, R. M. Walsh, K. A. Aromolaran, A. J. Gonzalez-Hernandez, H. Munguba, S. Y. Koo, N. Xu, K. J. Sevilla, P. A. Goldstein, J. Levitz, C. S. Leslie, R. P. Koche, L. Studer, An epigenetic barrier sets the timing of human neuronal maturation. *Nature* **626**, 881–890 (2024).

44. J. A. Piedrahita, The role of imprinted genes in fetal growth abnormalities. *Birt. Defects Res. A. Clin. Mol. Teratol.* **91**, 682–692 (2011).

45. Q.-L. Ying, M. Stavridis, D. Griffiths, M. Li, A. Smith, Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. *Nat. Biotechnol.* **21**, 183–186 (2003).

46. T. Yamaguchi, S. Hamanaka, H. Nakauchi, The generation and maintenance of rat induced pluripotent stem cells. *Methods Mol. Biol. Clifton NJ* **1210**, 143–150 (2014).

47. D. G. Brownstein, *Manipulating the Mouse Embryo: A Laboratory Manual . Third Edition. By* Andras Nagy, Marina Gertsenstein, Kristina Vintersten, and Richard Behringer. *Cold Spring Harbor (New York): Cold Spring Harbor Laboratory Press* . $195.00 (hardcover); $115.00 (paper). x + 764 p; ill.; index. ISBN: 0-87969-574-9 (hc); 0-87969-591-9 (pb). 2003. *Q. Rev. Biol.* **78**, 365–365 (2003).

48. A. D. Yates, J. Allen, R. M. Amode, A. G. Azov, M. Barba, A. Becerra, J. Bhai, L. I. Campbell, M. Carbajo Martinez, M. Chakiachvili, K. Chougule, M. Christensen, B. Contreras-Moreira, A. Cuzick, L. Da Rin Fioretto, P. Davis, N. H. De Silva, S. Diamantakis, S. Dyer, J. Elser, C. V. Filippi, A. Gall, D. Grigoriadis, C. Guijarro-Clarke, P. Gupta, K. E. Hammond-Kosack, K. L. Howe, P. Jaiswal, V. Kaikala, V. Kumar, S. Kumari, N. Langridge, T. Le, M. Luypaert, G. L. Maslen, T. Maurel, B. Moore, M. Muffato, A. Mushtaq, G. Naamati, S. Naithani, A. Olson, A. Parker, M. Paulini, H. Pedro, E. Perry, J. Preece, M. Quinton-Tulloch, F. Rodgers, M. Rosello, M. Ruffier, J. Seager, V. Sitnik, M. Szpak, J. Tate, M. K. Tello-Ruiz, S. J. Trevanion, M. Urban, D. Ware, S. Wei, G. Williams, A. Winterbottom, M. Zarowiecki, R. D. Finn, P. Flicek, Ensembl Genomes 2022: an expanding genome resource for non-vertebrates. *Nucleic Acids Res.* **50**, D996–D1003 (2022).

49. F. A. Wolf, P. Angerer, F. J. Theis, SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).

50. I. Korsunsky, N. Millard, J. Fan, K. Slowikowski, F. Zhang, K. Wei, Y. Baglaenko, M. Brenner, P. Loh, S. Raychaudhuri, Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).

51. V. A. Traag, L. Waltman, N. J. Van Eck, From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* **9**, 5233 (2019).

52. D. J. Di Bella, E. Habibi, R. R. Stickels, G. Scalia, J. Brown, P. Yadollahpour, S. M. Yang, C. Abbate, T. Biancalani, E. Z. Macosko, F. Chen, A. Regev, P. Arlotta, Molecular logic of cellular diversification in the mouse cerebral cortex. *Nature* **595**, 554–559 (2021).

53. M. A. Lipiec, J. Bem, K. Koziński, C. Chakraborty, J. Urban-Ciećko, T. Zajkowski, M. Dąbrowski, Ł. M. Szewczyk, A. Toval, J. L. Ferran, A. Nagalski, M. B. Wiśniewska, TCF7L2 regulates postmitotic differentiation programs and excitability patterns in the thalamus. *Development*, dev.190181 (2020).

54. R. Witschi, T. Johansson, G. Morscher, L. Scheurer, J. Deschamps, H. U. Zeilhofer, *Hoxb8-Cre* mice: A tool for brain-sparing conditional gene deletion. *genesis* **48**, 596–602 (2010).

