A distinct expression profile in the cerebellum and striatum for genes under selection within introgression deserts

Raül Buisan[1,*], Juan Moriano[1,2,*], Alejandro Andirkó[1,2], and Cedric Boeckx[1,2,3,**]

[1]Universitat de Barcelona
[2]Universitat de Barcelona Institute of Complex Systems
[3]Catalan Institute for Research and Advanced Studies (ICREA)
[*]Contributed equally
[**]Correspondence: cedric.boeckx@ub.edu

Abstract

Analyses of ancient DNA from extinct hominins have provided unique insights into the complex evolutionary history of *Homo sapiens*, intricately related to that of the Neanderthals and the Denisovans as revealed by several instances of admixture events. These analyses have also allowed the identification of introgression deserts: genomic regions in our species that are depleted of 'archaic' haplotypes. The presence of genes like *FOXP2* in these deserts has been taken to be suggestive of brain-related functional differences between *Homo* species. Here, we seek a deeper characterization of these regions, taking into account signals of positive selection in our lineage. Analyzing publicly available transcriptomic data from the human brain at different developmental stages, we found that structures outside the cerebral neocortex, and especially the cerebellum and the striatum at prenatal stages, show the most divergent transcriptomic profiles when considering genes under positive selection within introgression deserts.

Keywords: *Homo sapiens*; deserts of introgression; positive selection; cerebellum; striatum; gene expression.

1 Introduction

The availability of high-coverage genomes from our closest extinct relatives, the Neanderthals and Denisovans, constitutes a significant advance in the range of questions one can ask about the deep history of our species [1–4]. One of the main themes emerging from this progress is interbreeding. In recent years, a fairly large number of admixture events between Neanderthals, Denisovans and Sapiens populations have been postulated. A recent review [5] considers that at least four such events are supported by strong evidence.

While it is important to ask whether our species benefited from these admixture events (so-called adaptive introgression, where alleles inherited from other hominins rose to high frequency as a result of positive selection after gene flow), it is also worth examining regions of the genomes that are depleted of alleles resulting from gene flow from other hominins [6–10]. Such regions are called introgression deserts (sometimes also 'genomic islands of divergence/speciation' [11]).

There are multiple reasons why genetic differences that arose after the divergence of populations may not be well tolerated [12]: there could be negative selection on 'archaic' variants (deleterious changes on the 'archaic' lineage), or positive selection on human-specific variants (adaptive changes on the human lineage), or it may be due to drift. It is reasonable to expect, and indeed has been shown, that the X chromosome constitutes such a desertic region (not only in our species [13]). This could be due to repeated selective sweeps on this chromosome: genes involved in reproduction on this chromosome might act as strong reproductive barriers between populations [14].

In the case of modern humans, other genomic regions are devoid of Neanderthal and Denisovan introgression, for reasons that are perhaps less obvious. A recent study [8] identifies four large deserts depleted of Neanderthal introgression, partially overlapping with a previous independent study [7]. As pointed out in [12, 15], since there quite possibly were several different pulses of gene flow between us and our closest relatives, the depletion observed in these four regions must have been reinforced repeatedly, and given the size of the deserts, it is reasonable that the 'archaic' haplotype was purged within a short time after the gene flow event.

The presence of *FOXP2*, a gene known for its role in language [16, 17], in one of these deserts has attracted attention [15], as it raises the possibility that the incompatibility between Sapiens and other hominin in such persistent introgression deserts may point to (subtle, but real) cognitive differences. Indeed, the presence in such deserts of not only *FOXP2* but also other genes like *ROBO1* and *ROBO2*, both independently associated with language traits [18, 19], together with the observation that genes within deserts are significantly enriched in the developing cerebral cortex and in the adult striatum [7], suggest a possible point of entry into some of the most distinctive aspects of the human condition [20]. Such considerations, combined with independent evidence that introgressed Neanderthal alleles show significant downregulation in brain regions [21], motivated us to focus on the brain in this study.

Specifically, we investigated the four largest genomic regions that resisted 'archaic' introgression reported in [8] jointly with the most comprehensive signals of positive selection in our lineage [22], a combination that, to our knowledge, has not been previously studied in detail. We sought to characterize the expression profile of genes within our regions of interest by analyzing transcriptomic data from several brain regions at different developmental stages from [23], which allows for greater resolution than the Allen Brain Atlas data used in [7], especially for early stages of development. Three of the brain regions under study showed the highest transcriptomic divergence (as assessed via pairwise comparisons of their Euclidean distances based on the two first principal components): the cerebellum, the striatum and the prenatal thalamus. Both the cerebellum and striatum stand out when focusing on genes present at the intersection of regions under positive selection and deserts of introgression. Among the genes found in these overlapping regions that contribute to the observed divergence we found *CADPS2*, *ROBO2*, and *SYT6*, involved in neurotrophin release, axon guidance and neuronal proliferation.

2 Results

We first confirmed the distinctive, derived status of the regions under consideration. The percentage of *Homo sapiens*-derived fixed alleles (in relation to total high-frequency alleles, that is, with >90% frequency in present-day human populations) in deserts of introgression is nearly twice as high the percentage of those in non-desertic regions (16.2% to 8.6%). In addition, almost half of the *Homo sapiens*-derived high-frequency alleles present in putative positively-selected regions within deserts considered here are fixed (46.22%) (all pairwise comparison proved significant; Chi-squared test with Benjamini-Hochberg correction for multiple comparisons; see Table S1). Accordingly, the mean frequency in these overlapping regions is highest, indeed almost fixed (with mean frequency: 0.994), in comparison to desertic and non-desertic regions (see Table S1).

