Loss of synteny to conserved noncoding elements impact the fetus-specific gene expression in mammalian brain

Meenakshi Bagadia[§], Keerthivasan Raanin Chandradoss[§], Yachna Jain, Harpreet Singh, Mohan Lal, Kuljeet Singh Sandhu*

Department of Biological Sciences
Indian Institute of Science Education and Research (IISER) - Mohali
Knowledge City, Sector - 81, SAS Nagar 140306, India

[§]*Equal contribution*
**To whom correspondence should be addressed*
Kuljeet Singh Sandhu
Assistant Professor
Department of Biological Sciences
Indian Institute of Science Education and Research (IISER) - Mohali
E. mail: sandhuks@iisermohali.ac.in

Abstract

Conserved noncoding elements (CNEs) have significant regulatory influence on their neighbouring genes. Loss of synteny to CNEs through genomic rearrangements can, therefore, impact the transcriptional states of the cognate genes. Yet, the evolutionary implications of such chromosomal position effects have not been studied. Through genome-wide analysis of CNEs and the cognate genes of representative species from 5 different mammalian orders, we observed significant loss of synteny to CNEs in rat lineage. The CNEs and genes losing synteny had significant association with the fetal, but not the post-natal, brain development as assessed through ontology terms, developmental gene expression, chromatin marks and genetic mutations. The loss of synteny correlated with the independent evolutionary loss of fetus-specific upregulation of genes in rat brain. DNA-breakpoints implicated in brain abnormalities of germline origin had significant representation between CNE and the gene that exhibited loss of synteny, signifying the underlying developmental tolerance of genomic rearrangements that had allowed the evolutionary splits of CNEs and the cognate genes in rodent lineage. These observations highlighted the non-trivial impact of chromosomal position-effect in shaping the evolutionary dynamics of mammalian brain development and might explain loss of brain traits, like cerebral folding of cortex, in rodent lineage.

Author Summary

Expression of genes is regulated by proximally located non-coding regulatory elements. Loss of linear proximity between gene and its regulatory element thus can alter the expression of gene. Such a phenomenon can be tested at whole genome scale using evolutionary methods. We compared the positions of genes and regulatory elements in 5 different mammals and identified the significant loss of proximities between gene and their regulatory elements in rat during evolution. Brain development related function was selectively enriched among the genes and regulatory elements that had lost the proximity in rat. The observed separation of genes and their regulatory elements was strongly associated with the evolutionary loss of developmental gene expression pattern in rat brain, which coincided with the loss of brain traits in rodents. The study highlighted the importance of relative chromosomal positioning of genes and their gene regulatory elements in the evolution of phenotypes.

Keywords: conserved non-coding elements, enhancer, genome organization, synteny, evolution, position-effect.

Introduction

Around 4-8% of the human genome is evolutionary constrained, of which coding elements contribute only about 1.5%, while rest is non-coding [1-3]. Massive data produced by ENCODE and Epigenome Roadmap projects have confirmed that majority of the evolutionary constrained non-coding DNA serve as protein binding sites[4, 5]. These conserved noncoding elements (CNEs) are interwoven with the protein coding genes in a complex manner. Ample evidence converges to non-trivial regulatory impact of CNEs on proximal gene. Deletion of a non-coding region between sclerostin (SOST) gene, a negative regulator of bone formation, and MEOX1 impacts the expression of SOST and is strongly associated with Van Buchmen disease characterized by progressive overgrowth of bones [6]. Similarly, deletion of a 10kb non-coding region downstream to stature homeobox (SHOX) gene is associated with Leri Weill dyschondrosteosis syndrome, a skeletal dysplasias condition [7]. Mutations in CNEs downstream to PAX6 gene prevent its expression and are associated with Aniridia, a congenital eye malformation. Genetic errors in locus control region (LCR) at alpha and beta globin loci strongly associate with alpha/beta-thalassemia [8, 9]. Maternal deletion of Igf2/H19 ICR disrupts the Igf2 imprinting leading to bi-allelic expression of Igf2, which is strongly associated with Beckwith Weidman syndrome [10]. Loss of a CNE proximal to androgen receptor is strongly associated with evolutionary loss of penile spines and sensory vibrissae in human [11].

Around 200,000 human-anchored Conserved Non-coding Elements (CNEs) have been identified in mammals, which are likely to exhibit gene regulatory potential, as measured through enhancer-associated chromatin marks [12-14]. Most CNEs position around developmental genes [14-16]. However, establishing causal relationship between CNE and the phenotype remains a daunting task. Though genome wide association studies (GWAS) have uncovered a whole repertoire of non-coding variants with phenotypic associations [17], it is difficult to identify the causal variants. More recently, pooled CRISPR-Cas technique has been implemented to alter the non-coding elements to assess their function more precisely [18]. These methods are difficult to be scaled up for high throughput genotype-phenotype associations. With the availability of whole genome sequences of multiple species, evolutionary methods are instrumental in deciphering genotype-phenotype associations. Through comprehensive multi-species comparison, it has been inferred that most CNEs are syntenic to the nearest gene in linear proximity and are likely to regulate the same [14, 19]. Attempts have been made to link evolutionary loss and sequence divergence of CNEs to lineage specific traits, like auditory system in echo-locating mammals and adaptively morphed pectoral flippers in marine mammals [20, 21]. In this study, we asked the question whether the lineage-specific evolutionary alterations in relative chromosomal positions of CNEs are associated with lineage-specific changes in

101 gene expression. Through analysis of chromosomal positions of orthologous CNEs and genes from 5
102 different mammals, we observed that a significant number of genes had lost synteny to their
103 adjacent CNEs independently in rat lineage. This loss of synteny was significantly associated with the
104 down-regulation of genes involved in neuorogenesis and neuronal migration during fetal brain
105 development, which coincided with the evolutionary loss of several brain traits independently in rat
106 lineage. The study suggested significant contribution of chromosomal position-effect in the
107 evolutionary divergence of developmental gene expression trajectories in mammals.
108

109 **Results**

110 **Loss of synteny between CNE and the proximal gene**

111 Using chromosomal position data of CNEs and genes from representative primate (human), rodent
112 (rat), carnivore (dog), perrisodactyl (horse), and artiodactyls (cow), we obtained 51434 'syntenic'
113 CNE-gene pairs (4241 genes), wherein the CNE and the nearest gene-TSS were <1 Mb distance apart
114 in all 5 species. There were 3579 'non-syntenic' CNE-gene pairs (334 genes), wherein the CNE and
115 the gene-TSS were on different chromosomes or were >2Mb apart independently in one of the
116 species (Figure 1A, Figure S1, Methods). The rationale of 1Mb distance cutoff for the synteny was
117 based on the observation that the distribution of all CNE-gene distances saturated when approached
118 1Mb range (Figure 1B). Similar approach has been used earlier to infer enhancer-promoter linkage
119 based on evolutionary syntney between the two in 1Mb range[19]. The distance cut-off of 2Mb for
120 loss of synteny made sure that the minimal expansion in CNE-gene distance would at-least be 2 fold.
121 To test if the CNEs in syntenic and non-syntenic sets were comparable, we assessed their lengths
122 and degree of conservation in mammalian genomes. Figure 1C showed insignificant differences in
123 the degree of sequence conservation of syntenic and non-syntenic CNEs, suggesting that the
124 sequence of non-syntenic CNEs had not diverged among mammals as compared to that of syntenic
125 CNEs. The length distribution of syntenic and non-syntenic CNEs showed only marginal difference
126 towards slightly longer CNEs in non-sytenic set (Figure 1D). However, the syntenic and non-syntenic
127 CNEs were located in the genomic domains of distinct sequence properties. We analysed the
128 enrichment of SINE, LINE and LTR retrotransposons, which covers upto 50% of mammalian genome,
129 around syntenic and non-syntenic CNEs. Syntenic CNEs were enriched in the region of open
130 chromatin, as signified through greater enrichment SINE content, and might have more wide-spread
131 role across different cell-lineages as compared with the non-syntenic CNEs (Figure 1E). In contrast,
132 the non-syntneic CNEs were located in the domains enriched with long terminal repeats (LTR),
133 marking their susceptibility to genomic rearrangements through mechanisms like non-allelic
134 homologous recombination (NAHR) [22, 23] (Figure 1E). LINE elements, in general, did not exhibit

135 significant difference in the two sets. These observations were largely consistent across species,
136 marking an ancestral property (Figure 1E), except in the rat wherein the LINE elements were
137 enriched around non-syntenic CNEs. This exception can be explained by the fact that the rodents
138 retained least of the ancestral LINE elements (1%) as compared to most other mammals (7-8%) and
139 had accumulated new LINE elements[24].
140
141 Relatively large number of syntenic CNE-gene pairs (93.5%) confirmed the widespread conservation
142 of linear proximity between CNE and its adjacent gene [14]. Among the total 3579 non-syntenic
143 instances, 2711 (75%) were associated with the rat genome alone, coherent with the significantly
144 greater number of structural variations in rodent clade[25] (Figure 2A-B). Positive scaling between
145 number of non-syntenic instances and the break-point distances of species from the common
146 ancestor signified that CNE-gene synteny was an ancestral trait (Figure 2B). Due to significant loss of
147 synteny in rat lineage as compared to others, we focussed on rat instances in this study. By 'Loss of
148 synteny' or 'non-syntenic' set, we referred to loss of CNE-gene synteny in rat from figure 2C
149 onwards. To directly assess the proportion of non-syntenic CNE-gene pairs associated with structural
150 variations, we analysed the rodent-specific evolutionary break-points (Methods). We observed that
151 930 (34%) of all non-syntenic instances in rat had at-least one rodent-specific break-point in
152 between the gene-TSS and the CNE as compared to 319 (11%) on an average for the random null
153 prepared through distance-controlled bootstrap sampling of syntenic CNE-gene pairs (Figure 2C,
154 Methods). This suggested that the loss of CNE-gene synteny in rat could largely be explained through
155 rodent-specific genomic rearrangements (Figure 2C). We further argue that the sequence alignment
156 based annotations of evolutionary breakpoints might not represent the entire repertoire of genomic
157 rearrangements and, therefore, analysed the neighbouring genes on either side of non-syntenic
158 CNEs to map the various rearrangement scenario through which CNE-gene synteny was lost. We
159 found that that the translocation like scenario, as marked by (i) in Figure 2D, largely explained the
160 inter-chromosomal (*trans*) splits of CNE and the adjacent gene. The scenario, which reflected the
161 mapping artefacts, like in panel (iii), was under-represented (5%, Figure 2D). Analysis of intra-
162 chromosomal (*cis*) splits suggested inversion-like events separating the CNE-gene pairs. Scenario (iv)
163 and (v) showed events where region adjacent to CNE (on left side in scenario-iv and right side in
164 scenario-v) had undergone local rearrangements, of which 30% and 90% events respectively were
165 confirmed as inversion events by analysing the change in relative strand orientation of neighboring
166 genes. We illustrated examples of *trans* and *cis* splits of CNE and the genes in Figure 2E. Gene
167 POU3F2 on human chromosome 6 was syntenic to a CNE, which was 45Kb upstream. The
168 orthologous CNE and the gene in rat were on chromosome 8 and 5 respectively marking the *trans*

