Human-Specific Histone Methylation Signatures at Transcription Start Sites in Prefrontal Neurons


1 Program in Bioinformatics and Integrative Biology, University of Massachusetts Medical School, Worcester, Massachusetts, United States of America, 2 Department of Human Genetics and Genomics, Vavilov Institute of General Genetics, Moscow, Russian Federation, 3 Budnick Neuropsychiatric Research Institute, University of Massachusetts Medical School, Worcester, Massachusetts, United States of America, 4 New England Primate Center, Southborough, Massachusetts, United States of America, 5 Department of Neurology, Boston University, Boston, Massachusetts, United States of America, 6 Yerkes National Primate Research Center/Emory University, Atlanta, Georgia, United States of America, 7 Research Center of Mental Health, Russian Academy of Medical Sciences, Moscow, Russian Federation, 8 School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, 9 Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Russian Federation, 10 Departments of Psychiatry and Neuroscience, Friedman Brain Institute, Mount Sinai School of Medicine, New York, New York, United States of America

Abstract

Cognitive abilities and disorders unique to humans are thought to result from adaptively driven changes in brain transcriptomes, but little is known about the role of cis-regulatory changes affecting transcription start sites (TSS). Here, we mapped in human, chimpanzee, and macaque prefrontal cortex the genome-wide distribution of histone H3 trimethylated at lysine 4 (H3K4me3), an epigenetic mark sharply regulated at TSS, and identified 471 sequences with human-specific enrichment or depletion. Among these were 33 loci selectively methylated in neuronal but not non-neuronal chromatin from children and adults, including TSS at DPP10 (2q14.1), CNTN4 and CHL1 (3p26.3), and other neuropsychiatric susceptibility genes. Regulatory sequences at DPP10 and additional loci carried a strong footprint of hominid adaptation, including elevated nucleotide substitution rates and regulatory motifs absent in other primates (including archaic hominins), with evidence for selective pressures during more recent evolution and adaptive fixations in modern populations. Chromosome conformation capture at two neurodevelopmental disease loci, 2q14.1 and 16p11.2, revealed higher order chromatin structures resulting in physical contact of multiple human-specific H3K4me3 peaks spaced 0.5–1 Mb apart, in conjunction with a novel cis-bound antisense RNA linked to Polycomb repressor proteins and downregulated DPP10 expression. Therefore, coordinated epigenetic regulation via newly derived TSS chromatin could play an important role in the emergence of human-specific gene expression networks in brain that contribute to cognitive functions and neurological disease susceptibility in modern day humans.


Academic Editor: David Penny, Massey University, New Zealand

Received May 28, 2012; Accepted October 12, 2012; Published November 20, 2012

Copyright: © 2012 Shulha et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: Supported by Yerkes Base Grant, P51RR000165, NEPRC Base Grant P51RR000168, US NIH grants R01MH081943, R21NS076958, R01071476, R01DG005153, R01DA021420, R01 AG029360, and Ministry of Education and Science of the RF 16.512.11.2102, 02.740.11.0854; EU FP7 242257-ADAMS; RFBR 10-04-02078. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Abbreviations: 3C, chromosome conformation capture; H3K4me3, trimethyl-H3-lysine 4; HEK, human embryonic kidney; HP, human-specific peak; HSA, human-specific sequence alteration; PFC, prefrontal cortex; RT, reverse transcriptase; TSS, transcription start site.

* E-mail: schahram.akbarian@msm.edu (SA); jeffrey.jensen@epfl.ch (JDJ); evgeny.rogaev@umassmed.edu (EIR); zhiping.weng@umassmed.edu (ZW)

* These authors contributed equally to this work.

Introduction

Cognitive abilities and psychiatric diseases unique to modern humans could be based on genomic features distinguishing our brain cells, including neurons, from those of other primates. Because protein coding sequences for synaptic and other neuron-specific genes are highly conserved across the primate tree [1,2], a significant portion of hominid evolution could be due to DNA sequence changes involving regulatory and non-coding regions at the 5’ end of genes [3,4]. Quantifying these differences, however, is ultimately a daunting task, considering that, for example, the chimpanzee–human genome comparison alone reveals close to 35 x 10^6 single bp and 5 x 10^4 multi-bp substitutions and insertion/deletion events [3]. While a large majority of these are likely to reflect genetic drift and are deemed “non-consequential” with respect to fitness, the challenge is to identify the small subset of regulatory sequence alterations impacting brain function and behavior.

Here, we combine comparative genomics and population genetics with genome-scale comparisons for histone H3-trimethyl-lysine 4 (H3K4me3), an epigenetic mark sharply regulated at transcription start sites (TSS) and the 5’ end of transcriptional units in brain and other tissues [5–8] that is stably maintained in brain specimens collected postmortem [7,9]. Our rationale to focus on TSS chromatin was also
Author Summary

