

Novel H3K4me3 marks are enriched at human- and chimpanzee-specific cytogenetic structures

Giuliana Giannuzzi,¹ Eugenia Migliavacca,^{1,2} and Alexandre Reymond¹

¹Center for Integrative Genomics, University of Lausanne, 1015 Lausanne, Switzerland; ²Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland

Human and chimpanzee genomes are 98.8% identical within comparable sequences. However, they differ structurally in nine pericentric inversions, one fusion that originated human chromosome 2, and content and localization of heterochromatin and lineage-specific segmental duplications. The possible functional consequences of these cytogenetic and structural differences are not fully understood and their possible involvement in speciation remains unclear. We show that subtelomeric regions—regions that have a species-specific organization, are more divergent in sequence, and are enriched in genes and recombination hotspots—are significantly enriched for species-specific histone modifications that decorate transcription start sites in different tissues in both human and chimpanzee. The human lineage-specific chromosome 2 fusion point and ancestral centromere locus as well as chromosome 1 and 18 pericentric inversion breakpoints showed enrichment of human-specific H3K4me3 peaks in the prefrontal cortex. Our results reveal an association between plastic regions and potential novel regulatory elements.

[Supplemental material is available for this article.]

Chromosomes, the DNA–protein structures that carry genetic information, undergo structural rearrangements, including fusion and fission events, as well as inversions, translocations, duplications, and deletions. Human and chimpanzee karyotypes differ by one chromosomal fusion that gave rise to human chromosome 2 (HSA2) from two ancestral chromosomes coupled to the inactivation of one of the two centromeres, at least nine pericentric inversions, and in the content of constitutive heterochromatin (Yunis et al. 1980; Ijdo et al. 1991; Baldini et al. 1993; Nickerson and Nelson 1998). Seven of these inversions, mapping to human chromosomes 4, 5, 9, 12, 15, 16, and 17, are specific to the chimpanzee lineage (Marzella et al. 2000; Kehrer-Sawatzki et al. 2002; Locke et al. 2003; Goidts et al. 2005; Kehrer-Sawatzki et al. 2005a,b,c; Shimada et al. 2005; Szamalek et al. 2005), while the remaining two, mapping to HSA1 and HSA18, appeared in the human lineage after separation from the chimpanzee (Yunis and Prakash 1982; McConkey 1997; Dennehey et al. 2004; Weise et al. 2005; Szamalek et al. 2006). These reorganized structures became fixed during evolution either by providing an advantage or by mere genetic drift.

Human subtelomeric regions, as well as pericentromeric ones, are hotspots of segmental duplications that were reshaped over recent evolutionary time (Horvath et al. 2000; Mefford and Trask 2002; She et al. 2004; Linardopoulou et al. 2005). Indeed, while human and chimpanzee genomes are 98.77% identical within comparable sequences, they show an increased divergence (15%) in the terminal 10 Mbp (millions of base pairs) of chromosomes (The Chimpanzee Sequencing and Analysis Consortium 2005). These highly plastic segments of the human genome show qualitative and quantitative differences in the distribution of segmental duplications when compared with the great apes, consistent with their recent origin and human-specific sequence transfers (Horvath et al. 2001; Bailey et al. 2002; Horvath et al. 2003; Linardopoulou et al. 2005; Locke et al. 2005). In addition, regions enriched in

segmental duplications are more prone to both interspecies and intraspecies structural variation (Newman et al. 2005; Sharp et al. 2005), since these repeated segments may mediate nonallelic homologous recombination (NAHR) (Hastings et al. 2009).

It is still unclear whether chromosomal rearrangements and structurally different loci played a role in the human–chimpanzee speciation. Indeed the hypothesis that they affected the rate of genetic divergence between humans and chimpanzees does not have enough support (Kehrer-Sawatzki and Cooper 2007). Previous studies revealed no evidence of accelerated evolution for genes on rearranged versus colinear chromosomes (Lu et al. 2003; Navarro and Barton 2003; Vallender and Lahn 2004; Zhang et al. 2004; The Chimpanzee Sequencing and Analysis Consortium 2005; Marques-Bonet et al. 2007) and showed that chromosomal rearrangements have generally no impact on gene expression except in a few particular cases (Munoz and Sankoff 2012). However, chromosomal rearrangements appear to be associated with higher divergence in gene-expression levels in the brain (Marques-Bonet et al. 2004) and genes located on rearranged chromosomes showed reduced recombination rate compared with colinear ones (Farré et al. 2013).

In this study we analyzed the chromosomal distribution of human- and chimpanzee-specific enrichment/depletion of H3K4me3 histone modifications in the prefrontal cortex (Shulha et al. 2012) and lymphoblastoid cell lines (LCLs) (Cain et al. 2011) and tested their accumulation at genomic regions with species-specific structure. H3K4me3 is an epigenetic mark broadly associated with RNA polymerase II occupancy at transcription start sites and RNA expression levels (Wang et al. 2008; The ENCODE Project Consortium 2012; Kilpinen et al. 2013). We detected a higher density of human- and chimpanzee-specific H3K4me3 peaks in subtelomeric regions both in the prefrontal cortex and LCLs. The

Corresponding authors: giuliana.giannuzzi@unil.ch, alexandre.reymond@unil.ch

Article published online before print. Article, supplemental material, and publication date are at <http://www.genome.org/cgi/doi/10.1101/gr.167742.113>.

© 2014 Giannuzzi et al. This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see <http://genome.cshlp.org/site/misc/terms.xhtml>). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

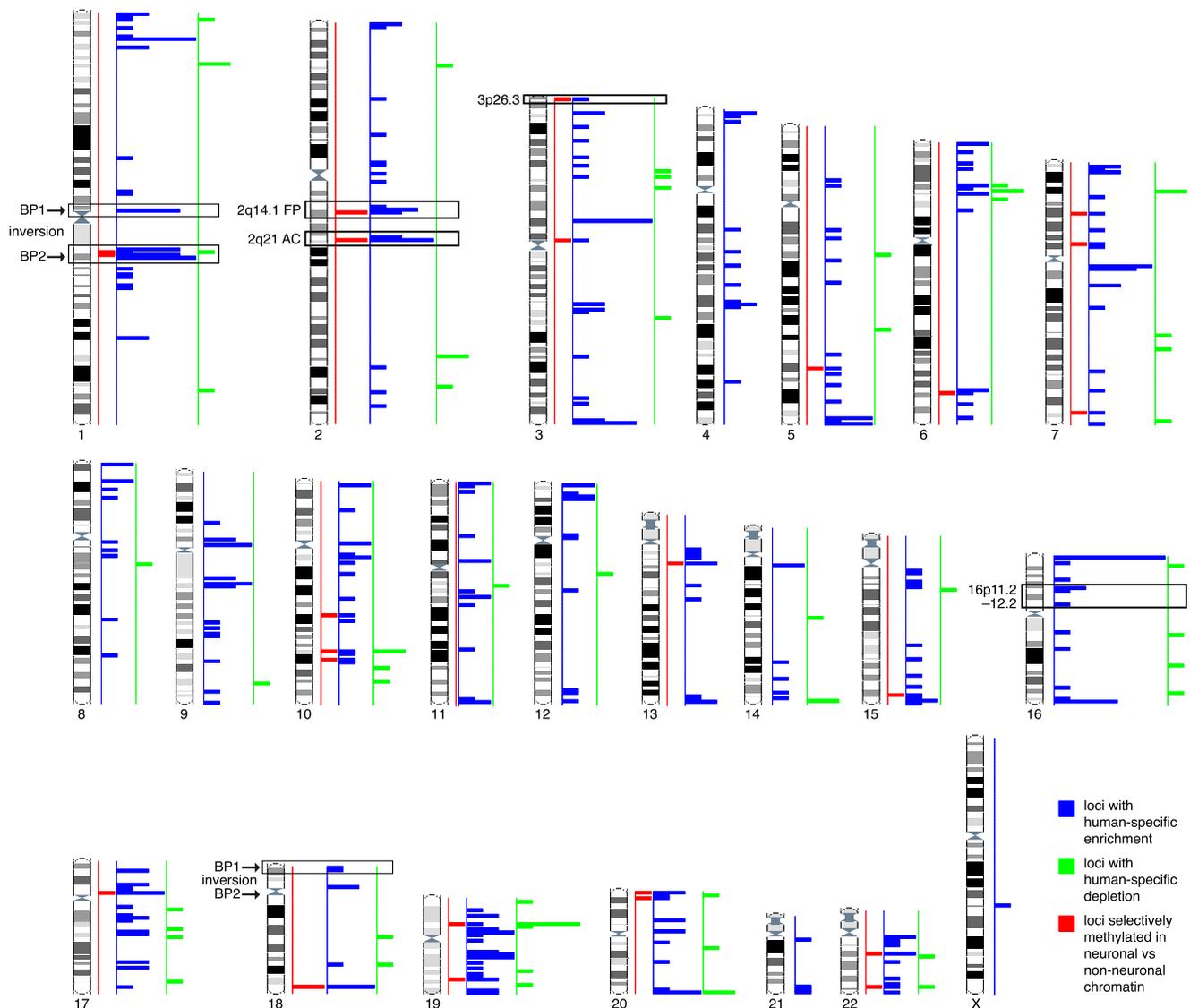


Figure 1. Karyotype-wide mapping of the regions with human-specific enrichment ($n = 410$, blue) and depletion ($n = 61$, green) of H3K4me3 modifications in prefrontal neurons. The localizations of the 33 human-specific loci selectively methylated in neuronal versus nonneuronal cells are similarly pinpointed in red ($n = 33$ regions). The human-specific pericentric inversion breakpoints (BP1 and BP2) of HSA1 and HSA18, the fusion point (FP) and ancestral centromere (AC) of HSA2, and the 16p11.2–12.2 and 3p26.3 regions mentioned in the text are indicated.

human prefrontal cortex similarly showed a higher density of species-specific H3K4me3 marks at other human-specific genomic structures. Our results provide evidence for a possible functional and regulatory role in human and chimpanzee evolution of recently acquired structural and chromosomal differences.

