Title
Enhancer Divergence and cis-Regulatory Evolution in the Human and Chimp Neural Crest

Authors
Prescott, SL
Srinivasan, R
Marchetto, MC
et al.

Cell

Enhancer Divergence and *cis*-Regulatory Evolution in the Human and Chimp Neural Crest

Graphical Abstract
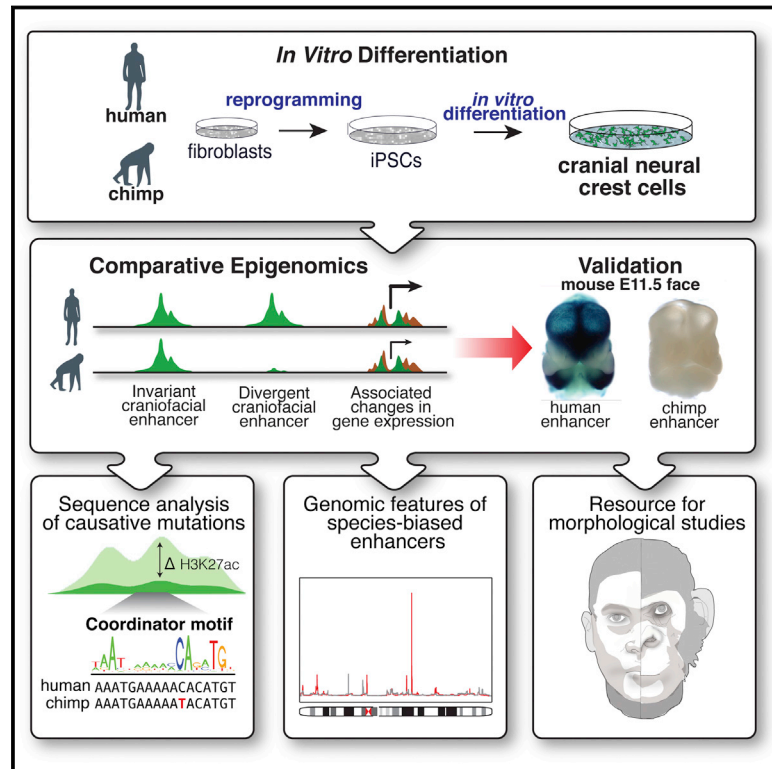
Authors

Sara L. Prescott, Rajini Srinivasan, Maria Carolina Marchetto, ..., Fred H. Gage, Tomek Swigut, Joanna Wysocka

Correspondence

swigut@stanford.edu (T.S.), wysocka@stanford.edu (J.W.)

In Brief

Epigenomic and transcriptome profiling from in-vitro-derived human and chimpanzee cranial neural crest cells (CNCCs) allows for the identification and mechanistic exploration of recent *cis*-regulatory landscape divergence underlying human craniofacial evolution.

Highlights

- Aspects of higher primate embryogenesis can be recapitulated in vitro using iPSCs

- Epigenomic profiling from human and chimp CNCCs reveals divergent facial enhancers

- Recently diverging CNCC enhancers have distinct sequence features

- Species-biased enhancers cluster near loci affecting intra-human facial variation

CrossMark

CellPress

Article

Enhancer Divergence and *cis*-Regulatory Evolution in the Human and Chimp Neural Crest

Sara L. Prescott,[1] Rajini Srinivasan,[1] Maria Carolina Marchetto,[2] Irina Grishina,[3] Iñigo Narvaiza,[2] Licia Selleri,[3] Fred H. Gage,[2,4] Tomek Swigut,[1,*] and Joanna Wysocka[1,5,6,*]

[1]Department of Chemical and Systems Biology and Department of Developmental Biology, Stanford University School of Medicine, Stanford, CA 94305, USA
[2]Laboratory of Genetics, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, CA 92037, USA
[3]Department of Cell and Developmental Biology, Weill Cornell Medical College, Cornell University, New York, NY 10065, USA
[4]Center for Academic Research and Training in Anthropogeny (CARTA), University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA
[5]Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, CA 94305, USA
[6]Institute of Stem Cell Biology and Regenerative Medicine, Stanford University School of Medicine, Stanford, CA 94305, USA
*Correspondence: swigut@stanford.edu (T.S.), wysocka@stanford.edu (J.W.)
http://dx.doi.org/10.1016/j.cell.2015.08.036

SUMMARY

cis-regulatory changes play a central role in morphological divergence, yet the regulatory principles underlying emergence of human traits remain poorly understood. Here, we use epigenomic profiling from human and chimpanzee cranial neural crest cells to systematically and quantitatively annotate divergence of craniofacial *cis*-regulatory landscapes. Epigenomic divergence is often attributable to genetic variation within TF motifs at orthologous enhancers, with a novel motif being most predictive of activity biases. We explore properties of this *cis*-regulatory change, revealing the role of particular retroelements, uncovering broad clusters of species-biased enhancers near genes associated with human facial variation, and demonstrating that *cis*-regulatory divergence is linked to quantitative expression differences of crucial neural crest regulators. Our work provides a wealth of candidates for future evolutionary studies and demonstrates the value of "cellular anthropology," a strategy of using in-vitro-derived embryonic cell types to elucidate both fundamental and evolving mechanisms underlying morphological variation in higher primates.

INTRODUCTION

Since the discovery that the protein-coding regions of the genome remain largely conserved between humans and chimpanzees, it has long been postulated that morphological divergence between closely related species is driven principally through quantitative and spatiotemporal changes in gene expression, mediated by alterations in *cis*-regulatory elements (Carroll, 2008; King and Wilson, 1975; Wray, 2007). A number of excellent case studies have validated these early predictions and demonstrated that mutations or deletions affecting distal regulatory elements called enhancers can alter ecologically relevant traits (Gompel et al., 2005; Shapiro et al., 2004; Attanasio et al., 2013). Recent successes in full-genome sequencing and epigenomic strategies have enabled the first genome-wide comparisons of transcription factor (TF) binding and regulatory landscapes in closely related species, demonstrating the value of comparative epigenomics in the context of high-genome orthology for understanding principles of *cis*-regulatory evolution (Bradley et al., 2010; He et al., 2011; Stefflova et al., 2013). Nonetheless, despite the availability of human and chimpanzee genomes, our knowledge of *cis*-regulatory divergence between humans and our closest evolutionary relatives remains fairly speculative. Previous efforts have relied heavily on computational approaches to pinpoint conserved non-coding elements that were either deleted or had undergone accelerated change specifically in the human lineage (McLean et al., 2011; Pollard et al., 2006; Prabhakar et al., 2006). Functional epigenomic comparisons between humans and other primates have been largely limited to lymphoblastoid cell lines (Cain et al., 2011; Shibata et al., 2012; Zhou et al., 2014) or to profiling whole organs from more distantly related species (Cotney et al., 2013; Villar et al., 2015).

Recently, iPSCs were made available from our nearest living evolutionary relative, the chimpanzee (Marchetto et al., 2013), offering an opportunity to derive developmentally relevant and previously inaccessible tissue types in vitro. This allows aspects of species-specific development to be recapitulated in a dish, facilitating "cellular anthropology" through the discovery of cell-type-specific regulatory changes that occurred during recent human evolution. Here, we focus on the neural crest (NC), one of the embryonic cell populations most relevant to emergence of uniquely human traits. In vivo, NC cells (NCCs) arise during weeks ~3–5 of human gestation from the dorsal part of the neural tube ectoderm and migrate into the branchial arches and what will later become the embryonic face, consequently establishing the central plan of facial morphology (Bronner and LeDouarin, 2012; Cordero et al., 2011; Jheon and Schneider, 2009). Within our recent evolutionary history, the modern human craniofacial complex has undergone dramatic changes in shape and sensory organ function, which helped to build a recognizably

human face and were required to accommodate the transition to bipedal posture, enlargement of the brain, extension of the larynx for speech, and compensatory rotations of the orbits, olfactory bulb, and nasomaxillary complex (Bilsborough and Wood, 1988; Lieberman, 1998; Spoor et al., 1994).

To overcome the inability to obtain cranial NCCs (CNCCs) directly from higher primate embryos, we here employ a pluripotent stem-cell-based in vitro differentiation model in which specification, migration, and maturation of human and chimpanzee CNCCs are recapitulated in the dish (Bajpai et al., 2010; Rada-Iglesias et al., 2012; this study). We compared TF and coactivator binding, histone modifications, and chromatin accessibility genome-wide to annotate the divergent regulatory element repertoire of human and chimpanzee CNCCs. This information allowed us to explore, with unprecedented comprehensiveness and resolution, the mechanisms of tissue-specific enhancer landscape evolution within a developmentally relevant tissue type in humans and our nearest evolutionary relative.

RESULTS

Derivation of Human and Chimpanzee CNCCs

Given the similarities in hominid gestational environment, we hypothesized that non-human primate CNCCs could be derived from pluripotent cells using the same cell culture conditions that we have previously applied to human embryonic stem cells (ESCs)/iPSCs (Bajpai et al., 2010; Rada-Iglesias et al., 2012). Chimp iPSCs have recently become available and can be maintained in vitro under identical conditions as human ESCs/iPSCs (Marchetto et al., 2013). Upon differentiation of our chimp iPSCs, we observed formation of highly mobile stellate cells that were morphologically indistinguishable from human CNCCs, expressed a broad range of migratory NC markers at levels equivalent to those seen in human cells, and had a very low level of *HOX* gene expression, a profile consistent with CNCC identity (Figures 1A-1C and S1A). To characterize staging and homogeneity of our human and chimp CNCC populations, we identified a panel of five cluster of differentiation (CD) markers, whose expression is sensitive to the developmental progression of CNCC (see Experimental Procedures and Figure S1B). These markers provided a platform for us to monitor and optimize our cell culture protocol for derivation and maintenance of primate CNCCs achieving metrics of homogeneity greater than 90% regardless of the genetic background, initial cell source (e.g., iPSC versus ESC), or species (human versus chimp); see Figure S1C and Experimental Procedures. Cultured primate CNCCs show a high correlation of expression signatures and epigenomic profiles with CNCCs isolated from chick embryos, reinforcing the NC identity of these in-vitro-derived cells (Figures S2A and S2B). Importantly, derived human and chimp CNCCs are both capable of prolonged maintenance (for up to 18 passages) and sustained differentiation capacity into both mesenchymal and non-mesenchymal lineages (Figure S2C). Furthermore, xenotransplantation of cultured human and chimp CNCCs into the dorsal neural tube of early chick embryos demonstrates their ability to engraft and then follow endogenous migration cues into the distal branchial arches (Figures S2D and S2E).

Epigenomic Profiling of Human and Chimpanzee CNCCs

For epigenomic profiling, we derived CNCCs from H9 hESCs and from iPSCs from two humans and two chimpanzees (Marchetto et al., 2013). We subsequently performed chromatin immunoprecipitation and sequencing (ChIP-seq) using antibodies against CNCC TFs (TFAP2A and NR2F1), a general coactivator (p300), and histone modifications associated with active regulatory elements (H3K4me1, H3K4me3, and H3K27ac) (Figures 1A and 1E). In parallel, we mapped genome-wide chromatin accessibility using an assay for transposase-accessible chromatin (ATAC-seq) (Buenrostro et al., 2013).