Primate and human genomes comprise billions of base pairs, but we are unlikely to gain a deeper understanding of brain functions unique to human (including cognitive abilities and psychiatric diseases) merely by comparing linear DNA sequences. Such determinants of species-specific function might instead be found in the so-called “epigenetic” characteristics of genomic regions; differences in the protein-packaged chromatin state in which genomic DNA exists in the cell. Here, we examine neurons from the prefrontal cortex, a brain region closely associated with the evolution of the primate brain, and identify hundreds of short DNA sequences defined by human-specific changes in chromatin structure and function when compared to non-human primates. These changes included species-specific regulation of methylation marks on the histone proteins around which genomic DNA is wrapped. Sequences subject to human-specific epigenetic regulation showed significant spatial clustering, and despite being separated by hundreds of thousands of base pairs on the linear genome, were in direct physical contact with each other through chromatosomal looping and other higher order chromatin features. This observation raises the intriguing possibility that coordinated epigenetic regulation via newly derived chromatin features at gene transcription start sites could play an important role in the emergence of human-specific gene expression networks in the brain. Finally, we identified a strong genetic footprint of hominid evolution in a small subset of transcription start sites defined by human-specific gains in histone methylation, with particularly strong enrichment in prefrontal cortex neurons. For example, the base pair sequence of motifs absent in non-human primates and archaic hominins specifically in the human branch of the primate tree, regulatory mechanisms remains unclear. Here, we report that cell type-specific epigenome mapping in prefrontal cortex neurons (including cytoarchitectonic Brodmann Area BA10 and the higher association cortex subject to disproportionate morphological expansion during primate evolution [13], and are involved in cognitive operations important for informed choice and creativity [14,15], among other executive functions. Given that histone methylation in neuronal and non-neuronal chromatin is differentially regulated at thousands of sites genome-wide [7], we avoided chromatin studies in tissue homogenates because glia-to-neuron ratios are 1.4- to 2-fold higher in mature human PFC as compared to chimpanzee and macaque [16]. Instead, we performed cell type-specific epigenome profiling for each of the three primate species, based on NeuN (“neuron nucleus”) antigen-based immunotagging and fluorescence-activated sorting, followed by deep sequencing of H3K4me3-tagged neuronal nucleosomes.

Prefrontal H3K4me3 epigenomes from NeuN+ nuclei of 11 humans, including seven children and four adults [7], were compared to four chimpanzees and three macaques of mature age (Table S1). Sample-to-sample comparison, based on a subset of highly conserved Refseq TSS with one mismatch maximum/36bp, consistently revealed the highest correlations between neuronal epigenomes from the same species (Table S2). Strikingly, however, the H3K4me3 landscape in human neurons was much more similar to chimpanzee and macaque neurons, when compared to non-neuronal (NeuN−) cells [7] from the same specimen/donor or to blood (Figure 1A). Therefore, PFC neuronal epigenomes, including their histone methylation landscapes at TSS, carry a species-specific signature, but show an even larger difference when compared to their surrounding glial and other NeuN− cells.

Several Hundred Loci Show Human-Specific Gain, or Loss, of Histone Methylation in PFC Neurons

To identify loci with human-specific H3K4me3 enrichment in PFC neurons, we screened 34,639 H3K4me3 peaks that were at least 500 bp long and showed a consistent $>2$-fold H3K4me3 increase for the 11 humans as compared to the average of the seven chimps and macaques and (ii) minimum length of 500 bp. We identified 410 peaks in the human genome (HG19) with significant enrichment compared to the two non-human primate species (with reads also mapped to HG19) after correcting for false discovery (FDR), and we call these peaks “HP” hereafter for “human-specific peaks” (Figure 1D; Table S3). We had previously reported that infant and child PFC neurons tend to have stronger peaks at numerous loci, compared to the adult [7]. To better age-match the human and non-human primate cohorts, we therefore repeated the analysis with our entire, recently published cohort of nine adult humans without known neurological or psychiatric disease [7,8]. Using the same set of filter criteria ($>2$-fold increase in humans compared to chimpanzees and macaques), we identified 425 peaks and 296 of them overlapped with the original 410 HP (Table S3). Furthermore, 345 of the 410 peaks overlapped with the overlapped with the peaks with $>1.5$-fold increase for nine adult humans (compared to non-human primates; with correction for FDR) (Table S4), indicating that HPs can be detected reliably. To obtain human depleted peaks we used a reciprocal approach where initial peaks were detected in chimpanzee and macaque. For the original cohort of 11 children and adult humans, this resulted in 61 peaks with a significant, at least 2-fold depletion in
human PFC neurons (Table S5). 50 peaks defined by human-specific depletion in the mixed cohort of 11 children and adults were part of the total of 177 peaks with >1.5-fold decrease in the cohort of nine adults (compared to each of the two non-human primate species; Table S6). From this, we conclude that at least 471 loci in the genome of PFC neurons show robust human-specific changes (gain, 410; loss, 61) in histone methylation across a very wide postnatal age range.

We further explored chimpanzee-specific changes in the H3K4me3 landscape of PFC neurons by comparing human and chimpanzee peaks within the chimpanzee genome. To this end, we constructed a mono-nucleosomal DNA library from chimpanzee PFC to control for input, and mapped the neuronal H3K4me3 landscapes in PFC neurons by comparing human and chimpanzee peaks within the chimpanzee genome (PT2). We identified 551 peaks in the PT2 genome that were subject to >2-fold gain and 337 peaks subject to >2-fold depletion, compared to human regardless of the H3K4me3 level in macaque (Tables S7 and S8). A substantial portion of these PT2-annotated peaks (133 and 40 peaks, respectively) with gain or loss in chimpanzee PFC neurons matched loci with the corresponding, reciprocal changes specific to human PFC neurons in HG19 (410 and 61 peaks as described above). Genetic differences among these genomes and additional, locus-specific differences in nucleosomal organization (leading to differences in background signal in the input libraries) are potential factors that would lead to only partial matching of peaks when species-specific H3K4me3 signals are mapped within the human, or chimpanzee genome, respectively. These findings, taken together, confirm that genome sequence differences in cis are one important factor for the species-specific histone methylation landscapes in PFC neurons.

