Cross cell-type systems genetics reveals the influence of eQTL at multiple points in the developmental trajectory of mouse neural progenitor cells

Selcan Aydin[1,*], Daniel A. Skelly[1], Hannah Dewey[1,2], J. Matthew Mahoney[1], Ted Choi[3], Laura G. Reinholdt[1,2], Christopher L. Baker[1,2], and Steven C. Munger[1,2,*]

[1] The Jackson Laboratory, Bar Harbor, ME 04609 USA
[2] Graduate School of Biomedical Sciences, Tufts University, Boston, MA 02111 USA
[3] Predictive Biology, Inc., Carlsbad, CA 92010 USA
[*]Correspondence: selcan.aydin@jax.org; steven.munger@jax.org

Short title: Genetic dissection of the neural progenitor cell transcriptome

Keywords: Genetic diversity, eQTL, systems genetics, neural progenitor cells

Abstract

Genetic variation leads to phenotypic variability in pluripotent stem cells that presents challenges for regenerative medicine. Although recent studies have investigated the impact of genetic variation on pluripotency maintenance and differentiation capacity, less is known about how genetic variants affecting the pluripotent state influence gene regulation in later stages of development. Here, we characterized expression of more than 12,000 genes in 127 donor-matched Diversity Outbred (DO) mouse embryonic stem cell (mESC) and neural progenitor cell (mNPC) lines. Quantitative trait locus (QTL) mapping identified 2,947 expression QTL (eQTL) unique to DO mNPCs and 1,113 eQTL observed in both mNPCs and mESCs with highly concordant allele effects. We mapped three eQTL hotspots on Chromosomes (Chrs) 1, 10, and 11 that were unique to mNPCs. Target genes of the Chr 1 hotspot were overrepresented for those involved in mRNA processing, DNA repair, chromatin organization, protein degradation, and cell cycle. Mediation analysis of the Chr 1 hotspot identified *Rnf152* as the best candidate mediator expressed in mNPCs, while cross-cell type mediation using mESC gene expression along with partial correlation analysis strongly implicated genetic variant(s) affecting *Pign* expression in the mESC state as regulating the mNPC Chr 1 eQTL hotspot. Together these findings highlight that many local eQTL confer similar effects on gene expression in multiple cell states; distant eQTL in DO mNPCs are numerous and largely unique to that cell state, with many co-localizing to mNPC-specific hotspots; and mediation analysis across cell types suggests that expression of *Pign* early in development (mESCs) shapes the transcriptome of the more specialized mNPC state.

Introduction

Pluripotent stem cells (PSCs) have the capacity to self-renew indefinitely in culture and differentiate into any cell type. These properties make them a valuable resource for the study of gene regulation in various cellular contexts, giving us access to cell types that are otherwise difficult to obtain or follow through developmental time. Recently, systems genetics approaches have combined the power of genetic diversity with the differentiation potential of PSCs to identify genetic variants that influence gene expression regulation across different cellular contexts (Warren and Cowan 2018; Swanzey et al. 2021; Farbehi et al. 2024). For example, mouse embryonic stem cell lines (mESCs) derived from diverse strains have been used to study the maintenance of the pluripotent state, early lineage specification, and the stability of genomic imprinting (Skelly et

47   al. 2020; Byers et al. 2022; Aydin et al. 2023; Parikh et al. 2024). Similarly, the influence of genetic
48   variation on human induced PSCs (iPSCs) has been characterized at multiple molecular levels
49   (Kilpinen et al. 2017; Panopoulos et al. 2017; Banovich et al. 2018; Mirauta et al. 2020) and
50   throughout differentiation into various cell types (Cuomo et al. 2020; Jerber et al. 2021; Elorbany et
51   al. 2022; Wells et al. 2023; Popp et al. 2024). Despite these examples, our understanding of the
52   downstream consequences of pluripotent state variation on developmental trajectories and later
53   stages of development remains largely undefined. To address this gap, we turned to a system where
54   we could conduct systems genetics inquiry in both the pluripotent as well as a subsequent
55   developmental stage using cells from the same donors.
56
57   PSCs can be differentiated into neural progenitor cells (NPCs) that can proliferate in culture
58   continuously and have the capacity to further differentiate into specialized neuronal cell types such
59   as neurons and astrocytes (Conti et al. 2005). Similar to their pluripotent ESC progenitors, NPCs
60   represent a dynamic cell state maintained through external signaling molecules and growth factors
61   in cell culture rather than a distinct cell type observed in vivo. NPCs are an increasingly important
62   resource for cellular modeling of neurodevelopment and neurodegenerative disorders. For
63   example, Young-Pearse and colleagues recently analyzed neuronal cell types differentiated from a
64   set of 50 iPSC lines from the Religious Orders Study and Memory and Aging Project (ROSMAP) aging
65   cohorts to identify a neuron-specific interaction between the Alzheimer's Disease risk genes
66   SORL1, APOE, and CLU (Lee et al. 2023). In a separate study, genetically diverse NPCs derived from
67   human iPSCs were recently used to identify the genetic variants driving differences in susceptibility
68   to Zika virus (Wells et al. 2023). Finally, another recent study profiled hPSC differentiation to mature
69   neurons with single cell RNA sequencing (scRNA-seq) across a fine time course that captured
70   progenitor-like cell states. The authors compared genetically identical lines in different cell states
71   to identify molecular signatures that could predict differentiation efficiency into neuronal cells
72   (Jerber et al. 2021). However, genetic diversity was limited in these human cell studies, and they
73   lacked the sample sizes necessary to map trans eQTL interactions among genes that drive neural
74   lineage specification.
75
76   Powerful mouse resources including the Diversity Outbred (DO) heterogeneous stock combine
77   genetic variation from multiple parent strains in a randomized breeding design that minimizes
78   population stratification and ensures balanced allele frequencies across the genome, providing
79   optimal power for genetic mapping studies (Churchill et al. 2012). Recently, we applied a systems
80   genetics approach to mESCs derived from DO mice to infer the complex gene regulatory network
81   and genetic interactions that underlie ground state pluripotency (Skelly et al. 2020; Aydin et al.
82   2023). Specifically, we profiled chromatin accessibility, gene expression, and protein abundance in
83   a large panel of DO mESCs and mapped thousands of quantitative trait loci (QTLs) that affected
84   chromatin accessibility (caQTLs), transcript abundance (eQTLs) and protein abundance (pQTL)
85   (Skelly et al. 2020; Aydin et al. 2023). In this study, we extend our systems genetics approach to
86   characterize variation in transcript abundance and identify cell-state specific and conserved eQTLs
87   in mouse NPCs derived from many of the same DO mESC lines used in our previous studies (Skelly
88   et al. 2020; Aydin et al. 2023). In line with our findings in DO mESCs, we observe high expression
89   variation among the DO mNPCs. Moreover, despite large expression differences overall between
90   the mESC and mNPC cell states, we observed high co-variation in mESC and mNPC
91   transcriptomes from the same DO donors. Genetic mapping identified thousands of significant
92   local and distant eQTL, of which a quarter — mostly local eQTLs — were also detected previously in
93   DO mESCs. Most distant eQTL, by contrast, were uniquely detected in mNPCs, including three
94   eQTL hotspots on Chrs 1, 10, and 11. Targets of the Chr 1 hotspot were enriched for roles in

95   chromosome segregation and the spindle assembly checkpoint (SAC). Cross cell-type mediation
96   and partial correlation analysis identified two potential mediator genes of this hotspot, one
97   expressed in the mNPCs (*Rnf152*) and the other earlier in the pluripotent mESCs (*Pign*). Both
98   *Rnf152* and *Pign* have been implicated in cell-cycle regulation with *Pign* having a direct role in the
99   SAC (Okamoto et al. 2020; Teye et al. 2021), suggesting that these genes drive variation in cell-cycle
100   regulation in DO mNPCs. More broadly, our study demonstrates the power of applying a systems
101   genetics approach across timepoints in a developmental trajectory, which allowed us to map
102   eQTLs and link gene expression variation in one cell state to a causal variant affecting expression of
103   a regulatory gene in a progenitor cell.

104

105   **Results**

106   **Gene expression in genetically diverse mNPCs is highly variable and co-varies with gene**
107   **expression in mESCs**
108   To better understand how the abundant genetic variation segregating in the DO population affects
109   gene expression during neural differentiation, we quantified the transcript abundance of 14,163
110   genes by RNA-seq in 186 DO mNPC lines (Figure 1A). We first performed principal component
111   analysis (PCA) on mNPC gene expression (PCA-N) to identify the genes and pathways that are most
112   variable across the DO mNPC lines. The first principal component (PC1-N, Figure S1A) explained
113   14% of expression variation among DO mNPCs, and functional enrichment analysis showed that
114   PC1-N driver genes were overrepresented for those involved in the cell cycle, mRNA processing,
115   translation, response to leukemia inhibitory factor (LIF), *in utero* embryonic development, and
116   targets of transcription factors with roles in neural differentiation such as *Otx2* (Vernay et al. 2005)
117   (Table S1). A closer look at the cell cycle-related processes revealed genes with roles in almost all
118   of the cell cycle checkpoints, underscoring the proliferative nature of mNPCs under expansion
119   culture conditions.

120

121   Next, we sought to better understand how transcriptional variation in DO mNPC lines compared to
122   variation in their progenitor pluripotent DO mESC lines. We previously characterized gene
123   expression in mESCs from the same DO donor for 127 of the mNPC lines in the current study (Skelly
124   et al. 2020). Over 12,000 genes were expressed in both cell states, and we observed significant
125   covariation between the mESC and mNPC transcriptomes of these 127 cell lines (median intraline
126   NPC-ESC correlation = 0.73, permutation test p < 2.2e-16; Figure 1B, Figure S1B), likely indicating a
127   large role for genetic background in conferring expression variability. Similarly, we observed high
128   concordance between the mean and variance of genes expressed in both cell states, with a
129   correlation of 0.77 (p < 2.2e-16) and 0.74 (p < 2.2e-16), respectively (Figure S1C-D).

