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.

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 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,d, 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 Mb (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 (Farre 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
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 nonhuman 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 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.
Species-specific structure and chromatin marks

![Figure 2](image_url)

**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 highly plastic subtelomeric and pericentromeric sections of the human genome (density, one-way \(P = 4 \times 10^{-73}\) and \(3 \times 10^{-15}\) respectively; fraction, Fisher’s exact test, \(P = 2 \times 10^{-56}\) and \(2 \times 10^{-13}\) 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).

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, \(P = 0.001\) and sequence coverage (15- and fivefold, \(P = 0.001\) 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 (\(P = 0.001\)), while the HSA18 inversion breakpoints show 10-fold \(P = 0.006\) and 18-fold \(P = 0.001\) 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 (\(P = 1\) 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 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 of recurrent pathogenic rearrangements (Fig. 1; 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 (\(P = 2 \times 10^{-21}\), permutation \(P = 0.001\)) and amount of base pairs covered (\(P = 7 \times 10^{-13}\), permutation \(P = 0.001\)) 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, \(P = 0.001\) and sequence coverage (15- and sixfold, \(P = 0.001\) 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 (\(P = 0.001\)), while the HSA18 inversion breakpoints show 10-fold \(P = 0.006\) and 18-fold \(P = 0.001\) 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 (\(P = 1\) of H3K4me3 lineage-specific peaks compared with the genome average (Table 1).
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.
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 \( \leq 4 \times 10^{-5} \)), 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 peaks—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

---

**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 subtelomeric regions | 295 | 1.15 | 5.90 | 2.11 \( \times 10^{-93} \) | 0.001 | 380,760 | 1.49 | 133,324 | 9.27 | 6.22 | 4.05 \( \times 10^{-67} \) | 0.001 |
| Pericentromeric regions | 98 | 6.81 | 9.00 | 2.29 \( \times 10^{-6} \) | 0.001 | 26,054 | 2.93 | 15,810 | 2.87 | 1.53 | 2.55 \( \times 10^{-3} \) | 0.001 |
| Human-specific seg dup. Chr1 inversion BPs | 25 | 2.21 | 1.91 | 2.24 \( \times 10^{-3} \) | 0.06 | 26,899 | 2.37 | 15,180 | 1.59 | 0.05 | 0.288 |
| Chimpanzee inversion BPs | — | — | — | — | — | — | — | — | — | — | — |
| 16p11.2–12.2 Mb FP Chr2:113–116 | 0 | 0.00 | 0.00 | Not tested | 1 | 0 | 0.00 | 0.00 | Not tested | 1 |
| 2q21 AC Chr1 inversion | 3 | 18.91 | 17.38 | Not tested | 1 | 0 | 0.00 | 0.00 | Not tested | 1 |
| 18p inversion Chr18 inversion BPs | 3 | 15.25 | 13.20 | Not tested | 0.08 | 7,248 | 36.84 | 24.71 | Not tested | 0.001 |
| Chimpanzee inversion BPs | 5 | 2.29 | 1.98 | 0.35 | 0.272 | 4,821 | 28.77 | 19.30 | Not tested | 0.001 |

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

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

Breakpoints of chimpanzee lineage-specific inversions mapping to HSA4, 5, 9, 12, 15, 16, and 17 (see text for details).
elements. Our analyses stemmed from the observation that the DPP10 locus, featured in Shulha et al. (2012) because it is characterized by both human-specific and neuronal-specific epigenetic marks in the prefrontal cortex, maps 1 Mb apart from the HSA2 fusion site. Besides gauging enrichment of human specifically enriched marks at the HSA2 fusion point, we also evaluated the number of human-specific H3K4me3-marked sites at other human lineage-specific genomic structures such as the HSA2 ancestral centromere locus as well as the HSA1 and HSA18 pericentric inversion breakpoints. Additionally, we considered segments of the human genome known to be structurally different between apes and humans, like subtelomeric and pericentromeric intervals as well as segmental duplications.