Results

A recent analysis compared the genome-wide profiles of H3K4me3 histone modifications in the prefrontal cortex of human, chimpanzee, and macaque (Shulha et al. 2012). It identified 471 loci with significant changes in histone modification rates in human (enrichment, 410; loss, 61) when compared with the two non-human primate species. Additionally, the investigators detected 33 human specifically enriched loci that were selectively methylated in neuronal versus nonneuronal cells (Shulha et al. 2012),

among which they pinpoint *DPP10* (chromosomal band 2q14.1), *CNTN4*, and *CHL1* (both at 3p26.3), three genes conferring susceptibility to neurological disease (Sakurai et al. 2002; Fernandez et al. 2004; Marshall et al. 2008; Glessner et al. 2009; Roohi et al. 2009; Salyakina et al. 2011). They describe and further analyze the *DPP10* and 16p11.2–12.2 loci, the latter among the 410 loci with human-specific enrichment in H3K4me3 modifications. They conclude that “coordinated epigenetic regulation via newly derived transcription start site chromatin could play an important role in the emergence of human-specific gene expression networks in brain” (Shulha et al. 2012). It is noteworthy that all four of the featured loci map to regions that were modified in recent human evolution. *CNTN4* and *CHL1* are within the subtelomeric region of the HSA3 short arm. The *DPP10* locus (2q14.1) maps only 1 Mbp away from the above-mentioned fusion point of HSA2 (Ijdo et al. 1991; Fan et al. 2002; The Chimpanzee Sequencing and Analysis Consortium

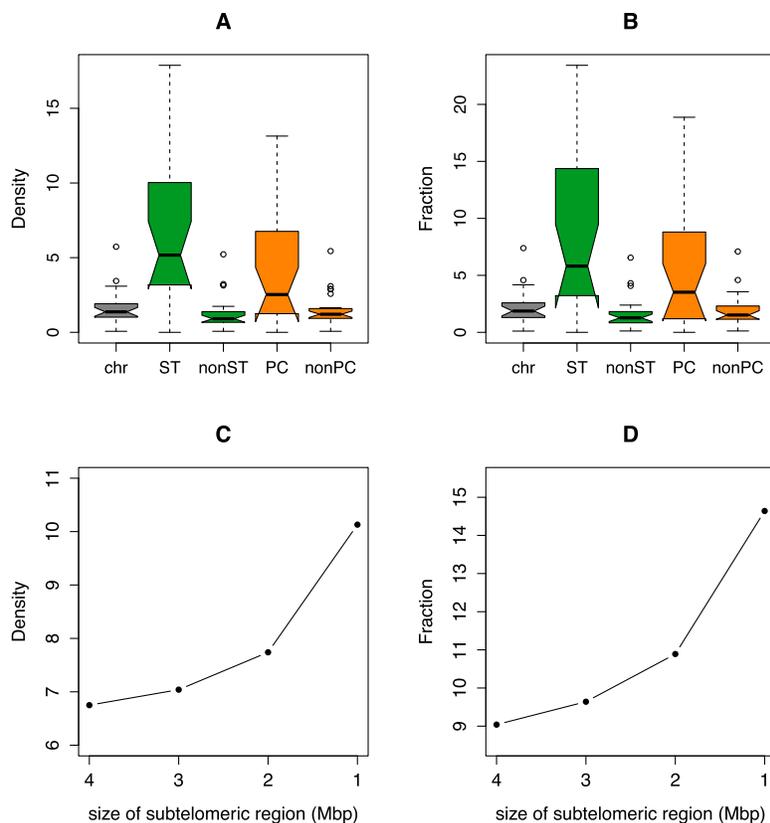


Figure 2. Distribution of the density (A) and fraction (B) of H3K4me3-enriched human-specific regions in all chromosome (chr), subtelomeric (ST), nonsubtelomeric (nonST), pericentromeric (PC), and nonpericentromeric (nonPC) regions. Note the significant concentration of these epigenetic decorations in the highly plastic subtelomeric and pericentromeric sections of the human genome (density, one-way χ^2 test, P -values = 4×10^{-73} and 3×10^{-15} , respectively; fraction, Fisher's exact test, P -values = 2×10^{-56} and 2×10^{-13} , respectively). See Table 1 for nomenclature. Comparison of density (C) and fraction (D) of human-specific H3K4me3 modifications in subtelomeric regions of different size (4, 3, 2, and 1 Mbp).

2005), while the 16p11.2–12.2 region has seen a rapid integration of segmental duplications in the last 15 million years of hominoid evolution that contributed to a profound modification of these chromosomal bands (Johnson et al. 2001; Antonacci et al. 2010), putting them at risk for recurrent pathogenic rearrangements (Fig. 1; Girirajan et al. 2010; Walters et al. 2010; Jacquemont et al. 2011). These observations suggest that we should consider the possibility that the differences reported in regulatory footprints (Shulha et al. 2012) might be associated with the specific genomic organization of these loci in human and/or with their localization to or closeness to highly plastic sections of the human genome.

The chromosome-wide distribution of the 410 H3K4me3 peaks with human-specific enrichment in prefrontal neurons is not uniform (Fig. 1). We assessed the possible association between regions with significantly different epigenetic panorama in humans when compared with other primates (Shulha et al. 2012) and genomic segments around loci that were structurally modified during the recent evolution of the human genome, i.e., HSA2 fusion point and ancestral centromere, as well as HSA1 and HSA18 inversion breakpoints (Yunis et al. 1980; Dennehey et al. 2004; Szamalek et al. 2006), with HSA1 also encompassing pericentromeric heterochromatin that is absent in its chimpanzee homolog (Yunis et al. 1980). Additionally, we assessed highly plastic segments

such as human-specific segmental duplications (Sudmant et al. 2013) together with subtelomeric and pericentromeric regions (Yunis et al. 1980; Bailey et al. 2001; Bailey et al. 2002; Horvath et al. 2003; The Chimpanzee Sequencing and Analysis Consortium 2005; Linardopoulou et al. 2005; Locke et al. 2005).

The human-specific prefrontal cortex H3K4me3-enriched sites significantly accumulate at subtelomeric (fivefold) and pericentromeric sites (threefold) in both the number of peaks (P -value = 4×10^{-73} and 3×10^{-15} , respectively; permutation P -value = 0.001) and the amount of base pairs covered (P -value = 2×10^{-56} and 2×10^{-13} , respectively; permutation P -value = 0.001) (Fig. 2A,B; Table 1). For example, 87% and 82% of the subtelomeric and pericentromeric regions of autosomes (34/39 and 18/22, respectively) contain at least one H3K4me3 human-enriched peak (Fig. 1). Interestingly, the density and fraction of these peaks show a sharp increase toward the chromosomal ends (Fig. 2C,D).

We detected a significant eightfold increase in both the density (P -value = 2×10^{-21} , permutation P -value = 0.001) and amount of base pairs covered (P -value = 7×10^{-11} , permutation P -value = 0.001) of human specifically enriched histone modification marks in human-specific segmental duplications (Sudmant et al. 2013). Albeit not significant, the segmental duplication-rich 16p11.2–12.2 cytogenetic bands showed a consistent twofold enrichment of these epigenetic marks when compared with the genome average.

Despite a certain overlap of highly plastic genome sections such as subtelomeric, pericentromeric, and duplicated regions with human lineage-specific rearrangement breakpoints (e.g., HSA1 p-arm inversion breakpoint overlaps with its pericentromeric region and HSA18 p-arm inversion breakpoint overlaps with its subtelomeric region), we independently assessed possible enrichment in the latter. The fusion point (chr2:113–116 Mbp) and ancestral centromere (2q21) loci of HSA2 overlap clusters with higher density (14- and sixfold, respectively, P -value = 0.001) and sequence coverage (15- and sixfold, P -value = 0.001 and 0.002, respectively) of H3K4me3 peaks when compared with the genome-wide average (Figs. 1, 3; Table 1). The HSA1 inversion breakpoints show 21-fold enrichment in both density and sequence coverage (P -value = 0.001), while the HSA18 inversion breakpoints show 10- (P -value = 0.006) and 18-fold (P -value = 0.001) higher density and sequence coverage than the genome average, respectively (Figs. 1, 3; Table 1). As a negative control, we tested chimpanzee-specific inversion breakpoints for which the human genome organization preserves the ancestral state (i.e., the ones mapping to chromosomes 4, 5, 9, 12, 16, and 17), and found no difference in either density or sequence coverage (P -values = 1) of H3K4me3 lineage-specific peaks compared with the genome average (Table 1).

