RESOURCE

nature neuroscience

Epigenomic annotation of gene regulatory alterations during evolution of the primate brain

Marit W Vermunt[1], Sander C Tan[1], Bas Castelijns[1], Geert Geeven[1], Peter Reinink[1], Ewart de Bruijn[1], Ivanela Kondova[2], Stephan Persengiev[2], Netherlands Brain Bank[3], Ronald Bontrop[2], Edwin Cuppen[1], Wouter de Laat[1] & Menno P Creyghton[1]

Although genome sequencing has identified numerous noncoding alterations between primate species, which of those are regulatory and potentially relevant to the evolution of the human brain is unclear. Here we annotated *cis*-regulatory elements (CREs) in the human, rhesus macaque and chimpanzee genomes using chromatin immunoprecipitation followed by sequencing (ChIP-seq) in different anatomical regions of the adult brain. We found high similarity in the genomic positioning of rhesus macaque and human CREs, suggesting that the majority of these elements were already present in a common ancestor 25 million years ago. Most of the observed regulatory changes between humans and rhesus macaques occurred before the ancestral separation of humans and chimpanzees, leaving a modest set of regulatory elements with predicted human specificity. Our data refine previous predictions and hypotheses on the consequences of genomic changes between primate species and allow the identification of regulatory alterations relevant to the evolution of the brain.

Despite representing a scaled up primate brain[1], the human brain has several qualitative differences from closely related primate species that could underlie advances in cognitive function[2,3]. To shed light on the genetic basis of these differences, genome sequencing analyses of primate genomes have uncovered a host of species-specific events including gene copy number variations, insertions and deletions as well as specific alterations of coding sequence[4,5]. However, given the overall similarity between coding sequences among primates, most of the phenotypic variation has been predicted to occur as a result of changes in the control of gene expression[6]. Especially in the brain, adaptive changes have been predicted to occur primarily in noncoding DNA[7]. In support of this, alterations in gene expression networks across the primate lineage have now been described, including those that are specific to the human brain[8]. However, linking these changes to specific gene-regulatory elements altered during evolution has proven difficult. This is primarily due to the fact that functional annotation of regulatory elements in the nonhuman primate genome is sparse.

Noncoding enhancer elements control transcription of genes from large genomic distances in a cell type–specific manner[9]. This inherent tissue-specific activity, as well as the presence of redundant enhancers[10,11], potentially renders these elements more amenable to changes during evolution[12–14]. Accordingly, overall sequence conservation at enhancers across species is generally low, and functional conservation as determined by the location of transcription factors or specific histone signatures across mammals is limited[15,16]. Nevertheless, specific sequence changes at enhancer elements have been linked to major phenotypic changes during mammalian evolution[12,13,17,18]. Therefore, functional comparison of regulatory elements in relevant tissues as well as relevant species is required to understand their role in speciation. In previous work, regulatory changes between the human and rhesus macaque embryonic cortex and limb have been analyzed[19,20]. However, whether these regulatory alterations are specific to humans or occurred earlier in primate evolution is still unclear.

Here we annotated CREs in several distinct anatomical regions of adult brain tissue from human, rhesus macaque and chimpanzee. We found that the overall positional conservation of CREs is high, but observed differences in their usage. Furthermore, we show that most regulatory changes occurred before the divergence of human and chimpanzee. These data are a substantial resource from which human-specific regulatory changes in the brain can be predicted.

RESULTS

Annotation of CREs in the rhesus and chimpanzee brain

The location of active CREs in the genome can be predicted by the presence of histone modifications including histone H3 lysine 27 acetylation (H3K27ac) for promoters as well as enhancers[16,21–25]. Recently, we had used this modification to annotate CREs in the human brain in two complete donor hemispheres using ChIP-seq[26]. As significant redundancy in H3K27ac enrichment existed between the areas that we had analyzed previously, for analysis in this study we selected eight anatomical regions corresponding to the major anatomical subdivisions of the brain, including cerebellum, caudate nucleus, thalamic nuclei, putamen, white matter, precentral gyrus, prefrontal cortex and occipital pole (**Supplementary Fig. 1a**, and **Supplementary Tables 1** and **2**). For comparative purposes, we generated additional data sets from a third human donor (**Supplementary Tables 1** and **2**).

[1]Hubrecht Institute-KNAW and University Medical Center Utrecht, Utrecht, the Netherlands. [2]Biomedical Primate Research Center, Rijswijk, the Netherlands. [3]Netherlands Institute for Neuroscience, Amsterdam, the Netherlands. Correspondence should be addressed to M.P.C. (m.creyghton@hubrecht.eu).

Received 12 November 2015; accepted 21 December 2015; published online 25 January 2016; doi:10.1038/nn.4229

We analyzed the same eight anatomical structures by ChIP-Seq for H3K27ac in hemispheres obtained from three different rhesus macaque monkeys and two chimpanzees (**Supplementary Fig. 1a**,**b**, and **Supplementary Tables 1** and **2**). Rhesus macaques are evolutionary more distant than great apes, sharing the last common ancestor with humans around 25 million years ago[27] (**Fig. 1a**). In contrast, the chimpanzee is the closest related primate species to human, with a common ancestor that lived approximately 6 million years ago (**Fig. 1a**). This enabled the identification of human-specific regulatory changes, as opposed to changes that occurred before the separation of the human and chimpanzee lineages.

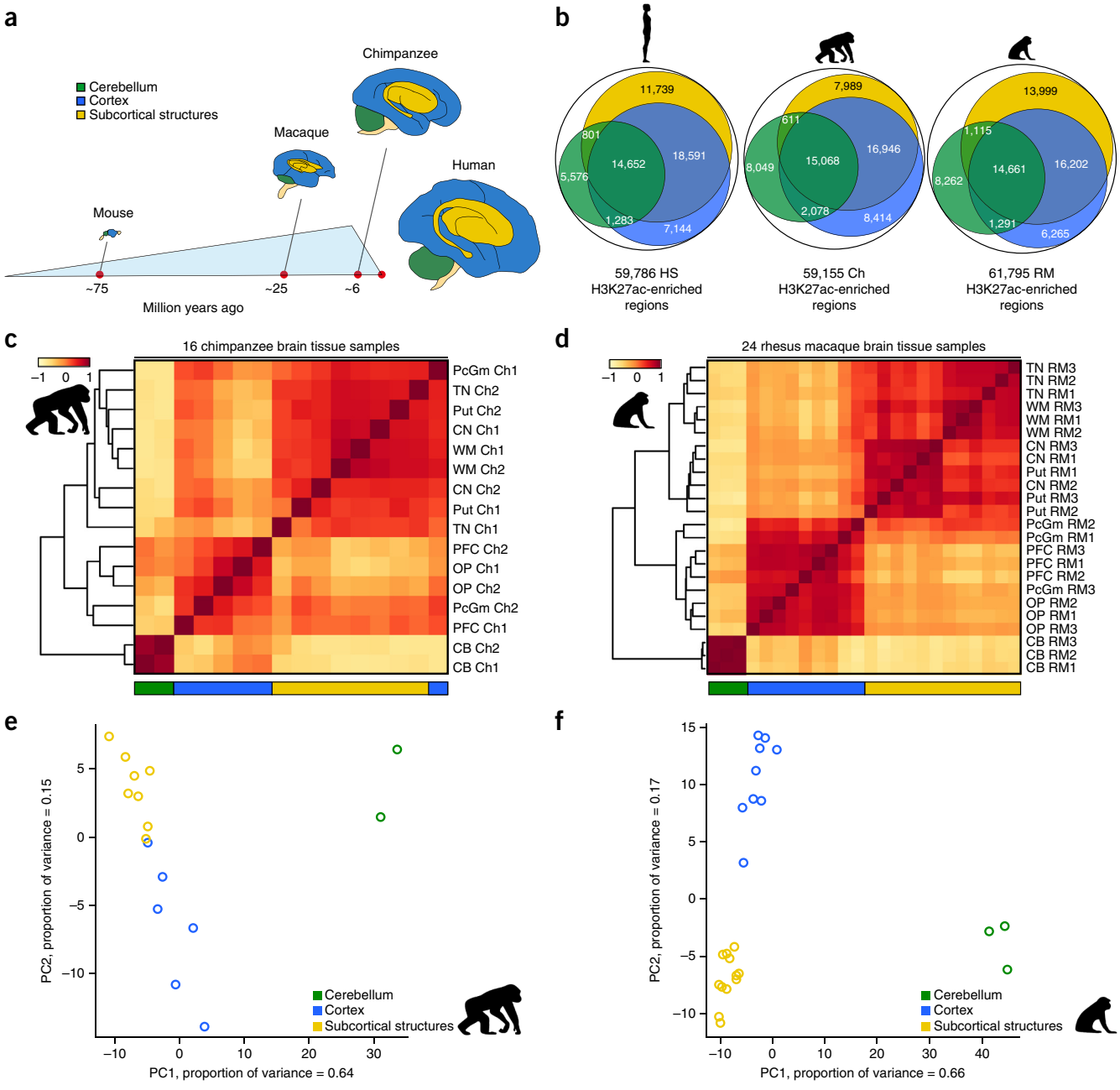

Figure 1 Annotation of active *cis*-regulatory elements in the rhesus macaque and chimpanzee brain. (**a**) Schematic of 100 million years of brain evolution in mammals. Brains are drawn to approximate scale, with main anatomical subdivisions color-coded as indicated. (**b**) Overlap of H3K27ac-enriched regions between the main anatomical subdivisions of the brain (color-coded as in **a**) for human (HS), chimpanzee (Ch) and rhesus macaque (RM). Outer black circles are scaled to the total number of enriched regions, listed below the diagrams. (**c**,**d**) Hierarchical clustering of 16 chimpanzee brain samples based on normalized H3K27ac enrichment of 59,155 predicted chimpanzee CREs (**c**) and of 24 rhesus macaque brain samples based on normalized H3K27ac enrichment of 61,795 predicted rhesus macaque CREs (**d**). Color scale depicts the Pearson correlation coefficient between samples. Bar at the bottom indicates main anatomical structures (**Supplementary Fig. 1a**): cortex (blue; prefrontal cortex (PFC), precentral gyrus (PcGm) and occipital pole (OP)), subcortical structures (yellow; caudate nucleus (CN), putamen (Put), white matter (WM) and thalamic nuclei (TN)) and cerebellum (CB; green). (**e**) PCA of 16 chimpanzee brain samples on normalized H3K27ac enrichment of 59,155 predicted CREs in chimpanzee. (**f**) PCA of 24 rhesus macaque brain samples on normalized H3K27ac enrichment for 61,795 predicted CREs in rhesus macaque. Results of a similar analysis on matched human samples are shown in **Supplementary Figure 1c**,**d**.