Our results suggest that a significant fraction of the newly acquired human topological domains characterized by lineage-specific epigenetic decorations in the prefrontal cortex (Shulha
et al. 2012) overlaps with domains having new cytogenetic architecture generated by evolutionary chromosomal rearrangements or with rapidly evolving sites like subtelomeric 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 subtelomeric 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 subtelomeric location. Of note, increased epigenetic diversity at subtelomeric 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 duplcon 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 subtelomeric/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.
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; Reymond 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. Summary of density and fraction enrichments of human and chimpanzee prefrontal cortex (PC) and LCLs in enriched H3K4me3-marked sites

| Region                        | PC Density enrich. | PC P-value | PC P-value perm. | PC Fraction enrich. | PC P-value | PC P-value perm. | LCL Density enrich. | LCL P-value | LCL P-value perm. | LCL Fraction enrich. | LCL P-value | LCL P-value perm. |
|-------------------------------|--------------------|------------|------------------|---------------------|------------|------------------|--------------------|-------------|-------------------|----------------------|-------------|-------------------|
| Subtelomeric regions          | 4.67               | 4.19 × 10^-23 | 0.001            | 4.79                | 2.01 × 10^-16 | 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 dup        | 8.05               | 2.49 × 10^-21 | 0.001            | 7.82                | 6.76 × 10^-11 | 0.001            | —                 | —          | —                 | —                    | —           | —                 |
| Chimpanzee seg dup            | 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                   | 13.83              | Not tested   | Not tested       | 15.52               | Not tested   | 0.001            | 6.18               | Not tested | 0.149             | 4.19                 | Not tested | 0.149             |
| Mbp FP                         | 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)
**Table 2. Continued**

|                  | PC                        | Chimpanzee                  | LCL                        |
|------------------|---------------------------|-----------------------------|----------------------------|
|                  | Density enrich. | P-value | P-value perm. | Density enrich. | P-value | P-value perm. | Density enrich. | P-value | P-value perm. | Density enrich. | P-value | P-value perm. |
| Subtelomeric regions | 4.40 | $8.74 \times 10^{-84}$ | 0.001 | 4.40 | $8.70 \times 10^{-49}$ | 0.001 | 2.21 | $1.78 \times 10^{-3}$ | 0.001 | 2.02 | $1.56 \times 10^{-4}$ | 0.003 |
| Pericentromeric regions | 1.33 | 0.298 | 0.042 | 1.33 | 0.259 | 0.039 | 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 | — | — | — | — | — | — |
| Chimpanzee-specific seg dups | 11.00 | $7.01 \times 10^{-19}$ | 0.001 | 10.76 | $1.10 \times 10^{-6}$ | 0.001 | — | — | — | — | — | — |
| 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 | 0.016 | 4.93 | Not tested | 0.025 | 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 | 0.009 | 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.