One crucial advantage of performing comparative epigenomics between human and chimpanzee, as opposed to a more distant primate relative, is the large similarity between genomes, which permits reciprocal mapping of sequencing reads to the reference genomes of both species. This allows for quantification of read enrichments from each species in the context of both reference genomes, removing otherwise difficult-to-control-for biases due to mappability, ambiguous liftOver, and other technical caveats. Importantly, we could unambiguously assign one-to-one orthology between genomes for >95% of all enhancer candidates from either species, with the remaining 4%–5% representing enhancers that fall within putative species-specific structural variants. We found that enrichments for all ChIP-ed factors and for chromatin accessibility were largely independent of the chosen reference genome and excluded all candidate elements for whom enrichment divergence was dependent upon the reference (< 0.1%) or that did not map uniquely in both genomes (see Experimental Procedures). Globally, the observed epigenomic patterns at candidate regions were highly correlated for human and chimp CNCCs (Figures 1E and Figure S4A).

Genome-wide Annotation of Human and Chimpanzee CNCC Regulatory Elements Uncovers Enhancers with Craniofacial Activity

To annotate enhancers genome wide, we promiscuously identified candidate *cis*-regulatory regions by the presence of TF or p300 enrichment and/or increased chromatin accessibility. We then restricted our analysis primarily to enhancers by assessing the ratio of H3K4me1/H3K4me3 enrichment at these candidate sites, which distinguishes distal enhancers from promoters (Heintzman et al., 2007), and further using H3K27ac enrichment to differentiate active from inactive elements (Creyghton et al., 2010; Rada-Iglesias et al., 2011). The resulting enhancer candidates had enriched conservation signatures compared to surrounding genomic regions and were near genes annotated with craniofacial ontologies—consistent with bona fide NC enhancer status (Figures S3A–S3C). Furthermore, cross-referencing our enhancer list with the VISTA Enhancer Browser database (Visel et al., 2007) identified 247 regions overlapping CNCC enhancers that were functionally tested for activity in mouse embryos. Of those 247 regions: (1) 208 were active at E11.5 (odds ratio 6.33 and $p < 5 \times 10^{-32}$), and (2) these 208 active enhancers were significantly enriched for activity in NC-derived head tissues (branchial arches and facial mesenchyme; Figure 1D, examples are shown in Figures 1E [right], and Figure S3D). Thus, our analysis captures regulatory regions relevant for distinct
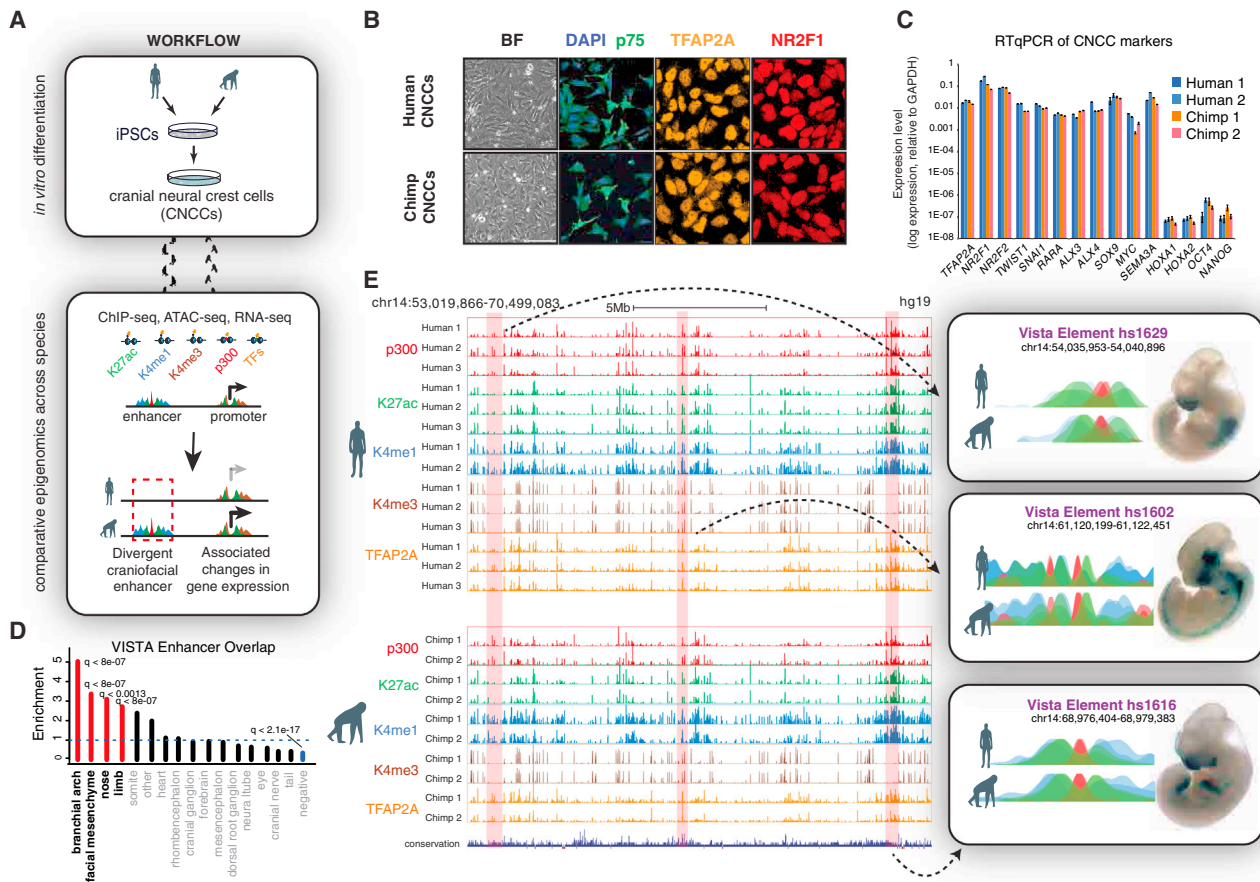
Figure 1. Derivation of Human and Chimpanzee CNCCs and Epigenomic Annotation of Craniofacial Enhancers

(A) Workflow of comparative epigenomic strategy.

(B) Confocal immunofluorescence detection of NC markers p75, TFAP2A, and NR2F1 in human and chimp CNCCs at passage 4.

(C) RT-qPCR of NC markers, HOXs, and pluripotency markers *OCT4* and *NANOG* in derived human and chimp CNCCs from two genetic backgrounds of each species. Error bars represent one SD.

(D) Enrichment of annotated expression domain categories from overlap of top 15,000 enhancer calls with regions in the VISTA enhancer database. p values were calculated with Fisher's exact test and corrected for pFDR. Categories with q value < 0.05 are indicated in red (enrichment) or blue (depletion).

(E) Representative UCSC Genome Browser tracks showing ChIP-seq profiles for p300 (red), H3K27ac (green), H3K4me1 (blue), H3K4me3 (brown), and TFAP2A (orange) from both species aligned to hg19 reference genome. Representative elements tested through the VISTA enhancer database (Visel et al., 2007) displayed on the right next to the reported *lacZ* expression domains.

spatial identities in the developing face in vivo (Figure 1D). Taken together, our epigenomic approach thus comprehensively annotates putative human and chimp NC enhancers, at least a subset of which is active in facial structures during embryogenesis.

Quantitative Analysis of H3K27ac Enrichments Predicts Species-Biased Enhancers

We hypothesized that, in closely related species, quantitative modulation of activity at orthologous regions is a major form of enhancer divergence. To identify such divergence, we used H3K27ac enrichment data in biological quadruplicate (i.e., independent CNCC derivations from each individual) to quantitatively approximate activity at all annotated CNCC enhancers detected for either species. Global comparisons of H3K27ac enrichments between individuals of the same species revealed high concordance of signals, with some minor variation due to either differ-

ences in genetic background or experimental variability (Figures 2A, highlighted in red, and S4A). Human and chimpanzee CNCC H3K27ac enrichment was also highly correlated when mapped to the same reference genome, and human and chimpanzee CNCC H3K27ac profiles clustered together distinctly from 49 other human cell types (Figures S4A and S4B). Despite this high conservation of profiles, a substantial subset of elements demonstrated a significant species bias (Figure 2A, FDR < 0.01 highlighted in blue), which we thereafter considered to be our species-biased enhancer candidates. H3K27ac ChIP-qPCR at select candidate enhancers from independent CNCC derivations recapitulated this species bias (Figure S4C).

Importantly, consistent with the premise that H3K27ac is a suitable readout of enhancer activity, the bias in H3K27ac status alone was highly predictive of biases in TF and p300 binding, as well as chromatin accessibility (Figure 2C; examples are shown
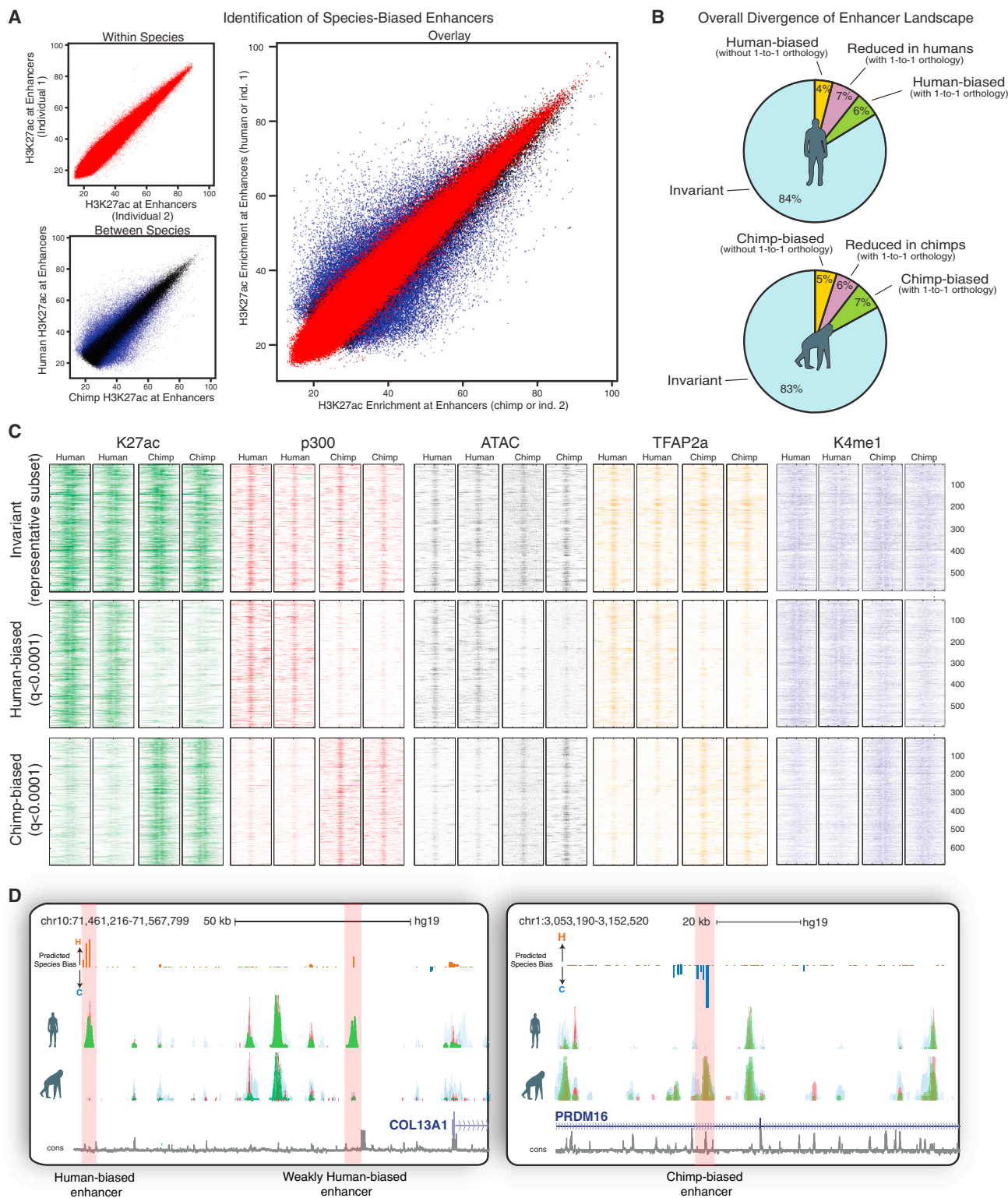
Figure 2. Identification of Species-Biased Enhancers Using H3K27ac Enrichments at Orthologous Loci

(A) Enrichment of H3K27ac at candidate enhancer elements compared within individuals of the same species (red) or across species (blue/black), with overlay shown on the right. Enhancers with significant inter-species divergence indicated in blue ($p_{adj}<0.01$).