2.1 Genes in deserts of introgression have different expression levels relative to the rest of the genome

We set to understand whether the mean expression of genes in deserts of introgression [8] and the positively selected regions within them (from [22]) have a significantly different mean expression compared to the rest of the genome, using publicly available transcriptomic data from the human brain [23]. To this end, we selected random regions of the genome ($n = 1000$), excluding the deserts, of the same average length, i.e., 15 million base-pairs, with a possible deviation of 1 million base-pairs to account for the length variability between different deserts of introgression. The mean expression of genes lying in random regions of the genome was summarized for each brain structure (and log2-transformed). A repeated-measures two-way ANOVA shows that the mean expression of both sets of these regions is significantly different to the rest of the genome ($p < 0.01$ for both sets). A post-hoc pairwise ANOVA (with Bonferroni correction) shows the difference between a gene expression value in a brain region as derived from the control and that obtained from the genes in our two sets of interest is significant for most structures. An outlier's Grubbs test shows that the structures with the highest and lowest mean gene expression values in deserts of introgression and the positively-selected windows within them fall inside the expected range of variability given the data.

2.2 The cerebellar cortex, the striatum, and the prenatal thalamus show the most divergent profiles when considering genes within deserts and positively-selected regions.

We then investigated the temporal progression of expression profiles of genes within deserts of introgression and putative positively-selected regions analyzing RNA-seq data of different human brain regions at different developmental stages [23]. We found that the median expression of genes within deserts and positively-selected regions is higher than those present in deserts alone, the former peaking at prenatal stages in neocortical areas and decreasing later on. Outside the cerebral neocortex, this pronounced prenatal peak is not observed and, specifically for the cerebellar cortex, the expression profile of these genes increases before birth and reaches the highest median expression from childhood to adulthood in comparison to the rest of structures (see Figure1 A and Figure S1).

In order to statistically evaluate the differences observed for each structure and developmental stage (see Figure1B,C), we performed a principal component analysis and calculated the pairwise Euclidean distances between brain regions for each developmental stage using the first two computed principal components.

At prenatal stages, for genes within deserts, the highest differences (per statistical significance; pairwise Wilcoxon rank sum test with Bonferroni correction) were found for the striatum (fetal stage 1), thalamus (first two fetal stages), and the cerebellar cortex (fetal stage 2). For genes found within both deserts and positively-selected regions, the cerebellar cortex (fetal stages 1 and 2), followed by the striatum (fetal stage 1), stand out (Figures 2 and 3; also Figure S4). These differences are not observed when evaluating the global expression profile (n=9358 genes, see Figure S5). Nor do we find the same effect when looking at the totality of regions of the modern human genome under positive selection, where the thalamus stands out prenatally, and the striatum and cerebellum only do so at later stages (Figure S6).

Postnatally, two structures most significantly differ for both genes within deserts and genes within deserts and positively-selected regions: the striatum at infancy/early childhood (deserts) and at birth/early infancy and adulthood (deserts and positively-selected regions); and the cerebellar cortex, at adolescence (deserts; $p = 3\text{x}10^{-4}$) and at adolescence and adulthood (deserts and positively-selected regions) (Figures 2 and 3; also Figure S4). The pattern observed for the cerebellar cortex, but not for the striatum, at postnatal stages is also found in the global expression profile (see Figure S5), and in the global dataset for genes within positively selected regions (Figure S6) , though.

For the sake of comparison with previous reports, we also used the Allen Brain Atlas dataset and performed a similar exercise: a principal component analysis and pairwise Euclidean distances between brain regions for each developmental stage using the first two computed principal components. The results are reported in Figures S10, S11, S12, and S13. While the cerebellum continues to stand out for both deserts and selected regions within deserts, other structures, especially the striatum are less prominent. We suspect this may be partly due to the lower resolution found in the Allen Brain Atlas dataset at prenatal stages.

The trajectories of expression across developmental stages in genes within deserts of introgression might be affected by positive selection. To control for this, we analyzed the contrast between a control group of genes not under positive selection but within deserts of introgression compared to those under positive selection in these same regions. We found that, within deserts of introgression, genes under positive selection have an overall lower expression than those in regions not under positive selection ($p = 0.0007$, Kruskal-Wallis test, Supp. Fig. 2). A linear regression model predicts that this effect is not structure-specific (p = 0.655), and that overall variability in the data is not explained by between-structure differences ($p = 0.9904$, ANOVA test between fitted models that do and do not include brain regions as a variable). Expression linked to specific developmental stages diverges

significantly between genes under positive selection and those that are not (0.0001, linear regression). However, a post-hoc Tukey test (corrected for repeated measures, Fig. S3) reveals that this difference holds only at the fetal stages. In portions of deserts not under selection, the fetal period of development is significantly different from most posterior stages, while in genes under selective pressures only the first fetal stage is significantly different from post-fetal stages (with a significance threshold of $p < 0.05$).