Between 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, 53 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 peaks 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 $\chi^2$ 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 | Protein-coding gene density ratio |
| Genome                  | 2,835,679,040               | 7%                      | 50%                        | 20           | 1.61                | 7                           | 1.94                           |
| Human                   | 159,894,210                 | 11%                     | 44%                        | 33           | 1.43                | 6                           | 0.91                           |
| subtelomeric regions    | 160,315,523                 | 4%                      | 57%                        | 29           | 1.59                | 13                          | 1.88                           |
| Human                   | 167,712,210                 | 10%                     | 44%                        | 32           | 1.48                | 7                           | 1.00                           |
| pericentromeric regions | 167,265,523                 | 4%                      | 56%                        | 30           | 1.48                | 7                           | 1.00                           |
| Chimpanzee              |                             |                         |                             |              |                     |                             |                                |
| subtelomeric regions    |                             |                         |                             |              |                     |                             |                                |
|                       |                             |                         |                             |              |                     |                             |                                |
| Human-specific seg dup  | 12,881,225                  | 3%                      | 51%                        | 72           | 3.56                | 20                          | 2.78                           |
|                       |                             |                         |                             |              |                     |                             |                                |
| Chim-specific seg dup   | 4,436,370                   | 6%                      | 49%                        | 85           | 4.19                | 44                          | 6.18                           |
| 16p11.2–12.2            | 13,250,000                  | 6%                      | 57%                        | 41           | 2.04                | 14                          | 1.93                           |
| Chr2:113–116           | 3,000,000                   | 6%                      | 48%                        | 26           | 1.27                | 9                           | 1.26                           |
| Mb p FP                 |                             |                         |                             |              |                     |                             |                                |
| 2q21 AC                 | 6,900,000                   | 7%                      | 50%                        | 26           | 1.31                | 6                           | 0.89                           |
| Chr inversion BPs       | 3,680,803                   | 3%                      | 49%                        | 48           | 2.37                | 13                          | 1.83                           |
| Chr1 inversion BPs      | 2,111,989                   | 8%                      | 54%                        | 32           | 1.57                | 8                           | 1.13                           |
| Chr18 inversion BPs     | 23,739,142                  | 7%                      | 52%                        | 30           | 1.51                | 10                          | 1.41                           |

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.
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 $P$-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-
randomly distributed along the genome. Lineage-specific peaks—except the human-depleted ones—were not randomly distributed along the genome. We also performed permutation tests by replacing subtelomeric, pericentromeric, HSA2 ancestral centromere, HSA1 and HSA18 inversion breakpoint regions as well as chimpanzee inversion breakpoint point, HSA2 ancestral centromere, HSA1 and HSA18 inversion breakpoint regions as well as chimpanzee inversion breakpoint regions across the genome using shuffled 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 Nicola 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 I, 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 JW, Eichler EE. 2002. Recent segmental duplications in the human genome. Science 297: 1003–1007.
Baldini A, Ried T, Shridhar V, Ogura K, D’Aulio L, Rocchi M, Ward DC. 1993. An aphil 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, Stawara P, Akgis CA, Eichler EE. 2012. Primate segmental duplication creates novel promoters for the LRRC37 gene family within the 17g21.31 inversion polymorphism region. Genome Res 22: 1050–1058.
Cain CE, Bleikham 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 I, Yap VB, Miller M. 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.
Dennehy 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, Linasopoulou 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.
Farre M, Micheletti D, Ruiz-Herrera A. 2013. Recombination rates and genome 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: 397–407.
Flick P, Amode MR, Barrell D, Beal K, Bills 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 LRRC37 gene family. Genome Res 23: 46–59.
Girirajan S, Rosenfeld JA, Cooper GM, Antonacci F, Siswara P, Issara 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.
Gissner 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 ubiquitinn and neuronal genes. Nature 459: 569–573.
Girirajan S, 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.
Gittwood J, Gordon LA, Ohlen A, Terry A, Schmutz J, Lamedri 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, Chaingat E, Reymond A. 2009a. Copy number variants, diseases and gene expression. Hum Mol Genet 18: R1–R8.
Henrichsen CN, Vinckenbosch N, Zollner S, Chiangfat 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 centromeric 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: 2215–2224.
Horvath JE, Gulden CI, Bailey JA, Yohn C, McPherson JD, Prescott A, Roe BA, de Jong PJ, Ventura M, Micoseo D, et al. 2003. Using a pericentromeric interspersed repeat to recapitulate the phylogeny and expansion of human centromeric segmental duplications. Mol Biol Evol 20: 1465–1479.
Ijdo JW, Baldini A, Ward DC, Reeder ST, Wells RA. 1991. Origin of human chromosome 2: an ancestral telomere-telomere fusion. Proc Natl Acad Sci USA 88: 9051–9055.
The International HapMap Consortium. 2005. A haplotype map of the human genome. Nature 437: 1299–1320.

