Evolutionary Loss of Genomic Proximity to Conserved Noncoding Elements Impacted the Gene Expression Dynamics During Mammalian Brain Development

Meenakshi Bagadia,¹ Keerthivasan Raanin Chandradoss,¹ Yachna Jain, Harpreet Singh, Mohan Lal, and Kuljeet Singh Sandhu²

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

ORCID ID: 0000-0001-7632-6339 (K.S.S.)

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

KEYWORDS noncoding elements; genome organization; brain development; position effect; enhancer

AROUND 4–8% of the human genome is evolutionarily constrained, of which coding elements contribute only ~1.5% while the rest is noncoding (Lindblad-Toh *et al.* 2011; Thurman *et al.* 2012; Rands *et al.* 2014). The massive amount of data produced by the ENCODE (the Encyclopedia of DNA Elements) and Epigenome Roadmap projects have confirmed that the majority of the evolutionarily constrained noncoding DNA serves as protein-binding sites (ENCODE Project Consortium 2012; Skipper *et al.* 2015). These conserved noncoding elements (CNEs) are interwoven with the protein-coding

genes in a complex manner. Several lines of evidence converge to support the nontrivial regulatory impact of CNEs on proximal genes (Table 1). Around 200,000 human-anchored CNEs have been identified in mammals, which likely exhibit gene regulatory potential as measured through enhancer-associated chromatin marks (Roh *et al.* 2007; Seridi *et al.* 2014; Babarinde and Saitou 2016). Most CNEs cluster around developmental genes in relatively gene-poor regions of the genome (Woolfe *et al.* 2005; Akalin *et al.* 2009; Babarinde and Saitou 2016). These clusters, termed as gene regulatory blocks, have constrained linear and spatial genome organization (Kikuta *et al.* 2007; Harmston *et al.* 2017; Polychronopoulos *et al.* 2017). Most CNEs contain clusters of overlapping binding sites of developmental transcription factors, which might explain their extreme conservation (Viturawong *et al.* 2013; Warnefors *et al.* 2016).

Establishing genome-wide association between CNEs and phenotypes remains a daunting task. “Forward genetics”

Copyright © 2019 by the Genetics Society of America

doi: <https://doi.org/10.1534/genetics.119.301973>

Manuscript received September 24, 2018; accepted for publication February 14, 2019; published Early Online February 22, 2019.

Supplemental material available at Figshare: <https://doi.org/10.25386/genetics.7750919>.

¹These authors contributed equally to this work.

²Corresponding author: Department of Biological Sciences, Indian Institute of Science Education and Research (IISER) Mohali, SAS Nagar 140306, India. E-mail: sandhuks@iisermohali.ac.in

Table 1 Examples of loss-of-function mutations in CNEs

#	Gene	Disease/phenotype	Reference
1	<i>SOST</i>	Van buchmen disease	Loots <i>et al.</i> (2005)
2	<i>SHOX</i>	Leri weill dyschondrosteosis syndrome	Sabherwal <i>et al.</i> (2007)
3	<i>PAX6</i>	Aniridia	Plaisancié <i>et al.</i> (2018)
4	<i>IGF2</i>	Beckwith weidman syndrome	Sparago <i>et al.</i> (2004)
5	α/β -globins	α/β -thalassemia	Driscoll <i>et al.</i> (1989), Hatton <i>et al.</i> (1990)
6	<i>AR</i>	Evolutionary loss of penile spines and sensory vibrissae in human	McLean <i>et al.</i> (2011)
7	<i>ADGRL3</i>	Attention-deficit/Hyperactivity disorder	Martinez <i>et al.</i> (2016)
8	<i>MEIS1</i>	Restless legs syndrome	Spieler <i>et al.</i> (2014)
9	<i>BMP2</i>	Brachydactyly type A2	Dathe <i>et al.</i> (2009)
10	<i>SOX9</i>	Brachydactyly-anonychia; pierre robin sequence	Kurth <i>et al.</i> (2009)
11	<i>IHH</i>	Syndactyly and craniosynostosis	Klopocki <i>et al.</i> (2011)
12	<i>POU3F4</i>	X-linked deafness type 3	de Kok <i>et al.</i> (1996)
13	<i>FOXL2</i>	Blepharophimosis syndrome	Beysen <i>et al.</i> (2005)
14	<i>SOX10</i>	Vaardenburg syndrome type 4	Bondurand <i>et al.</i> (2012)
15	<i>GDF6</i>	Evolutionary loss of digit shortening of human feet.	Indjeian <i>et al.</i> (2016)
16	<i>IRF6</i>	Cleft lip	Rahimov <i>et al.</i> (2008)
17	<i>ZIC2</i>	Holoprosencephaly	Roessler <i>et al.</i> (2012)
18	<i>IRX</i>	Familial idiopathic scoliosis and kyphosis	Justice <i>et al.</i> (2016)
19	<i>RET</i>	Hirschsprung disease	Emison <i>et al.</i> (2005)
20	<i>CUX1</i> , <i>PTBP2</i> , <i>GPC4</i> , <i>CDKL5</i>	Autism spectrum disorders	Doan <i>et al.</i> (2016)
21	<i>SHH</i>	Preaxial polydactyly	Lettice <i>et al.</i> (2003)

The table represents the genes affected by the genetic mutations in the proximal conserved noncoding regulatory elements, associated diseases or phenotypes, and the citations to the original studies.

approaches, like genome-wide association studies (GWAS), and the “reverse genetics” approaches, like mouse mutagenesis, are notoriously difficult to scale up for high-throughput genotype–phenotype associations (Welter *et al.* 2014). With the availability of whole-genome sequences of multiple species, evolutionary methods are instrumental in deciphering genotype–phenotype associations at the genome scale. 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 (Naville *et al.* 2015; Babarinde and Saitou 2016). Attempts have been made to link evolutionary loss and sequence divergence of CNEs to lineage-specific traits, like the auditory system in echolocating mammals, adaptively morphed pectoral flippers in marine mammals, and the loss of penile spines and sensory vibrissae in humans (McLean *et al.* 2011; Davies *et al.* 2014; Marcovitz *et al.* 2016). In this study, we asked whether the lineage-specific evolutionary alterations in relative chromosomal positions of CNEs are associated with lineage-specific changes in gene expression. Through analysis of chromosomal positions of orthologous CNEs and genes from five different mammals, we observed that a significant number of genes had lost proximity to their adjacent CNEs independently in the rat lineage. This loss-of-proximity (LOP) was significantly associated with the down-regulation of genes involved in neurogenesis and neuronal migration during fetal brain development, and coincided with the independent evolutionary loss of several brain traits in the rat lineage. The study suggested a significant contribution of chromosomal rearrangements in the evolutionary divergence of developmental gene expression trajectories in mammals.

Materials and Methods

Compilation of chromosomal position data

Human (hg19), rat (rn5), dog (camFam3), horse (equCab2), and cow (bosTau6) genome assemblies were used in the analysis. CNEs were taken from Marcovitz *et al.* (2016). Marcovitz *et al.* identified CNEs by filtering the 46-way phastCons track for the coding potential as per knownGene, Ensembl, the Mammalian Gene Collection, RefSeq, Exoniphy, the Vertebrate Genome Annotation database, the Yale Pseudogene database, the miRNA registry, and snoRNA-LBME-db. Minimum length of CNEs was set to 50 and all the CNEs within 20 bp were merged to get the longer ones. Only the CNEs that were present in at least 7 of the 19 mammals having genome sequences as well as having morphological trait data from MorphoBank were kept in the final data set of 266,115 CNEs (~1.5% of the human genome). Mean and maximum lengths of the final CNEs were 174 and 2191 bp, respectively. While phastCons-based identification of CNEs does not include percent identity criteria in the definition, the least and the average percent identities for human–rat CNE alignments were 61 and 85% respectively in the rat-LOP set. The least percent identity of 61% was higher than the expected value of 45% calculated as $1 - b$, where b is the branch length ($b = 0.55$) between human and rat in the sequence alignment-driven phylogenetic tree.

Our choice of species and the CNE data set was constrained by the following considerations: (1) we wanted sufficient evolutionary depth in the analysis and Marcovitz *et al.* had considered the criteria of conservation in at least 7 of the 19 mammalian genomes when compiling the CNEs; (2) since

our analysis considered the chromosomal positions of CNEs and the genes, we only considered the genomes for which complete chromosome assemblies were available (for example, chromosome assemblies for the orders Cetacean, Chiroptera, and Proboscidea *etc.* are not presently available); (3) to obtain the sufficient number of orthologous genes across species, we restricted our analysis to fewer mammalian lineages as considering multiple species would have compromised the total number of orthologous genes to start with; and (4) to assess the loss of phenotypes, it was important to consider the species for which morphological trait data were available in the structured form, like those in the MorphoBank database. Among the sequenced and extant rodents, trait data were only available for rat and not for mouse, partly due to the slightly bigger brain size of the rat compared to that of the mouse, which had allowed the systematic morphological dissection of rat brains in the past. Similarly, analysis of time course gene expression data for pre- and postnatal brain development was an important component of this study. Such a data set was available for the rat with multiple fetal and postnatal time points, but not for mouse.

