ARTICLE

A hominid-specific shift in cerebellar expression, upstream retrotransposons, and a potential *cis*-regulatory mechanism: bioinformatics analyses of the mu-opioid receptor gene

Orna Levran [1] · Eran Even-Tov[2] · Li Zhao[2]

Abstract

The mu-opioid receptors (MOR, *OPRM1*) mediate the effects of beta-endorphin and modulate many biological functions including reward processing and addiction. The present study aimed to use bioinformatics to determine *OPRM1* brain expression profiles in higher primates and to look for regulatory mechanisms. We used the same computational pipeline to analyze publicly available expression data from postmortem brain regions across humans, chimpanzees, and rhesus macaques. The most intriguing finding was high *OPRM1* cerebellar expression in humans and chimpanzees and low expression in macaques. Together with previous reports of low cerebellar *OPRM1* expression in mice, this suggests an evolutionary shift in the expression profiles. Bioinformatic analysis of the *OPRM1* upstream region revealed a functional CTCF-binding region that evolved from tandem insertions of retrotransposons L1P1 and L1PA1 upstream (−60 kb) of *OPRM1*. The insertions arose in different time points after the split of small apes from great apes, and their combined sequence is unique. Furthermore, the derived G allele of SNP rs12191876, in the inserted region, is associated with an increased *OPRM1* expression in the cerebellum of postmortem human brains ($p = 4.7e–5$). The derived G allele became the major allele (60–90%) in the populations represented in the 1000 Genomes Project and may be beneficial. This study provides a foundation for building new knowledge about evolutionary differences in *OPRM1* brain expression. Further investigations are needed to elucidate the role of the inserted region and its SNPs in *OPRM1* expression, and to assess the biological function and relevance of *OPRM1* expression in the cerebellum.

Introduction

Evolutionary shifts in the regulation of brain gene expression of neuropeptides and their receptors are thought to contribute to the evolution of behavior. It is of interest to identify these changes and the genomic mechanisms that have the potential to drive them (e.g., regulatory elements).

✉ Orna Levran
  levrano@rockefeller.edu

1 Laboratory on the Biology of Addictive Diseases, The Rockefeller University, New York, NY, USA

2 Laboratory of Evolutionary Genetics and Genomics, The Rockefeller University, New York, NY, USA

Comparisons across primates can further elucidate the biological basis of behavioral evolution.

The basis of gene expression regulation is an important biological question for understanding phenotypic and behavioral evolution. How a *cis*-regulatory element plays a role in the expression regulation has been an exciting topic in human comparative and population biology. Transposon elements are a major component of mammalian genomes. Recent studies suggest that retrotransposon expansion plays an important role in the regulatory binding site evolution and regulation (Schmidt et al. 2012). Many of the newly evolved regulatory elements or binding sites are enriched in the young transposon elements (Trizzino et al. 2017). Transposable elements were shown to acquire new physiological functions as regulatory elements (Ito et al. 2017).

This type of study is now possible thanks to the comprehensive data from different sources that are publicly available, including the RNA-Seq data from postmortem human (GTEx) (Lonsdale et al. 2013) and nonhuman primates (NCBI project PRJNA236446) (Sousa et al. 2017),

the complete genomic sequences from these species (UCSC Genome Browser), and the chromatin immunoprecipitation (ChIP) with massively parallel DNA sequencing (ChIP-Seq) (ENCODE, Roadmap Epigenomics and Blueprint) that are integrated in Ensembl.

Opioid peptides and receptors are expressed throughout the central and peripheral nervous system (Valentino and Volkow 2018). The mu-opioid receptor (MOR) has an essential role in mediating natural rewards, addiction, and pain perception. The MOR gene (OPRM1) has been identified in many species, ranging from nonmammalian vertebrates to humans, and was shown to undergo extensive alternative splicing (Pasternak and Pan 2013). OPRM1 is composed of several alternative exons and undergoes extensive alternative splicing that creates isoforms with unique characteristics (Regan et al. 2016). OPRM1 is flanked by a 1-Mb upstream gene desert. An alternatively-spliced exon in the 5′ upstream region (~28–30 kb from the canonical first exon) was identified in rodents and humans (Xu et al. 2009). An extended OPRM1 locus with additional regulatory regions was also suggested, based on haplotype analysis (Levran et al. 2011).

The present study aimed to explore the brain expression of OPRM1 across humans, chimpanzees, and rhesus macaques and to identify and characterize functional cis-regulatory elements in the OPRM1 upstream region that were formed relatively recently in the course of human evolution.

Methods

Differential gene expression analysis

Humans

RNA-Seq data from postmortem human brains that are part of the Genotype-Tissue Expression (GTEx) project (Lonsdale et al. 2013) were obtained from dbGaP (study accession number phs000424.v3.p1). Donor criteria and processes are described in the GTEx project website. Brains were sampled at the University of Miami Brain Endowment Bank. All samples with RNA Integrity Number of 6.0 or higher qualified for RNA Sequence analysis. The time between death and tissue collection was <24 h. Briefly, we downloaded the RNA-Seq data of 1426 RNA-Seq libraries from 13 regions (11 unique regions) of the human brain. The 13 regions are spinal cord (cervical c-1), nucleus accumbens (basal ganglia), frontal cortex (BA9), anterior cingulate cortex (BA24), cerebellar hemisphere, hippocampus, amygdala, cerebellum, hypothalamus, caudate (basal ganglia), putamen (basal ganglia), substantia nigra, and cortex. We aligned each of the RNA-Seq datasets to the

human genome (GRCh38/hg38) using HISAT2 (Kim et al. 2015) with default parameters. After that, we used StringTie (Pertea et al. 2015) (parameter: -e -A -B -p 8 -G) and Ballgown (Pertea et al. 2016) to estimate the expression of each gene and each transcript.

Chimpanzees and rhesus macaques

RNA-Seq data from chimpanzees and rhesus macaques were obtained from the NCBI project PRJNA236446 (Sousa et al. 2017). The primate specimens were described in detail elsewhere (Sousa et al. 2017). Common chimpanzee (Pan troglodytes) and rhesus macaque (Macaca mulatta) brain samples were collected postmortem from five adult specimens each. Brain specimens were obtained from the Alamogordo Primate Facility (Holloman Air Force Base, New Mexico, USA). All chimpanzees suffered sudden death with no prolonged agonal state.

Fifteen brain tissues were analyzed: ventrolateral prefrontal cortex, primary visual (V1) cortex, striatum, superior temporal cortex, primary somatosensory (S1) cortex, orbital prefrontal cortex, medial prefrontal cortex, mediodorsal nucleus of the thalamus, primary motor (M1) cortex, inferior temporal cortex, posterior inferior parietal cortex, hippocampus, cerebellar cortex, amygdala, and primary auditory (A1) cortex. In total, there were 78 chimpanzee datasets and 77 macaque datasets. Similar to human data, we aligned each of the RNA-Seq datasets to their genomes (macaque genome Ensembl version Mmul_8.0.1; chimpanzee genome Ensembl version Pan_tro_3.0) using HISAT2 (Kim et al. 2015) with default parameters. After that, we used StringTie (Pertea et al. 2015) (parameter: -e -A -B -p 8 -G) and Ballgown (Pertea et al. 2016) to estimate the expression of each gene and each transcript.