Human-Specific H3K4me3 Peaks in PFC Neurons Overlap with DNA Methylation Signatures in the Male Germline

Both catalytic and non-catalytic subunits of H3K4 methyltransferase complex are associated with transgenerational epigenetic inheritance in the worm, Caenorhabditis elegans, and other simple model organisms [17], and furthermore, H3K4me3 and other epigenetic markings such as DNA cytosine methylation are readily detectable in non-somatic (“germline”-related) cells such as sperm, potentially passing on heritable information to human offspring [18]. Therefore, we wanted to explore whether a subset of the 410 loci with at least 2-fold H3K4me3 enrichment in human neurons are subject to species-specific epigenetic regulation in germ tissue. To this end, we screened a human and chimpanzee sperm database on DNA methylation [19], in order to find out which, if any of the 410 sequences with human-specific H3K4me3 gain in brain overlap with a set of >70,000 sequences defined by very low, or non-detectable DNA methylation in human and chimpanzee sperm (termed [DNA] “hypomethylated regions” in [19]). Of note, the genome-wide distribution of H3K4me3 and DNA cytosine methylation is mutually exclusive in germ and embryonic stem cells, and gains in DNA methylation generally are associated with loss of H3K4me3 in differentiated tissues [20,21]. Unsurprisingly therefore, 300/410 HP peaks in brain matched a DNA hypomethylated sequence in sperm of both species. Strikingly, however, 90/410, or approximately 22% of HP were selectively (DNA) hypomethylated in human but not in chimpanzee sperm (Table S3), a ratio that is approximately 4-fold higher than the expected 5.7% based on 10,000 simulations (p<0.00001; see also Text S1) (Figure 1B). Conversely, the portion of HP lacking DNA hypomethylation in male germ cells of either species altogether (18/410 or 4%), or with selective hypomethylation in chimpanzee
sperm (2/410 or 0.5%), showed a significant, 5-fold underrepresentation in our dataset (Figure 1B). Thus, approximately one-quarter of the 410 loci with human-specific gain in histone methylation in PFC neurons also carry species-specific DNA methylation signatures in sperm, with extremely strong bias towards human (DNA) hypomethylated regions (22%) compared to chimpanzee-specific (DNA) hypomethylated regions (0.5%).

In striking contrast, fewer than ten of the 61 loci with human-specific H3K4me3 depletion in PFC neurons showed species-specific differences in sperm DNA methylation between species (six human- and three chimpanzee-specific DNA hypomethylated regions; Table S5).

H3K4 Methylation Sites with Human-Specific Gain Physically Interact in Megabase-Scale Higher Order Chromatin Structures and Provide an Additional Layer for Transcriptional Regulation

We noticed that, at numerous chromosomal loci, HP tended to group in pairs or clusters (Table S3). There were more than 245 (163) from the total of 410 HP spaced less than 1 (or 0.5) Mb apart, which is a highly significant, 2- (or 3-) fold enrichment compared to random distribution within the total pool of 34,639 peaks (Figure 1C; Text S1). Therefore, sequences with human-specific gain in H3K4me3 depletion in PFC neurons showed species-specific differences in sperm DNA methylation between species (six human- and three chimpanzee-specific DNA hypomethylated regions; Table S5).

This type of non-random distribution due to pairing or clustering of the majority of human-enriched sequences broadly resonates with the recently introduced concept of Mb-sized topological domains as a pervasive feature of genome organization, including increased physical interactions of sequences carrying the same set of epigenetic decorations within a domain [22]. Of note, H3K4 trimethylation of nucleosomes is linked to the RNA polymerase II transcriptional initiation complex, and sharply increased around TSS and broadly correlated with “open chromatin” and gene expression activity [5,6]. Therefore, we reasoned that a subset of human-enriched “paired” H3K4me3 peaks could engage in chromatin loopings associated with transcriptional regulation. This is a very plausible hypothesis given that promoters and other regulatory sequences involved in transcriptional regulation are often tethered together in loopings and other higher order chromatin [23,24].

To explore this, we screened a database obtained on chromatin interaction analysis by paired-end tag sequencing (ChIA-PET) for RNA polymerase II, a technique designed to detect chromosomal loopings bound by the Pol II complex [25]. Indeed, we identified at least three interactions that matched to our H3K4me3 peaks with human-specific gain in PFC neurons (Table S9), including a loop interspersed by approximately 2.5 Mb of sequence in
chromosome 16p11.2–12.2. This is a risk locus for microdeletions that are linked to a wide spectrum of neurodevelopmental disease including autism spectrum disorder (ASD), intellectual disability (ID), attention deficit hyperactivity disorder (ADHD), seizures, and schizophrenia [26–31]. We were able to validate this interaction by chromosome conformation capture (3C), a technique for mapping long range physical interactions between chromatin segments [32], in 2/2 human PFC specimens and also in a human embryonic kidney (HEK) cell line (Figure 2). We conclude that human-specific H3K4me3 peaks spaced as far apart as 1 Mb are potentially co-regulated and physically interact via chromatin loopings and other higher order chromatin structures.

Neuronal Antisense RNA LOC389023 Originating from a DPP10 (Chromosome 2q14) Higher Order Chromatin Structure Forms a Stem-Loop and Interacts with Transcriptional Repressors