We obtained the ortholog positions of human CNEs in query species using a standard approach of mapping through LiftOver (<https://genome-store.ucsc.edu/>) chains at 0.95 mapping coverage (Marcovitz *et al.* 2016). The “deleted,” “partially deleted,” and “duplicated” mappings were removed from the data set. Finally, we compiled 114,219 CNEs that had orthologous positions in all five species. We independently obtained the table of orthologous genes across five mammals from Ensembl (<https://www.ensembl.org/index.html>). Using CNE and gene tables, a list of the nearest genes that were within 1 Mb of the CNEs was obtained for humans. The positions of orthologous CNEs and genes in other mammals were assessed, and CNE–gene pairs were classified as syntenic if the distance between the two was < 1 Mb in all five species, and as LOP if the CNE and the gene were > 2 Mb apart, or were on different chromosomes in one of the species and remained within 1 Mb in rest of the species. If there were multiple orthologs for the same gene, we took the nearest gene to the CNE on the same chromosome to ensure that a syntenic pair would not be classified as LOP due to ortholog redundancy. The distance cutoff of 1 Mb was determined based on the distribution of the number of CNE–gene pairs at different distance cutoffs. At ~1 Mb, the overall distribution approached a plateau and the numbers did not increase significantly after that (Figure 1B). The 2-Mb cutoff for LOP ensured that CNE and the gene were distant by at least twofold in their LOP form when compared to their syntenic form. Larger distance cutoff was also likely to be robust against the annotation artifacts of gene coordinates. A flow chart illustrating the overall strategy is given in Supplemental Material, Figure S1. All the data are available in the supplemental data file.

To assess the genome assembly artifacts, we mapped the rat-LOP CNE pairs to known problematic regions of the rat genome (<https://github.com/shwetaramdas/rataccessibleregions/>). Out of 2711 CNEs and out of a total of 245 genes,

only three CNEs (0.1%) and three genes (*ABCC6*, *FOS*, and *BNIP2*; 1.2%) mapped to these regions, respectively. Exclusion of these regions was unlikely to change our claims. We further mapped the rat-LOP CNE–gene pairs of the rn5 rat assembly to the rn6 assembly. Out of 2711 CNE–gene pairs, 2667 pairs (98.4%) were successfully lifted over to rn6. In total, 2227 (83.5%) pairs maintained rat-LOP status in rn6 as well (Figure S11A). Removing the ambiguous pairs did not alter the significance of brain association (Figure S11B). We also replaced the Ensembl ortholog information by *other_refseq* data in the above analysis to assess the correctness of ortholog mapping. Therefore, the concordance of 83.5%, and the persistence of brain association, confirmed that the observations presented in the article were robust against the technical artifacts of genome assembly and gene orthology.

Analysis of genomic attributes

Chromosomal coordinates of repeat elements were downloaded from the UCSC table browser. Repeat elements were mapped ± 50 kb around syntenic and LOP CNEs, and the average values of enrichment in 2-kb bins were plotted. For conservation analysis, PhyloP scores of placental mammals (<http://cgc.vital-it.ch/mga/hg19/phyloP/phyloP.html>) were mapped ± 1 kb to CNEs.

Functional enrichment analysis

Gene ontology (GO) and mammalian phenotype ontology (MPO) analyses were performed using GREAT (<http://bejerano.stanford.edu/great/public/html/>). Tissue-specificity analysis was performed using the Tissue Specific Expression Analysis (TSEA) (<http://genetics.wustl.edu/jdlab/tsea/>) and Cell-type Specific Expression Analysis (CSEA) (<http://genetics.wustl.edu/jdlab/csea-tool-2/>) resources. The tissue-specificity index (pSI) score of a gene i in a tissue k , over the given tissues $j = 1, 2, \dots, m$ was calculated as per Dougherty *et al.* (2010) using the following equation:

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

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 pSI cutoff of 0.05 was taken for the analysis. For the syntenic gene set, we randomly sampled 245 genes (the size of the rat-LOP gene set) from 4241 syntenic genes 100 times, and plotted the mean and the SE of the significance ($-\log_{10}$ of corrected P -value) of overlap between the candidate gene sets and the tissue-specific genes in the genome. The random sample of syntenic genes that exhibited the 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 the developing cerebral cortex and heart of humans, rats, and sheep were taken from BrainSpan (human cortex; <http://www.brainspan.org/static/download.html>), GSE71148 (human heart), Stead *et al.* 2006 (rat cortex), GSE53512 (rat heart), Clark *et al.*

Figure 1 Conservation of proximity and lack thereof between CNEs and the nearest genes. (A) Illustration of the strategy to infer the synteny and the LOP between CNEs and the neighboring genes across five representative mammalian orders. CNE–gene pairs were classified as syntenic if they remained proximal (< 1 Mb) in all the five species, and as LOP if they departed by > 2 Mb or were on different chromosomes in one of the species while maintaining synteny in the other four species. (B) Pdf of all CNE–gene distances in the human genome. Most CNE–gene pairs were < 1 Mb apart and therefore a cutoff of 1 Mb was applied for CNE–gene synteny. (C) Sequence conservation, as measured

through mammalian PhyloP scores, and (D) length distribution of CNEs in syntenic and LOP sets. *P*-values were calculated using Mann–Whitney *U*-test. (E) Enrichment of retrotransposons $\pm$ 50 kb around syntenic and LOP CNEs. Asterisk indicates significant *P*-values (< 0.05) calculated using a Mann–Whitney *U*-test of enrichment values $\pm$ 10 kb around CNEs. CNE, conserved noncoding element; cons., conservation; LINE, long interspersed elements; LOP, loss-of-proximity; pdf, probability distribution function; SINE, short interspersed elements; Syn, syntenic.

(sheep cortex), GSE66725 (sheep heart), and GSE63482 (mouse cortex). Average gene expression was plotted across a developmental time course. Null distributions were represented by mean and 95% C.I.s of 500 random samples of the same size as the original gene sets.

Enhancer analysis

The regulatory potential of CNEs was assessed by mapping ChromHMM data obtained from the Epigenome roadmap (http://egg2.wustl.edu/roadmap/web_portal/imputed.html#chr_imp) and ENCODE (<http://hgdownload.soe.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeBroadHmm/>) projects onto CNEs. Enhancer coordinates from FANTOM (<http://enhancer.binf.ku.dk/presets/>) were also mapped to CNEs. Cumulative overlap across the aforementioned three resources was calculated. Data sets for H3K4me1 (histone-3-lysine-4 monomethylation) for fetal and postnatal/adult human tissues were obtained from the Epigenome Roadmap (<http://www.roadmapepigenomics.org/data/>) with following accession identifiers (IDs) and age groups: fetal brain (E081 and E082; 17GW), adult brain (E067, E068, E069, E071, E072, E073, and E074; pooled 73Yr/75Yr/81Yr), fetal muscle (E089 and E090; 15GW), postnatal muscle (E107; pooled 54Yr/72Yr) and fetal thymus (E093; 15GW), postnatal thymus (E112; 3Yr), fetal heart (E083, 91 days), postnatal heart (E95, E104, and E105, pooled 3Yr/34Yr), fetal small intestine (E085, 15GW), and postnatal small intestine (E109, pooled 3Yr/30Yr). H3K4me1 data for fetal and adult mouse brain were obtained from ENCODE (ENCSR000CCZ and ENCSR000CAI for E14.5 embryos and 8-week-old adults, respectively). Fold-change over input DNA was used for aggregation plots. The Washington University (WashU) epigenome browser was used for visualization. Motif analysis was performed through “peak-motif” package from Regulatory Sequence Analysis Tool (RSAT)

(http://rsat.sb-roscoff.fr/peak-motifs_form.cgi) using JASPAR core matrices for vertebrate genomes. Syntenic CNEs were taken as background control sequences.

Mapping of proxy GWAS SNPs

A total 251,835 GWAS SNPs were obtained from GWASdb (<http://jjwanglab.org/gwasdb>). From these data, 71,990 brain-related SNPs were obtained by analyzing the Human Phenotype Ontology (HPO) terms associated with brain-associated phenotypes. We extended these data to total 533,388 nearby SNPs (proxy) that were in linkage disequilibrium with 71,990 brain-related GWAS SNPs based on 1000 genomes project using the SNAP algorithm (<https://www.broadinstitute.org/snap/snap>). Random null was prepared by picking CNE samples, of the same sample size and CNE–gene distance as the rat-LOP set, from the syntenic set 1000 times and mapping SNPs to each of these samples. The number of CNEs with at least one SNP was counted for each sample. The distribution of these numbers was regarded as random null. The *P*-value was calculated as follows:

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

where *B* = number of resampling iterations (1000), *k_b* = number of sampled syntenic CNEs having at least one SNP, and *k* = number of observed rat-LOP CNEs having at least one SNP.

Chromosome Conformation Capture data

Sequence Read Archive (SRA) files of high-throughput chromosome conformation capture (HiC) data sets for fetal and adult brains were obtained from GSE77565 and GSE87112. Data sets were processed using HiC User

Pipeline (HiCUP) (<https://www.bioinformatics.babraham.ac.uk/projects/hicup/>), and contact maps were normalized using the iterative correction and eigenvector decomposition method (<https://github.com/hiclib/iced>). To generate interaction profiles similar to Circular Chromosome Conformation Capture (4C), the transcription start site (TSS) in each CNE–gene pair was taken as bait (reference point) and its intrachromosomal interactions were obtained from HiC matrices. A Loess regression line was fitted to the HiC counts as a function of genomic distance from the bait. Significant interactions with the bait were identified by applying a cutoff of a 3-SD distance from the regression line (Klein *et al.* 2015).

DNA breakpoint analysis

We obtained 552 germline breakpoints associated with congenital disorders having brain abnormalities and 68,018 somatic cancer breakpoints from van Heesch *et al.* (2014). The matching RNA-sequencing (RNA-Seq) data of peripheral blood of the patient and the mother were obtained from the European Nucleotide Archive (<https://www.ebi.ac.uk/ena>) with the accession IDs ERX358048 and ERX358046, respectively.