Genomic analysis

Repeat analysis was conducted by RepeatMasker (http://www.repeatmasker.org) through the UCSC Genome Browser. We found that CTCF ChIP-Seq results suggest that there are CTCF binding motifs located within the L1P1 site (http://hgdownload.soe.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeAwgTfbsUniform/wgEncodeAwgTfbsUtaH1hescCtcfUniPk.narrowPeak.gz). To independently verify this, we performed CTCF site prediction using MEME Suite (version 4.10.0; FIMO), and the reference human CTCF site MA0139.1 from JASPAR (http://jaspar.genereg.net/matrix/MA0139.1/). In addition, we used CTCFBSDB 2.0 (http://insulatordb.uthsc.edu/search_new.php) to verify the motif-finding results.

Regulatory features were predicted by the Ensembl regulatory build (integration of data from ENCODE, Roadmap Epigenomics and Blueprint projects) (Zerbino et al. 2015).

Comparative genomic analysis was performed using the University of California, Santa Cruz (UCSC) Genome Browser (http://genome.ucsc.edu/) and was based on the MULTIZ program. BLAST searches were performed using Ensembl. Linkage disequilibrium and phased haplotype data were obtained from Ensembl (Genomes Project Consortium et al. 2015). The sample populations used were CEU (Utah residents with Northern and Western European ancestry from the CEPH collection) and YRI (Yoruba in Ibadan, Nigeria).

Results

OPRM1 mRNA expression

Publicly available *OPRM1* mRNA expression data from postmortem brain regions across humans, chimpanzees (*Pan troglodytes*, pt), and rhesus macaques (*Macaca mulatta*, mm) were analyzed. The analyses were performed for total mRNA expression and individual isoforms (alternatively spliced transcripts).

RNA-Seq data from 11 brain regions in 173 human subjects were obtained from the GTEx project. A very low *OPRM1* expression was found in most brain regions analyzed, and the only region with relatively high expression was the cerebellum (CB) (Fig. 1a). *OPRM1* expression in chimpanzees (ENSPTRG00000018723) and rhesus macaques (ENSMMUG00000022691) was analyzed in 15 brain regions from five brains each. The highest expression in macaque was in the mediodorsal nucleus of the thalamus (MD) followed by the amygdala and the striatum (Fig. 1b). The pattern of expression was different in chimpanzees, with the highest expression in the MD followed by the CB (Fig. 1c). Notably, in macaques only, there was low cerebellar *OPRM1* expression compared with the other regions.

Cerebellum OPRM1 isoforms (alternatively spliced transcripts)

In humans, the main isoforms expressed in the cerebellum are the canonical MOR-1 (*OPRM1*-202; ENST00000330432; NM_000914) and MOR-1G1 (*OPRM1*-219, ENST00000522555; NM_001285526.1; the transcript variant is also described as MOR-1K2) (Fig. 2 and Supplementary Fig. S1). In addition, there are two untranslated transcripts: *OPRM1*-213 (ENST00000519613) and *OPRM1*-221 (ENST00000523520). Exons 2 and 3 (8 and 9 in Ensembl), which transcribe all but the first transmembrane (TM) domain, as well as the ligand-binding site, are included in all main isoforms. Isoform MOR-1G1 is a 6-TM protein that utilizes an alternate promoter and an alternative start codon in a unique downstream exon (exon 7
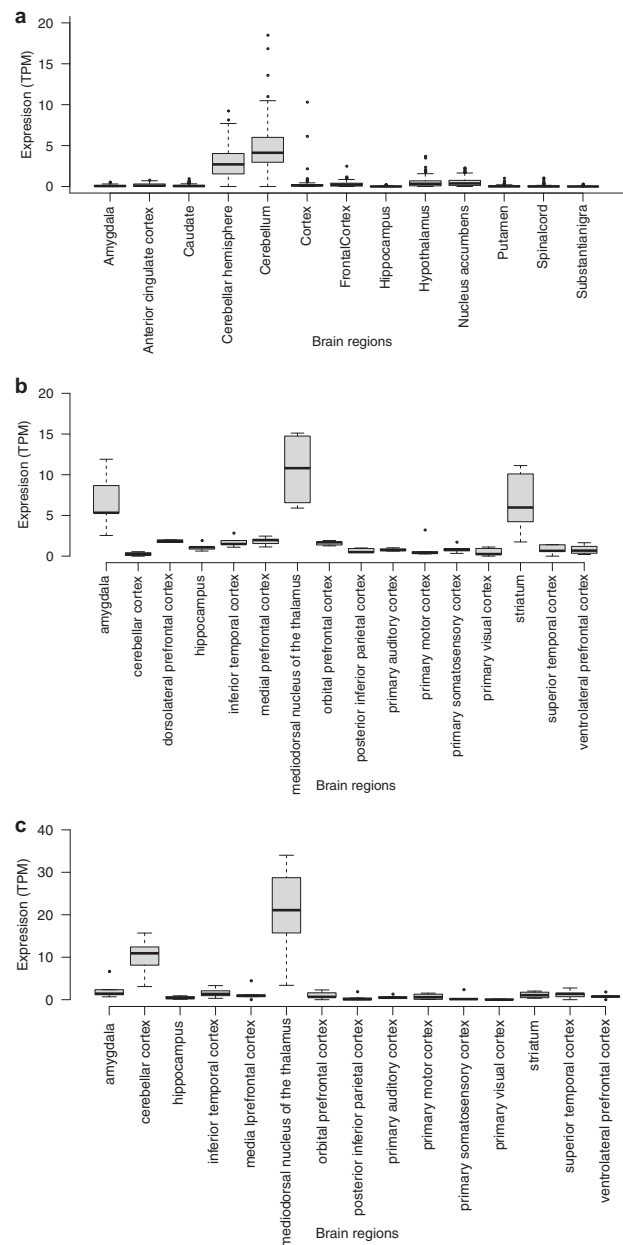
Fig. 1 *OPRM1* expression in postmortem brain across humans (**a**), macaques (**b**), and chimpanzees (**c**). Estimates of *OPRM1* expression in the specific postmortem brain regions based on RNA-Seq datasets obtained from dbGaP (the human GTEx project) and NCBI (for chimpanzees and rhesus macaques)

in Ensembl). Exon 1 is not transcribed in this isoform leading to the absence of the first TM domain.