(legend continued on next page)

in Figures 2D and S4D). Furthermore, this approach enabled genome-wide assignment of signed significance scores on a per-enhancer basis, visualizable as a genome browser track (Figure 2D, "Predicted Species Bias" track).

Altogether, of all annotated active human CNCC enhancers (n = 14,606), 84% were invariant, 4% fell at non-orthologous sites, and 6% and 7% demonstrated quantitative increase or decrease, respectively (Figure 2B). One limitation is the low number of currently available chimpanzee iPSC lines, especially given the high reported degree of polymorphism among chimps (Kaessmann et al., 1999). To estimate false positive rate for identifying true fixed inter-species differences, we applied our strategy to previously published ChIP-seqs from chimp lymphoblastoid cell lines and estimated a conservative FDR of 0.15 when using only two chimp genetic backgrounds. This suggests that the vast majority of identified differences represent functionally fixed differences across species (the rest represent enhancers that are still divergent but remain polymorphic within one of the species). Our observations agree with the emerging notion that quantitative modulation of enhancer activity is the prevalent source of regulatory landscape divergence among closely related species.

cis-Sequence Changes Drive Species-Biased Enhancer Activity In Vitro and In Vivo

To functionally validate our predictions, we used a luciferase reporter assay to examine activity of a selected set of orthologous pairs of species-biased human and chimpanzee enhancers. We found that >80% of tested enhancers had correlated species bias in luciferase expression, which was consistent regardless of whether the reporter assays were performed in human or chimpanzee CNCCs (Figures 3A and 3B). These results further validate that H3K27ac identifies both enhancer activity and bona fide species bias; thus, for simplicity, we refer to H3K27ac enrichment interchangeably with "activity." Importantly, these results also demonstrate that enhancer divergence can be largely explained by cis-sequence changes rather than differences in the trans regulatory environments of the human and chimp CNCCs.

The conservation of trans-environments across species facilitates testing of human and chimp regulatory elements in vivo using a mouse LacZ transgenic reporter assay. We selected two predicted human-biased enhancers near CNTNAP2 (enhancer 1) and PAPPA (enhancer 2), respectively (Figures 3C and 3D). For both predicted human-biased enhancers we observed gains of additional expression domains in head regions, as well as quantitative gains in enhancer strength, as evidenced by the overall higher LacZ staining intensity for the human sequence compared to the chimp ortholog (Figures 3C–3H and S5). Notably, to ensure that the negative/weak staining results obtained with the chimp sequences were not a result of undersampling, we performed surplus embryo injections with both chimp enhancer reporters (Figure S5A). Thus, species-biased enhancers identified in our in vitro analysis drive distinct expression patterns within CNCC-derived tissues in vivo.

Human Accelerated Regions Overlap with Distal CNCC Enhancers

Our results suggest that DNA sequence is the predominant driver of enhancer divergence; therefore, we began examining sequence properties of species-biased enhancers. Although species-biased enhancers were similar in H3K27ac enrichment levels when compared to invariant enhancers, they showed a distinct reduction of sequence conservation signatures (Figure 4A). Furthermore, we identified 163 "human accelerated regions" (HARs; Hubisz and Pollard, 2014) overlapping active chromatin features in CNCCs, of which 20 showed species-biased activity (at a cutoff of q < 0.001; n = 48 with a cutoff of q < 0.1) (Figures 4B and S6A–S6D), representing a significant enrichment relative to the whole enhancer set (p < 0.025, odds ratio 1.81). It is possible that the HAR-overlapping regions without species bias in CNCC could manifest divergence in another tissue type, as exemplified by HAR2 (a.k.a., HACNS1), which overlaps an invariant CNCC enhancer (Figure S6D, p value of species bias = 0.339) that has a pharyngeal arch activity domain that is conserved in primates but has human-specific activity in the embryonic limb (Prabhakar et al., 2008).

Species-Biased Enhancers Are Enriched for Specific Classes of Retroelements

Given that nearly half of the human genome is composed of transposable elements, the majority of which invaded the primate lineage prior to the separation of humans and chimpanzees (Cordaux and Batzer, 2009), we hypothesized that a subset of species-biased orthologous enhancers may be transposon derived. Interestingly, we found that, while CNCC enhancers overlapped with many different classes of repeats, specific subclasses of endogenous retroviruses (ERV1, ERVL-MaLR, and ERVK) as well as L1 elements were preferentially enriched at species-biased enhancers (Figure 4C), suggesting that these specific subclasses may harbor progenitor sequences that are prone to acquire CNCC enhancer activity over relatively short evolutionary distances.

(B) Pie charts showing the percentage of total active CNCC enhancers classified as either species-biased enhancers with gained activity (green), species-biased enhancers with decreased activity (purple), enhancers without clear orthology across genomes (yellow), or invariant enhancers (blue) using a human reference genome (above) or chimp reference genome (below).
(C) Heatmap of raw ChIP-seq and ATAC-seq counts across species-biased and invariant CNCC enhancers for two human and two chimp genetic backgrounds. Each row represents a 2 kb window (1 kb each direction) centered around the middle of human-biased (n = 598, q < 0.0001), chimp-biased (n = 691, q < 0.0001), or invariant (n = 584 representative subset, q > 0.95) enhancers for H3K27ac (green), p300 (red), TFAP2A (yellow), K4me1 (blue), and ATAC-seq (gray). All reads were aligned to hg19.
(D) Representative browser tracks showing overlaid H3K4me1 (blue), p300 (red), and H3K27ac (green) from human and chimp CNCCs mapped to hg19. Examples of strongly human-biased, weakly human-biased, or strongly chimp-biased enhancers highlighted in pink. Predicted species-bias track shown above for candidate enhancers; the magnitude of the bias track represents −log10 (adjusted p value of divergence) with negative sign (indigo) representing chimp bias and positive (bronze) human bias.
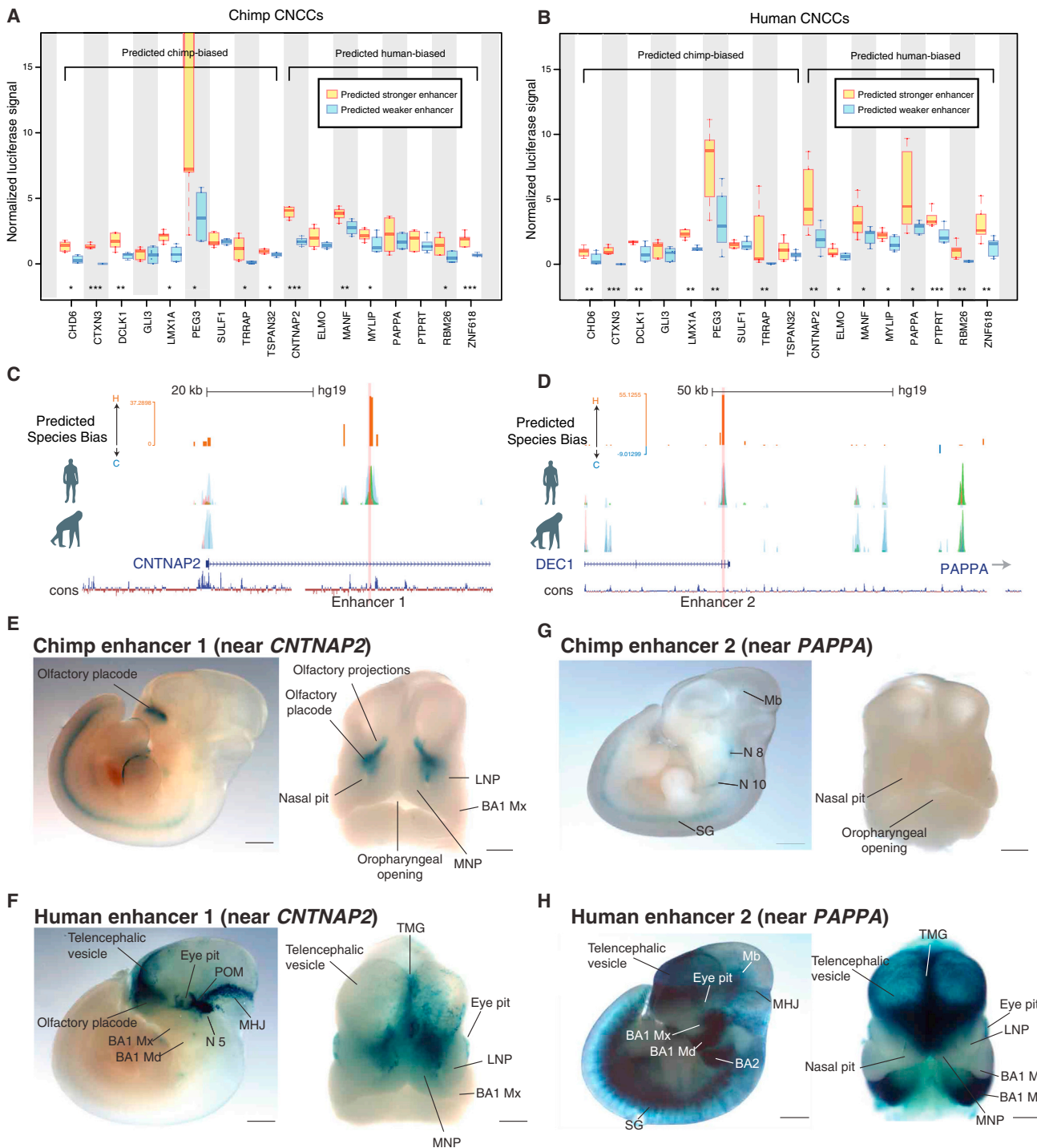
Figure 3. In Vitro and In Vivo Validations of Species-Biased Enhancers

(A and B) Luciferase reporter assays performed in chimp CNCCs (A) or human CNCCs (B) for 9 chimp-biased regions (and orthologous human regions) and 8 human-biased regions (and orthologous chimp regions). Luciferase signal was normalized to renilla transfection control. Significance tested from three biological replicates from each species with ANOVA followed by residuals testing with Student's t test. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. Central bar represents the median, box outline represents first and third quartile, and whiskers extend to furthest datapoint within 1.5× box length way from the box. Tested enhancers are named by nearest gene.

(C and D) Genome browser tracks showing human-biased enhancer 1 (near *CNTNAP2* gene; C) and enhancer 2 (near *PAPPA* gene; D) selected for a *lacZ* reporter mouse transgenesis assay.