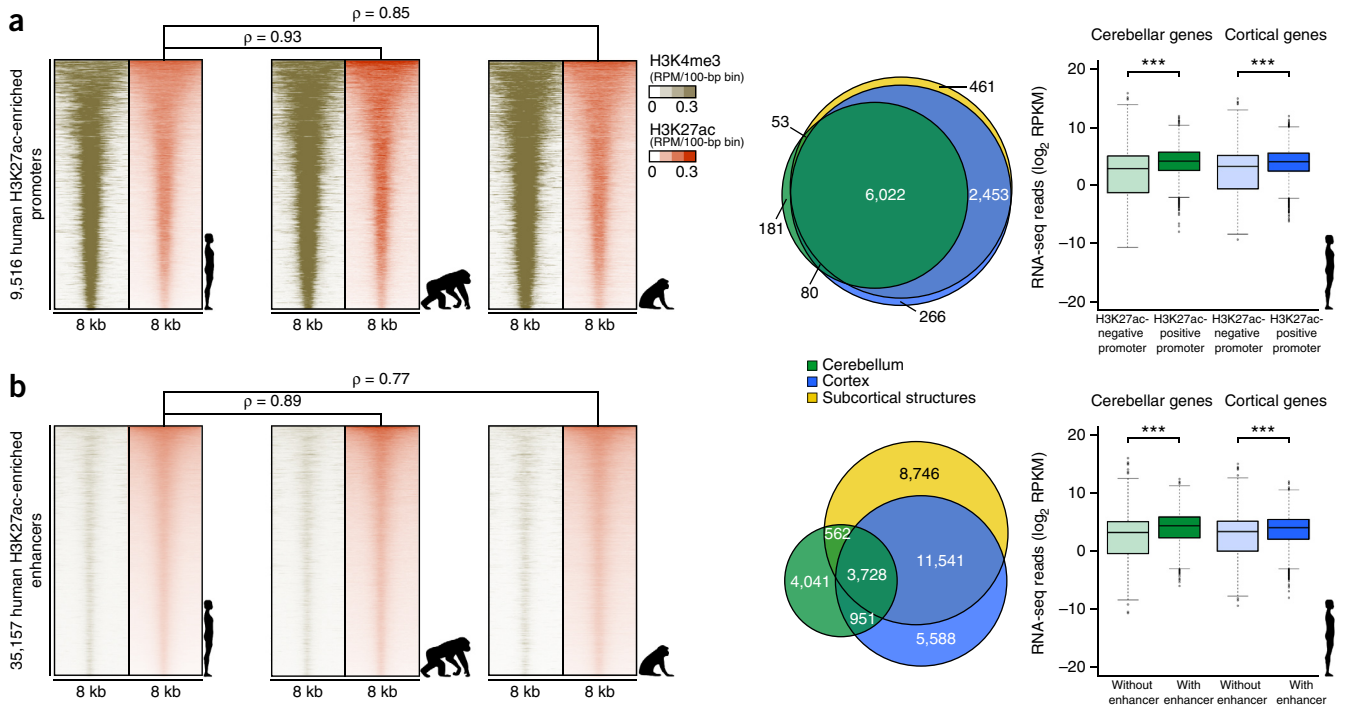

Figure 2 Comparison of histone enrichment at human brain enhancers and promoters. (**a**) Heatmaps of H3K4me3 and H3K27ac reads per million (RPM) normalized read counts at human promoters in all three species (left). Spearman's rank correlation coefficients for H3K27ac enrichment between the species are shown. Overlap of H3K27ac enrichment at promoters between the main anatomical subdivisions of the brain (middle). Normalized RNA read counts from human brain samples showing genes with and without H3K27ac enrichment at their promoter (right). Bottom and top of the boxes are the first and third quartiles. The line within the boxes represents the median and whiskers denote the interval within 1.5× the interquartile range from the median. Outliers are plotted as dots. Data are shown for cerebellum and cortex[33]. Dissimilarity between distributions was calculated using a Wilcoxon rank-sum test; $***P < 0.0005$. (**b**) Same analyses as in **a** for predicted human enhancers.

All data sets were well within quality parameters (**Supplementary Table 2** and Online Methods), and across the eight distinct brain areas 59,786, 59,155 and 61,795 genome-scale significant H3K27ac-enriched regions were identified in human, chimpanzee and rhesus macaque, respectively (**Supplementary Tables 3–5**). Overlap of H3K27ac-enriched regions between the three larger anatomical subdivisions of the brain (cerebellum, cortex and subcortical structures) confirmed substantial regulatory redundancy between anatomical brain areas (54–59% of the CREs) for each of the different species (random sampling, $P < 0.001$; **Fig. 1b**). Hierarchical clustering (**Fig. 1c,d**) and principal component analysis (PCA; **Fig. 1e,f**) for all chimpanzee and rhesus macaque data sets based on their predicted CREs demonstrated robust clustering according to overall anatomical position in the brain, similar to the human data (**Supplementary Fig. 1c,d** and **Supplementary Data 1**). Furthermore, 91% of the predicted CREs in the human cortex contained enhancer and/or promoter states based on chromatin signatures defined by ChromHMM[28–30] in the dorsolateral prefrontal cortex[24] (random sampling, $P < 0.001$; **Supplementary Fig. 2**). Conversely, 66–76% of the ChromHMM-predicted active enhancer states and 95% of the active promoter states were enriched for H3K27ac in our data ($P < 0.001$; **Supplementary Fig. 2**). Taken together, these data sets represent a comprehensive annotation of regulatory elements in the adult human, chimpanzee and rhesus macaque brains.

Validation of putative CREs across three primate genomes

To compare predicted CREs among the three species, we combined human and rhesus macaque enriched regions that were reproducible in multiple specimens and could be mapped to both other primate genomes[31,32]. This excluded chimpanzee-specific regulatory DNA, which was not the focus of our study. For 93% of the putative *cis*-regulatory elements, we found sequence orthologs in the human, chimpanzee and rhesus genome (assemblies hg38, panTro4 and rheMac3, respectively) hg38, panTro4 and rheMac3. After filtering for regions of poor genome quality and regions containing an excess of ambiguous reads or no reads (**Supplementary Fig. 3a** and Online Methods), we defined a final set of 60,702 H3K27ac-enriched regions with comparable mappability and size across the three species' genomes (**Supplementary Fig. 3b,c** and **Supplementary Table 6**).

As H3K27ac is enriched at promoter and enhancer CREs, the 60,702 enriched regions were intersected with known human transcriptional start sites (TSSs) to distinguish between them. 9,966 enriched regions were present at known TSSs, whereas we defined 50,736 as putative enhancers. Of those, 9,516 promoters and 35,157 enhancers were enriched for H3K27ac in the human brain (**Supplementary Fig. 3a**). The majority (86%) of promoters were also enriched for histone H3 lysine 4 trimethylation (H3K4me3, random sampling, $P < 0.001$), a histone mark selectively present at promoters (**Fig. 2a**)[24]. Similar fractions of predicted promoters were enriched for H3K4me3 in rhesus macaque (86%) and chimpanzee (85%). Although promoter enrichment was highly comparable between different anatomical subdivisions of the human brain (89% enriched in more than one subdivision; **Fig. 2a**), this similarity was substantially lower for predicted enhancers (47%; **Fig. 2b**). In addition, observed comparable histone signatures among the three species for H3K27ac at promoters (Spearman's rank correlation coefficient, $\rho = 0.93$, and $\rho = 0.85$ for human-chimp and human-rhesus, respectively; **Fig. 2a**) and to a lesser extend at enhancers ($\rho = 0.89$, $\rho = 0.77$; **Fig. 2b**). These data thus corroborate the tissue-specific nature of these CREs[17].