HSA19 is the human chromosome with the highest gene density (Grimwood et al. 2004) and the highest density of

Table 1. Distribution of the 410 regions with human-specific enrichment of H3K4me3 modifications in prefrontal cortex^a (including and excluding segmental duplications and chromosome X from the counts)

	With segmental duplications									
	# Peaks	Density	Density enrichment	P-value	P-value perm.	Peak size	Fraction	Fraction enrichment	P-value	P-value perm.
Genome-wide	410	1.45								
Subtelomeric regions	108	6.75	4.67	4.19×10^{-73}	0.001	535,422	1.89	4.79	2.01×10^{-56}	0.001
Pericentromeric regions	61	3.80	2.63	3.05×10^{-15}	0.001	78,214	4.88	2.58	1.59×10^{-13}	0.001
Human-specific seg dups	15	11.64	8.05	2.49×10^{-21}	0.001	19,022	14.77	7.82	6.76×10^{-11}	0.001
Chimp-specific seg dups	4	9.02	6.24	1.35×10^{-4}	0.008	3,751	8.46	4.48	5.28×10^{-2}	0.032
16p11.2–12.2	4	3.02	2.09	Not tested	0.114	4,843	3.66	1.94	Not tested	0.126
Chr2:113–116 Mbp FP	6	20.00	13.83	Not tested	0.001	8,790	29.30	15.52	Not tested	0.001
2q21 AC	6	8.70	6.01	Not tested	0.001	7,619	11.04	5.85	Not tested	0.002
Chr1 inversion BPs	11	29.88	20.67	Not tested	0.001	14,379	39.06	20.69	Not tested	0.001
Chr18 inversion BPs	3	14.20	9.82	Not tested	0.006	7,248	34.32	18.18	Not tested	0.001
Chimpanzee inversion BPs ^b	5	2.11	1.46	1	0.245	4,821	2.03	1.08	1	0.399

(continued)

H3K4me3-specific peaks (fourfold enrichment when compared with the average results of autosomes) (Fig. 1). Although it has no large-scale structural differences when compared with its chimpanzee and macaque homologs (The Chimpanzee Sequencing and Analysis Consortium 2005; Rhesus Macaque Genome Sequencing and Analysis Consortium 2007), HSA19 is the autosome with the highest human–chimpanzee sequence divergence besides HSA21 (The Chimpanzee Sequencing and Analysis Consortium 2005) and one of the chromosomes with the highest segmental duplication density (Bailey et al. 2001; Bailey et al. 2002).

We observed similar results for the 33 H3K4me3 human-enriched regions selectively methylated in neuronal versus non-neuronal chromatin (Shulha et al. 2012) as 10 (fivefold enrichment) map to subtelomeric regions, two map to the ancestral fusion locus of HSA2 (see above), and two are at the ancestral centromere locus of the same chromosome (Fig. 1; Supplemental Table S1). Thus 51% of the methylated regions (17 out of 33) overlap at least one of the features. Conversely, the 61 human-specific prefrontal cortex H3K4me3-depleted sites show no enrichment at these regions (Fig. 1; Supplemental Table S2).

We evaluated whether enrichments of human-specific H3K4me3 sites at genome structures unique to humans were specific to the prefrontal cortex or also present in other tissues. We used H3K4me3 ChIP-seq data of human, chimpanzee, and macaque LCLs (Cain et al. 2011) to identify autosomal human specifically enriched or depleted peaks. We found significant fourfold enrichments of H3K4me3-enriched peaks at subtelomeric regions in LCLs (density P -value = 2×10^{-23} , permutation P -value = 0.001; fraction P -value = 4×10^{-19} , permutation P -value = 0.001), together with enrichment trends of density (sixfold) and fraction (fourfold) at the HSA2 fusion point and threefold enrichment at 16p11.2–12.2 (Fig. 4; Table 2; Supplemental Table S3). These results show that the enrichment of human-specific H3K4me3 sites at subtelomeric regions is not unique to brain structures. Similar to prefrontal cortex, the 109 autosomal regions with human-specific depletion of H3K4me3 marks

in LCLs showed no enrichment at these regions (Supplemental Table S4).

Next we assessed if concentrations of lineage-specific H3K4me3 sites at lineage-specific genome structures were unique to humans. Toward this goal we determined the chimpanzee-specific H3K4me3 peaks in LCLs ($n = 224$ enriched and $n = 36$ depleted) and used the published chimpanzee and nonhuman peaks of the prefrontal cortex ($n = 523$ enriched and $n = 327$ depleted) (Shulha et al. 2012). We identified significant accumulation at subtelomeric regions of H3K4me3 sites that are chimpanzee-enriched (fourfold enrichment) or chimpanzee-depleted (twofold) in the prefrontal cortex and of chimpanzee-enriched H3K4me3 sites in LCLs (twofold) (Fig. 4; Table 2; Supplemental Tables S5–S8). For example, the segment orthologous to the fusion point of chromosome 2, which has a subtelomeric location in the chimpanzee genome (Fig. 4C), showed a fivefold enrichment of density and fraction in the prefrontal cortex. The situation is less clear pericentromerically and at species' inversion breakpoints, because we detected no enrichment at these sites in chimpanzee (Fig. 4B; Table 2; Supplemental Tables S5, S6).

As we observe a general increase of species-specific H3K4me3 peaks in segmental duplications, we then asked whether these observed accumulations at structurally different regions were possibly an indirect effect of their high duplication content. Alternatively, we considered that part of the signal might arise from the erroneous inclusion of false positive species-specific peaks due to poor annotation of multicopy sequences in some species/regions (Pickrell et al. 2011). To this end, we repeated our enrichment analyses in the prefrontal cortex excluding duplicated regions (Table 1; Supplemental Tables S1, S2, S5–S7). Barring all duplicated regions of the genome did not abrogate the reported enrichment at subtelomeric regions for the two species (human and chimpanzee) studied and at the HSA2 fusion point. Conversely, removing these regions abolished the enrichment of human-specific prefrontal cortex H3K4me3 sites at the HSA2 ancestral centromere locus.

Table 1. *Continued*

	Without segmental duplications									
	# Peaks	Density	Density enrichment	P-value	P-value perm.	Peak size	Fraction	Fraction enrichment	P-value	P-value perm.
Genome-wide	295	1.15				380,760	1.49			
Subtelomeric regions	98	6.81	5.90	2.11×10^{-93}	0.001	133,324	9.27	6.22	4.05×10^{-67}	0.001
Pericentromeric regions	25	2.21	1.91	2.24×10^{-3}	0.06	26,899	2.37	1.59	0.05	0.288
Human-specific seg dups	—	—	—	—	—	—	—	—	—	—
Chimp-specific seg dups	—	—	—	—	—	—	—	—	—	—
16p11.2–12.2	0	0.00	0.00	Not tested	1	0	0.00	0.00	Not tested	1
Chr2:113–116 Mbp FP	4	15.87	13.74	Not tested	0.004	7,252	28.77	19.30	Not tested	0.001
2q21 AC	0	0.00	0.00	Not tested	1	0	0.00	0.00	Not tested	1
Chr1 inversion BPs	3	18.91	16.38	Not tested	0.003	2,408	15.18	10.18	Not tested	0.025
Chr18 inversion BPs	3	15.25	13.20	Not tested	0.008	7,248	36.84	24.71	Not tested	0.001
Chimpanzee inversion BPs ^b	5	2.29	1.98	0.35	0.272	4,821	2.21	1.48	0.80	0.394

“Density” specifies the number of enriched regions in 10 Mbp of sequence. “Fraction” delineates the number of nucleotides overlapped per 10 kbp. Enrichments are relative to the genome-wide values. (Seg dups) Segmental duplications; (FP) fusion point; (AC) ancestral centromere; (BPs) breakpoints. P-values are Bonferroni corrected.

^aHuman specifically enriched H3K4me3 histone modification peaks as defined in Shulha et al. (2012).

^bBreakpoints of chimpanzee lineage-specific inversions mapping to HSA4, 5, 9, 12, 15, 16, and 17 (see text for details).

To gain insights into the possible cause(s) of these enrichments, we then studied the features of both the affected genomic regions and the human-specific peaks they encompass after ensuring first that the observed increased densities could not be explained simply by the wide heterogeneity of the genome. We found that the distributions of the lineage-specific marks were not random (P -values $\leq 4 \times 10^{-6}$), except the human depleted one (P -value = 0.4) (Supplemental Table S9; Methods). In the genomic regions we investigated the concentration of repetitive elements, genes, protein-coding genes, and recombination hotspots (Table 3). The latter were associated with testis-specific trimethylation of H3K4 in mouse (Smagulova et al. 2011), while repetitive elements were shown to create novel regulatory elements (Feschotte 2008). We observed no correlation between repetitive elements and the increases of H3K4me3 marks. Similarly, gene density alone cannot explain the totality of the observed enrichments of species-specific H3K4me3 peaks, as its increase compared with genome average is narrower than that of the chromatin marks. For example, while subtelomeric regions display a 1.61-fold inflated gene density (twofold if considering coding genes), the corresponding increase of H3K4me3 lineage-specific peaks varies from two- to fivefold depending on tissue and species. As subtelomeric regions are both gene-richer and the cradle of an abundance of recombination hotspots, it is possible that these two characteristics together with the higher divergence of these portion vis-a-vis the chimpanzee genome explain the observed enrichments. Additionally, we computed the content in repetitive elements, assessed sequence similarity with the chimpanzee and overlap with mapped recombination hotspots of human-enriched peaks at subtelomeric regions, HSA2 fusion point and HSA1 and HSA18 inversion breakpoints (the latter in prefrontal cortex only), and compared them with those of the complete set of H3K4me3 human-enriched peaks. We found that the peaks at

subtelomeric regions and rearrangement break- and fusion points presented greater sequence divergence (measured both through identity percentage and aligned fraction) (Fig. 5). As these concomitantly show no enrichment/depletion in their repetitive element content, it suggests that the emergence of these lineage-specific peaks might be favored by sequence changes but that the latter are not due to lineage-specific inclusion of repeated elements. The higher percentage of specific peaks overlapping recombination hotspots at subtelomeric regions ($\sim 1.7\times$ higher than the genome average) is correlated to the higher density of recombination hotspots in these portions ($1.6\times$ higher) (Table 3), confirming that, similarly to gene density, recombination hotspots do not explain the observed enrichments of specific peaks at these segments entirely (Table 4).

