Enhancing Evolution

Charting regulatory changes across
primate evolution

Bas Castelijns

ISBN: 978-90-393-7286-9

Cover Art: Charting evolution, by B. Castelijns

Printing: Ridderprint | www.ridderprint.nl

The work described in this thesis was performed at the Hubrecht Institute for Developmental Biology and Stem Cell Research of the Royal Dutch Academy of Arts and Sciences (KNAW), within the framework of the Cancer, Stem Cells and Developmental Biology (CSnD) graduate program of Utrecht University.

Copyright 2020 by B. Castelijns. All rights reserved. No parts of this book may be reproduced, stored in retrieval system, or transmitted in any form or by any means, without prior permission of the author.

Enhancing Evolution

Charting regulatory changes across primate evolution

Evolutie Eleveren

In kaart brengen van regulatoire veranderingen in primaat evolutie

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht
op gezag van de rector magnificus, prof.dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op

donderdag 12 november 2020 des ochtends te 11.00 uur

door

Bas Castelijns

geboren op 2 juli 1989 te Alkmaar

Promotor:

Prof. dr. W.L. de Laat

Copromoter:

Dr. M.P. Creyghton

**“Nothing in Biology Makes Sense Except in the
Light of Evolution”**

Theodosius Dobzhansky

Table of Contents

Chapter 1	9
General introduction	
 Chapter 2	 25
Hominin-specific regulatory elements selectively emerged in oligodendrocytes and are disrupted in autism patients	
 Chapter 3	 69
Recently evolved enhancers emerge with high interindividual variability and less frequently associate with disease	
 Chapter 4	 103
Cell type resolved gene expression reveals no distinct oligodendrocyte subpopulation affected by hominin-specific regulatory rewiring	
 Chapter 5	 123
General discussion	
 Addendum	 139
Nederlandse samenvatting	
Acknowledgements	
List of publications	
Curriculum Vitea

Chapter 1

General Introduction

B. Castelijns¹

¹Hubrecht Institute-KNAW & University Medical Center Utrecht,
Utrecht, The Netherlands

Genome structure

The human genome contains all the genetic information needed to create a complete body from a single fertilized cell. This genome consists of long strands of DNA, constructed from 4 different nucleotides; adenine (A), thymine (T), cytosine (C) and guanine (G). For humans, this is roughly 3 billion nucleotides long and if completely stretched out, would measure 2 meters from end to end. To ensure proper fitting inside the cell nucleus, the genome is tightly packed but nevertheless highly structured. The basic building block is the nucleosome, which is ~150 nucleotides wrapped around 8 histone proteins¹. These nucleosomes are packaged in 30-nanometer fibers^{2,3} and further condensed into the well-known chromosomal structure (Fig. 1a).

Since the first report of the human genome sequence in 2001⁴ and its completion in 2004⁵, we have an almost complete picture of the human genome sequence at every position⁶. Analysis of this sequence has revealed that our species holds roughly 20,000 protein coding genes, which is fewer than the ~100,000 predicted earlier⁷. This number is similar to other mammals, such as non-human primates and mice^{8,9}, but fewer than found in several plants or other eukaryotic organisms^{10,11}. This disparity indicates that gene count or genome size alone are not good proxies for organism complexity¹². Moreover, the most remarkable finding was that most of our genome does not encode for any protein coding gene and therefore seemed to be devoid of any biological function^{6,13}.

These non-coding parts of our genome have long been thought of as so-called “junk-DNA”, genomic pieces that were passed along during evolution but did not harbor any functionality. However, it recently became clear that some regions are more conserved across evolution than would be expected based on random mutational shift, and that there are ultra-conserved genomic stretches between humans and rodents, indicating potential important biological roles^{14,15}. This became more evident with the release of the ENCODE project in 2012, which systematically mapped biological activity and function across our genome¹⁶. It is now clear that even though most parts of the genome do not code for proteins, it is littered with regulatory elements with an estimated count of over 400,000¹⁶⁻¹⁹. Nevertheless, the precise function for most of these elements remains unknown.

Gene regulatory elements and their function

Every cell in our body originates from a single stem cell and, in principle, carries the same genetic information. However, we comprise many different cell types with a diverse set of functions. Therefore, each cell needs to access different parts of the genome in order to activate expression of the right set of genes while ensuring that other parts are silenced. This is controlled by gene regulatory elements (GREs), such as enhancers and promoters, which act in a cell-type-specific and temporal-specific manner²⁰⁻²². These elements can be found in close proximity to genes, such as promoters on transcriptional start sites (TSSs) or as intronic enhancers. However,

many GREs are located in non-coding stretches of the genome as enhancers and can be located up to 1 million bases away from their target gene (Fig. 1b)^{23,24}.

One of the first GREs to be discovered were enhancers controlling the expression of genes in the immunoglobulin heavy chain locus^{24,25}. Later, more evidence for regulatory elements and their effect on gene expression was found with the discovery of the locus-control-regions (LCR)²⁶. For example, expression in the beta-globin locus was found to be regulated by a single region located upstream of the globin genes²⁷. The importance of this LCR was confirmed in patients lacking a 20 kb region upstream of these genes. While the genes themselves were intact, all patients lacking the control region had no expression of any of the globin genes^{28,29}.

GREs vastly outnumber the known protein coding genes. The expression of our 20,000 genes is controlled by over 400,000 different GREs^{16,18,19,30–33}. However, in any given cell type only a minority of these elements are active^{34–36}. Activity of promoters is typically more pleiotropic compared to that of enhancers^{37,38}, as their activity is directly coupled to the expression of their target gene³⁹. Enhancers, on the other hand, generally act more specifically and regulate gene expression in only a few cell types or developmental stages. In addition, genes can be regulated by different enhancers depending on the cell type or environmental cues³⁴.