In total, 2061 evolutionary DNA breakpoints for rodents were taken from Bourque *et al.* (2004, 2006), Larkin *et al.* (2009), and Lemaitre *et al.* (2009). These breakpoints were then mapped onto interspacer regions between CNEs and the nearest gene-TSSs. The random null was obtained by picking CNE–gene pairs, of the same sample size and CNE–gene distance as that of the rat-LOP set, from the syntenic set 1000 times and mapping the breakpoint in the interspacer regions. The number of CNE–gene pairs with at least one breakpoint in between were counted for each sample. The distribution of these numbers was regarded as random null. *P*-values were calculated using the equation used in the GWAS SNP analysis. The breakpoint distances from the mammalian ancestor were obtained from Luo *et al.* (2012).

Mammalian traits

The statuses of morphological traits in five mammals were obtained from project ID P773 of the MorphoBank database (<https://morphobank.org/>). Traits that exhibited the same status in at least three of the mammals including rats, but that showed a different status in humans, were classified as independently modified traits in humans. Similarly, the traits that had the same status in at least three species including humans, but had changed status in rats, were denoted as independently modified traits in the rat.

Data availability

The authors affirm that all data necessary for confirming the conclusions of this article are represented fully within the article, its tables and figures, and the associated supplemental material. Supplemental material available at Figshare: <https://doi.org/10.25386/genetics.7750919>.

Results

Evolutionary loss of genomic proximity between the CNE and the proximal gene

Using chromosomal position data of CNEs and representative primate (human), rodent (rat), carnivore (dog), perrisodactyl (horse), and artiodactyl (cow) protein-coding genes, we obtained 51,434 syntenic CNE–gene pairs (4241 genes), wherein CNEs and the nearest gene-TSSs were < 1 Mb distant in all five species. There were 3579 LOP cases (334 genes) wherein the CNEs and the gene-TSSs were on different chromosomes, or were > 2-Mb apart independently only in one of the species (Figure 1A and Figure S1, *Materials and Methods*). The rationale of 1-Mb distance cutoff for the synteny was based on the observation that the distribution of all CNE–gene distances saturated when they approached a 1-Mb range (Figure 1B). A similar approach has been used previously to infer enhancer–promoter linkage based on evolutionary synteny between the two in a 1-Mb range (Naville *et al.* 2015). The distance cutoff of 2 Mb for LOP ensured that the minimal expansion in CNE–gene distance would at least be twofold. To test if the CNEs in syntenic and LOP sets were comparable, we assessed their lengths and degree of conservation in mammalian genomes. Figure 1C showed an insignificant difference in the degree of sequence conservation of syntenic and LOP CNEs, suggesting that the sequence of LOP CNEs had not diverged among mammals as compared to that of syntenic CNEs. The length distribution of syntenic and LOP CNEs showed only a marginal difference toward slightly longer CNEs in the LOP set (Figure 1D). The density of CNEs per gene (within 1 Mb) was also not significantly different in the two sets ($P = 0.15$, Figure S2). However, the syntenic and LOP CNEs were located in the genomic domains of distinct sequence properties. We analyzed the enrichment of Short Interspersed Elements (SINE), Long Interspersed Elements (LINE), and Long Terminal Repeats (LTR), which cover ~ 50% of the mammalian genome, around syntenic and LOP CNEs. The syntenic CNEs were enriched in the regions of open chromatin, as signified through greater enrichment of SINE content ($P < 7e-04$) as compared with the LOP CNEs, and might have a more widespread role across different cell lineages (Figure 1E). In contrast, the LOP CNEs were located in the domains enriched with LTR repeats ($P < 0.05$), marking their susceptibility to genomic rearrangements through mechanisms like nonallelic homologous recombination (Roeder and Fink 1980; Chan and Kolodner 2011) (Figure 1E). LINE elements, in general, did not exhibit significant differences in the two sets (except in rat). These observations were largely consistent across species, marking an ancestral property (Figure 1E), except in rat wherein the LINE elements were enriched ($P = 3.4e-04$) around LOP CNEs. This exception can be explained by the fact that the rodents retained the least of the ancestral retrotransposons when compared to other mammals and instead accumulated novel retroelements (Buckley *et al.* 2017).

Figure 2 Genomic rearrangements underlying the loss of proximity. (A) Number of CNEs and genes (in parentheses) exhibiting LOP independently in different mammals. The P -value for the LOP instances in rat and the next highest value (in dog) was calculated using Fisher's exact test. (B) Scaling between the number of LOP CNE-gene pairs and the evolutionary breakpoint distance of species from the common mammalian ancestor. A breakpoint distance signifies adjacency of two genes that were together in one genome, but not in its neighbor in the breakpoint-based phylogenetic tree. Dashed line: least squares regression. (C) The number of rat-LOP CNE-gene pairs having at least one rodent-specific breakpoint between a CNE and a gene-TSS, overlaid onto the null distribution prepared from the syntenic set. (D) Distinct *trans* and *cis* chromosomal rearrangements as inferred from the analysis of gene orthology flanking the rat-LOP CNEs. Shown are the neighboring genes around CNEs. Red color represents the target gene and orange color represents the nearest gene on the other side of the CNE. White box represents the neighboring gene of the target gene. The bar plot represents the proportion of total rat-LOP CNE-gene pairs in each scenario. The red color in the bar plot marks the proportion for which inversion could be confirmed through analysis of gene orientations. (E) Shown are the two examples illustrating the loss of CNE-gene proximity in the rat. In the first example, a CNE was located 45-kb upstream of the gene *POU3F2* in human but was split on different chromosomes in the rat. The second example shows that an inversion event had distanced the CNE and the proximal *ADAM23* gene by up to 2.4 Mb in the rat genome. chr, chromosome; CNE, conserved noncoding element; LOP, loss-of-proximity; Syn, syntenic.

A relatively large number of syntenic CNE-gene pairs (93.5%) confirmed the widespread conservation of linear proximity between CNEs and their adjacent genes (Babarinde and Saitou 2016). Of the total 3579 LOP instances, a significant number (2711, 75%, $P < 2.2e-16$) were associated with the rat genome alone, in line with the significantly greater number of structural variations in the rodent clade (Luo *et al.* 2012) (Figure 2, A and B). Positive scaling ($\rho = 0.53$) between the number of LOP CNE-gene pairs and the breakpoint distances of species from the common ancestor signified that the CNE-gene proximity was an ancestral trait (Figure 2B). Due to the significant loss of

CNE-gene proximity in the rat lineage as compared to others, we focused on rat instances in this study. By rat-LOP, we refer to the "loss of CNE-gene proximity in rat" from Figure 2C onward. To directly assess the association of rat-LOPs with structural variations in the genome, we analyzed the rodent-specific evolutionary breakpoints (Materials and Methods). We observed that 930 (34%) of all rat-LOP instances had at least one rodent-specific breakpoint between the gene-TSS and the CNE, as compared to 319 (11%) on average for the random null prepared through distance-controlled bootstrap sampling of syntenic CNE-gene pairs ($P < 0.001$, Figure 2C, Materials and Methods). This suggested that the

rodent-specific genomic rearrangements nonrandomly coincided with the loss of CNE–gene proximity in the rat, but could not explain the entire repertoire of rat-LOP cases (Figure 2C). We argue that the sequence alignment-based annotations of evolutionary breakpoints might not represent the entire repertoire of genomic rearrangements and, therefore, analyzed the gene orthology on either side of LOP CNEs to map the various rearrangement scenarios through which CNE–gene proximity was lost. We found that the translocation-like scenario, as marked by “i” in Figure 2D, largely explained the interchromosomal (*trans*) splits of CNE and the adjacent gene. The scenario that reflected the mapping artifacts, as in panel iii, was underrepresented (5%, Figure 2D). Analysis of intrachromosomal (*cis*) splits suggested inversion-like events separating the CNE–gene pairs. Scenarios iv and v showed events wherein regions adjacent to CNEs (on the left side in scenario iv and right side in scenario v) underwent local rearrangements, of which 30 and 90% events, respectively, were confirmed as inversion-like events (red-colored bars in Figure 2D) by analyzing the changes in relative strand orientations of neighboring genes. Scenario vi represented 1.4% of the total cases wherein the in-between region of CNE and the gene was expanded without involving any apparent rearrangement. We illustrated examples of *trans* and *cis* splits of CNEs and genes in Figure 2E. Gene *POU3F2* on human chromosome 6 was syntenic to a CNE, which was 45-kb upstream. The orthologous CNE and the gene in the rat were on chromosomes 8 and 5, respectively, marking the *trans* split of the CNE and the gene through translocation (Figure 2E). Another CNE was 18-kb upstream of gene *ADAM23* on human chromosome 2. The rat orthologs were separated by 2.4 Mb on chromosome 9 through an inversion (Figure 2E).

We concluded that the rodent-specific genomic rearrangements, as inferred from gene orthology approach, largely explained the loss of CNE–gene proximity in the rat.

Genes that had lost proximity to CNEs in rats were associated with fetal brain development

Significant differences in the genomic attributes around syntenic and rat-LOP CNEs hinted at their distinct functional roles. To assess their functions, syntenic and rat-LOP gene-lists were subjected to GO and MPO analyses. The analysis of GO terms revealed enrichment of general as well as various tissue-specific development-related terms in the syntenic set ($P < 6e-06$), while the rat-LOP set was specifically enriched with nervous system development-related terms ($P < 0.03$, Figure 3A). In MPO analysis, syntenic set exhibited enrichment of neonatal lethality and skeletal phenotypes ($P < 9e-12$), while the rat-LOP set was associated with brain morphology-related phenotypes ($P < 0.04$, Figure S3A). Species, other than rat, did not exhibit enrichment of any particular functional term, owing to smaller sample size.