130

131   To discern the sources of expression variation that were unique to mNPCs or shared with mESCs,
132   we performed PCA on the combined gene expression counts from both cell types (PCA-C). Cell
133   state appeared to be the primary driver of expression variation in the combined dataset (PC1-C,
134   48%; Figure 1C), despite the high within-line co-variation of expression in mNPCs and mESCs.
135   Functional enrichment analysis of PC1-C implicated ubiquitin mediated protein degradation,
136   nervous system development, signaling pathways regulating pluripotency, and cellular
137   differentiation as the major drivers of gene expression variation between the mESC and mNPC lines
138   (Table S2). Of note, PC1-C showed high agreement with PC1-N values (cor = -0.9) for the 127 DO
139   mNPC lines included in both analyses. However, we find little overlap in the genes contributing to
140   PC1-N and PC1-C, with only 39 genes shared among 709 driver genes for PC1-N and 605 driver
141   genes for PC1-C. While PC1-C clearly corresponded to cell state expression differences, no other

142 PC in the combined dataset appeared to distinguish between cell states; indeed, PC2-C and PC3-C
143 did not show any separation, exhibiting a range of values that spanned both cell states, and likely
144 capturing variation common to both (Figure 1C). Drivers of PC2-C and PC3-C were enriched for
145 genes involved in essential cellular processes including the mitotic cell cycle, mitochondrial cell
146 function, mRNA processing, membrane lipid metabolism and chromatin organization (Table S2).
147 PC4-C appears to differentiate mESC lines by their chromosomal sex, which was previously
148 observed and attributed to the two active X chromosomes in the pluripotent mESC state (Skelly et
149 al. 2020; Aydin et al. 2023). This strong sex effect is not present in the mNPCs where the second X
150 chromosome is known to be inactivated (Figure 1C). Variance decomposition analysis confirms this
151 difference in sex effect on gene expression in mESCs and mNPCs (Figure S1E).
152
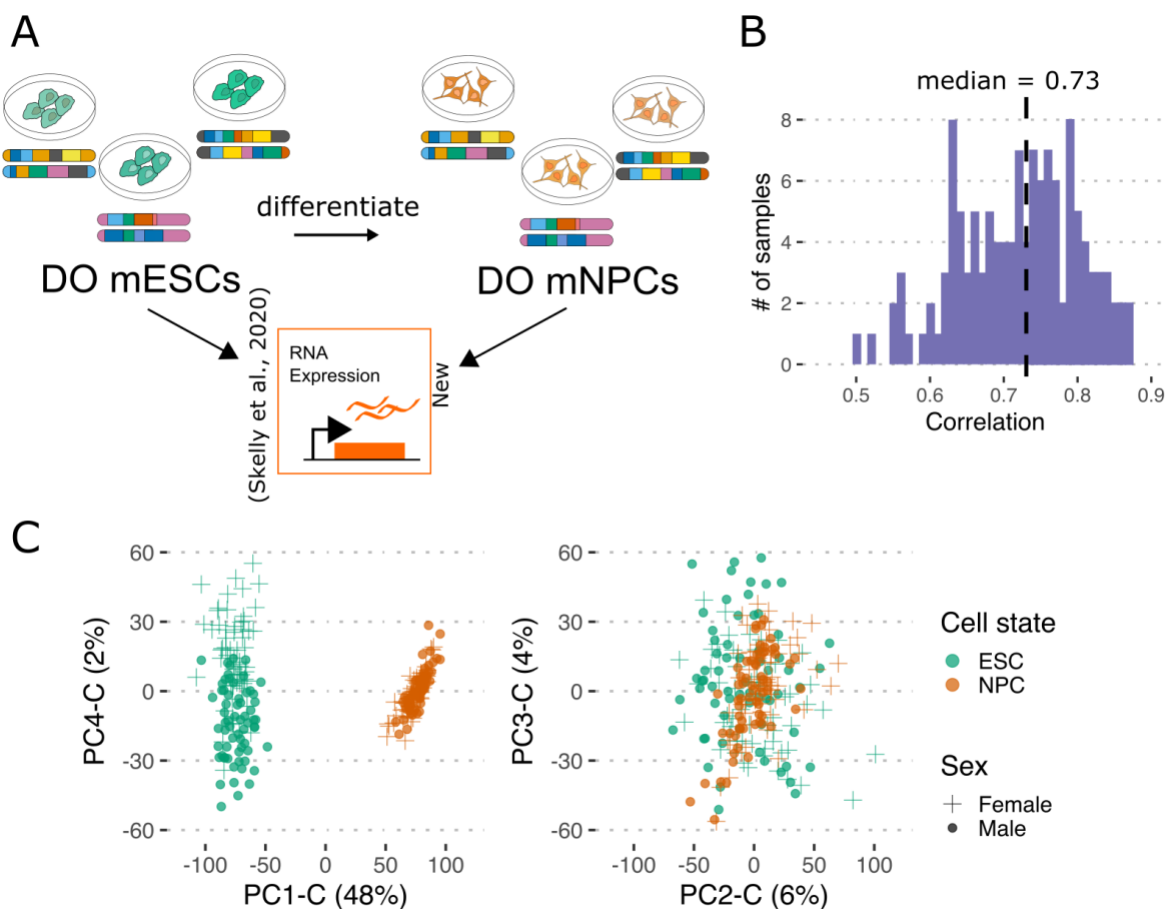
153
154 **Figure 1**. Variation in DO mNPC transcriptome and covariation to DO mESCs. (A) Experimental overview of
155 the data. (B) Histogram of correlation coefficients between the ESC and NPC transcriptomes of 127 samples.
156 (C) Principal component analysis results of the combined ESC, NPC transcriptomes of 127 samples and
157 12,095 genes.

158 To better characterize biological processes that were differentially regulated across the cell states,
159 we performed differential gene expression analysis. We identified 1,409 upregulated and 773
160 downregulated genes in mNPCs relative to mESCs (adjusted $p < 0.05$, abs(log2 fold change) > 2).
161 Genes upregulated in mNPCs were enriched for biological processes including nervous system
162 development and pathways involved in neural differentiation such as WNT signaling. Conversely,
163 genes downregulated in mNPCs were enriched for biological processes including ribosome

164   biogenesis and pathways regulating pluripotency such as response to LIF (Table S3), confirming the
165   PCA results.
166
167   **Genetic architecture of the mNPC transcriptome**
168   To further characterize the sources of interline expression differences in mNPCs, we estimated
169   narrow sense heritability. We estimated the transcript abundance of over 90% of expressed genes
170   in mNPCs to be heritable (median h^2 = 0.23) — confirming that genetic diversity among DO mNPCs
171   is a major driver of transcriptome variation. Genetic mapping identified 4,060 significant eQTL (LOD
172   > 7.5, alpha = 0.05, FDR = 0.075), with almost one-third of expressed transcripts having one or more
173   significant eQTL (Table S4, Figure 2A). The majority of eQTL are local (n = 2,998), meaning the
174   genetic variant impacting the abundance of the transcript is within close proximity (± 10Mb) to the
175   gene itself. The remaining quarter are distant eQTL (n = 1,062), where the peak is further away from
176   the gene or on a different chromosome and likely influences the target gene's expression indirectly
177   through direct effects on the expression or function of another "mediator" gene. Similar to previous
178   eQTL analyses, we observed that many distant eQTL colocalized to eQTL "hotspots". We detected
179   three such eQTL hotspots on Chrs 1, 10, and 11 that were unique to mNPCs (Figure 2B, Table S5).
180   Interestingly, we also mapped a significant QTL for PC1-N values (LOD > 7.11, alpha = 0.1) to the
181   same hotspot on Chr 1, suggesting that this hotspot is a causal driver of global transcriptional
182   variation in mNPCs (Figure S2A).
183
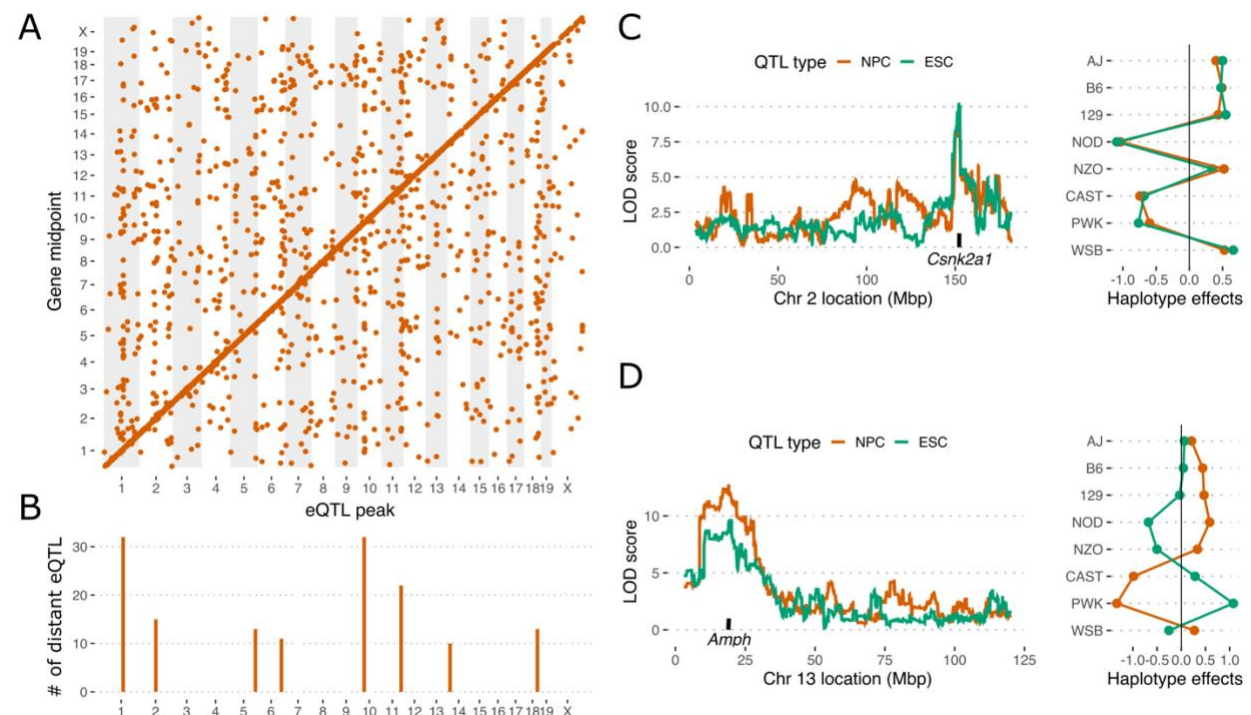
184
185   **Figure 2.** Genetic architecture of DO mNPC transcriptomes. (A) Map of DO mNPC eQTL containing 4,060
186   significant peaks (LOD > 7.5) from 3,806 unique genes. (B) eQTL hotspots identified across the genome. (C)
187   eQTL scan for *Csnk2a1* gene on the left and the haplotype effects at the eQTL peak on Chr 2 on the right. (D)
188   eQTL scan for *Amph* gene on the left and the haplotype effects at the eQTL peak on Chr 13 on the right.