55. J. Lee, C. C. Rabbani, H. Gao, M. R. Steinhart, B. M. Woodruff, Z. E. Pflum, A. Kim, S. Heller, Y. Liu, T. Z. Shipchandler, K. R. Koehler, Hair-bearing human skin generated entirely from pluripotent stem cells. *Nature* **582**, 399–404 (2020).

56. M. Cardoso-Moreira, J. Halbert, D. Valloton, B. Velten, C. Chen, Y. Shao, A. Liechti, K. Ascenção, C. Rummel, S. Ovchinnikova, P. V. Mazin, I. Xenarios, K. Harshman, M. Mort, D. N. Cooper, C. Sandi, M. J. Soares, P. G. Ferreira, S. Afonso, M. Carneiro, J. M. A. Turner, J. L. VandeBerg, A. Fallahshahroudi, P. Jensen, R. Behr, S. Lisgo, S. Lindsay, P. Khaitovich, W. Huber, J. Baker, S. Anders, Y. E. Zhang, H. Kaessmann, Gene expression across mammalian organ development. *Nature* **571**, 505–509 (2019).

57. P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. Van Der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. Van Mulbregt, SciPy 1.0 Contributors, A. Vijaykumar, A. P. Bardelli, A. Rothberg, A. Hilboll, A. Kloeckner, A. Scopatz, A. Lee, A. Rokem, C. N. Woods, C. Fulton, C. Masson, C. Häggström, C. Fitzgerald, D. A. Nicholson, D. R. Hagen, D. V. Pasechnik, E. Olivetti, E. Martin, E. Wieser, F. Silva, F. Lenders, F. Wilhelm, G. Young, G. A. Price, G.-L. Ingold, G. E. Allen, G. R. Lee, H. Audren, I. Probst, J. P. Dietrich, J. Silterra, J. T. Webber, J. Slavič, J. Nothman, J. Buchner, J. Kulick, J. L. Schönberger, J. V. De Miranda Cardoso, J. Reimer, J. Harrington, J. L. C. Rodríguez, J. Nunez-Iglesias, J. Kuczynski, K. Tritz, M. Thoma, M. Newville, M. Kümmerer, M. Bolingbroke, M. Tartre, M. Pak, N. J. Smith, N. Nowaczyk, N. Shebanov, O. Pavlyk, P. A. Brodtkorb, P. Lee, R. T. McGibbon, R. Feldbauer, S. Lewis, S. Tygier, S. Sievert, S. Vigna, S. Peterson, S. More, T. Pudlik, T. Oshima, T. J. Pingel, T. P. Robitaille, T. Spura, T. R. Jones, T. Cera, T. Leslie, T. Zito, T. Krauss, U. Upadhyay, Y. O. Halchenko, Y. Vázquez-Baeza, SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).

58. T. Zeng, J. P. Spence, H. Mostafavi, J. K. Pritchard, "Bayesian estimation of gene constraint from an evolutionary model with gene features" (preprint, Genetics, 2023); https://doi.org/10.1101/2023.05.19.541520.

59. The ENCODE Project Consortium, F. Abascal, R. Acosta, N. J. Addleman, J. Adrian, V. Afzal, R. Ai, B. Aken, J. A. Akiyama, O. A. Jammal, H. Amrhein, S. M. Anderson, G. R. Andrews, I. Antoshechkin, K. G. Ardlie, J. Armstrong, M. Astley, B. Banerjee, A. A. Barkal, I. H. A. Barnes, I.

