Evolutionary dynamics of genome-wide position effects in mammals

Meenakshi Bagadia, Keerthivasan Raanin Chandradoss, Yachna Jain, Kuljeet Singh Sandhu*

Department of Biological Sciences

Indian Institute of Science Education and Research (IISER) - Mohali

Knowledge City, Sector - 81, SAS Nagar 140306, India

To whom correspondence should be addressed

Kuljeet Singh Sandhu

Assistant Professor and Group Leader

Genome Biology Laboratory, Department of Biological Sciences

Indian Institute of Science Education and Research (IISER) - Mohali

E. mail: sandhuks@iisermohali.ac.in

Abstract

Alterations in genomic proximity of a gene and its regulatory elements can impact transcriptional state of the gene. Through genome-wide analysis of Conserved Noncoding Elements (CNEs) and their cognate genes from representative mammals, we show that the genes syntenic to their adjacent CNEs were associated with developmentally essential functions while the ones that had lost synteny independently in one of the non-human lineage were associated with fetal, but not post-natal, brain development in human. Accordingly, associated CNEs exhibited specific enhancer activity in fetal brain and contained SNPs associated with brain disorders. Relatively greater representation of DNA-breakpoints of germ-line origin between CNE and the gene signified the underlying developmental tolerance of CNE-gene split during loss of synteny. While between closely related species, like rat and mouse, linear split of CNE-gene synteny was compensated by their spatial proximity to maintain gene-expression, the gain and loss of genes' proximity to CNEs between distant species strikingly correlated with the gain and loss of tissue-specific fetal gene-expression as exemplified through developing brain and heart in human and mouse. These observations highlight the nontrivial contribution of position-effect in the evolution of gene-expression pattern during development and have implications in evolutionary gain and loss of lineage-specific traits.

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

Introduction

It has been estimated that around 4-8% of the human genome is evolutionary constrained, of which coding elements contribute about 1.5% and rest is non-coding DNA (1-3). Through the observations from massive data produced by ENCODE and Epigenome roadmap projects, it has been shown that most of the evolutionary constrained non-coding DNA serve as protein binding sites and not as non-coding RNA (4,5). Protein coding genes are interwoven to these binding sites in a complex manner. Several lines of evidence converge to non-trivial regulatory impact of non-coding DNA proximal or distal to a target gene. The deletions or the malfunctioning of some of the distal regulatory elements have often shown observable phenotypes. 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 regulatory elements downstream to PAX6 gene are associated with Aniridia, a congenital eye malformation (3' deletions cause aniridia by preventing PAX6 gene expression). 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 conserved regulatory element of androgen receptor is strongly associated with evolutionary loss of penile spines and sensory vibrissae in human (11). The non-coding elements identified in these examples showed strong conservation when subjected to multi-species sequence alignments and most of the elements exhibited strong enhancer activity in *in-vivo* reporter assays.

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). However, establishing causal relationship between CNE and the phenotype remains a daunting task. Although genome wide association studies (GWAS) have uncovered a whole repertoire of non-coding variants with phenotypic associations (15), 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 (16). However the scope of these studies remains limited to a rather small number of candidate sites. Attempts have been made to link evolutionary loss or divergence of CNEs to lineage specific traits, like auditory system in echo-locating mammals and adaptively morphed pectoral flippers in marine mammals(17,18). While the loss and sequence divergence of CNEs has been subjected to thorough genome-wide analysis, the

CNE-to-gene synteny and lack thereof has not been studied in this context. Through comprehensive human-mouse comparison, it has been inferred that the CNEs generally regulate the nearest gene in linear proximity (14). Could there be phenotypic consequences if the proximity between CNE and its cognate gene is lost or gained in the evolution? In this study, we attempted to address the above question through comprehensive analysis of chromosomal position data of CNEs and genes across mammalian lineages. The study shows striking differences in functional associations of syntenic and non-syntenic CNE-gene pairs and highlights the significant contribution of loss and gain of CNE-gene proximity in the evolutionary divergence of developmental trajectories among mammals.

Materials and Methods

Compilation of chromosomal position data

Human (hg19), rat (rn5), rabbit (oryCun2), dog (camFam3), horse (equCab2), cow (bosTau6) and pig (susScr3) genome assemblies were used in the analysis. A total of 266115 Conserved Non-coding Elements (CNEs) were taken from Marcovitz et al (Marcovitz et al.2016), which in turn were obtained by curating mammalian CNEs anchored to the human genome (hg19). The CNE set was rigorously filtered for coding regions as per UCSC knownGene, Ensembl, Mammalian Gene Collection, Refseq, Exoniphy, VEGA, Yale Pseudogene Database, miRNA registry, snoRNA-LBME-db by the authors(18). Minimum length of CNE was set to 50 and all the CNEs within 20bp were merged to get the longer ones(18). We obtained the orthologue positions of human CNEs in query species using standard approach of mapping through LiftOver (https://genome-store.ucsc.edu/) chains at 0.95 mapping coverage(18). Finally, we compiled 114219 CNEs that were having orthologous positions in all the species. Next, the list of nearest genes that were within 100kb to the CNEs was obtained for human and corresponding orthologues in other mammals were obtained from Ensembl database. If there were multiple orthologues for the same gene, we took the nearest gene to the CNE on the same chromosome to ensure that a syntenic pair should not have classified as non-syntenic because of mapping errors. The CNE-gene pairs were considered as syntenic if they resided within the proximity (<100 kb in all 7 species) and non-syntenic when they are separated by more than 2Mb from each other or are located on different chromosomes in one of the mammalian species and remain proximal (< 1Mb) in rest of the 6 species and within 100kb in at least in two of the species. The distance cut-off of >2mb in one species and <1Mb in rest of the species assures reasonable linear separation of CNE and the gene independently in one of the species, while the cut-off of <100kb for atleast two of the other species assures that in some, if not all the rest, species CNE and the cognate gene remained syntenic. Further to avoid the erroneous orthologue mapping, we cross-checked the non-syntenic CNE-gene pairs manually through UCSC browser

(https://genome.ucsc.edu/), BLAT (https://genome.ucsc.edu/cgi-bin/hgBlat) searches and OMA database (http://omabrowser.org/). We removed the instances that exhibited discrepancies. This led to 21786 syntenic and 1036 non-syntenic CNE-gene pairs with 4454 and 265 unique genes respectively. Further, to nullify any bias due distinct representation of intronic CNEs in synentic and non-syntenic sets, we only focussed on inter-genic CNEs and the associated genes. The final set contained 9026 syntenic and 701 non-syntenic CNEs associated with 2349 and 184 unique genes respectively. A flow-chart illustrating the overall strategy is given in figure S13. All the data are available in supplementary data file.

Analysis of genomic attributes

Human genome sequence (hg19) was downloaded from UCSC and GC percentage was calculated +/- 50Kb around syntenic and non-syntenic CNEs. Average values of % GC in 2kb bins were used for the analysis. Chromosomal coordinates of repeat elements were downloaded from UCSC table browser for human assembly hg19. Repeat elements were mapped +/- 50kb around syntenic and non-syntenic CNEs and average value of enrichment in 2kb bins were plotted. For conservation analysis PhyloP scores of placental mammals (http://ccg.vital-it.ch/mga/hg19/phylop/phylop.html) were mapped +/- 5kb to CNE.

Functional enrichment analysis

Functional enrichment analysis was done using GREAT (http://bejerano.stanford.edu/great/public/html/), MamPHEA (http://evol.nhri.org.tw/phenome/index.jsp?platform=mmus) and Enrichr (http://amp.pharm.mssm.edu/Enrichr/) servers. GREAT analysis was done independently for one nearest gene within 100kb, two nearest genes within 100kb either side and for the syntenic vs. non-syntenic comparison by taking syntenic CNE coordinates as background and non-syntenic CNE coordinates as foreground. For MamPHEA and Enrichr analysis, list of the nearest genes within 100kb to CNE was considered. Tissue specificity analysis was done using TSEA (http://genetics.wustl.edu/jdlab/tsea/), CSEA (http://genetics.wustl.edu/jdlab/csea-tool-2/) and Bgee (http://bgee.org/?page=top_anat#/) tools. The tissue specificity index (pSI) score of a gene i in a tissue 1, over the given tissues $j=1,2,...m$ was calculated as per Dougherty et al (19) using following equation:

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

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 j. A stringent pSI cut-off of 0.0001 was taken for the anaysis. For syntenic gene set, we randomly sampled 184 genes (the size of non-syntenic gene set) from 2239 syntenic genes 100 times and plotted the mean and the standard error of the significance (-log10 of corrected p-value) of overlap between the candidate gene-sets and the tissues specific genes in the genome. The random sample of syntenic genes that exhibited most significant overlap with the brain specific genes was taken for the expression specificity analysis among brain tissues across developmental stages.

Normalized gene expression data for developing brain, heart and lung of human, mouse and rat were taken from BRAINSPAN (mixed brain tissues, http://www.brainspan.org/static/download.html), GSE21278 (hypothalamus), GSE71148(mixed left/right ventricles and atria), GSE51483 (mixed whole/left/right heart ventricles), GSE43767 (whole lung), GSE74243 (whole lung) and Stead et al (mixed brain tissues) respectively. For inter-species comparison of expression levels, the gene expression data from both the species were compiled into one table and quantile normalized to acquire the comparable distributions. Such strategy has been implemented earlier by others too (20).

Enhancer analysis

Datasets for H3K4me1 methylation for pre-natal and post-natal/adult human tissues were obtained from Epigenome Roadmap (http://www.roadmapepigenomics.org/data/) with following accession IDs and ages: fetal brain (E081; 17GW), adult brain (E067, E068, E069, E071, E072, E073, E074; pooled 73Yr/81Yr), fetal muscle (E090; 15GW), post-natal muscle (E100; pooled 3Yr/34Yr) and fetal thymus (E093; 15GW), post-natal thymus (E112; 3Yr). Fold-change over input DNA was used for aggregation plots. WashU epigenome browser was used for visualization. DNase-seq data for pre/post-natal tissues of human and mouse were downloaded from Epigenome Roadmap and ENCODE (https://www.encodeproject.org/matrix/?type=Experiment) with following accession IDs: human fetal brain (E081), human fetal heart (average of ENCSR366EGE & ENCSR911LTI), mouse fetal brain (ENCSR000COE) and mouse 0-day post-natal heart (ENCSR005WPU). CNE-to-neighborhood fold-change was calculated by averaging DNase-Seq signal 500bp +/- to CNE center and dividing it by average value in the neighbourhood (500-1500bp from CNE center). Data for in-vivo enhancer activity was obtained from VISTA enhancer database (https://enhancer.lbl.gov/). CNE coordinates were mapped onto the coordinates of total 2658 VISTA enhancers. Syntenic and non-syntenic CNEs exhibiting activity in different embryonic tissues were counted and compared statistically using

Fisher's exact test. Discriminant motif analysis was performed through MEME-suite package (http://meme-suite.org/) using JASPAR matrices for vertebrate genomes.

Mapping of proxy GWAS SNPs

Total of 251835 GWAS SNPs were obtained from GWASdb (http://jjwanglab.org/gwasdb). From this data, 71990 brain related SNPs were obtained by analyzing the HPO terms associated with brain associated phenotypes. Only 3 of the brain associated GWAS SNPs were mapped to non-syntenic CNEs. We therefore obtained the 600607 nearby SNPs (proxy) that were in linkage disequilibrium to GWAS SNPs based on 1000 genome data using SNAP algorithm(https://personal.broadinstitute.org/plin/snap/index.php), in addition to GWAS SNPs.

Analysis of Hi-C data

Coordinates of TAD domains in 22 distinct human cell-types were obtained from Schmitt et al(21). TAD coordinates for human fetal brain cortex were taken from Won et al(22). SRA files for Hi-C data of mouse brain cortex were downloaded from GEO accession (GSE34587). SRA files were converted into fastq files using fastq-dump (SRA Toolkit) with parameter "split-files" for separating paired reads into separate files. They were mapped onto mouse indexed reference genome (mm10) using bowtie2 with parameter "sensitive" for mapping. Paired reads were mapped separately and then joined together to get the Hi-C interaction pairs. The reads were binned into 250kb bins to get the chromatin interaction frequency. Density plot of Interaction frequency (log scale) was evaluated to define a rational cut-off of read-count 16 per 250 Kb bin to obtain total 143292 significant inter-chromosomal interactions (Figure S14). For significance analysis, we generated random null distribution by randomly sampling the same number of bin pairs as the original set while preserving the chromosomal distribution (i.e., CNE bin and cognate gene bin were taken from respective chromosomes) same as the original set. P- value was calculated using following equation:

$$p = \frac{1}{B}\left(1 + \sum_{b=1}^{B} (1)_{k_b > k}\right)$$

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

k_b = Number of resampled pairs exhibiting significant *trans* interactions.

k = Number of observed pairs exhibiting significant *trans* interactions.