189   Next, we compared our mNPC eQTL map to the mESC eQTL map from our previous study (Skelly et
190   al. 2020). Over a quarter of eQTL in mNPCs are also detected in mESCs and show concordant allele

191 effects (abs(cor) > 0.7 and adjusted p < 0.1; n = 1,113 concordant eQTL / 4,060 total eQTL, Figure
192 S2A-B); these concordant eQTL are primarily local variants (963 local eQTL out of 1,113 total
193 concordant eQTL). For example, *Csnk2a1* has a shared local eQTL with highly similar effects in both
194 mNPCs and mESCs (Figure 2C). *Csnk2a1* is known to regulate various cellular processes including
195 WNT signaling, cell cycle progression, apoptosis and transcription, and mutations in this gene have
196 recently been linked to neurodevelopmental abnormalities in humans (Okur et al. 2016). Of note,
197 we also map concordant local eQTL with opposing allele effects on gene expression in mNPCs and
198 mESCs. Although much fewer in number, one such example is *Amph* (Figure 2D), a gene involved in
199 synaptic vesicle endocytosis where loss of expression leads to cognitive deficits in mouse models
200 (Paolo et al. 2002; Zhang et al. 2021). The remaining three quarters of eQTL are unique to mNPCs (n
201 = 2,947 unique / 4,060 total eQTL) and affect target genes that are enriched for roles in cell-cell
202 signaling and nervous system development (Table S6).
203
204 **Cross-cell type mediation analysis uncovers the influence of genetic variation across cell**
205 **states**
206 To characterize the eQTL hotspots uniquely observed in DO mNPCs, we performed functional
207 enrichment analysis to identify affected pathways and mediation analysis to identify candidate
208 regulatory genes in the hotspot regions. For these analyses, we relaxed our inclusion criteria to
209 include all target genes that mapped with a suggestive eQTL (LOD > 6) in the three hotspot regions
210 on Chrs 1, 10, and 11. Targets of the Chr1 hotspot were significantly enriched for genes involved in
211 mRNA processing, DNA repair, chromatin organization, protein degradation and cell cycle (adjusted
212 p-value< 0.01; Table S7). Genes involved in protein processing in the ER were significantly
213 overrepresented for the hotspot on Chr 10, while the Chr 11 hotspot did not show any significant
214 functional overrepresentation. Mediation analysis failed to identify strong candidate genes
215 underlying the Chrs 10 and 11 eQTL, suggesting that these hotspots may stem from genetic
216 variant(s) affecting abundance of an RNA not profiled (e.g., miRNA), a post-transcriptional
217 modification of RNA, function or stability of a protein, or any mechanism that does affect the
218 transcript abundance of the mediator gene. We excluded these loci from further analysis.
219
220 Measurement error — including read mapping errors in RNA-seq datasets — can adversely affect
221 individual gene expression estimates and introduce noise into eQTL mapping. Dimensionality
222 reduction techniques like PCA are an effective way to summarize the main component of
223 expression variation among eQTL hotspot target genes to infer founder allele effects and identify
224 candidate mediator genes. Accordingly, we performed PCA on the transcript abundance of hotspot
225 target genes and mapped QTL for the eigengene value (Eigen-Chr1). As expected, Eigen-Chr1
226 mapped with a significant QTL to the Chr 1 hotspot and exhibited founder haplotype effects similar
227 to the observed individual target gene eQTL. Specifically, DO mNPC lines that inherited the Chr 1
228 hotspot region from B6, 129, NOD, or CAST founder strains showed similar expression of the eQTL
229 target genes, while those lines that inherited this locus from AJ, NZO, PWK, or WSB showed a
230 different pattern of gene expression (Figure 3A, B). Of note, the PC1-N Chr 1 QTL also showed the
231 same 4:4 split in founder haplotype effects (Figure S3A). PC1-N values showed high agreement to
232 Eigen-Chr1 (cor = -0.99), which likely reflects the significant overlap between the individual gene
233 targets of the Chr 1 hotspot (overlap n = 193 / 322) and the gene drivers of PC1-N (overlap n = 193 /
234 709, hypergeometric p = 1e-86).
235
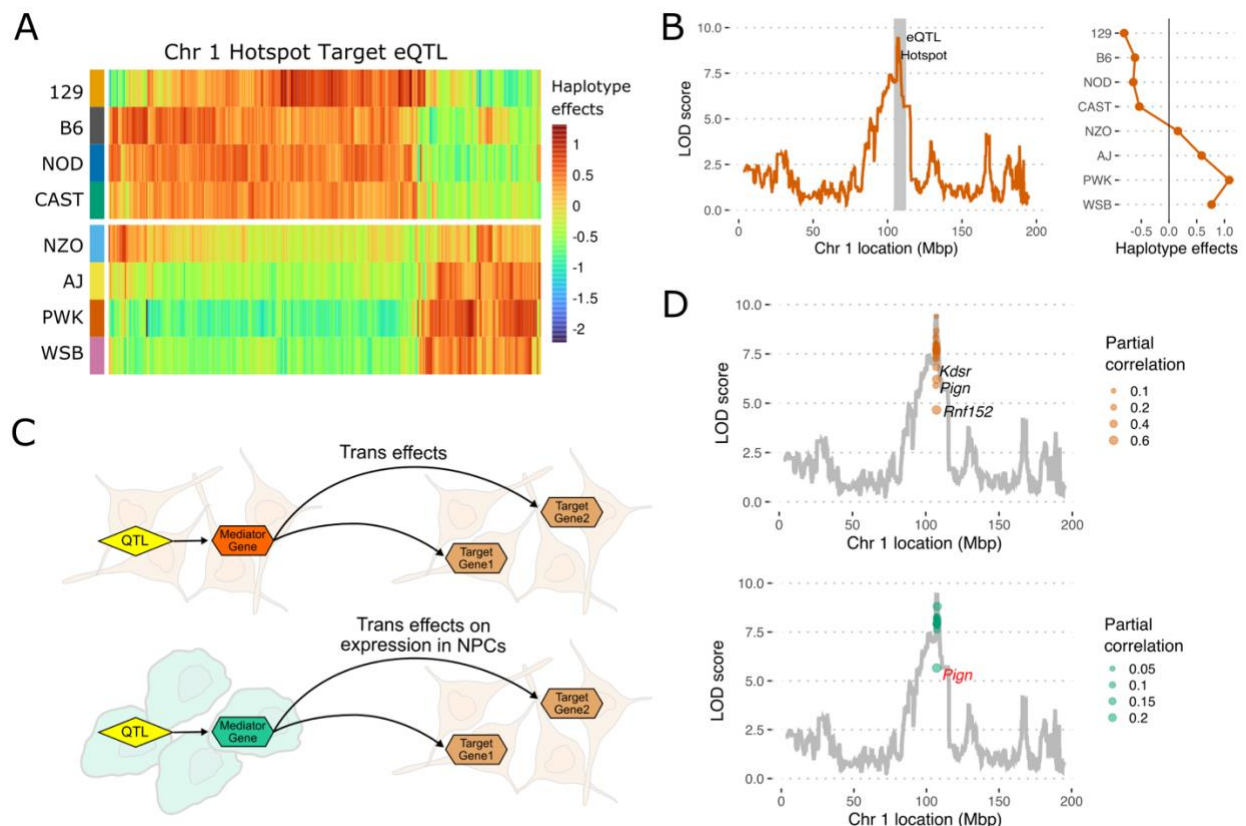
Figure 3. Details of the Chr 1 eQTL hotspot. (A) Heatmap of founder haplotype effects at the eQTL peaks of the target eQTL (LOD >6, n = 322) within the Chr 1 hotspot. (B) QTL scan of PC1-T of the Chr 1 target eQTL on the left, founder haplotype effects at the PC1-T QTL peak on the right. (C) Cartoon depicting mediation analysis using ESC and NPC expression to identify potential regulators of an eQTL hotspot. (D) In gray, PC1-T QTL scan is plotted overlaid with the new LOD scores obtained after mediation with genes in the region. Red gene name depicts genome wide significance for the mediation LOD drop.