(E and F) Analysis of enhancer activity for chimpanzee and human enhancer 1 in a *lacZ* reporter transgenic mouse assay. (E) Representative E11.5 transgenic embryo obtained for the chimpanzee enhancer 1 reporter, shown in lateral view (left) or frontal view (right) of the embryonic head. (F) Representative E11.5

(legend continued on next page)

Sequence Substitutions within TF Binding Motifs at Species-Biased Enhancers Contribute to Epigenomic Divergence

Consistent with the expectation that species-specific biases are largely sequence driven, we observed that the variance in H3K27ac between species at each enhancer scales proportionally with the degree of sequence dissimilarity (i.e., Levenshtein distance) at those orthologous sites, while the intra-species variance at the same regions remains unchanged (Figure 4D). Nonetheless, even at enhancers with detectable species bias, sequence substitutions were still infrequent—only ~3–6 substitutions per 500 bp enhancer—suggesting that a small number of mutations can confer substantial effects on overall enhancer activity, likely by affecting binding of key sequence-dependent TFs. We therefore interrogated how frequently sequence substitutions fall within particular classes of TF motifs and to what degree these mutations correlate, either positively or negatively, with changes in enhancer activity or other chromatin modifications (Figure 4E). This, in essence, leverages preexisting genetic variation like a large-scale mutagenesis screen.

Through this approach, we identified a large set of both known and novel motifs for which deviation from the consensus was correlated with species bias of H3K27ac and other epigenomic marks, implying functional consequences for these mutations. As expected, the correlations vary in frequency and in effect, with some motifs being frequent and having small effects (e.g., Forkhead factors) and others being infrequent but conferring large effects (e.g., TFAP2A), with one outlier motif being both very frequent and conferring large effects when mutated (see description of the "Coordinator" motif below) (Figure 4F). Among our top hits, we identified many motifs for TFs with known effects in NC regulation, including a set of TFAP2 motif variants that serve as a positive control for our approach, as we see a high correlation between TFAP2 motif mutations and inter-species divergence in TFAP2A ChIP signals at these sites (Figure 4G, group 3). We previously showed that TFAP2A participates in establishment of active chromatin states at NC enhancers (Rada-Iglesias et al., 2012), and consistently we observed that divergence from the TFAP2A consensus also correlates with the loss of H3K27ac, co-activator binding, and chromatin accessibility. Notably, TFAP2 motifs are depleted from species-biased sites, likely due to strong selective pressure to conserve TFAP2A function in the NC and possibly in other pleiotropic contexts (Figure 4F). Another interesting set of motifs, which are both frequent at species-biased sites and positively correlated with permissive chromatin states, are those recognized by ALX homeobox factors that are highly expressed in the face and mutated in severe frontonasal dysplasias in humans (Twigg et al., 2009) (Figures 4F and 4G, group 2).

Intriguingly, we also identified a group of motifs whose mutations away from the consensus were correlated with a gain in chromatin accessibility and H3K27ac, suggesting that these motifs may recruit repressive factors with negative effects on overall enhancer activity. Examples of such motifs included the SNAI2 motif, which is bound by a known transcriptional repressor, the TBX family motif bound by T-box factors, and other candidate negative regulators representing distinct TF classes, e.g., HIC1/2, MESP1, TCF3/4, and GLIS1 (Figure 4G, group 1). These results suggest an unappreciated prevalence of repressive inputs in quantitative modulation of enhancer activity.
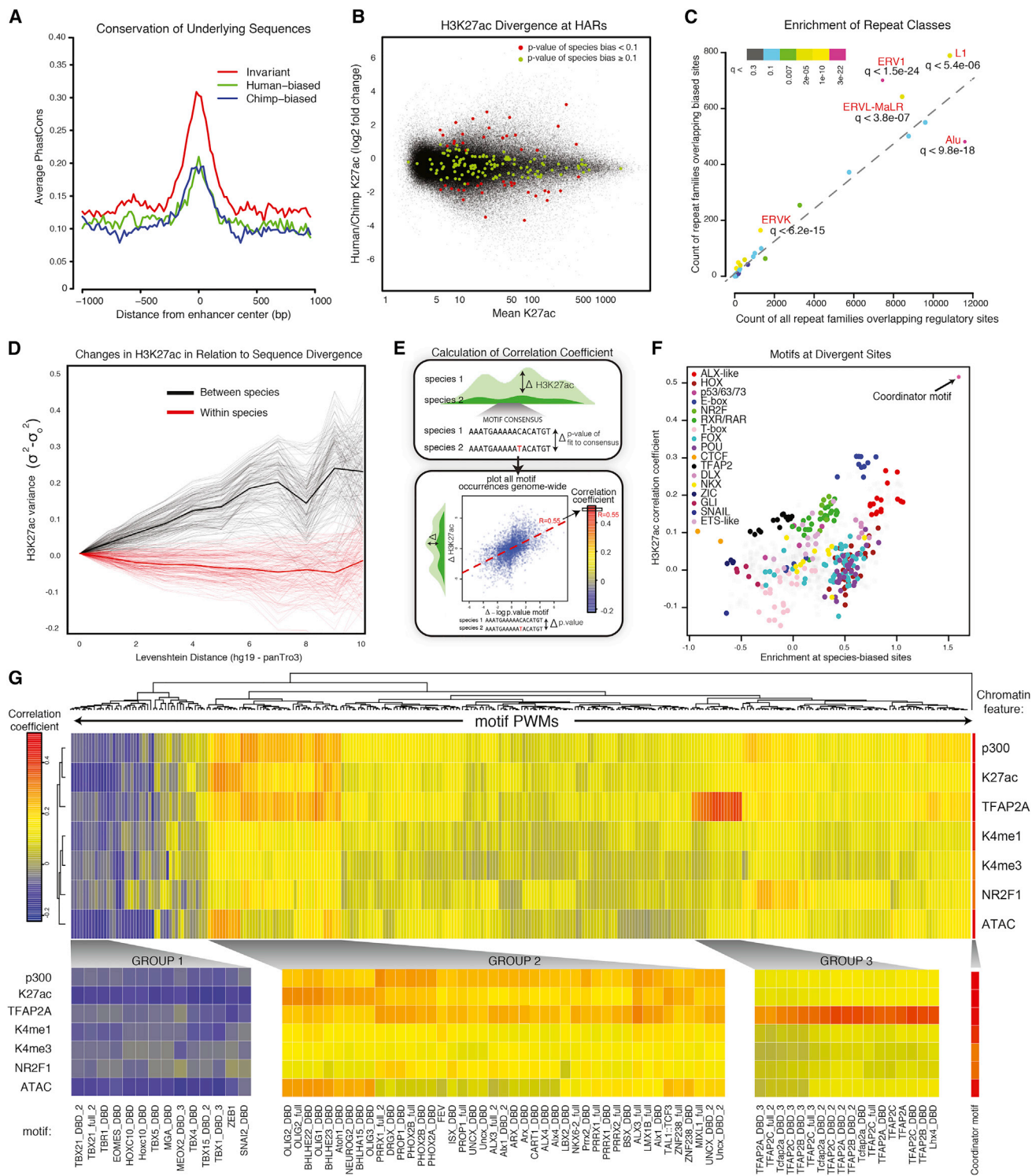
"Coordinator": A novel Motif that Is Highly Predictive of Active Chromatin States and Species Bias

Surprisingly, one motif stood out as an outlier in this analysis, as it was exceptionally enriched at divergent sites and was the most correlated with changes in all examined active chromatin features (Figures 4F, upper-right, and 4G, far-right). This sequence, which we termed the "Coordinator" motif, is a 17-bp-long motif, which we identified through de novo motif discovery from our CNCC-specific enhancers and was not previously annotated to a known regulatory complex. We note that portions of the Coordinator resemble an E box and HOX-like motifs; however, these represent large protein families, and the particular factors that bind at this element remain to be identified.

Sequence analysis using INSIGHT, a tool to infer signatures of recent natural selection using human polymorphism data (Gronau et al., 2013), found evidence of positive selection at the Coordinator motif occurrences within species-biased enhancers, but not within invariant enhancers, suggesting that the motif and its cognate binder(s) have played a privileged role in recent enhancer divergence in primate CNCCs (Figure 5A). When we further dissected the motif by individual bases, we found that the correlations of each nucleotide with ChIP enrichments (both for histone modifications and TF ChIPs) recapitulated the information content of the motif itself, as would be expected if Coordinator motif mutations were causal for the observed chromatin changes (Figure 5B). Fittingly, we found human mutations that strengthen the Coordinator motif within both human-biased enhancers tested in mouse transgenesis (Figure S6E). Globally, the Coordinator motif was preferentially enriched at distal regulatory elements rather than at promoters (Figure S6F) and was further enriched at enhancers that were CNCC specific as opposed to those that shared measurable H3K27ac in other tissue types (Figure 5C). Interestingly, we observe that LTR9 elements, a retroelement class enriched at species-biased enhancers, are 5× more likely to harbor a Coordinator motif variant than MER52A elements, a similar repeat

Figure 4. Global Features of Species-Biased Enhancers and Correlation of Mutations within TF Binding Motifs with Epigenomic Divergence

(A) Average PhastCons scores are shown for strong invariant enhancers (q > 0.98), strongly human-biased enhancers (q < 0.0001), or strongly chimp-biased enhancers (q < 0.0001) for 1 kb surrounding each enhancer center.

(B) Degree of species bias (log$_2$ fold change H3K27ac human/chimp, y axis) relative to enhancer strength (human-chimp-averaged H3K27ac enrichment, x axis) for bulk CNCC elements (black) and elements overlapping HARs (color representing q value of species bias: q < 0.1 in red, q $\geq$ 0.1 in green).

(C) Counts of repeat families overlapping species-biased enhancers (y axis) relative to counts of repeat families overlapping all CNCC regulatory sites (x axis) are plotted. q values of enrichment for different repeat classes is indicated by color.
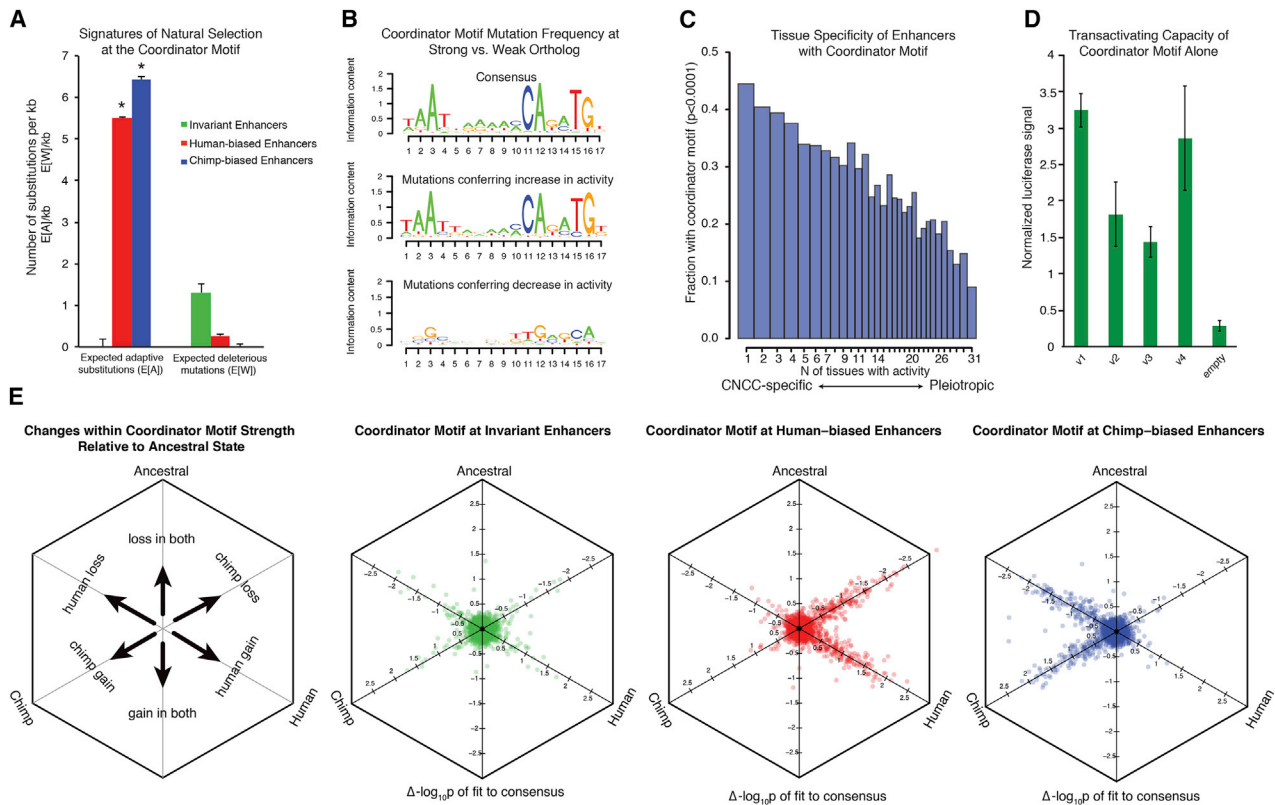
(legend continued on next page)

Figure 5. Properties of the Novel "Coordinator" Motif

(A) Expected number of adaptive substitutions (E[A]) per kilobase and expected number of deleterious mutations E[W] per kilobase were calculated for all sites of the Coordinator motif at invariant enhancers (green), at human-biased enhancers (red), and at chimp-biased enhancers (blue) using default INSIGHT parameters (Gronau et al., 2013). Significance indicated by * (p < 0.01). Overall fractions of nucleotides under selection (ρ) not shown (ρ$_{invariant}$ = 0.66, p < 0.01; ρ$_{human-biased}$ = 0.015, p < 0.01; ρ$_{chimp-biased}$ = 0.019, p < 0.01). Error bars represent approximate SE.