169 split of CNE and the gene through translocation (Figure 2E). Another CNE was 18kb upstream to
170 gene ADAM23 on human chromosome 2. The rat orthologues were separated by 2.4Mb on
171 chromosome 9 through an inversion (Figure 2E).
172
173 We concluded that the rodent-specific genomic rearrangements largely explained the loss of CNE-
174 gene synteny in rat.
175
176 **Genes that had lost synteny to CNEs in rat were associated with the fetal brain development**
177 Significant differences in the genomic attributes around syntenic and non-syntenic CNEs hinted at
178 their distinct functional roles. To assess their functions, syntenic and non-syntenic gene-lists were
179 subjected to Gene Ontology (GO) and Mammalian Phenotype Ontology (MPO) analyses. The analysis
180 of GO terms revealed enrichment of general as well as various tissue-specific development related
181 terms in the syntenic set, while non-syntenic set was specifically enriched with nervous system
182 development related terms (Figure 3A). In MPO analysis, syntenic set exhibited of enrichment of
183 neonatal lethality and skeletal phenotypes, while non-syntenic set was associated with brain
184 morphology related phenotypes (Figure S2A). Species, other than rat, did not exhibit enrichment of
185 any particular functional term, owing to smaller sample size.
186
187 We further followed the above observations through tissue-specific gene expression analysis in
188 human. The syntenic set had widespread representation of genes expressed in different cell-lineages
189 and, therefore, did not exhibit significant tissue-specificity, while genes in non-syntenic set were
190 specifically expressed in brain (Figures 3B). The brain-specific expression of genes in non-syntenic
191 set was also confirmed through enrichment analysis of anatomical terms from *bgee* database (Figure
192 S2B). Within brain, non-syntenic set was enriched with the genes specifically expressed in cerebral
193 cortex during fetal, but not post-natal development (Figure 3C). In contrast, the genes in syntenic set
194 did not exhibit any specificity for brain tissues and developmental stages (Figures 3C). These
195 observations highlighted fetal brain-specific roles of genes in the non-syntenic set.
196
197 **Non-syntenic CNEs function as fetal brain-specific enhancers**
198 To test whether the differences between syntenic and non-syntenic sets observed through
199 functional analysis of genes, were coherent with the associated CNEs, we tested the regulatory
200 potential of CNEs by analyzing their epigenomic properties across tissues. Through analysis of
201 enhancer associated chromatin state annotations from Epigenome Roadmap, ENCODE and Fantom
202 consortia (Methods), we observed that 74% of syntenic and 61% of non-syntenic CNEs overlapped

203 with the enhancer-associated regulatory sites in at-least one of the tissues or cell-types, marking the
204 enhancer potential of CNEs. Relatively less representation of enhancers in the non-syntenic set
205 might relate to their tissue or developmental stage specific functions, a hypothesis that we further
206 reconciled through detailed analysis of histone modification associated with enhancers, namely
207 Histone-3-Lysine-4-mono-methylation (H3K4me1). We chose this mark because of its strong
208 association with the enhancer potential and the availability of genome-wide datasets for all the cell-
209 lineages we were interested in. We observed that: 1) the CNEs in syntenic set exhibited consistent
210 H3K4me1 enrichment across several fetal and adult tissues like thymus (endodermal), muscle
211 (mesodermal), heart (mesodermal), intestine(endodermal) and brain (ectodermal) (Figure 4A, Figure
212 S3A-B); 2) H3K4me1 enrichment over CNEs in non-syntenic set was specifically higher (comparable
213 to that of syntenic CNEs) in fetal, but not in adult brain (Figure 4A). We further observed the
214 significant enrichment of binding sites of ectoderm-specific transcription factors, which were
215 specifically upregulated in fetal brain, in non-syntenic CNEs as compared to syntenic CNEs (Figure S4,
216 Methods). These observations were largely coherent with our proposal that non-syntenic CNE-gene
217 pairs were associated with fetal brain development.
218
219 To assess the physical enhancer-promoter association, we generated virtual 4C data by processing
220 available HiC datasets of fetal and adult human brains (Figure S3C). Figure 4B showed the significant
221 fetal-to-adult ratio of the proportion of non-syntenic CNE-gene pairs showing significant physical
222 interactions as compared to that of syntenic CNE-gene pairs. We illustrated the physical interactions
223 between CNE and the gene through examples (Figure 4C, Figure S3D). Transcription start sites of
224 GPR85 and FEZF2 genes, both associated with neurological phenotypes[26-28], showed significant
225 interaction frequency to their cognate CNEs in human fetal brain, but not in adult brain. The
226 H3K4me1 signals at TSSs and CNEs were also significant in fetal brain as compared to adult.
227 Epigenomic analyses thus suggested that the majority of the non-syntenic CNEs exhibited enhancer
228 associated hallmarks in fetal brain.
229
230 By mapping the trait/disease associated SNPs from Genome Wide Associated Studies (GWAS) and
231 the nearby SNPs (proxy) in the linkage disequilibrium based on 1000 genome data, we observed that
232 105 of the non-syntenic CNEs were having at-least one brain related SNP (Figure 4D). This
233 representation was statistically significant when compared with that of syntenic set (Figure 4D).
234 These observations represented genetic evidence of brain-specific roles of non-syntenic CNEs. We
235 highlighted the example of EPHA4 gene, which is required for radial neuron migration and is
236 involved in the pathways leading to lissencephaly and schizophrenia in human[29, 30]. Upstream

237 CNE to this gene had a schizophrenia associated SNP. Fetal brain specificity of CNE and gene
238 expression was illustrated using H3K4me1 and RNA-seq tracks of fetal and post-natal human brains
239 (Figure 4E).
240
241 Therefore, our observations through enhancer datasets, epigenomic marks, differential motif
242 enrichment analysis and brain associated SNPs concomitantly established that the non-syntenic
243 CNEs were specific to fetal brain development in human.
244
245 **Developmental tolerance loss of synteny events**
246 While we have shown that the genes and the CNEs that had lost synteny in rat were associated with
247 fetal brain development, whether or not CNE-gene proximity was causally linked with the brain-
248 specific expression of the cognate gene remained to be addressed. Towards this, we assessed the
249 representation of germ-line breakpoints associated with the congenital disorders exhibiting brain
250 abnormalities and the somatic cancer breakpoints, between CNE and gene-TSS in syntenic and non-
251 syntenic sets. Since the observed germ-line break-points are the ones that had survived through
252 germ-line and the embryonic development, their presence and absence between CNE and the
253 adjacent gene signifies developmental tolerance and intolerance respectively of loss of CNE-gene
254 synteny. On the contrary, the cancer breakpoints of somatic origin do not undergo such selection
255 and hence do not indicate the developmental tolerance or lack thereof. Figure 5A showed relative
256 proportion of non-syntenic CNE-gene pairs having at-least one DNA breakpoint between gene-TSS
257 and the CNE superimposed onto random null prepared from syntenic CNE-gene pairs of similar
258 distance distribution as that of non-syntenic. We observed significantly greater representation of
259 germ-line breakpoints in non-syntenic set as compared to syntenic set, while representation of
260 somatic breakpoints showed insignificant difference (Figure 5A). We interpreted that DNA
261 breakpoints between non-syntenic CNE and the genes were developmentally tolerable and genomic
262 rearrangements thereof in the ancestral genome might have served as an evolutionary substrate for
263 position effect. We further showed a few examples of germ-line chromosomal rearrangements that
264 had split the CNE and the adjacent gene in congenital disorders with brain abnormalities (Figure 5B).
265 Example (i) in Figure 5B showed a chromothripsis event wherein an inversion had split the CNE-gene
266 pair. The involved gene BCL11A regulates cortical neuron migration and mutations therein associate
267 with microcephaly and intellectual disability in human[31]. BCL11A gene also exhibited 3.6 fold loss
268 of expression in peripheral blood of the patient having genomic rearrangement as compared with
269 the normal mother of the patient. Example (ii) showed a translocation event splitting a CNE and
270 CCDC68 gene. Genetic mutations in CCDC68 are associated with schizophrenia, bipolar disorder and

271 autism[32]. In example (iii), an inversion had split the CNE from DNAJB6 gene, which has role in
272 neuritogenesis and neuroprotective functions[33].
273
274 **Loss of synteny to CNEs coincided with the loss of fetus-specific upregulation of genes in rat brain**
275 An important question was whether or not the evolutionary loss of CNE-gene synteny in rat was
276 associated with the loss of expression. To assess the functional fate of associated genes, we
277 compared their time-course gene expression trajectories for developing cerebral cortex of human,
278 rat and sheep (as an out-group). Sheep was inducted in the analysis due to the availability of gene
279 expression datasets for pre- and post-natal tissues. We found that 99.4% of CNE-gene pairs that had
280 lost synteny in rat were syntenic in sheep too, confirming the independent loss of synteny in rat
281 lineage. We observed relative loss of fetus-specific gene expression in rat brain as compared to that
282 of human and sheep, suggesting that the loss of synteny correlates with the loss of fetus-specific
283 gene expression in developing rat brain (Figure 6). Enrichment of neurogenesis related genes and
284 down-regulation thereof in fetal brain of rat has implications in understanding loss of brain traits in
285 rat lineage, as discussed in detail in the discussion section.
286
287 Taken together, our analysis suggested a strong association between evolutionary dynamics of
288 chromosomal positions of gene regulatory elements and the gain or loss of gene expression, aligning
289 to the notion of 'position effect'. Tissue and developmental stage specific impact of position effect
290 highlighted the possibility of its significant role in altering developmental dynamics towards
291 evolutionary gain or loss of lineage-specific traits.
292
293 **Discussion**
294 It is not always the change in number and the sequence of protein coding regions in the genome
295 that leads to the phenotype alternation in evolution, the dynamics of gene expression is equally
296 relevant in the context. One way the gene expression is altered is through position-effect, i.e.,
297 relative chromosomal position of the gene in the genome can alter its expression through regulatory
298 elements and chromatin states in the neighbourhood. Position effect was first discovered through
299 the observation that the chromosomal arrangement of duplicated copies of *bar* gene in *bar*-mutant
300 flies had influence on its expression and consequently causes the relative decrease in number of eye
301 facets [34, 35]. Similarly, *white* gene when localized near heterochromatin gives mottled eye
302 phenotype with red and white patches in drosophila eye [36]. Despite its significance, the role of
303 position effect in evolution of traits has not been investigated thoroughly. Through comparative
304 genomic analysis, we showed that the CNE-gene pairs that were syntenic in most mammals, but lost