We further followed the above observations through tissue-specific gene expression analysis in humans. The syntenic set had a widespread representation of genes expressed in

different cell-lineages and, therefore, did not exhibit significant tissue specificity, while genes in the rat-LOP set were specifically expressed in the brain ($P = 0.001$, Figure 3B). The brain-specific expression of genes in the rat-LOP set was also confirmed through enrichment analysis of anatomical terms from the Bgee database (Figure S3B). Within the brain, the rat-LOP set was enriched with the genes specifically expressed in the cerebral cortex during fetal, but not post-natal, development ($P < 0.025$, Figure 3C). In contrast, the genes in the syntenic set did not exhibit any specificity for brain tissues and developmental stages (Figure 3C). These observations highlighted fetal brain-specific roles of genes in the rat-LOP set.

CNEs exhibiting LOP to genes in rats function as fetal brain-specific enhancers

To test whether the differences between syntenic and rat-LOP sets observed through the functional analysis of genes were coherent with the associated CNEs, we tested the regulatory potential of CNEs by analyzing their epigenomic properties across tissues. Through analysis of enhancer-associated chromatin state annotations from the Epigenome Roadmap, and the ENCODE and Fantom consortia (*Materials and Methods*), we observed that 74% of syntenic and 61% of rat-LOP CNEs overlapped with the enhancer-associated regulatory sites in at least one of the tissues or cell types, marking the enhancer potential of CNEs. Relatively less representation of enhancers in the rat-LOP set might relate to their tissue or developmental stage-specific functions, a hypothesis that we further reconciled through detailed analysis of histone modification associated with enhancers, namely H3K4me1. We chose this mark because of its strong association with the enhancer potential and the availability of genome-wide data sets for all the cell lineages that we were interested in. We observed that: (1) the CNEs in the syntenic set exhibited consistent H3K4me1 enrichment across several fetal and adult tissues like the thymus (endodermal), muscle (mesodermal), heart (mesodermal), intestine (endodermal), and brain (ectodermal) (Figure 4A and Figure S4, A and B); (2) H3K4me1 enrichment over CNEs in the rat-LOP set was specifically higher (comparable to that of syntenic CNEs) in the fetal, but not adult, brain ($P_{\text{fetal}} = 0.03$ vs. $P_{\text{adult}} = 2.8e-06$ for comparison with the syntenic CNEs, Figure 4A). We further observed the significant enrichment of binding sites of ectoderm-specific transcription factors, which were specifically upregulated in fetal brain in rat-LOP CNEs as compared to syntenic CNEs ($P < 0.05$, Figure S5, *Materials and Methods*). These observations were largely coherent with our proposal that rat-LOP CNE–gene pairs were associated with fetal brain development.

To assess the physical association between CNEs and cognate genes, we generated virtual 4C data by processing available HiC data sets of fetal and adult human brains (Figure S4C). Figure 4B shows the significant ($P = 0.0069$) fetal-to-adult ratio of the proportion of rat-LOP CNE–gene pairs showing significant physical interactions as compared to that

Syntenic

Rat-LOP

Figure 3 Functional characterization of genes in syntenic and rat-LOP sets. (A) Enrichment of gene ontology terms among genes in syntenic and rat-LOP sets. *P*-values were calculated using a hypergeometric test and were corrected using the Benjamini–Hochberg method. (B) Tissue-specific expression analysis of genes in the syntenic and rat-LOP sets. 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) < 0.05. The horizontal gray-colored line represents a *P*-value of 0.01. For the syntenic set, mean values and SE of significance for 100 random samples of 245 genes (the size of the rat-LOP set) from syntenic sets were plotted. (C) Expression specificity of genes in syntenic and rat-LOP sets across brain regions and across developmental stages. For the syntenic set, the sample that exhibited maximum significance for brain specificity in (B) was taken. Size of the nested hexagons represents the proportion of all genes specifically expressed in a particular tissue at a particular developmental stage. Hexagons are nested inward based on relative stringency of tissue-specificity scores (pSI = 0.05, 0.01, 0.001, and 0.0001, respectively). Color gradient represents the magnitude of corrected *P*-values of Fisher's exact test. dev., development; LOP, loss-of-proximity; pos., possible; pSI, tissue-specificity index; reg., regulation; sys., system; transc., transcription.

Figure 4 Enhancer properties of CNEs in syntenic and rat-LOP sets. (A) H3K4me1 Chromatin immunoprecipitation enrichment (over input), on and around CNEs in syntenic and rat-LOP sets, in fetal and postnatal tissues. Shaded regions around the curves represent 95% C.I.s. *P*-values for the difference between syntenic and rat-LOP CNEs were calculated using the Mann–Whitney *U*-test of H3K4me1 enrichment values in 1-kb spanning windows on either side of the CNEs. (B) Virtual 4C analysis of CNE–gene interactions in fetal and adult human brains. The barplot shows the fetal-to-adult ratio of the proportion of CNE–gene pairs exhibiting significant (above 3σ distance from the Loess regression fit) physical chromatin interactions. *P*-value was calculated using Fisher’s exact test. (C) The examples of *GPR85* and *FEZF2* genes are shown for illustration. Vertical gray (reference point) and yellow bars represent the TSS and CNE positions, respectively. Red and gray curves show virtual 4C and H3K4me1 signals, respectively. The black smooth line represents the Loess fit of the 4C signal as a function of genomic distance from the reference point. The dotted line represents the 3σ distance from the Loess regression line. (D) The proportion of rat-LOP CNEs having at least one brain-associated SNP superimposed onto the null distribution obtained from syntenic CNEs. *P*-value was calculated using the bootstrap method by randomly sampling 2711 CNEs (size of the rat-LOP set) from the syntenic set 1000 times. (E) An example of the *EPHA4* gene and its proximal CNE having a schizophrenia-associated GWAS SNP is shown. The tracks for PhylP conservation score, H3K4me1, and RNA-sequencing data of fetal and postnatal brains are aligned accordingly. The orthologous CNE and gene were 7.2-Mb apart on chromosome 9 in the rat. Cons, conservation; CNE, conserved noncoding element; cons., conservation;; GWAS, genome-wide association studies; LOP, loss-of-proximity; RNA-seq, RNA-sequencing; Syn., syntenic; TSS, transcription start site; 4C, circular chromosome conformation capture.

of syntenic CNE–gene pairs. We illustrated the physical interactions between CNEs and genes through examples (Figure 4C and Figure S4D). TSSs (reference point) of the *GPR85* and *FEZF2* genes, both associated with neurological phenotypes (Matsumoto *et al.* 2008; Chen *et al.* 2011; Eckler *et al.* 2014), showed significant interaction frequency (above 3 SD from regression fit) to their cognate CNEs in human fetal brain, but

not in the adult brain. The H3K4me1 signals at TSSs and CNEs were also significant in fetal brain as compared to the adult. Epigenomic analyses thus suggested that the majority of the rat-LOP CNEs exhibited enhancer-associated hallmarks in the fetal brain.

By mapping the trait/disease-associated SNPs from GWAS and the nearby SNPs (proxy) in the linkage disequilibrium

based on 1000 genome data, we observed that 105 of the rat-LOP CNEs had at least one brain-related SNP (Figure 4D). This representation was statistically significant ($P = 0.006$) when compared with that of the syntenic set (Figure 4D). These observations represented genetic evidence of brain-specific roles of rat-LOP CNEs. We highlighted the example of the *EPHA4* gene, which is required for radial neuron migration, and is involved in the pathways leading to lissencephaly and schizophrenia in humans (Sentürk *et al.* 2011; Steinecke *et al.* 2014). An upstream CNE to this gene had a schizophrenia-associated SNP. Fetal brain specificity of CNEs and gene expression were illustrated using H3K4me1 and RNA-Seq tracks of fetal and postnatal human brains (Figure 4E).

Therefore, our observations through enhancer data sets, epigenomic marks, differential motif enrichment analysis, and brain-associated SNPs concomitantly established that the rat-LOP CNEs were specific to fetal brain development in humans.

Developmental tolerance of LOP to CNEs

While we have shown that the genes and the CNEs that had lost proximity in the rat were associated with fetal brain development, whether or not CNE–gene proximity itself was linked with the brain-specific expression of the cognate gene remained to be addressed. Toward this, we assessed the representation of germline breakpoints associated with the congenital disorders exhibiting brain abnormalities, and the somatic cancer breakpoints between CNEs and gene-TSSs, in syntenic and rat-LOP sets. Since the observed germline breakpoints are the ones that had survived through germline and the embryonic development, their presence and absence between CNEs and the adjacent genes signifies developmental tolerance and intolerance, respectively, of loss of CNE–gene synteny. In contrast, the cancer breakpoints of somatic origin do not undergo such selection, and hence do not indicate the developmental tolerance or lack thereof. Figure 5A shows the number of rat-LOP CNE–gene pairs having at least one DNA breakpoint between the gene-TSS and the CNE. To test the significance, we prepared the random null from syntenic CNE–gene pairs of a similar distance distribution as that of the rat-LOP. We observed significantly greater representation ($P < 0.001$) of germline breakpoints in the rat-LOP set as compared to the syntenic set, while the representation of somatic breakpoints showed an insignificant difference ($P = 0.92$, Figure 5A). We interpreted that DNA breaks between rat-LOP CNEs and the proximal genes were developmentally tolerable, and possibly mediated the genomic rearrangements in the genome of the rodent ancestor, which consequently allowed the LOP between CNEs and genes in rats. We further elaborated an example of germline chromosomal rearrangements that had split the CNE and the adjacent gene in congenital disorders with brain abnormality (Figure 5B). The example in Figure 5B shows a chromothripsis event wherein an inversion has split a CNE–gene pair. The involved gene, *BCL11A*, regulates cortical neuron migration, and mutations therein associate with microcephaly and intellectual disability in human (Wiegrefe *et al.* 2015). The

Figure 5 Tolerance and intolerance of CNE–gene split. (A) The number of rat-LOP CNE–gene pairs flanking at least one germline breakpoint associated with the congenital disorders having brain abnormalities (left panel) and the somatic cancer breakpoint (right panel), superimposed onto null distributions obtained from the syntenic set of the same CNE–gene distance distribution as that of the rat-LOP set. P -values were calculated using the bootstrap method with 1000 random samplings. (B) An example illustrating a chromothripsis event breaking CNE–gene proximity in a congenital disorder with brain abnormalities. The right panel shows the difference in expression level of involved gene (*BCL11A*) in the patient having genomic rearrangement and in the normal mother. chr, chromosome; CNE, conserved noncoding element; LOP, loss-of-proximity; RPKM, reads per kilobase per million; syn., syntenic.