(B) Position weight matrix of the Coordinator consensus sequence from top 3,000 CNCC specific enhancers is shown (top) relative to logo of mutations preferred at more acetylated (H3K27ac) alleles (middle) versus mutations preferred at less acetylated alleles (bottom).

(C) Enhancers were scored for H3K27ac ChIP-seq enrichments from 30 public data set cell types and binned by number of tissues with activity (1 to 31). The fraction of enhancers per bin with recognizable Coordinator motif (p < 0.0001) is indicated on y axis.

(D) Four different versions (V1–V4) of the Coordinator motif were cloned in tandem into luciferase reporter vectors and were tested for transactivation activity in human CNCCs. Luciferase was normalized relative to renilla transfection control. Error bars represent one SD.

(E) Comparison of sequence changes within the Coordinator motif with a reconstructed human-chimp ancestral outgroup. Changes in fit to the Coordinator consensus compared to the ancestral ortholog (−log$_{10}$ p value) were plotted as orthographic projections along space diagonals for all occurrences of the motif for both human and chimpanzee lineages at different classes of sites. Overlapping data points were jittered for better visualization. Schematic is shown on the far left.

class depleted from species-biased sites. Even at sites without activity in CNCCs, LTR9 sequences are 3.7× more likely to harbor a Coordinator-like motif than MER52A, consistent with the idea that a preexisting Coordinator-like progenitor sequence contributed to the recent adaptation of some retroelements for CNCC enhancer function. Lastly, we found that the Coordinator motif alone was able to drive activity in luciferase reporter assays in CNCCs (Figure 5D).

(D) Pairwise H3K27ac variance σ2-σ2$_{ld = 0}$ at enhancers across samples, ranked by increasing sequence dissimilarity counted by Levenshtein distance (ld) between human (hg19) and chimp (panTro3) orthologous 200 bp enhancers, relative to ld = 0. Comparison between samples of different species shown in black; same species shown in red (means represented by thick lines).

(E) Schematic showing method for deriving the correlation coefficient. For a given motif, each occurrence genome wide containing a genetic change across species is plotted as Δ-log10 p value (human/chimp) of the fit to consensus (x axis) versus ΔH3K27ac for the overlying enhancer region (human/chimp) (y axis), then a line is fit. The slope of the line represents the correlation coefficient for that given motif and epigenomic modification genome wide.

(F) Enrichments of classes of motifs at species-biased enhancers over all enhancers (log odds ratio, x axis) plotted relative to genome-wide correlation coefficient calculated for each motif (using H3K27ac), as described in E (y axis).

(G) Genome-wide correlation coefficients were calculated for whole databases of annotated motifs and multiple chromatin features, revealing motifs with large influence on epigenomic profiles. Correlation coefficients are bi-clustered per motif, and resulting changes in enrichment of chromatin features (p300, K27ac, TFAP2A, H3K4me1, H3K4me3, NR2F1, ATAC) at all enhancers containing mutated PWMs are represented by color. Individual subclusters are magnified below with corresponding motifs indicated.
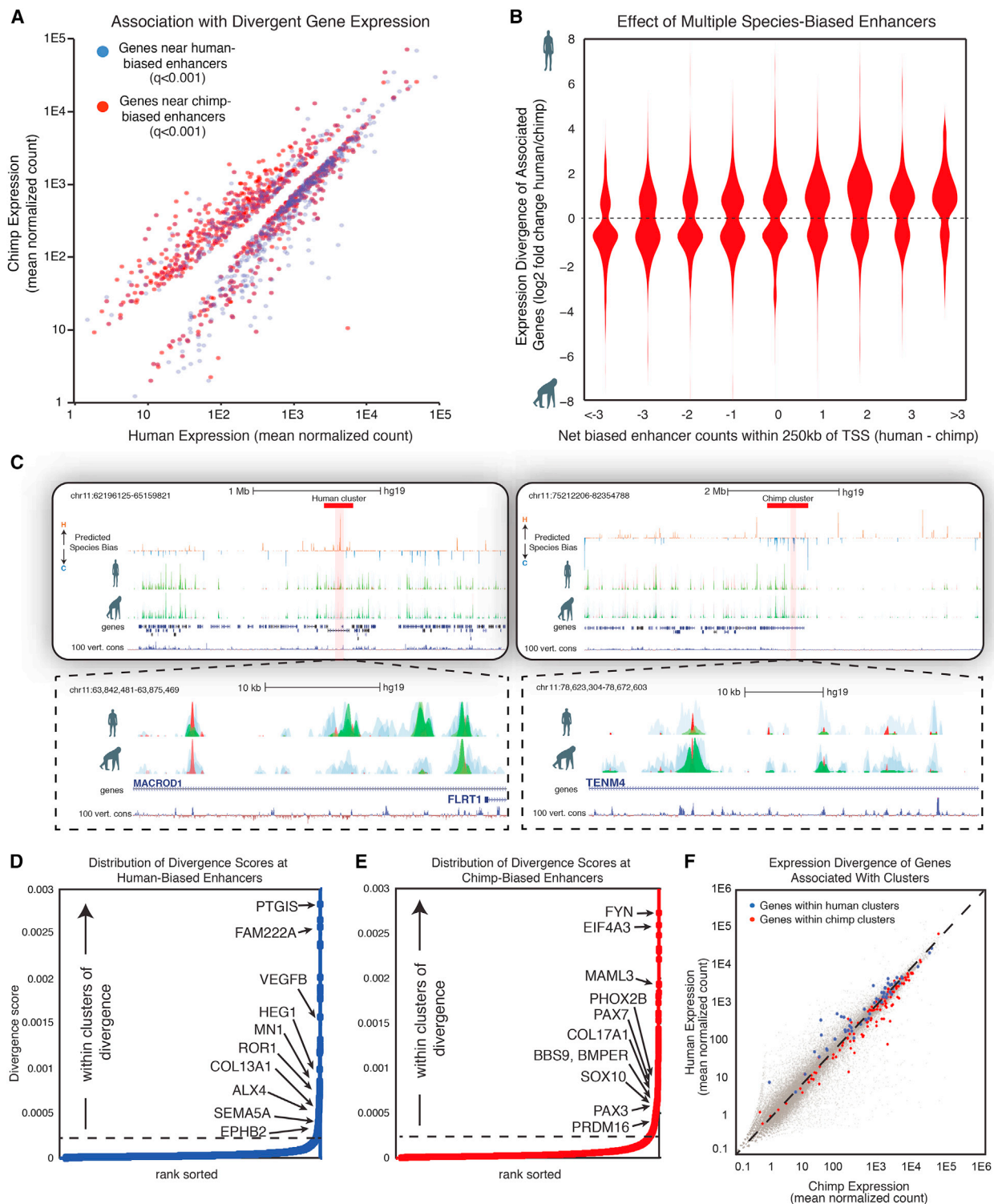
Figure 6. Clusters of Regulatory Divergence Overlap Loci with Crucial Roles in Trait Variation and Are Predictive of Expression Bias

(A) Mean normalized human expression (x axis) versus mean normalized chimp expression (y axis) for genes associated with human-biased enhancers (q < 0.001, blue) or with chimp-biased enhancers (q < 0.001, red). Only genes with significant inter-species expression change (p_{adj} value < 0.1) are shown.

(B) Violin plots showing $\log_2$ fold change human/chimp H3K27ac enrichment at orthologous enhancers binned by total count of biased enhancers (total number of human-biased enhancers minus total number of chimp-biased enhancers) within 250 kb of promoter regions for genes with significant differences in expression across species (p_{adj} value < 0.1).

Sequence Analysis Reveals the Recent Evolutionary History of Coordinator Motif Changes

Our results suggest that nucleotide changes within Coordinator motif sites represent an important class of "causative" mutations predictably associated with gain or loss of CNCC enhancer activity. Thus, by comparing the fit to the consensus for Coordinator-like motifs with a reconstructed ancestral outgroup, we can infer the polarity of enhancer activity change in each lineage relative to the common human-chimp ancestor. Using this strategy, we observed that human-biased enhancers contain Coordinator-like sequences that were equally prone to: (1) a gain in the fit in the human lineage (n = 300) or (2) a loss in the fit in the chimp lineage (n = 255) relative to the ancestral state (Figure 5E). However, human-biased enhancers contain almost no examples in which there was a gain of Coordinator fit in the chimp lineage or loss in the human lineage, an important validation of our analysis. Conversely, we see that chimp-biased enhancers are similarly prone to gains of the Coordinator motif in the chimp lineage (n = 218) versus losses in the human lineage (n = 255) and again, with almost no gains in human or losses in chimp. Thus, there appears to be no preferred direction of enhancer divergence in either lineage since the split from our common ancestor for this class of sites. We also applied our analysis to hominin outgroups such as Denisovans and Neanderthals and found that, as expected given the much more recent split from the common ancestor, these lineages primarily share the human-like variants of the Coordinator motif at species-biased sites (Figure S6G). Therefore, even for individuals substantially more diverged than any modern human, most changes are present in the hominin lineage relative to the human-chimp ancestor. However, there is a small set of changes that are unique to modern humans compared to other hominins, and those clearly merit further exploration.

Species-Biased Enhancers Flank Genes that Show Species-Biased Expression

Recent studies suggest that gene expression levels are more evolutionarily conserved than utilization of *cis*-regulatory elements and can be buffered by redundant or compensatory elements regulating the same loci (Hong et al., 2008; Odom et al., 2007; Schmidt et al., 2010; Vierstra et al., 2014; Wong et al., 2014). Nonetheless, at least some of the species-biased enhancers should be associated with transcriptional changes at nearby genes if they are responsible for morphological variation. To test this, we performed RNA-seq analyses of our human and chimp CNCC populations and identified genes whose expression significantly diverged between, but not within, species. We found that genes with significantly divergent expression between humans and chimpanzees are strongly enriched for nearby species-biased enhancers, with human-biased genes flanked by human-biased enhancers and chimp-biased genes flanked by chimp-biased enhancers (Figure 6A). In addition, we observed that the fraction of species-biased genes (but not the degree of the expression bias) scales with the number of flanking enhancers biased toward the same species (Figure 6B).