305 the close linear proximity independently in rat were associated with the alteration in the
306 transcriptional program during fetal brain development, presenting evidence how the position-effect
307 might have impacted the evolution of lineage-specific phenotypes by modulating the developmental
308 trajectories in early stages.
309
310 Enhancers can function at distance longer than several Mbs and spatial synteny has been observed
311 among genomic regions that had been rearranged in the evolution[37, 38]. How might then the loss
312 of linear proximity to CNEs downregulate the expression of genes? Evidence suggests that such long-
313 range or trans enhancer promoter interactions occur at the cost of increased expression noise[39-
314 41]. As a result, the overall expression level in a tissue is expected to decline due to increased
315 stochastic fluctuations in gene expression across cells. Therefore, we hypothesized that the loss of
316 linear proximity between CNE and the gene would have compromised with the expression level of
317 the gene by allowing stochastic variations in enhancer-promoter interactions.
318
319 Enrichment of brain development related genes in the non-syntenic set might relate to
320 developmental plasticity of brain as compared to other tissues. Genomic alterations at the loci which
321 are important in the development of basic body plan and functioning would be embryonic lethal,
322 which largely explained the significant representation of skeletal/heart development and neonatal
323 lethality related genes in the syntenic set (Figure 3A, S2A). Brain, despite having neurodevelopment
324 plasticity, exhibits least genome-wide expression divergence across mammalian species[42, 43], but
325 within the space of small non-syntenic gene-set the expression divergence was observed. This
326 suggested that the least expression divergence observed for brain were due to cellular functions that
327 needed to be precisely regulated to maintain delicately shaped brain tissues of all the mammals in
328 general, while the ones that exhibited divergence would implicate in developmental functions
329 specific to fetal brain. Our data showed that one of the ways, such expression divergence was
330 modulated in the evolution was through alteration of genomic proximity between CNEs and the
331 neighbouring genes. Fetus brain-specific downregulation of neurogenesis related genes that had lost
332 synteny to CNEs in rat aligned to the hypothesis that observed genomic alterations might link to
333 brain traits that were lost in rodent lineage. We showed evidence that among the species taken in
334 this analyses, rat exhibited most number of independently modified brain traits, including the ones
335 directly associated with neurogenesis, like absence of cerebral folding of cortex, absence of
336 claustrum separation from cortex, absence of lateral geniculate nucleus magnocellular layer etc
337 (Figure S5). Of these, loss of cerebral folding of cortex, i.e., lissencephalic or smooth brain phenotype
338 is the largest visible alteration in the rodent brains. Folded or gyrencephalic brain, in general, is

339 considered as adaptation for the mammals with greater encephalization quotient, intelligence and
340 complex behavioural traits[44, 45]. It can, therefore, be contended that the CNE-gene proximity and
341 the associated fetal brain-specific expression was not lost in rat, but were rather gained in other
342 mammals that had bigger and gyrencephalic brains. However, we argue that significant non-
343 uniformity in cerebral cortex has been observed across several different mammalian species[46] and
344 the assumption that the common ancestor of placental mammals had a smaller and simpler brain
345 has been challenged recently[47, 48]. Evidence converges to gyrencephalic brain of eutherian
346 ancestor and the subsequent loss of cortical gyration in some lineages including rodents has been
347 supported[49]. The enrichment of genes associated with brain morphology phenotypes (Figure S2),
348 ECM-receptor interactions & actin cytoskeleton regulation (ACTN, ITGA1, ITGA11, ASAP1, LAMB4,
349 CD36 etc), and the ones implicated in human cortical malformations (MYCN, NRXN1, RASA1, DDX11,
350 FEZF2, EFNA5, GLI3 etc) [50]in the non-syntenic set further supported the loss of synteny in rat
351 rather than gain of synteny in other mammals (Figure S6). We also emphasized that the loss of
352 synteny in rat was inferred by filtering the CNE-gene pairs which were syntenic in all other species,
353 hence were evolutionarily constrained, except in rat. Assessing gain of synteny was difficult because
354 a CNE-gene pair that was non-syntenic in all species except one cannot be considered as
355 evolutionarily constrained CNE-gene pair. We suspected that gain of synteny inferred in this flawed
356 manner would not have shown any functional association. This indeed was observed through an
357 independent analysis (Figure S6).
358
359 It remains arguable whether or not the alterations in brain traits in rodent lineage represented the
360 adaptive selection or was a product of neutral drift. Some studies have suggested that smaller and
361 lissencephalic brain was adaptively selected among mammals with distinct life history traits, like
362 narrow habitat and smaller social groups, than that of gyrencephalic species[48]. Distinct neurogenic
363 potential of gyrencephalic and lissencephalic species has been attributed to the observed
364 difference[48]. Increased proliferative potential of basal progentior cells is necessary and sufficient
365 to explain the gyrencephalic brains[48]. The loss of such proliferative potential, which was likely an
366 ancestral trait, might have caused inefficient neurogenic program in lissencephalic species. Our
367 observation that the genes that had lost the synteny to CNEs in rat were involved in neuronal
368 differentiation and were downregulated in fetal rat brain is largely coherent with the above
369 proposal.
370
371 Altogether, our observations highlighted the link between genome order and the evolutionary
372 dynamics of temporal gene expression pattern associated with mammalian brain development. The

373 study also suggested that the genomic rearrangements, without any change in the genomic content,
374 might impact the developmental trajectories and shape the evolution of phenotypes.
375

376 ## Acknowledgement

377 Financial support to KSS from Department of Science and Technology (EMR/2015/001681) is duly
378 acknowledged. MB and ML were financially supported by IISER-Mohali. KRC, YJ and HS thank ICMR,
379 CSIR and SERB-DST agencies respectively for their fellowships.
380

381 ## Methods

382 **Compilation of chromosomal position data**

383 Human (hg19), rat (rn5), dog (camFam3), horse (equCab2) and cow (bosTau6) genome assemblies
384 were used in the analysis. Conserved Non-coding Elements (CNEs) were taken from Marcovitz et al
385 (Marcovitz et al.2016), which in turn were obtained by curating mammalian CNEs anchored to the
386 human genome (hg19). Our choice of the aforementioned species and the CNE dataset was
387 constrained by following considerations: i) We wanted sufficient evolutionary depth in the analysis
388 and Marcovitz el al had considered 20 sequenced mammalian genomes to identify CNEs; 2) Since our
389 analysis considered the chromosomal positions of CNEs and the genes, we only considered the
390 genomes for which complete chromosome assemblies were available. For example, chromosome
391 assemblies for the orders Cetacean, Chiroptera and Proboscidea etc. are not presently available; 3)
392 In order to obtain the sufficient number of orthologous genes across species, we restricted our
393 analysis to fewer mammalian lineages only. Considering multiple species would have compromised
394 the total number of orthologous genes to start with.
395

396 We obtained the orthologue positions of human CNEs in query species using standard approach of
397 mapping through LiftOver (https://genome-store.ucsc.edu/) chains at 0.95 mapping coverage[21].
398 Finally, we compiled 114219 CNEs that were having orthologous positions in all 5 species. We
399 independently obtained the table of orthologous genes across 5 mammals from Ensembl. Using CNE
400 and gene tables, the list of nearest genes that were within 1Mb to the CNEs was obtained for
401 human. The position of orthologous CNEs and genes in other mammals were assessed and CNE-gene
402 pairs were classified as syntenic if the distance between the two was less than 1Mb in all 5 species
403 and as non-syntenic if the CNE and the gene were >2Mb apart or were on different chromosomes in
404 one of the species and remained within 1Mb in rest of the species. If there were multiple
405 orthologues for the same gene, we took the nearest gene to the CNE on the same chromosome to
406 ensure that a syntenic pair should not have classified as non-syntenic because of mapping errors.

407 The distance cut-off of 1Mb was determined based on distribution of number of CNE-gene pairs at
408 different distance cut-offs. At around 1Mb, the overall distribution approaches a plateau and the
409 numbers do not increase significantly after that (Figure 1B). The lack of synteny cut-off of 2Mb
410 ensured that CNE and the gene were distant at-least by 2 fold in their non-syntenic form when
411 compared to their syntenic form. Larger distance cut-off was also likely to be robust against the
412 annotation artefacts of gene coordinates. A flow-chart illustrating the overall strategy is given in
413 Figure S1. All the data are available in supplementary data file.
414
415 To assess the genome assembly artefacts, we mapped the non-syntenic CNE-gene pairs of rn5 rat
416 assembly to rn6 assembly. Out of 2711 CNE-gene pairs, 2667 pairs (98.4%) were successfully lifted
417 over to rn6. All pairs maintained non-sytenic status in rn6 too (Figure S7). We further mapped the
418 non-syntenic CNE-pairs to known problematic regions of rat genome. Out of 2711 CNE, only 3 CNEs
419 (0.1%) and out of total 245 genes, only three genes (ABCC6, FOS, BNIP2; 1.2%) mapped to these
420 regions. Exclusion of these regions was unlikey to change our claims.
421
422 **Analysis of genomic attributes**
423 Chromosomal coordinates of repeat elements were downloaded from UCSC table browser. Repeat
424 elements were mapped +/- 50kb around syntenic and non-syntenic CNEs and average value of
425 enrichment in 2kb bins were plotted. For conservation analysis PhyloP scores of placental mammals
426 (http://ccg.vital-it.ch/mga/hg19/phylop/phylop.html) were mapped +/- 1kb to CNE.
427
428 **Functional enrichment analysis**
429 Gene Ontology and Mammalian Phenotype Ontology analysis was performed using GREAT
430 (http://bejerano.stanford.edu/great/public/html/). Tissue specificity analysis was done using TSEA
431 (http://genetics.wustl.edu/jdlab/tsea/), CSEA (http://genetics.wustl.edu/jdlab/csea-tool-2/). The
432 tissue specificity index (pSI) score of a gene i in a tissue k, over the given tissues $j=1,2,...m$ was
433 calculated as per Dougherty et al [51] using following equation:

$$pSI_{i,k} = \frac{\sum_{j=2}^{m} \left(rank\left(\frac{x_{i,k}}{x_{i,j}}\right) \right)}{m-1}$$

434 Where $x_{i,1}$ is the expression level of gene i in tissue 1 and $x_{i,j}$ is the expression level of gene i in tissue
435 j. A stringent pSI cut-off of 0.05 was taken for the analysis. For syntenic gene set, we randomly
436 sampled 245 genes (the size of non-syntenic gene set) from 4241 syntenic genes 100 times and
437 plotted the mean and the standard error of the significance (-log10 of corrected p-value) of overlap
438 between the candidate gene-sets and the tissues specific genes in the genome. The random sample

439 of syntenic genes that exhibited most significant overlap with the brain-specific genes was taken for
440 the expression specificity analysis among brain tissues across developmental stages.
441
442 Normalized gene expression data for developing cerebral corte and heart of human, rat and sheep
443 were taken from BRAINSPAN (human cortex; http://www.brainspan.org/static/download.html),
444 GSE71148(human heart), Stead et al (rat cortex), GSE53512 (rat heart), Clark et al (sheep cortex) and
445 GSE66725 (sheep heart). Average gene expression with 90% confidence intervals were plotted across
446 development time course.
447
448 **Enhancer analysis**
449 Regulatory potential of CNEs was assessed by mapping ChromHMM data obtained from Epigenome
450 roadmap (http://egg2.wustl.edu/roadmap/web_portal/imputed.html#chr_imp) and ENCODE
451 (http://hgdownload.soe.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeBroadHmm/) projects
452 onto CNEs. Enhancer coordinates from FANTOM (http://enhancer.binf.ku.dk/presets/) were also
453 mapped to CNEs. Cumulative overlap across aforementioned three resources was calculated.
454 Datasets for H3K4me1 methylation for fetal and post-natal/adult human tissues were obtained from
455 Epigenome Roadmap (http://www.roadmapepigenomics.org/data/) with following accession IDs and
456 age groups: fetal brain (E081, E082; 17GW), adult brain (E067, E068, E069, E071, E072, E073, E074;
457 pooled 73Yr/75Yr/81Yr), fetal muscle (E089, E090; 15GW), post-natal muscle (E107; pooled
458 54Yr/72Yr) and fetal thymus (E093; 15GW), post-natal thymus (E112; 3Yr), fetal heart (E083, 91
459 days), post-natal heart (E95, E104, E105, pooled 3Yr/34Yr), fetal small intestine(E085, 15GW) and
460 post-natal small intestine(E109, pooled 3Yr/30Yr). Fold-change over input DNA was used for
461 aggregation plots. WashU epigenome browser was used for visualization. Motif analysis was
462 performed through RSAT's 'peak-motif' package (http://rsat.sb-roscoff.fr/peak-motifs_form.cgi)
463 using JASPAR core matrices for vertebrate genomes. Syntenic CNEs were taken as background
464 control sequences.
465
466 **Mapping of proxy GWAS SNPs**
467 Total 251835 GWAS SNPs were obtained from GWASdb (http://jjwanglab.org/gwasdb). From this
468 data, 71990 brain related SNPs were obtained by analyzing the HPO terms associated with brain
469 associated phenotypes. We extended this data to total 533388 nearby SNPs (proxy) that were in
470 linkage disequilibrium to 71990 brain related GWAS SNPs based on 1000 genome data using SNAP
471 algorithm (https://personal.broadinstitute.org/plin/snap/index.php). Random null was prepared by
472 picking CNE samples, of same sample size and CNE-gene distance as of non-syntenic set, from the

473 syntenic set 1000 times and mapping SNPs to each of these samples. Number of CNEs with at-least
474 one SNP was counted for each sample. The distribution of these numbers was regarded as random
475 null. P-value was calculated as following:

—

476 Where B = number of re-sampling iterations (1000)

477 k_b = Number of sampled syntenic CNEs having at-least one SNP.