Shulha and colleagues found that four of the 33 neuronal peaks (12%) were associated with novel RNA expression specific to human prefrontal cortex and 18 of the 410 human-enriched peaks (4%) with different RNA levels (Shulha et al. 2012). We similarly assessed how many human/chimpanzee enriched/depleted H3K4me3 LCL peaks corresponded to changes in expression of embedded exons. We uncovered an overlap between such exons and lineage-specific peaks ranging from 13% to 26% (Supplemental Table S10), suggesting that a noteworthy fraction of the human–chimpanzee variation in H3K4me3 peaks has an effect on RNA expression.

Discussion

We analyzed the location of human-specific and chimpanzee-specific H3K4me3 histone modification marks—a proxy for promoters/TSSs—in prefrontal neurons (Shulha et al. 2012) to assess if particular chromosomal portions and lineage-specific chromosomal rearrangements provide fertile ground for new regulatory

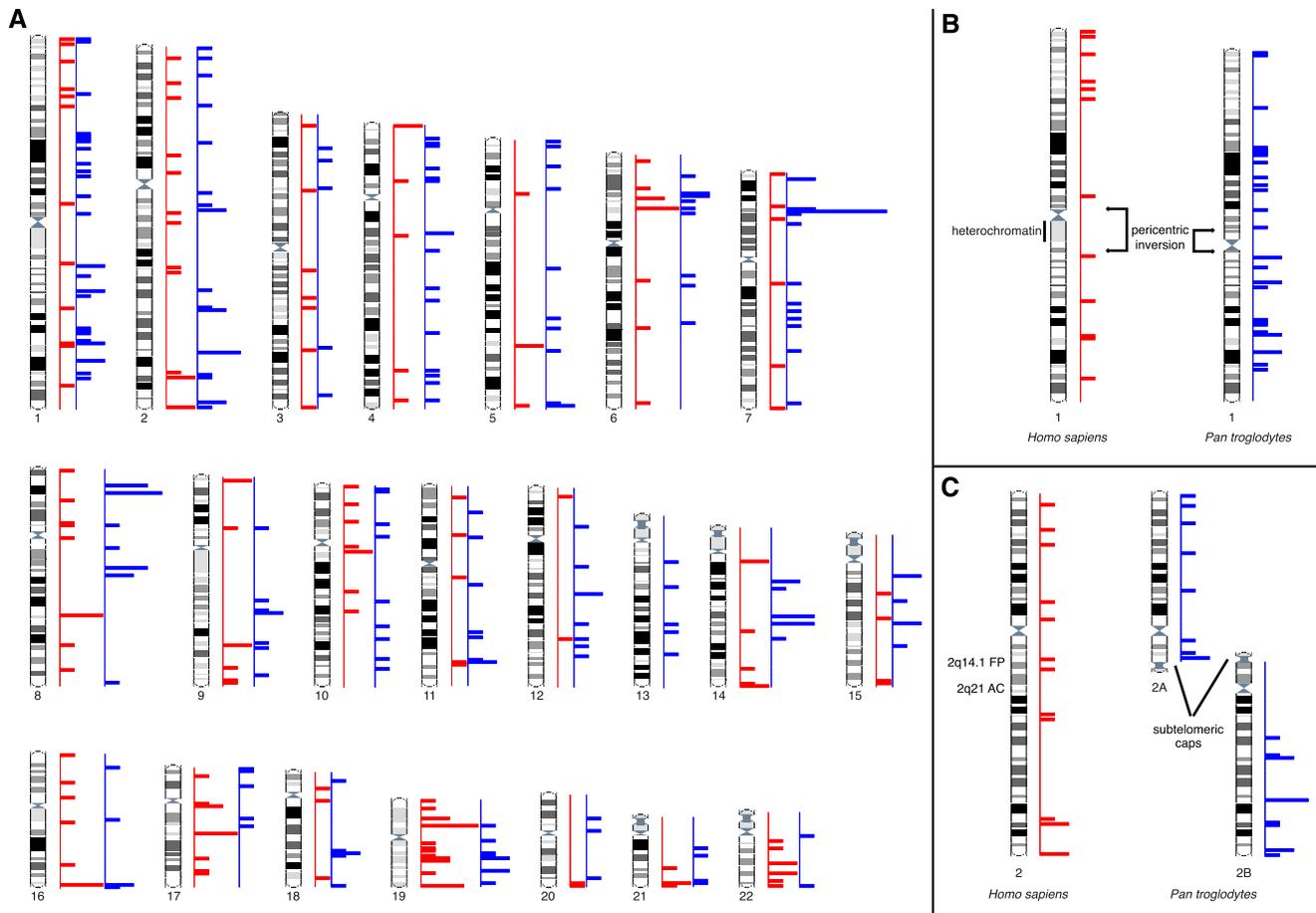


Figure 4. (A) Human karyotype-wide mapping of the regions with human-specific ($n = 164$, red) and chimpanzee-specific ($n = 224$, blue) enrichment of H3K4me3 modifications in LCLs. Chimpanzee-specific regions in panel A are positioned on the human karyotype; however, some chromosomes are structurally different between these species. For example, chromosome 1 differs because of a pericentric inversion and heterochromatin content (B); chromosome 2 differs because of a chromosomal fusion in human and the presence of subteleromic heterochromatic caps (C).

et al. 2012) overlaps with domains having new cytogenetic architecture generated by evolutionary chromosomal rearrangements or with rapidly evolving sites like subteleromic and pericentromeric regions and segmental duplications. For example, the majority of regions selectively methylated in neurons identified in Shulha et al. (2012) map to highly plastic chromosome regions and/or with a known human-specific organization (Fig. 1; Supplemental Table S1).

To understand whether this propensity exists in other tissue and species as well, we analyzed another cell type, LCLs, and another species, the chimpanzee. We found a consistent and general increase of diversity and novelty in H3K4me3 epigenetic marks at subteleromic regions in both human and chimpanzee and both prefrontal cortex and LCLs. In human prefrontal cortex, both the density and fraction of specifically enriched H3K4me3 sites sharply increase as we approach the chromosomal ends (Fig. 2C, D), suggesting that the conjunction of an increased divergence between the human and chimpanzee genome sequences, as we progress toward the telomere (The Chimpanzee Sequencing and Analysis Consortium 2005) and distinctive contents in recombination hotspots, heterochromatin, and duplicated regions at these sites, supports novelty in regulatory elements. The novel human epigenetic marks at the chromosome 2 fusion point might

also reflect its ancestral subteleromic location. Of note, increased epigenetic diversity at subteleromic regions was observed in two recent comparisons: the 5-hydroxymethylcytosine epigenetic mark between induced pluripotent stem cells (iPSCs) and embryonic stem cells (ESCs) in human (Wang et al. 2013) and higher order chromatin structure between human and mouse (Chambers et al. 2013). Moreover, an example of the emergence of novel promoters and expression modules through segmental duplications was described for the human core duplicon *LRRC37* (Bekpen et al. 2012; Giannuzzi et al. 2013).

The loss of enrichments at pericentromeric regions, and HSA2 ancestral centromere locus when excluding human prefrontal cortex peaks mapped to duplications, suggests that the higher concentration of species-specific H3K4me3 peaks at these sites may derive from the high content in duplicated sequences rather than their cytogenetic localization. Similarly, the lack of enrichment of chimpanzee-specific peaks at chimpanzee inversion breakpoints suggests that the human concentration at these sites may not be associated with the structural change per se but rather with their subteleromic/pericentromeric localization or presence of duplicated blocks. Of note, the correct identification of H3K4me3 peaks in duplicated regions might require separate efforts due to the misannotation of recent and highly similar segmental duplications in genomes.

Table 2. Summary of density and fraction enrichments of human and chimpanzee prefrontal cortex (PC) and LCLs in enriched H3K4me3-marked sites

	Human											
	PC						LCL					
	Density enrich.	<i>P</i> -value	<i>P</i> -value perm.	Fraction enrich.	<i>P</i> -value	<i>P</i> -value perm.	Density enrich.	<i>P</i> -value	<i>P</i> -value perm.	Fraction enrich.	<i>P</i> -value	<i>P</i> -value perm.
Subtelomeric regions	4.67	4.19×10^{-73}	0.001	4.79	2.01×10^{-56}	0.001	4.22	2.00×10^{-23}	0.001	4.09	3.63×10^{-19}	0.001
Pericentromeric regions	2.63	3.05×10^{-15}	0.001	2.58	1.59×10^{-13}	0.001	1.51	0.472	0.116	1.33	0.826	0.191
Human-specific seg dups	8.05	2.49×10^{-21}	0.001	7.82	6.76×10^{-11}	0.001	—	—	—	—	—	—
Chimp-specific seg dups	6.24	1.35×10^{-4}	0.008	4.48	5.28×10^{-2}	0.032	—	—	—	—	—	—
16p11.2–12.2	2.09	Not tested	0.114	1.94	Not tested	0.126	3.40	Not tested	0.116	3.39	Not tested	0.069
Chr2:113–116 Mbp FP	13.83	Not tested	0.001	15.52	Not tested	0.001	6.18	Not tested	0.149	4.19	Not tested	0.149
2q21 AC	6.01	Not tested	0.001	5.85	Not tested	0.002	0.00	Not tested	1	0.00	Not tested	1
Chr1 inversion BPs	20.67	Not tested	0.001	20.69	Not tested	0.001	0.00	Not tested	1	0.00	Not tested	1
Chr18 inversion BPs	9.82	Not tested	0.006	18.18	Not tested	0.001	0.00	Not tested	1	0.00	Not tested	1
Chimpanzee inversion BPs	1.46	1	0.245	1.08	1	0.399	1.43	1	0.409	1.12	1	0.394

(continued)

While the enrichment at subtelomeric regions was consistently seen across species and tissues, at pericentromeric sites enrichment was only detected in the human prefrontal cortex, not in human LCLs or in the chimpanzee. Is this human-specific enrichment in cortex truly singular or a mere reflection of the lower quality achieved by the chimpanzee compared with the human genome, especially within pericentromeric regions? Further studies are warranted to confirm/refute these differences.