Given the large physical distance between eQTL hotspots and their target genes, it is generally assumed that the trans effects of these variants are conferred via direct proximal effects (i.e. cis effects) on the expression or function of a regulatory gene in the hotspot region. In cases where the causal genetic variant alters the expression level of the regulatory gene, mediation analysis can be a powerful tool to identify this "mediator" gene (Chick et al. 2016; Skelly et al. 2020). Importantly, an eQTL hotspot in one cell state (e.g. mNPCs) could stem from a genetic variant(s) that influences the expression of a regulatory gene in that same cell state *or* in a progenitor stem cell (e.g. mESCs). We performed mediation analysis on the Eigen-Chr1 values to identify potential mediators of the Chr 1 hotspot target genes, searching first in the mNPC expression data and then in the mESC expression data (Figure 3C). Initially, we used a standard regression model to identify this potential local regulator within the Chr 1 hotspot (Chick et al. 2016). To gain more insight into the relationship between candidate mediators and the Chr 1 hotspot, we also performed partial correlation analysis — assessing the correlation of expression levels between candidate mediators and target genes after accounting for the observed effect of the distant eQTL — with the expectation that a true causal mediator should still co-vary in expression with the target genes even in the absence of the observed QTL effect. We calculated the partial correlation between Eigen-Chr1 and the expression of candidate mediators after accounting for the genetic ancestry at the QTL peak.

261 Mediation analysis with the mNPC expression data identified *Rnf152* as the best candidate mNPC
262 mediator in the Chr 1 hotspot. That is, regressing out *Rnf152* expression in mNPCs caused the
263 largest LOD drop in the most target eQTL and the Eigen-Chr1 QTL (Figure 3D, Figure S3B). None of
264 the candidate mediators, including *Rnf152*, were statistically significant as assessed by genome-
265 wide scaling, but *Rnf152* showed the highest partial correlation to Eigen-Chr1 after regressing out
266 the QTL effect (correlation = 0.6, adjusted p = 3.5e-12). *Rnf152* encodes a ubiquitin ligase residing in
267 the lysosome; it is reported to be involved in mammalian target of rapamycin (mTOR) signaling and
268 has been shown to regulate autophagy and cell proliferation (Deng et al. 2015; Okamoto et al.
269 2020). *Rnf152* has a local mNPC eQTL with a split in founder allele effects similar to the eQTL
270 hotspot, consistent with its role as a Chr 1 hotspot mediator. (Figure S3C).

272 Next, we expanded our mediation analysis across cell states to identify potential mediators of the
273 mNPC Chr 1 hotspot in the mESC expression data. We identified *Pign* as the best candidate mESC
274 mediator for the mNPC Chr 1 hotspot, causing the largest LOD drops in the most target eQTL and
275 the Eigen-Chr1 QTL. Further, mESC *Pign* expression showed the highest partial correlation to Eigen-
276 Chr1 after regressing out the Chr 1 eQTL effect (correlation = 0.3, adjusted p = 0.1) (Figure 3D,
277 Figure S3B). The best mNPC mediator, *Rnf152*, was found to be expressed at very low levels in
278 mESCs and did not pass our filtering threshold — consistent with previous reports (Lienert et al.
279 2011). *Pign* encodes for an enzyme involved in glycosylphosphatidylinositol (GPI)-anchor
280 biosynthesis, and more recently was linked to cell cycle regulation through interaction with the
281 spindle assembly checkpoint (SAC) complex (Teye et al. 2021). In DO mESCs and mNPCs, *Pign* has
282 a significant local eQTL with effects that match the 4:4 founder allele split observed for the eQTL
283 hotspot (Figure S3D). Importantly, target genes of the Chr 1 hotspot include those involved in proper
284 chromosome segregation at mitosis such as *Mad2l1* (MAD2), a member of the SAC complex
285 (Dobles et al. 2000).

287 **Discussion**

288 Over the past two decades, genome-wide association studies (GWAS) in large human cohorts have
289 identified thousands of common variants associated with variation in numerous complex traits and
290 disease phenotypes (Abdellaoui et al. 2023). Many of these variants likely influence adult
291 phenotypes through direct effects on gene expression or cell differentiation early in development,
292 making them difficult to assay directly in humans (Maurano et al. 2012). In this study, we focused
293 on neural differentiation, and sought to better understand how genetic variation can influence gene
294 expression using a large panel of mouse neural progenitor cells derived from the powerful Diversity
295 Outbred heterogeneous stock. We analyzed interline variation in mNPC gene expression,
296 performed eQTL mapping and mediation analysis, and compared our results to pluripotent mESC
297 lines from the same donors. We identified genetic variation to be a significant driver of gene
298 expression variability in DO mNPCs. Although they represent very different cellular states and this
299 difference is clearly evident in their transcriptomes, genetically identical mESC and mNPC lines
300 showed high covariance in gene expression (median intraline correlation = 0.73) — indeed,
301 considerably higher agreement than we observed between transcript and protein abundance within
302 genetically identical mESC lines (median intraline correlation = 0.36) (Aydin et al. 2023). Gene
303 expression differences between cell states highlight known differentiation pathways, as genes
304 involved in nervous system development and differentiation are upregulated in mNPCs while genes
305 with roles in pluripotency maintenance are downregulated in mNPCs relative to mESCs.
306 Meanwhile, genes involved in essential cellular functions like the cell cycle, mitochondrial function,
307 RNA processing, and membrane lipid metabolism were found to be variably expressed in both cell

308   states. Given that mESCs and mNPCs are both self-renewing in culture, and given the dynamics
309   and requirements of actively dividing cells, we were not surprised to observe high expression
310   variability for genes involved in cell cycle regulation and cellular metabolism.
311
312   Expression QTL mapping added context to these expression patterns and uncovered thousands of
313   genetic variants responsible for their expression variation. We found that over a quarter of
314   significant eQTL detected in mNPCs were also mapped in mESCs and showed similar allele effects.
315   These concordant eQTL were primarily locally-acting and affected many genes with essential
316   cellular functions. Population variation at these loci likely accounts for much of the high expression
317   covariation observed in mNPCs and mESCs from the same DO donor. The remaining three quarters
318   of mNPC eQTL appear unique to this cell state; of note, a similar proportion of mESC eQTL were
319   also cell state-specific (4,854 / 6,069 or 80%). Many of these cell state-specific distant eQTL co-
320   localize to genomic hotspots and likely play significant roles in cell lineage specification. For
321   example, the previously identified Chr 15 eQTL hotspot regulating pluripotency maintenance is only
322   observed in DO mESCs (Skelly et al. 2020; Aydin et al. 2023), while the hotspots on Chrs 1, 10, and
323   11 are only observed in DO mNPCs. Further, we found that genes involved in cell-cell signaling and
324   nervous system development are significantly overrepresented among the unique distant eQTL in
325   DO mNPCs. Together, these results highlight the abundance and widespread distribution of
326   expression-modulating variants in genetically diverse populations, underscore a role for conserved
327   eQTL in driving interindividual variation within cell states, and provide further support for the
328   importance of distant eQTL to lineage specification driving cell state-specific functions.
329
330   We identified three eQTL hotspots on Chrs 1, 10, and 11 with distinct genetic effects in DO mNPCs.
331   Furthermore, mediation analysis helped us identify potential regulators for the hotspot on Chr 1.
332   One of the top candidates, *Rnf152*, is a ubiquitin ligase that regulates autophagy. Although the drop
333   in LOD scores obtained through mediation analysis did not achieve genome-wide significance for
334   the target eQTL, *Rnf152* showed the highest partial correlation to the PC1 of the expression of target
335   genes when controlled for the founder strain ancestry at the QTL peak. Interestingly, the other
336   candidate regulator, *Pign,* was identified from the ESC state. *Pign* was shown to regulate the spindle
337   assembly checkpoint (SAC) by forming a complex with the SAC proteins MAD1, MAD2, and the
338   mitotic kinase MPS1 where loss of *Pign* expression led to an increase in errors in chromosome
339   segregation and dysregulation of the SAC (Teye et al. 2021). In addition, *Pign* is linked to a variety of
340   human disorders including some with neurological symptoms like developmental delay, epilepsy
341   and seizures (Maydan et al. 2011). Although neither *Rnf152* nor *Pign* are transcription factors and
342   are unlikely to directly regulate gene expression, they could be indirectly regulating the expression
343   of downstream target genes through cellular signaling and influencing the progression of the cell
344   cycle. *Rnf152* and *Pign* are separated by less than 200kb on Chr 1, and we identified four candidate
345   SNPs in the region that match the observed 4:4 strain allele effect (rs31971829, rs47676250,
346   rs37516825, rs47457121) and may influence expression of one or both genes. Future experiments
347   will seek to validate the causal variant(s) underlying this mNPC hotspot eQTL and establish the cell
348   state of origin (i.e., mNPC or mESC) for its direct effects.
349
350   In this study, we obtain the first comprehensive look at the variation in the transcriptomes of a
351   genetically diverse set of mNPCs. In addition, we characterized the influence of genetic variation
352   across cell states using a large panel of genetically identical mESC and mNPC lines. Our
353   experimental design allowed us to trace the impacts of genetic variation across two states in a
354   developmental trajectory, which showed that primarily local genetic effects are shared between
355   cell states. However, trans effects, even ones with large influences (e.g. *Lifr* allele in mESCs (Skelly

356   et al. 2020)), are not conserved. Combined with the overrepresentation of neural development
357   genes among those with unique distant eQTL in mNPCs, our results emphasize the specialized
358   regulation of gene expression across cell states.
359
360   Finally, we acknowledge a number of study limitations that temper the strength of our conclusions.
361   Although comparable in size to other published eQTL studies, our DO mNPC panel is not optimally
362   powered to detect variants that confer only subtle effects. Given that essential regulatory genes
363   may not tolerate large effect variants, we concede that our inability to detect small effect eQTL may
364   cause us to miss important genetic interactions. Mediation analysis is a powerful tool but limited to
365   cases where the causal variant(s) underlying the distant eQTL modulates the transcript abundance
366   of the mediator gene. Distant eQTL that stem from causal variants that influence post-
367   transcriptional regulation, protein function or stability, or any other mechanism that does not
368   impact transcript abundance of the mediator gene will be missed by mediation analysis. Similarly,
369   partial correlation between genes does not necessarily indicate a direct regulatory or functional
370   interaction.
371
372   **Materials and Methods**