478 k = Number of observed non-syntenic CNE having atleast one SNP.

479

480 **Virtual 4C data**

481 SRA files of HiC datasets for fetal and adult brains were obtained from GSE77565 and GSE87112.
482 Datasets were processed using HiCUP (https://github.com/theaidenlab/juicer/wiki/HiCCUPS) and
483 contact maps were normalized using iterative correction and eigen vector decomposition (ICE)
484 method (https://github.com/hiclib/iced). TSS in each CNE-gene pair was taken as bait (reference
485 point) and its intra-chromosomal interactions were obtained from HiC matrices. Loess regression
486 line was fit to the HiC counts as a function of genomic distance from the bait. Signifcant interactions
487 with the bait were identified by applying cut-off of 3-standard deviation distance from the regression
488 line [52].

489

490 **DNA breakpoint analysis**

491 We obtained 552 germline breakpoints associated with congenital disorders having brain
492 abnormalities and 68018 somatic cancer breakpoints from van Heesch et al [53]. The matching RNA-
493 seq data of perpheral blood of patient and the mother were obtained from European Nucleotide
494 Archive(https://www.ebi.ac.uk/ena) with the accession IDs ERX358048 and ERX358046 respectively.

495 Total 2061 evolutionary DNA break-points for rodents were taken from Bourque etal (2004 & 2006),
496 Larkin etal and Lemaitre etal et al [54-57]. These breakpoints were then mapped onto inter-spacer
497 regions between CNE and the nearest gene-TSS. The random null was obtained by picking CNE-gene
498 pairs, of same sample size and CNE-gene distance as of non-syntenic set, from the syntenic set 1000
499 times and mapping the breakpoint in the inter-spacer regions. Number of CNE-gene pairs with at-
500 least one break-point in between was counted for each sample. The distribution of these numbers
501 was regarded as random null. P-values was calculated using equation as in GWAS SNP analysis. The
502 break-point distances from the ancestor were obtained from Luo et al[25].

503

504 **Mammalian traits**

505 Status of traits in 5 mammals were obtained from project ID P773 of Morphobank database
506 (https://morphobank.org/). Traits that exhibited same status atleast in 3 of the mammals including
507 rat, but showed a different status in human were classified as independently modified traits in
508 human. Similarly the traits that had same status in atleast 3 species including human, but changed
509 status in rat were denoted as independently modified traits in rat.
510
511 **Availability of data.** All datasets presented in this article are available as supplementary data:
512
513 **References**

514 1. Rands CM, Meader S, Ponting CP, Lunter G: **8.2% of the Human genome is constrained:**
515 **variation in rates of turnover across functional element classes in the human lineage.** *PLoS*
516 *Genet* 2014, **10:**e1004525.
517 2. Lindblad-Toh K, Garber M, Zuk O, Lin MF, Parker BJ, Washietl S, Kheradpour P, Ernst J, Jordan
518 G, Mauceli E, et al: **A high-resolution map of human evolutionary constraint using 29**
519 **mammals.** *Nature* 2011, **478:**476-482.
520 3. Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, Sheffield NC,
521 Stergachis AB, Wang H, Vernot B, et al: **The accessible chromatin landscape of the human**
522 **genome.** *Nature* 2012, **489:**75-82.
523 4. Consortium EP: **An integrated encyclopedia of DNA elements in the human genome.**
524 *Nature* 2012, **489:**57-74.
525 5. Skipper M, Eccleston A, Gray N, Heemels T, Le Bot N, Marte B, Weiss U: **Presenting the**
526 **epigenome roadmap.** *Nature* 2015, **518:**313.
527 6. Loots GG, Kneissel M, Keller H, Baptist M, Chang J, Collette NM, Ovcharenko D, Plajzer-Frick
528 I, Rubin EM: **Genomic deletion of a long-range bone enhancer misregulates sclerostin in**
529 **Van Buchem disease.** *Genome Res* 2005, **15:**928-935.
530 7. Sabherwal N, Bangs F, Roth R, Weiss B, Jantz K, Tiecke E, Hinkel GK, Spaich C, Hauffa BP, van
531 der Kamp H, et al: **Long-range conserved non-coding SHOX sequences regulate expression**
532 **in developing chicken limb and are associated with short stature phenotypes in human**
533 **patients.** *Hum Mol Genet* 2007, **16:**210-222.
534 8. Driscoll MC, Dobkin CS, Alter BP: **Gamma delta beta-thalassemia due to a de novo mutation**
535 **deleting the 5' beta-globin gene activation-region hypersensitive sites.** *Proc Natl Acad Sci U*
536 *S A* 1989, **86:**7470-7474.
537 9. Hatton CS, Wilkie AO, Drysdale HC, Wood WG, Vickers MA, Sharpe J, Ayyub H, Pretorius IM,
538 Buckle VJ, Higgs DR: **Alpha-thalassemia caused by a large (62 kb) deletion upstream of the**
539 **human alpha globin gene cluster.** *Blood* 1990, **76:**221-227.
540 10. Sparago A, Cerrato F, Vernucci M, Ferrero GB, Silengo MC, Riccio A: **Microdeletions in the**
541 **human H19 DMR result in loss of IGF2 imprinting and Beckwith-Wiedemann syndrome.**
542 *Nat Genet* 2004, **36:**958-960.
543 11. McLean CY, Reno PL, Pollen AA, Bassan AI, Capellini TD, Guenther C, Indjeian VB, Lim X,
544 Menke DB, Schaar BT, et al: **Human-specific loss of regulatory DNA and the evolution of**
545 **human-specific traits.** *Nature* 2011, **471:**216-219.
546 12. Seridi L, Ryu T, Ravasi T: **Dynamic epigenetic control of highly conserved noncoding**
547 **elements.** *PLoS One* 2014, **9:**e109326.
548 13. Roh TY, Wei G, Farrell CM, Zhao K: **Genome-wide prediction of conserved and**
549 **nonconserved enhancers by histone acetylation patterns.** *Genome Res* 2007, **17:**74-81.
550 14. Babarinde IA, Saitou N: **Genomic Locations of Conserved Noncoding Sequences and Their**
551 **Proximal Protein-Coding Genes in Mammalian Expression Dynamics.** *Mol Biol Evol* 2016,
552 **33:**1807-1817.

553 15. Woolfe A, Goodson M, Goode DK, Snell P, McEwen GK, Vavouri T, Smith SF, North P,
554 Callaway H, Kelly K, et al: **Highly conserved non-coding sequences are associated with**
555 **vertebrate development.** *PLoS Biol* 2005, **3:**e7.
556 16. Akalin A, Fredman D, Arner E, Dong X, Bryne JC, Suzuki H, Daub CO, Hayashizaki Y, Lenhard B:
557 **Transcriptional features of genomic regulatory blocks.** *Genome Biol* 2009, **10:**R38.
558 17. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, Klemm A, Flicek P, Manolio T,
559 Hindorff L, Parkinson H: **The NHGRI GWAS Catalog, a curated resource of SNP-trait**
560 **associations.** *Nucleic Acids Res* 2014, **42:**D1001-1006.
561 18. Rajagopal N, Srinivasan S, Kooshesh K, Guo Y, Edwards MD, Banerjee B, Syed T, Emons BJ,
562 Gifford DK, Sherwood RI: **High-throughput mapping of regulatory DNA.** *Nat Biotechnol*
563 2016, **34:**167-174.
564 19. Naville M, Ishibashi M, Ferg M, Bengani H, Rinkwitz S, Krecsmarik M, Hawkins TA, Wilson
565 SW, Manning E, Chilamakuri CS, et al: **Long-range evolutionary constraints reveal cis-**
566 **regulatory interactions on the human X chromosome.** *Nat Commun* 2015, **6:**6904.
567 20. Davies KT, Tsagkogeorga G, Rossiter SJ: **Divergent evolutionary rates in vertebrate and**
568 **mammalian specific conserved non-coding elements (CNEs) in echolocating mammals.**
569 *BMC Evol Biol* 2014, **14:**261.
570 21. Marcovitz A, Jia R, Bejerano G: **"Reverse Genomics" Predicts Function of Human Conserved**
571 **Noncoding Elements.** *Mol Biol Evol* 2016, **33:**1358-1369.
572 22. Chan JE, Kolodner RD: **A genetic and structural study of genome rearrangements mediated**
573 **by high copy repeat Ty1 elements.** *PLoS Genet* 2011, **7:**e1002089.
574 23. Roeder GS, Fink GR: **DNA rearrangements associated with a transposable element in yeast.**
575 *Cell* 1980, **21:**239-249.
576 24. Buckley RM, Kortschak RD, Raison JM, Adelson DL: **Similar Evolutionary Trajectories for**
577 **Retrotransposon Accumulation in Mammals.** *Genome Biol Evol* 2017, **9:**2336-2353.
578 25. Luo H, Arndt W, Zhang Y, Shi G, Alekseyev MA, Tang J, Hughes AL, Friedman R: **Phylogenetic**
579 **analysis of genome rearrangements among five mammalian orders.** *Mol Phylogenet Evol*
580 2012, **65:**871-882.
581 26. Matsumoto M, Straub RE, Marenco S, Nicodemus KK, Matsumoto S, Fujikawa A, Miyoshi S,
582 Shobo M, Takahashi S, Yarimizu J, et al: **The evolutionarily conserved G protein-coupled**
583 **receptor SREB2/GPR85 influences brain size, behavior, and vulnerability to schizophrenia.**
584 *Proc Natl Acad Sci U S A* 2008, **105:**6133-6138.
585 27. Chen L, Zheng J, Yang N, Li H, Guo S: **Genomic selection identifies vertebrate transcription**
586 **factor Fezf2 binding sites and target genes.** *J Biol Chem* 2011, **286:**18641-18649.
587 28. Eckler MJ, Larkin KA, McKenna WL, Katzman S, Guo C, Roque R, Visel A, Rubenstein JL, Chen
588 B: **Multiple conserved regulatory domains promote Fezf2 expression in the developing**
589 **cerebral cortex.** *Neural Dev* 2014, **9:**6.
590 29. Steinecke A, Gampe C, Zimmer G, Rudolph J, Bolz J: **EphA/ephrin A reverse signaling**
591 **promotes the migration of cortical interneurons from the medial ganglionic eminence.**
592 *Development* 2014, **141:**460-471.
593 30. Senturk A, Pfennig S, Weiss A, Burk K, Acker-Palmer A: **Ephrin Bs are essential components**
594 **of the Reelin pathway to regulate neuronal migration.** *Nature* 2011, **472:**356-360.
595 31. Wiegreffe C, Simon R, Peschkes K, Kling C, Strehle M, Cheng J, Srivatsa S, Liu P, Jenkins NA,
596 Copeland NG, et al: **Bcl11a (Ctip1) Controls Migration of Cortical Projection Neurons**
597 **through Regulation of Sema3c.** *Neuron* 2015, **87:**311-325.
598 32. Cross-Disorder Group of the Psychiatric Genomics C: **Identification of risk loci with shared**
599 **effects on five major psychiatric disorders: a genome-wide analysis.** *Lancet* 2013,
600 **381:**1371-1379.
601 33. Calabrese V, Scapagnini G, Ravagna A, Giuffrida Stella AM, Butterfield DA: **Molecular**
602 **chaperones and their roles in neural cell differentiation.** *Dev Neurosci* 2002, **24:**1-13.