Clusters of Regulatory Divergence Flank Loci Involved in Intra-Human Facial Variation

Interestingly, we found that strongly divergent enhancers were not distributed at random throughout the genome but instead were likely to fall in close genomic proximity to other species-biased enhancers matching in polarity (Figure S7A), suggesting that divergent enhancers fall into regulatory clusters. To systematically locate these clusters, we calculated a genome-wide divergence score using a moving window over the nearest ~10 enhancers for each species, integrating both the degree and genomic span of divergent enhancers in series (Figure S7B). This strategy revealed a low baseline encompassing the bulk of interspersed species-biased enhancers (examples of Chr11 in Figures S7C and S7D, top) but exposed a subset of regions throughout the genome (~1–4 per chromosome), with a marked increase in their divergence score resulting from presence of dense clusters of strongly biased enhancers (Figure 6C). Importantly, we find that these clusters of divergence do not emerge simply by chance due to increased frequency of enhancers near highly active CNCC genes (Figures S7C and S7D).

When ranking all human- and chimp-biased enhancers according to their divergence score, we observed an inflection in the distribution (Figures 6D for human, 6E for chimp). Using this inflection point as a cutoff, we identified 32 human and 65 chimp clusters of divergence, spanning genomic windows of, on average, ~500 kb and encompassing ~11.9% of all species-biased enhancers. Of note, while some clusters overlapped super-enhancers in CNCCs, most super-enhancers were not identified as a species-biased cluster and many species-biased clusters did not encompass super-enhancers, indicating that these two entities are distinct (Whyte et al., 2013).
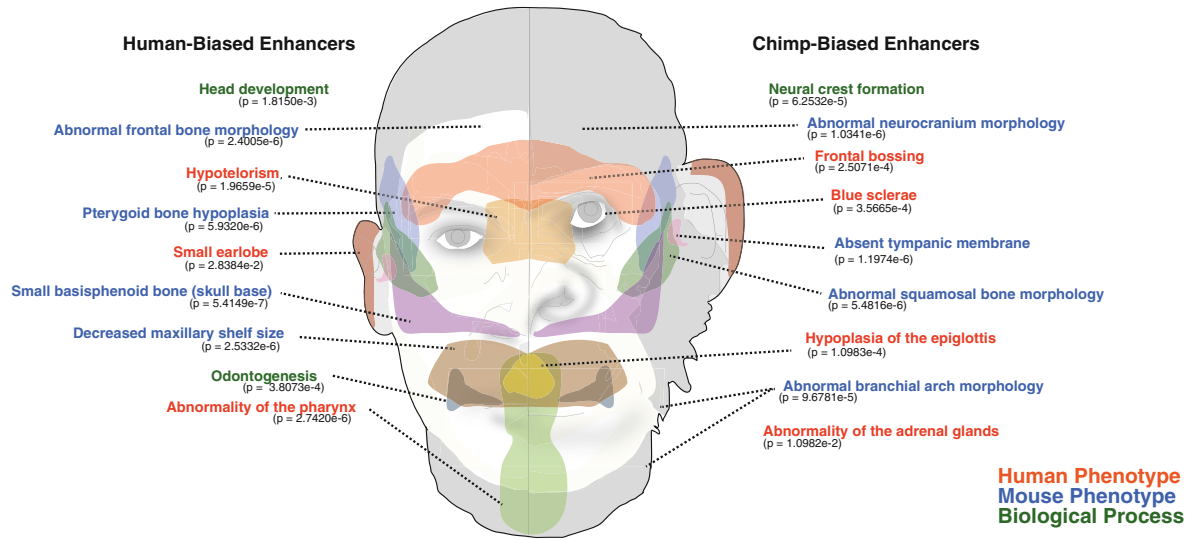
We speculate that these species-biased enhancer clusters represent broad *cis*-regulatory regions under strong evolutionary pressure to diverge and hypothesize that they may contain genes with central roles in the regulation of NC-associated phenotypes. Indeed, these regions fall immediately over or next to genes that are critical in facial morphogenesis, including *PRDM16*, *MN1*, *COL17A1*, *EDNRA*, *PAX3*, *PAX7*, *SOX10*, and *ALX4*. Intriguingly, of five chromosomal regions linked to normal-range human facial variation in GWAS, three (*PRDM16*, *COL17A1*, and *PAX3*) fall directly within these regions of high divergence. Importantly, the clusters were highly predictive of changes in nearby gene

(C) Representative browser tracks showing clusters of species-biased enhancers. Top panel shows broad view with predicted species-bias track (human-biased in orange, chimp-biased in blue) and the corresponding H3K4me1 (blue), p300 (red), and H3K27ac (green) from two individuals of each species shown in overlay. Boundaries of the cluster are indicated by a red block. Close-up of an individual cluster of biased enhancers is shown below. All chromatin features are mapped to hg19.

(D and E) Distribution of divergence scores at human-biased enhancers (D) and chimp-biased enhancers (E). Selected genes falling within identified clusters are highlighted next to the enhancer in the cluster with highest divergence score.

(F) Mean normalized human expression (x axis) versus mean normalized chimp expression (y axis) for genes within or flanking human-biased enhancer clusters (blue) or chimp-biased enhancer clusters (red).

A

Human-Biased Enhancers

Head development
(p = 1.8150e-3)

Abnormal frontal bone morphology
(p = 2.4005e-6)

Hypotelorism
(p = 1.9659e-5)

Pterygoid bone hypoplasia
(p = 5.9320e-6)

Small earlobe
(p = 2.8384e-2)

Small basisphenoid bone (skull base)
(p = 5.4149e-7)

Decreased maxillary shelf size
(p = 2.5332e-6)

Odontogenesis
(p = 3.8073e-4)

Abnormality of the pharynx
(p = 2.7420e-6)

Chimp-Biased Enhancers

Neural crest formation
(p = 6.2532e-5)

Abnormal neurocranium morphology
(p = 1.0341e-6)

Frontal bossing
(p = 2.5071e-4)

Blue sclerae
(p = 3.5665e-4)

Absent tympanic membrane
(p = 1.1974e-6)

Abnormal squamosal bone morphology
(p = 5.4816e-6)

Hypoplasia of the epiglottis
(p = 1.0983e-4)

Abnormal branchial arch morphology
(p = 9.6781e-5)

Abnormality of the adrenal glands
(p = 1.0982e-2)

Human Phenotype
Mouse Phenotype
Biological Process

B

Gene	Expression bias:	DESEQ q value	Coordinates of nearby species-biased enhancers (hg19)	Relevant genetic phenotypes, disease associations and comments	References
PAX3	C	4.1584E-06	chr2:223983203-223983402; chr2:223157099-223157298 ; chr2:223148418-223148617 ; chr2:223084744-223084943 ; chr2:223006007-223006206 ; chr2:223005367-223005566; (see Figure S4B);	Heterozygous loss-of-function mutations cause Waardenburg syndrome, characterized by craniofacial, auditory and pigmentation defects; in model organisms Pax3 is involved in induction, specification and differentiation of neural crest cells and craniofacial development; in GWAS studies the locus was associated with normal-range variation of facial morphology in Europeans.	Asher et al., 1996; Conway et al., 1997; Liu et al., 2012; Milet et al., 2013; Paternoster et al., 2012; Tassabehji et al., 1992; Tremblay et al., 1995;
PAX7	C	0.11724596	chr1:19267843-19268042 ; chr1:19268294-19268493 ; chr1:19268848-19269047 ; chr1:19271215-19271414 ; chr1:19275542-19275741 ; chr1:19276342-19276541	Involved in early specification of the neural crest in the embryo; loss of function in the mouse leads to reduction of the maxilla and a pointed snout.	Basch et al., 2006; Mansouri et al., 1996;
EDNRA	C	1.1308E-04	chr4:148262113-148262312; chr4:148276381-148276580 ; chr4:148454528-148454727 ; chr4:148455160-148455359 ; chr4:148507330-148507529 ; chr4:148508238-148508437; chr4:148510911-148511110 (see Figure S4B);	Mouse deficient for Ednra exhibit cranial and cardiac neural crest defects. Most lower jaw structures in Ednra−/− embryos undergo a homeotic transformation into maxillary-like structures; other defects include absence of tympanic rings, malleus, and incus, and the rostral relocation of the hyoid bone.	Clouthier et al., 1998, 2010; Ruest et al., 2004
EDN3	H	5.2338E-03	chr20:57723838-57724037 ; chr20:57724804-57725003 ; chr20:57726203-57726402 ; chr20:57894099-57894298 ; chr20:57988186-57988385 ; chr20:58028164-58028363 ; chr20:58028993-58029192 ; chr20:58154612-58154811 (see Figure S4B);	Heterozygous or homozygous mutations associated with several human neurocristopathies, including Waardenburg syndrome type 4B, Hirschprung's disease, and Congenital Central Hypoventilation Syndrome (CCHS); in animal models EDN3 is involved in the regulation of coat pigmentation and enteric neuron function.	Baynash et al., 1994; Bidaud et al., 1997; Dupin et al., 2000; Edery et al., 1996; Lahav et al., 1996; Pingault et al., 2001;
NRP2	H	4.7551E-02	chr2:206515229-206515428 ; chr2:206436082-206436281 ; chr2:206266143-206266342 ; chr2:206250975-206251174	Involved in guidance of NCC migration and restricting migratory paths of cranial and trunk NCCs, positioning sensory neurons and organizing their projections.	Gammill et al., 2007; Schwarz et al., 2008; Yu and Moens, 2005;
EPHB2	H	5.1251E-02	chr1:23164766-23164965 ; chr1:23164254-23164453 ; chr1:23162983-23163182 ; chr1:23162487-23162686	Ephrin B signaling is involved in targeting and restricting neural crest migration within branchial arches; compound EphB2/B3 knockout in mice leads to cleft palate.	Orioli et al., 1996; Risley et al., 2009; Smith et al., 1997;
BMP4	H	5.4181E-02	chr14:54911474-54911673 ; chr14:55093544-55093743; chr14:55094758-55094957 ; chr14:55095533-55095732	Proposed to be a crucial mediator of beak shape changes in Darwin's finches and of craniofacial shape and morphological adaptive radiation in Cichlid fish. CNCC-specific overexpression of BMP4 during mouse development results in the dramatic change of facial shape, with shortening in both the mandible and maxilla, rounding of the skull shape, and more anterior orientation of the eyes. In humans, mutations/variants in *BMP4* are associated with orofacial clefts, microphthalmia, and age of the primary tooth eruption.	Abzhanov et al., 2004; Albertson et al., 2005; Boell et al., 2013; Fatemifar et al., 2013; Tucker et al., 1998; Wu et al., 2004;
BMPER	C	2.2758E-05	chr7:33525807-33526006 ; chr7:33526320-33526519 ; chr7:33527256-33527455 ; chr7:33540319-33540518 ; chr7:33540928-33541127; (see Figure S4B)	Negative regulator of BMP4 function in osteoblast and chondrocyte differentiation (see also *BMP4*). In humans, homozygous or heterozygous mutations in *BMPER* are associated with a skeletal disorder, diaphanospondylodysostosis, whose consistent craniofacial features include ocular hypertelorism, epicanthal folds, depressed nasal bridge with short nose, and low-set ears.	Binnerts et al., 2004; Funari et al., 2010; Moser et al., 2003;
PITX2	C	1.2564E-11	chr4:111230391-111230590; chr4:111230942-111231141 ; chr4:111820988-111821187	PITX2 haploinsufficiency is associated with Axenfeld-Rieger syndrome involving ocular anterior segment dysgenesis, tooth anomalies, and craniofacial anomalies such as maxillary hypoplasia with mid-face flattening and prominent forehead; in mice, ocular manifestations are largely recapitulated by the neural crest-specific knockout of Pitx2; genetically interacts with *FOXC1*, see also *FOXC1*.	Evans and Gage, 2005; Lu et al., 1999; Matt et al., 2005; Semina et al., 1996
FOXC1	C	3.4267E-02	chr6:1744897-1745096	Heterozygous mutations in *FOXC1* are associated with Axenfeld-Rieger syndrome (A-RS) involving ocular anterior segment dysgenesis, tooth anomalies, and maxillary and mandibular hypoplasia; dosage-dependent interactions with another A-RS gene *PITX2* have been observed; in mice, loss of FoxC1 results in bony syngnathia, defects in maxillary and mandibular structures, and agenesis of the temporomandibular joint; see also *PITX2* .	Berry et al., 2006; Inman et al., 2013; Kelberman et al., 2011; Mears et al., 1998; Tümer and Bach-Holm, 2009;
POU3F3	H	2.1354E-02	chr2:105024721-105024920 ; chr2:104990082-104990281 ; chr2:104989534-104989733 ; chr2:104937657-104937856	In mouse knockout leads to loss of squama temporalis and stapes fusion to styloid process.	Dheedene et al., 2014; Jeong et al., 2008;