Barozzi, D. Barrell, G. Barson, D. Bates, U. K. Baymuradov, C. Bazile, M. A. Beer, S. Beik, M. A. Bender, R. Bennett, L. P. B. Bouvrette, B. E. Bernstein, A. Berry, A. Bhaskar, A. Bignell, S. M. Blue, D. M. Bodine, C. Boix, N. Boley, T. Borrman, B. Borsari, A. P. Boyle, L. A. Brandsmeier, A. Breschi, E. H. Bresnick, J. A. Brooks, M. Buckley, C. B. Burge, R. Byron, E. Cahill, L. Cai, L. Cao, M. Carty, R. G. Castanon, A. Castillo, H. Chaib, E. T. Chan, D. R. Chee, S. Chee, H. Chen, H. Chen, J.-Y. Chen, S. Chen, J. M. Cherry, S. B. Chhetri, J. S. Choudhary, J. Chrast, D. Chung, D. Clarke, N. A. L. Cody, C. J. Coppola, J. Coursen, A. M. D'Ippolito, S. Dalton, C. Danyko, C. Davidson, J. Davila-Velderrain, C. A. Davis, J. Dekker, A. Deran, G. DeSalvo, G. Despacio-Reyes, C. N. Dewey, D. E. Dickel, M. Diegel, M. Diekhans, V. Dileep, B. Ding, S. Djebali, A. Dobin, D. Dominguez, S. Donaldson, J. Drenkow, T. R. Dreszer, Y. Drier, M. O. Duff, D. Dunn, C. Eastman, J. R. Ecker, M. D. Edwards, N. El-Ali, S. I. Elhajjajy, K. Elkins, A. Emili, C. B. Epstein, R. C. Evans, I. Ezkurdia, K. Fan, P. J. Farnham, N. P. Farrell, E. A. Feingold, A.-M. Ferreira, K. Fisher-Aylor, S. Fitzgerald, P. Flicek, C. S. Foo, K. Fortier, A. Frankish, P. Freese, S. Fu, X.-D. Fu, Y. Fu, Y. Fukuda-Yuzawa, M. Fulciniti, A. P. W. Funnell, I. Gabdank, T. Galeev, M. Gao, C. G. Giron, T. H. Garvin, C. A. Gelboin-Burkhart, G. Georgolopoulos, M. B. Gerstein, B. M. Giardine, D. K. Gifford, D. M. Gilbert, D. A. Gilchrist, S. Gillespie, T. R. Gingeras, P. Gong, A. Gonzalez, J. M. Gonzalez, P. Good, A. Goren, D. U. Gorkin, B. R. Graveley, M. Gray, J. F. Greenblatt, E. Griffiths, M. T. Groudine, F. Grubert, M. Gu, R. Guigó, H. Guo, Y. Guo, Y. Guo, G. Gursoy, M. Gutierrez-Arcelus, J. Halow, R. C. Hardison, M. Hardy, M. Hariharan, A. Harmanci, A. Harrington, J. L. Harrow, T. B. Hashimoto, R. D. Hasz, M. Hatan, E. Haugen, J. E. Hayes, P. He, Y. He, N. Heidari, D. Hendrickson, E. F. Heuston, J. A. Hilton, B. C. Hitz, A. Hochman, C. Holgren, L. Hou, S. Hou, Y.-H. E. Hsiao, S. Hsu, H. Huang, T. J. Hubbard, J. Huey, T. R. Hughes, T. Hunt, S. Ibarrientos, R. Issner, M. Iwata, O. Izuogu, T. Jaakkola, N. Jameel, C. Jansen, L. Jiang, P. Jiang, A. Johnson, R. Johnson, I. Jungreis, M. Kadaba, M. Kasowski, M. Kasparian, M. Kato, R. Kaul, T. Kawli, M. Kay, J. C. Keen, S. Keles, C. A. Keller, D. Kelley, M. Kellis, P. Kheradpour, D. S. Kim, A. Kirilusha, R. J. Klein, B. Knoechel, S. Kuan, M. J. Kulik, S. Kumar, A. Kundaje, T. Kutyavin, J. Lagarde, B. R. Lajoie, N. J. Lambert, J. Lazar, A. Y. Lee, D. Lee, E. Lee, J. W. Lee, K. Lee, C. S. Leslie, S. Levy, B. Li, H. Li, N. Li, S. Li, X. Li, Y. I. Li, Y. Li, Y. Li, Y. Li, J. Lian, M. W. Libbrecht, S. Lin, Y. Lin, D. Liu, J. Liu, P. Liu, T. Liu, X. S. Liu, Y. Liu, Y. Liu, M. Long, S. Lou, J. Loveland, A. Lu, Y. Lu, E. Lécuyer, L. Ma, M. Mackiewicz, B. J. Mannion, M. Mannstadt, D. Manthravadi, G. K. Marinov, F. J. Martin, E. Mattei, K. McCue, M. McEown, G. McVicker, S. K. Meadows, A. Meissner, E. M. Mendenhall, C. L. Messer, W. Meuleman, C. Meyer, S. Miller, M. G. Milton, T. Mishra, D. E. Moore, H. M. Moore, J. E. Moore, S. H. Moore, J. Moran, A. Mortazavi, J. M. Mudge, N. Munshi, R. Murad, R. M. Myers, V. Nandakumar, P. Nandi, A. M. Narasimha, A. K. Narayanan, H. Naughton, F. C. P. Navarro, P. Navas, J. Nazarovs, J. Nelson, S. Neph, F. J. Neri, J. R. Nery, A. R. Nesmith, J. S. Newberry, K. M. Newberry, V. Ngo, R. Nguyen, T. B. Nguyen, T. Nguyen, A. Nishida, W. S. Noble, C. S. Novak, E. M. Novoa, B. Nuñez, C. W. O'Donnell, S. Olson, K. C. Onate, E. Otterman, H. Ozadam, M. Pagan, T. Palden, X. Pan, Y. Park, E. C. Partridge, B. Paten, F. Pauli-Behn, M. J. Pazin, B. Pei, L. A. Pennacchio, A. R. Perez, E. H. Perry, D. D. Pervouchine, N. N. Phalke, Q. Pham, D. H. Phanstiel, I. Plajzer-Frick, G. A. Pratt, H. E. Pratt, S. Preissl, J. K. Pritchard, Y. Pritykin, M. J. Purcaro, Q. Qin, G. Quinones-Valdez, I. Rabano, E. Radovani, A. Raj, N. Rajagopal, O. Ram, L. Ramirez, R. N. Ramirez, D. Rausch, S. Raychaudhuri, J. Raymond, R. Razavi, T. E. Reddy, T. M. Reimonn, B. Ren, A. Reymond, A. Reynolds, S. K. Rhie, J. Rinn, M. Rivera, J. C. Rivera-Mulia, B. S. Roberts, J. M. Rodriguez, J. Rozowsky, R. Ryan, E. Rynes, D. N. Salins, R. Sandstrom, T. Sasaki, S. Sathe, D. Savic, A. Scavelli, J. Scheiman, C. Schlaffner, J. A. Schloss, F. W. Schmitges, L. H. See, A. Sethi, M. Setty, A. Shafer, S. Shan, E. Sharon, Q. Shen, Y. Shen, R. I. Sherwood, M. Shi, S. Shin, N. Shoresh, K. Siebenthall, C. Sisu, T. Slifer, C. A. Sloan, A. Smith, V. Snetkova, M. P. Snyder, D. V.