In chimpanzees, the main isoforms expressed in the cerebellum are pt-201 and pt-202 (Table 1, Fig. 2, Supplementary Tables S1 and S2). Chimpanzee pt-202 (ENSPTRT00000088023) is similar to the human MOR-1X and has a shorter N-terminus and a longer and distinct C-terminus. Chimpanzee pt-201 (ENSPTRT00000094459) is similar to the human MOR-1G2 (ENST00000518759.5,
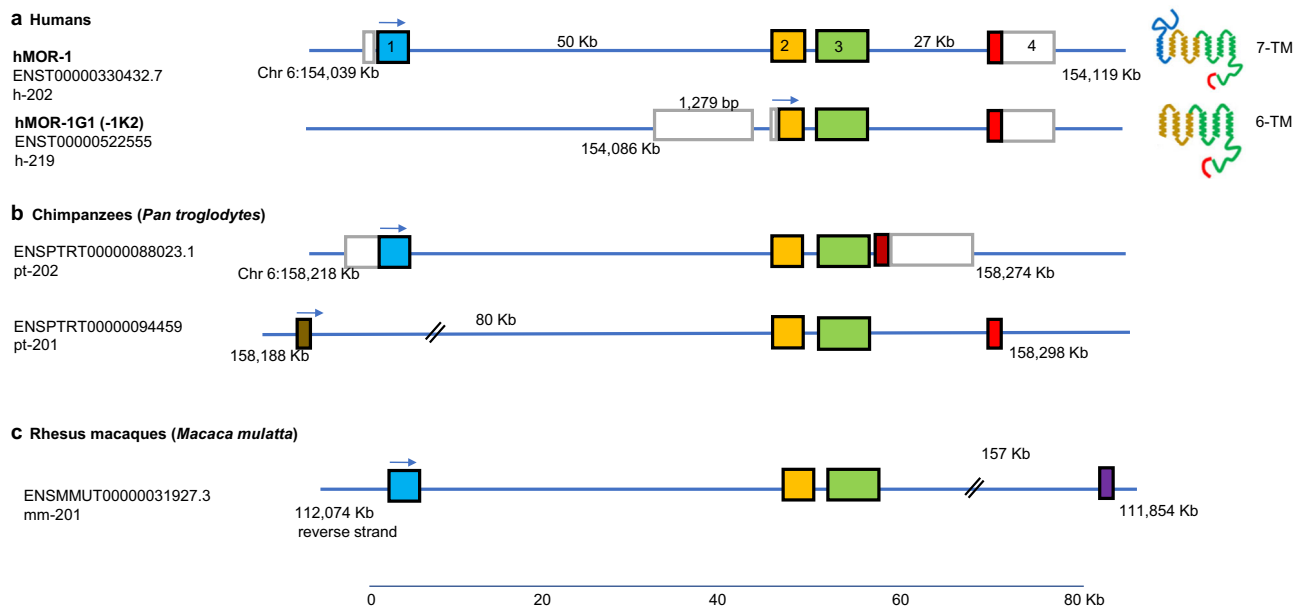
Fig. 2 Schematic representation of the main *OPRM1* isoforms expressed in postmortem brains of humans, chimpanzees, and macaques. Schematic representation of the exons comprising the main isoforms in postmortem brains of humans, chimpanzees, and macaques. Filled colored boxes represent translated exons. Empty gray boxes represent untranslated sequences. Exons 2 and 3 are part of all isoforms. Exons sizes are not in scale

Table 1 Main brain *OPRM1* isoforms in chimpanzees and macaques

Transcript	Splice Variant	Size		Ave. FPKM		
		(bp)	MD	AMY	STR	CB
ptOPRM1						
1 pt-201 ENSPTRT00000094459	1G2	960	7.1 ± 4.7	1.6 ± 1.4	0.5	3.3 ± 0.8
2 pt-202 ENSPTRT00000088023	1×	3293	2.7 ± 1.2	0.5 ± 0.0	1.1	1.6 ± 1.2
mmOPRM1						
1 mm-201 ENSMMUT00000031927	1O	1257	3.5 ± 1.7	1.74 ± 0.0	2.1 ± 0.9	0.11 ± 0.2
2 mm-202 ENSMMUT00000074867	"1"[a]	1461	1.9 ± 0.3	4.1 ± 0.4	1.5	n.d.

n.d. not detected, *MD* the mediodorsal nucleus of the thalamus, *AMY* amygdala, *STR* striatum, *CB* cerebellum,

FPKM fragments per kilobase of transcripts per million mapped reads

[a]This isoform uses an alternative translation site and has additional amino acids in the N-terminus

h-211) and has a novel N-terminus that replaces exon 1. In macaques, the only isoform detected in low levels in the cerebellum is mm-201 (ENSMMUT00000031927.3) that is equivalent to the human isoform MOR-1O, with the canonical exons 1–3, but a distinct C-terminus (exon 4) (Fig. 2, Supplementary Fig. S1, Supplementary Tables S1 and S2). The canonical MOR-1 isoform is not highly expressed in the cerebellum of both species. Exons 2 and 3 are included in all transcripts.

The comparison revealed several differences in the isoforms' expression pattern. Due to the limitations of the study (i.e., a small number of brains of chimpanzees and macaques, the use of postmortem brains, and the generally

low level of *OPRM1* expression), we have focused on one clear specific difference of low *OPRM1* cerebellar expression in macaques compared with human and chimpanzees.

L1 insertions and a CTCF-binding site

The upstream region (~100 kb) of *OPRM1* was explored for regions that were introduced relatively recently in evolution. Tandem insertions of two repetitive elements L1P1 (Chr 6:153,979,014–153,979,696, hg38) and L1PA16 (6:153,979,694–153,982,894) about 60 kb upstream of *OPRM1* were identified (Fig. 3). These two repetitive elements are common in primates and are part of the non-long-

terminal repeats (LTR) retrotransposon long interspersed nuclear element (LINE) family. Sequence alignments revealed that the L1PA16 insertion appeared before the L1P1 insertion. The L1P1 insertion is unique to nonhuman great apes (e.g., gorillas, bonobos, and chimpanzees) indicating a relatively recent evolutionary event that arose sometime after the split of smaller apes (e.g., gibbons and macaques) from the other great apes.

Detailed analysis of the region revealed the presence of a 200-bp binding region for CTCF (CCCTC-binding factor) (ENSR00000323803, 6:153,979,601–153,979,800, hg38) that overlaps the two repetitive elements (Fig. 4). Publicly available data (Ensembl) from ChIP-Seq studies revealed that the site is active in a cell line-specific manner. Several motifs were identified in the region including a CTCF-

binding motif (TGCCCAGTAGGGGCAGACT) at position 6:153,979,659–153,979,677 (hg38) (Jasper, Fimo MA0139.1, score 7.7, $p = 0.0001$, $q = 0.08$) with peak location at position 6:154,979,675 (UCSC accession wgEncodeEH000560) (Fig. 4, Supplementary Tables S3 and S4).