2.3 Gene-specific expression trajectories of genes in the overlapping desertic and positively-selected regions.

Focusing on the specific trajectories of genes at the intersection of desertic and positively-selected regions (Figure S1), we performed a segmented regression analysis (using the Trendy package [24]) filtering out genes with an adjusted R^2 less than 0.5. For the genes present in these overlapping regions (n=12), the analysis showed a marked increase of transcriptomic divergence for two structures when compared to the rest, the cerebellum and the striatum, at both prenatal and postnatal stages.

For the cerebellum, *CADPS2* expression is the one that most closely mimics the observed pattern, with highest postnatal expression and a marked increased of its expression around birth and infancy (R^2 0.82; see Figure S7 and S8). This Ca^{2+}-dependent activator protein is known to regulate exocytosis in granule cells, particularly neurotrophic factors BDNF and NT-3 release, and its knockout disrupts normal cerebellar development and causes an autistic-like behavioral phenotype in mice [25, 26]. In addition, increasing expression through developmental stages was also found for *GPR22* and *BCAP29* (R^2 0.80 and 0.65, respectively; see Figure S8). Two other genes, *KCND2* and *ST7*, with comparatively high expression postnatally, did not pass however the adjusted R^2 threshold (Figure S7).

Regarding the striatum, four genes within the overlapping desertic and positively-selected regions could be fitted with an adjusted R^2 higher than 0.5. Two of them, *ROBO2* and *SYT6*, show sustained higher expression at prenatal stages followed by a marked decreased around birth, ending with low expression at later postnatal stages (R^2 0.82 and 0.68, respectively; see Figure S9). Robo2 is a receptor of the Slit/Robo signaling pathway which is critically involved in axon guidance [27], but also in the proliferation and differentiation of neural progenitors with possible different roles in dorsal and ventral telencephalon [28, 29]. *Syt6* is another synapse-related gene expressed in the developing basal ganglia [30], and in fact linked to the distinctive expression profile of this structure [31]. Additionally, *Syt6* shows a similar expression profile in the cerebellar cortex although at lower levels (see Figure S8), where is differentially expressed in *Cadps2* knockout mice [32]. Trajectories for *BCAP29*, with increasing expression starting at around birth (R^2 0.77), and *ST7*, with decreasing expression from early fetal stages (R^2 0.58), could also be fitted above the set threshold.

3 Discussion

The current study brings out two main findings: the importance of structures beyond the cerebral neocortex in the attempt to characterize some of the most derived features of our species' brain, and the fact that the strongest effect in these regions takes place at early stages of development. This adds complementary evidence for the perinatal globularization phase as a species-specific ontogenic innovation [33]. It also provides new evidence for the claim that the cerebellum, the striatum, and the thalamus significantly contribute to this phenotype [21, 34–37].

To our knowledge this is the first study to characterize the effect of the cerebellum in the context of introgression deserts. For the striatum, previous studies have already highlighted the relevance of this structure: genes carrying Neanderthal-derived changes and expressed in the striatum during adolescence exhibit a higher McDonald-Kreitman ratio [4]. In addition, using a different range of introgressed regions, it had already been noted [7] that genes within deserts are significantly enriched in the developing in the striatum at adolescence and adult stages. For us, though, the effect of the striatum is more salient at earlier stages of development.

Naturally, the functional effects of these divergent developmental profiles for the cerebellum, striatum, and prenatal thalamus remain to be understood. But the fact that prenatal stages are key holds the promise of using brain organoid technology to probe the nature of these differences, since such *in vitro* techniques best track these earliest developmental windows [38–40].

The fact that FOXP2 expression is known to be particularly high in the brain regions highlighted here [41] may help shed light on why *FOXP2* is found in introgression deserts in modern human genomes. As pointed out in [15], this portion of chromosome 7 is not a desert for introgression in other great apes, nor did it act as a barrier for gene flow from Sapiens into Neanderthals. As such, it may therefore capture something genuinely specific about our species.

4 Methods

Processing a catalog of *Homo sapiens*-specific high-frequency variants. To evaluate the degree of fixation of *Homo sapiens*-specific variants in regions of interest, we processed a dataset containing single-nucleotide changes derived in the modern human lineage at fixed or nearly fixed frequency (above 90% in present-day human populations) where high-quality Neanderthal/Denisovan genomes carry the ancestral allele [42]. This dataset was then intersected with genomic data of regions depleted of archaic haplotypes [8] and putative positively-selected regions [22]. Statistical evaluation of differences in mean frequency and percentage of fixed alleles between set of regions was performed with a Wilcoxon rank sum test and Chi-squared test, respectively, with Benjamini-Hochberg correction for multiple comparisons. Significant differences were considered if $p < .01$. Genes within deserts of introgression or putative positively-selected regions were obtained via BioMart R package [43], using the respective genomic region coordinates as input and filtering by protein-coding genes.

mRNA-seq analysis. Publicly available transcriptomic data of the human brain at different developmental stages was retrieved from [23]. Reads per kilo base per million mapped reads (RPKM) normalized counts were log-transformed and then subsetted to select genes either in deserts of introgression or in both deserts and putative positively-selected regions. The complete log-transformed, RPKM normalized count matrix was subsetted to select genes with median expression value > 2, as in [23], while no median filtering was employed for the subsets of genes within deserts and positively-selected regions, due to the potential relevance of the outliers in these specific regions for the purposes of our study. To assess transcriptomic variability between brain regions (following strategy in [23]) accounted for by genes either in deserts or in deserts and positively-selected regions, we performed principal component analysis and calculated the pairwise Euclidean distances between brain regions for each dataset. We then statistically evaluated such differences at each developmental stage using pairwise Wilcoxon tests with Bonferroni correction. Significant differences were considered if $p < .01$. Additionally, to increase the robustness of our results, we also analyzed RNA-seq data from the Allen Brain Atlas (using [44]), which contains transcriptomic data for similar brain regions but fewer developmental stages. These results are discussed in Figures S10, S11, S12 and S13.