603 34. Sturtevant AH: **A Further Study of the so-Called Mutation at the Bar Locus of Drosophila.**
604 *Genetics* 1928, **13:**401-409.
605 35. Sturtevant AH: **The Effects of Unequal Crossing over at the Bar Locus in Drosophila.**
606 *Genetics* 1925, **10:**117-147.
607 36. Martin-Morris LE, Csink AK, Dorer DR, Talbert PB, Henikoff S: **Heterochromatic trans-**
608 **inactivation of Drosophila white transgenes.** *Genetics* 1997, **147:**671-677.
609 37. Veron AS, Lemaitre C, Gautier C, Lacroix V, Sagot MF: **Close 3D proximity of evolutionary**
610 **breakpoints argues for the notion of spatial synteny.** *BMC Genomics* 2011, **12:**303.
611 38. Bagadia M, Singh A, Singh Sandhu K: **Three Dimensional Organization of Genome Might**
612 **Have Guided the Dynamics of Gene Order Evolution in Eukaryotes.** *Genome Biol Evol* 2016,
613 **8:**946-954.
614 39. Singh A, Bagadia M, Sandhu KS: **Spatially coordinated replication and minimization of**
615 **expression noise constrain three-dimensional organization of yeast genome.** *DNA Res*
616 2016, **23:**155-169.
617 40. Sandhu KS: **Did the modulation of expression noise shape the evolution of three**
618 **dimensional genome organizations in eukaryotes?** *Nucleus* 2012, **3:**286-289.
619 41. Kustatscher G, Grabowski P, Rappsilber J: **Pervasive coexpression of spatially proximal**
620 **genes is buffered at the protein level.** *Mol Syst Biol* 2017, **13:**937.
621 42. Khaitovich P, Enard W, Lachmann M, Paabo S: **Evolution of primate gene expression.** *Nat*
622 *Rev Genet* 2006, **7:**693-702.
623 43. Strand AD, Aragaki AK, Baquet ZC, Hodges A, Cunningham P, Holmans P, Jones KR, Jones L,
624 Kooperberg C, Olson JM: **Conservation of regional gene expression in mouse and human**
625 **brain.** *PLoS Genet* 2007, **3:**e59.
626 44. Prothero JW, Sundsten JW: **Folding of the cerebral cortex in mammals. A scaling model.**
627 *Brain Behav Evol* 1984, **24:**152-167.
628 45. Toro R, Perron M, Pike B, Richer L, Veillette S, Pausova Z, Paus T: **Brain size and folding of**
629 **the human cerebral cortex.** *Cereb Cortex* 2008, **18:**2352-2357.
630 46. Herculano-Houzel S, Collins CE, Wong P, Kaas JH, Lent R: **The basic nonuniformity of the**
631 **cerebral cortex.** *Proc Natl Acad Sci U S A* 2008, **105:**12593-12598.
632 47. Rowe TB, Macrini TE, Luo ZX: **Fossil evidence on origin of the mammalian brain.** *Science*
633 2011, **332:**955-957.
634 48. Lewitus E, Kelava I, Kalinka AT, Tomancak P, Huttner WB: **An adaptive threshold in**
635 **mammalian neocortical evolution.** *PLoS Biol* 2014, **12:**e1002000.
636 49. Kelava I, Lewitus E, Huttner WB: **The secondary loss of gyrencephaly as an example of**
637 **evolutionary phenotypical reversal.** *Frontiers in Neuroanatomy* 2013, **7:**16.
638 50. Pinero J, Bravo A, Queralt-Rosinach N, Gutierrez-Sacristan A, Deu-Pons J, Centeno E, Garcia-
639 Garcia J, Sanz F, Furlong LI: **DisGeNET: a comprehensive platform integrating information**
640 **on human disease-associated genes and variants.** *Nucleic Acids Res* 2017, **45:**D833-D839.
641 51. Dougherty JD, Schmidt EF, Nakajima M, Heintz N: **Analytical approaches to RNA profiling**
642 **data for the identification of genes enriched in specific cells.** *Nucleic Acids Res* 2010,
643 **38:**4218-4230.
644 52. Klein FA, Pakozdi T, Anders S, Ghavi-Helm Y, Furlong EE, Huber W: **FourCSeq: analysis of 4C**
645 **sequencing data.** *Bioinformatics* 2015, **31:**3085-3091.
646 53. van Heesch S, Simonis M, van Roosmalen MJ, Pillalamarri V, Brand H, Kuijk EW, de Luca KL,
647 Lansu N, Braat AK, Menelaou A, et al: **Genomic and functional overlap between somatic**
648 **and germline chromosomal rearrangements.** *Cell Rep* 2014, **9:**2001-2010.
649 54. Larkin DM, Pape G, Donthu R, Auvil L, Welge M, Lewin HA: **Breakpoint regions and**
650 **homologous synteny blocks in chromosomes have different evolutionary histories.**
651 *Genome Res* 2009, **19:**770-777.

652 55. Bourque G, Zdobnov EM, Bork P, Pevzner PA, Tesler G: **Comparative architectures of**
653 **mammalian and chicken genomes reveal highly variable rates of genomic rearrangements**
654 **across different lineages.** *Genome Res* 2005, **15:**98-110.
655 56. Bourque G, Pevzner PA, Tesler G: **Reconstructing the genomic architecture of ancestral**
656 **mammals: lessons from human, mouse, and rat genomes.** *Genome Res* 2004, **14:**507-516.
657 57. Lemaitre C, Zaghloul L, Sagot MF, Gautier C, Arneodo A, Tannier E, Audit B: **Analysis of fine-**
658 **scale mammalian evolutionary breakpoints provides new insight into their relation to**
659 **genome organisation.** *BMC Genomics* 2009, **10:**335.

660

661 ## Figure legends

662 **Figure 1. Synteny and lack thereof between CNE and the nearest gene.** (a) Illustration of the strategy to infer
663 the synteny and lack thereof between CNE and the neighboring gene across 5 representative mammalian
664 orders. CNE-gene pairs were classified as syntenic if remained proximal (<1Mb) in all the 5 species and as non-
665 syntenic if departed by > 2mb or on different chromosomes in on of the species while maintaining synteny in
666 other 4 species. (b) Cumulative distribution of all CNE-gene distances in the human genome. Most CNE-gene
667 pairs were <1Mb distance and, therefore, cut-off of 1Mb was applied for CNE-gene synteny. (c) Sequence
668 conservation, as measured through mammalian PhyloP scores, and (d) length distribution of CNEs in syntenic
669 and non-syntenic sets. (e) Enrichment of retrotransposons +/- 50Kb around syntenic and non-syntenic CNEs.
670 Asterisk indicate significant p-values (<0.05) calculated using Mann Whitney U test of enrichment values +/-
671 10kb around CNEs.

672

673 **Figure 2. Genomic rearrangements underlying the loss of synteny** (a) Number of non-syntenic instances and
674 genes (in parentheses) in different mammals. P-value for the non-syntenic instances in rat and the next highest
675 value (in dog) was calculated using Fisher's exact test. (b) Scaling between number of non-syntenic cases and
676 the evolutionary break-point distance from the common mammalian ancestor. (c) Number of non-syntenic
677 CNE-gene pairs having atleast one rodent-specific break-point inbetween CNE and gene-TSS, overlaid onto null
678 distribution prepared from syntenic set. (d) Distinct *trans* and *cis* chromosomal rearrangements as inferred
679 from the analysis of genes flanking the non-syntenic CNEs. Shown are the neighboring genes around CNE. Red
680 color represents the target gene and orange color represents the nearest gene on the other side of the CNE.
681 Red color in the barplot marks the proportion for which inversion could be confirmed through analysis of gene
682 orientations. Shown are the two examples illustrating loss of CNE-gene synteny in rat. In first example, a CNE
683 was located 45kb upstream to the gene POU3F2 in human, but were split on different chromosomes in rat.
684 Second example shows that an inversion event had distanced the CNE and the proximal ADAM23 gene upto
685 2.4Mb in rat genome.