BCL11A gene also exhibited a 3.6-fold loss-of-expression in the peripheral blood of the patient having genomic rearrangement as compared with the normal mother of the patient.

LOP to CNEs coincided with the loss of fetus-specific upregulation of genes in rat brains

An important question was whether or not the evolutionary loss of CNE–gene proximity in the rat was associated with a loss of gene expression. To assess the functional fate of associated genes, we analyzed their time course gene expression trajectories for the developing cerebral cortices of humans, rats, and sheep (as an out-group). Sheep were included in the analysis due to the availability of gene expression data sets for pre- and postnatal tissues. We found that 99.4% of CNE–gene pairs that had lost proximity in the rat were syntenic in sheep as well, again confirming the independent LOP in the rat lineage. We observed a relative loss of fetus-specific gene expression in the rat brain as compared to that of humans and sheep, suggesting that the LOP correlated with a loss of fetus-specific gene expression in the developing rat brain (Figure 6A). These observations were also robust against varying CNE–gene distances in the rat-LOP set (Figure S6). Enrichment of neurogenesis-related genes and downregulation

Figure 6 Evolutionary dynamics of developmental gene expression associated with the loss of CNE–gene proximity. (A) Red curves in the plots represent the mean expression of genes in the rat-LOP set and black line with 95% C.I.s (gray bars) represents the random null prepared from the syntenic set. Fetal samples are highlighted in gray background. Asterisk (*) indicates significant ($P < 0.01$) expression of rat-LOP genes in the fetal stages of brain development as compared to postnatal. Left panel represents cerebral cortex and right panel is for heart data sets (control). Numbers of fetal and postnatal time points are indicated in parentheses at the bottom of each plot. P -values are calculated using a one-tailed Student’s t -test for the fetal upregulation of rat-LOP genes when compared to postnatal gene expression in the same species and data set. Note that the expression units are not comparable across plots. (B) Bar plot representing the rat-to-human ratio of the proportion of traits modified independently in humans and rats. Top three traits that were independently modified in humans and rats are marked. A greater number of brain traits were independently modulated in rats (#9) as compared to humans (#3) ($P = 0.07$, Fisher’s exact test). Brain traits independently modified in rats are listed on the right. CNE, conserved noncoding element; LOP, loss-of-proximity; RPKM, reads per kilobase per million; TPM, transcripts per million.

thereof in the fetal brains of rats has implications in understanding the loss of brain traits in the rat lineage. Indeed, an analysis of mammalian trait data suggested that a significant number of brain traits were independently modified in rats, but were preserved in other mammals analyzed in this study ($P = 0.07$, Figure 6B).

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” through loss of enhancer-promoter associations. The tissue and developmental stage-specific impact of structural variations highlighted the possibility of their significant role in altering developmental dynamics toward evolutionary gain or loss of lineage-specific traits.

Discussion

It is not always the change in number and the sequence of protein-coding regions in the genome that leads to phenotype alternation in evolution; the dynamics of gene expression is equally relevant in this context. One of the ways the gene expression is altered is through position effect, *i.e.*, the relative chromosomal position of the gene in the genome can alter its expression through altered proximities to regulatory elements, chromatin states in the neighborhood, and spatial localization to different subnuclear compartments. Position

effect was first discovered through the observation that the chromosomal arrangement of duplicated copies of the *bar* gene in *bar*-mutant flies had an influence on its expression, and consequently caused a relative decrease in the number of eye facets (Sturtevant 1925, 1928). Similarly, the *white* gene, when localized near heterochromatin, gives a mottled eye phenotype with red and white patches in the *Drosophila* eye (Wallrath and Elgin 1995; Martin-Morris *et al.* 1997). Despite its significance, the role of position effect in the evolution of traits has not been investigated thoroughly. In this study, we analyzed a kind of position effect due to evolutionary loss of proximities of genes to the conserved noncoding regulatory elements in mammals. Through comparative genomic analysis, we showed that the CNE–gene pairs that were syntenic in most mammals, but lost their close linear proximity independently in rats, were associated with alterations in the transcriptional program during fetal brain development. We presented evidence regarding how the genomic rearrangements might have impacted the evolution of lineage-specific phenotypes by modulating the developmental trajectories in early stages.

Enhancers can function at distances longer than several Megabases and spatial synteny has been observed among genomic regions that have been rearranged during evolution (Véron *et al.* 2011; Bagadia *et al.* 2016). How then does the loss of linear proximity to CNEs downregulate the expression of genes? Position effect significantly alters the expression

noise of the genes (Chen and Zhang 2016). Evidence also suggests that long-range or *trans* enhancer–promoter interactions occur at the cost of increased expression noise (Sandhu 2012; Singh *et al.* 2016; Kustatscher *et al.* 2017). As a result, the overall expression level in the tissue is expected to decline due to increased stochastic fluctuations in gene expression across cells. Therefore, we hypothesized that the loss of linear proximity between a CNE and a gene would have compromised the expression level of the gene by allowing stochastic variations in enhancer–promoter interactions. An alternate explanation can stem from the functional divergence of orthologous CNEs. While CNEs have sufficient depth in sequence conservation across lineages, their developmental expression pattern can vary in a lineage-specific manner (Polychronopoulos *et al.* 2017). Therefore, the CNEs losing proximity to their cognate genes in rats might have developmentally diverged functions elsewhere in the genome.

Enrichment of brain development-related genes in the rat-LOP set might relate to the developmental plasticity of the brain as compared to other tissues. Genomic alterations at the loci important for the development of the basic body plan and functioning would be embryonic lethal, which largely explains the significant representation of skeletal/heart development- and neonatal lethality-related genes in the syntenic set (Figure 3A and Figure S3A). The brain, despite having neurodevelopment plasticity, exhibits the least genome-wide expression divergence across mammalian species (Khaitovich *et al.* 2006; Strand *et al.* 2007). However, within the space of a small rat-LOP gene set, expression divergence was observed. This suggested that the least expression divergence observed for brain is due to cellular functions that need to be precisely regulated to maintain the delicately shaped brain tissues of all mammals in general, while the ones that exhibit divergence would be implicated in developmental functions specific to the fetal brain. Our data showed that one of the ways that such expression divergence was modulated in evolution was through the alteration of genomic proximity between CNEs and neighboring genes. Fetus brain-specific downregulation of neurogenesis-related genes that had lost proximity to CNEs in rats aligned to the hypothesis that observed genomic alterations might link to brain traits that were lost in the rodent lineage. To test this hypothesis, we obtained the quantitative data of mammalian traits from MorphoBank. Through analysis of the trait data, we showed that among the species in this analyses, rats exhibited the greatest number of independently modified brain traits, including the ones directly associated with neurogenesis, like an absence of cerebral folding of the cortex, absence of claustrum separation from the cortex, and an absence of the lateral geniculate nucleus magnocellular layer *etc.* (Figure 6). Among the rodent brain traits, loss of cerebral folding of the cortex, *i.e.*, a lissencephalic or smooth brain phenotype, is the largest visible alteration in rodent brains. Folded or gyrencephalic brain, in general, is considered to be an adaptation in mammals with a greater encephalization quotient, intelligence, and complex behavioral traits (Prothero and

Sundsten 1984; Toro *et al.* 2008). It can, therefore, be contended that CNE–gene proximity and associated fetal brain-specific expression was not lost in rats, but was rather gained in other mammals that had bigger, gyrencephalic brains. However, we argue that significant nonuniformity in the cerebral cortex has been observed across several different mammalian species (Herculano-Houzel *et al.* 2008), and the assumption that the common ancestor of placental mammals had a smaller and simpler brain has been challenged recently (Rowe *et al.* 2011; Lewitus *et al.* 2014). Evidence has supported the gyrencephalic brain of the eutherian ancestor and the subsequent loss of cortical gyration in rodents (Kelava *et al.* 2013). The enrichment of genes associated with brain morphology phenotypes (Figure S3), extracellular matrix–receptor interactions, and actin cytoskeleton regulation (*ACTN*, *ITGA1*, *ITGA11*, *DCLK2*, *ASAP1*, *LAMB4*, *LZTS1*, *CD36*, *Frabin*, *TWISTNB*, *DDX11*, *SGCG*, and *SCIN* *etc.*), as well as ones implicated in human cortical malformations (*MYCN*, *NRXN1*, *RASA1*, *FEZF2*, *EFNA5*, and *GLI3* *etc.*) (Piñero *et al.* 2017), in the rat-LOP set further support the LOP to CNEs in rat rather than a gain-of-proximity in other mammals (Figure S7).