373   **Neural differentiation of DO mESCs**
374   Diversity Outbred ES cell lines were produced as previously described (Skelly et al. 2020). Neural
375   stem cell lines were produced by differentiation of ES cell lines following previously established
376   protocols (Pollard et al. 2006). Briefly, cryopreserved ES lines were thawed and carried for 3 to 6
377   passages prior to differentiation. ES cells were trypsinized and 50K to 75K cells were plated per well
378   of a laminin treated 12 well plate in NS Media with FGF2 and EGF (Pollard et al. 2006). The medium
379   was changed daily for 8 days, then detached with accutase, washed with NS media and replated
380   into laminin coated 6 well plates with 2mls per well of NS media. Neural stem cells were expanded
381   in clusters for 3 days then serially expanded into 6cm and then 10cm dishes until subconfluent. NS
382   cells were passaged for an additional 3 weeks and then cryopreserved. NS lines were screened
383   positive for neural stem cell markers such as nestin and GLAST, and negative for ES cell markers
384   such as SSEA1. The lines were further validated by differentiation to terminal neural cell types such
385   as astrocytes and neurons, as demonstrated by flow cytometry after cell type specific staining.
386
387   **Diversity Outbred mNPC RNA-seq**
388   Total RNA was isolated from each of 186 DO mNPC lines and quantitated by paired-end RNA
389   sequencing. Briefly, for each mNPC line, one 15cm dish of cells was grown to near confluence,
390   washed 3x with PBS, and mechanically harvested to yield 10M cells. About 100k cells from each
391   frozen cell pellet were used for RNA sequencing. Next, total RNA was extracted using the Quick-
392   RNA 96 well format kit (Zymo Research) with in-column DNase treatment. Sequencing libraries
393   were prepared by Akesogen using the TruSeq Stranded mRNA HT kit (Illumina, Cat no. 20020595)
394   and included ribosomal RNA reduction and poly-A selection, enzymatic fragmentation, cDNA
395   synthesis from random hexamer priming, adaptor ligation and PCR amplification steps to generate
396   indexed, stranded mRNA-seq libraries. Libraries were checked for quality and quantitated with the
397   Agilent Bioanalyzer, and samples that failed QC were repeated starting from the cryovial stage.
398   Finally, pooled libraries were sequenced on the NextSeq platform (Illumina) using the NextSeq
399   500/550 High Output v2 150-cycle kits (Illumina, Cat no. FC-404-2002). To minimize technical
400   variation, samples were randomly assigned to lanes prior to sample processing steps, barcoded,
401   and multiplexed at 16 samples per flow cell, yielding 6M-55M 2x75bp paired-end (PE) reads per
402   sample.

403
404 We aligned paired-end 75bp reads with bowtie v1.1.2 (Langmead et al. 2009) to a pooled "8-way"
405 transcriptome containing strain-specific isoform sequences from all eight DO founder strains as
406 previously described (Skelly et al. 2020). In order to identify and correct sample mix-ups, we
407 inferred sample genotypes from the RNA-seq data using GBRS v0.1.6 (Choi et al. 2020) and
408 compared GBRS-derived genotypes to our DNA genotypes obtained from GigaMUGA arrays.
409 Correlations between genotypes inferred from GBRS and GigaMUGA for the same sample were
410 typically on the order of 0.8-0.9 whereas for different samples they were less than 0.5. We resolved
411 17 sample mix ups for DO mNPC lines showing an incongruent genotype. For resolving multi-
412 mapping reads and quantifying transcript- and gene- level expression counts, we utilized EMASE as
413 implemented in GBRS v0.1.6 (Raghupathy et al. 2018; Choi et al. 2020). We filtered genes with a
414 median TPM (transcripts per million) value smaller than 0.5 or zero value (i.e., not expressed) in
415 more than half of the samples. To account for differences in library size, we normalized gene-level
416 counts to the upper quartile value then applied the ComBAT function from R/sva package to remove
417 batch effects caused by library preparation (Johnson et al. 2007). We transformed normalized and
418 batch corrected values to rank normal scores using rankZ normalization (Gatti et al. 2014). Gene
419 annotations such as MGI symbol, gene location and gene biotype were added using v84 of the
420 Ensembl database.
421
422 **Diversity Outbred mESC RNA-seq**
423 Raw RNA-seq data was retrieved from ArrayExpress (E-MTAB-7728) and analyzed same as to the DO
424 mNPC RNA-Seq data.
425
426 **Statistical analysis**
427 All analyses and figures were generated with the R statistical programming language (R Core Team
428 2023). Unless otherwise stated, R/tidyr package was used for data processing, R/ggplot2 for
429 plotting, and R/pheatmap for heatmap plots.
430
431 **Functional enrichment analysis**
432 We performed functional enrichment analysis using the 'gost' function in the gProfiler2 package
433 (Raudvere et al. 2019; Kolberg et al. 2020) by controlling the version using
434 'set_base_url(set_base_url("https://biit.cs.ut.ee/gprofiler_archive3/e102_eg49_p15/"))' in R using
435 an appropriate universal background on a case-by-case basis and 'fdr' option for p-value
436 correction. For example, we used the list of all genes that are expressed in NPCs for enrichment
437 analysis of the DO mNPC Chr 1 hotspot target eQTL. For Gene Set Enrichment Analysis (GSEA) we
438 utilized the R/fgsea package (Korotkevich et al. 2021) with gene sets that belong to the Gene
439 Ontology Biological Processes (GO:BP) subcategory in the molecular signatures database (MSigDB)
440 and 10,000 permutations.
441
442 **Differential expression analysis**
443 Differential expression was calculated for the 12,095 genes expressed in both mESCs and mNPCs
444 from the same DO donor (n = 127 lines total). In addition to the removal of the batch effects, sex
445 effects were also removed before the comparison in both data sets using ComBAT as implemented
446 in the R/sva package (Johnson et al. 2007). We compared gene expression using the Wilcoxon Rank
447 Sum test, implemented as the 'wilcox.test' function in the R/rstatix package, and adjusted raw p-
448 values to correct for multiple testing using the Benjamini-Hochberg method "BH". The log2 fold
449 change was calculated by taking the log2 of the ratio between the mean expression of each gene in
450 mNPCs over mESCs.

451

452 **Quantitative Trait Locus mapping**
453 Genetic mapping was performed using the R/qtl2 package (Broman et al. 2019) with wrapper
454 functions utilizing parallelization for efficient large-scale eQTL analysis in R/QTLretrievR
455 (https://github.com/deweyhannah/QTLretrievR). Briefly, eQTL were mapped with a linear mixed
456 model — implemented as the 'scan1' function in R/qtl2 — using the upper quartile normalized,
457 batch corrected, rankZ transformed gene expression values, with chromosomal sex included as an
458 additive covariate and the Leave One Chromosome Out (loco) option used for kinship
459 correction(Yang et al. 2014). To estimate genome-wide significance, we permuted genotypes 1000
460 times while maintaining the relationship between the phenotype and covariates. For each
461 permutation we retained the maximum LOD score to generate a null distribution for the test
462 statistic (Churchill and Doerge 1994). To calculate thresholds for eQTL, we repeated this
463 permutation strategy for all transcripts and estimated a significance cutoff at LOD > 7.5 (alpha =
464 0.05), and a suggestive cutoff at LOD > 6. False discovery rates (q-values) were determined for each
465 permutation-derived p-value with R/qvalue software, using the bootstrap method to estimate pi0
466 and the default lambda tuning parameters (Storey et al. 2004) and our LOD threshold of 7.5
467 corresponds to an FDR of 0.075. We call a QTL 'local' if the QTL peak is within ±10 Mbp to the
468 midpoint of its corresponding gene and 'distal' if otherwise.
469
470 Founder allele effects were estimated as best linear unbiased predictors (BLUPs) at the QTL peak
471 using 'scan1blup' function in R/qtl2 package. To identify overlaps with significant NPC eQTL, we
472 used a relaxed threshold of LOD > 5 for ESC eQTL. They were classified as shared if the eQTL peaks
473 were within +/-5 Mb of each other and the correlation between haplotype effects was significant
474 (adjusted p-value < 0.1).
475
476 For hotspot calling, we first identified distal eQTL that reach genome-wide permutation-based
477 threshold (p < 0.05; LOD 7.5). Next, we applied a sliding window method to identify hotspots as
478 described in (Skelly et al. 2020). Briefly, we counted the number of distal eQTL within 1cM windows
479 (0.25 cM shift) across the genome and selected the top 0.5% of bins with the most distant eQTL
480 (0.5% bin threshold 9 distant eQTLs). Final coordinates for each hotspot were determined using the
481 Bioconductor package 'GenomicRanges' to merge adjacent bins into a single region (Lawrence et
482 al. 2013). Each hotspot contains >20 significant distant eQTL (>150 suggestive, LOD > 6) and shows
483 distinct allelic effect patterns.
484
485 **Partial correlation analysis**
486 We used the 'pcor' function from the R/ppcor package (Kim 2015) to calculate the partial
487 correlation between the expression of the mediator gene and the first principal component of the
488 expression of genes with the target eQTL (Eigen-Chr1) on Chr 1. To control for the genetic effects in
489 the partial correlation analysis, we classified the genotypes at the Eigen-Chr1 QTL peak based on
490 their ancestry from the 8 DO founder strains. Given the 4:4 split at the Eigen-Chr1 QTL peak, if the
491 cell line showed homozygous ancestry from any of the strains in the B6/129/NOC/CAST group they
492 were classified as "Ref", homozygous ancestry from any of the AJ/NZO/PWK/WSB strain group they
493 were classified as "Alt", and if they showed ancestry from both groups were classified as "Het". We
494 calculated the partial correlation for all the candidate mediators within ± 10Mb of the Eigen-Chr1
495 QTL peak using both gene expression from DO mESCs and mNPCs.
496
497 **Mediation analysis**