GREs function by recruitment and binding of transcription factors (TF) to their loci (Fig. 1b)^{40,41}. On promoters, these TFs can directly interact with the transcription machinery, thereby modulating the transcription rate^{23,42}. However, as enhancers are located relatively far from their target TSS, they need to bring their bound TFs in close proximity in order to affect expression^{43–45}. This is achieved via so-called enhancer-promoter loops (Fig. 1b), which bring the enhancer bound TF-complexes into the vicinity of the transcription machinery, allowing them to regulate expression^{46–48}. These enhancer-promoter loops are dynamically established, allowing enhancers to control target gene expression independently of orientation or distance^{34,49}.

Due to their inherent cell-type-specific and dynamic nature, enhancer detection remains a major challenge. Nevertheless, with recent efforts to identify GREs across multiple tissues and even species, many regulatory elements have now been discovered^{16,18,36–38,50}. For example, GREs have been annotated across 111 different human cell types¹⁸ and 87 different brain regions⁵¹. In addition, active regulatory elements have been identified in liver samples of 20 different mammals³⁷ and across several primate brains³⁸. Combined, these results elucidate how regulatory networks are used throughout the human body and how they differ between species.

Epigenetics and the identification of GREs

The term epigenetics was first described by Conrad Waddington in 1942⁵², in an effort to link genetics to developmental biology and initially describe the effects of genetic processes on human traits and their heritability. Today, the term epigenetics refers to any reversible modification on either the DNA or surrounding proteins that affect gene expression^{53,54}. Although for most of these modifications their biological

function is not fully understood, their regulatory potential seems to reside in either the ability to directly affect the packaging of chromatin or in the ability to recruit or repel specific transcription factors^{55–58}.

The first regulatory elements were found based on their relative sequence conservation across evolution compared to other stretches of non-coding DNA^{14,15}. However, as most GREs are not conserved, their locations on the genome are currently identified based on the pattern of epigenetic modifications on either the DNA or on the supporting histone proteins^{19,59–62} (Fig. 1a). These epigenetic modifications are transiently bound chemical modifications, such as acetyl or methyl groups, that do not alter the underlying DNA sequence. Their attachment is reversible and can have different affects depending on the deposited mark or location^{19,63}. For example, acetylation of a lysine will negate the positive charge of that residue, which reduces the interaction of the histone tail with the nucleosomal DNA and creates less tightly packed chromatin⁶⁴. Thus, creating an optimal environment for other proteins to bind the DNA. On the other hand, trimethylation of the same residue (e.g. H3K27me3) is known to repress genomic activity and has been shown to serve as a docking site to recruit factors that will establish this modification on the neighbouring histones, thereby propagating its own repressive signal⁶⁵. Recent efforts to comprehensively map these epigenetic modifications on a genome-wide scale using chromatin immunoprecipitation (ChIP-Seq)^{37,38}, has revealed that they mark diverse genomic regions^{16,20,21,66}. For example, acetylation of lysine 27 of histone 3 (H3K27ac) marks active regulatory elements such as enhancer and promoters^{20,37,38}, which can be separated by assessment of the methylation status of lysine 4 of histone 3. Mono-methylation of this residue (H3K4me1) marks enhancers while tri-methylation (H3K4me3) is specifically found on active promoters^{19,66}.

While classic ChIP-Seq is one of the most commonly used techniques to positionally map histone modifications on a genome-wide scale^{18,21,33,38,61,67}, several new techniques have been developed in recent years to refine or expand upon this approach, including ATAC-Seq, MNase-ChIP and CUT&RUN^{17,68–71}. ATAC-Seq for example, leverages the open configuration of active genomic regions to identify GREs by incubating samples with a transposase that preferentially cuts DNA in open chromatin regions and directly inserts sequencing adapters⁶⁹. This eliminates the need for mechanical shearing and specific antibodies to enrich for regions of interest, resulting in a much lower number of cells required per assay (e.g. 1 million for ChIP-Seq versus 500 for ATAC-Seq^{72–74}). Moreover, several sequencing assays have recently been developed that can be used to identify regulatory elements with single cell resolution^{31,36,75,76}. Together, these techniques have greatly expanded the catalogue of identified regulatory elements and will likely continue to do so as more rare cell types and species are assessed.

Genetic evolution

Evolution and speciation are driven by DNA mutations that either change

Figure 1. Chromatin structure and regulatory elements. **a**, Schematic overview of the packaging of DNA, with associated histone and DNA modifications. Double stranded DNA is wrapped around nucleosomes at a 150 bp interval, which is further condensed in to chromatin fibers and chromosomes. **b**, Schematic overview of the activation of transcription by an enhancer. In short, specific transcription factors are recruited to enhancer regions and brought into proximity of a gene promoter to activate transcription (TF: transcription factor). **c**, Schematic representation of enhancer divergence during evolution, how this leads to phenotypical differences between species and the underlying genetic changes.

the function of a particular gene product, by causing an amino acid change upon translation, or by changing the regulation of gene expression, by affecting the activity of enhancers and promoters. While the effect of coding evolution is relatively easily understood, by looking at functional changes in the encoded protein⁷⁷, the evolutionary effect of mutations in GREs is harder to dissect as little is known about the functional consequences of DNA mutations in TF-binding sites in regulatory elements (Fig. 1c)⁷⁸.