Gene-specific expression trajectories. The R package Trendy [24] was used to perform segmented regression analysis and characterize the expression trajectories of genes within both deserts of introgression and putative positively-selected regions (12 genes). The no log-transformed, normalized RPKM values (from [23]) in the form of a gene-by-time samples matrix was used to fit each gene expression trajectory to an optimal segmented regression model. Genes were considered if their adjusted R^2 was > 0.5. In addition, a maximum number of breakpoints (significant changes in gene expression trajectory) was set at 3, minimum number of samples in each segment at 2, and minimum mean expression, 2.

The permutation tests using gene expression data from [23] and [45] were done using the regioneR package [46] at $n = 1000$.

Data and material availability

Datasets and code are made available at `https://github.com/jjaa-mp/desertsHomo`.

Author Contributions

Conceptualization: CB & AA & JM & RGB; Data Curation: AA & JM & RGB; Formal Analysis: AA & JM & RGB; Visualization: CB & AA & JM & RGB; Writing – Original Draft Preparation: CB & AA & JM & RGB; Writing – Review & Editing: CB & AA & JM & RGB; Supervision: CB; Funding Acquisition: CB.

Funding statement

CB acknowledges support from the Spanish Ministry of Economy and Competitiveness (grant PID2019-107042GB-I00), MEXT/JSPS Grant-in-Aid for Scientific Research on Innovative Areas #4903 (Evolinguistics: JP17H06379), Generalitat de Catalunya (2017-SGR-341), and the Foundation BBVA (Leonardo fellowship). JM acknowledges financial support from the Departament d'Empresa i Coneixement, Generalitat de Catalunya (FI-SDUR 2020). AA acknowledges financial support from the Spanish Ministry of Economy and Competitiveness and the European Social Fund (BES-2017-080366).

References

[1] M. Meyer, *et al.*, "A High-Coverage Genome Sequence from an Archaic Denisovan Individual," *Science*, vol. 338, pp. 222–226, Oct. 2012.

[2] K. Prüfer, *et al.*, "The complete genome sequence of a Neanderthal from the Altai Mountains," *Nature*, vol. 505, pp. 43–49, Jan. 2014.

[3] K. Prüfer, *et al.*, "A high-coverage Neandertal genome from Vindija Cave in Croatia," *Science*, vol. 358, pp. 655–658, Nov. 2017.

[4] F. Mafessoni, *et al.*, "A high-coverage Neandertal genome from Chagyrskaya Cave," *Proceedings of the National Academy of Sciences*, vol. 117, pp. 15132–15136, June 2020. Publisher: National Academy of Sciences Section: Biological Sciences.

[5] A. Bergström, *et al.*, "Origins of modern human ancestry," *Nature*, vol. 590, pp. 229–237, Feb. 2021.

[6] S. Sankararaman, *et al.*, "The Combined Landscape of Denisovan and Neanderthal Ancestry in Present-Day Humans," *Current Biology*, vol. 26, pp. 1241–1247, May 2016.

[7] B. Vernot, *et al.*, "Excavating Neandertal and Denisovan DNA from the genomes of Melanesian individuals," *Science*, vol. 352, pp. 235–239, Apr. 2016. Publisher: American Association for the Advancement of Science Section: Report.

[8] L. Chen, *et al.*, "Identifying and Interpreting Apparent Neanderthal Ancestry in African Individuals," *Cell*, vol. 180, pp. 677–687.e16, Feb. 2020.

[9] L. Skov, *et al.*, "The nature of Neanderthal introgression revealed by 27,566 Icelandic genomes," *Nature*, vol. 582, pp. 78–83, June 2020.

[10] D. C. Rinker, *et al.*, "Neanderthal introgression reintroduced functional ancestral alleles lost in Eurasian populations," *Nature Ecology & Evolution*, vol. 4, pp. 1332–1341, Oct. 2020.

[11] S. Wang, *et al.*, "Selection on a small genomic region underpins differentiation in multiple color traits between two warbler species," *Evolution Letters*, vol. 4, no. 6, pp. 502–515, 2020.

[12] A. B. Wolf *et al.*, "Outstanding questions in the study of archaic hominin admixture," *PLOS Genetics*, vol. 14, p. e1007349, May 2018.

[13] M. Kuhlwilm, *et al.*, "Ancient admixture from an extinct ape lineage into bonobos," *Nature Ecology & Evolution*, vol. 3, pp. 957–965, June 2019. Number: 6 Publisher: Nature Publishing Group.

[14] C. Fontsere, *et al.*, "Admixture in Mammals and How to Understand Its Functional Implications," *BioEssays*, vol. 41, no. 12, p. 1900123, 2019.