Next, we wanted to explore whether sequences with human-specific gain in histone methylation, including those that show evidence for pairing and physical interactions, could affect the regulation of gene expression specifically in PFC neurons. To this end, we first identified which portion from the total of 410 human-specific peaks showed much higher H3K4me3 levels selectively in PFC neurons, when compared to their surrounding non-neuronal cells in the PFC. Thus, in addition to the aforementioned filter.
criteria (2-fold increase in human PFC neurons compared to non-human primate PFC neurons), we searched for peaks with differential regulation among PFC neurons and non-neurons (see Text S1). We found 33 HP with selective enrichment in neuronal PFC chromatin (termed neuHP in the following) (Figure S1; Table S10). Among these were two HP spaced less than 0.5 Mb apart within the same gene, DPP10 (chr2q14.1), encoding a dipeptidyl peptidase-related protein regulating potassium channels and neuronal excitability (Figure 3A–3B) [33]. Interestingly, rare structural variants of DPP10 confer strong genetic susceptibility to autism, while some of the gene’s more common variants contribute to a significant risk for bipolar disorder, schizophrenia, and asthma [34–36]. Histone methylation at DPP10 was highly regulated in species- and cell type-specific manner, with both DPP10-1 and DPP10-2 peaks defined by a very strong H3K4me3 signal in human PFC neurons (Figure 3A), but only weak or non-detectable peaks in their surrounding NeuN2 (non-neuronal) nuclei (Figure S1; Table S10) or blood-derived epigenomes [7]. We then employed 3C assays across 1.5 Mb of the DPP10 (chr2q14.1) in PFC of four humans. To increase the specificity in each 3C PCR assay, we positioned both the forward and reverse primer in the same orientation on the sense strand, and samples
processed for 3C while omitting the critical DNA ligation step from the protocol served as negative control (Figure 3A-3B). Indeed, 3C assays on four of four human PFC specimens demonstrated direct contacts between the DPP10-1 and -2 peaks (Figure 3A). As expected for neighboring fragments [32], DPP10-1 also interacted with portions of the interspersed sequence (CR2 in Figure 3A). These interactions were specific, because several other chromatin segments within the same portion of chr2q14.1 did not show longer range interactions with DPP10-1 (CR1, CR3 in Figure 3A). We further verified one of the DPP10-1/2 physical interactions (the sequences captured by primers 6 and 17 in Figure 3A) in four of five brains using 3C-qPCR with a TaqMan probe positioned in fragment 6. Furthermore, DPP10-2 interacted with a region (“CR3” in Figure 3A) 400 kb further downstream positioned in close proximity to a blood-specific H3K4me3 peak. No interactions at the DPP10 locus were observed in cultured cells derived from the H9 embryonic stem cell line (H9ESC in Figure 3A), suggesting that these chromatin architectures are specific for differentiated brain tissue. Of note, similar types of DPP10 physical interactions were found in 3C assays conducted on PFC tissue of three of three macaques (Figure 3B). Because macaque PFC, in comparison to human, shows much weaker H3K4 methylation at these DPP10 sequences, we conclude that the corresponding chromatin tetherings are not critically dependent on human-specific H3K4me3 dosage.

Next, we wanted to explore whether human-specific H3K4 methylation at the DPP10 locus is associated with a corresponding change in gene expression at that locus. Notably, H3K4me3 is on a genome-wide scale broadly correlated with transcriptional activity, including negative regulation of RNA expression by generating very short (50–200 nt) promoter-associated RNAs.
neuronal functions that may have contributed not only to the emergence of human-specific executive and social-emotional functions, but also for increased susceptibility for developmental brain disease [42]. In this context, we noticed that the 33 **HP** (which are defined by two criteria which are (i) human-specific gain compared to non-human primates and (ii) high H3K4me3 in PFC neurons but not their surrounding non-neuronal cells) included multiple genes conferring susceptibility to neurological disease. Three loci, including **DPP10** on chromosome 2q14.1 and two genes in close proximity on chromosome 3p26.3, **CNTN4** and **CHLI**, both encoding cell adhesion molecules [34,43–45], confer very strong susceptibility to autism, schizophrenia, and related disease. Other disease-associated loci with human-specific gain selectively in PFC neurons include **ADCAPI**, a schizophrenia [46,47] and movement disorder gene [48] that is part of a cAMP-activating pathway also implicated in posttraumatic stress [49]. **PDE4DIP** (**MYOMEALIN**) (Figure 1D) encodes a centrosomal regulator of brain size and neurogenesis [50] that in some studies was 9-fold higher expressed in human as compared to chimpanzee cortex [51,52]. **SORC5** is implicated in beta amyloid processing and Alzheimer disease [53,54] and attention deficit hyperactivity disorder [55], which again are considered human-specific neurological conditions [10]. Because four of 33, or 12% of **HP** overlapped with neurodevelopmental susceptibility genes (**CNTN4**, **CHLI**, **DPP10**, **SORC5**), we then checked whether the entire set of 410 human-specific peaks is enriched for genes and loci conferring genetic risk for autism, intellectual disability, and related neurological disease with onset in early childhood. However, there was only minimal overlap with the Simons Foundation Autism Research Initiative database (SFARI) [56], and Human unidentified Gene Encoded protein database (HuGE) for pervasive developmental disorder (including autism) associated polymorphism [57], and recent reference lists for mental retardation and/ or autism-related genes (each of these databases five or fewer of the human-enriched peaks) [58]. Likewise, there was minimal, and non-significant overlap with the set of 61 human- and 337 chimpanzee-depleted peaks, or the 551 chimpanzee-enriched in PFC neurons (five or fewer of peaks/database). None of the lists of peaks with human- or chimpanzee-specific gain or loss of H3K4me3 revealed statistical significance for any associations with the Gene Ontology (GO) database. We conclude that DNA sequences subject to differential histone methylation in human or chimpanzee PFC neurons are, as a group, not clustered together into specific cellular signaling pathways or functions. Table 1 presents examples of disease-associated genes associated with human-specific gain, or loss of H3K4-trimethylation.

**Evolutionary Footprints at Sites Defined by Human-Specific Histone Methylation**

We then asked whether the subset of DNA sequences with species- and cell type-specific epigenetic regulation, including the **HP** peaks mentioned above carry a strong footprint of hominid evolution. Indeed, nucleotide substitution analysis revealed that both **DPP10** peaks **DPP10** -1/2, as well as **ADCAPI**, **CHLI**, **CNTN4**, **NR5V2**, and **SIRPA** show a significantly elevated rate, with 2- to 5-fold increase specifically in the human branch of the primate tree, when compared to four other anthropoid primate species (Pan troglodytes, Gorilla gorilla, Pongo abelii, Macaca mulatta) (Table S11). The finding that both **DPP10** peaks, **DPP10**-1 and -2 showed a significant, >4-fold increase in nucleotide substitution rates in the human branch of the primate tree—indicating “co-evolution” (or coordinated loss of constraint)—is very plausible given that chromatin structures surrounding these DNA sequences are in direct physical contact (discussed above), reflecting a