We also emphasized that the LOP in the rat was inferred by filtering the CNE–gene pairs, which were syntenic in all other species, and hence were evolutionarily constrained, except in rats. Assessing gain-of-proximity was difficult because a CNE–gene pair that was distant in all species except one cannot be considered as an evolutionarily constrained CNE–gene pair. We suspected that gain-of-proximity inferred in this flawed manner would not have shown any functional association. Indeed, this was observed through an independent analysis (Figure S7).

While our analysis was limited to the species for which morphological trait data were available, assessment of rat-LOPs in the mouse was warranted to infer if the LOP to CNEs was a property of the rodent ancestor. We inducted mice in the analysis for this purpose. The overlap of mouse-LOP with that of rat-LOP was small (62 genes and 596 CNEs, Figure S8). While rats and mice are phylogenetically close, their genomes were independently rearranged from the common rodent ancestor (Luo *et al.* 2012). Therefore, lower overlap was not completely unexpected. Though the genes that had commonly lost proximity to CNEs in rats and mice did not show significant enrichment of any ontology term owing to a smaller sample size, the genes exhibited fetal brain-specific gene expression in humans, but not in rats and mice (Figure S8). Several genes related to cytoskeleton regulation (*DCLK2*, *TWISTNB*, *DDX11*, *FEZ1*, *Frabin*, *SGCG*, and *SCIN*) commonly lost proximity to CNEs in rats and mice. Among these genes, the *DCLK2*, *FEZ1*, *Frabin*, *TWISTNB*, and *DDX11* genes have known association with cortical malformations (Kosan and Kunz 2002; Chen *et al.* 2005; Nakanishi and Takai 2008; van der Lelij *et al.* 2010; Kang *et al.* 2011; Stouffer *et al.* 2016; Romero *et al.* 2018). Yet another explanation for the lower overlap of rat-LOP and mouse-LOP genes could be the limited number of orthologous mappings across six species. To expand the ortholog data, we performed an

independent analysis of primates (humans and chimps) and rodents (mice and rats), and fetched the instances where genes were proximal to CNEs in both primates but lost proximity in both rodents. As shown in Figure S8, the rodent-LOPs were enriched with cell migration and actin filament regulation-related functions. The rodent-LOP genes obtained in this analysis also exhibited fetal brain-specific upregulation in humans, but not in mice and rats (Figure S8). These observations suggested that the association of LOP with cell migration, cytoskeleton regulation, extracellular matrix organization, and fetal brain-specific upregulation of genes was common to the rodent ancestor. The CNEs in the rodent-LOP sets also showed significant enrichment of the H3K4me1 mark when compared to syntenic CNEs in human, but not in mouse, fetal brains (Figure S9). Therefore, the concurrently higher fetal brain-specific activities of LOP-CNEs and genes in a syntenic configuration (in human), and loss thereof in an LOP configuration (mouse), strongly supported our major claims.

We also assessed our observations with the genes that exhibited LOP to CNEs in rats but not in mice, and *vice versa*. In adherence to our hypothesis, the genes exhibited loss of fetal brain-specific expression only in the case of LOP to CNEs either in rats or in mice, ruling out the possibility that the lack of fetus-specific expression was a general property of all the genes in rodent brains (Figure S10).

It remains debatable whether or not the alterations in the brain traits in the rodent lineage represent adaptive selection, or were products of neutral drift. Some studies have suggested that smaller and lissencephalic brains were adaptively selected among mammals with distinct life history traits, like narrow habitats and smaller social groups, than those of gyrencephalic species (Lewitus *et al.* 2014). The distinct neurogenic potentials of gyrencephalic and lissencephalic species have been attributed to the observed differences (Lewitus *et al.* 2014). The increased proliferative potential of basal progenitor cells is necessary and sufficient to explain gyrencephalic brains (Lewitus *et al.* 2014). The loss of such proliferative potential, which was likely an ancestral trait, might have caused an inefficient neurogenic program in lissencephalic species. Our observation that the genes that had lost their proximity to CNEs in rats were involved in neuronal differentiation and were downregulated in the fetal rat brain is largely coherent with the above proposal.

Altogether, our observations highlight a link between genome order and the evolutionary dynamics of temporal gene expression patterns associated with mammalian brain development. The study also suggests that the genomic rearrangements, without any change in the genomic content, might have impacted the developmental trajectories and shaped the evolution of phenotypes.

Acknowledgments

K.S.S. received financial support from the Department of Science and Technology (EMR/2015/001681) and the Department

of Biotechnology (BT/PR16366/BID/7/598/2016), India. M.B. and M.L. were financially supported by the Indian Institute of Science Education and Research Mohali. K.R.C., Y.J., and H.S. thank the University Grant Commission (UGC), Department of Biotechnology (DBT), and Science and Engineering Research Board (SERB) respectively for their fellowships.

Literature Cited

- Akalin, A., D. Fredman, E. Arner, X. Dong, J. C. Bryne *et al.*, 2009 Transcriptional features of genomic regulatory blocks. *Genome Biol.* 10: R38. <https://doi.org/10.1186/gb-2009-10-4-r38>
- Babarinde, I. A., and N. Saitou, 2016 Genomic locations of conserved noncoding sequences and their proximal protein-coding genes in mammalian expression dynamics. *Mol. Biol. Evol.* 33: 1807–1817. <https://doi.org/10.1093/molbev/msw058>
- Bagadia, M., A. Singh, and K. Singh Sandhu, 2016 Three dimensional organization of genome might have guided the dynamics of gene order evolution in eukaryotes. *Genome Biol. Evol.* 8: 946–954. <https://doi.org/10.1093/gbe/evw050>
- Beysen, D., J. Raes, B. P. Leroy, A. Lucassen, J. R. Yates *et al.*, 2005 Deletions involving long-range conserved nongenic sequences upstream and downstream of FOXL2 as a novel disease-causing mechanism in blepharophimosis syndrome. *Am. J. Hum. Genet.* 77: 205–218. <https://doi.org/10.1086/432083>
- Bondurand, N., V. Fouquet, V. Baral, L. Lecerf, N. Loundon *et al.*, 2012 Alu-mediated deletion of SOX10 regulatory elements in Waardenburg syndrome type 4. *Eur. J. Hum. Genet.* 20: 990–994. <https://doi.org/10.1038/ejhg.2012.29>
- Bourque, G., P. A. Pevzner, and G. Tesler, 2004 Reconstructing the genomic architecture of ancestral mammals: lessons from human, mouse, and rat genomes. *Genome Res.* 14: 507–516. <https://doi.org/10.1101/gr.1975204>
- Bourque, G., E. M. Zdobnov, P. Bork, P. A. Pevzner, and G. Tesler, 2005 Comparative architectures of mammalian and chicken genomes reveal highly variable rates of genomic rearrangements across different lineages. *Genome Res.* 15: 98–110. <https://doi.org/10.1101/gr.3002305>
- Buckley, R. M., R. D. Kortschak, J. M. Raison, and D. L. Adelson, 2017 Similar evolutionary trajectories for retrotransposon accumulation in mammals. *Genome Biol. Evol.* 9: 2336–2353. <https://doi.org/10.1093/gbe/evx179>
- Chan, J. E., and R. D. Kolodner, 2011 A genetic and structural study of genome rearrangements mediated by high copy repeat Ty1 elements. *PLoS Genet.* 7: e1002089. <https://doi.org/10.1371/journal.pgen.1002089>
- Chen, J. G., M. R. Rasin, K. Y. Kwan, and N. Sestan, 2005 Zfp312 is required for subcortical axonal projections and dendritic morphology of deep-layer pyramidal neurons of the cerebral cortex. *Proc. Natl. Acad. Sci. USA* 102: 17792–17797. <https://doi.org/10.1073/pnas.0509032102>
- Chen, L., J. Zheng, N. Yang, H. Li, and S. Guo, 2011 Genomic selection identifies vertebrate transcription factor Fezf2 binding sites and target genes. *J. Biol. Chem.* 286: 18641–18649. <https://doi.org/10.1074/jbc.M111.236471>
- Chen, X., and J. Zhang, 2016 The genomic landscape of position effects on protein expression level and noise in yeast. *Cell Syst.* 2: 347–354. <https://doi.org/10.1016/j.cels.2016.03.009>
- Clark, E. L., S. J. Bush, M. E. B. McCulloch, I. L. Farquhar, R. Young *et al.*, 2017 A high resolution atlas of gene expression in the domestic sheep (*Ovis aries*). *PLoS Genet* 13: e1006997. <https://doi.org/10.1371/journal.pgen.1006997>