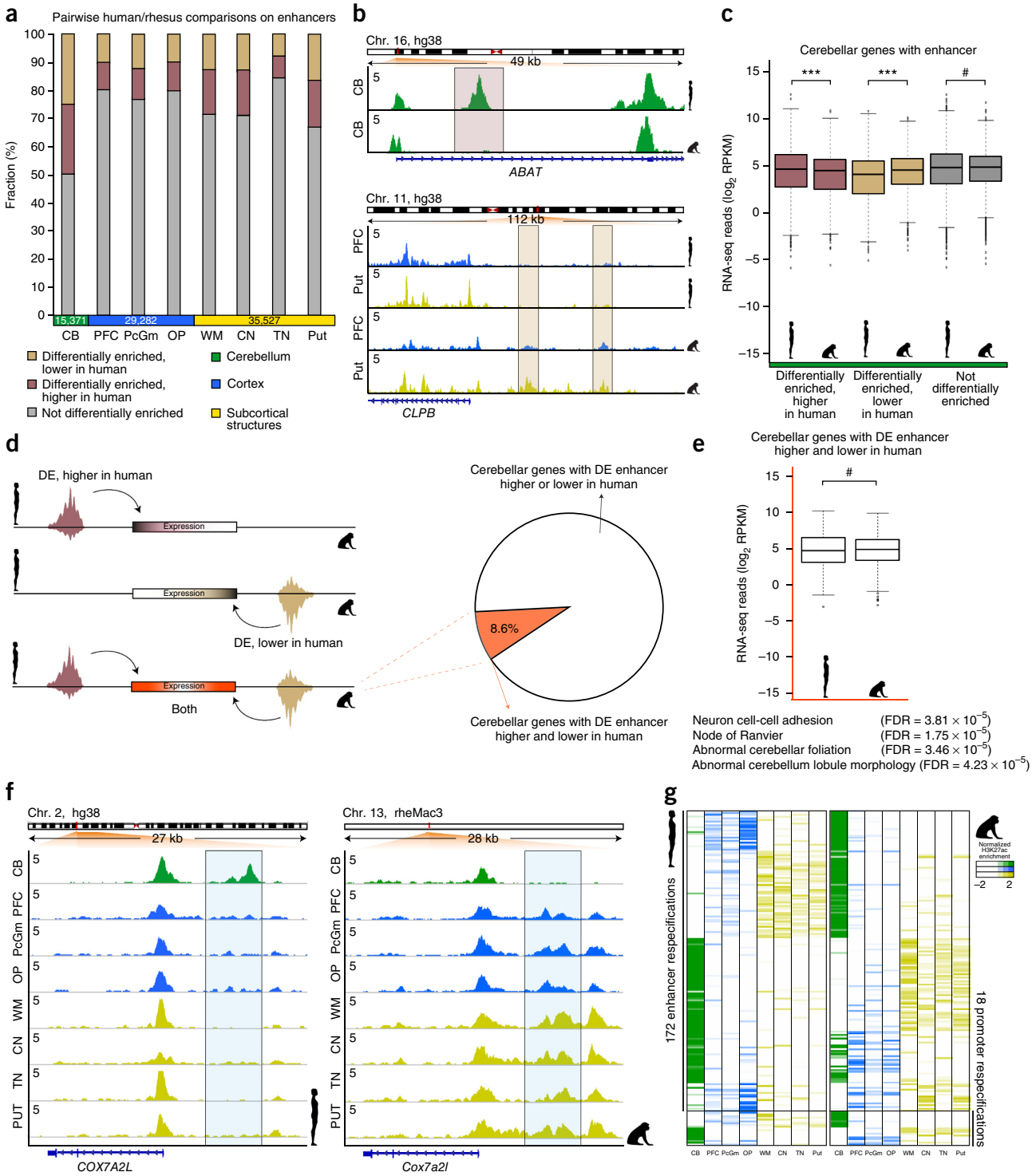

Figure 3 Differences in H3K27ac enrichment between human and rhesus macaque. (**a**) Fraction of predicted enhancers per brain region that is or is not significantly differentially enriched (DE) between human and rhesus defined by DESeq2. The number of analyzed enhancers per subdivision is indicated below the graph. (**b**) RPM normalized ChIP-seq reads for H3K27ac (axis limit 5 RPM) for a DE enhancer ('human gain', purple shading) in cerebellum (top). Analysis (bottom) for two DE enhancers ('human loss', light brown shading) in prefrontal cortex (blue tracks) and putamen (yellow tracks). (**c**) Comparison of normalized RNA read counts between human and rhesus for cerebellar genes linked to a DE or not DE enhancer. Dissimilarity between distributions was calculated using a Wilcoxon rank-sum test; $^{***}P < 0.0005$, $^{\#}P > 0.05$. (**d**) Cartoon depiction of gene regulation by single or multiple compensatory DE enhancers. Inset, fraction of cerebellar genes linked to a DE, higher and/or lower enhancer. (**e**) Normalized RNA read counts for human and rhesus showing cerebellar genes with a higher and lower DE enhancer. Bottom and top of the boxes are the first and third quartiles. The line within the boxes represents the median and whiskers denote the interval within 1.5× the interquartile range from the median. Outliers are plotted as dots. Wilcoxon rank-sum test; $^{\#}P > 0.05$. (**f**) RPM normalized ChIP-seq reads for H3K27ac at a genomic region containing the *COX7A2L* gene. Blue shading highlights the repurposed enhancer. (**g**) Heatmap of *z*-normalized RPKM values showing 172 predicted enhancer and 18 promoter respecifications. Columns depict the eight different brain regions that were analyzed for enrichment and used to *z*-normalize RPKM values for each enhancer (average of the three biological replicates per species per brain region is shown).

When comparing brain enhancers and co-regulated gene modules previously identified in different primate tissues[33], we found a relative enrichment for genes coupled to brain enhancers in brain-specific modules as compared to genes in modules specific to other tissues (Chi-squared test, $P = 5.39 \times 10^{-19}$; **Supplementary Fig. 3d** and Online Methods). Additionally, genes in cortical modules were significantly more often linked to putative enhancers that were exclusively present in cortex compared to unrelated tissues[24] (Fisher's exact test, for all pairwise comparisons $P < 0.0005$; **Supplementary Fig. 3e**). The same was true for genes in cerebellar modules that were linked to enhancers specific to the cerebellum ($P < 0.0005$, **Supplementary Fig. 3e**). Furthermore, genes in the vicinity of predicted brain-specific enhancers were overall linked to neural processes confirming a cell type–related gene regulatory function for these elements (**Supplementary Fig. 3f**). Finally, the genes that were closest to an enhancer were expressed at significantly higher levels than genes without a nearby enhancer (Wilcoxon rank-sum test, $P < 0.0005$; **Fig. 2b**). We also observed a similar enhancement of gene expression for genes with their promoters enriched for H3K27ac, which is consistent with previous findings[34] (Wilcoxon rank-sum test, $P < 0.0005$; **Fig. 2a**). Thus, the genomic regions identified here represent a set of putative promoters and enhancers with comparable sequence content across the three primate species.

Positional conservation of CREs across primates

As human and rhesus macaque are evolutionary most distant, we started by comparing the differences in active chromatin at CREs between these two species. We compared predicted human and rhesus macaque enhancers and promoters by overlapping the data for each anatomical subdivision of the brain (shown for CB and PFC in **Supplementary Fig. 4a**). As direct overlap of enriched regions suffers from thresholding biases[35], we also compared normalized read counts for each brain region between the human and rhesus macaque triplicate data sets. This allowed for the quantitative pairwise assessment of ChIP-seq signals between the two species using DESeq2 (ref. 36). A substantial amount (37–71% for enhancers, 55–73% for promoters in CB and PFC, respectively) of the regulatory differences predicted by direct overlap of the data sets based on peak calling were not significantly different as determined using DESeq2 (less than twofold-change and/or false discovery rate (FDR) ≥ 0.01), and thus we considered them to be false positives (**Supplementary Fig. 4a**). Indeed, expression of genes near differentially enriched CREs was overall significantly distinct between human and rhesus macaque brain tissue (Wilcoxon rank-sum test, $P < 0.05$; **Supplementary Fig. 4a**). We did not observe this effect for putative CREs that we considered false positives on the basis of DESeq2 analysis ($P > 0.05$; **Supplementary Fig. 4a**), thus validating this refinement.

We found significant changes in regulatory chromatin at promoter regions with lower frequency compared to putative enhancers, which is consistent with earlier reports proposing a role for enhancer alterations as a more prominent determinant of evolutionary change[16]. Nonetheless, most predicted CREs (57–86% for enhancers and 90–97% for promoters; **Supplementary Fig. 4b**) were either enriched in both species or not significantly different between human and rhesus macaque, demonstrating that the positional information of the regulatory landscape in the brain is largely conserved.

Despite this overall positional conservation, PhastCons analysis of these regions revealed minimal sequence conservation at CREs. Although promoters and putative enhancers showed significant conservation as compared to random DNA (Wilcoxon rank-sum test, $P < 0.0005$ for promoters and $P < 0.05$ for enhancers), the magnitude of this effect was marginal for putative enhancers (**Supplementary Fig. 4c**). This is consistent with earlier findings demonstrating limited sequence conservation at enhancers in spite of functional conservation[18,37,38]. However, with the resolution of putative enhancers identified here, we potentially missed conservation of small stretches of sequence containing transcription factor binding sites. Indeed, the fraction of reads in positionally conserved CREs that could map with one mismatch onto the genome of the other species was significantly higher in comparison to species-specific enhancers and promoters (Wilcoxon rank-sum test, $P < 0.0005$; **Supplementary Fig. 4d** and Online Methods). Thus, although sequence change associates with functional change, the overall functional conservation of CREs identified here cannot solely be explained by sequence conservation.

Quantitative assessment of H3K27ac enrichment at CREs

Several elements that we found enriched in both the human and rhesus macaque brain also demonstrated significant differences in H3K27ac enrichment (≥2-fold change and FDR <0.01; **Supplementary Fig. 4e**). Previous work has included these elements as regulatory gains of activity between species with evolutionary importance[19,20]. When assessing expression of nearby genes in our data sets, these corresponded to the quantitative increase or decrease of H3K27ac enrichment at those CREs (Wilcoxon rank-sum test, $P < 0.05$; **Supplementary Fig. 4e**), demonstrating that quantitative changes in enrichment are overall functional.

Therefore, we analyzed all enriched CREs per anatomical region by quantitative assessment of H3K27ac regardless of differences in peak calling and found that a considerable fraction of the regulatory elements was significantly different between human and rhesus macaque (≥ twofold change and FDR < 0.01, 16–50% for putative enhancers, 5–16% for promoters; **Fig. 3a**,**b**, **Supplementary Fig. 5a**,**b** and **Supplementary Tables 7–10**). Comparison of gene expression levels in cerebellum and prefrontal cortex between human and rhesus macaque[33] again demonstrated that gene expression changes correspond to differences in H3K27ac enrichment for both predicted enhancers (Wilcoxon rank-sum test, $P < 0.05$; **Fig. 3c** and **Supplementary Fig. 5c**) as well as promoters ($P < 0.05$; **Supplementary Fig. 5d**). CRE size changes between human and rhesus macaque were not more prominent in differentially enriched categories, suggesting this was unlikely to be an underlying cause for the observed changes in enrichment (**Supplementary Fig. 5e**).

Although the absolute number of changes per brain region was comparable, we found CREs in the cerebellum overall more frequently differentially enriched. This could have resulted from the fact that the cerebellum is a more homogeneous structure (>70% granule neurons)[1,39] resulting in better-resolved data sets, higher read counts per enriched region and thus more confident calling of differentially enriched regions. Skewing between the prefrontal cortex and cerebellum has also been observed in gene expression data comparing humans and primates, with the cerebellum counting most of the changes[33].

Across brain regions, 4.9–10.3% of the genes with an enhancer were associated with gain of enrichment at one element as well as a loss of enrichment at another (**Fig. 3d**). Observations in *Drosophila*[40] and across mammals[18] have indicated this phenomenon could represent compensatory mechanisms to correct for genetic alterations at enhancers near critical genes. In agreement with this, gene expression levels in cerebellum were not significantly different for genes that were associated with both regulatory gains and losses (Wilcoxon rank-sum test, $P > 0.05$; **Fig. 3e**). We did not observe this effect when we selected random sets of differentially enriched enhancers (permutation test, $P < 0.05$). Functional analysis of these genes confirmed an overall association with important neuronal processes that are linked to the

anatomical subdivision analyzed (**Fig. 3e** and **Supplementary Table 11**). Taken together, these data suggest that regulatory fine-tuning of existing CRE activity is widespread across the brain between human and rhesus macaque, and preferentially affects predicted enhancers. Furthermore, these results confirmed earlier observations of regulatory compensation near genes with critical functions (**Fig. 3d**,**e**).