Since the initial sequencing of the human genome, the genomes of many other species have been resolved, from our closest relative the chimpanzee to more distant

species such as mouse or even budding yeast^{79–81}. This has led to the realization that, while only a small percentage of the genome encodes for proteins, the sequences are well conserved between animals; 98% similarity with chimpanzee, 90% with mouse and 50% with zebrafish^{9,80,82}. Although many sequence variations have been discovered in genes that could explain some of the phenotypic differences between species^{83–85}, many of the identified differences reside in non-coding parts of the genome^{8,86–89} and thus potentially affect distal gene regulatory elements. It has therefore been postulated that changes in gene expression patterns are a driving force of evolution^{90–93}.

Recent studies have indeed shown that enhancers display a high evolutionary turnover^{37,38,94}. For example, while promoters are relatively conserved across evolution, every species deploys around 10,000 specific enhancers³⁷. Nonetheless these changes only mildly impact gene expression compared to more conserved enhancers⁹⁴. This might be explained by the fact that the gain of an enhancer can be offset against the loss of another³⁸, or due to buffering by other elements^{94,95}. Nevertheless, rapid turnover of enhancers may create an ideal playfield for evolution to tinker with expression patterns within a specific cell type without changing the actual proteins.

While regulatory changes across evolution generally have only a minor impact on gene expression, examples of regulatory changes with high impact have also been described.^{87,88} For instance, one enhancer controlling *Pitx1* expression was shown to control pelvic development and to have adaptive activity depending on the living environment of the stickleback fish⁸⁷. Another classic example is an enhancer involved in *Ssh* expression and limb development, which was specifically deleted in the snake lineage and contributed to the loss of limbs in these species⁸⁸. Moreover, a 13 bp insertion within the same enhancer has been identified in patients with polydactyly⁹⁶. These examples indicate how a change in an enhancer can completely alter the development of an organ, without any functional gene change.

The most complex organ to have evolved is the human brain with its cognitive abilities. While it is not yet fully clear what sets our brain apart from those of other non-human primates⁹⁷, many human-specific changes have been identified in non-coding regions, indicating an important role for changes at gene regulatory elements^{38,97–99}. For example, a difference in activity between human and chimpanzee was observed for an enhancer controlling *FZD8* expression⁹⁸, a Wnt pathway receptor implicated in brain development and size. This enhancer was identified due to its human-specific elevated mutation rate and the authors were able to show that the human sequence drove gene expression more extensively than its chimpanzee counterpart. Moreover, mice harbouring the human-specific enhancer element had a 12% increase in brain size and accelerated neural progenitor cell cycle⁹⁸.

Similarly, multiple changes in enhancer activity have been identified in non-coding regions that were ultra-conserved across evolution but showed an elevated mutation rate within the primate or human lineage^{14,15,100–102}. For example, human

accelerated region 1 (HAR1) has been shown to be part of an RNA gene that is specifically expressed in the developing human neocortex⁹⁹. Additionally, recent assessments of differences in GREs between humans and other mammals has identified numerous human-specific regulatory gains and losses in both the brain and other organs^{37,38,103–105}. One of these, HAR87, was shown to be located in a human-specific gene regulatory element associated with *CADM1*³⁸, a gene expressed in purkinje cells and involved in vocalization¹⁰⁶. Overall, these multi-species comparisons revealed that, in accordance with sequence conservation, regulatory activity is more conserved at promoters than at enhancers^{37,38} and that non-coding regulatory changes can be linked to functional differences at a cell-type-specific level^{107–109}. Moreover, it highlights the importance of mapping regulatory changes to fully understand evolutionary processes.

Disease association of regulatory elements

With more and more human genomes sequenced it became clear that although we share most of our DNA sequence, a lot of genetic variation still exists between individuals and that this variation could explain phenotypic differences such as diseasesusceptibility¹¹⁰. These variations range from single nucleotide polymorphisms (SNPs), small insertions and deletions (indels) to larger structural differences such as copy number variations¹¹¹. For example, repeat expansion in the *C9orf72* locus has been linked to ALS^{112,113}, while copy number variation of the *SNCA* gene underlies familial early onset Parkinson's disease^{114,115}. Therefore, in recent years a lot of effort has been put into mapping this variation across human individuals, resulting in large scale project such as the 1000 genomes project, the UK10K consortium and even the complete sequencing of 10,000 human genomes^{116–118}, leading to the realization that every human carries around 8000 unique DNA variants¹¹⁸. However, the functional relevance for most of these variants remains to be addressed.

To link DNA variants to specific traits, massive genome-wide association studies (GWAS) have been performed for many different diseases and phenotypes^{119,120}. To date, more than 3000 GWAS studies have been published, studying an equal number of traits, with the largest cohort including more than 1,000,000 individuals¹²¹. Some of these GWAS have linked genetic variations between human individuals to diseases such as Parkinson's disease¹²² and Alzheimer's disease¹²³ while others assessed traits such as height¹²⁴ or intelligence¹²⁵. These studies have let to extensive meta-analysis and databases, such as GRAPS and the NHGRI-EBI GWAS database^{126,127}, which try to summarize and combine GWAS studies to increase statistical power to identify rare variants or variants with a small effect size. Nevertheless, for most of the identified variants it remains unclear as to how they exactly influence the associated trait or susceptibility to disease, as most reside in non-coding genomic regions¹²⁸ and only explain a minor proportion of the observed variation^{110,129}.