DNA breakpoint data

Cancer associated DNA breakpoint data from 64 cancer genomes spanning over 7 different tumor types was downloaded from Malhotra, A. et al. (Malhotra, A. et al. 2013). We got 21389 germline

breakpoints and 5297 somatic breakpoints. These breakpoints were then mapped onto inter-spacer regions between CNE and the nearest gene-TSS. Fisher's exact test was used to check the statistical significance. Total 2061 evolutionary DNA break-points for rodents were taken from Bourque etal (2004 & 2006), Larkin etal and Lemaitre etal et al (23-26). These breakpoints were mapped to the inter-spacer regions between non-syntenic CNEs and the nearest gene-TSSs within 100kb in the human genome.

Availability of data. All datasets presented in this article are available in supplementary data file.

Results

Conservation of genomic proximity between CNE and the nearest gene

Using chromosomal position data of CNEs and genes from representative primate (human), rodent (rat), lagomorph (rabbit), carnivore (dog), perrisodactyl (horse) and artiodactyls (cow and pig), we obtained 9026 'syntenic' CNE-gene pairs, wherein the inter-genic CNE and the nearest gene-TSS were <100kb distance apart in all 7 species. There were 701 'non-syntenic' CNE-gene pairs, wherein the inter-genic CNE and the gene-TSS were >2Mb apart or were on different chromosomes independently in one of the non-human species (Materials and Methods). There were 2349 and 184 unique genes associated with syntenic and non-syntenic CNEs respectively (Figure 1a-c). Relatively large number of syntenic CNE-gene pairs (92.7%) highlighted the widespread conservation of linear proximity between CNE and its cognate gene (Figure 1b). We also confirmed that the CNE and the nearest gene in our final dataset were preferably (96.2%) located within the same topologically associated domain (TAD) in 22 human cell-types, suggesting that the arbitrary distant cut-off of 100kb was largely coherent with topology based compartmentalization of genome (Figure 1d, Table S1). To test if the CNEs in syntenic and non-syntenic sets were comparable, we assessed their lengths and degree of conservation in mammalian genomes. Figures 1e-f show insignificant differences in the length and the degree of sequence conservation of syntenic and non-syntenic CNEs, suggesting that the length and the sequence of non-syntenic CNEs had not diverged among mammals as compared to that of syntenic CNEs. Further, we overlaid several human genomic attributes on to syntenic and non-syntenic CNEs and observed that the syntenic CNEs were located in the regions of higher overall GC and SINE content as compared to non-syntenic CNEs in human genome (p-values: 7e-18 & 1e-17 respectively, Figure 1g), while non-syntenic CNEs were located primarily in regions enriched with LTRs, LINEs, satellite, simple repeat, DNA transposons and low complexity repeats (p-value=0.0005 to 1e-15, Figure 1g). These striking differences clearly suggested two things: 1) the syntenic CNEs were enriched in the region of open chromatin, as signified through

greater enrichment of GC and SINE content, and might have more wide-spread role across different cell-lineages as compared with the non-syntenic CNEs; 2) abundance of different kind of repeat elements around non-syntenic CNEs mark the susceptibility of the underlying loci for the genomic rearrangements through mechanisms like non-allelic homologous recombination (NAHR), which might explain the non-syntenic nature of these CNEs.

Functional and phenotypic attributes of genes that were syntenic and non-syntenic to their CNEs

Significant differences in the genomic attributes around syntenic and non-syntenic CNEs hinted at their distinct functional roles. To assess their functions, we first retrieved the lists of unique genes nearest (≤100kb) to each kind of CNE. These syntenic and non-syntenic gene-lists were then subjected to Mammalian Phenotype Ontology (MPO), Gene Ontology (GO) and pathway enrichment analysis (Material and Methods). As shown in the figure 2a, the analysis of MPO terms revealed enrichment of lethality and mortality related terms in syntenic set (p-value=0.0), while nervous system and craniofacial development related terms were enriched in non-syntenic set (p-value=1e-38 to 3e-57). We also confirmed these observations by comparing non-syntenic gene list to syntenic gene list while taking the latter as background dataset (p-value=1e-09 to 3e-24 , Figure 2b). Further, the syntenic genes were enriched with development related GO terms in general (p-value=0.0), while non-syntenic genes were enriched with nervous system related GO terms (p-value= 8e-14 to 1e-57, Figure S1a). GO term 'limb morphogeneis' also appeared among the top-10 functions, however the genes associated to this term were rather fewer as compared to the ones related to nervous system development (6% vs. 33% among top-10). Among pathways, evolutionarily conserved developmental pathways like Wnt, Integrin, TGF-β, Notch etc. and the growth pathways like IGF, FGF and PDGF were enriched in the syntenic gene-set (p-value=3e-25 to 1e-184, Figure S1b), while brain related pathways like axon guidance, glutamate receptor, acetylcholine receptor etc were enriched in non-syntenic gene-set (p-value=3e-03 to 3e-06, Figure S1b). These observations highlighted the developmentally 'ubiquitous' nature of syntenic genes and 'brain/craniofacial' association of non-syntenic genes. This conclusion was also supported by comparatively lower sequence divergence of proteins in syntenic set as compared to non-syntenic set (p-value= 1e-02, Figure 2c). Our observations were robust against the presumption that CNEs regulate the nearest proximal gene. We obtained the two proximal genes on either side of the CNE and subjected to functional analysis. The analysis confirmed the brain/craniofacial association of non-syntenic genes (p-value = 1e-24 to 9e-45 , Figures S2a-b).

We further followed the above observations through tissue-specific gene expression analysis. The syntenic set had widespread representation of genes expressed in different cell-lineages and therefore does not exhibit significant tissue specificity, while genes in non-syntenic set were specifically expressed in brain (p-value<0.01, Figures 2d-e). Marginal significance was observed for testes-specific expression of non-syntenic genes, which is likely due to known breadth of expression divergence of testes-specific genes among mammals aligning to the speciation through sexual conflict and sperm competition(27,28). The brain specific expression of genes in non-syntenic set was also confirmed through enrichment analysis of anatomical terms from *bgee* database (p-value=3e-07 to 3e-153 , Figure S3). Within brain, the genes in non-syntenic set were expressed in most brain tissues (p-value<0.05, Figures 2f-g). However, with respect to developmental time, the genes in non-syntenic set were specifically expressed during pre-natal development. In contrast, the genes in syntenic set did not exhibit any specificity for brain tissues and developmental stages (Figures 2f-g). Overall, these observations highlight pre-natal brain specific roles of genes that were non-syntenic to their proximal CNEs in non-primate species.

Distinct enhancer activities of syntenic and non-syntenic CNEs

To test whether the differences between syntenic and non-syntenic sets observed through functional analysis of genes, were coherent with the associated CNEs, we first tested the regulatory potential of CNEs through histone modification associated with enhancers, namely Histone-3-Lysine-4-mono-methylation (H3K4me1). We chose this mark because of the availability of genome-wide data for all the cell-lineages we were interested in. We observed that the: 1) CNEs in syntenic set exhibited consistent H3K4me1 enrichment across several fetal and adult tissues like thymus (endodermal), muscle (mesodermal) and brain (ectodermal) lineages (Figure 3a); 2) H3K4me1 enrichment over CNEs in non-syntenic set was specifically higher in fetal, but not in adult, brain only (Figure 3a).These observations were largely coherent with our proposal that the loss of CNE-gene synteny in the non-human lineage was associated with fetal brain development in human. Through analysis of in-vitro differentiation, we observed that unlike syntenic CNEs that had significant enrichment of H3K4me1 in embryonic stem cells and the derived cell-lineages, the non-sytenic CNEs acquired H3K4me1 enrichment specifically during differentiation of ES cells to neuronal progenitors and subsequently to neurons, and exhibited significantly lower enrichment during mesodermal, endodermal and trophoblast commitment, further strengthening their specific role in nervous system development (Figure 3b).

We validated the above observations through *in-vivo* enhancer assays. By mapping CNEs that were syntenic in all the species and the ones that were syntenic in human and mouse, but were non-syntenic in one of the other species, to VISTA enhancers, we showed that the non-syntenic CNEs had *in-vivo* activity primarily in facial mesenchyme, forebrain and nose (p-value=8e-03, 1e-02 & 2e-02 resp.), while the syntenic CNEs did not exhibit significant specificity towards brain tissues (Figure 3c-d). We have highlighted some of the cases in the figure 3b and figure S4. A non-syntenic CNE with brain specific activity was proximal to FOXG1 gene, which is involved in brain development and is strongly associated with severe intellectual disability, absent speech, autistic phenotypes, epilepsy and Rett syndrome (29). SP9, a zinc finger transcription factor having essential role in embryonic development of striatopallidal projection neurons (30) was accompanied by a non-syntenic CNE specifically active in forebrain, hindbrain and neural tube. CNE proximal to GRIP1 gene, a gene involved in synaptic targeting of AMPA receptor (31,32), was active in hindbrain and eye. Another brain specific CNE was located near FOXP4, which along with FOXP1 and FOXP2 regulates early neuronal development and is implicated in speech and language disorders (33,34). Similarly, CNE near ALX4, a gene important for craniofacial development, exhibited activity in eye and trigeminal V ganglion. Genetic errors at this locus are associated with Parietal Foramina, Frontonasal Dysplasia, Craniosynostosis etc (35). On the other hand, developmentally essential genes such as WNT5A, HHEX, LMO1, EMX2 etc (36-40) were syntenic to their cognate CNEs. These examples present visual snap-shots of the observed functional differences between syntenic and non-syntenic CNEs.

By mapping the trait/disease associated SNPs from Genome Wide Associated Studies (GWAS) and the nearby SNPs (proxy) in the linkage disequilibrium based on 1000 genome data, we observed that 23 of the non-syntenic CNEs were having at-least one brain related SNP. This representation was statistically significant when compared with that of syntenic set (p-value=1e-02, Figure 4a-b, Table S2). We highlight three of the cases, where GWAS SNPs mapping to non-syntenic CNEs were associated with amyotrophic lateral sclerosis (ALS), parent of origin language impairment and alcohol dependence traits (Figure 4c). It was interesting to note that these CNEs exhibited greater H3K4me1 enrichment in fetal brain as compared to adult brain suggesting that these disorders might have origin during fetal brain development. The genes proximal to these CNEs were GRIP1, MSX2 and NXPE2 respectively, which are known to be implicated in neurological disorders (41-43). These observations present genetic evidence of brain specific roles of non-syntenic CNEs in human.

Through motif finding algorithm MEME, we further observed that the non-syntenic CNEs were specifically enriched with binding sites of PAX3 and PAX7 transcription factors as compared to

syntenic CNEs (p-value= 4e-04 & 5e-04, Figure 4d). Interestingly, these factors were also expressed, but not restricted to, in brain tissues during fetal development, and not during post-natal development in concordance to our other observations (Figures 4e-f). PAX3 and PAX7 have been shown to be associated with neuronal development in vertebrates and implicate in several brain/behaviour related disorders in human (Figure S5).

We further illustrated the association of non-syntenic CNEs with the human fetal brain by plotting epigenomic profiles of some example loci using browser snap-shots (Figure 5, S6). As shown in the figure 4a, an upstream element to ADAM23 gene exhibited H3K4me1 and DNase-Seq peaks only in the fetal brain track and ADAM23 showed specific transcriptional activity in fetal brain as measured through RNA-Seq. Shown CNE was -18748 bp upstream to ADAM23 gene start site in human genome, but was distant by 2.4Mb in rat genome. The evolutionary expansion of distance between TSS and CNE was correlated with the chromosomal inversion of the region that contained the gene, but not the CNE. This rearrangement inverts the gene orientation in a manner that expands the distance between the TSS and the CNE. Similarly, in figure 4b, we showed an example of *trans* split of gene and the CNE. The CNE was located at -45287 bp upstream to POU3F2 gene in human, but was on different chromosome in rat. Again, the CNE and the gene exhibited correlated activity that was specific to fetal brain. We further compiled the quantile scaled gene expression data for human, mouse and rat fetal cortex cells and normalized by expression of tubulin-β gene. Interestingly, in both the examples that we cited, the orthologous genes exhibited higher expression level in fetal brain of human and mouse, where CNE was proximal to gene. The expression was significantly lower in fetal brain of rat, where CNEs and the genes were distant.

Therefore, our observations through epigenomic marks, *in-situ* enhancer assays and differential motif enrichment analysis concomitantly established that the non-syntenic CNEs were specific to pre-natal brain development in human.

Developmental tolerance and intolerance of CNE-gene split events

We have independently shown that the genes and CNEs in syntenic set had widespread role in development, it remained to be tested whether the proximity between CNE and the gene in syntenic set was important for the survival during development and hence not observed in any of the mammals analysed. To test this, we hypothesized that if the proximity between CNE and the gene was to be developmentally indispensable, then the DNA breakpoints of germ-line origin would be

underrepresented between CNE and the gene as compared to the scenario where the proximity between CNE and gene is developmentally dispensable. On the contrary, DNA break-points of somatic origin would not show any difference between two scenarios. Towards this, we obtained cancer associated DNA breakpoints of germ-line and somatic origins, spanning over 7 different tumour types. Figure 6 shows relative proportion of syntenic and non-syntenic CNE-gene pairs having at-least one germ-line and somatic DNA breakpoint between gene-TSS and the CNE. We observed significantly less representation of germ-line breakpoints in syntenic set as compared to non-syntenic set (p-value=8e-18), while representation of somatic breakpoints showed insignificant difference, aligning to our proposed hypothesis. These results highlight developmentally indispensable role of chromosomal synteny between CNE and the gene in the syntenic set and largely explains why loss of synteny for non-syntenic CNE-gene pairs survived in the evolution and might have served as a substrate for phenotypic changes.