Our results support the existence of an evolutionary role for chromosomal rearrangement loci and subtelomeric regions. These segments of the genome harbor new sequences, which arose both from increased divergence and species-specific organization, and/or more recombination events. The convergence of these features possibly allows chromatin reconfiguration and thus the appearance of novel H3K4me3 sites, partly associated with modification in gene expression. Our findings suggest that evolutionary novelties and neighboring sequences should be investigated not only for gene expression differences and as fertile ground for the emergence of novel genes and transcripts, but also in the quest for lineage-specific epigenetic and regulatory changes. Our results also indicate how the duplicated regions that border copy number variants could play a role in the modification of the expression of normal copy number flanking genes (Merla et al. 2006; Raymond et al. 2007; Henrichsen et al. 2009a,b), an effect that can extend over the entire length of the affected chromosome (Ricard et al. 2010).

Methods

Genomic regions

Coordinates refer to the human reference sequence hg19/GRCh37. Subtelomeric regions were defined as the first and the

last 4 Mbp of chromosome sequence (for acrocentric chromosomes only the last 4 Mbp were considered). Similarly, pericentromeric regions were the first 4 Mbp on either side (p and q chromosomal arms) of the centromere and, when present, the heterochromatin gaps, and the first 4 Mbp on the q-side for acrocentric chromosomes. Coordinates of human-specific and chimpanzee-specific (i.e., present only in human and chimpanzee, respectively) segmental duplications (both fixed duplications and expansions) (Sudmant et al. 2013) were converted from the hg18 to the hg19 release using the liftOver tool with default parameters. We retrieved coordinates of HSA1 inversion breakpoints from Szamalek et al. (2006); coordinates of HSA18, PTR4 (*Pan troglodytes*), PTR5, PTR12, PTR16, and PTR17 inversion breakpoints from the UCSC Genome Browser (Chiaromonte et al. 2002; Kent et al. 2002, 2003; Schwartz et al. 2003); coordinates of PTR9 inversion breakpoint on HSA9q21 from Kehrer-Sawatzki et al. (2005c); coordinates of PTR15 inversion breakpoint on HSA15q13 from Locke et al. (2003). A window spanning 1 Mbp upstream of and 1 Mbp downstream from the breakpoints was considered. We note that some of these features overlap: (1) HSA1 p-arm inversion breakpoint overlaps with its pericentromeric region; (2) HSA18 p-arm inversion breakpoint overlaps with its subtelomeric region; (3) HSA16 pericentromeric region overlaps with 16p11.2–12.2 cytogenetic bands; (4) human-specific segmental duplications overlap with all other features, i.e., human inversion breakpoints, HSA2 fusion point and ancestral centromere loci, 16p11.2–12.2, and subtelomeric and pericentromeric regions. Sequence gaps were excluded in all calculations. Chromosome Y and unplaced contigs were excluded from all analyses. The same regions with the exclusion of human and chimpanzee segmental duplications and chromosome X were considered as well. In the computation of enrichments for the chimpanzee, we redefined the coordinates of subtelomeric and pericentromeric regions considering the structural differences be-

Table 2. *Continued*

	Chimpanzee											
	PC						LCL					
	Density enrich.	<i>P</i> -value	<i>P</i> -value perm.	Fraction enrich.	<i>P</i> -value	<i>P</i> -value perm.	Density enrich.	<i>P</i> -value	<i>P</i> -value perm.	Fraction enrich.	<i>P</i> -value	<i>P</i> -value perm.
Subtelomeric regions	4.40	<i>8.74 × 10⁻⁸⁴</i>	<i>0.001</i>	4.40	<i>8.70 × 10⁻⁴⁹</i>	<i>0.001</i>	2.21	<i>1.78 × 10⁻⁵</i>	<i>0.001</i>	2.02	<i>1.56 × 10⁻⁴</i>	<i>0.003</i>
Pericentromeric regions	1.33	0.298	<i>0.042</i>	1.33	0.259	<i>0.039</i>	0.49	0.27	0.976	0.46	0.076	0.961
Human-specific seg dups	0.42	1	0.908	0.08	0.928	0.903	—	—	—	—	—	—
Chimp-specific seg dups	11.00	<i>7.01 × 10⁻¹⁹</i>	<i>0.001</i>	10.76	<i>1.10 × 10⁻⁶</i>	<i>0.001</i>	—	—	—	—	—	—
16p11.2–12.2	2.05	Not tested	0.105	1.91	Not tested	0.111	0.00	Not tested	1	0.00	Not tested	1
Chr2:113–116 Mbp FP	5.42	Not tested	<i>0.016</i>	4.93	Not tested	<i>0.025</i>	0.00	Not tested	1	0.00	Not tested	1
2q21 AC	0.79	Not tested	0.732	0.64	Not tested	0.574	0.00	Not tested	1	0.00	Not tested	1
Chr1 inversion BPs	0.00	Not tested	1	0.00	Not tested	1	0.00	Not tested	1	0.00	Not tested	1
Chr18 inversion BPs	5.13	Not tested	0.059	7.91	Not tested	<i>0.009</i>	0.00	Not tested	1	0.00	Not tested	1
Chimpanzee inversion BPs	1.37	1	0.276	1.40	1	0.211	1.57	1	0.283	1.61	0.651	0.186

See Table 1 and Supplemental Tables S3, S5, S6 for details. *P*-values < 0.05 are in italics.

tween human and chimpanzee chromosomes, i.e., chromosomal fusion and pericentric inversions.

Locations of recombination hotspots from Phase II HapMap data (release 21) (McVean et al. 2004; The International HapMap Consortium 2005; Winckler et al. 2005) were converted from hg17 to hg19 using liftOver (minMatch = 0.9). Repeat annotation for the GRCh37/hg19 human genome release was downloaded from the UCSC Genome Browser. Gene annotation refers to Ensembl v74 (Flicek et al. 2014).

H3K4me3 species-specific peaks

We obtained coordinates of 410 and 61 H3K4me3 peaks with human-specific enrichment/depletion, respectively, in prefrontal cortex, 33 human-specific neuronal H3K4me3 peaks, and 551 and 337 H3K4me3 peaks with chimpanzee-specific enrichment/depletion, respectively, in prefrontal cortex from Shulha et al. (2012).

We identified autosomal H3K4me3 peaks in LCLs using ChIP-seq data from Cain et al. (2011). We mapped human, chimpanzee, and macaque reads to the human genome (GRCh37/hg19) using Bowtie (version 0.12.9) (Langmead et al. 2009) and called human and chimpanzee peaks using MACS (Zhang et al. 2008). We filtered peaks with FDR < 0.1, >500 bp, mapped on autosomes, and not overlapping gaps or human/chimpanzee duplications (Sudmant et al. 2013). We defined enriched regions 500 bp around the peak summit as suggested in Bardet et al. (2012) and merged the human- and chimpanzee-enriched regions. We extended mapped reads of 100 bp and counted the coverage for the enriched regions. We identified the regions with FDR < 0.01 and at least twofold enriched or depleted in human and chimpanzee using limma (Law et al. 2014).

The coordinates of chimpanzee-specific (enriched and depleted) regions in the prefrontal cortex (Supplemental Tables S7, S8

of Shulha et al. 2012) were converted from panTro2 to GRCh37/hg19 using the liftOver tool (minMatch = 0.7). We manually checked through BLAT regions with a size difference >20% of the original size in the chimpanzee genome. This procedure allowed the conversion of 523 out of 551 and 327 out of 337 prefrontal cortex chimpanzee-enriched/depleted peaks.

We computed the intersection between peaks and genomic regions, recombination hotspots, repetitive elements, and genes using BEDTools (Quinlan and Hall 2010). To estimate the divergence of human peaks versus chimpanzee, we aligned the peak sequences to the chimpanzee genome (panTro4) using BLAT with default parameters. For each peak, we considered the best alignment and analyzed the percent identity and ratio between the alignment length and the size of the peak sequence. Boxplots and density curves were drawn using the R package (R Development Core Team 2014).

RNA-seq data analysis

We aligned LCL RNA-seq reads from Cain et al. (2011) to the human genome (hg19) using TopHat2 (Kim et al. 2013), assembled transcripts using Cufflinks (Trapnell et al. 2010), and joined human and chimpanzee transcripts using Cuffmerge. We counted reads in the transcripts using two python scripts from Anders et al. (2012) and identified differentially expressed exons (FDR < 0.01 and at least twofold change) among those that overlap with H3K4me3 marked regions using limma (Law et al. 2014).