As many variants reside in non-coding regions of the genome, they potentially affect regulatory elements and the integration of GWAS hits with regulatory

information is therefore gaining attention^{130,131}. For example, it has been shown that disease associated SNPs are preferentially located in GREs active in tissues relevant for that disease^{18,132}. A recent study revealed that Parkinson's disease associated SNPs are enriched in brain-specific enhancers while this was not observed for SNPs associated with heart failure or obesity⁵¹. Moreover, functional assessment of one of these SNPs, located in an intronic enhancer, revealed that it altered *SNCA* expression by modulating transcription factor binding and subsequently enhancer activity¹³³. Additionally, a SNP located in the promoter of the *VSTM1* gene, which has been implicated in rheumatoid arthritis and atopic dermatitis^{134,135} was shown to influence expression specifically in monocytes and not in other immune cells¹³⁴. These results demonstrate how non-coding DNA variants can influence phenotypic traits and disease, but also show the importance of assessing these variants in the right cell types. Large scale initiatives such as PsychENCODE and the international epigenome consortium^{136,137} are therefore crucial in order to chart regulatory elements across all human cell types and tissues. Combined with more disease-specific studies^{138–140}, these will be essential to achieve a better understanding on the interplay between genetic variation, regulatory elements and the visible phenotypic outcome.

Thesis outline

As GREs vastly outnumber coding-genes and have crucial roles during development and evolution, mutations at non-coding regions can have detrimental effects causing human disease. In this thesis we have studied GREs in different systems in order to further unravel their role during human evolution and their link to human disease susceptibility.

In **chapter 2**, we focus on the evolution of the brain, specifically the prefrontal cortex and cerebellum. By annotating GREs in the brains of common marmoset we were able to map enhancers across the whole primate branch. We identified a set of regulatory elements that were gained upon the separation of the great-apes from the other primates. We termed these elements hominin-specific gains and showed that they specifically emerged in oligodendrocytes. Surprisingly, we found that a subset of these elements is lost in the brains of autism patients, providing a link between the evolution of the human brain and the susceptibility to neurological disorders. In **chapter 3** we assess GRE evolution on a wider scale, by annotating GREs in different tissues of the common marmoset. Comparison with similar tissues of both human and mouse revealed that the tissue specificity of most elements is conserved between the species, with a subset switching activity between different tissues across evolution. Further analysis of these GREs revealed a link between elements that are evolutionary novel and high interindividual variation. Moreover, novel elements less frequently interact with promoter regions or other GREs indicating a reduced integration in the gene regulatory network. We also observed that GREs that newly emerge during evolution are depleted from trait associated variants, which are more often located in conserved elements. This suggests that the rapid

evolution of enhancers does not greatly affect the value of model species, such as the common marmoset. **Chapter 4** continues with the identified hominin-specific GREs in a more cell-type-specific manner by linking gene expression levels to DNA accessibility in both neuronal, non-neuronal and single cells of the different primate species. Analyzing hominin-specific elements, we observe that accessibility of these regions was obtained at an earlier timepoint in evolution, compared to the hominin-specific H3K27ac enrichment. However, expression of genes is only elevated human and chimpanzees, suggesting that gain of activity is required for gene expression changes. Using single cell data, we confirmed the oligodendrocyte selectivity of these elements, as expression of linked genes was elevated in oligodendrocytes specifically. Moreover, we were able to show that this gain in activity is not due to changes in a specific subset of oligodendrocyte. Finally, in **chapter 5** the results of this thesis are summarized and discussed in a broader context, while providing leads that could help address remaining questions in the field.

References

1. Richmond, T. J. & Davey, C. A. The structure of DNA in the nucleosome core. *Nature* 423, 145–50 (2003).
2. Song, F. et al. Cryo-EM study of the chromatin fiber reveals a double helix twisted by tetranucleosomal units. *Science* (New York, N.Y.) 344, 376–80 (2014).
3. Zhu, P. & Li, G. Structural insights of nucleosome and the 30-nm chromatin fiber. *Current opinion in structural biology* 36, 106–15 (2016).
4. Lander, E. S. et al. Initial sequencing and analysis of the human genome. *Nature* 409, 860–921 (2001).
5. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* 431, 931–45 (2004).
6. Lander, E. S. Initial impact of the sequencing of the human genome. *Nature* 470, 187–197 (2011).
7. Clamp, M. et al. Distinguishing protein-coding and noncoding genes in the human genome. *Proceedings of the National Academy of Sciences* 104, 19428–19433 (2007).
8. Kronenberg, Z. N. et al. High-resolution comparative analysis of great ape genomes. *Science* 360, eaar6343 (2018).
9. Breschi, A., Gingeras, T. R. & Guigó, R. Comparative transcriptomics in human and mouse. *Nature reviews. Genetics* 18, 425–440 (2017).
10. Schnable, P. S. et al. The B73 maize genome: complexity, diversity, and dynamics. *Science* (New York, N.Y.) 326, 1112–5 (2009).
11. Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815 (2000).
12. Gregory, T. R. Synergy between sequence and size in large-scale genomics. *Nature reviews. Genetics* 6, 699–708 (2005).
13. Venter, J. C. et al. The sequence of the human genome. *Science* (New York, N.Y.) 291, 1304–51 (2001).
14. Bejerano, G. et al. Ultraconserved elements in the human genome. *Science* (New York, N.Y.) 304, 1321–5 (2004).
15. Visel, A. et al. Ultraconservation identifies a small subset of extremely constrained developmental enhancers. *Nature genetics* 40, 158–60 (2008).
16. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74 (2012).
17. Andersson, R. et al. An atlas of active enhancers across human cell types and tissues. *Nature* 507, 455–461 (2014).
18. Kundaje, A. et al. Integrative analysis of 111 reference human epigenomes. *Nature* 518, 317–330 (2015).
19. Calo, E. & Wysocka, J. Modification of enhancer chromatin: what, how, and why? *Molecular cell* 49, 825–37 (2013).
20. Creyghton, M. P. et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences of the United States of America* 107, 21931–21936 (2010).
21. Ernst, J. et al. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* 473, 43–9 (2011).
22. FANTOM Consortium and the RIKEN PMI and CLST (DGT) et al. A promoter-level mammalian expression atlas. *Nature* 507, 462–70 (2014).
23. Schoenfelder, S. & Fraser, P. Long-range enhancer-promoter contacts in gene expression control. *Nature reviews. Genetics* (2019). doi:10.1038/s41576-019-0128-0
24. Banerji, J., Rusconi, S. & Schaffner, W. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell* 27, 299–308 (1981).
25. Banerji, J., Olson, L. & Schaffner, W. A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. *Cell* 33, 729–40 (1983).
26. Li, Q., Peterson, K. R., Fang, X. & Stamatoyannopoulos, G. Locus control regions. *Blood* 100, 3077–86 (2002).