Positional dynamics of non-syntenic CNEs was correlated with the tissue-specific fetal gene expression

An important question was whether or not loss of synteny between gene and CNE was associated with the loss of expression. Indeed, the genes shown in the browser snap-shots in figure 4 were specifically expressed in fetal brain of the species where CNE-gene synteny was maintained (human and mouse in this case) as compared to the species where synteny was lost (rat in this case), hinting that loss of CNE-gene synteny might correlate with loss of expression. To test this statistically, we obtained the CNE-gene pairs that were syntenic in human but non-syntenic (on different chromosomes) in mouse. Mouse was included in the comparison because of the availability of epigenomic profiles of its fetal and adult tissues. As expected the gene-list was enriched with nervous systems related MPO terms (43% of genes, p-value=4e-03, Figure 7a) and the associated CNEs exhibited relatively greater DNase-seq signal in human fetal brain, as compared to that of mouse (2.1 vs. 1.4 fold enrichment over neighbourhood, p=0.004, Figure 7b, Materials and Methods). Through comparison of gene expression data for developing brain of human and mouse, we observed the relative loss of foetus-specific gene expression in mouse as compared to that in human, suggesting that the loss of synteny correlates with the loss of foetus specific gene expression in developing brain (p-value=1e-93 & 0.4 for fetal specific expression in human and mouse brain resp., Figure 7c).

To find out what role might non-syntenic CNEs be playing at their rearranged loci in mouse, we focussed on the genes which gained proximity (<100kb) to non-syntenic CNEs in mouse. We

observed that evolutionary rearrangement of CNEs and gene positions in mouse shifted the proximity of CNEs from nervous system related genes in human to cardiovascular morphology related genes (30% of genes, p-value 7e-03, Figure 7d) in mouse. Further, the DNase-Seq data of fetal heart suggested greater activity of non-syntenic CNEs in mouse as compared to that in fetal heart of human, which was contrasting to the pattern observed for fetal brain in the Figure 6b (Figure 7e). We reconciled the observed functional shift by comparing gene expression data of developing brain and heart in human and mouse. As shown in the figure 6f, the genes that gained proximity to rearranged non-syntenic CNEs in mouse were primarily expressed in mouse fetal heart (p-value=1e-08), but not in human heart (p-value=0.8), suggesting a correlation between gain of linear proximity of CNEs to genes and the tissue specific fetal gene expression. As a control, we obtained the gene expression data of developing lung in human and mouse. We did not observe any foetus-specific expression in mouse lung (p-value=0.19, 0.07 resp. for human and mouse, Figure S7), suggesting that the observed gain of foetus-specific expression was constrained to mouse fetal heart. These observations largely explain why the CNE sequences were conserved despite the loss of synteny to the proximal genes. While it is not clear why the CNEs that were proximal to brain related genes in human were largely shifted to heart development related loci in mouse despite striking similarities between human and mouse heart development, appearance of cardiac electrical conductance related genes like RYR2, ATP1B1, and MYL4 in our data was noticeable. The developmental regulation of genes related to electrical conductance is highly species-specific(44). This observation also has physiological relevance of more efficient excitation/contraction coupling in mouse(45). Our earlier observation that the binding sites of PAX3 and PAX7 transcription factors, which exhibited expression in fetal brain, were enriched among non-syntenic CNEs is not incoherent in the present context because PAX factors are also known to be implicated in fetal heart development too(46-48). Indeed, we also observed the expression of PAX3 and PAX7 in fetal heart development of mouse, though at distinct developmental stage than that in fetal brain (Figure S8).

It is noteworthy that the observations through human-mouse analysis in figures 6c-f might not necessarily translate to comparisons between other mammals. We performed similar analysis for CNE-gene pairs that were proximal in rat, but distant (on different chromosomes) in mouse. We first confirmed that this gene-set also exhibited enrichment of nervous system related phenotypic terms (45% of genes, p-value=3e-02, Figure S9). However, surprisingly, we did not observe any loss of gene expression in fetal brain of mouse (p-value=1e-23, 1e-05 for rat and mouse resp., Figure 7g). The expression levels of orthologous genes in fetal brain of rat and mouse showed good correlation (Pearson's ρ=0.64. Figure 7h) as compared to that of human and mouse, which exhibited expression

bias towards human fetal brain (Pearson's ρ=0.37, Figure 7h). We hypothesize that since rat and mouse are evolutionarily closer than human and mouse, the overall three-dimensional organization of their genomes might not have diverged radically as in the case of human and mouse. By extrapolation, split CNE and the gene could still be communicating through spatial interactions. To test this hypothesis, we obtained the chromosome conformation data for human and mouse brain cortex, available in public domain. We first showed that the CNE and the gene in the human-mouse comparison were mostly within the same TAD in human fetal brain cortex (93%, Figure 8a). Corresponding CNE and the gene were located on different chromosomes in mouse and, therefore, we assessed their spatial connectivity by analysing inter-chromosomal (trans) interactions in mouse cortex genome. We did not observe significant proportion of CNE-gene pairs that exhibited *trans* interaction (p-value=0.2, Figure 8b). In contrast to human-mouse comparison, significant proportion of distant CNEs and the genes were spatially proximal in mouse in rat-mouse comparison (p-value<0.001), highlighting that the spatial proximity might have compensated for the lack of linear proximity between CNE and the gene (Figure 8c). These observations suggested that the observed position effect needed sufficient evolutionary divergence of higher order three dimensional organizations of genomes and might not be reflected in closely related species, like the species from the same order.

To further test the phenotypic association of position effect, we performed an inverse analysis wherein we focussed on the genes that were syntenic to their CNEs in rodents (mouse and rat), but were distant in human genome. Phenotype enrichment analysis of this gene-set suggested significant association with ear morphology, craniofacial morphology, cardiovascular and immune systems (p-value=0.02 to 0.01, Figure S10). Association with ear morphology was interesting given that the rodents have larger ears compared to body mass and have greater audible and frequency discrimination thresholds than primates (49). The analysis also validates the association of rearranged CNEs with heart related functions in rodents. Appearance of immune system related phenotype was an important observation owing to significant divergence of immunity related genes in rodents and primates (50).

Taken together, our analysis suggested a strong association between evolutionary dynamics of chromosomal positions of gene regulatory elements and the gain or loss of gene expression, aligning to the notion of 'position effect'. Moreover, tissue and developmental stage specific position effect observed in our analysis, highlights the possibility of its significant role in altering developmental dynamics towards evolutionary gain or loss of lineage/clade specific traits.

Discussion

It is not always the change in number and the sequence of proteins in the genome that leads to the phenotype alternation in evolution, the dynamics of gene expression is equally relevant in the context. One way the gene expression is altered is through position effect, i.e., relative chromosomal position of the gene in the genome can alter its expression through regulatory elements and chromatin states in the neighbourhood. Position effect was first discovered through the observation that the chromosomal arrangement of duplicated copies of *bar* gene in *bar*-mutant flies had influence on its expression and consequently causes the relative decrease in number of eye facets (51,52). Similarly, *white* gene when localized near heterochromatin gives mottled eye phenotype with red and white patches in drosophila eye (53). Despite its significance, the role of position effect in evolution of traits has not been investigated at large scale. Through comprehensive genomic analysis, we showed that the CNE-gene pairs that were syntenic in human but lost synteny in at-least one of the other mammalian species were specifically associated with pre-natal brain development in human, an observation that has several implications: 1) It is known that brain, as compared to other tissues, exhibits least genome-wide expression divergence across mammalian species(54,55), but within the space of our non-syntenic gene-set the expression divergence (particularly relative loss of expression) was observable. This suggested that the least expression divergence observed for brain might be due to cellular functions that are needed be precisely regulated to maintain delicately shaped brain tissues of all the mammals in general, while the ones that exhibit divergence would implicate in developmental functions specific to fetal brain. 2) The expression divergence of brain expressed genes in human and mouse is known to be primarily limited to fetal development and does not extend to post-natal stages(56,57). Coherently, our observations highlight the foetus specific role of non-coding regulatory elements, which might contribute to the observed expression divergence. 3) Several brain related diseases like Schizophrenia, Alzheimer and Autism are now considered to have origin during pre-natal brain development (58-63). Our analysis suggests that these disorders might also relate to genetic or epigenetic errors at noncoding regulatory elements flanking some of the genes that are expressed during pre-natal development. Indeed majority of the genes flanking non-syntenic CNEs were associated to some of these brain disorders (Figure S11).

Further, the evolutionary alteration in genomic proximity of genes and their regulatory elements was strikingly correlated with the alteration in the transcriptional program during fetal development, presenting evidence how the position effect would have impacted the evolution of lineage specific phenotypes by modulating the developmental trajectories in early stages. It was interesting to note

that the impact of such positional dynamics on transcription was observed when distant species like human and mouse were compared. Comparison of closely related species like rat and mouse suggested that the loss of linear proximity between CNE and the gene could have been compensated through spatial interactions in the nucleus. Indeed, it is well established that genomically rearranged loci remain spatially proximal in the evolution and in cancer genomes (64-67).

How would have been the relative positions of CNE and the cognate gene altered in the evolution? It would be reasonable to assume that the long range genomic rearrangement might have played role in repositioning of CNEs and genes. As shown in the figure 4a and S6, inversion of a large domain containing a gene, but not the proximal CNE, could increase the distance between the CNE and the promoter of the gene. The distance between CNE and the gene-promoter increased from ~20Kb in human genome to >2Mb in rat genome in the given examples. Quantifying such instances suggest that 25.8% of the *cis* alterations can partly be explained by inversion events. Similarly inter-chromosomal translocation would separate the gene and CNE to two different chromosomes. Mapping of evolutionary break-point data suggested 30.2 % of total human-rat inter-chromosomal changes were accounted by translocations. It was difficult to obtain direct evidence for segmental duplication followed by deletion or sequence divergence of one of the copies that might explain rest of the cases. In fact, it has been proposed as one of the main mechanisms that alters the gene order in the genome (68). We also argue that known evolutionary break-points among mammalian genome do not represent the complete space of all the break-point happened in the evolution, and therefore cannot be taken as entirety for this purpose.

It can be questioned that the structural variations that disrupt the CNE-gene synteny in evolution might be as common as within species variation and the non-syntenic CNE-gene pairs might not have any evolutionary significance. However, we argue that this might not be case because: 1) There was non-random association of only certain kind of biological functions with non-syntenic CNE-gene pairs, which strikingly coincided with the gain and loss of developmental stage specific transcriptional program. 2) Cancer associated germ-line DNA breakpoints between CNEs and the genes were present only in 20% of the non-syntenic cases. Removing these instances does not alter overall functional associations claimed in the study (Figure S12). Another artefact that might have possibly affected the observations is the error while mapping the orthologous genes across mammalian genomes. Issues like incorrect orthologue mapping and one-to-many orthologue mapping etc could lead to incorrect non-syntenic set. To ensure the correct mapping, we have subjected the non-syntenic set to a thorough manual scrutiny via BLAT searches across genomes.

For one-to-many orthologue mapping, we have taken the orthologue that has been on the same chromosome as the cognate CNE. In case there were multiple orthologues from the same chromosome as the CNE, we took the orthologue which was nearest to the CNE making sure that a syntenic CNE-gene pair should not get classified as non-syntenic because of redundant orthologue mappings. These additional measures, along with extensive manual curation through UCSC and WashU epigenome browsers, largely ensure the robustness of our observations against mapping artefacts.

Altogether, using relative positioning of conserved coding and non-coding elements and through large-scale epigenomic and functional data, we demonstrated the link between genome order and the evolutionary dynamics of temporal gene expression pattern associated with mammalian development. Our observations suggested that certain CNEs function as evolutionarily labile transcriptional accelerators having specificity towards certain tissues during fetal development and might have served as substrate for natural selection that subsequently fixed their preferred chromosomal positions towards optimization of certain traits in lineage- or clade-specific manner. Such approach might assist in annotating functionally diverged non-coding elements and might help explaining diverged phenotypic associations of orthologous non-coding sequences in different mammals. Our observations also establish the importance of genome order in comparative genomics studies, which has been underappreciated in the past.

Acknowledgement

Financial support to KSS from Department of Science and Technology (EMR/2015/001681) is duly acknowledged.

References

1. Rands, C.M., Meader, S., Ponting, C.P. and Lunter, G. (2014) 8.2% of the Human genome is constrained: variation in rates of turnover across functional element classes in the human lineage. *PLoS Genet*, **10**, e1004525.
2. Lindblad-Toh, K., Garber, M., Zuk, O., Lin, M.F., Parker, B.J., Washietl, S., Kheradpour, P., Ernst, J., Jordan, G., Mauceli, E. *et al.* (2011) A high-resolution map of human evolutionary constraint using 29 mammals. *Nature*, **478**, 476-482.
3. Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., Vernot, B. *et al.* (2012) The accessible chromatin landscape of the human genome. *Nature*, **489**, 75-82.
4. ENCODE. (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**, 57-74.
5. Skipper, M., Eccleston, A., Gray, N., Heemels, T., Le Bot, N., Marte, B. and Weiss, U. (2015) Presenting the epigenome roadmap. *Nature*, **518**, 313.