Repurposing of CRE activity between brain structures

Repurposing of CREs has recently been shown between human and mouse using DNase hypersensitivity assays in different cell types[41]. We found that enrichment at CREs could switch between anatomical regions of the human and rhesus macaque brain (**Fig. 3f**). 172 predicted respecifications occurred at putative enhancers and 18 involved known human promoters (**Fig. 3g**). Most of these (90.7% for enhancers and 94.4% for promoters) involved repurposing between the cerebellum and other anatomical regions of the brain. For example, repurposing at an enhancer linked to the *DFNB31* gene coincided with promoter repurposing and a gene expression switch between cerebellum and cortex (**Supplementary Fig. 5f**). Thus, although CREs displayed overall positional conservation, a subset

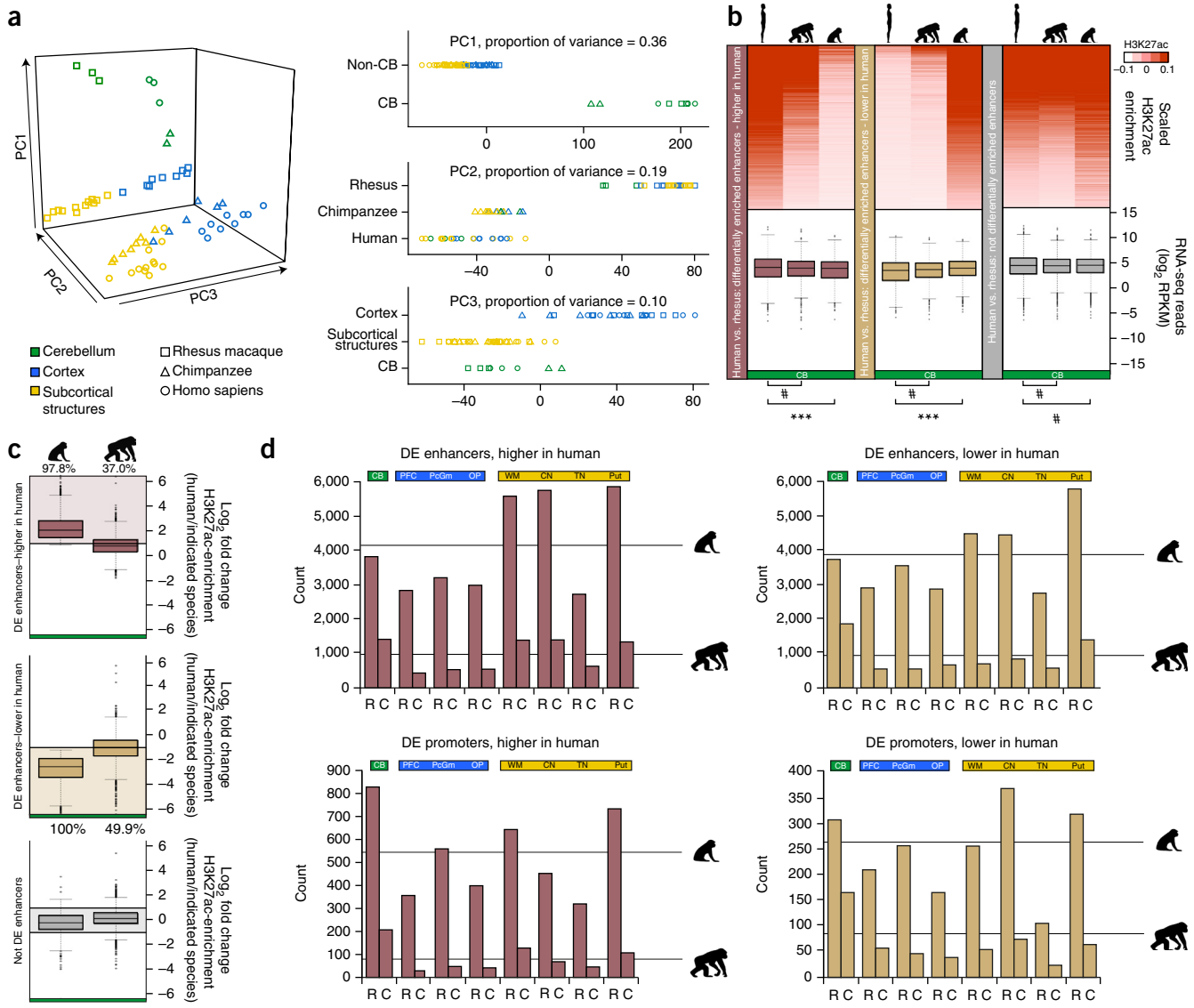

Figure 4 Changes in chromatin enrichment occurred primarily before the divergence of human and chimpanzee. (**a**) PCA of all human, chimpanzee and rhesus macaque brain samples based on normalized H3K27ac enrichment at 60,702 predicted human and rhesus CREs with orthologs on all three genomes (left). Variable driving the association plotted against the PC for each component separately (right). Bottom and top of the boxes are the first and third quartiles. The line within the boxes represents the median and whiskers denote the interval within 1.5× the interquartile range from the median. Outliers are plotted as dots. (**b**) Scaled H3K27ac enrichment of DE and not DE enhancers between human and rhesus macaque in cerebellum (CB) as indicated (top). The heat map displays regions that are more enriched in human (purple), more enriched in rhesus (light brown) and similarly enriched (gray). Normalized RNA read counts from human, rhesus macaque and chimpanzee brain samples for genes close to DE and not DE enhancers (bottom). Dissimilarity between distributions was calculated using a Wilcoxon rank-sum test; $^{***}P < 0.0005$, $^{\#}P > 0.05$. (**c**) Fraction of DE and not DE enhancers in cerebellum between human and rhesus macaque with at least a twofold difference in normalized read counts (mean of replicates) is highlighted for human gains (purple) and human losses (light brown). The fraction of enhancers that reach to twofold threshold is specified by the percentages. The same enhancers are shown in chimpanzee. Boxplots show the log2 fold change in H3K27ac enrichment at enhancers between human and the indicated species. (**d**) Number of DE enhancers and promoters per brain region with a twofold difference in normalized read counts between human and rhesus macaque (R), and human and chimpanzee (C). Horizontal lines indicate the mean number of differentially enriched CREs across the eight brain regions.

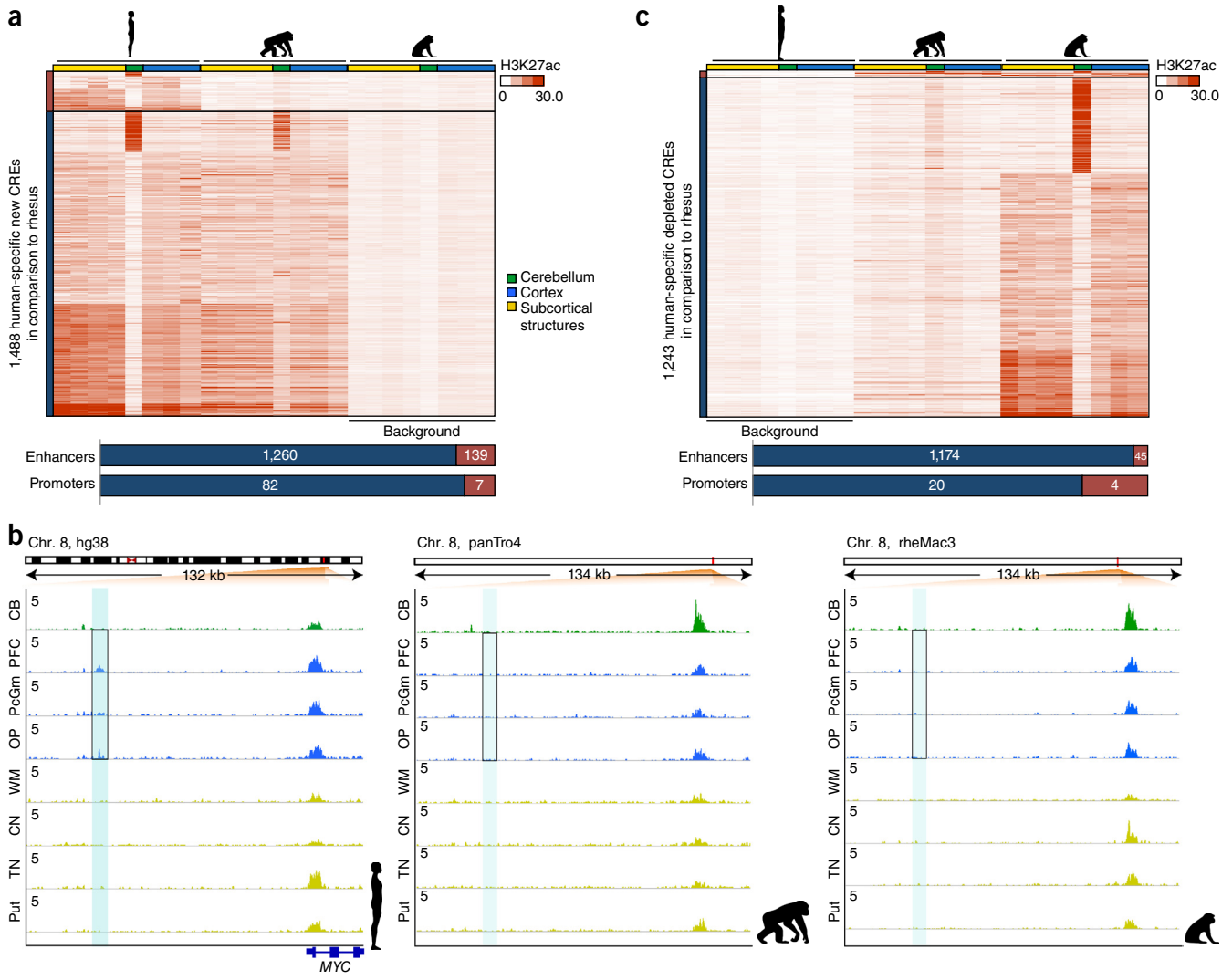

Figure 5 Newly introduced and specifically depleted CREs in human. (**a**) Heatmap of 1,488 human-specific CREs based on comparison with rhesus macaque. Read counts were normalized for number of reads in peaks, and the average value of replicates is plotted for the eight brain regions in human, chimpanzee and rhesus macaque. The upper section contains the human-specific gains (146 CREs) in comparison to both rhesus macaque and chimpanzee. The scale bar indicates RPM normalized H3K27ac enrichment. The bar plot illustrates the number of predicted enhancers and promoters that are specifically gained in human as defined by comparison to rhesus macaque only (blue), and to both rhesus macaque and chimpanzee (red). (**b**) RPM-normalized ChIP-seq read distribution (axis limit 5 RPM) for H3K27ac at a genomic region spanning 132 kilobases (kb) and containing the *MYC* gene in human (HS1), chimpanzee (Ch1) and rhesus macaque (RM1). Boxed regions highlight the cortical human-specific enhancer. (**c**) Same analysis as in **a** for human-specific depletions.

was differentially enriched and several CREs were repurposed between distinct anatomical regions.