498 We performed mediation analysis to identify candidate causal genes in mNPC eQTL hotspots.
499 Mediation analysis was performed using the R/intermediate package
500 (https://github.com/simecek/intermediate) by regressing target eQTL on the expression of a
501 candidate mediator in the QTL and adjusting for covariates. We applied the 'double-lod-diff'
502 method to reduce the effects of missing values. For mediation of QTL with the matching cell state
503 we used the full sample set, e.g., eQTL mediation by NPC transcripts were done using all the 186
504 samples. Mediation across cell states was performed with mESC expression data from the
505 common set of 127 DO donors. To assess the significance of a LOD drop, we mediated the QTL
506 against all expressed genes genome-wide in that cell state to establish a null distribution of LOD
507 score drops, converted the recorded LOD scores to normal scores, and checked if the score fell
508 below 4 standard deviations from the mean of the null distribution. Mediators were further filtered
509 to narrow down top candidates to include genes with midpoints that are found within ± 10Mb of the
510 QTL peak.
511
512 **Data availability**
513 The raw and processed RNAseq files are deposited to the Sequencing Read Archive and Gene
514 Expression Omnibus (GSE285231).
515

524
525 **Author Notes**
526 **Conflict of Interest** T.C. has an equity interest in Predictive Biology, Inc. All other authors report no
527 conflicts of interest.
528
529 **References**
530
531 Abdellaoui A, Yengo L, Verweij KJH, Visscher PM. 2023. 15 years of GWAS discovery: Realizing the
532 promise. The American Journal of Human Genetics. 110(2):179–194.
533 doi:10.1016/j.ajhg.2022.12.011.

534 Aydin S, Pham DT, Zhang T, Keele GR, Skelly DA, Paulo JA, Pankratz M, Choi T, Gygi SP, Reinholdt LG,
535 et al. 2023. Genetic dissection of the pluripotent proteome through multi-omics data integration.
536 Cell Genomics. 3(4). doi:10.1016/j.xgen.2023.100283. https://doi.org/10.1016/j.xgen.2023.100283.

537 Banovich NE, Li YI, Raj A, Ward MC, Greenside P, Calderon D, Tung PY, Burnett JE, Myrthil M,
538 Thomas SM, et al. 2018. Impact of regulatory variation across human iPSCs and differentiated cells.
539 Genome Res. 28(1):122–131. doi:10.1101/gr.224436.117.

540 Broman KW, Gatti DM, Simecek P, Furlotte NA, Prins P, Sen Ś, Yandell BS, Churchill GA. 2019.
541 R/qtl2: Software for Mapping Quantitative Trait Loci with High-Dimensional Data and Multiparent
542 Populations. Genetics. 211(2):495–502. doi:10.1534/genetics.118.301595.

543 Byers C, Spruce C, Fortin HJ, Hartig EI, Czechanski A, Munger SC, Reinholdt LG, Skelly DA, Baker
544 CL. 2022. Genetic control of the pluripotency epigenome determines differentiation bias in mouse
545 embryonic stem cells. EMBO J. 41(2):e109445. doi:10.15252/embj.2021109445.

546 Chick JM, Munger SC, Simecek P, Huttlin EL, Choi K, Gatti DM, Raghupathy N, Svenson KL, Churchill
547 GA, Gygi SP. 2016. Defining the consequences of genetic variation on a proteome-wide scale.
548 Nature. 534(7608):500–505. doi:10.1038/nature18270.

549 Choi K, He H, Gatti DM, Philip VM, Raghupathy N, Gyuricza IG, Munger SC, Chesler EJ, Churchill GA.
550 2020. Genotype-free individual genome reconstruction of Multiparental Population Models by RNA
551 sequencing data. doi:10.1101/2020.10.11.335323.

552 Churchill GA, Doerge RW. 1994. Empirical threshold values for quantitative trait mapping. Genetics.
553 138(3):963–971. doi:10.1093/genetics/138.3.963.

554 Churchill GA, Gatti DM, Munger SC, Svenson KL. 2012. The diversity outbred mouse population.
555 Mamm Genome. 23(9–10):713–718. doi:10.1007/s00335-012-9414-2.

556 Conti L, Pollard SM, Gorba T, Reitano E, Toselli M, Biella G, Sun Y, Sanzone S, Ying Q-L, Cattaneo E,
557 et al. 2005. Niche-Independent Symmetrical Self-Renewal of a Mammalian Tissue Stem Cell. PLOS
558 Biology. 3(9):e283. doi:10.1371/journal.pbio.0030283.

559 Cuomo ASE, Seaton DD, McCarthy DJ, Martinez I, Bonder MJ, Garcia-Bernardo J, Amatya S, Madrigal
560 P, Isaacson A, Buettner F, et al. 2020. Single-cell RNA-sequencing of differentiating iPS cells reveals
561 dynamic genetic effects on gene expression. Nat Commun. 11(1):810. doi:10.1038/s41467-020-
562 14457-z.

563 Deng L, Jiang C, Chen L, Jin J, Wei J, Zhao L, Chen M, Pan W, Xu Y, Chu H, et al. 2015. The
564 ubiquitination of rag A GTPase by RNF152 negatively regulates mTORC1 activation. Mol Cell.
565 58(5):804–818. doi:10.1016/j.molcel.2015.03.033.

566 Dobles M, Liberal V, Scott ML, Benezra R, Sorger PK. 2000. Chromosome Missegregation and
567 Apoptosis in Mice Lacking the Mitotic Checkpoint Protein Mad2. Cell. 101(6):635–645.
568 doi:10.1016/S0092-8674(00)80875-2.

569 Elorbany R, Popp JM, Rhodes K, Strober BJ, Barr K, Qi G, Gilad Y, Battle A. 2022. Single-cell
570 sequencing reveals lineage-specific dynamic genetic regulation of gene expression during human
571 cardiomyocyte differentiation. PLOS Genetics. 18(1):e1009666. doi:10.1371/journal.pgen.1009666.

572 Farbehi N, Neavin DR, Cuomo ASE, Studer L, MacArthur DG, Powell JE. 2024. Integrating population
573 genetics, stem cell biology and cellular genomics to study complex human diseases. Nat Genet.
574 56(5):758–766. doi:10.1038/s41588-024-01731-9.

575 Gatti DM, Svenson KL, Shabalin A, Wu L-Y, Valdar W, Simecek P, Goodwin N, Cheng R, Pomp D,
576 Palmer A, et al. 2014. Quantitative Trait Locus Mapping Methods for Diversity Outbred Mice. G3
577 Genes|Genomes|Genetics. 4(9):1623–1633. doi:10.1534/g3.114.013748.

578   Jerber J, Seaton DD, Cuomo ASE, Kumasaka N, Haldane J, Steer J, Patel M, Pearce D, Andersson M,
579   Bonder MJ, et al. 2021. Population-scale single-cell RNA-seq profiling across dopaminergic neuron
580   differentiation. Nat Genet. 53(3):304–312. doi:10.1038/s41588-021-00801-6.

581   Johnson WE, Li C, Rabinovic A. 2007. Adjusting batch effects in microarray expression data using
582   empirical Bayes methods. Biostatistics. 8(1):118–127. doi:10.1093/biostatistics/kxj037.

583   Kilpinen H, Goncalves A, Leha A, Afzal V, Alasoo K, Ashford S, Bala S, Bensaddek D, Casale FP,
584   Culley OJ, et al. 2017. Common genetic variation drives molecular heterogeneity in human iPSCs.
585   Nature. 546(7658):370–375. doi:10.1038/nature22403.

586   Kim S. 2015. ppcor: An R Package for a Fast Calculation to Semi-partial Correlation Coefficients.
587   Communications for Statistical Applications and Methods. 22(6):665–674.
588   doi:10.5351/CSAM.2015.22.6.665.

589   Kolberg L, Raudvere U, Kuzmin I, Vilo J, Peterson H. 2020. gprofiler2 -- an R package for gene list
590   functional enrichment analysis and namespace conversion toolset g:Profiler. F1000Res. 9:ELIXIR-
591   709. doi:10.12688/f1000research.24956.2.

592   Korotkevich G, Sukhov V, Budin N, Shpak B, Artyomov MN, Sergushichev A. 2021. Fast gene set
593   enrichment analysis. doi:10.1101/060012.

594   Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of
595   short DNA sequences to the human genome. Genome Biol. 10(3):R25. doi:10.1186/gb-2009-10-3-
596   r25.

597   Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, Morgan MT, Carey VJ. 2013.
598   Software for Computing and Annotating Genomic Ranges. PLOS Computational Biology.
599   9(8):e1003118. doi:10.1371/journal.pcbi.1003118.

600   Lee H, Aylward AJ, Pearse RV, Lish AM, Hsieh Y-C, Augur ZM, Benoit CR, Chou V, Knupp A, Pan C, et
601   al. 2023. Cell-type-specific regulation of APOE and CLU levels in human neurons by the Alzheimer's
602   disease risk gene SORL1. Cell Rep. 42(8):112994. doi:10.1016/j.celrep.2023.112994.

603   Lienert F, Mohn F, Tiwari VK, Baubec T, Roloff TC, Gaidatzis D, Stadler MB, Schübeler D. 2011.
604   Genomic prevalence of heterochromatic H3K9me2 and transcription do not discriminate
605   pluripotent from terminally differentiated cells. PLoS Genet. 7(6):e1002090.
606   doi:10.1371/journal.pgen.1002090.

607   Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu
608   H, Brody J, et al. 2012. Systematic localization of common disease-associated variation in
609   regulatory DNA. Science. 337(6099):1190–1195. doi:10.1126/science.1222794.

610   Maydan G, Noyman I, Har-Zahav A, Neriah ZB, Pasmanik-Chor M, Yeheskel A, Albin-Kaplanski A,
611   Maya I, Magal N, Birk E, et al. 2011. Multiple congenital anomalies-hypotonia-seizures syndrome is
612   caused by a mutation in PIGN. J Med Genet. 48(6):383–389. doi:10.1136/jmg.2010.087114.