6. Loots, G.G., Kneissel, M., Keller, H., Baptist, M., Chang, J., Collette, N.M., Ovcharenko, D., Plajzer-Frick, I. and Rubin, E.M. (2005) Genomic deletion of a long-range bone enhancer misregulates sclerostin in Van Buchem disease. *Genome Res*, **15**, 928-935.
7. Sabherwal, N., Bangs, F., Roth, R., Weiss, B., Jantz, K., Tiecke, E., Hinkel, G.K., Spaich, C., Hauffa, B.P., van der Kamp, H. *et al.* (2007) Long-range conserved non-coding SHOX sequences regulate expression in developing chicken limb and are associated with short stature phenotypes in human patients. *Hum Mol Genet*, **16**, 210-222.
8. Driscoll, M.C., Dobkin, C.S. and Alter, B.P. (1989) Gamma delta beta-thalassemia due to a de novo mutation deleting the 5' beta-globin gene activation-region hypersensitive sites. *Proc Natl Acad Sci U S A*, **86**, 7470-7474.
9. Hatton, C.S., Wilkie, A.O., Drysdale, H.C., Wood, W.G., Vickers, M.A., Sharpe, J., Ayyub, H., Pretorius, I.M., Buckle, V.J. and Higgs, D.R. (1990) Alpha-thalassemia caused by a large (62 kb) deletion upstream of the human alpha globin gene cluster. *Blood*, **76**, 221-227.
10. Sparago, A., Cerrato, F., Vernucci, M., Ferrero, G.B., Silengo, M.C. and Riccio, A. (2004) Microdeletions in the human H19 DMR result in loss of IGF2 imprinting and Beckwith-Wiedemann syndrome. *Nat Genet*, **36**, 958-960.
11. McLean, C.Y., Reno, P.L., Pollen, A.A., Bassan, A.I., Capellini, T.D., Guenther, C., Indjeian, V.B., Lim, X., Menke, D.B., Schaar, B.T. *et al.* (2011) Human-specific loss of regulatory DNA and the evolution of human-specific traits. *Nature*, **471**, 216-219.
12. Seridi, L., Ryu, T. and Ravasi, T. (2014) Dynamic epigenetic control of highly conserved noncoding elements. *PLoS One*, **9**, e109326.
13. Roh, T.Y., Wei, G., Farrell, C.M. and Zhao, K. (2007) Genome-wide prediction of conserved and nonconserved enhancers by histone acetylation patterns. *Genome Res*, **17**, 74-81.
14. Babarinde, I.A. and Saitou, N. (2016) Genomic Locations of Conserved Noncoding Sequences and Their Proximal Protein-Coding Genes in Mammalian Expression Dynamics. *Mol Biol Evol*, **33**, 1807-1817.
15. Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., Klemm, A., Flicek, P., Manolio, T., Hindorff, L. *et al.* (2014) The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res*, **42**, D1001-1006.
16. Rajagopal, N., Srinivasan, S., Kooshesh, K., Guo, Y., Edwards, M.D., Banerjee, B., Syed, T., Emons, B.J., Gifford, D.K. and Sherwood, R.I. (2016) High-throughput mapping of regulatory DNA. *Nat Biotechnol*, **34**, 167-174.
17. Davies, K.T., Tsagkogeorga, G. and Rossiter, S.J. (2014) Divergent evolutionary rates in vertebrate and mammalian specific conserved non-coding elements (CNEs) in echolocating mammals. *BMC Evol Biol*, **14**, 261.
18. Marcovitz, A., Jia, R. and Bejerano, G. (2016) "Reverse Genomics" Predicts Function of Human Conserved Noncoding Elements. *Mol Biol Evol*, **33**, 1358-1369.
19. Dougherty, J.D., Schmidt, E.F., Nakajima, M. and Heintz, N. (2010) Analytical approaches to RNA profiling data for the identification of genes enriched in specific cells. *Nucleic Acids Res*, **38**, 4218-4230.
20. Tan, P.P., French, L. and Pavlidis, P. (2013) Neuron-Enriched Gene Expression Patterns are Regionally Anti-Correlated with Oligodendrocyte-Enriched Patterns in the Adult Mouse and Human Brain. *Front Neurosci*, **7**, 5.
21. Schmitt, A.D., Hu, M., Jung, I., Xu, Z., Qiu, Y., Tan, C.L., Li, Y., Lin, S., Lin, Y., Barr, C.L. *et al.* (2016) A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome. *Cell Rep*, **17**, 2042-2059.
22. Won, H., de la Torre-Ubieta, L., Stein, J.L., Parikshak, N.N., Huang, J., Opland, C.K., Gandal, M.J., Sutton, G.J., Hormozdiari, F., Lu, D. *et al.* (2016) Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature*, **538**, 523-527.

23. Larkin, D.M., Pape, G., Donthu, R., Auvil, L., Welge, M. and Lewin, H.A. (2009) Breakpoint regions and homologous synteny blocks in chromosomes have different evolutionary histories. *Genome Res*, **19**, 770-777.
24. Bourque, G., Zdobnov, E.M., Bork, P., Pevzner, P.A. and Tesler, G. (2005) Comparative architectures of mammalian and chicken genomes reveal highly variable rates of genomic rearrangements across different lineages. *Genome Res*, **15**, 98-110.
25. Bourque, G., Pevzner, P.A. and Tesler, G. (2004) Reconstructing the genomic architecture of ancestral mammals: lessons from human, mouse, and rat genomes. *Genome Res*, **14**, 507-516.
26. Lemaitre, C., Zaghloul, L., Sagot, M.F., Gautier, C., Arneodo, A., Tannier, E. and Audit, B. (2009) Analysis of fine-scale mammalian evolutionary breakpoints provides new insight into their relation to genome organisation. *BMC Genomics*, **10**, 335.
27. Turner, L.M., Chuong, E.B. and Hoekstra, H.E. (2008) Comparative analysis of testis protein evolution in rodents. *Genetics*, **179**, 2075-2089.
28. Khaitovich, P., Hellmann, I., Enard, W., Nowick, K., Leinweber, M., Franz, H., Weiss, G., Lachmann, M. and Paabo, S. (2005) Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. *Science*, **309**, 1850-1854.
29. Seltzer, L.E., Ma, M., Ahmed, S., Bertrand, M., Dobyns, W.B., Wheless, J. and Paciorkowski, A.R. (2014) Epilepsy and outcome in FOXG1-related disorders. *Epilepsia*, **55**, 1292-1300.
30. Zhang, Q., Zhang, Y., Wang, C., Xu, Z., Liang, Q., An, L., Li, J., Liu, Z., You, Y., He, M. *et al.* (2016) The Zinc Finger Transcription Factor Sp9 Is Required for the Development of Striatopallidal Projection Neurons. *Cell Rep*, **16**, 1431-1444.
31. Setou, M., Seog, D.H., Tanaka, Y., Kanai, Y., Takei, Y., Kawagishi, M. and Hirokawa, N. (2002) Glutamate-receptor-interacting protein GRIP1 directly steers kinesin to dendrites. *Nature*, **417**, 83-87.
32. Geiger, J.C., Lipka, J., Segura, I., Hoyer, S., Schlager, M.A., Wulf, P.S., Weinges, S., Demmers, J., Hoogenraad, C.C. and Acker-Palmer, A. (2014) The GRIP1/14-3-3 pathway coordinates cargo trafficking and dendrite development. *Dev Cell*, **28**, 381-393.
33. Takahashi, K., Liu, F.C., Hirokawa, K. and Takahashi, H. (2008) Expression of Foxp4 in the developing and adult rat forebrain. *J Neurosci Res*, **86**, 3106-3116.
34. Sin, C., Li, H. and Crawford, D.A. (2015) Transcriptional regulation by FOXP1, FOXP2, and FOXP4 dimerization. *J Mol Neurosci*, **55**, 437-448.
35. Kayserili, H., Uz, E., Niessen, C., Vargel, I., Alanay, Y., Tuncbilek, G., Yigit, G., Uyguner, O., Candan, S., Okur, H. *et al.* (2009) ALX4 dysfunction disrupts craniofacial and epidermal development. *Hum Mol Genet*, **18**, 4357-4366.
36. Goodings, C., Smith, E., Mathias, E., Elliott, N., Cleveland, S.M., Tripathi, R.M., Layer, J.H., Chen, X., Guo, Y., Shyr, Y. *et al.* (2015) Hhex is Required at Multiple Stages of Adult Hematopoietic Stem and Progenitor Cell Differentiation. *Stem Cells*, **33**, 2628-2641.
37. Hunter, M.P., Wilson, C.M., Jiang, X., Cong, R., Vasavada, H., Kaestner, K.H. and Bogue, C.W. (2007) The homeobox gene Hhex is essential for proper hepatoblast differentiation and bile duct morphogenesis. *Dev Biol*, **308**, 355-367.
38. van Amerongen, R. and Berns, A. (2006) Knockout mouse models to study Wnt signal transduction. *Trends Genet*, **22**, 678-689.
39. Tse, E., Smith, A.J., Hunt, S., Lavenir, I., Forster, A., Warren, A.J., Grutz, G., Foroni, L., Carlton, M.B., Colledge, W.H. *et al.* (2004) Null mutation of the Lmo4 gene or a combined null mutation of the Lmo1/Lmo3 genes causes perinatal lethality, and Lmo4 controls neural tube development in mice. *Mol Cell Biol*, **24**, 2063-2073.
40. Miyamoto, N., Yoshida, M., Kuratani, S., Matsuo, I. and Aizawa, S. (1997) Defects of urogenital development in mice lacking Emx2. *Development*, **124**, 1653-1664.
41. Lai, C., Xie, C., McCormack, S.G., Chiang, H.C., Michalak, M.K., Lin, X., Chandran, J., Shim, H., Shimoji, M., Cookson, M.R. *et al.* (2006) Amyotrophic lateral sclerosis 2-deficiency leads to

neuronal degeneration in amyotrophic lateral sclerosis through altered AMPA receptor trafficking. *J Neurosci*, **26**, 11798-11806.

42. Benitez-Burraco, A., Lattanzi, W. and Murphy, E. (2016) Language Impairments in ASD Resulting from a Failed Domestication of the Human Brain. *Front Neurosci*, **10**, 373.
43. Toma, C., Torrico, B., Hervas, A., Valdes-Mas, R., Tristan-Noguero, A., Padillo, V., Maristany, M., Salgado, M., Arenas, C., Puente, X.S. *et al.* (2014) Exome sequencing in multiplex autism families suggests a major role for heterozygous truncating mutations. *Mol Psychiatry*, **19**, 784-790.
44. Zimmer, T., Haufe, V. and Blechschmidt, S. (2014) Voltage-gated sodium channels in the mammalian heart. *Glob Cardiol Sci Pract*, **2014**, 449-463.
45. Milani-Nejad, N. and Janssen, P.M. (2014) Small and large animal models in cardiac contraction research: advantages and disadvantages. *Pharmacol Ther*, **141**, 235-249.
46. Blake, J.A. and Ziman, M.R. (2014) Pax genes: regulators of lineage specification and progenitor cell maintenance. *Development*, **141**, 737-751.
47. Olaopa, M., Zhou, H.M., Snider, P., Wang, J., Schwartz, R.J., Moon, A.M. and Conway, S.J. (2011) Pax3 is essential for normal cardiac neural crest morphogenesis but is not required during migration nor outflow tract septation. *Dev Biol*, **356**, 308-322.
48. Conway, S.J., Henderson, D.J. and Copp, A.J. (1997) Pax3 is required for cardiac neural crest migration in the mouse: evidence from the splotch (Sp2H) mutant. *Development*, **124**, 505-514.
49. Fay, R.R. (1988) Comparative psychoacoustics. *Hear Res*, **34**, 295-305.
50. Emes, R.D., Goodstadt, L., Winter, E.E. and Ponting, C.P. (2003) Comparison of the genomes of human and mouse lays the foundation of genome zoology. *Hum Mol Genet*, **12**, 701-709.
51. Sturtevant, A.H. (1928) A Further Study of the so-Called Mutation at the Bar Locus of Drosophila. *Genetics*, **13**, 401-409.
52. Sturtevant, A.H. (1925) The Effects of Unequal Crossing over at the Bar Locus in Drosophila. *Genetics*, **10**, 117-147.
53. Martin-Morris, L.E., Csink, A.K., Dorer, D.R., Talbert, P.B. and Henikoff, S. (1997) Heterochromatic trans-inactivation of Drosophila white transgenes. *Genetics*, **147**, 671-677.
54. Khaitovich, P., Enard, W., Lachmann, M. and Paabo, S. (2006) Evolution of primate gene expression. *Nat Rev Genet*, **7**, 693-702.
55. Strand, A.D., Aragaki, A.K., Baquet, Z.C., Hodges, A., Cunningham, P., Holmans, P., Jones, K.R., Jones, L., Kooperberg, C. and Olson, J.M. (2007) Conservation of regional gene expression in mouse and human brain. *PLoS Genet*, **3**, e59.
56. Clancy, B., Darlington, R.B. and Finlay, B.L. (2001) Translating developmental time across mammalian species. *Neuroscience*, **105**, 7-17.
57. Liscovitch, N. and Chechik, G. (2013) Specialization of gene expression during mouse brain development. *PLoS Comput Biol*, **9**, e1003185.
58. Vladeanu, M., Giuffrida, O. and Bourne, V.J. (2014) Prenatal sex hormone exposure and risk of Alzheimer disease: a pilot study using the 2D:4D digit length ratio. *Cogn Behav Neurol*, **27**, 102-106.
59. Stoner, R., Chow, M.L., Boyle, M.P., Sunkin, S.M., Mouton, P.R., Roy, S., Wynshaw-Boris, A., Colamarino, S.A., Lein, E.S. and Courchesne, E. (2014) Patches of disorganization in the neocortex of children with autism. *N Engl J Med*, **370**, 1209-1219.
60. Guillemot, J., Lukaszewski, M.A., Montel, V., Delahaye, F., Mayeur, S., Laborie, C., Dickes-Coopman, A., Dutriez-Casteloot, I., Lesage, J., Breton, C. *et al.* (2011) Influence of prenatal undernutrition on the effects of clozapine and aripiprazole in the adult male rats: relevance to a neurodevelopmental origin of schizophrenia? *Eur J Pharmacol*, **667**, 402-409.
61. Winter, C., Djodari-Irani, A., Sohr, R., Morgenstern, R., Feldon, J., Juckel, G. and Meyer, U. (2009) Prenatal immune activation leads to multiple changes in basal neurotransmitter

levels in the adult brain: implications for brain disorders of neurodevelopmental origin such as schizophrenia. *Int J Neuropsychopharmacol*, **12**, 513-524.