[15] M. Kuhlwilm, "The evolution of FOXP2 in the light of admixture," *Current Opinion in Behavioral Sciences*, vol. 21, pp. 120–126, June 2018.

[16] C. S. L. Lai, *et al.*, "A forkhead-domain gene is mutated in a severe speech and language disorder," *Nature*, vol. 413, pp. 519–523, Oct. 2001. Number: 6855 Publisher: Nature Publishing Group.

[17] S. E. Fisher, "Human Genetics: The Evolving Story of FOXP2," *Current Biology*, vol. 29, pp. R65–R67, Jan. 2019.

[18] R. Wang, *et al.*, "Convergent differential regulation of SLIT-ROBO axon guidance genes in the brains of vocal learners," *Journal of Comparative Neurology*, vol. 523, no. 6, pp. 892–906, 2015. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cne.23719.

[19] B. St Pourcain, *et al.*, "Common variation near ROBO2 is associated with expressive vocabulary in infancy," *Nature Communications*, vol. 5, p. 4831, Sept. 2014. Number: 1 Publisher: Nature Publishing Group.

[20] S. Pääbo, "The Human Condition—A Molecular Approach," *Cell*, vol. 157, pp. 216–226, Mar. 2014.

[21] R. C. McCoy, *et al.*, "Impacts of Neanderthal-Introgressed Sequences on the Landscape of Human Gene Expression," *Cell*, vol. 168, pp. 916–927.e12, Feb. 2017.

[22] S. Peyrégne, *et al.*, "Detecting ancient positive selection in humans using extended lineage sorting," *Genome Research*, vol. 27, pp. 1563–1572, Sept. 2017.

[23] M. Li, *et al.*, "Integrative functional genomic analysis of human brain development and neuropsychiatric risks," *Science*, vol. 362, Dec. 2018.

[24] R. Bacher, *et al.*, "Trendy: segmented regression analysis of expression dynamics in high-throughput ordered profiling experiments," *BMC Bioinformatics*, vol. 19, p. 380, Oct. 2018.

[25] T. Sadakata, *et al.*, "Autistic-like phenotypes in Cadps2-knockout mice and aberrant CADPS2 splicing in autistic patients," *The Journal of Clinical Investigation*, vol. 117, pp. 931–943, Apr. 2007.

[26] T. Sadakata, *et al.*, "Axonal Localization of Ca2+-Dependent Activator Protein for Secretion 2 Is Critical for Subcellular Locality of Brain-Derived Neurotrophic Factor and Neurotrophin-3 Release Affecting Proper Development of Postnatal Mouse Cerebellum," *PLOS ONE*, vol. 9, p. e99524, June 2014. Publisher: Public Library of Science.

[27] G. López-Bendito, *et al.*, "Robo1 and Robo2 Cooperate to Control the Guidance of Major Axonal Tracts in the Mammalian Forebrain," *Journal of Neuroscience*, vol. 27, pp. 3395–3407, Mar. 2007. Publisher: Society for Neuroscience Section: Articles.

[28] W. Andrews, *et al.*, "The role of Slit-Robo signaling in the generation, migration and morphological differentiation of cortical interneurons," *Developmental Biology*, vol. 313, pp. 648–658, Jan. 2008.

[29] V. Borrell, *et al.*, "Slit/Robo Signaling Modulates the Proliferation of Central Nervous System Progenitors," *Neuron*, vol. 76, pp. 338–352, Oct. 2012.

[30] J. E. Long, *et al.*, "Dlx1&2 and Mash1 Transcription Factors Control MGE and CGE Patterning and Differentiation through Parallel and Overlapping Pathways," *Cerebral Cortex*, vol. 19, pp. i96–i106, July 2009.

[31] G. Konopka, *et al.*, "Human-Specific Transcriptional Networks in the Brain," *Neuron*, vol. 75, pp. 601–617, Aug. 2012. Publisher: Elsevier.

[32] T. Sadakata, *et al.*, "Analysis of gene expression in Ca2+-dependent activator protein for secretion 2 (Cadps2) knockout cerebellum using GeneChip and KEGG pathways," *Neuroscience Letters*, vol. 639, pp. 88–93, Feb. 2017.

[33] P. Gunz, *et al.*, "Brain development after birth differs between Neanderthals and modern humans," *Current Biology*, vol. 20, pp. R921–R922, Nov. 2010.

[34] C. A. Boeckx *et al.*, "The shape of the human language-ready brain," *Frontiers in Psychology*, vol. 5, 2014. Publisher: Frontiers.

[35] S. Neubauer, *et al.*, "The evolution of modern human brain shape," *Science Advances*, vol. 4, p. eaao5961, Jan. 2018.

[36] P. Gunz, *et al.*, "Neandertal Introgression Sheds Light on Modern Human Endocranial Globularity," *Current Biology*, vol. 29, pp. 120–127.e5, Jan. 2019.

[37] C. V. Weiss, *et al.*, "The *cis* -regulatory effects of modern human-specific variants," preprint, Evolutionary Biology, Oct. 2020.

[38] S. K. Muchnik, *et al.*, "Modeling the Evolution of Human Brain Development Using Organoids," *Cell*, vol. 179, pp. 1250–1253, Nov. 2019. Publisher: Elsevier.