Limited regulatory change after human-chimp divergence

As chimpanzee is evolutionary closer to human than rhesus macaque, a fraction of the observed regulatory changes between these two species is likely to have occurred before the split between human and chimpanzee. This is supported by PCA analysis for all human, chimpanzee and rhesus macaque samples based on all 60,702 predicted CREs in which PC2 distinguishes among the species (**Fig. 4a** and **Supplementary Data 2**). We obtained similar results using *t*-distributed stochastic neighbor embedding (t-SNE)-based multidimensional scaling (**Supplementary Fig. 6a**). Only a fraction of the differentially enriched CREs defined between human and rhesus macaque showed a similar twofold H3K27ac enrichment (reads per kilobase per million; RPKM) difference between human and chimpanzee (**Fig. 4b**,**c**, **Supplementary Fig. 6b**,**c** and **Supplementary Tables 12–15**). Accordingly, expression levels of the genes close to these enhancers were more comparable between human and chimpanzee than between human and rhesus macaque (**Fig. 4b** and **Supplementary Fig. 5b**). However, the tests that we used to accurately define differential H3K27ac enrichment at enhancers between triplicate samples of human and rhesus macaque could not be performed with the same statistical power using duplicate samples for chimpanzee. Therefore, it is possible that twofold-changed differentially enriched regions between human and chimpanzee are slightly overrepresented. On average, 22.5% of the differences between human and rhesus were also present between human and chimpanzee (**Fig. 4d**, and **Supplementary Tables 7–10** and **12–15**), demonstrating that a large proportion of the quantitative variation between human and rhesus macaque is absent in chimpanzee.

To test whether specific gene sets were overrepresented near human-specific regulatory gains or losses in the different brain regions,

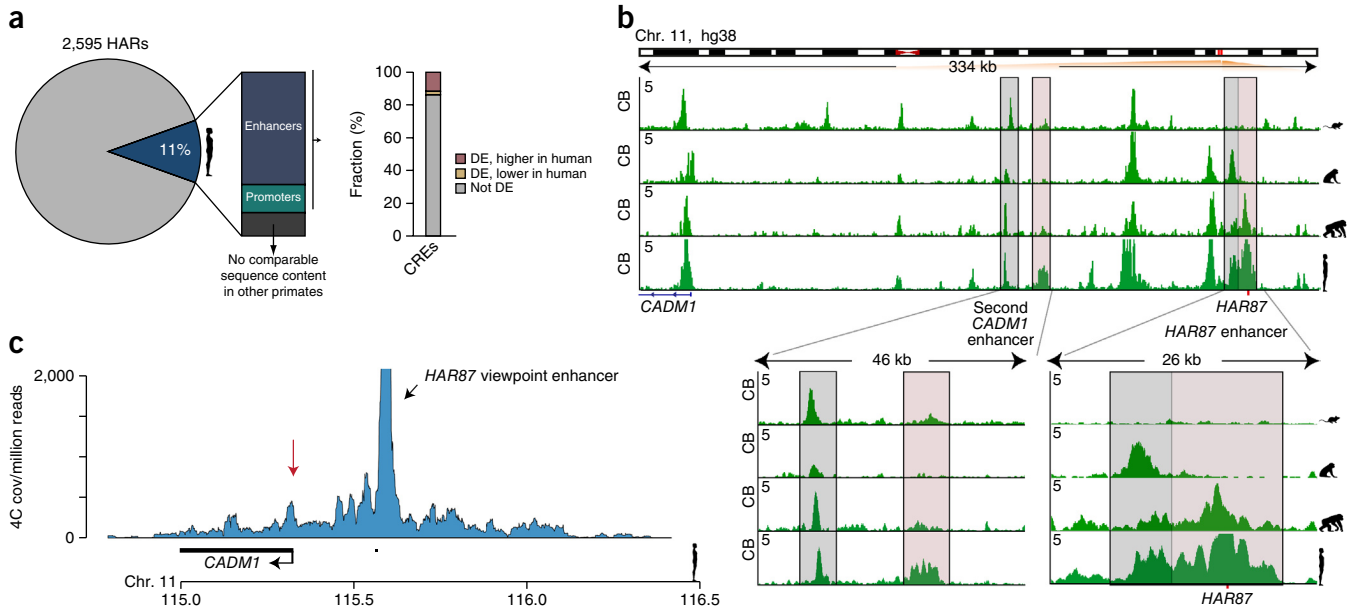

Figure 6 Functional alterations at accelerated DNA during primate evolution. (**a**) Percentage of HARs[44–46] that lie within putative brain CREs (left). Inset, distribution of enhancers, promoters and CREs that could not be mapped onto rheMac3 and panTro4 with similar sequence content. Percentages of CREs that were DE and were not DE in human versus rhesus macaque and chimpanzee (right). (**b**) RPM-normalized ChIP-seq read distribution (axis limit 5 RPM) for H3K27ac (top). Cerebellar tracks are shown for mouse, rhesus macaque (RM1), chimpanzee (Ch1) and human (HS1). Gray shades highlight shared enhancers among primates; purple shades highlight DE enhancers higher in human between human and rhesus, of which one contains *HAR87*, indicated in red. Bottom, zoom in on both enhancers upstream of the *CADM1* gene. (**c**) 4C-seq with the *HAR87* enhancer as viewpoint (black arrow). The red arrow (*CADM1* promoter) marks a local increase in 4C-seq coverage. A more detailed version of this panel matched to H3K27ac enrichment is available in **Supplementary Figure 8**.

we used gene ontology analysis to search for common functional classifications. We found no specific functional categories clearly enriched, and the statistical confidence for the identified terms was generally low (**Supplementary Table 16**). Although some of the associated terms were linked to genes with neural functions, many apparently unrelated processes were overrepresented at similar low confidence. This is not unlike related analyses previously done comparing human and chimpanzee using gene expression profiling[8,33]. It is possible that across the short evolutionary distance between human and chimpanzee, only a handful of altered elements with larger consequences are the predominant drivers of phenotypical divergence. Furthermore, the defining characteristics that separate the adult human brain from chimpanzee (which are unclear) might not be clearly represented in the gene ontology databases.

Rare newly introduced CREs in the human lineage

We next defined CREs predicted to be specifically active in human or rhesus macaque based on the strict absence of chromatin enrichment in the other species. We classified these as candidate enhancers and promoters for which activity was either newly introduced, or completely depleted in humans. 1,399 human enhancers and 89 promoters had enrichment levels comparable to the genome-wide background signal in all rhesus macaque brain regions, and we defined them as 'new' in the human brain (**Fig. 5a**, **Supplementary Tables 17** and **18**, and Online Methods). Only a small fraction of these putative CREs below background enrichment in rhesus macaque (~10%) was similarly below background enrichment in chimpanzee, leaving only a handful of CREs with real predicted human-specific activity (**Fig. 5a**). For example, we found a newly induced cortex-specific enhancer close to the *MYC* oncogene (**Fig. 5b**), a pleiotropic activator of transcriptional output[42,43]. Using similar stringent metrics, we found 1,243 H3K27ac-enriched regions in rhesus macaque to be enriched below background in human (**Fig. 5c**, and **Supplementary Tables 19** and **20**). The vast majority of these candidate depletions (~96%) were also not enriched in chimpanzee, demonstrating the activity of those CREs had been lost before the separation of human and chimpanzee (**Supplementary Fig. 7a**,**b**).

To assess whether newly introduced or completely depleted CREs occurred preferentially in particular brain regions, we compared the distribution of those CREs with those for all CREs across anatomical brain regions. Although predicted human-specific differentially expressed elements were enriched across all three major anatomical subdivisions of the brain, they were less frequently observed in multiple brain regions (permutation test, $P < 0.001$; **Supplementary Fig. 7c**). Depleted putative enhancers were similarly more often area-specific compared all enhancers ($P < 0.001$; **Supplementary Fig. 7d**). This suggests that the more pleiotropic effects of introducing or depleting CREs that are shared across multiple anatomical regions are also more often detrimental to the organism. These data support the hypotheses that enhancers are more amendable to evolutionary change at least in part because of their tissue specificity[12,14].

Enrichment of predicted CREs at human accelerated DNA

To further analyze predicted human CREs with changes in chromatin state, we cross-compared them with a set of previously defined ultraconserved genomic regions with accelerated sequence changes in the human lineage[44–46]. These human accelerated regions (HARs) were predicted to have altered regulatory elements near developmental and neuronal genes. 284 (out of 2,595)[44–46] predicted HARs were covered by an enriched region in human, which was significant compared to random genomic regions (random sampling, $P = 0.003$; **Fig. 6a** and **Supplementary Table 21**). 240 of those had

comparable sequence content onto rheMac3 and panTro4, and were further analyzed for differential enrichment and human specificity. 32 CREs (13.3%) were differentially enriched in human versus rhesus macaque and chimpanzee in at least one brain region (**Fig. 6a**). This is not significantly more than a random set of human CREs (permutation test, $P = 0.178$). In addition, none of the newly introduced CREs in comparison to both rhesus macaque and chimpanzee overlapped a HAR. Therefore, although HARs are overrepresented at regulatory elements of the adult brain, they were not found more often at differentially enriched enhancers. Similar results have been reported for the analysis of differentially enriched enhancers between human and rhesus macaque during embryogenesis in the brain[20]. Nevertheless, the intersection between the two data sets might still represent relevant evolutionary events. For instance, the 5′ region of the *CADM1* gene, a gene involved in synapse regulation of Purkinje cells in the cerebellum and ultrasonic vocalization in mice[45,47,48], gained enrichment next to an existing cerebellar enhancer in the rhesus macaque genome (**Fig. 6b**). This putative enhancer was absent in the mouse genome, demonstrating that it emerged in the primate lineage. We verified enhancer interaction with the promoter of *CADM1* as shown by circularized chromosome conformation capture (4C) sequencing (**Fig. 6c**) and found that the enhancer is likely part of a larger chromatin complex involving multiple promoter distal interaction sites within the region (**Fig. 6c** and **Supplementary Fig. 8**). However, despite the presence of a HAR, we also found the spreading of active chromatin covering this region in our chimpanzee data sets. Thus, our data allow the refinement of previously made predictions on the contribution of noncoding elements to the evolution of the human brain.

DISCUSSION

Understanding the evolutionary emergence and complexity of the human species requires the systematic annotation of coding as well as noncoding regulatory changes in the genome along the primate lineage. We used gain and loss of H3K27ac enrichment to describe differences at putative CREs[19,20] and demonstrate by gene expression analysis that this is relevant for both existing as well as new regulatory elements. As the presence of H3K27ac on a particular stretch of DNA is a binary event, the gains or losses of enrichment observed in this study will not necessarily translate directly into altered activity of the element in an episomal context. Recent single-cell chromatin accessibility data have demonstrated that population-based enrichment of open chromatin is best explained by the number of cells harboring the region in an open configuration[49,50]. It is therefore reasonable to assume that H3K27ac enrichment changes are mostly a reflection of the number of cells that contain the CRE in a stable active state. It will be interesting to combine these data with large-scale episomal analysis of CREs[40] in primates to couple activity to availability.