62. Torrey, E.F., Taylor, E.H., Bracha, H.S., Bowler, A.E., McNeil, T.F., Rawlings, R.R., Quinn, P.O., Bigelow, L.B., Rickler, K., Sjostrom, K. *et al.* (1994) Prenatal origin of schizophrenia in a subgroup of discordant monozygotic twins. *Schizophr Bull*, **20**, 423-432.
63. (2003) Transcription of live discussion: Alzheimer genes in cortical development: how do their prenatal functions relate to dementia? *J Alzheimers Dis*, **5**, 337-341.
64. Engreitz, J.M., Agarwala, V. and Mirny, L.A. (2012) Three-dimensional genome architecture influences partner selection for chromosomal translocations in human disease. *PLoS One*, **7**, e44196.
65. Roix, J.J., McQueen, P.G., Munson, P.J., Parada, L.A. and Misteli, T. (2003) Spatial proximity of translocation-prone gene loci in human lymphomas. *Nat Genet*, **34**, 287-291.
66. Martin, L.D., Harizanova, J., Zhu, G., Righolt, C.H., Belch, A.R., Mai, S. and Pilarski, L.M. (2012) Differential positioning and close spatial proximity of translocation-prone genes in nonmalignant B-cells from multiple myeloma patients. *Genes Chromosomes Cancer*, **51**, 727-742.
67. Veron, A.S., Lemaitre, C., Gautier, C., Lacroix, V. and Sagot, M.F. (2011) Close 3D proximity of evolutionary breakpoints argues for the notion of spatial synteny. *BMC Genomics*, **12**, 303.
68. Fischer, G., Neuveglise, C., Durrens, P., Gaillardin, C. and Dujon, B. (2001) Evolution of gene order in the genomes of two related yeast species. *Genome Res*, **11**, 2009-2019.

Figure legends

Figure 1. Synteny and lack thereof between CNE and the nearest gene. (a) CNE and the nearest gene within 100kb in human genome were mapped onto genomes of other mammals. The CNE-gene pairs were considered as syntenic if they resided within the proximity (<100 kb in all 7 species) and non-syntenic when they departed by more than 2Mb from each other or were located on different chromosomes in one of the mammalian species and remained proximal (<1Mb) in rest 6 species and within 100kb in at-least two of the other species. (b) Pie-charts representing relative numbers of CNEs and genes in syntenic and non-syntenic sets. (c) Binary heatmaps (brown for synteny and yellow for lack of synteny) show instances of CNE-gene pairs that had lost synteny in one of species. (d) Percentage of human CNE-gene pairs (<100kb) localized within and across topologically associated domains (TADs). Percentage shown is the average across 22 human cell-types (Table S1). (e) Violin plot shows length distribution of CNEs in syntenic and non-syntenic sets. (f) Relative sequence conservation, as measured through mammalian PhyloP scores, of CNEs in syntenic and non-syntenic sets. (g) Enrichment of GC content and repeat elements around syntenic and non-syntenic CNEs. Question mark (?) against certain repeat types signifies moderate sequence match with the corresponding family of repeat element. P-values were calculated using Mann Whitney U test.

Figure 2. ***Functional characterization of genes in syntenic and non-syntenic sets.*** (a) Enrichment of Mammalian Phenotype Ontology (MPO) terms among genes in syntenic and non-syntenic sets. (b) Enrichment of MPO terms among non-syntenic set when genes from syntenic set were taken as background. P-values of Fisher's exact test were corrected using Benjamini-Hochberg method. (c) Sequence divergence of syntenic

(across 7 mammals) and non-syntenic (rat) CNEs. P-value was calculated using *Mann Whitney U test*. (d-e) Tissue specific expression of genes in (d) syntenic and (e) non-syntenic set. Relative significance was plotted as negative of log10 transformed corrected p-values of Fisher's exact test for the overlap with the tissue specific genes at stringency score (pSI) < 0.0001. Vertical grey colored line represents the p-value of 0.01. For syntenic set, mean values and standard errors of significance for 100 random samples of 184 genes (the size of non-syntenic set) from syntenic sets were plotted. (f-g) Expression specificity of genes in (f) syntenic and (g) non-syntenic set across brain regions and across developmental stages. For sytenic set, the sample that exhibited maximum significance for brain specificity in panel-d was taken. Size of the nested hexagons represents the proportion of all genes specifically expressed in particular tissue at particular developmental stage. Hexagons are nested inwards based on relative stringency of tissue specificity scores (pSI=0.05, 0.01, 0.001 & 0.0001 respectively). Color gradient represents the magnitude of corrected p-values of Fisher's exact test.

Figure 3. Enhancer property of syntenic and non-syntenic CNEs. H3K4me1 enrichment on and around intergenic CNEs in syntenic and non-syntenic sets (a) for fetal and adult tissues, and (b) for *in-vitro* cultured embryonic stem cells and derived linreages. P-values for difference between syntenic and non-syntenic CNEs were calculated using Mann Whitney U test for the H3K4me1 enrichment values in 1kb spanning windows on either side of the center of the CNEs. (c-d) *In-vivo* validation of regulatory potential through VISTA enhancers that overlapped with the CNEs that were syntenic in human and mouse but non-syntenic in one of the other species (yellow) and the ones that were syntenic in all the species (brown) were taken for this analysis. (c) Top: Relative % of syntenic and non-syntenic CNEs that were active in different embryonic tissues. Bottom: Cartoon representation of relative enrichment of enhancer activity of non-syntenic CNEs over syntenic CNEs across mouse embryonic tissues. (d) Examples of In-vivo activity patterns in different parts of mouse E11.5 embryos. Specific embryonic tissues that exhibited the activity of CNEs are mentioned below each image. The number of embryos exhibiting activity and the total number of embryos examined are given in brackets against each embryonic tissue. The nearest gene is highlighted on the upper right corner of each panel.

Figure 4. Regulatory potential of syntenic and non-syntenic CNEs. (a) Proportion of syntenic and non-syntenic CNEs containing at-least one GWAS or proxy SNP. P-value was calculated using boot-strap method by randomly sampling 701 CNEs (size of non-syntenic set) from syntenic set 1000 times. (b) Examples of brain associated GWAS SNPs mapping to non-syntenic CNEs. Shown are the tracks of conservation (placental PhyloP score), H3K4me1 enrichment in fetal and adult human brain and the repeat elements. (c) Significantly enriched motifs in non-syntenic CNEs as compared to syntenic CNEs. The analysis was done using MEME-suite package and JASPAR matrices. P-values were calculated using Fisher's exact test and were corrected for multiple comparisons using Benjamini-Hochberg method. (d) Expression of transcription factors enriched among non-syntenic CNEs across different developmental stages and in different parts of human brain. Labeled brain samples exhibited consistently higher expression of both the factors in prenatal human brain. URL: Upper rhombic lip, CB: Cerebellum, CBC: Cerebrum Cortex. (f) Upper panel: gene expression pattern of Pax3 (left) and Pax7 (right) in different brain regions across different developmental stages in mouse. Lower panel: Cross-

sections of mouse embryos (E13.5) showing the spatial expression of patterns of Pax3 and Pax7. RSP: rostral secondary prosencephalon, Tel: telencephalic vesicle, PedHy: preduncular hypothalamus P1: prosmere 1 (pretectum), P2: prosomere 2 (thalamus), P3: prosomere 3 (pre-thalamus), M: midbrain, PPH: prepontine hindbrain, PH: pontine hindbrain, PMH: pontomeduallry hindbrain, MH: medullary hindbrain.

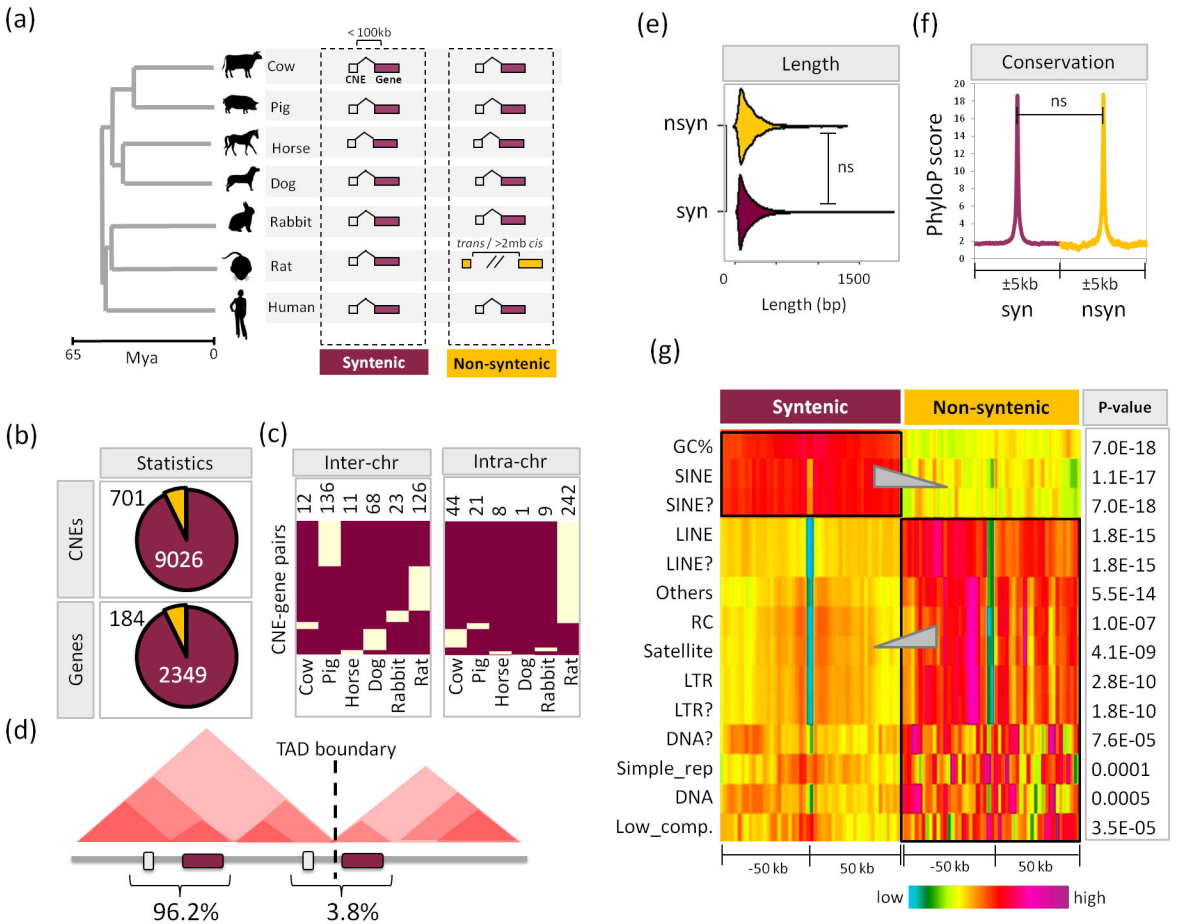
Figure 5. Illustrative examples of chromosomal position effect in cis and trans. Examples of CNE and the cognate gene that were within 100 kb in human genome, but were (a) >2Mb apart or (b) on different chromosome (*trans*) in rat genome. Shown are the different epigenomic tracks of H3K4me1, DNase hypersensitive sites, chromatin state (red: active TSS, orange: flanking active TSS, green: transcribed, yellow: enhancer, grey: repressed; detailed color-codes are given in Figure S6), and RNA-Seq in fetal brain, fetal leg muscle, fetal trunk muscle, fetal thymus and fetal lung. The bottom most tracks are of UCSC genes and human-rat synteny maps. (c) On the left, the strategy for gene-expression comparison of orthologous genes is depicted as cartoon. Bar-plots on right side show quantile normalized expression of ADAM3 and OU3F2 genes (relative to tubulin-β gene) in human, mouse and rat. The CNE-gene pairs shown here were syntenic in human and mouse genome, but were non-syntenic in rat.