Consistent with a possible function inside the nucleus, **LOC389023** was highly enriched in nuclear RNA fractions from extracted prenatal and normal (non-degenerative) adult human PFC, but not cerebellar cortex (Figure 5B). Indeed, in transiently transfected (human) SK-N-MC neuroblastoma cells, **LOC389023** showed a specific association with H3K4-trimethylated nucleosomes and SUZ12 (Figure 5D), a zinc finger protein and core component of PRC2 previously shown to interact with stem loop motifs [37]. These observations, taken together, are entirely consistent with the aforementioned findings that levels of **DPP10** transcript, including exons positioned downstream of the **DPP10**-2 peak from which **LOC389023** originates, are significantly decreased in human PFC as compared to macaque and chimpanzee. Conversely, these two primates show non-detectable (RNAseq) or much lower quantitative RT-PCR (qRT-PCR) **LOC389023** levels in the PFC, as compared to human (Figure 4A–4B). Taken together then, these findings strongly suggest that **LOC389023** emerged de novo in human PFC neurons and interacts with localized chromatin templates to mediate transcriptional repression at the **DPP10** locus (Figure 6).

**Association of Human-Specific H3K4-Methylation Sites with Disease**

The aforementioned human-specific gains in histone methylation at **DPP10** and the emergence of human RNA de novo at this locus could reflect a phylogenetically driven reorganization of
potential functional interaction and shared regulatory mechanisms between peaks.

To further confirm the role of phylogenetic factors in the emergence of human-specific H3K4me3 peaks, we focused on the set of 33 *H. sapiens* and calculated the total number of human-specific sequence alterations (HSAs), in a comparative genome analyses across five primates (*H. sapiens*, *P. troglodytes*, *G. gorilla*, *P. abelii*, *M. mulatta*). We recorded altogether 1,519 HSAs, with >90% as single-locus events.
nucleotide substitutions, five >100 bp INDELs, one (46k) retro-
transposon-like element at TRIB3 pseudokinase consistent with a
role of mobile elements in primate evolution [3], and gain or
loss of hundreds of regulatory motifs (Table S12). When compared to
a group of (neuronal) H3K4me3 peaks showing minimal changes
between the three primate species (Table S13), the "HP" as a
group, showed a significant, 2.5-fold increase in the number of
HSA (20.08 ± 5.52 HSAs versus 8.36 ± 2.44 HSAs per 1-kb
sequence, p = 2.4e−06, Wilcoxon rank sum test; Figure S3). The
findings further confirm that genetic differences related to
speciation indeed could play a major role for changes in the
brain's histone methylation landscape, particularly for H3K4me3
peaks that are highly specific for human neurons ("HP").
Interestingly, none of the above loci showed evidence for
accelerated evolution of neighboring protein coding sequences
(Table S11), reaffirming the view that protein coding sequences for
synaptic and other neuron-specific genes are extremely conserved
across the primate tree [1,2].

These DNA sequence alterations at sites of neuron-restricted
H3K4me3 peaks (with human-specific gain) point, at least for this
subset of loci, to a strong evolutionary footprint before the split of
human–chimpanzee lineage several million years ago [3]. Next,
we wanted to find out whether there is also evidence for more
recent selective pressures at these loci. Indeed, a subset of "HP
contain H. sapiens-specific sequences not only absent in rodents,
anthropoid primates, but even in extinct members of the genus
homo, including H. neanderthalensis and H. denisova [59]. Some of the
ancestral alleles (including MIA1, Sirpa, MSR3) shared with
archaic hominins exhibit very low frequencies at 0%–3% in all
modern populations, and therefore it remains possible that positive
selection for newly derived alleles contributed to their high
population frequencies in modern humans (Table S14). However,
for the entire set of "HP that are defined by high H3K4me3 levels in
PFC neurons (but not non-neurons), the number of HSAs that
emerged after the human lineage was split from H. denisova or H.
neanderthalensis were 3.31% and 1.75%, respectively, which is
approximately 2-fold lower as compared to 32 control H3K4me3
peaks with minimal differences among the three primate species
(5.03% and 3.77%). The 2-fold difference in the number of H.
sapiens-specific alleles ("HP compared to control peaks) showed a
strong trend toward significant (p = 0.067) for the Denisova, and
reached the level of significance (p = 0.034) for the Neanderal
genome (based on permutation test with 10,000 simulations [60]).
Taken together, these results suggest that at least a subset of the
TSS regions with H3K4me3 enrichment in human (compared to
non-human primates) were exposed to evolutionary driven DNA
sequence changes on a lineage of the common ancestor of H.
sapiens and the archaic hominins, but subsequently were stabilized
in more recent human evolution, after splitting from other
hominins.

To further test whether or not there were recent, perhaps even
ongoing selective pressures at loci defined by human-specific gain
in H3K4me3 peaks of PFC neurons, we searched for overlap
among the peaks in our study with hundreds of candidate regions
in the human genome showing evidence of selection during the
past 10–100,000 years from other studies. These loci typically
extend over several kb, and were identified in several recent studies
on the basis of criteria associated with a “selective sweep,” which
describes the elimination of genetic variation in sequences
surrounding an advantageous mutation while it becomes fixed
[61–64]. However, screening of the entire set of 410 human gain
and 61 human depleted H3K4me3 sequences against nine datasets
for putative selection in humans [65] revealed only five loci with
evidence for recent sweeps (Table S15). One of these matched to
the "HP on chromosome 2q14.1, corresponding to the second
DPP10 (DPP10-2) peak (see above). In independent analyses, using
the 1,000 genome database, we further confirmed recent adaptive
fixations around DPP10-2 (Table S16), as well as two other loci, POL1 and TSPAN4.
While it is presently extremely difficult to
determine how much of the genome has been affected by positive
selection (of note, a recent metaanalysis of 21 recent studies using
total genomic scans for positive selection using human polymor-
phism data) revealed unexpectedly minimal overlap between
studies [65], we conclude that the overwhelming majority of loci
associated with human-specific H3K4me3 gain or loss in PFC
neurons (compared to non-human primates) indeed does not show
evidence for more recent selective pressures.