[39] M. A. Mostajo-Radji, *et al.*, "Reverse engineering human brain evolution using organoid models," *Brain Research*, vol. 1729, p. 146582, Feb. 2020.

[40] C. Kyrousi *et al.*, "Using brain organoids to study human neurodevelopment, evolution and disease," *WIREs Developmental Biology*, vol. 9, no. 1, p. e347, 2020. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/wdev.347.

[41] C. S. L. Lai, *et al.*, "FOXP2 expression during brain development coincides with adult sites of pathology in a severe speech and language disorder," *Brain*, vol. 126, pp. 2455–2462, Nov. 2003.

[42] M. Kuhlwilm *et al.*, "A catalog of single nucleotide changes distinguishing modern humans from archaic hominins," *Scientific Reports*, vol. 9, p. 8463, June 2019.

[43] S. Durinck, *et al.*, "Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt," *Nature Protocols*, vol. 4, pp. 1184–1191, Aug. 2009. Number: 8 Publisher: Nature Publishing Group.

[44] S. Grote, *et al.*, "ABAEnrichment: an R package to test for gene set expression enrichment in the adult and developing human brain," *Bioinformatics*, vol. 32, pp. 3201–3203, Oct. 2016.

[45] M. J. Hawrylycz, *et al.*, "An anatomically comprehensive atlas of the adult human brain transcriptome," *Nature*, vol. 489, pp. 391–399, Sept. 2012.

[46] B. Gel, *et al.*, "regioneR: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests," *Bioinformatics*, vol. 32, pp. 289–291, Jan. 2016.

Set	Chromosome	Start	End
Deserts of introgression [8]	1	105400000	120600000
	3	74100000	89300000
	7	106200000	123200000
	8	49400000	66500000
Regions under positive selection [22] within deserts	1	113427676	113560554
	1	114641362	114645248
	1	119322276	119387279
	3	77027847	77034264
	7	106877730	107233808
	7	116762909	116773234
	7	120147456	120174406
	7	122320035	122406480