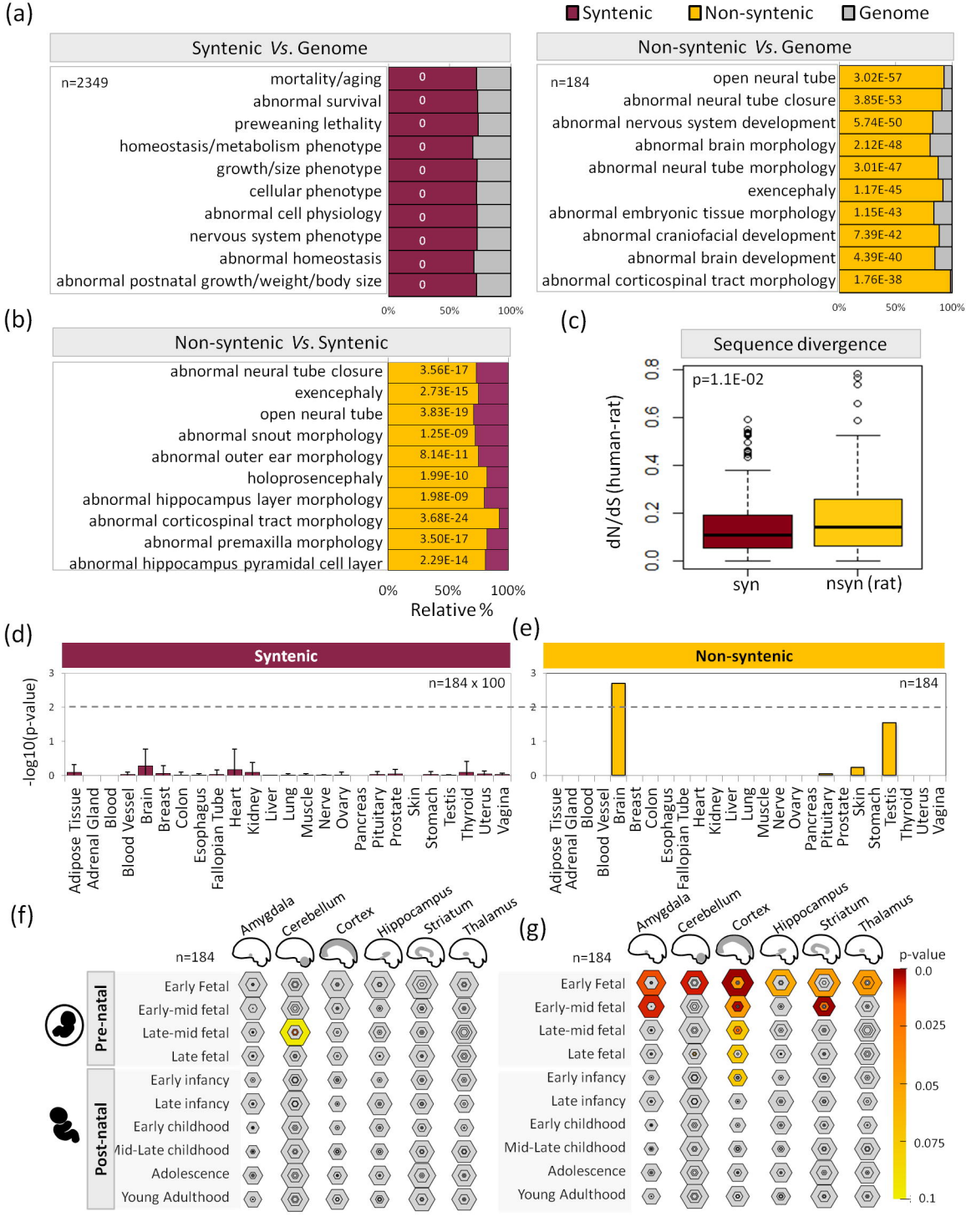
Figure 6. Tolerance and intolerance of CNE-gene split. Proportion of CNEs-gene pair having at-least one germ-line or somatic DNA breakpoint associated with cancer, in-between CNE and the gene-TSS. Fisher's exact test was used to calculate the p-values.

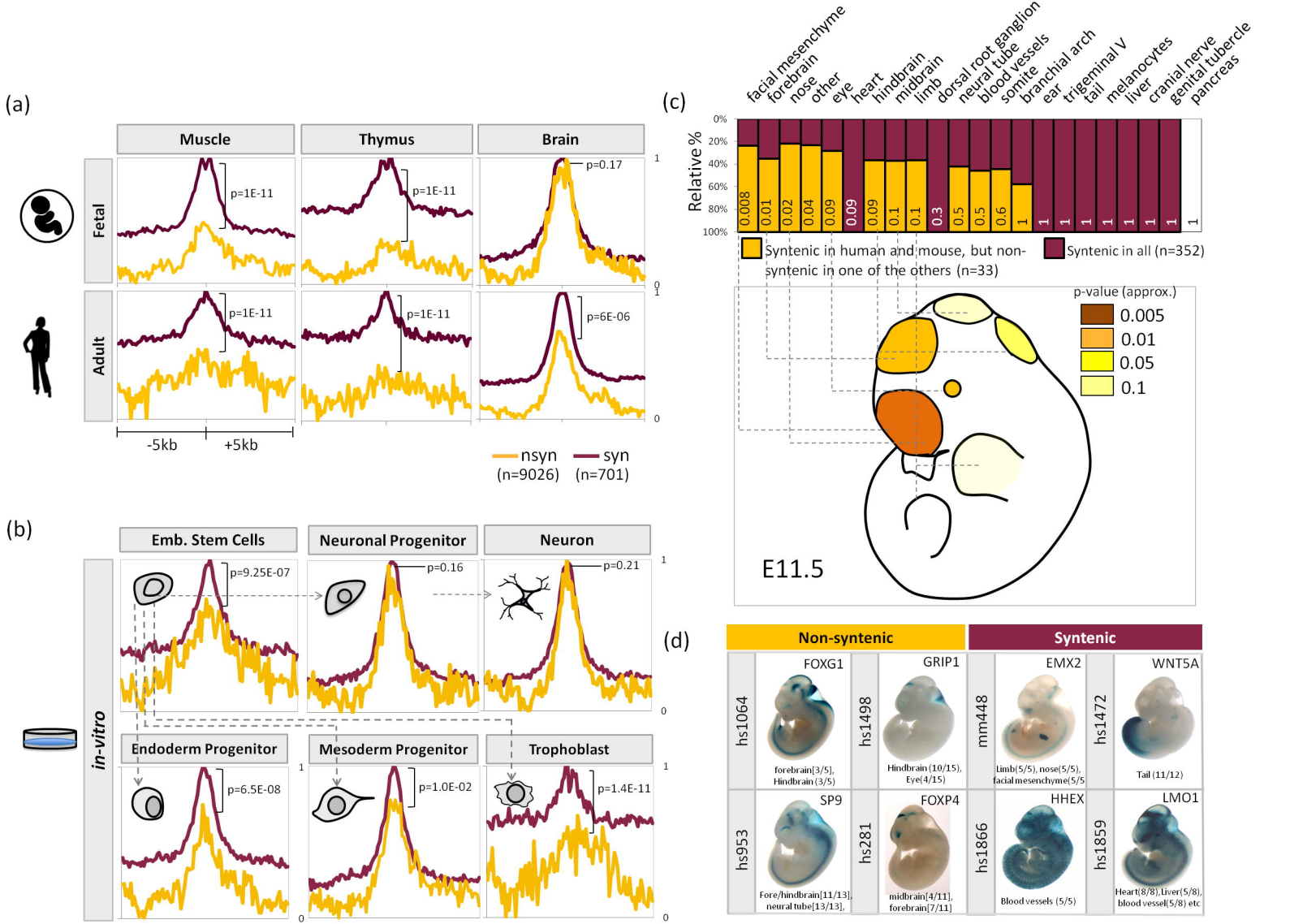
Figure 7. Functional fate of genes proximal to positionally rearranged CNEs in human, mouse and rat. (a) Enrichment of Mammalian Phenotype Ontology terms among genes that were proximal to CNEs in human, but were distant (other chromosomes) in mouse (b) CNE-to-neighborhood fold-change in DNase-seq signal on and around CNEs in fetal brain of human and mouse. (c) Expression dynamics of genes, that were proximal to CNEs in human but distant (different chromosome) in mouse, in developing brain of human (left panel) and of mouse (right panel). Curves represent the median expression of genes proximal to rearranged CNEs and bars represent standard errors. Heatmaps represent the complete expression data across all developmental stages. (d) Enrichment of Mammalian Phenotype Ontologies of genes which gained proximity to rearranged CNEs in mouse. (e) CNE-to-neighborhood fold-change in DNase-seq signal on and around CNEs in fetal heart of human and mouse. (f) Expression dynamics of genes, that gained proximity to CNEs in mouse, in developing heart of human (left panel) and of mouse (right panel). As a control, we analyzed gene expression data for developing human and mouse lungs (Figure S7). (g) Same as panel-c, but for rat and mouse comparison. (h) Contour plots for expression values of orthologous genes in fetal brain cortex of human (16-18 gestational weeks) vs. mouse (E18) (left) and rat (E18) vs. mouse (E18) (right).

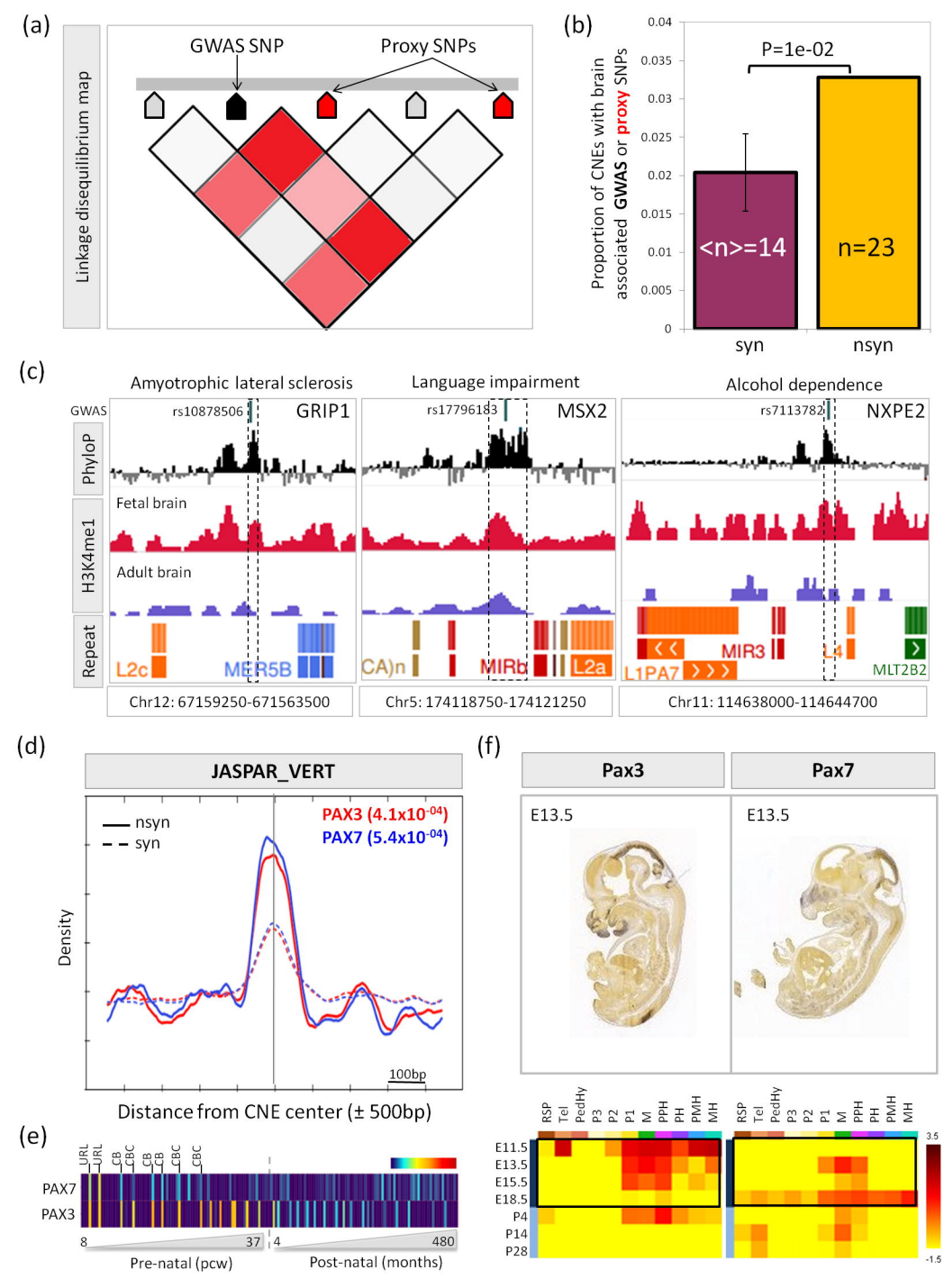
Figure 8. Chromosomal position effect and 3D genome organization (a) Percentage of CNE-gene pairs (syntenic in human and non-syntenic in mouse) that localized within the TAD and the ones that were across the TAD boundary in human fetal cortex. (b) Upper panel: Circos plot of spatial connectivity between genes