(legend on next page)

expression for the bulk of the associated genes in the region (Figure 6E), suggesting that either (1) multiple genes in the vicinity must be under coordinated selection for these super-divergent regions to emerge or, more likely, that (2) strong selection on one or a few target genes could drive changes in the local enhancer landscape that have secondary effects on other genes in the vicinity. Altogether, we provide evidence that highly divergent clusters of tissue-specific enhancers may promote interspecies and intra-species phenotypic variation.

Resource for Studies of Human Morphological Evolution

In addition to informing the basic mechanisms underlying the *cis*-regulatory divergence of human and chimpanzee NC, our study also provides a rich resource for future investigations of morphological evolution of human craniofacial traits. Ontology annotations of all significantly species-biased enhancers reveal strong associations with processes important for various craniofacial structures that are diverging in human and chimps (Figure 7A). As examples, we highlight some of the most interesting divergent candidate genes in Figure 7B. These featured loci show species-biased expression in our RNA-seq and also map to regions with species-biased enhancer divergence and are emphasized due to their known associations with CNCC development and/or facial morphology. Nonetheless, it is important to bear in mind that the biases in gene expression and enhancer states highlighted in Figure 7 refer to the relative change between human and chimpanzee CNCCs, without ascribing the polarity of the change with respect to the ancestral status.

Our divergently expressed genes are known to be involved in multiple, distinct developmental processes that cooperate to influence differential allocation of CNCCs in facial primordia and, in turn, contribute to species-specific morphology (Fish et al., 2014). These processes (and associated species-biased genes) include: (1) CNCC specification (e.g., *PAX3*, *PAX7*), (2) migration and guidance of CNCC migratory paths (e.g., *EPHB2*, *NRP2*, *EDNRA*, *EDN3*), (3) modulation of CNCC proliferation at facial primordia (e.g., *BMP4*), and (4) regulation of CNCC differentiation (e.g., *PITX2*). Moreover, heterozygous mutations in many of these genes (e.g., *PAX3*, *PITX2*, *FOXC1*, *EDN3*, *BMPER*) are associated with human syndromes that include craniofacial manifestations, suggesting that altered gene dosage can drive both morphological variation between species and, below a certain threshold, disease-associated malformations (Figure 7B). Furthermore, many phenotypes of the highlighted genes affect aspects of head morphology that have diverged between humans and chimps (e.g., size of the mandible and maxilla, skull shape, and pigmentation) (Figure 7B and Discussion). Altogether, our study provides a wealth of candidate loci for further deep exploration in studies of human evolution and variation.

DISCUSSION

Our study utilizes primate cellular models to provide a comprehensive map of human and chimp regulatory divergence in a tissue with central relevance to the development of the head and face. We show that a common mechanism of regulatory divergence in higher primates is quantitative modulation of orthologous elements, driven largely through small numbers of sequence changes that perturb tissue-specific TF binding motifs. This is consistent with previous studies from closely related *Drosophila* or mouse species demonstrating that large effects can be conferred by a small number of mutations affecting direct and cooperative binding of key TFs (Bradley et al., 2010; He et al., 2011; Stefflova et al., 2013). Interestingly, we find that not all TF binding sites contribute equally to regulatory divergence—in fact, we identify a broad spectrum of regulatory motifs that vary in frequency and effect, suggesting a mechanism through which evolution can fine-tune *cis*-regulation across an enhancer landscape. One outlier in our analysis is the Coordinator motif, a de novo consensus sequence that is strongly predictive of the surrounding chromatin features and is highly enriched at species-biased enhancers. We speculate that the factor(s) that recognize the Coordinator motif play a privileged role in the establishment of enhancer competence in this cell context, reminiscent of the *Drosophila* TAG-team motif bound by a pioneer factor Zelda (Liang et al., 2008; Satija and Bradley, 2012). Furthermore, we find evidence of repressive inputs into quantitative modulation of enhancer activity, with a sizable number of motifs whose gain in strength negatively correlates with acquisition of permissive chromatin states.

Our work provides a rich framework for future gene-centric studies on the developmental mechanisms of human morphological evolution. Indeed, our approach identified loci that are known to profoundly affect NC development and craniofacial morphology, often in a dosage-sensitive manner. For example, we observed that two genes involved in CNCC specification, *PAX3* and *PAX7*, are expressed at higher levels in chimps and are associated with clusters of chimp-biased enhancers. In mice, mutations of these TFs lead to reduction of pigmentation and snout length (*Pax3*) (Tremblay et al., 1995) and reduction of maxilla and pointed snout (*Pax7*) (Mansouri et al., 1996), features that are consistent with smaller jaw size and hypopigmentation of humans as compared to chimps. Furthermore, humans are sensitive to alterations of *PAX3* dosage, as haploinsufficiency of this gene is associated with craniofacial, auditory, and pigmentation defects (Waardenburg syndrome, OMIM #193510), and genetic variants at this locus have been identified in GWAS studies as regulators of normal-range facial shape (Liu et al., 2012; Paternoster et al., 2012). Thus, variation in

---

Figure 7. Species-Biased Enhancers Are Associated with Genes Affecting Craniofacial Structures

(A) GREAT term enrichments and associated facial regions indicated for human-biased enhancers (q < 0.01, baseMean > 300) and chimp-biased enhancers (q < 0.01, baseMean > 300); binomial raw p values are shown below. Ontology categories are color coded (human phenotypes, red; mouse phenotypes, blue; biological processes, green).

(B) Table of highlighted divergently expressed genes showing direction of bias (human-biased versus chimp-biased indicated by H or C, respectively), DESeq adjusted p value of expression divergence, coordinates of nearby species-biased enhancers with corresponding bias (hg19), description of genetic phenotypes, disease associations, comments, and relevant references. Full reference information can be found in Table S1.

PAX3 and PAX7 levels represents an attractive possible mechanism for mediating facial shape divergence between humans and chimpanzees.

We also find evidence that genes already known to affect facial morphology in other species, such as *BMP4*, are diverging in higher primates as well. BMP4 is the most well-understood example of a factor that influenced craniofacial morphological change during evolution, as it has been implicated in mediating changes in beak morphology in Darwin's finches (Abzhanov et al., 2004) and in jaw shape in Cichlid fish (Albertson et al., 2005). We were therefore intrigued to note that *BMP4* is associated with strongly human-biased enhancers and is expressed at higher levels in humans than in chimps. Conversely, expression of the BMP4 inhibitor *BMPER* was significantly chimp biased and showed dramatic strengthening of the local chimp enhancer landscape. What would be the potential effects of elevated BMP4 expression on primate facial development? Interestingly, in the mouse, CNCC-specific overexpression of *BMP4* results in a dramatic change of facial shape, with shortening of both the mandible and maxilla, rounding of the skull, and more anterior orientation of the eyes (Bonilla-Claudio et al., 2012)—morphological changes that resemble those observed between human and chimps. Thus, the same molecular mechanism that has been postulated to influence beak morphology in Darwin's finches may also contribute to our uniquely human facial features.

Even more intriguing, of five chromosomal regions that have been associated with normal-range human facial variation in GWAS, three (*PRDM16*, *COL17A1*, and *PAX3*) coincide with clusters of species-biased enhancers uncovered in our study (Liu et al., 2012; Paternoster et al., 2012), suggesting a significant overlap between loci regulating intra- and inter-species variation of facial shape in higher primates. We therefore hypothesize that other divergent clusters identified in our study represent novel candidates for loci involved in the regulation of facial shape in humans. More broadly, we suggest that comparisons of human regulatory landscapes with those of a closely related primate in any tissue of interest may provide an effective strategy to identify candidate loci involved in normal-range and disease-associated variation.

EXPERIMENTAL PROCEDURES

CNCC Derivation
Pluripotent lines were differentiated into CNCC as previously described (Rada-Iglesias et al., 2012). Details are provided in the Supplemental Experimental Procedures.

Chromatin Immunoprecipitation and Preparation of ChIP-Seq Libraries
Chromatin immunoprecipitation (ChIP) was performed using ~0.5–1 × 10^7 CNCCs per experiment, as previously described (Bajpai et al., 2010; Rada-Iglesias et al., 2011, 2012). Antibodies used for ChIPs are listed in the Supplemental Experimental Procedures. Sequencing libraries were prepared starting from 30 ng of ChIP DNA using the NEBNext Multiplex Oligos for Illumina kit (Cat# E7335S). Libraries were multiplexed four to six samples per lane for 1 × 50 bp next-gen sequencing on Illumina HiSeq platform.

Quantitative Analysis of H3K27ac ChIP-Seq and Identification of Divergence
All sequencing reads were aligned to both reference genomes (hg19 and pan-Tro3) using default settings with bowtie2.2.4, regardless of species of origin.

Modal peak positions for candidate regulatory elements were determined using a mean shift procedure, described in the Supplemental Experimental Procedures. To obtain count statistics for each H3K27ac ChIP alignment, we counted read coverage in a 1.6 kb window surrounding modal peak positions. ENCODE-blacklisted regions and outlier regions with high counts in control input sequences relative to ChIP were removed as artifacts. Scores for visualization and classification of remaining ChIPs were obtained using a kernel density estimate, as previously described (Buecker et al., 2014).

Calculations of species bias were inferred with DESeq2, based on the read counts from all replicates of H3K27ac at candidate enhancers from three human lines (one hESC, two iPSC) and two chimp lines (two iPSC). DESeq2 analysis was performed separately for panTro3 and hg19 counts; then conservatively, the higher p-adj value and lower abs(log2FoldChange) of the analysis from either hg19 or panTro3 were assigned to each region, while rare regions with discordant calls were excluded from list of biased sites (less than 0.1%).

REFERENCES

Abzhanov, A., Protas, M., Grant, B.R., Grant, P.R., and Tabin, C.J. (2004). Bmp4 and morphological variation of beaks in Darwin's finches. Science 305, 1462–1465.

Albertson, R.C., Streelman, J.T., Kocher, T.D., and Yelick, P.C. (2005). Integration and evolution of the cichlid mandible: the molecular basis of alternate feeding strategies. Proc. Natl. Acad. Sci. USA 102, 16287–16292.