The region is enriched for marks of open chromatin (e.g., DNase1 hypersensitivity), transcription factor binding, histone modifications, and DNA polymerase binding, some of which have been validated in vitro (Supplementary Table S5) (Ensembl). For example, histone methylation- and acetylation-enriched sites (H3K36me3, H3K9me3, and H4K8ac) were shown in human embryonic stem cells (H1-ESC) and H1-neuronal progenitor cells (Roadmap Epigenomics). The binding motif for transcription factors Fox A1, B1, C1, and C2
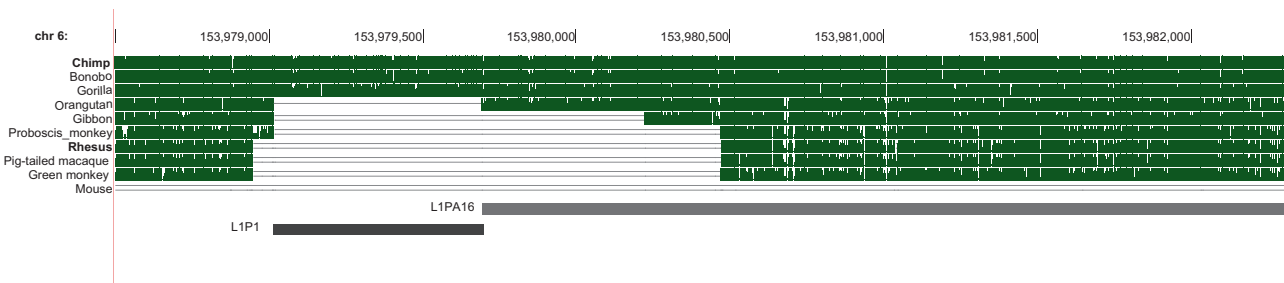

Fig. 3 Insertions of two L1 repetitive elements. A UCSC genome browser image showing multiple alignments and evolutionary conservation of the region of the L1 insertions in selected primates and mouse, compared with the human genome. Green bars indicate conserved regions. The L1PA16 insertion appeared before the L1P1 insertion that is unique to gorillas, bonobos, and chimpanzees. The repeat elements track was created by the RepeatMasker program
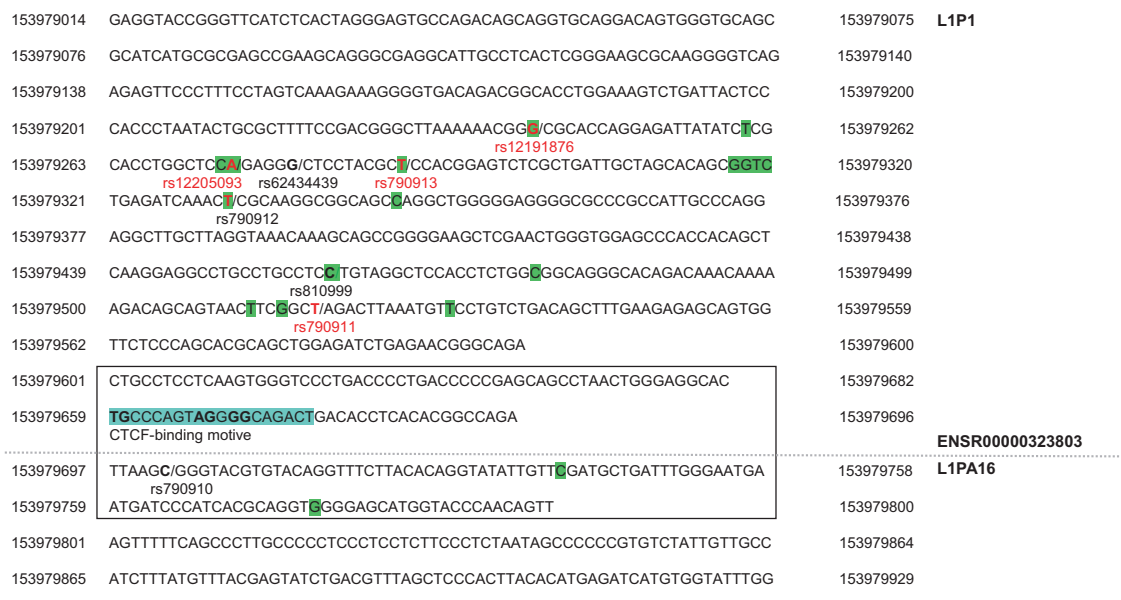

Fig. 4 Sequence of the L1P1 and part of the L1PA16 insertion. L1P1 and L1PA16 regions are marked. CTCF-binding site ENSR00000323803 is marked by a box. CTCF-binding motif is highlighted in turquoise. Human fixed bases and variant alleles are highlighted in green. SNPs in red fonts are part of a haplotype block

(ENSPFM0186; 6:153,979,386–153,979,396) was experimentally verified in the HepG2 cell line (Ensembl).

Homologous sequences in the human genome

The CTCF-binding region described in the current study is overlapping the tandem insertions of L1P1 and the L1PA16 that occurred at different times (Fig. 3). BLASTN analysis of the insertions sequence revealed no sequences with 100% homology in humans or other genomes (Supplementary Table S6 and Supplementary Fig. S2). Numerous sequences with 94% homology were identified in the human genome as well as in chimpanzees and Sumatran orangutans (*Pongo abelii*). Notably, all of them cover 85% of the sequence and are limited to the L1P1 insertion region (6:153,979,014–153,979,697). Analysis of a shorter region of 58-bp overlapping the L1P1 and L1PA16 insertion reveals similar results (Supplementary Table S7), confirming the unique combination of the two inserted elements at this location.

Notably, these homologous insertions did not create CTCF-binding sites. There are several common differences between the identified L1P1 insertion and the homologous sequences that may elucidate their different functionality. While there is a CTCF-binding motif (**TG**CCCAGTAGGGGCAGACT) at position 6:153,979,659–153,979,677 that exists in the chimpanzees insertion in this region, most of the homologous L1P1 insertions in humans, as well as in chimpanzees and orangutans, include the bases 'CC' in position 153,979,659–153,979,660 (**CC**CCCAGTAGGGGCAGACT), a change that reduces the motif score. Also, several common SNPs and fixed variants unique to this insertion appear in sites that are enriched in regulatory features.

SNPs and haplotypes

There are several common SNPs (MAF > 0.1) as well as human fixed variants in the insertion region (Table 2, Figs.