- Dathe, K., K. W. Kjaer, A. Brehm, P. Meinecke, P. Nurnberg *et al.*, 2009 Duplications involving a conserved regulatory element downstream of BMP2 are associated with brachydactyly type A2. *Am. J. Hum. Genet.* 84: 483–492. <https://doi.org/10.1016/j.ajhg.2009.03.001>
- Davies, K. T., G. Tsagkogeorga, and S. J. Rossiter, 2014 Divergent evolutionary rates in vertebrate and mammalian specific conserved non-coding elements (CNEs) in echolocating mammals. *BMC Evol. Biol.* 14: 261. <https://doi.org/10.1186/s12862-014-0261-5>
- de Kok, Y. J., E. R. Vossenaar, C. W. Cremers, N. Dahl, J. Laporte *et al.*, 1996 Identification of a hot spot for microdeletions in patients with X-linked deafness type 3 (DFN3) 900 kb proximal to the DFN3 gene POU3F4. *Hum. Mol. Genet.* 5: 1229–1235. <https://doi.org/10.1093/hmg/5.9.1229>
- Doan, R. N., B. I. Bae, B. Cubelos, C. Chang, A. A. Hossain *et al.*, 2016 Mutations in human accelerated regions disrupt cognition and social behavior. *Cell* 167: 341–354.e12. <https://doi.org/10.1016/j.cell.2016.08.071>
- Dougherty, J. D., E. F. Schmidt, M. Nakajima, and N. Heintz, 2010 Analytical approaches to RNA profiling data for the identification of genes enriched in specific cells. *Nucleic Acids Res.* 38: 4218–4230. <https://doi.org/10.1093/nar/gkq130>
- Driscoll, M. C., C. S. Dobkin, and B. P. Alter, 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. USA* 86: 7470–7474. <https://doi.org/10.1073/pnas.86.19.7470>
- Eckler, M. J., K. A. Larkin, W. L. McKenna, S. Katzman, C. Guo *et al.*, 2014 Multiple conserved regulatory domains promote Fezf2 expression in the developing cerebral cortex. *Neural Dev.* 9: 6. <https://doi.org/10.1186/1749-8104-9-6>
- Emison, E. S., A. S. McCallion, C. S. Kashuk, R. T. Bush, E. Grice *et al.*, 2005 A common sex-dependent mutation in a RET enhancer underlies Hirschsprung disease risk. *Nature* 434: 857–863. <https://doi.org/10.1038/nature03467>
- ENCODE Project Consortium, 2012 An integrated encyclopedia of DNA elements in the human genome. *Nature* 489: 57–74. <https://doi.org/10.1038/nature11247>
- Harmston, N., E. Ing-Simmons, G. Tan, M. Perry, M. Merkenschlager *et al.*, 2017 Topologically associating domains are ancient features that coincide with Metazoan clusters of extreme non-coding conservation. *Nat. Commun.* 8: 441. <https://doi.org/10.1038/s41467-017-00524-5>
- Hatton, C. S., A. O. Wilkie, H. C. Drysdale, W. G. Wood, M. A. Vickers *et al.*, 1990 Alpha-thalassemia caused by a large (62 kb) deletion upstream of the human alpha globin gene cluster. *Blood* 76: 221–227.
- Herculano-Houzel, S., C. E. Collins, P. Wong, J. H. Kaas, and R. Lent, 2008 The basic nonuniformity of the cerebral cortex. *Proc. Natl. Acad. Sci. USA* 105: 12593–12598. <https://doi.org/10.1073/pnas.0805417105>
- Indjeian, V. B., G. A. Kingman, F. C. Jones, C. A. Guenther, J. Greenwood *et al.*, 2016 Evolving new skeletal traits by cis-regulatory changes in bone morphogenetic proteins. *Cell* 164: 45–56. <https://doi.org/10.1016/j.cell.2015.12.007>
- Justice, C. M., K. Bishop, B. Carrington, J. C. Mullikin, K. Swindle *et al.*, 2016 Evaluation of IIRX genes and conserved noncoding elements in a region on 5p13.3 linked to families with familial idiopathic scoliosis and kyphosis. *G3 (Bethesda)* 6: 1707–1712. <https://doi.org/10.1534/g3.116.029975>
- Kang, E., K. E. Burdick, J. Y. Kim, X. Duan, J. U. Guo *et al.*, 2011 Interaction between FEZ1 and DISC1 in regulation of neuronal development and risk for schizophrenia. *Neuron* 72: 559–571. <https://doi.org/10.1016/j.neuron.2011.09.032>
- Kelava, I., E. Lewitus, and W. B. Huttner, 2013 The secondary loss of gyrencephaly as an example of evolutionary phenotypical reversal. *Front. Neuroanat.* 7: 16. <https://doi.org/10.3389/fnana.2013.00016>
- Khaitovich, P., W. Enard, M. Lachmann, and S. Paabo, 2006 Evolution of primate gene expression. *Nat. Rev. Genet.* 7: 693–702. <https://doi.org/10.1038/nrg1940>
- Kikuta, H., M. Laplante, P. Navratilova, A. Z. Komisarczuk, P. G. Engstrom *et al.*, 2007 Genomic regulatory blocks encompass multiple neighboring genes and maintain conserved synteny in vertebrates. *Genome Res.* 17: 545–555. <https://doi.org/10.1101/gr.6086307>
- Klein, F. A., T. Pakozdi, S. Anders, Y. Ghavi-Helm, E. E. Furlong *et al.*, 2015 FourCSeq: analysis of 4C sequencing data. *Bioinformatics* 31: 3085–3091. <https://doi.org/10.1093/bioinformatics/btv335>
- Klopocki, E., S. Lohan, F. Brancati, R. Koll, A. Brehm *et al.*, 2011 Copy-number variations involving the IHH locus are associated with syndactyly and craniosynostosis. *Am. J. Hum. Genet.* 88: 70–75. <https://doi.org/10.1016/j.ajhg.2010.11.006>
- Kosan, C., and J. Kunz, 2002 Identification and characterisation of the gene TWIST NEIGHBOR (TWISTNB) located in the microdeletion syndrome 7p21 region. *Cytogenet. Genome Res.* 97: 167–170. <https://doi.org/10.1159/000066618>
- Kurth, I., E. Klopocki, S. Stricker, J. van Oosterwijk, S. Vanek *et al.*, 2009 Duplications of noncoding elements 5' of SOX9 are associated with brachydactyly-anonychia. *Nat. Genet.* 41: 862–863. <https://doi.org/10.1038/ng0809-862>
- Kustatscher, G., P. Grabowski, and J. Rappsilber, 2017 Pervasive coexpression of spatially proximal genes is buffered at the protein level. *Mol. Syst. Biol.* 13: 937. <https://doi.org/10.15252/msb.20177548>
- Larkin, D. M., G. Pape, R. Donthu, L. Auvil, M. Welge *et al.*, 2009 Breakpoint regions and homologous synteny blocks in chromosomes have different evolutionary histories. *Genome Res.* 19: 770–777. <https://doi.org/10.1101/gr.086546.108>
- Lemaitre, C., L. Zaghloul, M. F. Sagot, C. Gautier, A. Arneodo *et al.*, 2009 Analysis of fine-scale mammalian evolutionary breakpoints provides new insight into their relation to genome organisation. *BMC Genomics* 10: 335. <https://doi.org/10.1186/1471-2164-10-335>
- Lettec, L. A., S. J. Heaney, L. A. Purdie, L. Li, P. de Beer *et al.*, 2003 A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum. Mol. Genet.* 12: 1725–1735. <https://doi.org/10.1093/hmg/ddg180>
- Lewitus, E., I. Kelava, A. T. Kalinka, P. Tomancak, and W. B. Huttner, 2014 An adaptive threshold in mammalian neocortical evolution. *PLoS Biol.* 12: e1002000. <https://doi.org/10.1371/journal.pbio.1002000>
- Lindblad-Toh, K., M. Garber, O. Zuk, M. F. Lin, B. J. Parker *et al.*, 2011 A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* 478: 476–482. <https://doi.org/10.1038/nature10530>
- Loots, G. G., M. Kneissel, H. Keller, M. Baptist, J. Chang *et al.*, 2005 Genomic deletion of a long-range bone enhancer misregulates sclerostin in Van Buchem disease. *Genome Res.* 15: 928–935. <https://doi.org/10.1101/gr.3437105>
- Luo, H., W. Arndt, Y. Zhang, G. Shi, M. A. Alekseyev *et al.*, 2012 Phylogenetic analysis of genome rearrangements among five mammalian orders. *Mol. Phylogenet. Evol.* 65: 871–882. <https://doi.org/10.1016/j.ympev.2012.08.008>
- Marcovitz, A., R. Jia, and G. Bejerano, 2016 “Reverse genomics” predicts function of human conserved noncoding elements. *Mol. Biol. Evol.* 33: 1358–1369. <https://doi.org/10.1093/molbev/msw001>
- Martin-Morris, L. E., A. K. Csink, D. R. Dorer, P. B. Talbert, and S. Henikoff, 1997 Heterochromatic trans-inactivation of Drosophila white transgenes. *Genetics* 147: 671–677.