Statistical analysis

We assessed the statistical significance of the density and fraction enrichments using the one-way χ^2 and Fisher's exact tests, respectively, and permutation tests (Davison and Hinkley 1997) by

Table 3. Features of the genomic regions analyzed

	With segmental duplications						Without segmental duplications					
	Size	Recombination hotspots	Repetitive elements	Gene density	Gene density ratio	Protein-coding gene density ratio	Size	Recombination hotspots	Repetitive elements	Gene density	Gene density ratio	Protein-coding gene density ratio
Genome	2,835,679,040	7%	50%	20	1.61	7	2,554,554,480	7%	50%	21	1.58	9
Human subtelomeric regions	159,894,210	11%	44%	33	1.61	14	143,867,784	11%	43%	32	1.58	15
Human pericentromeric regions	160,315,523	4%	57%	29	1.43	6	113,266,521	5%	56%	28	1.39	9
Chimpanzee subtelomeric regions	167,712,210	10%	44%	32	1.59	13	149,346,596	10%	43%	32	1.56	15
Chimpanzee pericentromeric regions	167,265,523	4%	56%	30	1.48	7	117,462,680	5%	56%	29	1.43	10
Human-specific seg dups	12,881,225	3%	51%	72	3.56	20	—	—	—	—	—	—
Chimp-specific seg dups	4,436,370	6%	49%	85	4.19	44	—	—	—	—	—	—
16p11.2-12.2	13,250,000	6%	57%	41	2.04	14	9,172,569	8%	58%	40	1.95	18
Chr2:113-116 Mbp FP	3,000,000	6%	48%	26	1.27	9	2,520,822	6%	49%	27	1.30	12
2q21 AC	6,900,000	7%	50%	26	1.31	6	5,076,173	7%	50%	17	0.83	6
Chr1 inversion BPs	3,680,803	3%	49%	48	2.37	13	1,586,461	5%	55%	44	2.15	18
Chr18 inversion BPs	2,111,989	8%	54%	32	1.57	8	1,967,546	8%	53%	30	1.49	8
Chimpanzee inversion BPs	23,739,142	7%	52%	30	1.51	10	21,832,369	7%	52%	29	1.44	13

The total size, percentage of sequences belonging to recombination hotspots and repetitive elements, and number of genes and protein-coding genes in 1 Mbp are reported.

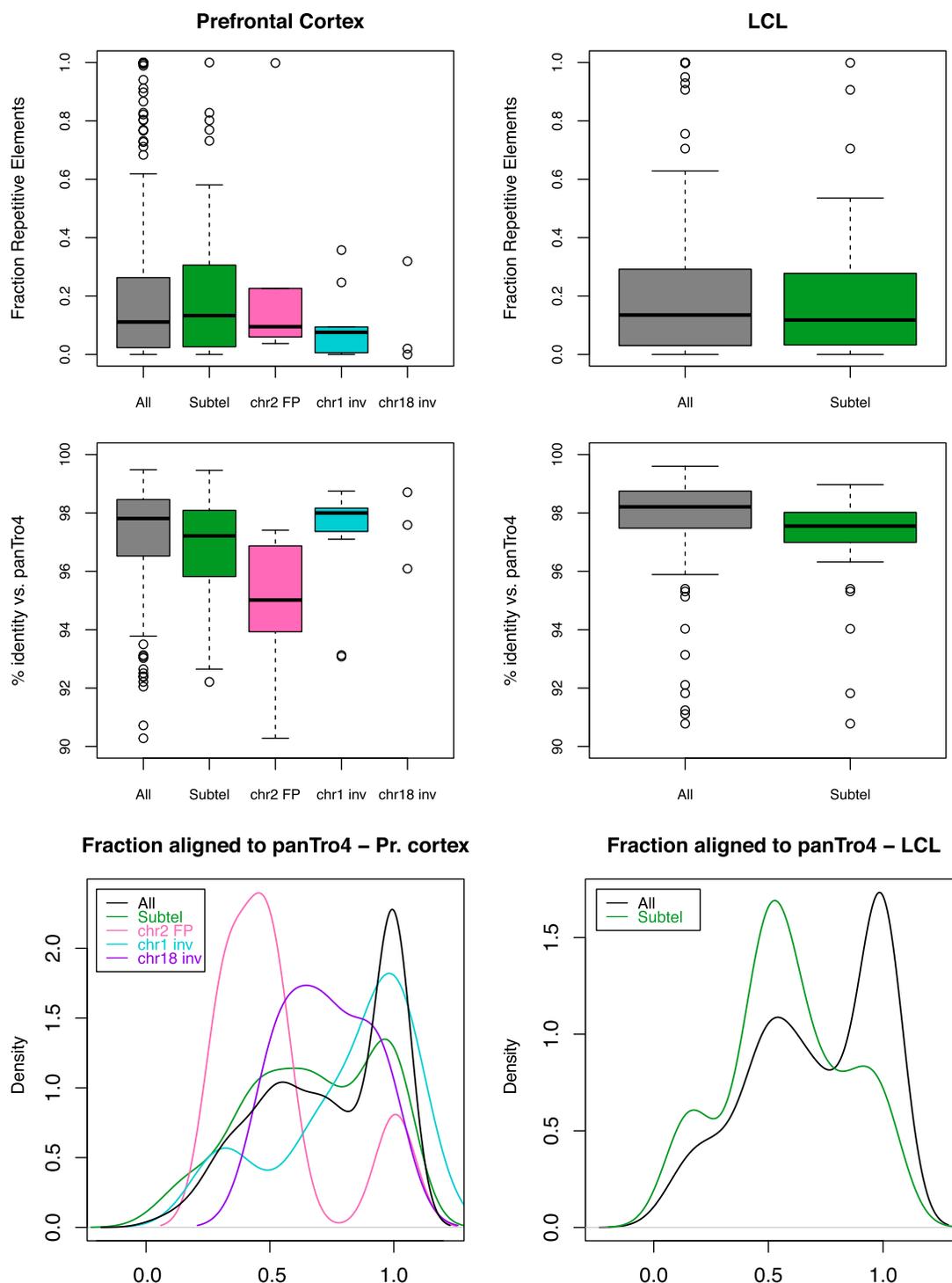


Figure 5. Human-enriched H3K4me3 peaks in prefrontal cortex (*left panels*) and LCLs (*right panels*), percentage of identity (*center panels*), and fraction aligned to the chimpanzee genome (panTro4) (*bottom panels*) of all human enriched peaks and peaks mapping to subtelomeric regions, HSA2 fusion points, HSA1 inversion breakpoints, and HSA18 inversion breakpoints (the rearranged points for prefrontal cortex only).

replacing peaks across the genome using shuffleBed 1000 times (Quinlan and Hall 2010). We adjusted P -values for multiple comparisons using the Bonferroni correction method. Permutation P -values were calculated by $P\text{-value} = (E + 1)/(R + 1)$, where R is the number of permutations, equal to 1000, and E is the number of

permutation test statistics that are greater than or equal to the observed test statistic.

The same enrichments were computed twice, including and excluding species-specific peaks mapped to duplicated sequences and chromosome X. To this end, human and chimpanzee dupli-

Table 4. Number and percentage of H3K4me3 human-enriched peaks in PC and LCLs mapped to a recombination hotspot

		Total	Mapped to recombination hotspots	Percentage mapped to recombination hotspots
Human PC enriched	All	410	49	12%
	Subtelomeric	108	23	21%
	chr2 FP	6	0	0%
	chr1 inv BP	11	0	0%
	chr18 inv BP	3	1	33%
Human LCL enriched	All	164	18	11%
	Subtelomeric	39	7	18%

cation data (both fixed duplications and expansions) (Sudmant et al. 2013) were converted from the hg18 to the hg19 release using the liftOver tool (-minMatch = 0.5 -minBlocks = 0.5).

We divided the human genome, excluding centromeric gaps, into 675 4-Mb segments and predicted the expected number of segments with different counts of marks using the Poisson distribution. We compared these predicted values with the observed ones (H3K4me3 human/chimpanzee-enriched/depleted peaks in prefrontal cortex and the 33 neuronal peaks) using a G-test and found that the human-enriched, chimpanzee-enriched, and chimpanzee-depleted peaks were not randomly distributed. We also performed permutation tests by replacing subtelomeric, pericentromeric, HSA2 fusion point, HSA2 ancestral centromere, HSA1 and HSA18 inversion breakpoint regions as well as chimpanzee inversion breakpoint regions across the genome using shuffleBed 1000 times (Quinlan and Hall 2010). The results of both approaches similarly showed that the lineage-specific peaks—except the human-depleted ones—were not randomly distributed along the genome.

Acknowledgments

We thank Maria Nicla Loviglio for discussions. This work was supported by the Swiss National Science Foundation (SNSF) and an SNSF Sinergia grant to A.R. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The computations were performed at the Vital-IT (<http://www.vital-it.ch>) Center for high-performance computing of the SIB Swiss Institute of Bioinformatics.

Author contributions: G.G. designed the study and conducted the analyses. G.G. and E.M. performed the statistical tests. G.G. and A.R. wrote the manuscript. All authors read and approved the final manuscript.

References

Anders S, Reyes A, Huber W. 2012. Detecting differential usage of exons from RNA-seq data. *Genome Res* **22**: 2008–2017.

Antonacci F, Kidd JM, Marques-Bonet T, Teague B, Ventura M, Girirajan S, Alkan C, Campbell CD, Vives L, Malig M, et al. 2010. A large and complex structural polymorphism at 16p12.1 underlies microdeletion disease risk. *Nat Genet* **42**: 745–750.

Bailey JA, Yavor AM, Massa HF, Trask BJ, Eichler EE. 2001. Segmental duplications: organization and impact within the current human genome project assembly. *Genome Res* **11**: 1005–1017.