27. Grosveld, F., van Assendelft, G. B., Greaves, D. R. & Kollias, G. Position-independent, high-level expression of the human beta-globin gene in transgenic mice. *Cell* 51, 975–85 (1987).
28. Forrester, W. C. et al. A deletion of the human beta-globin locus activation region causes a major alteration in chromatin structure and replication across the entire beta-globin locus. *Genes & development* 4, 1637–49 (1990).
29. Driscoll, M. C., Dobkin, C. S. & Alter, B. P. Gamma delta beta-thalassemia due to a de novo mutation deleting the 5' beta-globin gene activation-region hypersensitive sites. *Proceedings of the National Academy of Sciences of the United States of America* 86, 7470–4 (1989).
30. Kellis, M. et al. Defining functional DNA elements in the human genome. *Proceedings of the National Academy of Sciences of the United States of America* 111, 6131–8 (2014).
31. Cusanovich, D. A. et al. A Single-Cell Atlas of In Vivo Mammalian Chromatin Accessibility. *Cell* 1309–1324 (2018). doi:10.1016/j.cell.2018.06.052
32. Zhu, J. et al. Genome-wide chromatin state transitions associated with developmental and environmental cues. *Cell* 152, 642–54 (2013).
33. Shen, Y. et al. A map of the cis-regulatory sequences in the mouse genome. *Nature* 488, 116–20 (2012).
34. Bonn, S. et al. Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. *Nature Genetics* 44, 148–156 (2012).
35. Preissl, S. et al. Single-nucleus analysis of accessible chromatin in developing mouse forebrain reveals cell-type-specific transcriptional regulation. *Nature neuroscience* 21, 432–439 (2018).
36. Lake, B. B. et al. Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nature biotechnology* 36, 70–80 (2018).
37. Villar, D. et al. Enhancer evolution across 20 mammalian species. *Cell* 160, 554–566 (2015).
38. Vermunt, M. W. et al. Epigenomic annotation of gene regulatory alterations during evolution of the primate brain. *Nature neuroscience* 19, 494–503 (2016).
39. Guenther, M. G., Levine, S. S., Boyer, L. A., Jaenisch, R. & Young, R. A. A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* 130, 77–88 (2007).
40. Spitz, F. & Furlong, E. E. M. Transcription factors: from enhancer binding to developmental control. *Nature reviews. Genetics* 13, 613–26 (2012).
41. Lettice, L. A. et al. Opposing functions of the ETS factor family define Shh spatial expression in limb buds and underlie polydactyly. *Developmental cell* 22, 459–67 (2012).
42. de Jonge, W. J. et al. Molecular mechanisms that distinguish TFIID housekeeping from regulatable SAGA promoters. *The EMBO journal* 36, 274–290 (2017).
43. Vermunt, M. W., Zhang, D. & Blobel, G. A. The interdependence of gene-regulatory elements and the 3D genome. *The Journal of Cell Biology* jcb.201809040 (2018). doi:10.1083/jcb.201809040
44. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. *Science (New York, N.Y.)* 295, 1306–11 (2002).
45. Rao, S. S. P. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159, 1665–80 (2014).
46. Chen, H. et al. Dynamic interplay between enhancer-promoter topology and gene activity. *Nature genetics* 50, 1296–1303 (2018).
47. Deng, W. et al. Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell* 149, 1233–44 (2012).
48. Tolhuis, B., Palstra, R. J., Splinter, E., Grosveld, F. & de Laat, W. Looping and interaction between hypersensitive sites in the active beta-globin locus. *Molecular cell* 10, 1453–65 (2002).
49. Nord, A. S. et al. Rapid and pervasive changes in genome-wide enhancer usage during mammalian development. *Cell* 155, 1521–1531 (2013).
50. Fullard, J. F. et al. An atlas of chromatin accessibility in the adult human brain. *Genome research* 28, 1243–1252 (2018).
51. Vermunt, M. W. et al. Large-scale identification of coregulated enhancer networks in the adult human brain. *Cell Reports* 9, 767–779 (2014).
52. Waddington, C. H. CANALIZATION OF DEVELOPMENT AND THE INHERITANCE OF ACQUIRED CHARACTERS. *Nature* 150, 563–565 (1942).
53. Holliday, R. Epigenetics: a historical overview. *Epigenetics* 1, 76–80 (2006).