Table 2 Common SNPs in the insertion region

Variant ID	Location Chr 6	Alleles[a]	Global MAF
1 rs12191876	153,979,241	**G**/C	0.1–0.4
2 rs12205093	153,979,274	A/**G**	0.4–0.6
3 rs62434439	153,979,278	**G**/C	0.3–0.6
4 rs790913	153,979,287	**C**/T	0.4–0.5
5 rs790912	153,979,332	**T**/C	0.4–0.6
6 rs810999	153,979,459	C/**T**	0.3–0.4
7 rs790911	153,979,519	**A**/T	0.0–0.3
8 rs790910	153,979,702	C/**G**	0.2–0.4

The ancestor allele is in bold

[a]Alleles are ordered as major/minor in CEU (European ancestry)

4 and 5). SNP rs12191876 is associated with cerebellar *OPRM1* expression (eQTL) in human postmortem brains (GTEx, $p = 0.000047$, normalized effect size (NES) = -0.38, $n = 209$). The ancestral C allele is conserved in chimpanzees and bonobos, but not in gorillas. The human-specific variant G allele became the major allele in all world populations with frequencies of 60–90%. Most (97%) of the HapMap CEU (European ancestry) individuals carry at least one variant G allele.

Haplotype analysis was performed using the 1000 Genome Project (Genomes Project Consortium et al. 2015) phasing information for six common SNPs. The major finding is the strong LD between SNPs rs12191876, rs12205093, and rs790913. Major haplotypes in the 1000 Genome samples with European (CEU) and African (YRI) ancestry are shown in Fig. 5. The variant A allele of rs12205093 appears only with the variant G allele of rs12191876, indicating that it originated later than rs12191876 on the chromosome carrying the G allele. The variant T allele of rs790913 appears only with the two variant alleles indicating that it originated later, on the chromosome carrying both variant alleles. Since this pattern is found in the African sample as well, it probably occurred before the migration from Africa. The variant T allele of rs790911, which is absent in YRI, appears only in the haplotype with the three variant alleles (GAT; rs12191876/rs12205093/rs790913) indicating an even more recent event that occurred after the migration from Africa. The variant 'GAT' haplotype appears in 35% of the YRI chromosomes, while the two similar haplotypes (with and without rs790911, GAT rs12191876/rs12205093/rs790913 and GATT rs12191876/rs12205093/rs790913/rs790911) appear in 53% of the CEU sample. The majority (81%) of the CEU sample carries at least one chromosome with the 'GAT' haplotype.

Analysis of LD patterns of these SNPs revealed that rs790913 has a significantly different LD structure among different populations (Fig. 6). It is in strong LD (r^2) with several downstream SNPs that are closer to the N-terminus of *OPRM1* (-28 kb from MOR-1 and -2 kb from the alternative exon 11 of isoform MOR-1G2), mostly in European and Asian populations (Table 3). In African populations, it has a strong LD only with upstream SNPs further away from *OPRM1* (Fig. 6). This implies that the variant haplotypes 'GAT' and 'GATT' include additional variants closer to *OPRM1* in certain populations. One of these SNPs, the nonconserved rs1294103, is located in an open chromatin region (6:154,008,207–154,008,893; ENSR00000810172) near exon 11 that may be related to the alternative expression of exon 11-isoforms (Fig. 7). This region was shown to be active in A673 cell line (Encode) and originated from a more ancient mammalian L1MA2 insertion. The most 3′ SNP in this LD block is rs712242,
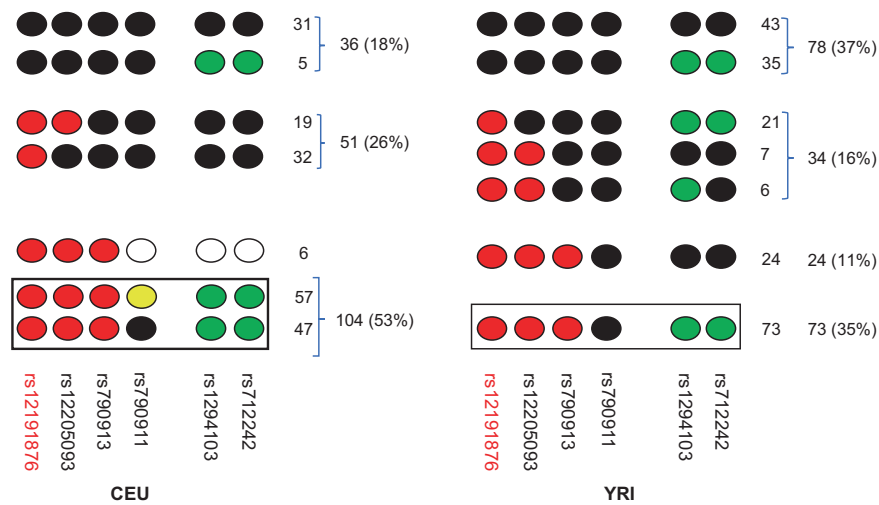
Fig. 5 Distribution of major haplotypes in CEU and YRI. Phased genotype data for selected SNPs in two representative samples (European ancestry; CEU, and African ancestry; YRI) from the 1000 Genomes Project. Red circles represent variant alleles, black circles represent ancestor alleles, yellow circle represents a non-African allele, and green circles represent variant alleles in downstream SNPs that are in strong LD with SNP rs79013 in CEU. Empty circles represent both alleles. Boxed haplotypes are the major haplotype in CEU that includes all the indicated variant alleles
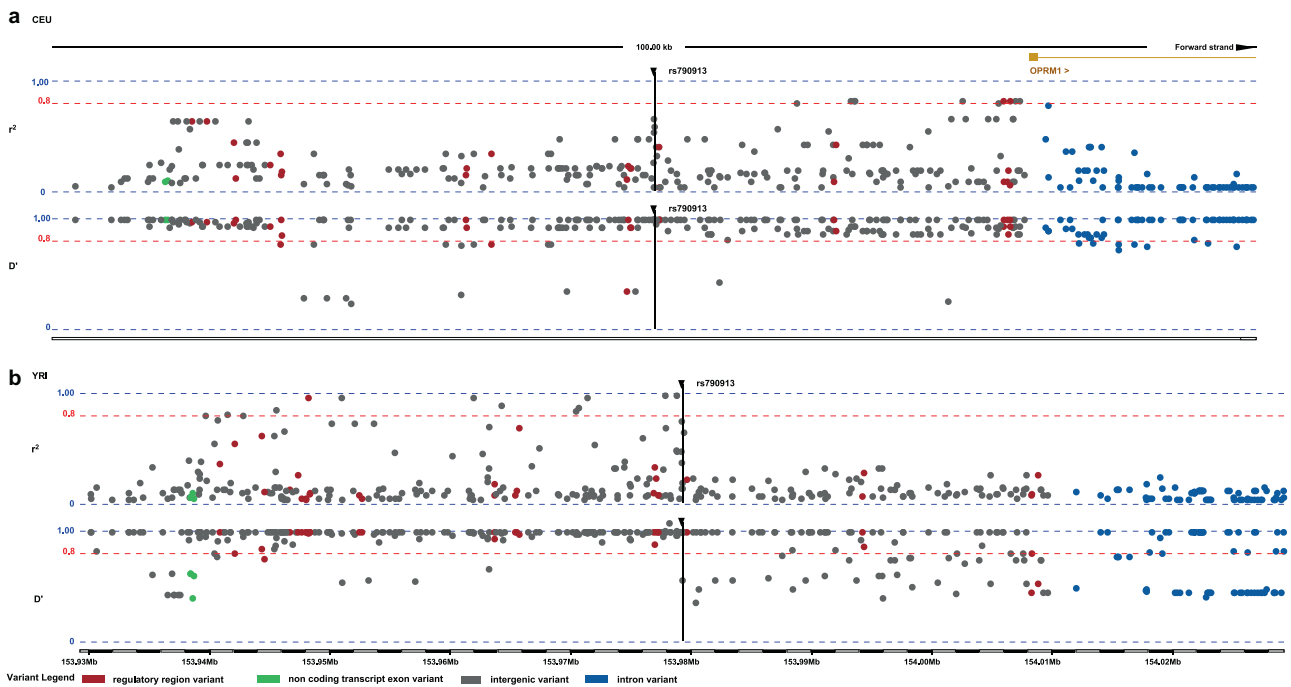