686

687 **Figure 3. Functional characterization of genes in syntenic and non-syntenic sets.** (a) Enrichment of Gene
688 Ontology (GO) terms among genes in syntenic and non-syntenic sets. P-values shown were corrected using
689 Benjamini-Hochberg method. (b) Tissue-specific expression analysis of genes in syntenic and non-syntenic set.
690 Relative significance was plotted as negative of log10 transformed corrected p-values of Fisher's exact test for

691 the overlap with the tissue-specific genes at stringency score (pSI) < 0.05. Horizontal grey colored dashed line
692 represents the p-value of 0.01. For syntenic set, mean values and standard errors of significance for 100
693 random samples of 245 genes (the size of non-syntenic set) from syntenic sets were plotted. (c) Expression
694 specificity of genes in syntenic and non-syntenic sets across brain regions and across developmental stages.
695 For syntenic set, the sample that exhibited maximum significance for brain specificity in panel-b was taken.
696 Size of the nested hexagons represents the proportion of all genes specifically expressed in particular tissue at
697 particular developmental stage. Hexagons are nested inwards based on relative stringency of tissue specificity
698 scores (pSI=0.05, 0.01, 0.001 & 0.0001 respectively). Color gradient represents the magnitude of corrected p-
699 values of Fisher's exact test.
700
701 **Figure 4. Enhancer properties of syntenic and non-syntenic CNEs. (a)** H3K4me1 ChIP enrichment (over input)
702 on and around CNEs in syntenic and non-syntenic sets in fetal and post-natal tissues. P-values for the
703 difference between syntenic and non-syntenic CNEs were calculated using Mann Whitney U test of H3K4me1
704 enrichment values in 1kb spanning windows on either side of the CNEs. (b) Virtual 4C analysis of CNE-gene
705 interactions in fetal and adult human brains. The barplot shows fetal-to-adult ratio of proportion of CNE-gene
706 pairs exhibiting significant physical chromatin interactions. (c) The examples of GPR85 and FEZF2 genes are
707 shown for illustration. Vertical grey and yellow bars represent the TSS and CNE positions respectively. Red and
708 grey curves show virtual 4C and h3K4me1 signals respectively. Black smooth line represents the loess fit of the
709 4C signal as function of genomic distance from the reference point. The dotted line represents 3σ distance
710 from the loess regression line. (d) Proportion of non-syntenic CNEs having at-least one brain associated SNP
711 superimposed onto the null distribution obtained from syntenic CNEs. P-value was calculated using boot-strap
712 method by randomly sampling 2711 CNEs (size of non-syntenic set) from syntenic set 1000 times. (e) An
713 example of EPHA4 gene and its proximal CNE having a Schizophrenia associated GWAS SNP is shown. The
714 tracks for PhyloP conservation score, H3K4me1 and RNA-seq data of fetal and post-natal brain aligned
715 accordingly. The orthologous CNE and gene were 7.2 Mb apart on chromosome 9 in rat.
716
717
718 **Figure 5. Tolerance and intolerance of CNE-gene split**. **(a)** Number of non-syntenic CNE-gene pairs flanking at-
719 least one germ-line breakpoint associated with the congenital disorders having brain abnormalities (upper
720 panel) and somatic cancer breakpoint (lower panel), superimposed onto null distributions obtained from the
721 syntenic set of same CNE-gene distance distribution as that of non-syntenic set. P-values were calculating
722 using boot-strap method with 1000 random samplings. (b) Examples illustrating chromothripsis, translocation
723 and inversion events breaking CNE-gene synteny in congenital disorders with brain abnormalities. Mutations in
724 BCL11A (i), CCDC68 (ii) and DNAJB6 (iii) genes are associated with abnormal brain phenotypes as elaborated in
725 the results section. The right most panel shows difference in expression level of BCL11A gene in the patient
726 having genomic rearrangement and the normal mother.
727

728 **Figure 6. Evolutionary dynamics of developmental gene expression associated with the loss of CNE-gene**
729 **synteny**. Red curves in the plots represent the mean expression of genes in the non-syntenic set and grey area
730 represents 95% confidence interval. Grey colored elongated triangles below the x-axis represent the fetal and
731 post-natal time course. Lack of triangles at some places denotes unavailability of multiple time points in the
732 data. Left panel represents cerebral cortex and right panel is for heart datasets (control).
733
734

735 ## Supplementary Figure Legends

736 **Figure S1.** Flowchart illustrating overall strategy to obtain syntenic and non-syntenic CNE-gene pairs.

737 **Figure S2.** Enrichment of (a) Mammalian Phenotype Ontology (MPO) terms , and (b) Bgee anatomical terms in
738 syntenic (left panel) and non-syntenic (right panel) gene-sets. P-values were corrected using Benjamini-
739 Hochberg method.
740

741 **Figure S3.** Extension of Epigenomic analyses of syntenic and non-syntenic CNEs. (A-B) H3K4me1 ChIP
742 enrichment (over inut DNA) around syntenic (grey) and non-syntenic (red) CNEs in (a) fetal, and (b) post-natal
743 tissues. (C) Strategy used to assess the CNE-TSS interactions using HiC datasets of human fetal and adult
744 brains. All-to-all HiC interactions were filtered for TSS-to-all interactions for the genes in syntenic and
745 nonsynteic sets. The resultant data was analogous to 4C and was analyzed using method atypical for 4C
746 analysis. Loess regression line was fit to 4C counts as a function of genomic distance from the refrence point
747 (TSS in this case). The distance of 3-standard deviation from this regression line was taken as signicance cut-off
748 for the interactions impinging onto TSS. (D) Example of CNE-TSS interactions identified in the non-syntenic set.
749 Upper and lower panels represent fetal and adult brain data. Red line: 4C signal; grey line: H3K4me1; Black
750 line: Loess fit; Dotted line: 3-standard deviation cut-off for signicance.

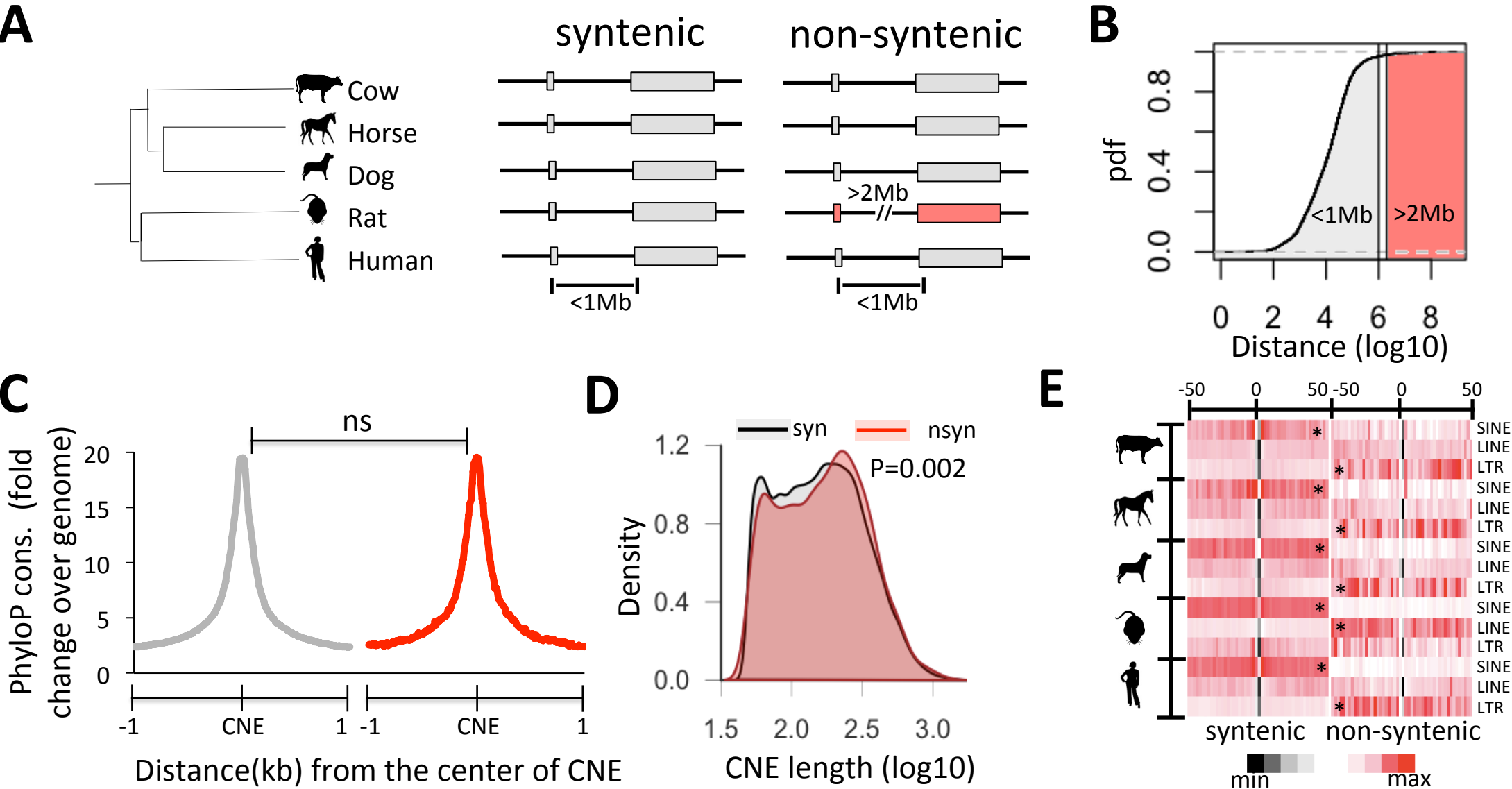
751 **Figure S4**. Sequence motif enrichment analysis for non-synteic CNEs. 'Peak-motifs' from RSAT was used to
752 identify over-represented sequence motifs in the non-synteic CNEs while taking syntenic CNEs as background
753 control. (a) Sequence motifs, their e-values and the matching transcription factors (TFs) from JASPAR. (b)
754 Tissue specificity analyses of TFs. Red bars represent the tissues, wherein the TFs exhibit significant specificity
755 ($P<0.05$). (c) Time course gene expression of TFs during human brain development. Red curve represents
756 average expression of TFs and grey colour denotes 90% confidence interval.

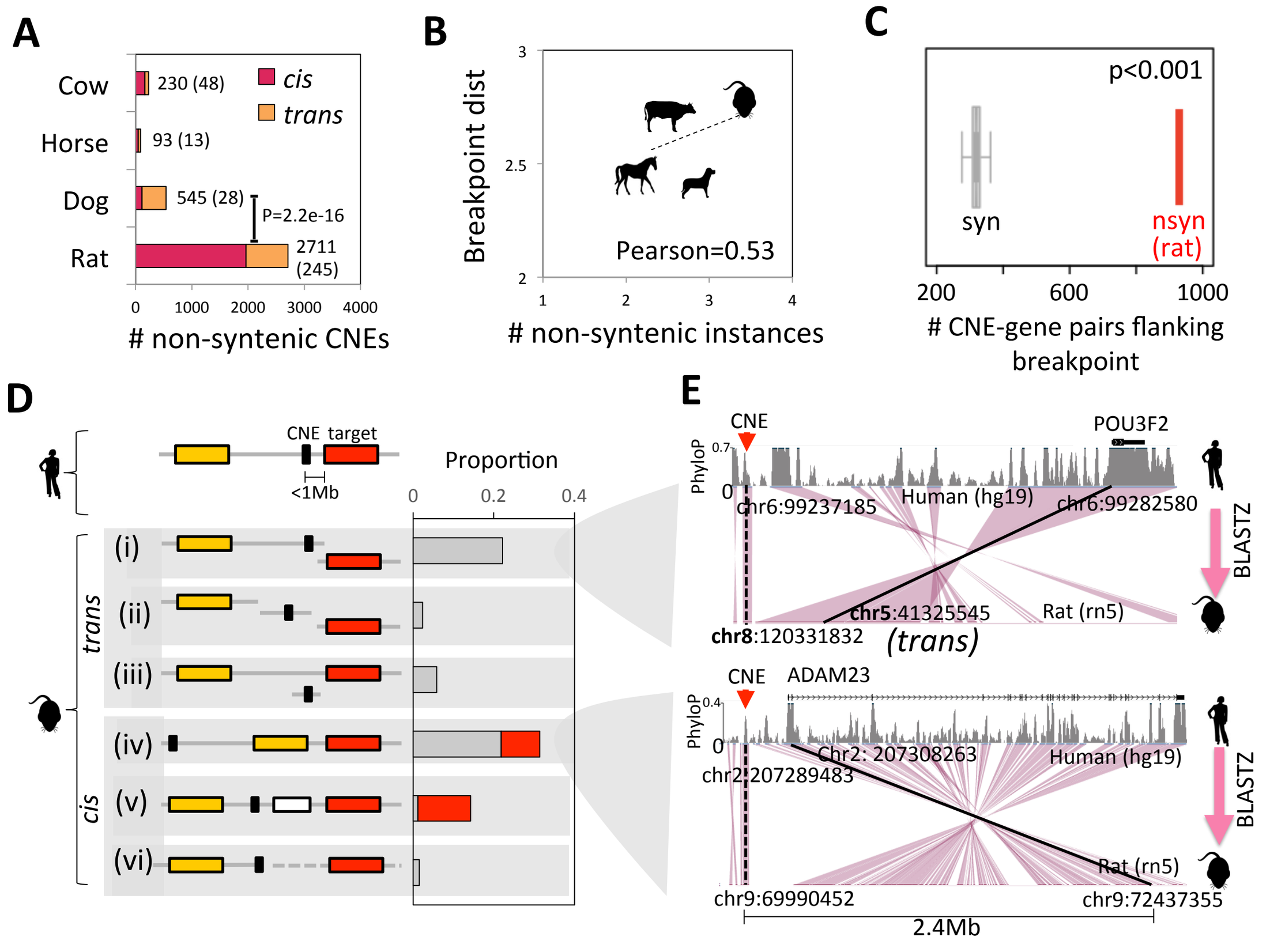
757 **Figure S5.** Independently modified traits in rat and human. (A) bar plot representing rat-to-human ratio for the
758 proportion of traits independently modified in each trait class. (B) List of traits that were independetly
759 modified in rat.