54. Felsenfeld, G. A brief history of epigenetics. *Cold Spring Harbor perspectives in biology* 6, 45–50 (2014).
55. Ruthenburg, A. J., Li, H., Patel, D. J. & Allis, C. D. Multivalent engagement of chromatin modifications by linked binding modules. *Nature reviews. Molecular cell biology* 8, 983–94 (2007).
56. Wysocka, J. et al. WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. *Cell* 121, 859–72 (2005).
57. Schneider, R., Bannister, A. J., Weise, C. & Kouzarides, T. Direct binding of INHAT to H3 tails disrupted by modifications. *The Journal of biological chemistry* 279, 23859–62 (2004).
58. Manohar, M. et al. Acetylation of histone H3 at the nucleosome dyad alters DNA-histone binding. *The Journal of biological chemistry* 284, 23312–21 (2009).
59. Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* 403, 41–5 (2000).
60. Heintzman, N. D. et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature genetics* 39, 311–8 (2007).
61. Heintzman, N. D. et al. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* 459, 108–12 (2009).
62. Stricker, S. H., Köferle, A. & Beck, S. From profiles to function in epigenomics. *Nature reviews. Genetics* 18, 51–66 (2017).
63. Zhao, Y. & Garcia, B. A. Comprehensive Catalog of Currently Documented Histone Modifications. *Cold Spring Harbor perspectives in biology* 7, a025064 (2015).
64. Zentner, G. E. & Henikoff, S. Regulation of nucleosome dynamics by histone modifications. *Nature structural & molecular biology* 20, 259–66 (2013).
65. Laugesen, A., Højfeldt, J. W. & Helin, K. Molecular Mechanisms Directing PRC2 Recruitment and H3K27 Methylation. *Molecular cell* 74, 8–18 (2019).
66. Zhou, V. W., Goren, A. & Bernstein, B. E. Charting histone modifications and the functional organization of mammalian genomes. *Nature reviews. Genetics* 12, 7–18 (2011).
67. Pennisi, E. ENCODE project writes eulogy for junk DNA. *Science (New York, N.Y.)* 337, 1159, 1161 (2012).
68. Arnold, C. D. et al. Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science (New York, N.Y.)* 339, 1074–7 (2013).
69. Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Current protocols in molecular biology* 109, 21.29.1–9 (2015).
70. Zaret, K. Micrococcal nuclease analysis of chromatin structure. *Current protocols in molecular biology* Chapter 21, Unit 21.1 (2005).
71. Jung, I. et al. A compendium of promoter-centered long-range chromatin interactions in the human genome. *Nature genetics* 51, 1442–1449 (2019).
72. Landt, S. & Marinov, G. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome ...* 1813–1831 (2012). doi:10.1101/gr.136184.111.
73. Kim, T. H. & Ren, B. Genome-wide analysis of protein-DNA interactions. *Annual review of genomics and human genetics* 7, 81–102 (2006).
74. Corces, M. R. et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nature Methods* 14, 959–962 (2017).
75. Shema, E., Bernstein, B. E. & Buenrostro, J. D. Single-cell and single-molecule epigenomics to uncover genome regulation at unprecedented resolution. *Nature Genetics* 51, 19–25 (2019).
76. Buenrostro, J. D. et al. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* 523, 486–90 (2015).
77. Enard, W. et al. Molecular evolution of FOXP2, a gene involved in speech and language. *Nature* 418, 869–72 (2002).
78. Franchini, L. F. & Pollard, K. S. Human evolution: The non-coding revolution. *BMC Biology* 15, 1–12 (2017).
79. Mouse Genome Sequencing Consortium et al. Initial sequencing and comparative analysis of the mouse genome. *Nature* 420, 520–62 (2002).
80. Mikkelsen, T. S. et al. Initial sequence of the chimpanzee genome and comparison with the human