Fig. 6 Linkage disequilibrium (LD) of rs790913 in the HapMap populations of CEU and YRI. Ensembl image of the LD plot for rs790913 in (**a**) CEU and (**b**) YRI. In CEU, there is a strong LD (r^2) between rs790913 and several downstream SNPs that are closer to the N-terminus of *OPRM1*. In YRI, there is strong LD only with upstream SNPs. LD values (D' and r^2) were calculated by a pairwise estimation between SNPs genotyped in the same samples and within a 100 kb window. D' is the difference between the observed and the expected frequency of a given haplotype. r^2 is the correlation between a pair of loci

which is highly conserved and is located 28 kb upstream of exon 1 of MOR-1 and in an intron of isoform MOR-1G2.

The region flanking rs712242 is highly conserved and did not originate from a repetitive element insertion.

Table 3 SNPs in strong LD with SNP rs790913 in CEU (European ancestry)

Variant	Location (Chr 6) GRCh38.p12	Distance from rs790913 (bp)	r^2	D'	Notes
rs1772945	153,991,150	11,863	0.801	0.913	
rs1675901	153,995,635	16,348	0.821	0.934	
rs1617666	153,995,915	16,628	0.821	0.934	
rs1675900	153,995,982	16,695	0.821	0.934	
rs1304725	154,004,918	25,631	0.821	0.934	
rs1294104	154,007,902	28,615	0.801	0.913	
rs1294103	154,008,319	29,032	0.821	0.934	Open chromatin
rs1294100	154,008,870	29,583	0.821	0.934	Open chromatin
rs1294095	154,009,307	30,020	0.821	0.934	
rs1294094	154,009,670	30,383	0.821	0.934	
rs712242	154,012,037	32,750	0.880	0.960	Intron of MOR-1G2
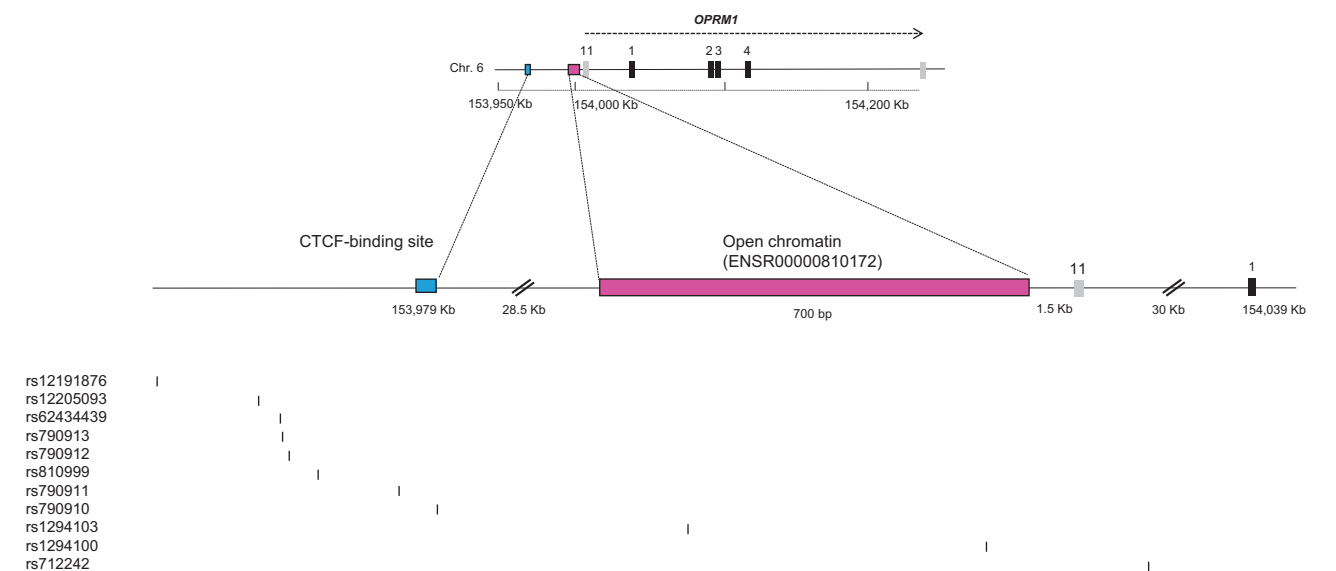
Fig. 7 The *OPRM1* gene and upstream region. Schematic representation of the *OPRM1* gene and its upstream region. The region is zoomed in on the CTCF-binding motif (in turquoise) and the open chromatin region ENSR00000810172 (in pink). The common SNPs are shown at their specific positions

Discussion

OPRM1 expression in the cerebellum

The finding of low cerebellar *OPRM1* expression in macaques compared with humans and chimpanzees is intriguing. To corroborate this finding, we studied in detail the data available from a previous high-throughput transcriptome sequencing (RNA-Seq) study comparing humans, chimpanzees, and rhesus macaques (Xu et al. 2010). Although that study focused on the human-specific genes and did not report findings on *OPRM1*, the study used cerebellar cortex samples from chimpanzees and macaques and their brain samples were different from the ones we have analyzed. Interestingly, they also found low cerebellar

OPRM1 expression in macaques compared with humans and chimpanzees (Supplementary file). Although chimpanzee, macaque, and human data were generated from samples at different ages, causes of death, and postmortem intervals, the observation that humans show a higher level of OPRM1 expression in the cerebellum is robust.

The RNA-Seq study of Sousa et al. (2017) was based on the same brain samples used in our study. Their main focus was on human-specific genes and they did not report findings on *OPRM1*. Nevertheless, they found a significant difference (adjusted $p = 0.0002$) in cerebellum *OPRM1* expression between chimpanzees and macaques, but not between humans and chimpanzees, supporting our analysis (Supplementary file). Interestingly, they reported a similar finding for cerebellar expression of the kappa opioid receptor (KOR,

OPRK1) with a lower but significant level. Notably, the chimpanzee/macaque difference in *OPRM1* expression was not identified in any other brain region in their study.