Bailey JA, Gu Z, Clark RA, Reinert K, Samonte RV, Schwartz S, Adams MD, Myers EW, Li PW, Eichler EE. 2002. Recent segmental duplications in the human genome. *Science* **297**: 1003–1007.

Baldini A, Ried T, Shridhar V, Ogura K, D'Aiuto L, Rocchi M, Ward DC. 1993. An alphoid DNA sequence conserved in all human and great ape chromosomes: evidence for ancient centromeric sequences at human chromosomal regions 2q21 and 9q13. *Hum Genet* **90**: 577–583.

Bardet AF, He Q, Zeitlinger J, Stark A. 2012. A computational pipeline for comparative ChIP-seq analyses. *Nat Protoc* **7**: 45–61.

Bekpen C, Tastekin I, Siswara P, Akdis CA, Eichler EE. 2012. Primate segmental duplication creates novel promoters for the LRR37 gene family within the 17q21.31 inversion polymorphism region. *Genome Res* **22**: 1050–1058.

Cain CE, Blekhnman R, Marioni JC, Gilad Y. 2011. Gene expression differences among primates are associated with changes in a histone epigenetic modification. *Genetics* **187**: 1225–1234.

Chambers EV, Bickmore WA, Semple CA. 2013. Divergence of mammalian higher order chromatin structure is associated with developmental loci. *PLoS Comput Biol* **9**: e1003017.

Chiaromonte F, Yap VB, Miller W. 2002. Scoring pairwise genomic sequence alignments. *Pac Symp Biocomput* **2002**: 115–126.

The Chimpanzee Sequencing and Analysis Consortium. 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437**: 69–87.

Davison AC, Hinkley DV. 1997. *Bootstrap methods and their application*. Cambridge University Press, Cambridge, UK.

Dennehey BK, Gutches DG, McConkey EH, Krauter KS. 2004. Inversion, duplication, and changes in gene context are associated with human chromosome 18 evolution. *Genomics* **83**: 493–501.

The ENCODE Project Consortium. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**: 57–74.

Fan Y, Linaropoulou E, Friedman C, Williams E, Trask BJ. 2002. Genomic structure and evolution of the ancestral chromosome fusion site in 2q13-2q14.1 and paralogous regions on other human chromosomes. *Genome Res* **12**: 1651–1662.

Farré M, Micheletti D, Ruiz-Herrera A. 2013. Recombination rates and genomic shuffling in human and chimpanzee—a new twist in the chromosomal speciation theory. *Mol Biol Evol* **30**: 853–864.

Fernandez T, Morgan T, Davis N, Klin A, Morris A, Farhi A, Lifton RP, State MW. 2004. Disruption of *contactin 4* (*CNTN4*) results in developmental delay and other features of 3p deletion syndrome. *Am J Hum Genet* **74**: 1286–1293.

Feschotte C. 2008. Transposable elements and the evolution of regulatory networks. *Nat Rev Genet* **9**: 397–405.

Flicek P, Amode MR, Barrell D, Beal K, Billis K, Brent S, Carvalho-Silva D, Clapham P, Coates G, Fitzgerald S, et al. 2014. Ensembl 2014. *Nucleic Acids Res* **42**: D749–D755.

Giannuzzi G, Siswara P, Malig M, Marques-Bonet T, Mullikin JC, Ventura M, Eichler EE. 2013. Evolutionary dynamism of the primate LRR37 gene family. *Genome Res* **23**: 46–59.

Girirajan S, Rosenfeld JA, Cooper GM, Antonacci F, Siswara P, Itsara A, Vives L, Walsh T, McCarthy SE, Baker C, et al. 2010. A recurrent 16p12.1 microdeletion supports a two-hit model for severe developmental delay. *Nat Genet* **42**: 203–209.

Glessner JT, Wang K, Cai G, Korvatska O, Kim CE, Wood S, Zhang H, Estes A, Brune CW, Bradfield JP, et al. 2009. Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature* **459**: 569–573.

Goidts V, Szamalek JM, de Jong PJ, Cooper DN, Chuzhanova N, Hameister H, Kehrer-Sawatzki H. 2005. Independent intrachromosomal recombination events underlie the pericentric inversions of chimpanzee and gorilla chromosomes homologous to human chromosome 16. *Genome Res* **15**: 1232–1242.

Grimwood J, Gordon LA, Olsen A, Terry A, Schmutz J, Lamerdin J, Hellsten U, Goodstein D, Couronne O, Tran-Gyamfi M, et al. 2004. The DNA sequence and biology of human chromosome 19. *Nature* **428**: 529–535.

Hastings PJ, Lupski JR, Rosenberg SM, Ira G. 2009. Mechanisms of change in gene copy number. *Nat Rev Genet* **10**: 551–564.

Henrichsen CN, Chaignat E, Reymond A. 2009a. Copy number variants, diseases and gene expression. *Hum Mol Genet* **18**: R1–R8.

Henrichsen CN, Vinckenbosch N, Zollner S, Chaignat E, Pradervand S, Schutz F, Ruedi M, Kaessmann H, Reymond A. 2009b. Segmental copy number variation shapes tissue transcriptomes. *Nat Genet* **41**: 424–429.

Horvath JE, Schwartz S, Eichler EE. 2000. The mosaic structure of human pericentromeric DNA: a strategy for characterizing complex regions of the human genome. *Genome Res* **10**: 839–852.

Horvath JE, Bailey JA, Locke DP, Eichler EE. 2001. Lessons from the human genome: transitions between euchromatin and heterochromatin. *Hum Mol Genet* **10**: 2215–2223.

Horvath JE, Gulden CL, Bailey JA, Yohn C, McPherson JD, Prescott A, Roe BA, de Jong PJ, Ventura M, Misceo D, et al. 2003. Using a pericentromeric interspersed repeat to recapitulate the phylogeny and expansion of human centromeric segmental duplications. *Mol Biol Evol* **20**: 1463–1479.

Ijdo JW, Baldini A, Ward DC, Reenders ST, Wells RA. 1991. Origin of human chromosome 2: an ancestral telomere-telomere fusion. *Proc Natl Acad Sci* **88**: 9051–9055.

The International HapMap Consortium. 2005. A haplotype map of the human genome. *Nature* **437**: 1299–1320.