- genome. *Nature* 437, 69–87 (2005).
81. Peter, J. et al. Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature* 556, 339–344 (2018).
82. Howe, K. et al. The zebrafish reference genome sequence and its relationship to the human genome. *Nature* 496, 498–503 (2013).
83. Worley, K. C. et al. The common marmoset genome provides insight into primate biology and evolution. *Nature Genetics* 46, 850–857 (2014).
84. Kukekova, A. v. et al. Red fox genome assembly identifies genomic regions associated with tame and aggressive behaviours. *Nature Ecology and Evolution* 2, 1479–1491 (2018).
85. Fischer, J. & Hammerschmidt, K. Ultrasonic vocalizations in mouse models for speech and socio-cognitive disorders: insights into the evolution of vocal communication. *Genes, brain, and behavior* 10, 17–27 (2011).
86. Trizzino, M. et al. Transposable element exaptation is the primary source of novelty in the primate gene regulatory landscape. *Genome research* 083980 (2016). doi:10.1101/083980
87. Chan, Y. F. et al. Adaptive Evolution of Pelvic Reduction of a Pitx1 Enhancer. *Science* 327, 302–305 (2010).
88. Kvon, E. Z. et al. Progressive Loss of Function in a Limb Enhancer during Snake Evolution. *Cell* 167, 633–642.e11 (2016).
89. Doan, R. N. et al. Mutations in Human Accelerated Regions Disrupt Cognition and Social Behavior. *Cell* 167, 341–354.e12 (2016).
90. Reilly, S. K. & Noonan, J. P. Evolution of Gene Regulation in Humans. *Annual review of genomics and human genetics* 17, 45–67 (2016).
91. Khaitovich, P., Enard, W., Lachmann, M. & Pääbo, S. Evolution of primate gene expression. *Nature reviews. Genetics* 7, 693–702 (2006).
92. Cotney, J. et al. The evolution of lineage-specific regulatory activities in the human embryonic limb. *Cell* 154, 185–96 (2013).
93. Nowick, K., Gernat, T., Almaas, E. & Stubbs, L. Differences in human and chimpanzee gene expression patterns define an evolving network of transcription factors in brain. *Proceedings of the National Academy of Sciences of the United States of America* 106, 22358–63 (2009).
94. Berthelot, C., Villar, D., Horvath, J. E., Odom, D. T. & Flicek, P. Complexity and conservation of regulatory landscapes underlie evolutionary resilience of mammalian gene expression. *Nature Ecology and Evolution* 2, 152–163 (2018).
95. Osterwalder, M. et al. Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature* 554, 239–243 (2018).
96. Laurell, T. et al. A novel 13 base pair insertion in the sonic hedgehog ZRS limb enhancer (ZRS/LMBR1) causes preaxial polydactyly with triphalangeal thumb. *Human mutation* 33, 1063–6 (2012).
97. Sousa, A. M. M., Meyer, K. A., Santpere, G., Gulden, F. & Sestan, N. Evolution of the Human Nervous System Function, Structure and Development. 2–7 (2017). doi:10.1007/978-0-387-76978-3
98. Boyd, J. L. et al. Human-chimpanzee differences in a FZD8 enhancer alter cell-cycle dynamics in the developing neocortex. *Current Biology* 25, 772–779 (2015).
99. Pollard, K. S. et al. An RNA gene expressed during cortical development evolved rapidly in humans. *Nature* 443, 167–72 (2006).
100. Bird, C. P. et al. Fast-evolving noncoding sequences in the human genome. *Genome biology* 8, R118 (2007).
101. Prabhakar, S., Noonan, J. P., Pääbo, S. & Rubin, E. M. Accelerated evolution of conserved noncoding sequences in humans. *Science (New York, N.Y.)* 314, 786 (2006).
102. Lindblad-Toh, K. et al. A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* 478, 476–82 (2011).
103. Zhou, X. et al. Epigenetic modifications are associated with inter-species gene expression variation in primates. *Genome Biology* 15, 1–19 (2014).
104. Reilly, S. K. et al. Evolutionary genomics. Evolutionary changes in promoter and enhancer activity during human corticogenesis. *Science (New York, N.Y.)* 347, 1155–9 (2015).
105. Chen, H., Li, C., Zhou, Z. & Liang, H. Fast-Evolving Human-Specific Neural Enhancers Are

- Associated with Aging-Related Diseases. *Cell Systems* 6, 604-611.e4 (2018).
106. Fujita, E., Tanabe, Y., Imhof, B. A., Momoi, M. Y. & Momoi, T. Cadm1-expressing synapses on Purkinje cell dendrites are involved in mouse ultrasonic vocalization activity. *PloS one* 7, e30151 (2012).
107. Kozlenkov, A. et al. A unique role for DNA (hydroxy)methylation in epigenetic regulation of human inhibitory neurons. *Science Advances* 4, 1–12 (2018).
108. Zhu, Y. et al. Spatiotemporal transcriptomic divergence across human and macaque brain development. *Science* 362, eaat8077 (2018).
109. Xu, C. et al. Human-specific features of spatial gene expression and regulation in eight brain regions. *Genome research* 28, 1097–1110 (2018).
110. Manolio, T. A. et al. Finding the missing heritability of complex diseases. *Nature* 461, 747–753 (2009).
111. 1000 Genomes Project Consortium et al. A global reference for human genetic variation. *Nature* 526, 68–74 (2015).
112. Iacoangeli, A. et al. C9orf72 intermediate expansions of 24-30 repeats are associated with ALS. *Acta neuropathologica communications* 7, 115 (2019).
113. Balendra, R. & Isaacs, A. M. C9orf72-mediated ALS and FTD: multiple pathways to disease. *Nature reviews. Neurology* 14, 544–558 (2018).
114. Fuchs, J. et al. Phenotypic variation in a large Swedish pedigree due to SNCA duplication and triplication. *Neurology* 68, 916–22 (2007).
115. Konno, T., Ross, O. A., Puschmann, A., Dickson, D. W. & Wszolek, Z. K. Autosomal dominant Parkinson's disease caused by SNCA duplications. *Parkinsonism & related disorders* 22 Suppl 1, S1-6 (2016).
116. 1000 Genomes Project Consortium et al. A map of human genome variation from population-scale sequencing. *Nature* 467, 1061–73 (2010).
117. UK10K Consortium et al. The UK10K project identifies rare variants in health and disease. *Nature* 526, 82–90 (2015).
118. Telenti, A. et al. Deep sequencing of 10,000 human genomes. *Proceedings of the National Academy of Sciences of the United States of America* 113, 11901–11906 (2016).
119. Mills, M. C. & Rahal, C. A scientometric review of genome-wide association studies. *Communications biology* 2, 9 (2019).
120. Visscher, P. M. et al. 10 Years of GWAS Discovery: Biology, Function, and Translation. *American journal of human genetics* 101, 5–22 (2017).
121. Lee, J. J. et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature genetics* 50, 1112–1121 (2018).
122. Chang, D. et al. A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci. *Nature Genetics* 49, 1511–1516 (2017).
123. Lambert, J. C. et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nature genetics* 45, 1452–8 (2013).
124. Marouli, E. et al. Rare and low-frequency coding variants alter human adult height. *Nature* 542, 186–190 (2017).
125. Savage, J. E. et al. Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nature genetics* (2018). doi:10.1038/s41588-018-0152-6
126. Buniello, A. et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic acids research* 47, D1005–D1012 (2019).
127. Leslie, R., Donnell, C. J. O. & Johnson, A. D. GRASP : analysis of genotype – phenotype results from 1390 genome-wide association studies and corresponding open access database. 30, 185–194 (2014).
128. Albert, F. W. & Kruglyak, L. The role of regulatory variation in complex traits and disease. *Nature reviews. Genetics* 16, 197–212 (2015).
129. Hindorff, L. A. et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proceedings of the National Academy of Sciences of the United States of America* 106, 9362–7 (2009).
130. Boyle, E. A., Li, Y. I. & Pritchard, J. K. An Expanded View of Complex Traits: From Polygenic to