613   Mirauta BA, Seaton DD, Bensaddek D, Brenes A, Bonder MJ, Kilpinen H, HipSci Consortium, Agu
614   CA, Alderton A, Danecek P, et al. 2020. Population-scale proteome variation in human induced
615   pluripotent stem cells. eLife. 9:e57390. doi:10.7554/eLife.57390.

616   Okamoto T, Imaizumi K, Kaneko M. 2020. The Role of Tissue-Specific Ubiquitin Ligases, RNF183,
617   RNF186, RNF182 and RNF152, in Disease and Biological Function. International Journal of
618   Molecular Sciences. 21(11):3921. doi:10.3390/ijms21113921.

619   Okur V, Cho MT, Henderson L, Retterer K, Schneider M, Sattler S, Niyazov D, Azage M, Smith S,
620   Picker J, et al. 2016. De novo mutations in CSNK2A1 are associated with neurodevelopmental
621   abnormalities and dysmorphic features. Hum Genet. 135(7):699–705. doi:10.1007/s00439-016-
622   1661-y.

623   Panopoulos AD, D'Antonio M, Benaglio P, Williams R, Hashem SI, Schuldt BM, DeBoever C, Arias
624   AD, Garcia M, Nelson BC, et al. 2017. iPSCORE: A Resource of 222 iPSC Lines Enabling Functional
625   Characterization of Genetic Variation across a Variety of Cell Types. Stem Cell Reports. 8(4):1086–
626   1100. doi:10.1016/j.stemcr.2017.03.012.

627   Paolo GD, Sankaranarayanan S, Wenk MR, Daniell L, Perucco E, Caldarone BJ, Flavell R, Picciotto
628   MR, Ryan TA, Cremona O, et al. 2002. Decreased Synaptic Vesicle Recycling Efficiency and
629   Cognitive Deficits in Amphiphysin 1 Knockout Mice. Neuron. 33(5):789–804. doi:10.1016/S0896-
630   6273(02)00601-3.

631   Parikh C, Glenn R, Shi Y, Chatterjee K, Swanzey E, Singer S, Do S, Zhan Y, Furuta Y, Tahiliani M, et al.
632   2024. Genetic variation modulates susceptibility to aberrant DNA hypomethylation and imprint
633   deregulation in naive pluripotent stem cells. doi:10.1101/2024.06.26.600805.

634   Pollard SM, Benchoua A, Lowell S. 2006. Neural Stem Cells, Neurons, and Glia. In: Methods in
635   Enzymology. Vol. 418. Academic Press. (Embryonic Stem Cells). p. 151–169.
636   https://www.sciencedirect.com/science/article/pii/S0076687906180106.

637   Popp JM, Rhodes K, Jangi R, Li M, Barr K, Tayeb K, Battle A, Gilad Y. 2024. Cell-type and dynamic
638   state govern genetic regulation of gene expression in heterogeneous differentiating cultures.
639   doi:10.1101/2024.05.02.592174. https://www.biorxiv.org/content/10.1101/2024.05.02.592174v1.

640   R Core Team. 2023. R: A Language and Environment for Statistical Computing. https://www.R-
641   project.org/.

642   Raghupathy N, Choi K, Vincent MJ, Beane GL, Sheppard KS, Munger SC, Korstanje R, Pardo-Manual
643   de Villena F, Churchill GA. 2018. Hierarchical analysis of RNA-seq reads improves the accuracy of
644   allele-specific expression. Valencia A, editor. Bioinformatics. 34(13):2177–2184.
645   doi:10.1093/bioinformatics/bty078.

646   Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H, Vilo J. 2019. g:Profiler: a web server for
647   functional enrichment analysis and conversions of gene lists (2019 update). Nucleic Acids
648   Research. 47(W1):W191–W198. doi:10.1093/nar/gkz369.

649   Skelly DA, Czechanski A, Byers C, Aydin S, Spruce C, Olivier C, Choi K, Gatti DM, Raghupathy N,
650   Keele GR, et al. 2020. Mapping the Effects of Genetic Variation on Chromatin State and Gene

651 Expression Reveals Loci That Control Ground State Pluripotency. Cell Stem Cell. 27(3):459-469.e8.
652 doi:10.1016/j.stem.2020.07.005.

653 Storey JD, Taylor JE, Siegmund D. 2004. Strong Control, Conservative Point Estimation and
654 Simultaneous Conservative Consistency of False Discovery Rates: A Unified Approach. Journal of
655 the Royal Statistical Society Series B (Statistical Methodology). 66(1):187–205.

656 Swanzey E, O'Connor C, Reinholdt LG. 2021. Mouse Genetic Reference Populations: Cellular
657 Platforms for Integrative Systems Genetics. Trends in Genetics. 37(3):251–265.
658 doi:10.1016/j.tig.2020.09.007.

659 Teye EK, Lu S, Chen F, Yang W, Abraham T, Stairs DB, Wang H-G, Yochum GS, Brodsky RA, Pu JJ.
660 2021. PIGN spatiotemporally regulates the spindle assembly checkpoint proteins in leukemia
661 transformation and progression. Sci Rep. 11:19022. doi:10.1038/s41598-021-98218-y.

662 Vernay B, Koch M, Vaccarino F, Briscoe J, Simeone A, Kageyama R, Ang S-L. 2005. Otx2 regulates
663 subtype specification and neurogenesis in the midbrain. J Neurosci. 25(19):4856–4867.
664 doi:10.1523/JNEUROSCI.5158-04.2005.

665 Warren CR, Cowan CA. 2018. Humanity in a Dish: Population Genetics with iPSCs. Trends in Cell
666 Biology. 28(1):46–57. doi:10.1016/j.tcb.2017.09.006.

667 Wells MF, Nemesh J, Ghosh S, Mitchell JM, Salick MR, Mello CJ, Meyer D, Pietilainen O, Piccioni F,
668 Guss EJ, et al. 2023. Natural variation in gene expression and viral susceptibility revealed by neural
669 progenitor cell villages. Cell Stem Cell. 30(3):312-332.e13. doi:10.1016/j.stem.2023.01.010.

670 Yang J, Zaitlen NA, Goddard ME, Visscher PM, Price AL. 2014. Advantages and pitfalls in the
671 application of mixed-model association methods. Nat Genet. 46(2):100–106. doi:10.1038/ng.2876.

672 Zhang X, Zou L, Meng L, Xiong M, Pan L, Chen G, Zheng Y, Xiong J, Wang Z, Duong DM, et al. 2021.
673 Amphiphysin I cleavage by asparagine endopeptidase leads to tau hyperphosphorylation and
674 synaptic dysfunction. Elife. 10:e65301. doi:10.7554/eLife.65301.

675

676 **Supplemental Figures**
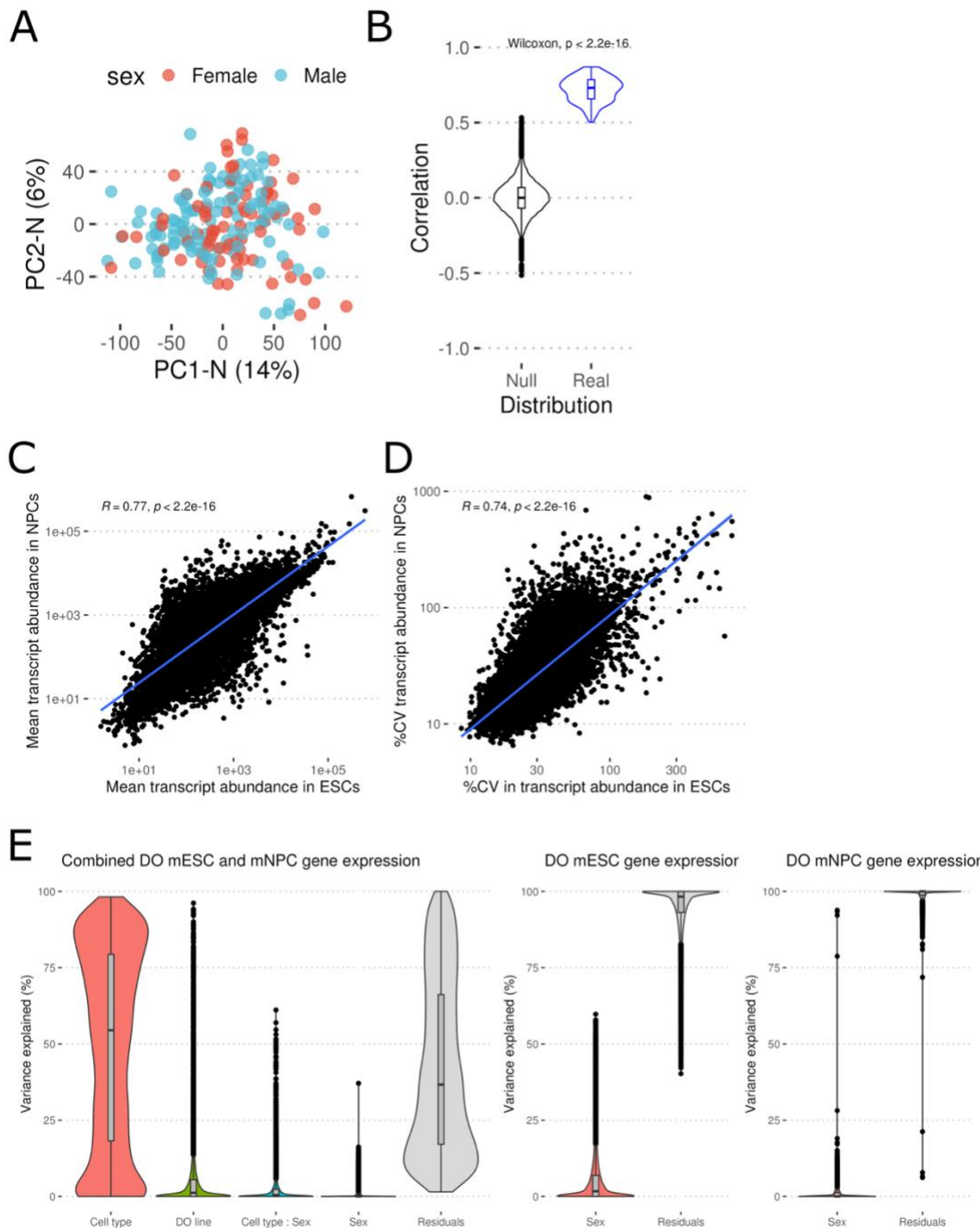
677
678 **Figure S1.** (A) Principal component analysis results of the DO mNPC data. (B) Genetically identical cell lines
679 show significantly higher correlation than what is expected by chance between the ESC and NPC
680 transcriptomes. Violin plots overlaid with boxplots depicting the distribution of Pearson correlation
681 coefficients between the transcriptomes of 127 genetically identical mESC and mNPCs (blue) and the null
682 distribution generated through 1000 permutations where the sample names are randomized (black). (C-D)
683 Scatterplots showing mean and coefficient of variation (% CV) for transcript abundance for genes with
684 measurements in both mESCs and mNPCs. (E) Violin plots overlaid with boxplots showing the variance
685 partition analysis results in the combined and individual DO mESC, mNPC transcriptomes.