- Jacquemont S, Reymond A, Zufferey F, Harewood L, Walters RG, Kutalik Z, Martinet D, Shen Y, Valsesia A, Beckmann ND, et al. 2011. Mirror extreme BMI phenotypes associated with gene dosage at the chromosome 16p11.2 locus. *Nature* **478**: 97–102.
- Johnson ME, Viggiano L, Bailey JA, Abdul-Rauf M, Goodwin G, Rocchi M, Eichler EE. 2001. Positive selection of a gene family during the emergence of humans and African apes. *Nature* **413**: 514–519.
- Kehrer-Sawatzki H, Cooper DN. 2007. Understanding the recent evolution of the human genome: insights from human-chimpanzee genome comparisons. *Hum Mutat* **28**: 99–130.
- Kehrer-Sawatzki H, Schreiner B, Tanzer S, Platzer M, Muller S, Hameister H. 2002. Molecular characterization of the pericentric inversion that causes differences between chimpanzee chromosome 19 and human chromosome 17. *Am J Hum Genet* **71**: 375–388.
- Kehrer-Sawatzki H, Sandig C, Chuzhanova N, Goidts V, Szamalek JM, Tanzer S, Muller S, Platzer M, Cooper DN, Hameister H. 2005a. Breakpoint analysis of the pericentric inversion distinguishing human chromosome 4 from the homologous chromosome in the chimpanzee (*Pan troglodytes*). *Hum Mutat* **25**: 45–55.
- Kehrer-Sawatzki H, Sandig C, Goidts V, Hameister H. 2005b. Breakpoint analysis of the pericentric inversion between chimpanzee chromosome 10 and the homologous chromosome 12 in humans. *Cytogenet Genome Res* **108**: 91–97.
- Kehrer-Sawatzki H, Szamalek JM, Tanzer S, Platzer M, Hameister H. 2005c. Molecular characterization of the pericentric inversion of chimpanzee chromosome 11 homologous to human chromosome 9. *Genomics* **85**: 542–550.
- Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. 2002. The human genome browser at UCSC. *Genome Res* **12**: 996–1006.
- Kent WJ, Baertsch R, Hinrichs A, Miller W, Haussler D. 2003. Evolution's cauldron: duplication, deletion, and rearrangement in the mouse and human genomes. *Proc Natl Acad Sci* **100**: 11484–11489.
- Kilpinen H, Waszak SM, Gschwind AR, Raghav SK, Witwicki RM, Orioli A, Migliavacca E, Wiederkehr M, Gutierrez-Arcelus M, Panousis N, et al. 2013. Coordinated effects of sequence variation on DNA binding, chromatin structure, and transcription. *Science* **342**: 744–747.
- Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **14**: R36.
- Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**: R25.
- Law CW, Chen Y, Shi W, Smyth GK. 2014. Voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol* **15**: R29.
- Linardopoulou EV, Williams EM, Fan Y, Friedman C, Young JM, Trask BJ. 2005. Human subtelomeres are hot spots of interchromosomal recombination and segmental duplication. *Nature* **437**: 94–100.
- Locke DP, Archidiacono N, Misceo D, Cardone MF, Deschamps S, Roe B, Rocchi M, Eichler EE. 2003. Refinement of a chimpanzee pericentric inversion breakpoint to a segmental duplication cluster. *Genome Biol* **4**: R50.
- Locke DP, Jiang Z, Pertz LM, Misceo D, Archidiacono N, Eichler EE. 2005. Molecular evolution of the human chromosome 15 pericentromeric region. *Cytogenet Genome Res* **108**: 73–82.
- Lu J, Li WH, Wu CI. 2003. Comment on "Chromosomal speciation and molecular divergence-accelerated evolution in rearranged chromosomes". *Science* **302**: 988.
- Marques-Bonet T, Caceres M, Bertranpetit J, Preuss TM, Thomas JW, Navarro A. 2004. Chromosomal rearrangements and the genomic distribution of gene-expression divergence in humans and chimpanzees. *Trends Genet* **20**: 524–529.
- Marques-Bonet T, Sanchez-Ruiz J, Armengol L, Khaja R, Bertranpetit J, Lopez-Bigas N, Rocchi M, Gazave E, Navarro A. 2007. On the association between chromosomal rearrangements and genic evolution in humans and chimpanzees. *Genome Biol* **8**: R230.
- Marshall CR, Noor A, Vincent JB, Lionel AC, Feuk L, Skaug J, Shago M, Moessner R, Pinto D, Ren Y, et al. 2008. Structural variation of chromosomes in autism spectrum disorder. *Am J Hum Genet* **82**: 477–488.
- Marzella R, Viggiano L, Miolla V, Storlazzi CT, Ricco A, Gentile E, Roberto R, Surace C, Fratello A, Mancini M, et al. 2000. Molecular cytogenetic resources for chromosome 4 and comparative analysis of phylogenetic chromosome IV in great apes. *Genomics* **63**: 307–313.
- McConkey EH. 1997. The origin of human chromosome 18 from a human/ape ancestor. *Cytogenet Cell Genet* **76**: 189–191.
- McVean GA, Myers SR, Hunt S, Deloukas P, Bentley DR, Donnelly P. 2004. The fine-scale structure of recombination rate variation in the human genome. *Science* **304**: 581–584.
- Mefford HC, Trask BJ. 2002. The complex structure and dynamic evolution of human subtelomeres. *Nat Rev Genet* **3**: 91–102.
- Merla G, Howald C, Henrichsen CN, Lyle R, Wyss C, Zobot MT, Antonarakis SE, Reymond A. 2006. Submicroscopic deletion in patients with Williams-Beuren syndrome influences expression levels of the nonhemizygous flanking genes. *Am J Hum Genet* **79**: 332–341.
- Munoz A, Sankoff D. 2012. Detection of gene expression changes at chromosomal rearrangement breakpoints in evolution. *BMC Bioinformatics (Suppl 3)* **13**: S6.
- Navarro A, Barton NH. 2003. Chromosomal speciation and molecular divergence-accelerated evolution in rearranged chromosomes. *Science* **300**: 321–324.
- Newman TL, Tuzun E, Morrison VA, Hayden KE, Ventura M, McGrath SD, Rocchi M, Eichler EE. 2005. A genome-wide survey of structural variation between human and chimpanzee. *Genome Res* **15**: 1344–1356.
- Nickerson E, Nelson DL. 1998. Molecular definition of pericentric inversion breakpoints occurring during the evolution of humans and chimpanzees. *Genomics* **50**: 368–372.
- Pickrell JK, Gaffney DJ, Gilad Y, Pritchard JK. 2011. False positive peaks in ChIP-seq and other sequencing-based functional assays caused by unannotated high copy number regions. *Bioinformatics* **27**: 2144–2146.
- Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841–842.
- R Development Core Team. 2014. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>.
- Reymond A, Henrichsen CN, Harewood L, Merla G. 2007. Side effects of genome structural changes. *Curr Opin Genet Dev* **17**: 381–386.
- Rhesus Macaque Genome Sequencing and Analysis Consortium. 2007. Evolutionary and biomedical insights from the rhesus macaque genome. *Science* **316**: 222–234.
- Ricard G, Molina J, Chrast J, Gu W, Gheldof N, Pradervand S, Schutz F, Young JI, Lupski JR, Reymond A, et al. 2010. Phenotypic consequences of copy number variation: insights from Smith-Magenis and Potocki-Lupski syndrome mouse models. *PLoS Biol* **8**: e1000543.
- Roohi J, Montagna C, Tegay DH, Palmer LE, DeVincent C, Pomeroy JC, Christian SL, Nowak N, Hatchwell E. 2009. Disruption of contactin 4 in three subjects with autism spectrum disorder. *J Med Genet* **46**: 176–182.
- Sakurai K, Migita O, Toru M, Arinami T. 2002. An association between a missense polymorphism in the close homologue of L1 (*CHL1*, *CALL*) gene and schizophrenia. *Mol Psychiatry* **7**: 412–415.
- Salyakina D, Cukier HN, Lee JM, Sacharow S, Nations LD, Ma D, Jaworski JM, Konidari I, Whitehead PL, Wright HH, et al. 2011. Copy number variants in extended autism spectrum disorder families reveal candidates potentially involved in autism risk. *PLoS ONE* **6**: e26049.
- Schwartz S, Kent WJ, Smit A, Zhang Z, Baertsch R, Hardison RC, Haussler D, Miller W. 2003. Human-mouse alignments with BLASTZ. *Genome Res* **13**: 103–107.
- Sharp AJ, Locke DP, McGrath SD, Cheng Z, Bailey JA, Vallente RU, Pertz LM, Clark RA, Schwartz S, Seagraves R, et al. 2005. Segmental duplications and copy-number variation in the human genome. *Am J Hum Genet* **77**: 78–88.
- She X, Horvath JE, Jiang Z, Liu G, Furey TS, Christ L, Clark R, Graves T, Gulden CL, Alkan C, et al. 2004. The structure and evolution of centromeric transition regions within the human genome. *Nature* **430**: 857–864.
- Shimada MK, Kim CG, Kitano T, Ferrell RE, Kohara Y, Saitou N. 2005. Nucleotide sequence comparison of a chromosome rearrangement on human chromosome 12 and the corresponding ape chromosomes. *Cytogenet Genome Res* **108**: 83–90.
- Shulha HP, Crisci JL, Reshetov D, Tushir JS, Cheung I, Bharadwaj R, Chou HJ, Houston IB, Peter CJ, Mitchell AC, et al. 2012. Human-specific histone methylation signatures at transcription start sites in prefrontal neurons. *PLoS Biol* **10**: e1001427.
- Smagulova F, Gregoretti IV, Brick K, Khil P, Camerini-Otero RD, Petukhova GV. 2011. Genome-wide analysis reveals novel molecular features of mouse recombination hotspots. *Nature* **472**: 375–378.
- Sudmant PH, Huddleston J, Catacchio CR, Malig M, Hillier LW, Baker C, Mohajeri K, Kondova I, Bontrop RE, Persengiev S, et al. 2013. Evolution and diversity of copy number variation in the great ape lineage. *Genome Res* **23**: 1373–1382.
- Szamalek JM, Goidts V, Chuzhanova N, Hameister H, Cooper DN, Kehrer-Sawatzki H. 2005. Molecular characterisation of the pericentric inversion that distinguishes human chromosome 5 from the homologous chimpanzee chromosome. *Hum Genet* **117**: 168–176.
- Szamalek JM, Goidts V, Cooper DN, Hameister H, Kehrer-Sawatzki H. 2006. Characterization of the human lineage-specific pericentric inversion that distinguishes human chromosome 1 from the homologous chromosomes of the great apes. *Hum Genet* **120**: 126–138.
- Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010. Transcript assembly and

- quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28**: 511–515.
- Vallender EJ, Lahn BT. 2004. Effects of chromosomal rearrangements on human-chimpanzee molecular evolution. *Genomics* **84**: 757–761.
- Walters RG, Jacquemont S, Valsesia A, de Smith AJ, Martinet D, Andersson J, Falchi M, Chen F, Andrieux J, Lobbens S, et al. 2010. A new highly penetrant form of obesity due to deletions on chromosome 16p11.2. *Nature* **463**: 671–675.
- Wang Z, Zang C, Rosenfeld JA, Schones DE, Barski A, Cuddapah S, Cui K, Roh TY, Peng W, Zhang MQ, et al. 2008. Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat Genet* **40**: 897–903.
- Wang T, Wu H, Li Y, Szulwach KE, Lin L, Li X, Chen IP, Goldlust IS, Chamberlain SJ, Dodd A, et al. 2013. Subtelomeric hotspots of aberrant 5-hydroxymethylcytosine-mediated epigenetic modifications during reprogramming to pluripotency. *Nat Cell Biol* **15**: 700–711.
- Weise A, Starke H, Mrasek K, Claussen U, Liehr T. 2005. New insights into the evolution of chromosome 1. *Cytogenet Genome Res* **108**: 217–222.
- Winckler W, Myers SR, Richter DJ, Onofrio RC, McDonald GJ, Bontrop RE, McVean GA, Gabriel SB, Reich D, Donnelly P, et al. 2005. Comparison of fine-scale recombination rates in humans and chimpanzees. *Science* **308**: 107–111.
- Yunis JJ, Prakash O. 1982. The origin of man: a chromosomal pictorial legacy. *Science* **215**: 1525–1530.
- Yunis JJ, Sawyer JR, Dunham K. 1980. The striking resemblance of high-resolution G-banded chromosomes of man and chimpanzee. *Science* **208**: 1145–1148.
- Zhang J, Wang X, Podlaha O. 2004. Testing the chromosomal speciation hypothesis for humans and chimpanzees. *Genome Res* **14**: 845–851.
- Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, et al. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**: R137.

Received October 1, 2013; accepted in revised form June 9, 2014.