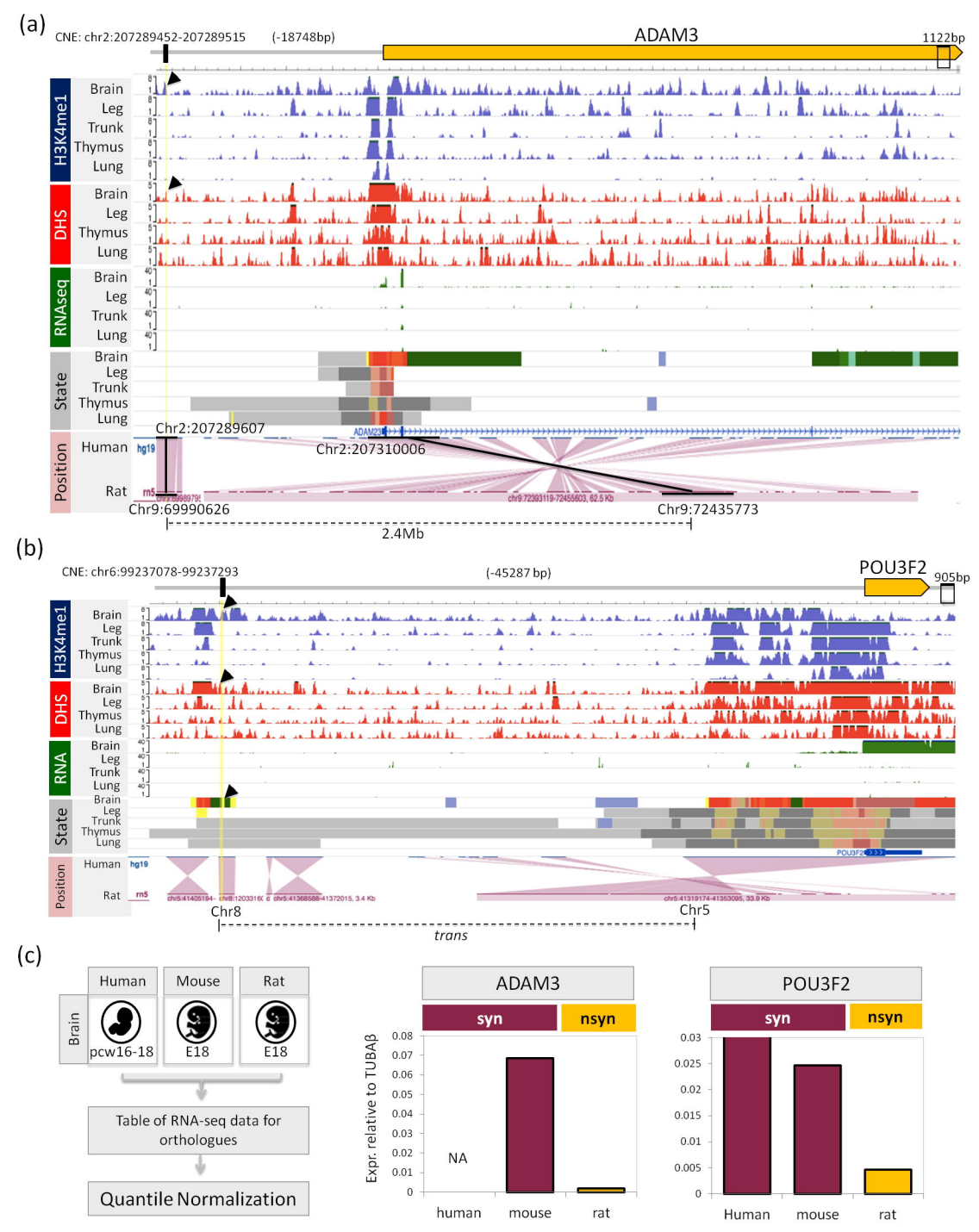
and their cognate CNEs that were proximal in human but were on different chromosomes in mouse. Grey colored edges represents the gene-CNE pairs, while black colored edges signify the CNE and the genes that exhibited spatial proximity in nuclei of mouse brain cortical cells. Lower panel: Observed number of *trans* interactions (vertical bar) overlaid upon the null distribution generated through bootstrap method. (c) Same as panel-b, but for rat-mouse comparison (d) Cartoon depicting interpretation of the data in Figure 6 and 7.

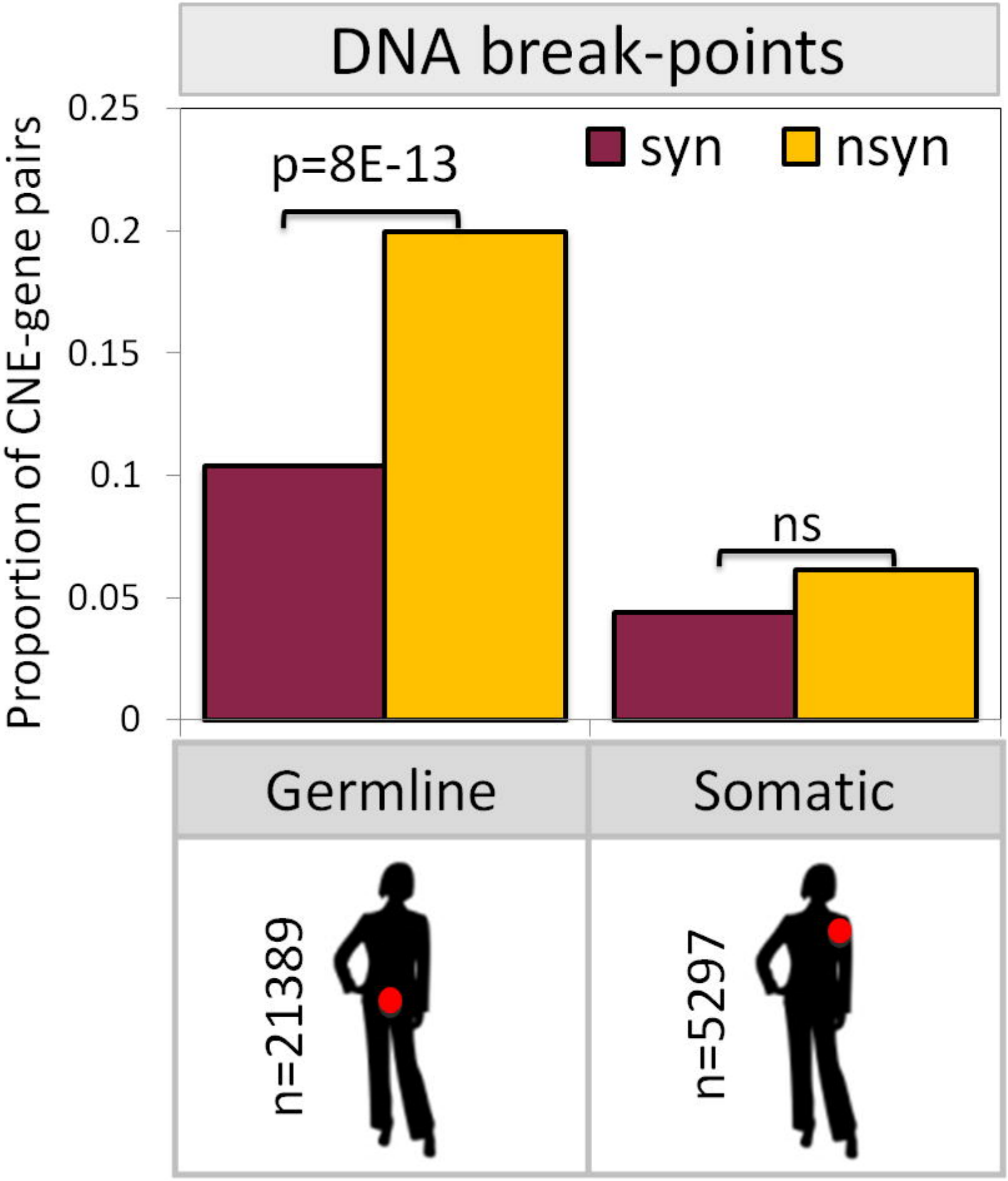

(a)
< 100kb
CNE
Gene
Cow
Pig
Horse
Dog
Rabbit
Rat
Human
trans / >2mb cis
65
Mya
0
Syntenic
Non-syntenic
(b)
Statistics
CNEs
Genes
701
9026
184
2349
(c)
Inter-chr
Intra-chr
CNE-gene pairs
12
136
11
68
23
126
44
21
8
1
9
242
Cow
Pig
Horse
Dog
Rabbit
Rat
(d)
TAD boundary
96.2%
3.8%
(e)
Length
nsyn
syn
ns
0
1500
Length (bp)
(f)
Conservation
PhyloP score
ns
±5kb
syn
±5kb
nsyn
(g)
Syntenic
Non-syntenic
P-value
GC%
SINE
SINE?
LINE
LINE?
Others
RC
Satellite
LTR
LTR?
DNA?
Simple_rep
DNA
Low_comp.
7.0E-18
1.1E-17
7.0E-18
1.8E-15
1.8E-15
5.5E-14
1.0E-07
4.1E-09
2.8E-10
1.8E-10
7.6E-05
0.0001
0.0005
3.5E-05
-50 kb
50 kb
-50 kb
50 kb
low
high