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

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

1466 Genome Research
www.genome.org
Jacquemont S, Reymond A, Zufferey F, Harewood L, Walters RG, Kutalik Z, Martinet D, Shen Y, Valsecchia A, Beckmann ND, et al. 2011. Mirror extreme BML phenotypes associated with gene dosage at the chromosome 16p11.2 locus. Nature 478: 97–102.

Johnson ME, Bailey JA, Abril 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.

Keher-Sawatzki H, Cooper DN. 2007. Understanding the recent evolution of the human genome: insights from human-chimpanzee genome comparisons. Hum Mutat 28: 99–130.

Keher-Sawatzki H, Schreiner B, Tanzer S, Platzer M, Muller S, Hameister H. 2002. Molecular characterization of the pericentric inversion that causes differences in chromosome 19 and human chromosome 17. Am J Hum Genet 71: 375–388.

Keher-Sawatzki H, Sandig C, Chuzhanova N, Goulds 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.

Keher-Sawatzki H, Sandig CA, Goulds 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.

Keher-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, Sugen CW, Furye TS, Roskin KM, Fringle 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: 11844–11849.

Kilpinen H, Waszak SM, Gschwind AR, Raghat SK, Witwicki RM, Orioli A, Miglavacca F, Wiederkem Guterres-Arcules 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 transcripts in the presence of insertions, deletions and gene fusions. Genome Biol 14: R36.

Lingmand 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. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. BMC Bioinformatics (Suppl 3) 13: S6.

Marshall CR, Noor A, Vincent JB, Lionel AC, Feuk L, Skaug J, Shago M, Marques-Bonet T, Sanchez-Ruiz J, Armengol L, Khaja R, Bertranpetit J, Lockman DP, Jiang Z, Pertz LM, Misceo D, Archidiacono N, Eichler EE. 2005. A genome-wide survey of structural variation between human and chimpanzee. Genome Res 15: 1354–1356.

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 JM. 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, H恩chnsen CN, Harewood L, Merla G. 2007. Side effects of genome structural changes. Curr Opin Genet Dev 17: 381–386.

Rheas Macaque Genome Sequencing and Analysis Consortium. 2007. Evolutionary and biomedical insights from the rhesus macaque genome. Science 316: 222–234.

Ricard G, Molina J, Carad 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, Teyg DH, Palmer LE, DeVincent C, Pomeroy JC, Christian SL, Nowak N, Hatchwell E. 2009. Disruption of contactin 3 in three subjects with autism spectrum disorder. J Med Genet 46: 176–182.

Sakurai C, Migtia O, Toru M, Azimami T. 2002. An association between a missense polymorphism in the close homologue of L1 (CHL1, CAL1) gene and schizophrenia. Mol Psychiatry 7: 412–415.

Salyakina D, Cukier HN, Lee JM, Sacharov S, Nations LD, Ma D, Jaworski JM, Konidari I, Whitehead PJ, 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 K, Kent WJ, Smith A, Zhang Z, Baertsch R, Hardison RC, Haussler D, Miller W. 2003. Human-mouse alignments with BLASTZ. Genome Res 13: 103–107.

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. Hum Genet 117: 189–191.

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, Majic K, Levy S, Eichler EE, Handsaker B, Korneliussen TS, Marth GT, Tarragona, et al. 2015. Characterization of the human lineage-specific pericentric inversion that distinguishes human chromosome 1 from the homologous chimpanzee chromosome. Hum Genet 133: 1373–1382.

Szamalek JM, Golds V, Chuzhanova N, Hameister H, Cooper DN, Keher-Sawatzki H. 2005. Molecular characterization of the pericentric inversion that distinguishes human chromosome 5 from the homologous chimpanzee chromosome. Hum Genet 117: 168–176.

Szamalek JM, Hameister H, Kondrashov SA, Kondrashov AS, Salzberg SL, Eichler EE. 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.

Taylor S, Williams BA, Pertea G, McVean 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, Andreix 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.

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