To provide an example on altered chromatin function due to an
alteration in a regulatory DNA sequence that occurred after the
human lineage split from the common ancestor with non-human
primates, we focused on a change in a GATA-1 motif (A/TGATTGA)
within a portion of DPP10-2 found in human, within
the otherwise deeply conserved sequence across many mammalian
lineages (Table S17). Gel shift assays demonstrate that the human-
specific sequence harboring the novel GATA-1 site showed much
higher affinity to HeLa nuclear protein extracts, compared to the
chimpanzee/other mammal sequence (Figure 4C). The emergence
of a novel GATA-1 motif at DPP10 is unlikely to reflect a systemic
trend because the motif overall was lost, rather than gained in
"HP (10/355 versus 4/375, χ2 p = 0.053). Therefore, evolution-
ary and highly specific changes in a small subset of regulatory
motifs at DPP10 and other loci could potentially result in profound
changes in nuclear protein binding at TSS and other regulatory
sequences, thereby affecting histone methylation and epigenetic
control of gene expression in humans, compared to other
mammals including monkeys and great apes. Of note, potentially
important changes in chromatin structure and function due to
human-specific sequence alterations at a single nucleotide within
an otherwise highly conserved mammalian sequence will be
difficult to “capture” by comparative genome analyses alone. For
example, when the total set of 410 HP was crosschecked against a
database of 202 sequences with evidence for human-specific
accelerated evolution in loci that are highly conserved between
primate lineages [66], only one of 410 HP matched
(Table S15).

Species-Specific Transcriptional Regulation
H3K4me3 is a transcriptional mark that on a genome-wide scale is
broadly associated with RNA polymerase II occupancies and
RNA expression [67]. However, it is also associated with repressive
chromatin remodeling complexes and at some loci the mark is
linked to short antisense RNAs originating from bidirectional
promoters, in conjunction with negative regulation of the (sense)
gene transcript [37,38]. Indeed, this is what we observed for the
DPP10 locus (Figure 6). Therefore, a comprehensive assessment of
all transcriptional changes associated with the evolutionary alter-
ations in H3K4me3 landscape of PFC neurons would require deep
sequencing of intra- and extranuclear RNA, to ensure full capture of
short RNAs and all other transcripts that lack polyadenylation and/
or export into cytoplasm. While this is beyond the scope of the
present study, we found several additional examples for altered
RNA expression at the site of human-specific H3K4me3 change.
There were four of 33 "HP loci associated with novel RNA
expression specific for human PFC, including the aforementioned
DPP10 locus. The remaining three human-specific transcripts
included two additional putative non-coding RNAs, LOC421321(chr7p14.3) and AX746692 (chr1p11.2). There was
also a novel transcript for ASPARATE DEHYDROGENASE

PLOS Biology | www.plosbiology.org
ISOFORM 2 (ASPDH)(chr19q13.33) (Figure S2). Furthermore, a fifth ASPDH, positioned within an intronic portion of the tetraspanin gene TSPAN4(chr11p13.5), was associated with a dramatic, human-specific decrease of local transcript, including the surrounding exons (Figure S2). Comparative analyses of prefrontal RNA-seq signals for the entire set of the 410 HP included at least 10 loci showing a highly consistent, at least 2-fold increase or decrease in RNA levels of human PFC, compared to the other two primate species (Table S10).

Discussion

In the present study, we report that on a genome-wide scale, 471 loci show a robust, human-specific change in H3K4me3 levels at TSS and related regulatory sequences in neuronal chromatin from PFC, in comparison to the chimpanzee and macaque. Among the 410 sequences with human-specific gain in histone methylation, there was a 4-fold overrepresentation of loci subject to species-specific DNA methylation in sperm [19]. This would suggest that there is already considerable “epigenetic distance” between the germline of H. sapiens and non-human primates (including the great apes), which during embryonic development and tissue differentiation is then “carried over” into the brain’s epigenome. The fact that many loci show species-specific epigenetic signatures both in sperm [19] and PFC neurons (Figure 1B) raises questions about the role of epigenetic inheritance [68] during hominid evolution. However, to further clarify this issue, additional comparative analysis of epigenetic markings in brain and germline will be necessary, including histone methylation maps from oocytes, which currently do not exist. However, the majority of species-specific epigenetic decorations, including those that could be vertically transmitted through the germline, could ultimately be driven by genetic differences. On the basis of DNA methylation analyses in three-generation pedigrees, more than 92% of the differences in methylcytosine load between alleles are explained by haplotype, suggesting a dominant role of genetic variation in the establishment of epigenetic markings, as opposed to environmental influences [69]. A broad overall correlation between genetic and epigenetic differences was also reported in a recent human–chimpanzee sperm DNA methylation study [19], and there is general consensus that the inherent mutability of methylated cytosine residues due to their spontaneous deamination to thymine is one factor contributing to sequence divergence at CpG rich promoters with differential DNA methylation and there is general consensus that the inherent mutability of methylated cytosine residues due to their spontaneous deamination to thymine is one factor contributing to sequence divergence at CpG rich promoters with differential DNA methylation and higher order chromatin structures. The confluence of these factors could then, in a subset of PFC neurons (Figure 5A), result in the expression of a novel antisense RNA, which associates with transcriptional repressors to regulate the target transcript in cis, DPP10 (Figures 5D and 6).