686
687
688
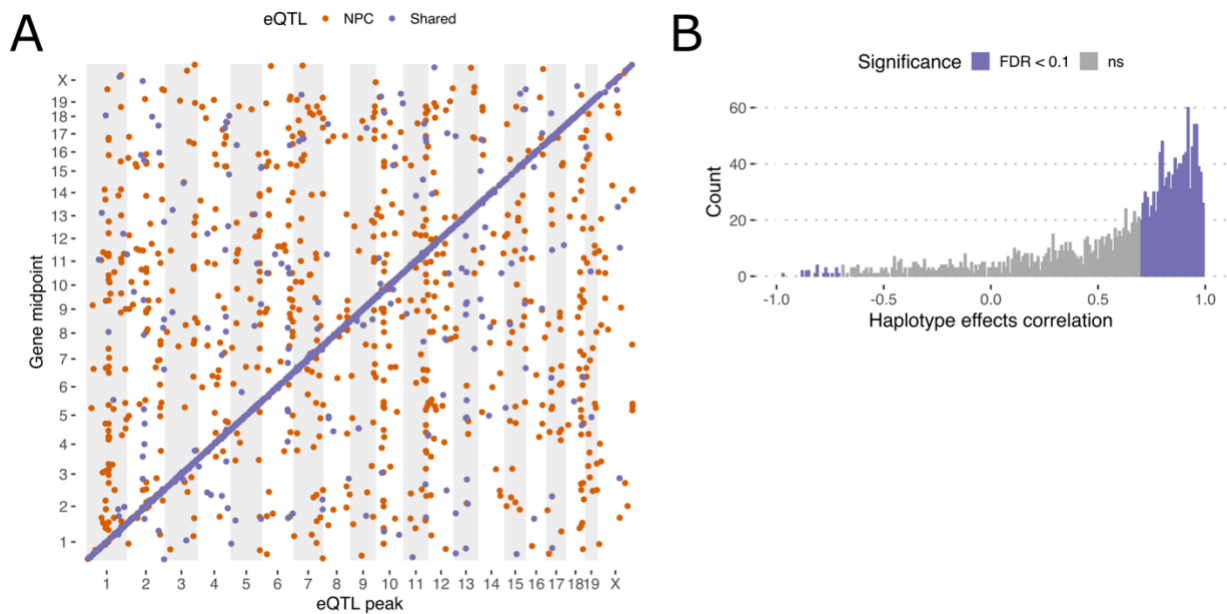
689
690 **Figure S2.** (A) Genetic mapping identifies 2,143 eQTL unique to NPCs and 2,529 eQTL shared between ESC
691 and NPCs. The location of the eQTL is plotted on the x axis against the midpoint of the gene on the y axis. (B)
692 Most co-mapping eQTL show high agreement in their haplotype effects. Histogram of pairwise correlation
693 coefficients between inferred allele effects from eQTL from ESC and NPC scans for all genes with co-
694 mapping QTLs. Bars are colored by significance of the correlation.

695
696
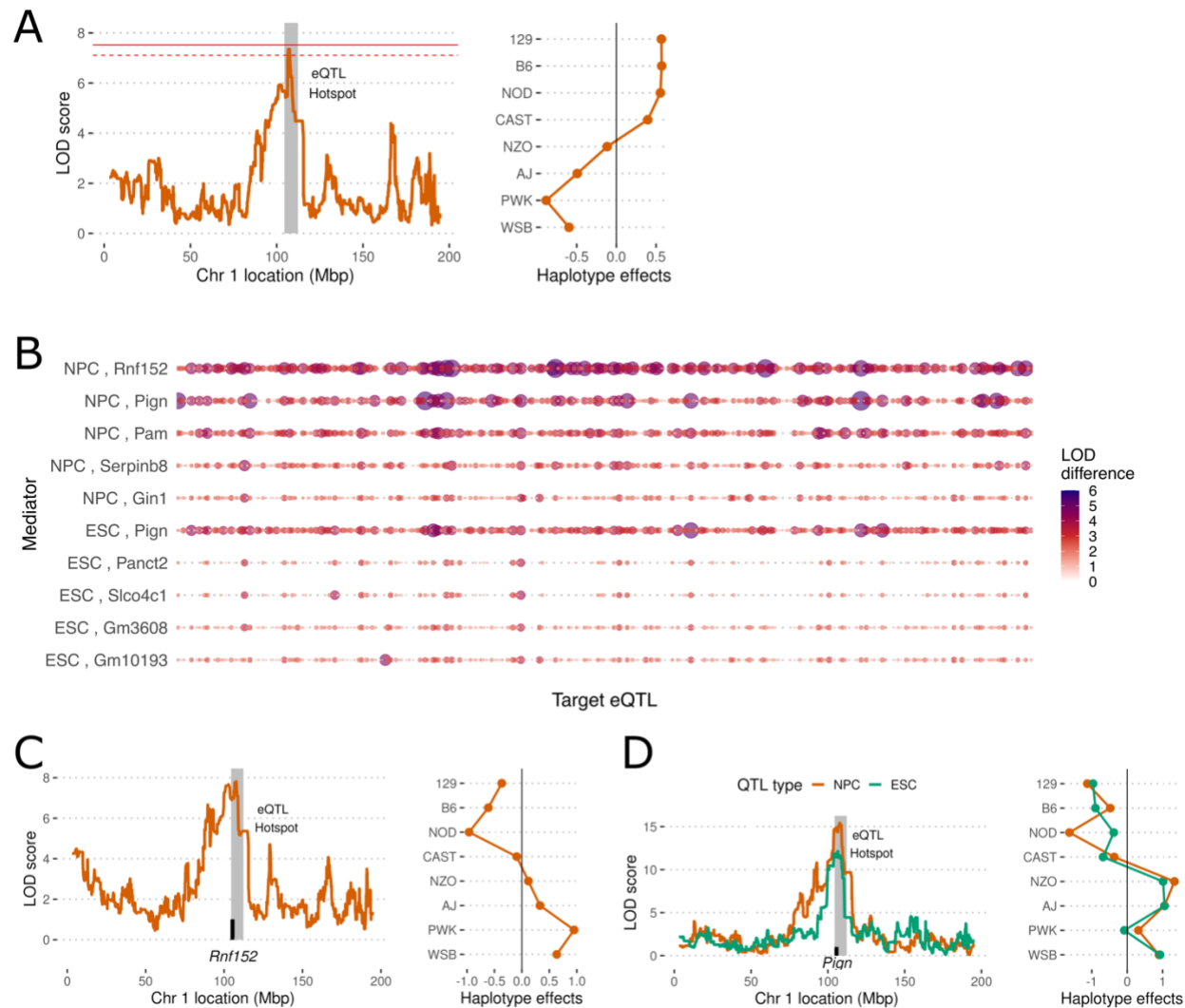
697
698 **Figure S3.** (A) Genetic mapping with PC1-N identifies a significant QTL on chromosome 1 with similar
699 haplotype effects to the eQTL hotspot on chromosome 1. The red lines correspond to LOD thresholds for
700 alpha = 0.05 (solid) and alpha = 0.1 (dashed). On the right, inferred haplotype effects at the QTL peak is
701 plotted. (B) drops obtained from mediation analysis using NPC and ESC expression is plotted for the
702 chromosome 1 target eQTL, showing results for the 5 best mediator genes within 10Mb of the eQTL hotspot.
703 (C) Genome scan of *Rnf152* expression in NPCs show a significant local eQTL peak on chromosome 1 with
704 inferred haplotype effects at the peak plotted on the right. (D) Genome scan of *Pign* expression in ESCs
705 (green) and NPCs (orange) show a significant local eQTL peak on chromosome 1. The inferred haplotype
706 effects at the peak are plotted on the right.

707
708
709
710
711
712
713
714

Supplemental Tables

Table S1. List of biological pathways and processes over-represented in DO mNPCs PCA PC1-N.

Table S2. List of biological pathways and processes over-represented in PC1-C, PC2-C and PC3-C drivers from the combined data PCA.

Table S3. List of biological pathways and processes enriched in genes differentially expressed between DO mNPCs and mESCs.

Table S4. List of all significant DO mNPC eQTL.

Table S5. Details of DO mNPC eQTL hotspots and target genes.

Table S6. List of biological pathways and processes over-represented in shared and unique DO mNPC eQTL.

Table S7. Biological pathways and processes over-represented in target genes of chromosome 1 and 10 eQTL hotspots.