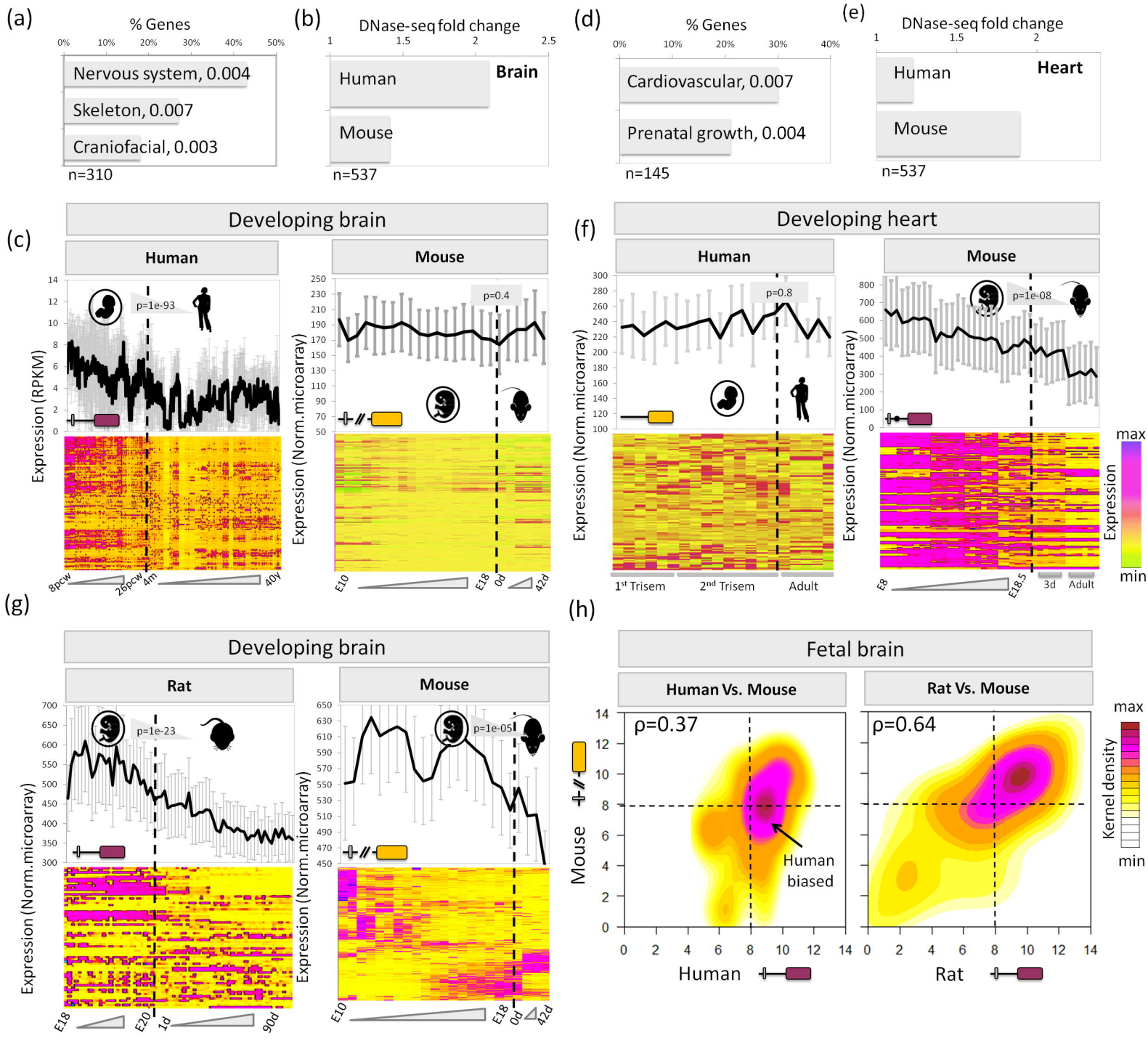

(a)
Syntenic
Non-syntenic
Genome
Syntenic Vs. Genome
n=2349
mortality/aging
abnormal survival
preweaning lethality
homeostasis/metabolism phenotype
growth/size phenotype
cellular phenotype
abnormal cell physiology
nervous system phenotype
abnormal homeostasis
abnormal postnatal growth/weight/body size
0
0%
50%
100%
Non-syntenic Vs. Genome
n=184
open neural tube
abnormal neural tube closure
abnormal nervous system development
abnormal brain morphology
abnormal neural tube morphology
exencephaly
abnormal embryonic tissue morphology
abnormal craniofacial development
abnormal brain development
abnormal corticospinal tract morphology
3.02E-57
3.85E-53
5.74E-50
2.12E-48
3.01E-47
1.17E-45
1.15E-43
7.39E-42
4.39E-40
1.76E-38
(b)
Non-syntenic Vs. Syntenic
abnormal neural tube closure
exencephaly
open neural tube
abnormal snout morphology
abnormal outer ear morphology
holoprosencephaly
abnormal hippocampus layer morphology
abnormal corticospinal tract morphology
abnormal premaxilla morphology
abnormal hippocampus pyramidal cell layer
3.56E-17
2.73E-15
3.83E-19
1.25E-09
8.14E-11
1.99E-10
1.98E-09
3.68E-24
3.50E-17
2.29E-14
0%
50%
100%
Relative %
(c)
Sequence divergence
p=1.1E-02
dN/dS (human-rat)
0.0
0.2
0.4
0.6
0.8
syn
nsyn (rat)
(d)
Syntenic
n=184 x 100
-log10(p-value)
0
1
2
3
Adipose Tissue
Adrenal Gland
Blood
Blood Vessel
Brain
Breast
Colon
Esophagus
Fallopian Tube
Heart
Kidney
Liver
Lung
Muscle
Nerve
Ovary
Pancreas
Pituitary
Prostate
Skin
Stomach
Testis
Thyroid
Uterus
Vagina
(e)
Non-syntenic
n=184
Adipose Tissue
Adrenal Gland
Blood
Blood Vessel
Brain
Breast
Colon
Esophagus
Fallopian Tube
Heart
Kidney
Liver
Lung
Muscle
Nerve
Ovary
Pancreas
Pituitary
Prostate
Skin
Stomach
Testis
Thyroid
Uterus
Vagina
(f)
n=184
Amygdala
Cerebellum
Cortex
Hippocampus
Striatum
Thalamus
Pre-natal
Post-natal
Early Fetal
Early-mid fetal
Late-mid fetal
Late fetal
Early infancy
Late infancy
Early childhood
Mid-Late childhood
Adolescence
Young Adulthood
(g)
n=184
Amygdala
Cerebellum
Cortex
Hippocampus
Striatum
Thalamus
p-value
0.0
0.025
0.05
0.075
0.1
Early Fetal
Early-mid fetal
Late-mid fetal
Late fetal
Early infancy
Late infancy
Early childhood
Mid-Late childhood
Adolescence
Young Adulthood

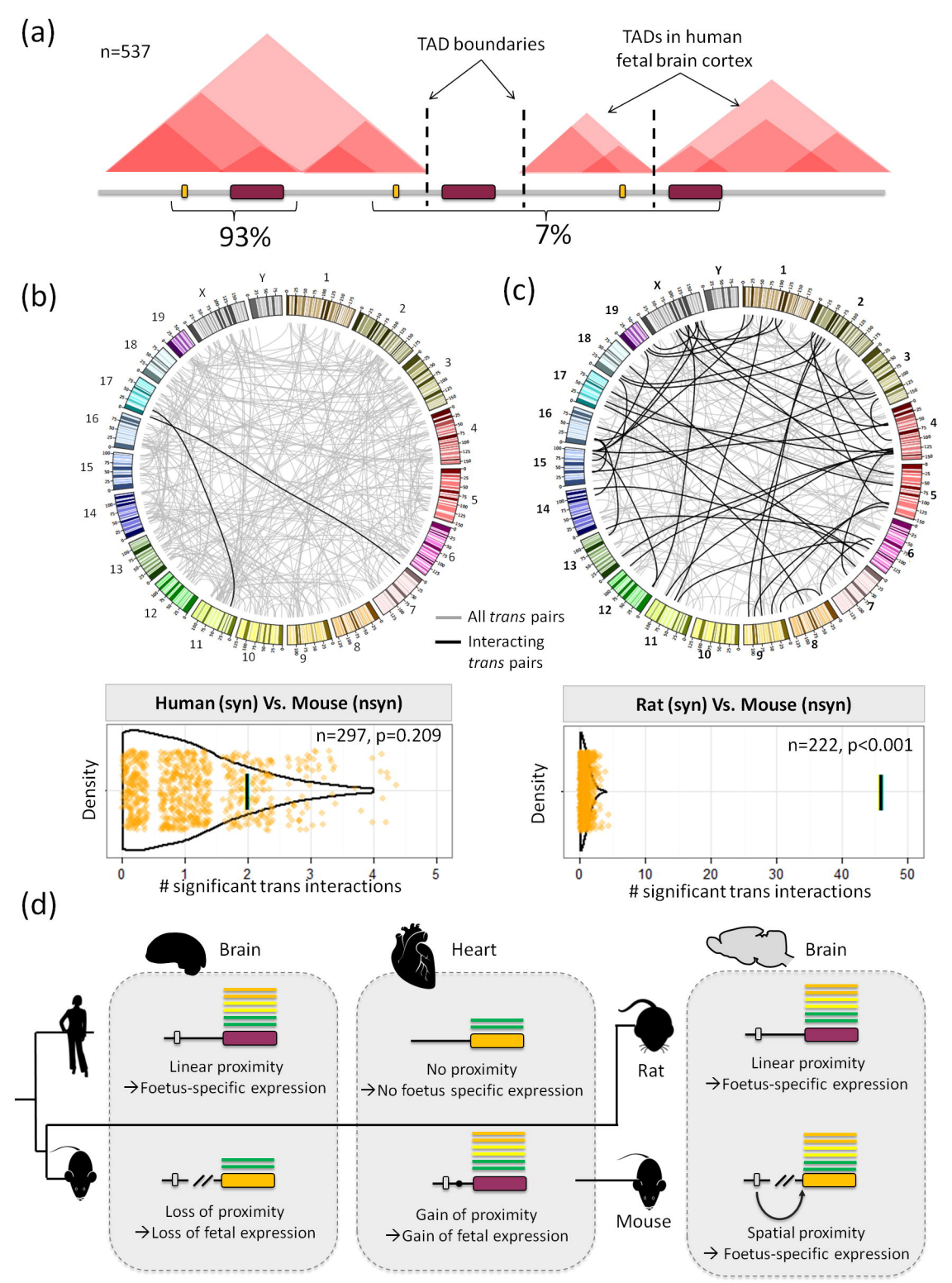

(a)
Muscle
Thymus
Brain
Fetal
Adult
p=1E-11
p=0.17
p=6E-06
-5kb
+5kb
nsyn (n=9026)
syn (n=701)
(b)
in-vitro
Emb. Stem Cells
Neuronal Progenitor
Neuron
Endoderm Progenitor
Mesoderm Progenitor
Trophoblast
p=9.25E-07
p=0.16
p=0.21
p=6.5E-08
p=1.0E-02
p=1.4E-11
(c)
facial mesenchyme
forebrain
nose
other
eye
heart
hindbrain
midbrain
limb
dorsal root ganglion
neural tube
blood vessels
somite
branchial arch
ear
trigeminal V
tail
melanocytes
liver
cranial nerve
genital tubercle
pancreas
Relative %
0.008
0.01
0.02
0.04
0.09
0.09
0.09
0.1
0.1
0.3
0.5
0.5
0.6
1
Syntenic in human and mouse, but non-syntenic in one of the others (n=33)
Syntenic in all (n=352)
p-value (approx.)
0.005
0.01
0.05
0.1
E11.5
(d)
Non-syntenic
Syntenic
hs1064
FOXG1
forebrain[3/5], Hindbrain (3/5)
hs1498
GRIP1
Hindbrain (10/15), Eye(4/15)
mm448
EMX2
Limb(5/5), nose(5/5), facial mesenchyme(5/5
hs1472
WNT5A
Tail (11/12)
hs953
SP9
Fore/hindbrain[11/13], neural tube[13/13],
hs281
FOXP4
midbrain[4/11], forebrain[7/11]
hs1866
HHEX
Blood vessels (5/5)
hs1859
LMO1
Heart(8/8),Liver(5/8), blood vessel(5/8) etc


(a)
Linkage disequilibrium map
GWAS SNP
Proxy SNPs
(b)
Proportion of CNEs with brain associated GWAS or proxy SNPs
P=1e-02
<n>=14
n=23
syn
nsyn
(c)
Amyotrophic lateral sclerosis
Language impairment
Alcohol dependence
GWAS
rs10878506
GRIP1
rs17796183
MSX2
rs7113782
NXPE2
PhyloP
H3K4me1
Fetal brain
Adult brain
Repeat
L2c
MER5B
CA)n
MIRb
L2a
L1PA7
MIR3
L4
MLT2B2
Chr12: 67159250-671563500
Chr5: 174118750-174121250
Chr11: 114638000-114644700
(d)
JASPAR_VERT
nsyn
syn
PAX3 (4.1x10-04)
PAX7 (5.4x10-04)
Density
100bp
Distance from CNE center (± 500bp)
(e)
PAX7
PAX3
Pre-natal (pcw)
Post-natal (months)
(f)
Pax3
Pax7
E13.5
E13.5
RSP
Tel
PedHy
P3
P2
P1
M
PPH
PH
PMH
MH
E11.5
E13.5
E15.5
E18.5
P4
P14
P28


(a)
CNE: chr2:207289452-207289515
(-18748bp)
ADAM3
1122bp
H3K4me1
DHS
RNAseq
State
Brain
Leg
Trunk
Thymus
Lung
Position
Human
Rat
hg19
rn5
Chr2:207289607
Chr2:207310006
Chr9:69990626
Chr9:72435773
2.4Mb
(b)
CNE: chr6:99237078-99237293
(-45287 bp)
POU3F2
905bp
RNA
Chr8
Chr5
trans
(c)
Human
Mouse
Rat
Brain
pcw16-18
E18
E18
Table of RNA-seq data for orthologues
Quantile Normalization
ADAM3
syn
nsyn
Expr. relative to TUBAβ
NA
human
mouse
rat
POU3F2
Human
mouse
rat


DNA break-points
Proportion of CNE-gene pairs
0.25
0.2
0.15
0.1
0.05
0
syn
nsyn
p=8E-13
ns
Germline
Somatic
n=21389
n=5297


(a)
% Genes
0%
10%
20%
30%
40%
50%
Nervous system, 0.004
Skeleton, 0.007
Craniofacial, 0.003
n=310
(b)
DNase-seq fold change
1
1.5
2
2.5
Human
Brain
Mouse
n=537
(d)
% Genes
0%
10%
20%
30%
40%
Cardiovascular, 0.007
Prenatal growth, 0.004
n=145
(e)
DNase-seq fold change
1
1.5
2
Human
Heart
Mouse
n=537
(c)
Developing brain
Human
Mouse
p=1e-93
p=0.4
Expression (RPKM)
Expression (Norm.microarray)
8pcw
26pcw
4m
40y
E10
E18
0d
42d
(f)
Developing heart
Human
Mouse
p=0.8
p=1e-08
Expression (Norm.microarray)
1st Trisem
2nd Trisem
Adult
E8
E18.5
3d
Adult
max
Expression
min
(g)
Developing brain
Rat
Mouse
p=1e-23
p=1e-05
Expression (Norm.microarray)
E18
E20
1d
90d
E10
E18
0d
42d
(h)
Fetal brain
Human Vs. Mouse
Rat Vs. Mouse
ρ=0.37
ρ=0.64
Mouse
Human
Rat
Human biased
max
Kernel density
min


(a)
n=537
TAD boundaries
TADs in human fetal brain cortex
93%
7%
(b)
(c)
All trans pairs
Interacting trans pairs
Human (syn) Vs. Mouse (nsyn)
n=297, p=0.209
Density
significant trans interactions
Rat (syn) Vs. Mouse (nsyn)
n=222, p<0.001
Density
significant trans interactions
(d)
Brain
Heart
Brain
Linear proximity
→Foetus-specific expression
No proximity
→No foetus specific expression
Rat
Linear proximity
→Foetus-specific expression
Loss of proximity
→Loss of fetal expression
Gain of proximity
→Gain of fetal expression
Mouse
Spatial proximity
→ Foetus-specific expression