Table 1: **Genomic coordinates used in this study.**

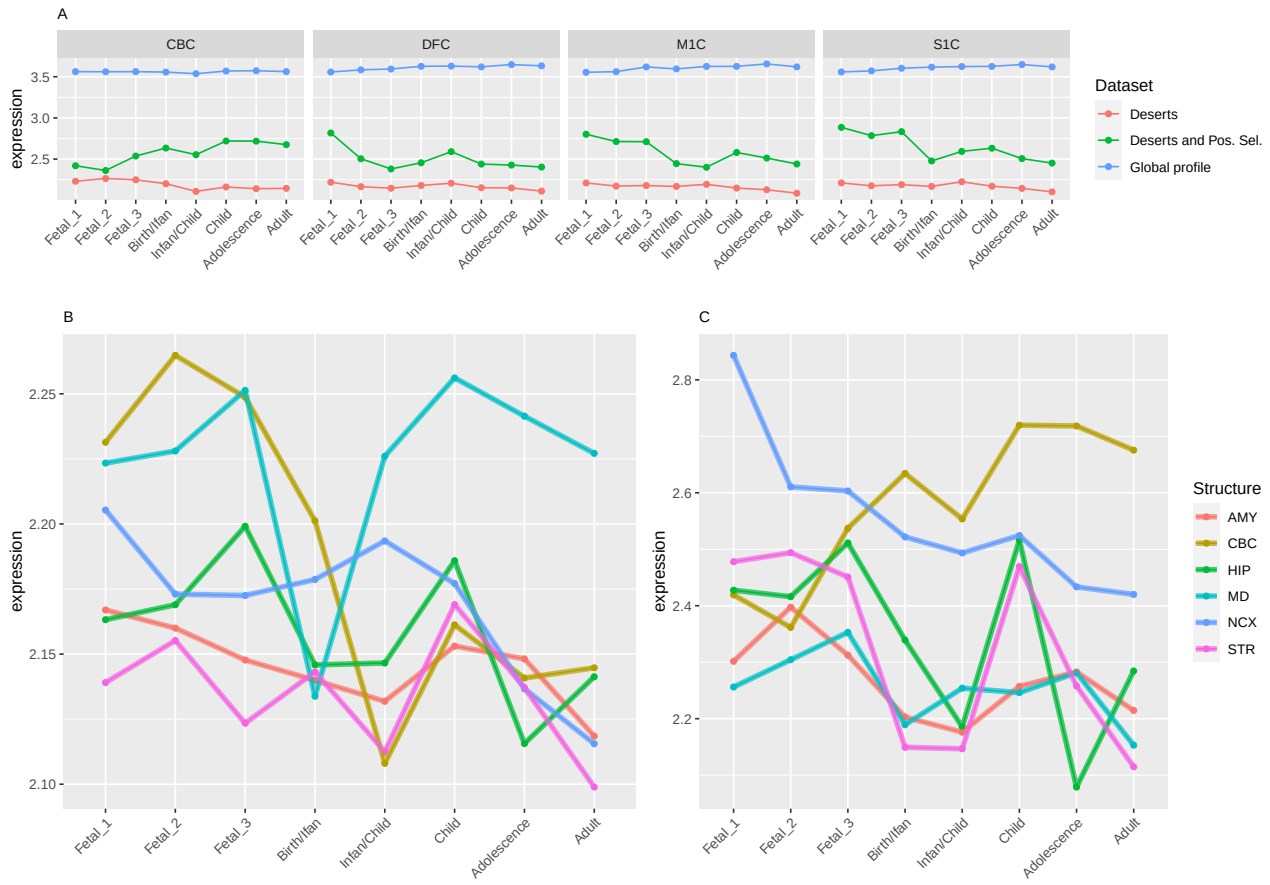

Figure 1: **Median expression of genes in deserts and deserts/positively-selected regions.** A) Selected examples comparing genes in the global dataset (n=9358 genes; requiring median > 2), in deserts of introgression (n=255) and in both deserts and positively-selected regions (n=12). A panel with all structures analyzed in this study is shown in Figure S1. In neocortical areas, genes within deserts and positively-selected regions peak at prenatal stages and decrease later, whereas the opposite pattern is found for the cerebellar cortex (see also C). Of note, the global dataset was filtered by setting a threshold of median expression higher than 2 (following [23]), to avoid genes with extremely low expression, while we kept any outlier for the two other datasets. B) Median expression profile of genes within deserts. The cerebellar cortex, prenatally, and the mediodorsal nucleus of the thalamus prenatally and postnatally present the highest expression. C) Median expression profile of genes within deserts and positively-selected regions. Genes expressed in neocortical areas reach the highest expression at the early fetal stages, whereas the cerebellar cortex, from birth until adulthood, remains with the highest expression profile.
Structures: AMY, amygdala; CBC, cerebellar cortex; HIP, hippocampus; MD, mediodorsal nucleus of thalamus; NCX, neocortex; STR, striatum.
Stages: Fetal 1: 12-13 post conception weeks (PCWs); Fetal 2: 16-18 PCW; Fetal 3: 19-22 PCW; Birth - Infancy: 35-37 PCW and 0-0.3 years; Infancy - Child: 0.5-2.5 years; Childhood: 2.8-10.7 years, Adolescence: 13-19 years; Adulthood: 21-64 years (as in [23])