While the present study identified a few loci, including the aforementioned DPP10 (chromosome 2q14.1), in which DNA sequences associated with a human-specific gain in neuronal histone methylation showed signs for positive selection in the human population, it must be emphasized that the overwhelming majority of sites with human-specific H3K4me3 changes did not show evidence for recent adaptive fixations in the surrounding DNA. Therefore, and perhaps not unsurprisingly, neuronal histone methylation mapping in human, chimpanzee, and macaque primarily reveals information about changes in epigenetic regulation of regulatory sequences in the hominid genome after our lineage split from the common ancestor shared with present-day non-human primates.

Moreover, according to the present study, the subset of 33 sequences with human-specific H3K4me3 gain and selective enrichment in neuronal (as opposed to non-neuronal) PFC chromatin show a significant, 3-fold increase in human-specific (DNA sequence) alterations in comparison to non-human primate genomes. This finding speaks to the importance of evolutionary changes in regulatory sequences important for neuronal functions. Strikingly, however, the same set of sequences show a significant, approximately 1.5- to 2-fold decrease in sequence alterations when compared to the two archaic hominins (H. denisova, H. neanderthalensis) genomes. This finding further reaffirms that sequences defined by differential epigenetic regulation in human and non-human primate brain, as a group, are unlikely to be of major importance for more recent evolution, including any (yet elusive) genetic alterations that may underlie the suspected differences in human and neanderthal brain development [73]. However, these general conclusions by no means rule out a critical role for a subset of human-specific sequence alterations on the single nucleotide level within any of the HPs described here, including the DPP10 locus.

Such types of single nucleotide alterations and polymorphisms may be of particular importance at the small number of loci with human-specific H3K4me3 gain that contribute to susceptibility of neurological and psychiatric disorders that are unique to human (though it should be noticed that as a group, the entire set of sequences subject to human-specific gain, or loss, of H3K4me3 are not significantly enriched for neurodevelopmental disease genes). The list would not only include the already discussed ADYCAP1, CHL1, CNTN4, and DPP10, which were among the narrow list of 33 human-specific peaks highly enriched in neuronal but not non-neuronal PFC chromatin), but also DGCR6, an autism and schizophrenia susceptibility gene [74,75] within the DiGeorge/
Velocardiofacial syndrome/22q11 risk locus, NOTCH1 and C4/RAHC encoding transmembrane signaling proteins linked to schizophrenia and bipolar disorder in multiple genome-wide association studies [76,77], SLCO2A3 encoding a neuronal glucose transporter linked to dyslexia and attention-deficit hyperactivity disorder [78,79] and the neuronal migration gene TUBB2B that has been linked to polymicrogria and defective neurodevelopment [80]. Furthermore, among the 61 peaks with human-specific loss of H3K4me3 is a 700-bp sequence upstream of the TSS of FOXP2, encoding a forkhead transcription factor essential for proper human speech and language capabilities [81] and that has been subject to accelerated evolution with amino acid changes leading to partially different molecular functions in human compared to great apes [82,83]. The homeobox gene LMX1B is another interesting disease-associated gene that is subject to human-specific H3K4me3 depletion (Table 1). While expression of many of these disease-associated genes is readily detectable even in mouse cerebral cortex [84], the neuropsychiatric conditions associated with them lack a correlate in anthropoid primates and other animals. This could speak to the functional significance of H3K4 methylation as an additional layer for transcriptional regulation, with adaptive H3K4me3 changes at select loci and TSS potentially resulting in improved cognition while at the same time in the context of genetic or environmental risk factors contribute to neuropsychiatric disease. More generally, our findings are in line with a potential role for epigenetic (dys)regulation in the pathophysiology of a wide range of neurological and psychiatric disorders [85–88].

Our study also faces important limitations. While we used child and adult brains for cross-species comparisons, human-specific signatures in the cortical transcriptome are thought to be even more pronounced during pre- and perinatal development [89]. Therefore, younger brains could show changes at additional loci or more pronounced alterations at the TSS of some genes identified in the present study, including the above mentioned susceptibility genes CNTV4 and myelopregnalin/PDE4DIP, which are expressed at very high levels in the human frontal lobe at midgestation [90]. In this context, our finding that a large majority, or 345 of 410 H3K4me3 peaks showed a human-specific gain both in children and adults, resonates with Somel and colleagues [11] who suggested that some of the age-sensitive or more pronounced alterations at the TSS of some genes due to trans-acting factors such as microRNA,s while cis-regulatory changes (which were the focus of the present study) primarily affect genes that are subject to a lesser regulation by developmental processes. More broadly, our studies supports the general view that transcriptional regulation of both of coding and non-coding (including antisense) RNAs could play a role in the evolution of the primary brain [91].

Furthermore, the cell type-specific, neuronal versus non-neuronal chromatin studies as presented here provide a significant advancement over conventional approaches utilizing tissue homogenate. However, pending further technological advances, it will be interesting to explore genome organization in select subsets of nerve cells that bear particularly strong footprints of adaptation, such as the Von Economo neurons, a type of cortical projection neuron highly specific for the hominid lineage of the primate tree and other mammals with complex social and cognitive-emotional skill sets [92]. Furthermore, our focus on PFC does not exclude the possibility that other cortical regions [93], or specialized sublayers such as within the fourth layer of visual cortex that shows a complex transcriptional architecture [94], show human-specific histone methylation gains at additional TSS that were missed by the present study.

More broadly, the approach provided here, which is region- and cell type-specific epigenome mapping in multiple primate species, highlights the potential of epigenetic markings to identify regulatory non-coding sequences with a potential role in the context of hominid brain evolution and the shaping of human-specific brain functions. Remarkably, a small subset of loci, including the aforementioned DPP10 (chromosome 2q14.1), shows evidence for ongoing selective pressures in humans, resulting in DNA sequence alterations and the remodeling of local histone methylation landscapes, after the last common ancestor of human and non-human primates.