Cerebellar MOR expression was documented in humans by different methods (e.g., Peckys and Landwehrmeyer 1999, Schadrack et al. 1999, Platzer et al. 2000, Saanijoki et al. 2018, Valentino and Volkow 2018) that reaffirmed the expression of functional receptors. Since rodents' cerebellum is typically devoid of the mu-opioid receptors (Darcq and Kieffer 2018), we hypothesize an evolutionary shift of increase in cerebellar *OPRM1* expression.

Traditionally, the cerebellum was known to be involved in motor control and was not associated with opioidergic mechanisms. Research on the opioidergic system in the cerebellum has been limited, and most research was focused on the classic addiction circuitry (e.g., the ventral tegmental area—the nucleus accumbens dopamine pathway). There is accumulating evidence from the preclinical, neuroimaging, and clinical data that the cerebellum is involved in addiction and reward processing in rodents, as well as in humans (for review, see Moulton et al. 2014, Miquel et al. 2016).

A recent optogenetic study in mice supports the connection between the cerebellum, reward, and social behavior by showing that monosynaptic excitatory projections from the cerebellar nuclei to the VTA activate the reward circuitry and modulate social behavior (Carta et al. 2019). In addition, cerebellar granule cells were shown to encode reward expectations, in mice (Wagner et al. 2017). These studies suggest that the cerebellum was part of the reward pathway even before *OPRM1* was expressed there and may indicate that cerebellar MOR is responsible for specific functions that evolved only in great apes and humans.

The cerebellum expanded rapidly during the hominid evolution, and it shares extensive functional and anatomical connections with the neocortex (MacLeod et al. 2003). There is evidence that the cerebellum is also involved in cognitive functions and social behavior (Schmahmann 2019). It was shown that social grooming and huddling initiated the release of beta-endorphin (Curley and Keverne 2005). A PET study in humans (Hsu et al. 2013) showed that social rejection and acceptance activate the MOR system in neuronal pathways regulating mood and motivation. In addition, naltrexone, an opioid antagonist, was shown to reduce feelings of connection in humans (Inagaki et al. 2016). One hypothesis is that beta-endorphin acquired the function of rewarding social bonding, and MORs are the key substrate for the control of social behavior (Darcq and Kieffer 2018).

MOR isoforms

OPRM1 is composed of several alternative exons and undergoes extensive alternative splicing that creates isoforms with unique characteristics (Regan et al. 2016). Human splicing corresponds closely with that in the mouse (Pasternak and Pan 2013, Xu et al. 2014). Despite the limitations of the current study (i.e., a small number of brains of chimpanzees and macaques, postmortem brain mRNA analysis, and the low level of *OPRM1* expression), we identified some evolutionary shift in *OPRM1* alternative splicing. The main isoforms in humans were the canonical MOR-1 and MOR-1G1. In comparison, MOR-1 was not highly expressed in chimpanzees and rhesus macaques. The finding of high expression of isoform MOR-1G1 in human cerebellum is intriguing. MOR-1G1 is a 6-TM (without the first TM) that exhibits distinct signaling properties and was shown to use Gs protein to increase cAMP production, in contrast to MOR-1 that utilizes Gi/o protein to inhibit cAMP levels (Shabalina et al. 2009, Gris et al. 2010). One of the main isoforms expressed in chimpanzees is MOR-1X. This isoform has a shorter N-terminus and a longer and distinct C-terminus that contains additional phosphorylation sites that are essential for initiating conformational changes and GPCR signal transduction. It exhibits distinct signaling and was shown to be upregulated by morphine (Regan et al. 2016).

L1 insertion

Relatively recent tandem insertions of two L1 transposable elements, which occurred in different times after the split of small apes from the ancestors of the great apes, were identified upstream of *OPRM1*. Interestingly, the region evolved into a CTCF-binding site that overlaps the two elements.

The major class of repetitive elements in the genome is transposable elements (TEs) that constitute ~45% of the human genome and can be subdivided based on their method of replication (Levin and Moran 2011). Retrotransposons are the most abundant class and can be further divided into LTR and non-LTR retrotransposons. Non-LTR elements can be further subdivided into SINE (e.g., Alu elements) and LINE (e.g., LINE1; L1 elements). L1 is a prominent retrotransposon in primate genomes that had an important role in the evolution and architecture of the genome (Konkel et al. 2010). A complete and active L1 element is 6 to 7 kb, but most L1 copies are mutated, rearranged, and/or truncated. The L1P1 and the L1PA16 insertions described in the current study are known to be amplified during the primate radiation (Smit et al. 1995).

Several transposable elements were shown to acquire new physiological functions as regulatory elements and have evolved into *cis*-regulatory sequences that mimic host regulatory elements (Chuong et al. 2017, Ito et al. 2017). CTCF is a DNA-binding protein with 11 zinc fingers that functions as a multifunctional transcription factor that contributes to the establishment of 3D long-range chromatin loop structures and could disrupt transcription by blocking

the connection between enhancer and promoter (Kim et al. 2007, Vietri Rudan and Hadjur 2015).

Although there are numerous sequences in the genome that are homologous to the L1P1 insertion, the specific and later insertion of L1P1 next to L1PA16 is unique to this position. This may explain why similar insertions did not evolve into CTCF-binding sites.

Analysis of the human genetic variability in the CTCF-binding region revealed that the region had acquired several common SNPs and fixed variants and has been actively mutated after the migration out of Africa. At least one variant was associated with cerebellar *OPRM1* expression. In addition, a different LD structure exists among different populations. There is a common variant haplotype in populations with African and European ancestry that has a higher frequency in European populations. Interestingly, in samples of European ancestry, this haplotype extends beyond the CTCF-binding region and includes regulatory variants closer to *OPRM1*. The high frequencies of the haplotype variant alleles compared with the ancestor alleles may suggest a positive selection. This haplotype may drive higher *OPRM1* expression that may have been beneficial.