Our data, which were generated from the adult brain, are highly complementary to recently published data describing the changes between fetal stages of brain development using rhesus macaque and human samples[20], and add complexity with the analysis of various functionally different anatomical regions. Comparing human and chimpanzee, we found no clear functional trends in the data for genes that are close to differentially enriched CREs. We speculate that the fraction of phenotypically defining changes in these data sets might be too low across such a short evolutionary distance. Small trends in gene-set analyses could be masked by a surplus of other alterations and/or changes with small phenotypical effects not directly linked to human brain function or by enhancer changes shared and selected for in different tissues. Indeed, alterations at individual CREs have been previously linked to significant phenotypical changes between species[12,13]. The data in this study will likely be of value to further identify and prioritize such elements.

In summary, our data allow the identification of regulatory changes that occurred in the brain during the final stages of human evolution. Combined with genomic and transcriptomic data, this work represents a framework from which single predicted regulatory elements can be further assessed using more targeted approaches.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Gene Expression Omnibus: GSE67978.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank S. Levine at the MIT BioMicro Center for managing the Solexa pipeline. This work was funded by the Dutch Royal Academy of Sciences (KNAW), the Dutch Cancer Foundation (KWF) grant HUBR2012-5392 (M.P.C.) and the Parkinson's foundation (Stichting ParkinsonFonds) (M.P.C.).

AUTHOR CONTRIBUTIONS

M.W.V. and M.P.C. conceived and designed the experiments. N.B.B. contributed human brain hemispheres. R.B., I.K. and S.P. provided nonhuman primate material from rhesus macaque and chimpanzee. M.W.V., B.C. and P.R. performed the experiments. M.W.V. and S.C.T. analyzed the data, and were supervised by M.P.C. G.G. analyzed 4C experiments, and was supervised by W.d.L. Sequencing at the Utrecht Sequencing Facility was performed by E.d.B., supervised by E.C. M.W.V. and M.P.C. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Azevedo, F.A. *et al.* Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *J. Comp. Neurol.* **513**, 532–541 (2009).
2. Rakic, P. Evolution of the neocortex: a perspective from developmental biology. *Nat. Rev. Neurosci.* **10**, 724–735 (2009).
3. Geschwind, D.H. & Rakic, P. Cortical evolution: judge the brain by its cover. *Neuron* **80**, 633–647 (2013).
4. Rogers, J. & Gibbs, R.A. Comparative primate genomics: emerging patterns of genome content and dynamics. *Nat. Rev. Genet.* **15**, 347–359 (2014).
5. O'Bleness, M., Searles, V.B., Varki, A., Gagneux, P. & Sikela, J.M. Evolution of genetic and genomic features unique to the human lineage. *Nat. Rev. Genet.* **13**, 853–866 (2012).
6. King, M.C. & Wilson, A.C. Evolution at two levels in humans and chimpanzees. *Science* **188**, 107–116 (1975).
7. Haygood, R., Babbitt, C.C., Fedrigo, O. & Wray, G.A. Contrasts between adaptive coding and noncoding changes during human evolution. *Proc. Natl. Acad. Sci. USA* **107**, 7853–7857 (2010).
8. Konopka, G. *et al.* Human-specific transcriptional networks in the brain. *Neuron* **75**, 601–617 (2012).
9. Shlyueva, D., Stampfel, G. & Stark, A. Transcriptional enhancers: from properties to genome-wide predictions. *Nat. Rev. Genet.* **15**, 272–286 (2014).
10. Hong, J.W., Hendrix, D.A. & Levine, M.S. Shadow enhancers as a source of evolutionary novelty. *Science* **321**, 1314 (2008).
11. Perry, M.W., Boettiger, A.N., Bothma, J.P. & Levine, M. Shadow enhancers foster robustness of *Drosophila* gastrulation. *Curr. Biol.* **20**, 1562–1567 (2010).
12. Wittkopp, P.J. & Kalay, G. Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nat. Rev. Genet.* **13**, 59–69 (2012).
13. Rubinstein, M. & de Souza, F.S. Evolution of transcriptional enhancers and animal diversity. *Phil. Trans. R. Soc. Lond. B* **368**, 20130017 (2013).
14. Wray, G.A. The evolutionary significance of cis-regulatory mutations. *Nat. Rev. Genet.* **8**, 206–216 (2007).
15. Schmidt, D. *et al.* Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. *Science* **328**, 1036–1040 (2010).
16. Villar, D. *et al.* Enhancer evolution across 20 mammalian species. *Cell* **160**, 554–566 (2015).

17. Bulger, M. & Groudine, M. Functional and mechanistic diversity of distal transcription enhancers. *Cell* **144**, 327–339 (2011).
18. Villar, D., Flicek, P. & Odom, D.T. Evolution of transcription factor binding in metazoans-mechanisms and functional implications. *Nat. Rev. Genet.* **15**, 221–233 (2014).
19. Cotney, J. *et al.* The evolution of lineage-specific regulatory activities in the human embryonic limb. *Cell* **154**, 185–196 (2013).
20. Reilly, S.K. *et al.* Evolutionary genomics. Evolutionary changes in promoter and enhancer activity during human corticogenesis. *Science* **347**, 1155–1159 (2015).
21. Visel, A. *et al.* A high-resolution enhancer atlas of the developing telencephalon. *Cell* **152**, 895–908 (2013).
22. Creyghton, M.P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. USA* **107**, 21931–21936 (2010).
23. Prescott, S.L. *et al.* Enhancer divergence and cis-regulatory evolution in the human and chimp neural crest. *Cell* **163**, 68–83 (2015).
24. Kundaje, A. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
25. Rada-Iglesias, A. *et al.* A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**, 279–283 (2011).
26. Vermunt, M.W. *et al.* Large-scale identification of coregulated enhancer networks in the adult human brain. *Cell Rep.* **9**, 767–779 (2014).
27. Kumar, S. & Hedges, S.B. A molecular timescale for vertebrate evolution. *Nature* **392**, 917–920 (1998).
28. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nat. Methods* **9**, 215–216 (2012).
29. Ernst, J. & Kellis, M. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat. Biotechnol.* **28**, 817–825 (2010).
30. Ernst, J. & Kellis, M. Large-scale imputation of epigenomic datasets for systematic annotation of diverse human tissues. *Nat. Biotechnol.* **33**, 364–376 (2015).
31. Gibbs, R.A. *et al.* Evolutionary and biomedical insights from the rhesus macaque genome. *Science* **316**, 222–234 (2007).
32. Chimpanzee Sequencing and Analysis Consortium. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437**, 69–87 (2005).
33. Brawand, D. *et al.* The evolution of gene expression levels in mammalian organs. *Nature* **478**, 343–348 (2011).
34. Wang, Z. *et al.* Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat. Genet.* **40**, 897–903 (2008).
35. Bardet, A.F., He, Q., Zeitlinger, J. & Stark, A. A computational pipeline for comparative ChIP-seq analyses. *Nat. Protoc.* **7**, 45–61 (2012).
36. Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
37. Hare, E.E., Peterson, B.K., Iyer, V.N., Meier, R. & Eisen, M.B. Sepsid even-skipped enhancers are functionally conserved in *Drosophila* despite lack of sequence conservation. *PLoS Genet.* **4**, e1000106 (2008).
38. Fisher, S., Grice, E.A., Vinton, R.M., Bessling, S.L. & McCallion, A.S. Conservation of RET regulatory function from human to zebrafish without sequence similarity. *Science* **312**, 276–279 (2006).
39. Andersen, B.B., Korbo, L. & Pakkenberg, B. A quantitative study of the human cerebellum with unbiased stereological techniques. *J. Comp. Neurol.* **326**, 549–560 (1992).
40. Arnold, C.D. *et al.* Quantitative genome-wide enhancer activity maps for five *Drosophila* species show functional enhancer conservation and turnover during cis-regulatory evolution. *Nat. Genet.* **46**, 685–692 (2014).
41. Vierstra, J. *et al.* Mouse regulatory DNA landscapes reveal global principles of cis-regulatory evolution. *Science* **346**, 1007–1012 (2014).
42. Rahl, P.B. & Young, R.A. MYC and transcription elongation. *Cold Spring Harb. Perspect. Med.* **4**, a020990 (2014).
43. Lin, C.Y. *et al.* Transcriptional amplification in tumor cells with elevated c-Myc. *Cell* **151**, 56–67 (2012).
44. Prabhakar, S., Noonan, J.P., Paabo, S. & Rubin, E.M. Accelerated evolution of conserved noncoding sequences in humans. *Science* **314**, 786 (2006).
45. Bird, C.P. *et al.* Fast-evolving noncoding sequences in the human genome. *Genome Biol.* **8**, R118 (2007).
46. Lindblad-Toh, K. *et al.* A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* **478**, 476–482 (2011).
47. Fujita, E., Tanabe, Y., Imhof, B.A., Momoi, M.Y. & Momoi, T. Cadm1-expressing synapses on Purkinje cell dendrites are involved in mouse ultrasonic vocalization activity. *PLoS One* **7**, e30151 (2012).
48. Boyd, J.L. *et al.* Human-chimpanzee differences in a FZD8 enhancer alter cell-cycle dynamics in the developing neocortex. *Curr. Biol.* **25**, 772–779 (2015).
49. Buenrostro, J.D. *et al.* Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486–490 (2015).
50. Cusanovich, D.A. *et al.* Epigenetics. Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* **348**, 910–914 (2015).

ONLINE METHODS

Selection of brain regions. In a previous study, in which 87 distinct anatomical regions of the human brain had been analyzed, three main subdivisions of brain regions could be distinguished by clustering and PCA analysis: cortex, subcortical structures and cerebellum[26] (**Supplementary Fig. 1a**), suggesting most variation existed between these structures. Therefore, in the current analyses we selected eight brain regions corresponding to these structures. 72% of the regulatory elements we had identified previously[26], were also recovered in these human data sets, demonstrating that the majority of regulatory elements can be characterized using a smaller set of diverse anatomical regions. Throughout the figures, cortical samples are depicted in blue, subcortical structures in yellow and cerebellum in green.