**Materials and Methods**

Text S1 contains detailed description for sample preparation for ChIP-seq and RNA-seq, qRT-PCR, gel shift, and 3C assays including primer sequences, RNA immunoprecipitation and in situ hybridization, bioinformatics and analyses of deep sequencing data, exploration of regulatory motifs, calculation of nucleotide substitution rates in the primate tree, and sweep analyses for polymorphic regions.

**Supporting Information**

Figure S1 H3K4me3 ChIP-seq browser tracks (UCSC) for H3K4me3 peaks with >2-fold gain in human PFC, compared to chimpanzee and macaque. y-Axis represents normalized tag densities (0–15, ppm) after annotation to the three reference genomes (HG19, rheMac2 = RM2, panTro2 = PT2).

PDF

Figure S2 RNAseq tag densities in human, chimpanzee, and macaque PFC for H3K4me3 peaks shown in Figure S1.

(PDF)

Figure S3 Comparison of number of human-specific DNA sequence alterations in H3K4me3 peaks with and without human-specific gain.

(PPTX)

Table S1 Sample information, including age, gender, and postmortem brain interval, and H3K4me3 ChIP-seq parameters.

(XLS)

Table S2 Sample-to-sample correlations of raw promoter tag counts of H3K4me3 ChIP-seq from PFC NeuN+ nuclei.

(XLSX)

Table S3 List of 410 H3K4me3 peaks with human-specific gain, with at least 2-fold higher normalized tag densities in 11 humans as compared to the three macaques and four chimpanzees, including human genome (HG) 19 coordinates, distance to nearest TSS, and overlap (1) or no overlap (0) with DNA hypomethylated regions in sperm DNA methylation database comparing human and chimpanzee [20].

(XLSX)

Table S4 List of 885 H3K4me3 peaks with human-specific gain, with at least 1.5-fold higher tag density in nine adult humans as compared to the three macaques and four chimpanzees.

(XLSX)

Table S5 List of 61 H3K4me3 peaks with human-specific depletion, with at least 2-fold lower normalized tag densities in 11 (seven children, four adult) humans as compared to the three macaques and four chimpanzees, including human genome (HG) 19 coordinates, distance to nearest TSS, and
overlap (1) or no overlap (0) with DNA hypomethylated regions in sperm DNA methylation database comparing human and chimpanzee [20].

(XLSX)

Table S6 List of 177 H3K4me3 peaks with human-specific enrichment, with at least 1.5-fold lower normalized tag densities in nine adult humans as compared to the three macaques and four chimpanzees.

(XLSX)

Table S7 List of 551 H3K4me3 peaks with chimpanzee-specific enrichment, with at least 2-fold higher tag density in four chimpanzees compared to 11 humans.

(XLSX)

Table S8 List of 337 H3K4me3 peaks selectively depleted in the chimpanzee, with at least 2-fold higher tag density in 11 humans compared to four chimpanzees.

(XLSX)

Table S9 Sequences with human-specific H3K4me3 gain in prefrontal neurons that were recently shown to be a part of chromatin loopings in conjunction with RNA polymerase II occupancy [25].

(XLSX)

Table S10 Genome coordinates of the subset of 33 neuHP that are significantly enriched in prefrontal neurons (NeuN+) as compared to lymphocytes (see Text S1, “ChiP-seq analyses”) and to non-neurons in the PFC (NeuN−; Figure S1) and that show human-specific gain in 11 humans as compared to the three macaques and four chimpanzees, including HG19 (and after liftover), Rhesus macaque 2 (RM2) and P. tresta 2 (PT2) genome coordinates, and species-specific enrichments.

(XLS)

Table S11 Nucleotide substitution rates for the 33 neuHP defined by human-specific gain and neuron-specific enrichment. Baseml (for genomic sequence) and codeml (for amino acid sequence) was to calculate branch- and site-specific nucleotide substitution rates, using human, chimpanzee, gorilla, orangutan, and macaque genome sequences.

(XLSX)

Table S12 Regulatory motifs (cis-Red database) gained and lost in the 1,519 human-specific DNA sequence alterations found in the 33 neuHP peaks, in comparison to four other primates and in comparison to archaic hominin genomes (H. neanderthalensis and H. densi). (XLSX)

Table S13 List of 32 control peaks with the least/no differences in H3K4me3 levels between human and non-human primates, to determine human-specific sequence alterations and compare with 33 neuHP. (XLSX)

Table S14 Ancestral allele frequencies for subset of neuHP peaks (Pilot 1000 Genome Project). (XLSX)

Table S15 Overlap between HP and sequences subject to adaptive fixations (sweep) in modern populations. Overlap between HP and human accelerated sequences within domains highly conserved between rodent and primates. (XLSX)

Table S16 Polymorphism-based sweep analyses for neuHP peaks, using 1,000 genomes pilot data. (XLSX)

Table S17 Novel human-specific GATA-1 motif in DPP10-2 promoter sequence otherwise deeply conserved across mammalian lineages. (XLSX)

Table S18 RNAseq normalized tag densities from human, chimpanzee, and macaque PFC, for HP sequences and their surrounding 1–2 kb. (XLSX)

Text S1 Methods. (DOC)

Acknowledgments

We thank R. Kozz, E. Kittler, and Maria Zapp and the University of Massachusetts Medical School (UMMS) deep sequencing and flow cytometry cores for excellent technical support.

Author Contributions

The author(s) have made the following declarations about their contributions: Conceived and designed the experiments: ER ZW JDJ SA. Performed the experiments: HPS JLC DR JST CJP IBH RB ACM. Analyzed the data: HPS JLC DR JST IC RB H-JC IBH CJP ACM. Contributed reagents/materials/analysis tools: ZW JDJ SA W-DY TMP RHM J-fC. Wrote the paper: ER JDJ ZW SA.

References