In conclusion, bioinformatics analysis of publicly available data revealed an evolutionary change in *OPRM1* cerebellar expression, with high *OPRM1* expression in the cerebellum in humans and chimpanzees but low in macaques and rodents. In addition, tandem insertions of L1 retrotransposons upstream of *OPRM1* that acquired a functional CTCF-binding site and a human-specific polymorphism that is associated with cerebellar *OPRM1* expression (eQTL) in human postmortem brains were revealed. This study provides a foundation for building new knowledge about evolutionary differences in *OPRM1* brain expression. Further investigations are needed to elucidate the role of the CTCF-binding region and its SNPs in *OPRM1* expression and to assess the biological function and relevance of *OPRM1* expression in the cerebellum.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

Carta I, Chen CH, Schott AL, Dorizan S, Khodakhah K (2019) Cerebellar modulation of the reward circuitry and social behavior. Science 363:pii: eaav0581

Chuong EB, Elde NC, Feschotte C (2017) Regulatory activities of transposable elements: from conflicts to benefits. Nat Rev Genet 18:71–86

Curley JP, Keverne EB (2005) Genes, brains and mammalian social bonds. Trends Ecol Evol 20:561–567

Darcq E, Kieffer BL (2018) Opioid receptors: drivers to addiction? Nat Rev Neurosci 19:499–514

Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Genomes Project Consortium et al. (2015) A global reference for human genetic variation. Nature 526:68–74

Gris P, Gauthier J, Cheng P, Gibson DG, Gris D, Laur O et al. (2010) A novel alternatively spliced isoform of the mu-opioid receptor: functional antagonism. Mol Pain 6:33

Hsu DT, Sanford BJ, Meyers KK, Love TM, Hazlett KE, Wang H et al. (2013) Response of the mu-opioid system to social rejection and acceptance. Mol Psychiatry 18:1211–1217

Inagaki TK, Ray LA, Irwin MR, Way BM, Eisenberger NI (2016) Opioids and social bonding: naltrexone reduces feelings of social connection. Soc Cogn Affect Neurosci 11:728–735

Ito J, Sugimoto R, Nakaoka H, Yamada S, Kimura T, Hayano T et al. (2017) Systematic identification and characterization of regulatory elements derived from human endogenous retroviruses. PLoS Genet 13:e1006883

Kim D, Langmead B, Salzberg SL (2015) HISAT: a fast spliced aligner with low memory requirements. Nat Methods 12:357–360

Kim TH, Abdullaev ZK, Smith AD, Ching KA, Loukinov DI, Green RD et al. (2007) Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. Cell 128:1231–1245

Konkel MK, Walker JA, Batzer MA (2010) LINEs and SINEs of primate evolution. Evol Anthropol 19:236–249

Levin HL, Moran JV (2011) Dynamic interactions between transposable elements and their hosts. Nat Rev Genet 12:615–627

Levran O, Awolesi O, Linzy S, Adelson M, Kreek MJ (2011) Haplotype block structure of the genomic region of the mu opioid receptor gene. J Hum Genet 56:147–155

Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S et al. (2013) The Genotype-Tissue Expression (GTEx) project. Nat Genet 45:580–585

MacLeod CE, Zilles K, Schleicher A, Rilling JK, Gibson KR (2003) Expansion of the neocerebellum in Hominoidea. J Hum Evol 44:401–429

Miquel M, Vazquez-Sanroman D, Carbo-Gas M, Gil-Miravet I, Sanchis-Segura C, Carulli D et al. (2016) Have we been ignoring the elephant in the room? Seven arguments for considering the cerebellum as part of addiction circuitry. Neurosci Biobehav Rev 60:1–11

Moulton EA, Elman I, Becerra LR, Goldstein RZ, Borsook D (2014) The cerebellum and addiction: insights gained from neuroimaging research. Addict Biol 19:317–331

Pasternak GW, Pan YX (2013) Mu opioids and their receptors: evolution of a concept. Pharm Rev 65:1257–1317

Peckys D, Landwehrmeyer GB (1999) Expression of mu, kappa, and delta opioid receptor messenger RNA in the human CNS: a 33P in situ hybridization study. Neuroscience 88:1093–1135

Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL (2016) Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat Protoc 11:1650–1667

Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL (2015) StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol 33:290–295

Platzer S, Winkler A, Schadrack J, Dworzak D, Tolle TR, Zieglgansberger W et al. (2000) Autoradiographic distribution of mu-, delta- and kappa 1-opioid stimulated [35S]guanylyl-5'-O-(gamma-thio)-triphosphate binding in human frontal cortex and cerebellum. Neurosci Lett 283:213–216

Regan PM, Langford D, Khalili K (2016) Regulation and functional implications of opioid receptor splicing in opioid pharmacology and HIV pathogenesis. J Cell Physiol 231:976–985

Saanijoki T, Tuominen L, Tuulari JJ, Nummenmaa L, Arponen E, Kalliokoski K et al. (2018) Opioid release after high-intensity interval training in healthy human subjects. Neuropsychopharmacology 43:246–254

Schadrack J, Willoch F, Platzer S, Bartenstein P, Mahal B, Dworzak D et al. (1999) Opioid receptors in the human cerebellum: evidence from [11C]diprenorphine PET, mRNA expression and autoradiography. Neuroreport 10:619–624

Schmahmann JD (2019) The cerebellum and cognition. Neurosci Lett 688:62–75

Schmidt D, Schwalie PC, Wilson MD, Ballester B, Goncalves A, Kutter C et al. (2012) Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. Cell 148:335–348

Shabalina SA, Zaykin DV, Gris P, Ogurtsov AY, Gauthier J, Shibata K et al. (2009) Expansion of the human mu-opioid receptor gene architecture: novel functional variants. Hum Mol Genet 18:1037–1051

Smit AF, Toth G, Riggs AD, Jurka J (1995) Ancestral, mammalian-wide subfamilies of LINE-1 repetitive sequences. J Mol Biol 246:401–417

Sousa AMM, Zhu Y, Raghanti MA, Kitchen RR, Onorati M, Tebbenkamp ATN et al. (2017) Molecular and cellular reorganization of neural circuits in the human lineage. Science 358:1027–1032

Trizzino M, Park Y, Holsbach-Beltrame M, Aracena K, Mika K, Caliskan M et al. (2017) Transposable elements are the primary source of novelty in primate gene regulation. Genome Res 27:1623–1633

Valentino RJ, Volkow ND (2018) Untangling the complexity of opioid receptor function. Neuropsychopharmacology 43:2514–2520

Vietri Rudan M, Hadjur S (2015) Genetic tailors: CTCF and Cohesin shape the genome during evolution. Trends Genet 31:651–660

Wagner MJ, Kim TH, Savall J, Schnitzer MJ, Luo L (2017) Cerebellar granule cells encode the expectation of reward. Nature 544:96–100

Xu AG, He L, Li Z, Xu Y, Li M, Fu X et al. (2010) Intergenic and repeat transcription in human, chimpanzee and macaque brains measured by RNA-Seq. PLoS Comput Biol 6:e1000843

Xu J, Lu Z, Xu M, Rossi GC, Kest B, Waxman AR et al. (2014) Differential expressions of the alternatively spliced variant mRNAs of the micro opioid receptor gene, OPRM1, in brain regions of four inbred mouse strains. PLoS ONE 9:e111267

Xu J, Xu M, Hurd YL, Pasternak GW, Pan YX (2009) Isolation and characterization of new exon 11-associated N-terminal splice variants of the human mu opioid receptor gene. J Neurochem 108:962–972

Zerbino DR, Wilder SP, Johnson N, Juettemann T, Flicek PR (2015) The ensembl regulatory build. Genome Biol 16:56