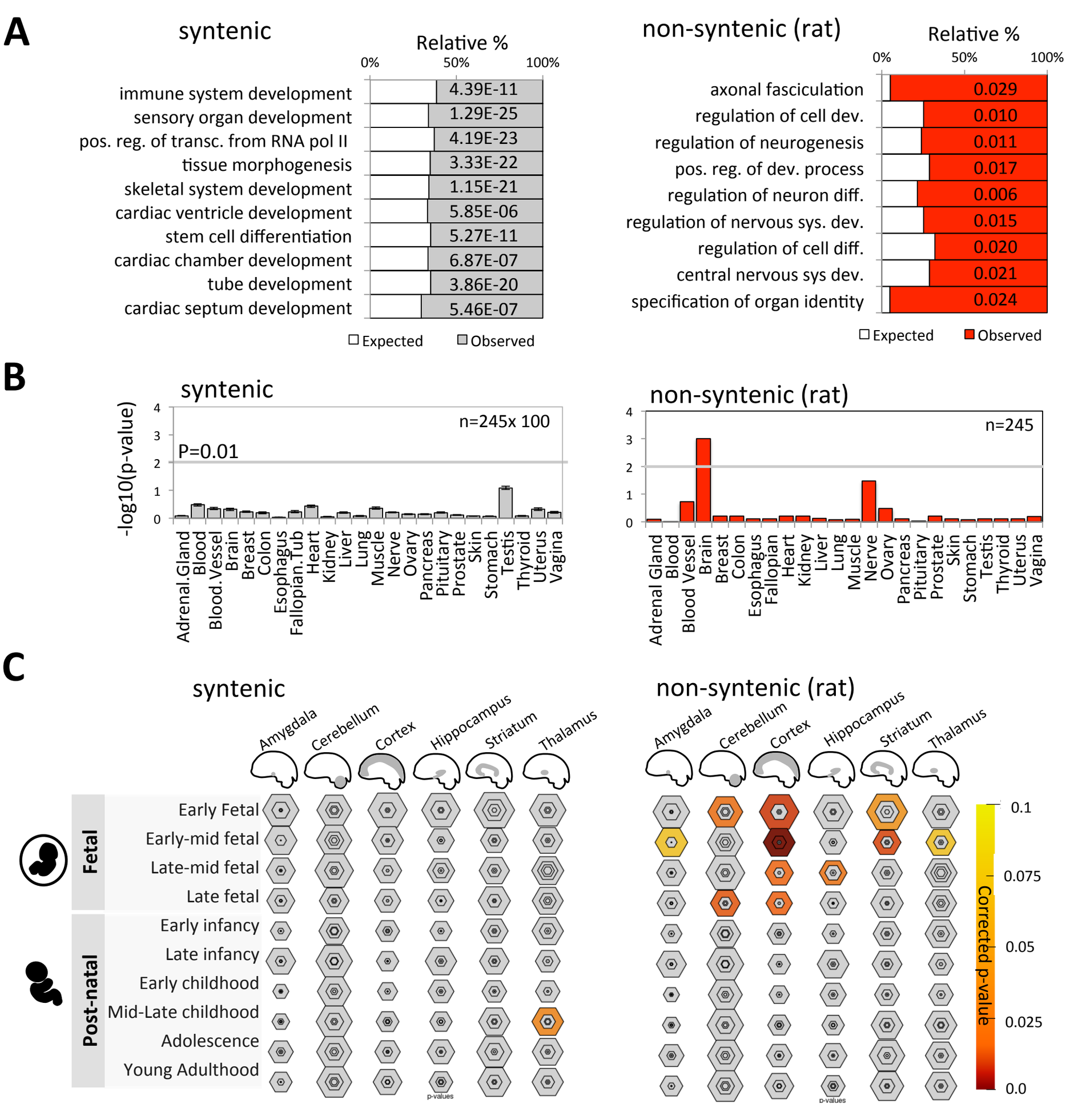
760 **Figure S6**. Analysis of gain of synteny instances in human. Gain of synteny instances were identified using the
761 strategy shown. No enrichment was found in the GO process terms. The most enriched GO term is shown.

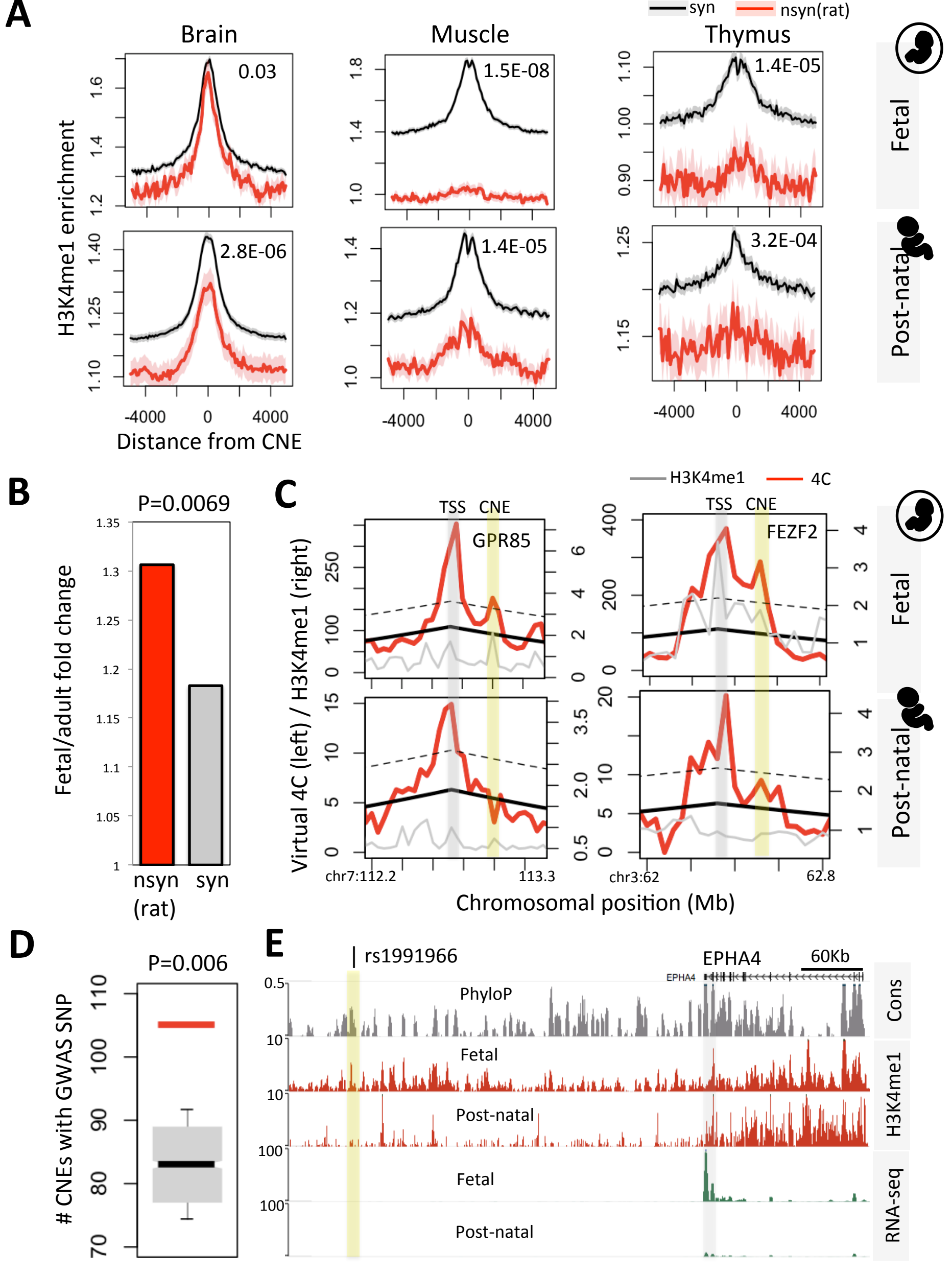
762 **Figure S7.**Comparison of non-syntenic (rat) CNE-gene pairs in rn5 and rn6 genome assemblies. Shown is the
763 scatter plot of CNE-gene distances (log10 scale) the non-syntenic set in rn5 and rn6 assemblies. All the non-
764 syntenic CNE-gene pairs in rn5 remained non-syntenic in rn6.