- Martinez, A. F., Y. Abe, S. Hong, K. Molyneux, D. Yarnell *et al.*, 2016 An ultraconserved brain-specific enhancer within ADGRL3 (LPHN3) underpins attention-deficit/hyperactivity disorder susceptibility. *Biol. Psychiatry* 80: 943–954. <https://doi.org/10.1016/j.biopsych.2016.06.026>
- Matsumoto, M., R. E. Straub, S. Marengo, K. K. Nicodemus, S. Matsumoto *et al.*, 2008 The evolutionarily conserved G protein-coupled receptor SREB2/GPR85 influences brain size, behavior, and vulnerability to schizophrenia. *Proc. Natl. Acad. Sci. USA* 105: 6133–6138. <https://doi.org/10.1073/pnas.0710717105>
- McLean, C. Y., P. L. Reno, A. A. Pollen, A. I. Bassan, T. D. Capellini *et al.*, 2011 Human-specific loss of regulatory DNA and the evolution of human-specific traits. *Nature* 471: 216–219. <https://doi.org/10.1038/nature09774>
- Nakanishi, H., and Y. Takai, 2008 Frabin and other related Cdc42-specific guanine nucleotide exchange factors couple the actin cytoskeleton with the plasma membrane. *J. Cell. Mol. Med.* 12: 1169–1176. <https://doi.org/10.1111/j.1582-4934.2008.00345.x>
- Naville, M., M. Ishibashi, M. Ferg, H. Bengani, S. Rinkwitz *et al.*, 2015 Long-range evolutionary constraints reveal cis-regulatory interactions on the human X chromosome. *Nat. Commun.* 6: 6904. <https://doi.org/10.1038/ncomms7904>
- Piñero, J., À. Bravo, N. Quéralt-Rosinách, A. Gutiérrez-Sacristán, J. Deu-Pons *et al.*, 2017 DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res.* 45: D833–D839. <https://doi.org/10.1093/nar/gkw943>
- Plaisancié, J., M. Tarilonte, P. Ramos, C. Jeanton-Scaramouche, V. Gaston *et al.*, 2018 Implication of non-coding PAX6 mutations in aniridia. *Hum. Genet.* 137: 831–846. <https://doi.org/10.1007/s00439-018-1940-x>
- Polychronopoulos, D., J. W. D. King, A. J. Nash, G. Tan, and B. Lenhard, 2017 Conserved non-coding elements: developmental gene regulation meets genome organization. *Nucleic Acids Res.* 45: 12611–12624. <https://doi.org/10.1093/nar/gkx1074>
- Prothero, J. W., and J. W. Sundsten, 1984 Folding of the cerebral cortex in mammals. A scaling model. *Brain Behav. Evol.* 24: 152–167. <https://doi.org/10.1159/000121313>
- Rahimov, F., M. L. Marazita, A. Visel, M. E. Cooper, M. J. Hitchler *et al.*, 2008 Disruption of an AP-2alpha binding site in an IRF6 enhancer is associated with cleft lip. *Nat. Genet.* 40: 1341–1347. <https://doi.org/10.1038/ng.242>
- Rands, C. M., S. Meader, C. P. Ponting, and G. Lunter, 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. <https://doi.org/10.1371/journal.pgen.1004525>
- Roeder, G. S., and G. R. Fink, 1980 DNA rearrangements associated with a transposable element in yeast. *Cell* 21: 239–249. [https://doi.org/10.1016/0092-8674\(80\)90131-2](https://doi.org/10.1016/0092-8674(80)90131-2)
- Roessler, E., P. Hu, S. K. Hong, K. Srivastava, B. Carrington *et al.*, 2012 Unique alterations of an ultraconserved non-coding element in the 3' UTR of ZIC2 in holoprosencephaly. *PLoS One* 7: e39026. <https://doi.org/10.1371/journal.pone.0039026>
- Roh, T. Y., G. Wei, C. M. Farrell, and K. Zhao, 2007 Genome-wide prediction of conserved and nonconserved enhancers by histone acetylation patterns. *Genome Res.* 17: 74–81. <https://doi.org/10.1101/gr.5767907>
- Romero, D. M., N. Bahi-Buisson, and F. Francis, 2018 Genetics and mechanisms leading to human cortical malformations. *Semin. Cell Dev. Biol.* 76: 33–75. <https://doi.org/10.1016/j.semcdb.2017.09.031>
- Rowe, T. B., T. E. Macrini, and Z. X. Luo, 2011 Fossil evidence on origin of the mammalian brain. *Science* 332: 955–957. <https://doi.org/10.1126/science.1203117>
- Sabherwal, N., F. Bangs, R. Roth, B. Weiss, K. Jantz *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. <https://doi.org/10.1093/hmg/ddl470>
- Sandhu, K. S., 2012 Did the modulation of expression noise shape the evolution of three dimensional genome organizations in eukaryotes? *Nucleus* 3: 286–289. <https://doi.org/10.4161/nucl.20263>
- Sentürk, A., S. Pfennig, A. Weiss, K. Burk, and A. Acker-Palmer, 2011 Ephrin Bs are essential components of the Reelin pathway to regulate neuronal migration. *Nature* 472: 356–360 (erratum: *Nature* 478: 274). <https://doi.org/10.1038/nature09874>
- Seridi, L., T. Ryu, and T. Ravasi, 2014 Dynamic epigenetic control of highly conserved noncoding elements. *PLoS One* 9: e109326. <https://doi.org/10.1371/journal.pone.0109326>
- Singh, A., M. Bagadia, and K. S. Sandhu, 2016 Spatially coordinated replication and minimization of expression noise constrain three-dimensional organization of yeast genome. *DNA Res.* 23: 155–169. <https://doi.org/10.1093/dnares/dsw005>
- Skipper, M., A. Eccleston, N. Gray, T. Heemels, N. Le Bot *et al.*, 2015 Presenting the epigenome roadmap. *Nature* 518: 313. <https://doi.org/10.1038/518313a>
- Sparago, A., F. Cerrato, M. Vernucci, G. B. Ferrero, M. C. Silengo *et al.*, 2004 Microdeletions in the human H19 DMR result in loss of IGF2 imprinting and Beckwith-Wiedemann syndrome. *Nat. Genet.* 36: 958–960. <https://doi.org/10.1038/ng1410>
- Spieler, D., M. Kaffé, F. Knauf, J. Bessa, J. J. Tena *et al.*, 2014 Restless legs syndrome-associated intronic common variant in Meis1 alters enhancer function in the developing telencephalon. *Genome Res.* 24: 592–603. <https://doi.org/10.1101/gr.166751.113>
- Stead, J. D., C. Neal, F. Meng, Y. Wang, S. Evans *et al.*, 2006 Transcriptional profiling of the developing rat brain reveals that the most dramatic regional differentiation in gene expression occurs postpartum. *J. Neurosci.* 26: 345–353. <https://doi.org/10.1523/JNEUROSCI.2755-05.2006>
- Steinecke, A., C. Gampe, G. Zimmer, J. Rudolph, and J. Bolz, 2014 EphA/ephrin A reverse signaling promotes the migration of cortical interneurons from the medial ganglionic eminence. *Development* 141: 460–471. <https://doi.org/10.1242/dev.101691>
- Stouffer, M. A., J. A. Golden, and F. Francis, 2016 Neuronal migration disorders: focus on the cytoskeleton and epilepsy. *Neurobiol. Dis.* 92: 18–45. <https://doi.org/10.1016/j.nbd.2015.08.003>
- Strand, A. D., A. K. Aragaki, Z. C. Baquet, A. Hodges, P. Cunningham *et al.*, 2007 Conservation of regional gene expression in mouse and human brain. *PLoS Genet.* 3: e59. <https://doi.org/10.1371/journal.pgen.0030059>
- Sturtevant, A. H., 1925 The effects of unequal crossing over at the bar locus in *Drosophila*. *Genetics* 10: 117–147.
- Sturtevant, A. H., 1928 A further study of the so-called mutation at the bar locus of *Drosophila*. *Genetics* 13: 401–409.
- Thurman, R. E., E. Rynes, R. Humbert, J. Vierstra, M. T. Maurano *et al.*, 2012 The accessible chromatin landscape of the human genome. *Nature* 489: 75–82. <https://doi.org/10.1038/nature11232>
- Toro, R., M. Perron, B. Pike, L. Richer, S. Veillette *et al.*, 2008 Brain size and folding of the human cerebral cortex. *Cereb. Cortex* 18: 2352–2357. <https://doi.org/10.1093/cercor/bhm261>
- van der Lelij, P., K. H. Chrzanowska, B. C. Godthelp, M. A. Rooimans, A. B. Oostra *et al.*, 2010 Warsaw breakage syndrome, a cohesinopathy associated with mutations in the XPD helicase family member DDX11/ChIR1. *Am. J. Hum. Genet.* 86: 262–266. <https://doi.org/10.1016/j.ajhg.2010.01.008>

- van Heesch, S., M. Simonis, M. J. van Roosmalen, V. Pillalamarri, H. Brand *et al.*, 2014 Genomic and functional overlap between somatic and germline chromosomal rearrangements. *Cell Rep.* 9: 2001–2010. <https://doi.org/10.1016/j.celrep.2014.11.022>
- Véron, A. S., C. Lemaitre, C. Gautier, V. Lacroix, and M. F. Sagot, 2011 Close 3D proximity of evolutionary breakpoints argues for the notion of spatial synteny. *BMC Genomics* 12: 303. <https://doi.org/10.1186/1471-2164-12-303>
- Viturawong, T., F. Meissner, F. Butter, and M. Mann, 2013 A DNA-centric protein interaction map of ultraconserved elements reveals contribution of transcription factor binding hubs to conservation. *Cell Rep.* 5: 531–545. <https://doi.org/10.1016/j.celrep.2013.09.022>
- Wallrath, L. L., and S. C. Elgin, 1995 Position effect variegation in *Drosophila* is associated with an altered chromatin structure. *Genes Dev.* 9: 1263–1277. <https://doi.org/10.1101/gad.9.10.1263>
- Warnefors, M., B. Hartmann, S. Thomsen, and C. R. Alonso, 2016 Combinatorial gene regulatory functions underlie ultra-conserved elements in *Drosophila*. *Mol. Biol. Evol.* 33: 2294–2306. <https://doi.org/10.1093/molbev/msw101>
- Welter, D., J. MacArthur, J. Morales, T. Burdett, P. Hall *et al.*, 2014 The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* 42: D1001–D1006. <https://doi.org/10.1093/nar/gkt1229>
- Wiegrefe, C., R. Simon, K. Peschkes, C. Kling, M. Strehle *et al.*, 2015 Bcl11a (Ctip1) controls migration of cortical projection neurons through regulation of Sema3c. *Neuron* 87: 311–325. <https://doi.org/10.1016/j.neuron.2015.06.023>
- Woolfe, A., M. Goodson, D. K. Goode, P. Snell, G. K. McEwen *et al.*, 2005 Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biol.* 3: e7. <https://doi.org/10.1371/journal.pbio.0030007>

Communicating editor: J. Birchler