Attanasio, C., Nord, A.S., Zhu, Y., Blow, M.J., Li, Z., Liberton, D.K., Morrison, H., Plajzer-Frick, I., Holt, A., Hosseini, R., et al. (2013). Fine tuning of craniofacial morphology by distant-acting enhancers. Science 342, 1241006.

Bajpai, R., Chen, D.A., Rada-Iglesias, A., Zhang, J., Xiong, Y., Helms, J., Chang, C.-P., Zhao, Y., Swigut, T., and Wysocka, J. (2010). CHD7 cooperates with PBAF to control multipotent neural crest formation. Nature 463, 958–962.

Bilsborough, A., and Wood, B.A. (1988). Cranial morphometry of early hominids: facial region. Am. J. Phys. Anthropol. 76, 61–86.

Bonilla-Claudio, M., Wang, J., Bai, Y., Klysik, E., Selever, J., and Martin, J.F. (2012). Bmp signaling regulates a dose-dependent transcriptional program to control facial skeletal development. Development 139, 709–719.

Bradley, R.K., Li, X.-Y., Trapnell, C., Davidson, S., Pachter, L., Chu, H.C., Tonkin, L.A., Biggin, M.D., and Eisen, M.B. (2010). Binding site turnover produces pervasive quantitative changes in transcription factor binding between closely related Drosophila species. PLoS Biol. *8*, e1000343.

Bronner, M.E., and LeDouarin, N.M. (2012). Development and evolution of the neural crest: an overview. Dev. Biol. *366*, 2–9.

Buecker, C., Srinivasan, R., Wu, Z., Calo, E., Acampora, D., Faial, T., Simeone, A., Tan, M., Swigut, T., and Wysocka, J. (2014). Reorganization of enhancer patterns in transition from naive to primed pluripotency. Cell Stem Cell *14*, 838–853.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods *10*, 1213–1218.

Cain, C.E., Blekhman, R., Marioni, J.C., and Gilad, Y. (2011). Gene expression differences among primates are associated with changes in a histone epigenetic modification. Genetics *187*, 1225–1234.

Carroll, S.B. (2008). Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. Cell *134*, 25–36.

Cordaux, R., and Batzer, M.A. (2009). The impact of retrotransposons on human genome evolution. Nat. Rev. Genet. *10*, 691–703.

Cordero, D.R., Brugmann, S., Chu, Y., Bajpai, R., Jame, M., and Helms, J.A. (2011). Cranial neural crest cells on the move: their roles in craniofacial development. Am. J. Med. Genet. A. *155A*, 270–279.

Cotney, J., Leng, J., Yin, J., Reilly, S.K., DeMare, L.E., Emera, D., Ayoub, A.E., Rakic, P., and Noonan, J.P. (2013). The evolution of lineage-specific regulatory activities in the human embryonic limb. Cell *154*, 185–196.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc. Natl. Acad. Sci. USA *107*, 21931–21936.

Fish, J.L., Sklar, R.S., Woronowicz, K.C., and Schneider, R.A. (2014). Multiple developmental mechanisms regulate species-specific jaw size. Development *141*, 674–684.

Gompel, N., Prud'homme, B., Wittkopp, P.J., Kassner, V.A., and Carroll, S.B. (2005). Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in Drosophila. Nature *433*, 481–487.

Gronau, I., Arbiza, L., Mohammed, J., and Siepel, A. (2013). Inference of natural selection from interspersed genomic elements based on polymorphism and divergence. Mol. Biol. Evol. *30*, 1159–1171.

He, Q., Bardet, A.F., Patton, B., Purvis, J., Johnston, J., Paulson, A., Gogol, M., Stark, A., and Zeitlinger, J. (2011). High conservation of transcription factor binding and evidence for combinatorial regulation across six Drosophila species. Nat. Genet. *43*, 414–420.

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat. Genet. *39*, 311–318.

Hong, J.-W., Hendrix, D.A., and Levine, M.S. (2008). Shadow enhancers as a source of evolutionary novelty. Science *321*, 1314.

Hubisz, M.J., and Pollard, K.S. (2014). Exploring the genesis and functions of Human Accelerated Regions sheds light on their role in human evolution. Curr. Opin. Genet. Dev. *29*, 15–21.

Jheon, A.H., and Schneider, R.A. (2009). The cells that fill the bill: neural crest and the evolution of craniofacial development. J. Dent. Res. *88*, 12–21.

Kaessmann, H., Wiebe, V., and Pääbo, S. (1999). Extensive nuclear DNA sequence diversity among chimpanzees. Science *286*, 1159–1162.

King, M.C., and Wilson, A.C. (1975). Evolution at two levels in humans and chimpanzees. Science *188*, 107–116.

Liang, H.-L., Nien, C.-Y., Liu, H.-Y., Metzstein, M.M., Kirov, N., and Rushlow, C. (2008). The zinc-finger protein Zelda is a key activator of the early zygotic genome in Drosophila. Nature *456*, 400–403.

Lieberman, D.E. (1998). Sphenoid shortening and the evolution of modern human cranial shape. Nature *393*, 158–162.

Liu, F., van der Lijn, F., Schurmann, C., Zhu, G., Chakravarty, M.M., Hysi, P.G., Wollstein, A., Lao, O., de Bruijne, M., Ikram, M.A., et al. (2012). A genome-wide association study identifies five loci influencing facial morphology in Europeans. PLoS Genet. *8*, e1002932.

Mansouri, A., Stoykova, A., Torres, M., and Gruss, P. (1996). Dysgenesis of cephalic neural crest derivatives in Pax7-/- mutant mice. Development *122*, 831–838.

Marchetto, M.C.N., Narvaiza, I., Denli, A.M., Benner, C., Lazzarini, T.A., Nathanson, J.L., Paquola, A.C.M., Desai, K.N., Herai, R.H., Weitzman, M.D., et al. (2013). Differential L1 regulation in pluripotent stem cells of humans and apes. Nature *503*, 525–529.

McLean, C.Y., Reno, P.L., Pollen, A.A., Bassan, A.I., Capellini, T.D., Guenther, C., Indjeian, V.B., Lim, X., Menke, D.B., Schaar, B.T., et al. (2011). Human-specific loss of regulatory DNA and the evolution of human-specific traits. Nature *471*, 216–219.

Odom, D.T., Dowell, R.D., Jacobsen, E.S., Gordon, W., Danford, T.W., MacIsaac, K.D., Rolfe, P.A., Conboy, C.M., Gifford, D.K., and Fraenkel, E. (2007). Tissue-specific transcriptional regulation has diverged significantly between human and mouse. Nat. Genet. *39*, 730–732.

Paternoster, L., Zhurov, A.I., Toma, A.M., Kemp, J.P., St Pourcain, B., Timpson, N.J., McMahon, G., McArdle, W., Ring, S.M., Smith, G.D., et al. (2012). Genome-wide association study of three-dimensional facial morphology identifies a variant in PAX3 associated with nasion position. Am. J. Hum. Genet. *90*, 478–485.

Pollard, K.S., Salama, S.R., Lambert, N., Lambot, M.-A., Coppens, S., Pedersen, J.S., Katzman, S., King, B., Onodera, C., Siepel, A., et al. (2006). An RNA gene expressed during cortical development evolved rapidly in humans. Nature *443*, 167–172.

Prabhakar, S., Poulin, F., Shoukry, M., Afzal, V., Rubin, E.M., Couronne, O., and Pennacchio, L.A. (2006). Close sequence comparisons are sufficient to identify human cis-regulatory elements. Genome Res. *16*, 855–863.

Prabhakar, S., Visel, A., Akiyama, J.A., Shoukry, M., Lewis, K.D., Holt, A., Plajzer-Frick, I., Morrison, H., Fitzpatrick, D.R., Afzal, V., et al. (2008). Human-specific gain of function in a developmental enhancer. Science *321*, 1346–1350.

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. Nature *470*, 279–283.

Rada-Iglesias, A., Bajpai, R., Prescott, S., Brugmann, S.A., Swigut, T., and Wysocka, J. (2012). Epigenomic annotation of enhancers predicts transcriptional regulators of human neural crest. Cell Stem Cell *11*, 633–648.

Satija, R., and Bradley, R.K. (2012). The TAGteam motif facilitates binding of 21 sequence-specific transcription factors in the Drosophila embryo. Genome Res. *22*, 656–665.

Schmidt, D., Wilson, M.D., Ballester, B., Schwalie, P.C., Brown, G.D., Marshall, A., Kutter, C., Watt, S., Martinez-Jimenez, C.P., Mackay, S., et al. (2010). Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. Science *328*, 1036–1040.

Shapiro, M.D., Marks, M.E., Peichel, C.L., Blackman, B.K., Nereng, K.S., Jónsson, B., Schluter, D., and Kingsley, D.M. (2004). Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. Nature *428*, 717–723.

Shibata, Y., Sheffield, N.C., Fedrigo, O., Babbitt, C.C., Wortham, M., Tewari, A.K., London, D., Song, L., Lee, B.-K., Iyer, V.R., et al. (2012). Extensive evolutionary changes in regulatory element activity during human origins are associated with altered gene expression and positive selection. PLoS Genet. *8*, e1002789.

Spoor, F., Wood, B., and Zonneveld, F. (1994). Implications of early hominid labyrinthine morphology for evolution of human bipedal locomotion. Nature *369*, 645–648.

Stefflova, K., Thybert, D., Wilson, M.D., Streeter, I., Aleksic, J., Karagianni, P., Brazma, A., Adams, D.J., Talianidis, I., Marioni, J.C., et al. (2013). Cooperativity

and rapid evolution of cobound transcription factors in closely related mammals. Cell *154*, 530–540.

Tremblay, P., Kessel, M., and Gruss, P. (1995). A transgenic neuroanatomical marker identifies cranial neural crest deficiencies associated with the Pax3 mutant Splotch. Dev. Biol. *171*, 317–329.

Twigg, S.R.F., Versnel, S.L., Nürnberg, G., Lees, M.M., Bhat, M., Hammond, P., Hennekam, R.C.M., Hoogeboom, A.J.M., Hurst, J.A., Johnson, D., et al. (2009). Frontorhiny, a distinctive presentation of frontonasal dysplasia caused by recessive mutations in the ALX3 homeobox gene. Am. J. Hum. Genet. *84*, 698–705.

Vierstra, J., Rynes, E., Sandstrom, R., Zhang, M., Canfield, T., Hansen, R.S., Stehling-Sun, S., Sabo, P.J., Byron, R., Humbert, R., et al. (2014). Mouse regulatory DNA landscapes reveal global principles of cis-regulatory evolution. Science *346*, 1007–1012.

Villar, D., Berthelot, C., Aldridge, S., Rayner, T.F., Lukk, M., Pignatelli, M., Park, T.J., Deaville, R., Erichsen, J.T., Jasinska, A.J., et al. (2015). Enhancer evolution across 20 mammalian species. Cell *160*, 554–566.

Visel, A., Minovitsky, S., Dubchak, I., and Pennacchio, L.A. (2007). VISTA Enhancer Browser–a database of tissue-specific human enhancers. Nucleic Acids Res. *35*, D88–D92.

Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell *153*, 307–319.

Wong, E.S., Thybert, D., Schmitt, B.M., Stefflova, K., Odom, D.T., and Flicek, P. (2014). Decoupling of evolutionary changes in transcription factor binding and gene expression in mammals. Genome Res.

Wray, G.A. (2007). The evolutionary significance of cis-regulatory mutations. Nat. Rev. Genet. *8*, 206–216.

Zhou, X., Cain, C.E., Myrthil, M., Lewellen, N., Michelini, K., Davenport, E.R., Stephens, M., Pritchard, J.K., and Gilad, Y. (2014). Epigenetic modifications are associated with inter-species gene expression variation in primates. Genome Biol. *15*, 547.