Spacek, S. Srinivasan, R. Srivas, G. Stamatoyannopoulos, J. A. Stamatoyannopoulos, R. Stanton, D. Steffan, S. Stehling-Sun, J. S. Strattan, A. Su, B. Sundararaman, M.-M. Suner, T. Syed, M. Szynkarek, F. Y. Tanaka, D. Tenen, M. Teng, J. A. Thomas, D. Toffey, M. L. Tress, D. E. Trout, G. Trynka, J. Tsuji, S. A. Upchurch, O. Ursu, B. Uszczynska-Ratajczak, M. C. Uziel, A. Valencia, B. V. Biber, A. G. Van Der Velde, E. L. Van Nostrand, Y. Vaydylevich, J. Vazquez, A. Victorsen, J. Vielmetter, J. Vierstra, A. Visel, A. Vlasova, C. M. Vockley, S. Volpi, S. Vong, H. Wang, M. Wang, Q. Wang, R. Wang, T. Wang, W. Wang, X. Wang, Y. Wang, N. K. Watson, X. Wei, Z. Wei, H. Weisser, S. M. Weissman, R. Welch, R. E. Welikson, Z. Weng, H.-J. Westra, J. W. Whitaker, C. White, K. P. White, A. Wildberg, B. A. Williams, D. Wine, H. N. Witt, B. Wold, M. Wolf, J. Wright, R. Xiao, X. Xiao, J. Xu, J. Xu, K.-K. Yan, Y. Yan, H. Yang, X. Yang, Y.-W. Yang, G. G. Yardımcı, B. A. Yee, G. W. Yeo, T. Young, T. Yu, F. Yue, C. Zaleski, C. Zang, H. Zeng, W. Zeng, D. R. Zerbino, J. Zhai, L. Zhan, Y. Zhan, B. Zhang, J. Zhang, J. Zhang, K. Zhang, L. Zhang, P. Zhang, Q. Zhang, X.-O. Zhang, Y. Zhang, Z. Zhang, Y. Zhao, Y. Zheng, G. Zhong, X.-Q. Zhou, Y. Zhu, J. Zimmerman, J. E. Moore, M. J. Purcaro, H. E. Pratt, C. B. Epstein, N. Shoresh, J. Adrian, T. Kawli, C. A. Davis, A. Dobin, R. Kaul, J. Halow, E. L. Van Nostrand, P. Freese, D. U. Gorkin, Y. Shen, Y. He, M. Mackiewicz, F. Pauli-Behn, B. A. Williams, A. Mortazavi, C. A. Keller, X.-O. Zhang, S. I. Elhajjajy, J. Huey, D. E. Dickel, V. Snetkova, X. Wei, X. Wang, J. C. Rivera-Mulia, J. Rozowsky, J. Zhang, S. B. Chhetri, J. Zhang, A. Victorsen, K. P. White, A. Visel, G. W. Yeo, C. B. Burge, E. Lécuyer, D. M. Gilbert, J. Dekker, J. Rinn, E. M. Mendenhall, J. R. Ecker, M. Kellis, R. J. Klein, W. S. Noble, A. Kundaje, R. Guigó, P. J. Farnham, J. M. Cherry, R. M. Myers, B. Ren, B. R. Graveley, M. B. Gerstein, L. A. Pennacchio, M. P. Snyder, B. E. Bernstein, B. Wold, R. C. Hardison, T. R. Gingeras, J. A. Stamatoyannopoulos, Z. Weng, Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **583**, 699–710 (2020).