Sample collection. Samples from *Homo sapiens* (HS) donors HS1 and HS2 have been described previously[26]. Brain regions from donors HS3 and HS4 were obtained from the Netherlands brain bank (http://www.brainbank.nl/) (**Supplementary Table 1**). Informed consent was obtained for all donors. HS3 died from complications while suffering from a brain tumor that was well-defined and not invasive. We omitted dissecting near the area where the tumor had been removed surgically before death. As we did not obtain cerebellum from HS3, the third cerebellum sample stems from a different donor (HS4). This HS4 sample is referred to as HS3 in the paper. Rhesus macaque (RM) 1, RM2 and RM3 brains were collected at the Biomedical Primate Research Centre (BPRC) in Rijswijk (http://www.bprc.nl/) and represent rest material involving no animal experimentation for the purpose of this work as determined by the Animal Experimental Committee (DEC). Chimpanzee samples (Ch1, Ch2 and Ch3) were obtained via the BPRC after the animals died in a local zoo due to natural causes (**Supplementary Table 1**). All tissues were frozen as fast as possible after death (post mortem delays are listed in **Supplementary Table 1**). Eight brain regions, covering anatomically and functionally different areas, were analyzed (**Supplementary Fig. 1a** and **Supplementary Table 2**). These include the occipital pole (OP, primary visual cortex), precentral gyrus (PcGm, primary motor cortex) and prefrontal cortex (PFC). Three other dissected brain regions covered diverse midbrain nuclei: thalamic nuclei (TN), putamen (Put) and caudate nucleus (CN). Finally, a subcortical section of white matter (WM) and part of the cerebellar cortex (CB) were analyzed. Dissections were done in a −20 °C climate-controlled room using precooled biopsy punches (Microtech), and regions that were taken out are listed in **Supplementary Table 2** and depicted in **Supplementary Figure 1a**,**b** for rhesus macaque (RM1) and chimpanzee (Ch3). For both chimpanzees (Ch1 and Ch3) one of the eight brain regions was not available. Therefore, replicate samples for CB and the TN were isolated from a third chimpanzee (Ch2) and added to complete Ch1 and Ch3. These two complete sets were given Ch1 and Ch2 designations throughout the manuscript.

Chromatin immunoprecipitation–sequencing. ChIP was carried out as described previously[26] with the following adaptations. 60 mg of tissue was used per ChIP and homogenized in a glass douncer (Kontes Glass Co.) for further analysis. After crosslinking and washing in cold PBS, cells were lysed as described before[26]. Nuclei were sonicated in 80 µl of buffer in microtubes using the Covaris S series. The following settings were used for 12 cycles of 60 s: intensity 3, duty cycle 20%, 200 cycles/bursts. After sonication, lysis buffer and Triton X-100 (to a final concentration of 1%) were added to a total volume of 550 µl. Samples were spun for 10 min at 21,139*g* at 4 °C. Supernatant was transferred onto DynaI Protein G beads, which were preincubated with antibody in PBS with 0.5% BSA at 4 °C for at least 4 h, and rotated overnight at 4 °C. Antibodies used were ab4729 from Abcam for H3K27ac and 07-473 from Millipore for H3K4me3. The next day, beads were washed four times with RIPA (50 mM Hepes pH 7.6, 1 mM EDTA, 0.7% DOC, 1% Igepal and 0.5 M LiCl) and once using TE with 50 mM NaCl. Elution and reverse crosslinking were done overnight at 65 °C in TE with 1% SDS. Beads were collected and supernatant was diluted 1:1 with TE. RNase A (final concentration of 0.2 µg/µl) was added for 2 h at 37 °C. After a subsequent proteinase K (final concentration of 0.2 µg/µl) treatment of 2 h at 55 °C, DNA was extracted with phenol/chloroform in MaXtract High Density gel tubes (Qiagen) and ethanol-purified. DNA was prepared for sequencing according to the Illumina Truseq DNA library protocol, and sequencing was done at the MIT BioMicro Center (http://openwetware.org/wiki/BioMicroCenter) or at the Utrecht DNA Sequencing facility (http://www.utrecht-sequencing-facility.nl) using the Illumina HiSeq 2000 or NextSeq 500 genome sequencer. Separate lanes, containing the eight different brain regions, were sequenced per primate specimen (Ch1, Ch2, RM1, RM2 and RM3) to avoid batch effects.

Data analysis. *ChIP-seq enrichment analysis.* Images acquired from the Illumina/Solexa sequencer were processed using the bundled Solexa image extraction pipeline (version 1.5 or 1.6 (cassava)). Sequences were aligned using Bowtie 1.1.0 (ref. 51; bowtie-bio.sourceforge.net) excluding reads with more than one mismatch (seed length 40) or with multiple alignments, unless stated otherwise. Mapping was done onto the following reference genomes: mouse mm10, rhesus macaque rheMac3, chimpanzee panTro4 and human hg38. Between 8 million and 36 million reads were mapped for each ChIP sample in the data sets. Fractions of reads in peaks scores all exceeded the 1% threshold used by the Encyclopedia of DNA Elements (ENCODE)[52], and the percentage of mapped and unique reads was overall high (**Supplementary Table 2**). Statistically genome-wide significant enriched regions for H3K27ac and H3K4me3 were called per sample using MACS2 version 2.1.0 (ref. 53) ($P = 10^{-5}$, extsize = 400, local lambda = 100,000). Whole-cell extract input controls were generated for each brain region of every species. However, we used the internal lambda control in the MACS algorithm to correct for local bias as whole-cell extract inputs often introduce sonication biases at open chromatin[54]. Identified peaks smaller than 2,000 base pairs (bp) were extended to a size of 2,000 bp (peak center ± 1,000 bp) to match peak resolution usually observed for this histone mark[16,26,55]. Reproducible enriched regions were defined as enriched regions present in at least two biological replicate samples stemming from the same larger anatomical subdivision of the brain (CB, cortex and subcortical structures) per species. As only two biological samples could be obtained for chimpanzee, all called peaks were considered for this species. This did not influence our analyses, as three way comparisons between species are based on read counts (see below). Lists of enriched regions per species (59,786, 59,155 and 61,795 for human, chimpanzee and rhesus macaque, respectively) were obtained by merging enriched regions from all brain samples per species (24 for human, 16 for chimpanzee and 24 for rhesus macaque), with overlapping regions being stitched together (**Supplementary Tables 3–5**).

Hierarchical clustering, PCA and t-SNE. Duplicate reads were removed from bam files using Samtools 0.1.19 (ref. 56), and read coverage within enhancer regions was calculated using Bedtools v2.20.0 (ref. 57). In the analysis involving single species (**Fig. 1c–f** and **Supplementary Fig. 1c**,**d** and **Supplementary Data 1**), read counts were normalized for number of reads in peaks and $\log_2$ transformed using the rlog (blind) function in DESeq2. Subsequently, values for each CRE were *z*-normalized across the eight different brain regions for every specimen (HS1, HS2, HS3, RM1, RM2 and RM3) individually. For hierarchical clustering (**Fig. 1c**,**d** and **Supplementary Fig. 1c**), Pearson correlations between samples were calculated. Samples were clustered based on Pearson distance with average linkage, and heatmaps were made using heatmap.2 from the gplots R package. For PCA (**Fig. 1e**,**f** and **Supplementary Fig. 1d**), the prcomp function in R (http://www.R-project.org/)[58] was used.

For the analysis involving multiple species simultaneously (**Fig. 3a** and **Supplementary Fig. 6a** and **Supplementary Data 2**), read counts were normalized for the number of reads in peaks and $\log_2$ transformed using the rlog (blind) function in DESeq2. PCA was then performed as described above, and plots were generated using scatter3D from the plot3D R package. For the analysis showing separate components, the variable driving the association was plotted against the sample value for that component. t-SNE multidimensional scaling coordinates were determined using the t-SNE R package[59].

Cross-species comparison of H3K27ac-enriched regions. Coordinates for rhesus macaque H3K27ac-enriched regions on hg38 were obtained using the UCSC liftOver tool (-minMatch = 0.5). Mapping onto the target genome and back to the source genome (reciprocal liftOver) had to be unique. Regions that changed more than 50% in size during liftOver were discarded. Rhesus macaque enriched regions mapped on hg38 were subsequently merged with human enriched regions with overlapping regions being combined into one to a total of 77,316 CREs (**Supplementary Fig. 3a**). To obtain three-way orthologs (hg38, panTro4 and rheMac3), reciprocal liftOver was performed centered from the hg38 coordinates to rheMac3 as well as panTro4. The intersection of regions with coordinates on both target species was used for further analysis. To correct for variation in genome quality, selected enriched regions with orthologs on all genomes had to meet the following criteria: first, ≥ 90% of the bases had to be known in all three

doi:10.1038/nn.4229

reference genomes (<10% overlap with UCSC Table Browser's Gap Locations lists). This setting was used because unknown bases typically appear in short stretches rather than being scattered. Second, CREs with zero reads were discarded. Third, enrichment scores for the regions were not allowed to change significantly in the target genome when allowing reads to map to multiple locations as these could represent duplicated regions that are highly susceptible to poor annotation in lower-quality genomes. To analyze this, bam files for all samples per species were merged, and reads were mapped to unique locations (bowtie:–best–strata -m 1) as well as multiple locations (bowtie:–best–strata -M 1). Genomic regions that were called as a peak after allowing multiple mapping but not with unique mapping were discarded as they potentially represent duplications or repeat elements that are not annotated at similar depth among the three genomes. 60,702 of 77,316 human and rhesus CREs were matched on all three genomes and passed all filters (**Supplementary Fig. 3a** and **Supplementary Table 6**). For most CREs (97.3%), the size change after reciprocal liftOver was below 25% (**Supplementary Fig. 3c**).

Promoter and enhancer analysis for human and rhesus macaque enriched regions. H3K27ac-enriched regions with coordinates on all genomes were grouped into enhancers and promoters using the hg38 RefSeq list. Enriched regions that overlapped within 1,000 bp from known transcriptional start sites (TSSs) were annotated as promoters, and enriched regions that were located more than 1,000 bp away from TSSs were considered putative enhancers. To check whether these sequences were indeed marked with histone signatures known to be selectively present at promoters, H3K4me3 ChIP-seq was performed on a subset of samples in human, chimpanzee and rhesus macaque (**Supplementary Table 2**). Duplicate reads were removed using Samtools 0.1.19 (ref. 56) and H3K4me3 bam files were merged per species. Reads were counted for a region of 8 kb (80 bins of 100 bp) around the 9,516 putative human TSSs and around the center of 35,157 predicted human enhancers using Bedtools v2.20.0 9 (ref. 57). The same was done for the three species on merged H3K27ac bam files per species. Read density profiles were RPM-normalized and sorted on total read counts per promoter or enhancer in human. Heatmaps were visualized using Java Treeview (http://jtreeview.sourceforge.net)[60]. Spearman's rank correlation coefficient for H3K27ac enrichment between the species was calculated on the total coverage per enhancer (sum of all bins).

PhastCons scores. Scores for nucleotide conservation were calculated for each CRE using the PhastCons 20 mammals track. Mean values of all scored nucleotides per CRE were plotted for the indicated elements in **Supplementary Figure 4c**. Significance for promoters and enhancers was calculated using a Wilcoxon rank-sum test against the same number (9,966 for promoters and 50,763 for enhancers) of random genomic fragments of the same size. Selection of random genomic DNA is further specified below.