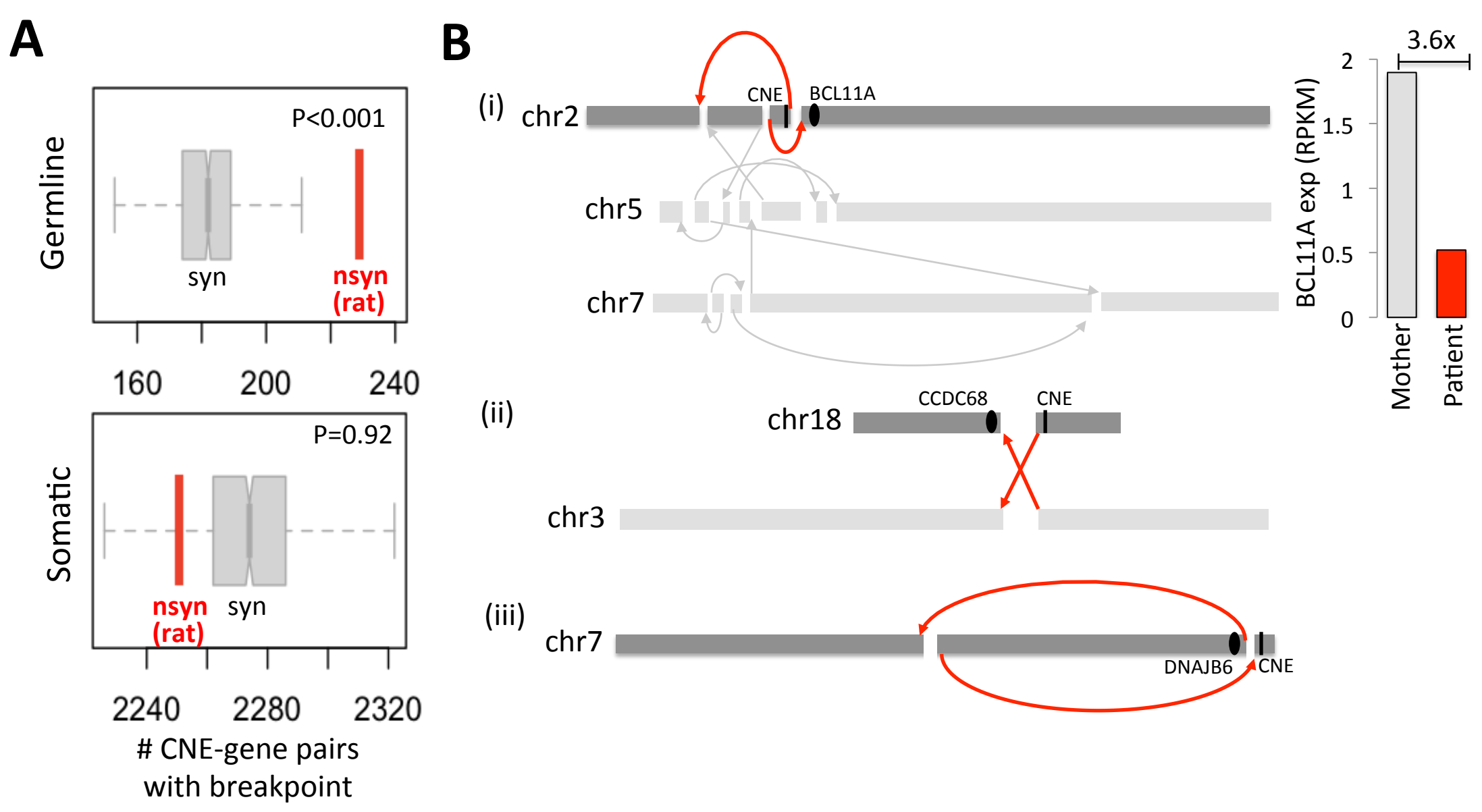
765

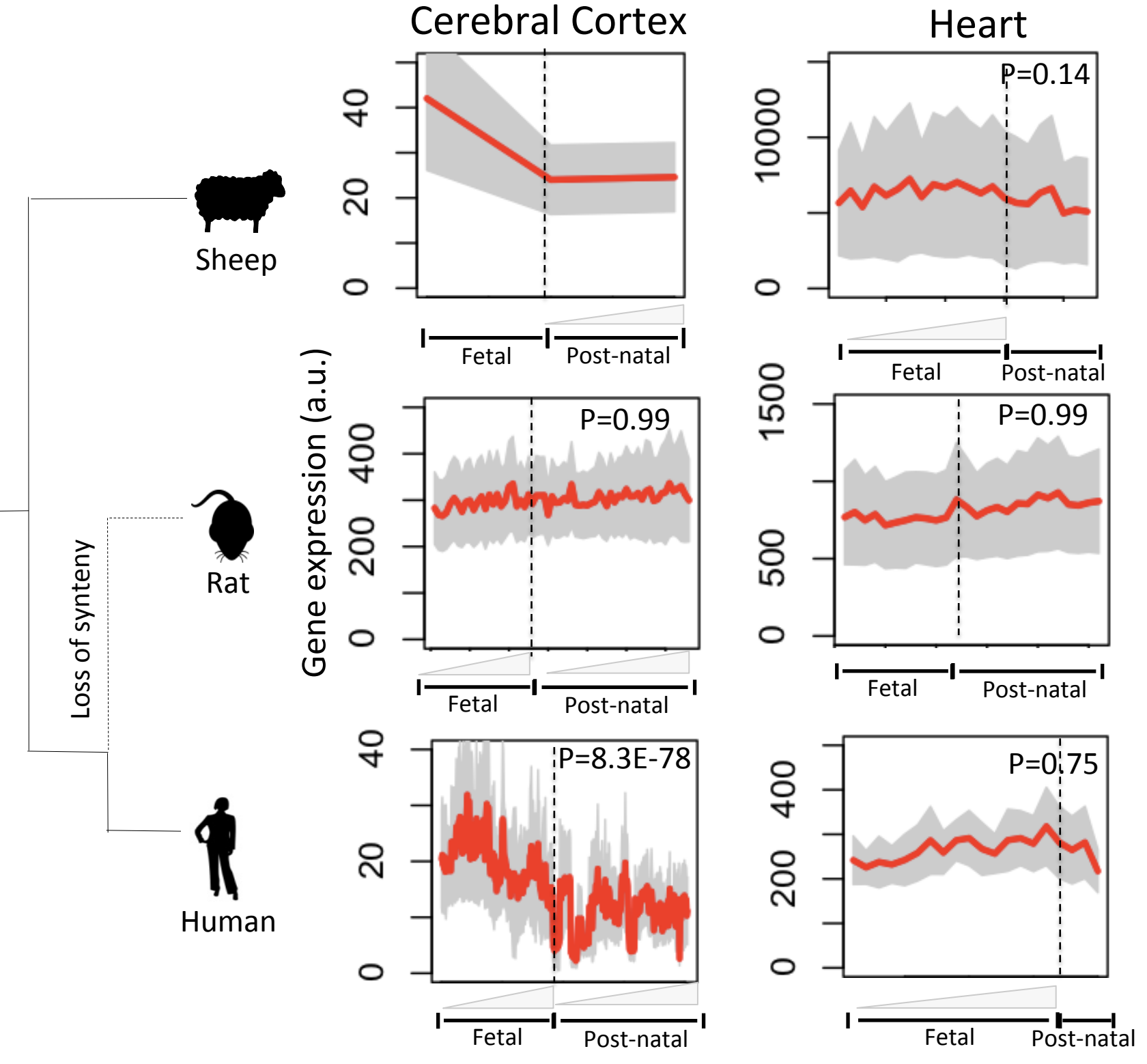

A
syntenic
non-syntenic
Cow
Horse
Dog
Rat
Human
>2Mb
<1Mb
<1Mb
B
pdf
0.0
0.4
0.8
<1Mb
>2Mb
0
2
4
6
8
Distance (log10)
C
ns
PhyloP cons. (fold change over genome)
20
15
10
5
0
-1
CNE
1
-1
CNE
1
Distance(kb) from the center of CNE
D
syn
nsyn
P=0.002
Density
1.2
0.8
0.4
0.0
1.5
2.0
2.5
3.0
CNE length (log10)
E
-50
0
50
-50
0
50
SINE
LINE
LTR
syntenic
non-syntenic
min
max


A
Cow
Horse
Dog
Rat
230 (48)
93 (13)
545 (28)
P=2.2e-16
2711 (245)
cis
trans
0
1000
2000
3000
4000
non-syntenic CNEs
B
Breakpoint dist
3
2.5
2
1
2
3
4
Pearson=0.53
# non-syntenic instances
C
p<0.001
syn
nsyn (rat)
200
600
1000
CNE-gene pairs flanking breakpoint
D
CNE target
<1Mb
Proportion
0
0.2
0.4
trans
cis
(i)
(ii)
(iii)
(iv)
(v)
(vi)
E
CNE
POU3F2
PhyloP
0.7
0
chr6:99237185
Human (hg19)
chr6:99282580
BLASTZ
chr5:41325545
Rat (rn5)
chr8:120331832
(trans)
CNE
ADAM23
PhyloP
0.4
0
Chr2: 207308263
Human (hg19)
chr2:207289483
BLASTZ
Rat (rn5)
chr9:69990452
chr9:72437355
2.4Mb


A
syntenic
Relative %
0%
50%
100%
immune system development
sensory organ development
pos. reg. of transc. from RNA pol II
tissue morphogenesis
skeletal system development
cardiac ventricle development
stem cell differentiation
cardiac chamber development
tube development
cardiac septum development
4.39E-11
1.29E-25
4.19E-23
3.33E-22
1.15E-21
5.85E-06
5.27E-11
6.87E-07
3.86E-20
5.46E-07
Expected
Observed
non-syntenic (rat)
Relative %
0%
50%
100%
axonal fasciculation
regulation of cell dev.
regulation of neurogenesis
pos. reg. of dev. process
regulation of neuron diff.
regulation of nervous sys. dev.
regulation of cell diff.
central nervous sys dev.
specification of organ identity
0.029
0.010
0.011
0.017
0.006
0.015
0.020
0.021
0.024
Expected
Observed
B
syntenic
-log10(p-value)
n=245x 100
P=0.01
Adrenal.Gland
Blood
Blood.Vessel
Brain
Breast
Colon
Esophagus
Fallopian.Tub
Heart
Kidney
Liver
Lung
Muscle
Nerve
Ovary
Pancreas
Pituitary
Prostate
Skin
Stomach
Testis
Thyroid
Uterus
Vagina
non-syntenic (rat)
n=245
Adrenal Gland
Blood
Blood Vessel
Brain
Breast
Colon
Esophagus
Fallopian
Heart
Kidney
Liver
Lung
Muscle
Nerve
Ovary
Pancreas
Pituitary
Prostate
Skin
Stomach
Testis
Thyroid
Uterus
Vagina
C
syntenic
Amygdala
Cerebellum
Cortex
Hippocampus
Striatum
Thalamus
Fetal
Early Fetal
Early-mid fetal
Late-mid fetal
Late fetal
Post-natal
Early infancy
Late infancy
Early childhood
Mid-Late childhood
Adolescence
Young Adulthood
p-values
non-syntenic (rat)
Amygdala
Cerebellum
Cortex
Hippocampus
Striatum
Thalamus
p-values
Corrected p-value
0.1
0.075
0.05
0.025
0.0


A
syn
nsyn(rat)
Brain
Muscle
Thymus
0.03
1.5E-08
1.4E-05
2.8E-06
1.4E-05
3.2E-04
Fetal
Post-natal
H3K4me1 enrichment
Distance from CNE
B
P=0.0069
Fetal/adult fold change
nsyn (rat)
syn
C
H3K4me1
4C
TSS
CNE
GPR85
FEZF2
Virtual 4C (left) / H3K4me1 (right)
chr7:112.2
113.3
chr3:62
62.8
Chromosomal position (Mb)
D
P=0.006
# CNEs with GWAS SNP
E
rs1991966
EPHA4
60Kb
PhyloP
Fetal
Post-natal
Cons
H3K4me1
RNA-seq


A
Germline
P<0.001
syn
nsyn
(rat)
160
200
240
Somatic
P=0.92
nsyn
(rat)
syn
2240
2280
2320
CNE-gene pairs
with breakpoint
B
(i)
chr2
CNE
BCL11A
chr5
chr7
3.6x
BCL11A exp (RPKM)
2
1.5
1
0.5
0
Mother
Patient
(ii)
chr18
CCDC68
CNE
chr3
(iii)
chr7
DNAJB6
CNE


Cerebral Cortex
Heart
Sheep
Rat
Human
Loss of synteny
Gene expression (a.u.)
P=0.14
P=0.99
P=0.99
P=8.3E-78
P=0.75
Fetal
Post-natal
0
20
40
200
400
500
1500
10000