60. I. Yanai, H. Benjamin, M. Shmoish, V. Chalifa-Caspi, M. Shklar, R. Ophir, A. Bar-Even, S. Horn-Saban, M. Safran, E. Domany, D. Lancet, O. Shmueli, Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinforma. Oxf. Engl.* **21**, 650–659 (2005).

61. M. Giurgiu, J. Reinhard, B. Brauner, I. Dunger-Kaltenbach, G. Fobo, G. Frishman, C. Montrone, A. Ruepp, CORUM: the comprehensive resource of mammalian protein complexes—2019. *Nucleic Acids Res.* **47**, D559–D563 (2019).

62. A. D. Rouillard, G. W. Gundersen, N. F. Fernandez, Z. Wang, C. D. Monteiro, M. G. McDermott, A. Ma'ayan, The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. *Database* **2016**, baw100 (2016).

63. M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, G. Sherlock, Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).

64. J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J.-Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, A. Cardona, Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).

**Acknowledgements:** We thank Alina Xiao, Daniel Pederick, Tom Hindmarsh-Sten, Colleen McLaughlin, Liqun Luo, and other Luo lab members for helpful feedback on the manuscript and immunofluorescence analysis. We also thank Leslie Magtanong, Gabriella Cale, and other members of the Fraser Lab for helpful discussions and feedback on the manuscript. We thank the Stanford Functional Genomics core facility for preparing the scRNA-seq libraries. We thank Xabier López Aranguren and Asier Ullate Agote for sharing their cell type annotations and other metadata for the heart embryonic dataset. We thank Jonathan Pritchard, Richard Schneider, and Vanessa Barrone for helpful feedback on the manuscript. Biorender was used in parts of the main and supplemental figures.

**Funding:** Funding was provided by NIH R01DK121851 (awarded to HN), the Japan Agency for Medical Research and Development (AMED) Grant Number JP22bm1004002 (awarded to HN), NIH R01HG012285 (awarded to HBF), and JSPS KAKENHI Grant Number 21H02378 (awarded to TN). ALS was supported by an fellowship under grant number FA9550-21-F-0003. KJI was supported by a National Science Foundation Graduate Research Fellowship under grant no. DGE-1656518.

**Authors contributions:** TN led all wet-lab work including cell culture, chimera generation and assessment of donor contribution, FACS, preparation of cells for input for scRNA-seq, and immunofluorescence. KJI helped with FACS and sampling of chimeric embryos. CF helped with immunofluorescent staining and optimization. ALS performed all bioinformatic analysis, visualization, validation, and writing of software with guidance from HBF. ALS wrote the main and supplementary text and created Figs. with input from HBF, TN, and HN. ALS and TN wrote the methods section with input from HBF and HN. HBF and HN provided funding and conceived the study.

**Competing interests:** HN is a co-founder and shareholder in ReproCELL, Megakaryon, and Century Therapeutics. All other authors declare no competing interests.

**Data and materials availability:** Sequencing data will be available on May 10th through the gene expression omnibus with accession GSE266218. All code needed to reproduce the analyses described in this study is available at https://github.com/astarr97/Chimera_Pilot.

**List of supplementary materials:**

Figs. S1-S29

Tables S1-S3

Supplementary Text 1