Identification of positionally conserved CREs. Promoters and enhancers in cerebellum, cortex and subcortical structures were compared to the enriched regions of those subdivisions in the other species (human and rhesus macaque). As peak calling suffers substantially from thresholding artifacts[35], read counts of enriched regions that were only called in one species, were compared using DEseq2 (below) for the eight brain regions. CREs that were not significantly DE and not DE between human and rhesus macaque, were assigned as false positive species-specific elements. CREs that were called as peak in both species were grouped with not DE CREs into positionally conserved elements.

Mappability of human and rhesus macaque reads onto the other genome. All bam files (24 samples) were merged per species for both human and rhesus macaque. Duplicate reads were removed using Samtools 0.1.19 (ref. 56) and read coverage within CREs was calculated using Bedtools v2.20.0 (ref. 57). The human reads that fell within a given CRE were selected and mapped onto rheMac3 using the mapping settings described above (one mismatch allowed). Percentages of human reads that mapped to the corresponding CRE on rheMac3 were plotted (**Supplementary Fig. 4d**). The same analysis was done for rhesus macaque reads within CREs onto the human genome.

Gene expression analysis. All CREs identified here were coupled to their closest gene (TSS) using the UCSC hg38 RefSeq list. Previously published gene expression data from six tissues in ten species (including prefrontal cortex and cerebellum in human, rhesus macaque and chimpanzee) was then used for further comparison[33]. RPKM RNA counts for 13,035 genes with orthologs across primates were extracted for the female prefrontal cortex and cerebellum samples (Constitutive Aligned Exons, Primate1to1Orthologs). Out of those, 12,885 were traced back on hg38 (Ensembl genes 81). Boxplots throughout the manuscript depict $\log_2$ RPKM RNA read count values for genes, within those 12,885 orthologous genes that were coupled to CREs based on proximity. Dissimilarity of RNA expression distributions was calculated using a nonparametric Wilcoxon rank-sum test (two-tailed) under the null hypothesis that the distributions are the same. For the permutation test in **Figure 3e**, the difference between the median of the human and rhesus distribution was compared to random sets of DE enhancers.

We used gene coexpression modules, previously identified in primates in six different tissues[33], to compare to our lists of CREs. All modules that were expressed in human, or in human and other primates, were selected for analysis. The percentage of genes within previously defined gene modules[33] that could be coupled to a CRE was calculated. To test the null hypothesis, which states that genes in any of the tissue modules (multiple modules of a single tissue) are equally likely to be linked to the brain enhancers identified here, we selected the 19,560 brain enhancers that had one of the 12,885 orthologous genes as a closest partner for further analyses. The number of unique genes for each tissue module (for example, all unique cerebellar module genes) was counted and for each of those, we analyzed whether the gene was linked to a brain enhancer or not. By doing so, a 2×5 contingency table of counts was created in which the row variable represents whether or not a gene is linked to a brain enhancer and the columns correspond to the 5 different tissue modules (cortex, cerebellum, heart, kidney and liver). A chi-squared test for contingency tables was performed to test the null hypothesis of independence between rows and columns. The enrichment of plotted genes (**Supplementary Fig. 3d**) is the square root of the contribution to the test statistic of the corresponding cell in the table. A comparable analysis was done to assess enrichment of cortical/cerebellar module genes linked to cortex/cerebellum-specific enhancers on the basis of overlap with 13 unrelated tissues. These include two replicates from colon crypt[61], adrenal gland, esophagus, gastric tissue, left ventricle of the heart, lung, pancreas, psoas muscle, sigmoid colon, small intestine, spleen, kidney and liver[24,62]. Link counts for all genes within separate modules per tissue linked to cortex/cerebellum brain-specific enhancers are shown in **Supplementary Figure 3e**. To make the comparisons more stringent, only unique genes per tissue module were considered for the statistical analysis of enrichment between two modules. Fisher's exact test was used to test the null hypothesis that enrichment for cortex/cerebellum brain-specific enhancers was equal in pairs of gene modules being compared.

RPKM normalized and scaled H3K27ac enrichment. H3K27ac enrichment within CREs was normalized for the number of reads in peaks and region size in **Figure 5a,c** and **Supplementary Figure 5a**. For the analyses in **Figure 4b** and **Supplementary Figure 6b**, these values were subsequently *z*-normalized per sample for each species using the scale function in R. Average values of all replicates per species were visualized using Java Treeview.

Analysis of variation in enrichment at promoters and enhancers. Pairwise comparison for each of the eight brain regions between human and rhesus macaque was done for all CREs using DESeq2 based on the three independent replicates[36]. CRE size-normalized read counts of the human replicates (HS1, HS2 and HS3; human group) were compared to those of the rhesus macaque replicates (RM1, RM2 and RM3; rhesus group). For each subdivision of the brain (cortex, subcortical structures and cerebellum), peaks identified in that subdivision were analyzed for differential enrichment. Significantly differentially enriched (DE) regions were defined as regions with at least a twofold change in enrichment and an FDR < 0.01. DE enhancers and promoters between human and rhesus are listed for the eight brain regions in **Supplementary Tables 7–10**. For human-chimpanzee comparisons, reads in peaks normalized read counts were compared. The mean was calculated for all replicates per species. A twofold-change cutoff was used to define regions as DE between both species (**Supplementary Tables 12–15**).

Identification of respecifications. DE CREs with higher expression in human ('gains') were merged for all eight brain regions. The same was done for DE CREs with lower expression in human ('losses'). When CREs were present in both lists, but not in a list containing regions that remained constant (not DE), they were assigned as respecifications. Read counts within repurposed CREs were normalized to the number of reads in peaks and region size. Subsequently, these values were *z*-normalized over the eight different brain regions and visualized using Java Treeview (**Fig. 3g**).

Definition of high confidence human-specific gains and losses by comparison with genome-wide background enrichment. To obtain high-confidence species-specific

enriched regions for both human and rhesus macaque, read counts within CREs were compared to the genome-wide enrichment for each sample in the other species. To define the background enrichment for each ChIP-seq sample, read counts were calculated in sliding (500-bp steps) windows of 3,000 bp (mean enhancer size) across the whole genome. Read count values for CREs were normalized to a 3 kb size and were considered background signal when they fell below the 80th percentile of the set of 3,000-bp background windows. In total, 1,399 human enhancers and 89 promoters were below 80th percentile threshold in all rhesus macaque samples. Of those, 139 and 7 were also below the background threshold in chimpanzee. 1,243 rhesus macaque CREs (24 promoters and 1,219 enhancers) were enriched below the 80th percentile of background enrichment in human. Of those, 45 enhancers and 4 promoters were defined human-specific losses as they were significantly enriched (as defined by peak calling) in chimpanzee.

Gene ontology analysis. Gene ontology (GO) analysis was done using Genomic Regions Enrichment of Annotations Tool (GREAT; http://bejerano.stanford.edu/great/public/html)[63] with the basal plus extension setting. Therefore, multiple genes are potentially assigned to enhancer regions. For the analyses of brain-specific enhancers (**Supplementary Fig. 3f**), we specifically analyzed regions that were not found enriched in a set of 13 tissues analyzed from encode data sets described above[24].

Analysis of accelerated regions. To link HARs with putative brain enhancers, predicted accelerated regions from three different sources[44–46] were merged totaling 2,595 predicted HARs on hg38. These were compared with the 59,786 human enriched regions. 284 human CREs covered a HAR (**Supplementary Table 21**). Out of those, 240 (192 enhancers and 48 promoters) were mappable onto panTro4 and rheMac3 with similar sequence content and analyzed for differential enrichment in human versus rhesus macaque and chimpanzee by overlap with the DE regions from all eight brain regions (**Supplementary Tables 12–15**).

Random sampling of genomic DNA. Significance of overlap between enriched regions throughout the manuscript was analyzed in comparison to 1,000 sets of random genomic DNA of comparable size (**Figs. 1b**, **2a**,**b** and **6a**, and **Supplementary Fig. 2**). To generate a H3K27ac-mappable reference genome, hg38, rheMac3 and panTro4 were divided in sliding (500-bp steps) windows of 3,000 bp and the following windows were excluded: windows with zero mappable reads using all H3K27ac reads of the indicated species, windows with more than 10% unknown bases (UCSC Table Browser's Gap Locations lists) and windows overlapping regions mapping to multiple positions (see above).

4C sequencing. The 4C experiment on the *CADM1* enhancer was performed as described previously[64,65], on human cerebellar tissue (HS1) using DpnII and NlaIII as restriction enzymes. For amplification, the following primers were used: 5′-GTTCTGGTGTCTGAGAACCA-3′ (reading primer) and 5′-TGTAACCAGACCCATTCTTC-3′ (non-reading primer). Sequencing was done on the Illumina Hiseq 2000 genome sequencer. Sequence reads were mapped onto hg19. Reads mapping to multiple fragment ends were removed, and 4C coverage was computed by averaging mapped reads in running windows of 41 fragment ends. With 71% of the total reads on the *cis* chromosome, and 77.88% of those within 200-kb of the viewpoint, quality scores met previously described conditions[65].

51. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
52. Landt, S.G. *et al.* ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res.* **22**, 1813–1831 (2012).
53. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
54. Auerbach, R.K. *et al.* Mapping accessible chromatin regions using Sono-Seq. *Proc. Natl. Acad. Sci. USA* **106**, 14926–14931 (2009).
55. Bonn, S. *et al.* Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. *Nat. Genet.* **44**, 148–156 (2012).
56. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
57. Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
58. R Development Core Team. *R: a Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria, 2012).
59. van der Maarten, L.J.P. & Hinton, G.E. Visualizing high-dimensional data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
60. Saldanha, A.J. Java Treeview--extensible visualization of microarray data. *Bioinformatics* **20**, 3246–3248 (2004).
61. Akhtar-Zaidi, B. *et al.* Epigenomic enhancer profiling defines a signature of colon cancer. *Science* **336**, 736–739 (2012).
62. Bernstein, B.E. *et al.* The NIH Roadmap Epigenomics Mapping Consortium. *Nat. Biotechnol.* **28**, 1045–1048 (2010).
63. McLean, C.Y. *et al.* GREAT improves functional interpretation of *cis*-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010).
64. Splinter, E., de Wit, E., van de Werken, H.J., Klous, P. & de Laat, W. Determining long-range chromatin interactions for selected genomic sites using 4C-seq technology: from fixation to computation. *Methods* **58**, 221–230 (2012).
65. van de Werken, H.J. *et al.* 4C technology: protocols and data analysis. *Methods Enzymol.* **513**, 89–112 (2012).