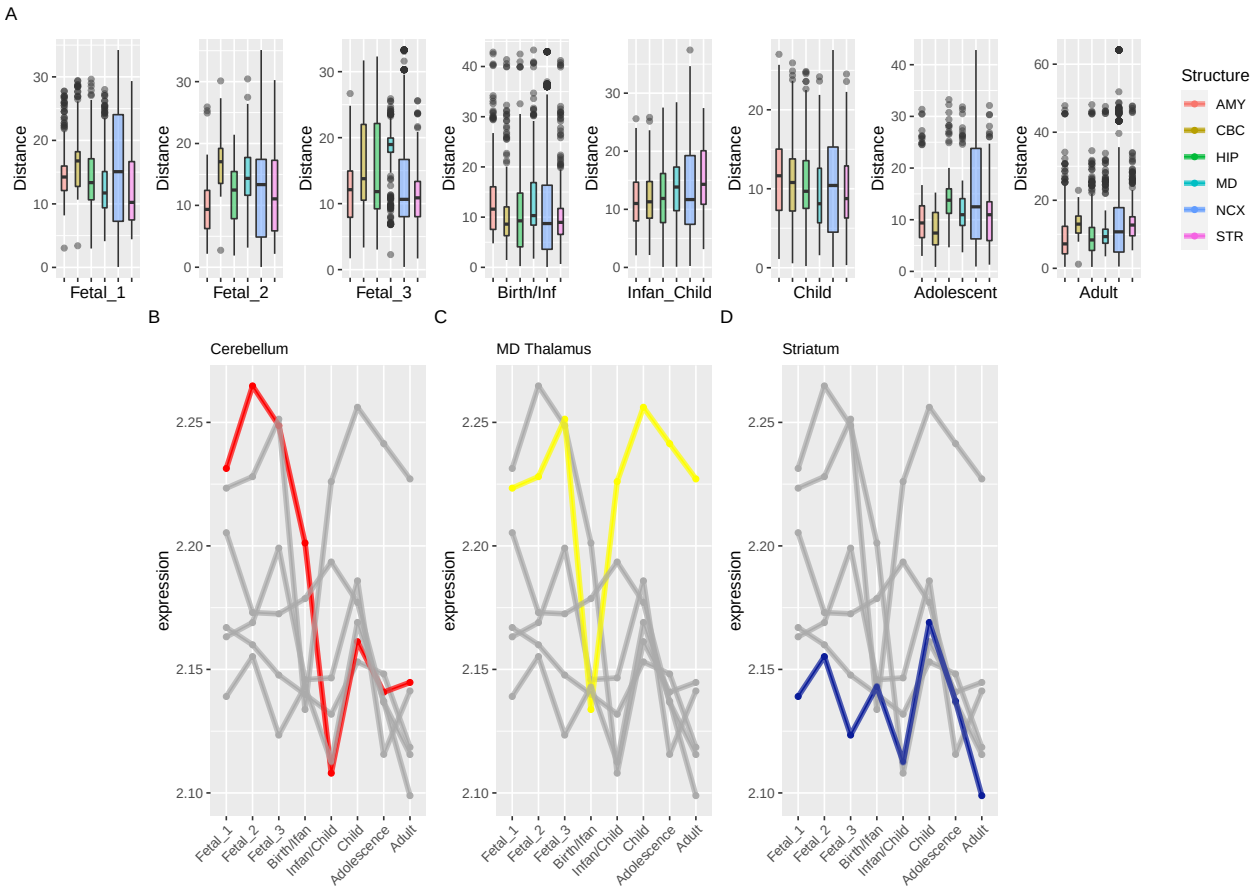

Figure 2: **Pairwise Euclidean distance comparisons between brain regions (genes within deserts).** A) Boxplots showing the distances between a given brain region and all other regions based on the first two principal components, grouped by stages. B, C, and D) Gene expression profile from brain regions showing the highest significance values after a pairwise Wilcoxon rank sum test with Bonferroni correction: Cerebellum (red, B), Mediodorsal nucleus of the thalamus (yellow, C), and Striatum (blue, D).
AMY, amygdala; CBC, cerebellar cortex; HIP, hippocampus; MD, mediodorsal nucleus of thalamus; NCX, neocortex; STR, striatum.

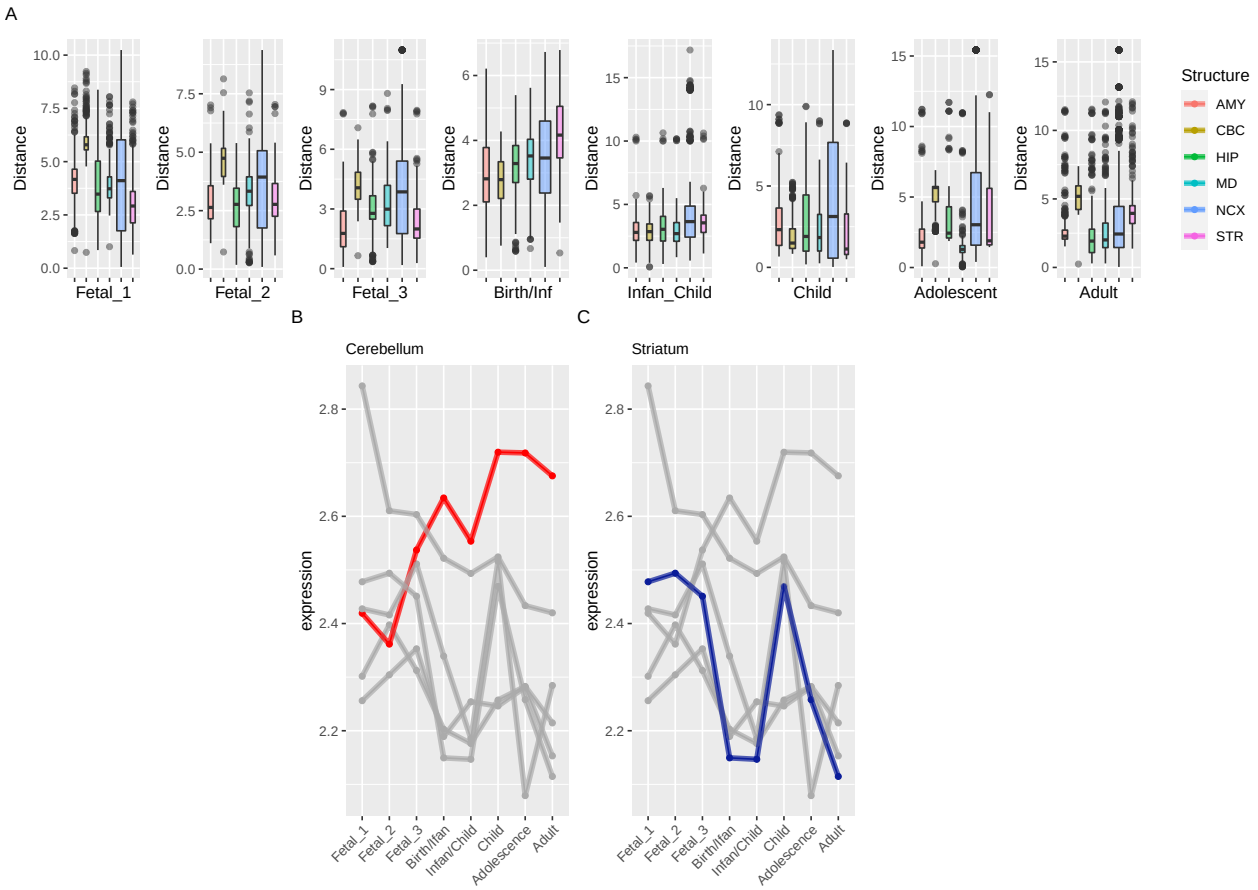

Figure 3: **Pairwise Euclidean distance comparisons between brain regions (genes within deserts and positively-selected regions)** Pairwise Euclidean distance comparisons between brain regions (genes within deserts). A) Boxplot pairwise distances between brain regions (based on first two principal components), grouped by stages. B and C) Gene expression profile from brain regions showing the highest significance values (pairwise Wilcoxon rank sum test with Bonferroni correction): Cerebellum (red, B) and Striatum (blue, C). AMY, amygdala; CBC, cerebellar cortex; HIP, hippocampus; MD, mediodorsal nucleus of thalamus; NCX, neocortex; STR, striatum.