- Omnigenic. *Cell* 169, 1177–1186 (2017).
131. Zhou, J. & Troyanskaya, O. G. Predicting effects of noncoding variants with deep learning-based sequence model. *Nature Methods* 12, 931–934 (2015).
 132. Maurano, M. et al. Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. *Science* 337, 1190–1195 (2012).
 133. Soldner, F. et al. Parkinson-associated risk variant in distal enhancer of α -synuclein modulates target gene expression. *Nature* 533, 95–99 (2016).
 134. Kumar, D. et al. A functional SNP associated with atopic dermatitis controls cell type-specific methylation of the VSTM1 gene locus. *Genome medicine* 9, 18 (2017).
 135. Wang, D., Li, Y., Liu, Y., He, Y. & Shi, G. Expression of VSTM1-v2 Is Increased in Peripheral Blood Mononuclear Cells from Patients with Rheumatoid Arthritis and Is Correlated with Disease Activity. *PloS one* 11, e0146805 (2016).
 136. Stunnenberg, H. G., International Human Epigenome Consortium & Hirst, M. The International Human Epigenome Consortium: A Blueprint for Scientific Collaboration and Discovery. *Cell* 167, 1145–1149 (2016).
 137. Wang, D. et al. Comprehensive functional genomic resource and integrative model for the human brain. *Science* 362, (2018).
 138. Sun, W. et al. Histone Acetylome-wide Association Study of Autism Spectrum Disorder. *Cell* 167, 1385–1397.e11 (2016).
 139. Marzi, S. J. et al. A histone acetylome-wide association study of Alzheimer’s disease identifies disease-associated H3K27ac differences in the entorhinal cortex. *Nature neuroscience* 21, (2018).
 140. Corces, M. R. et al. The chromatin accessibility landscape of primary human cancers. *Science (New York, N.Y.)* 362, (2018).

Chapter 2

Hominin-specific regulatory elements selectively emerged in oligodendrocytes and are disrupted in autism patients

Bas Castelijns¹, Mirna L. Baak¹, Ilia S. Timpanaro¹, Caroline R.M. Wiggers^{1,2},
Marit W. Vermunt¹, Peng Shang¹, Ivanela Kondova³, Geert Geeven¹,
Valerio Bianchi¹, Wouter de Laat¹, Niels Geijsen¹ and Menno P. Creyghton^{1,4}

¹Hubrecht Institute-KNAW & University Medical Center Utrecht,
Utrecht, The Netherlands

²Division of Pediatrics, University Medical Center Utrecht,
Utrecht, The Netherlands

³Biomedical Primate Research Center,
Rijswijk, The Netherlands

⁴Erasmus University Medical Center, Department of Developmental Biology,
Rotterdam, The Netherlands

Abstract

Speciation is associated with substantial rewiring of the regulatory circuitry underlying the expression of genes. Determining which changes are relevant and underlie the emergence of the human brain or its unique susceptibility to neural disease has been challenging. Here we annotate changes to gene regulatory elements (GREs) at cell type resolution in the brains of multiple primate species spanning most of primate evolution. We identify a unique set of regulatory elements that emerged in hominins prior to the separation of humans and chimpanzees. We demonstrate that these hominin gains preferentially affect oligodendrocyte function postnatally and are preferentially affected in the brains of autism patients. This preference is also observed for human-specific GREs suggesting this system is under continued selective pressure. Our data provide a roadmap of regulatory rewiring across primate evolution providing insight into the genomic changes that underlie the emergence of the brain and its susceptibility to neural disease.

Introduction

Understanding the emergence of the human brain and the unique properties that define our species in evolutionary history remains a major challenge¹. Furthermore, as several neuropsychiatric and neurodegenerative diseases are suspected to be linked to genetic changes that recently evolved²⁻⁵, unraveling evolution of the human brain may have consequences beyond a plain understanding of the human condition. However, as the human brain is the result of a process that spans the entirety of primate evolution, giving rise to primate brains of variable sizes and cognitive complexities^{1,6}, its understanding may also require a broader evolutionary context. This is especially relevant given the absence of pervasive data on neural disorders in great apes⁷ and the challenges in assessing their cognitive abilities⁸.

The process underlying brain development is controlled by gene expression programs that dictate cellular identity in a spatio-temporal manner^{9,10}. Gene regulatory elements (GREs), such as enhancers and promoters, function as transcriptional units that recruit specific transcription factor complexes to control the expression of genes in a cell type dependent manner¹¹. This cell specificity as well as functional redundancy of enhancers is likely linked to an enhanced evolutionary flexibility, avoiding some of the pleiotropic effects associated with changes in genes¹². As such, pervasive rewiring of the gene regulatory circuitry has been observed across mammalian evolution while genes remain mostly conserved¹³. This rewiring includes a host of regulatory changes that were specific to the human brain¹⁴⁻¹⁶. Despite these efforts, finding a cause-consequence relationship between important evolutionary alterations in gene regulatory elements, evolution of the brain and its susceptibility to neural disorders has remained difficult. Here we annotate regulatory changes in the brain across primate evolution. We identify a set of regulatory changes that

emerged after the separation from old world monkeys but prior to the separation between chimpanzee and human. These elements are referred to as hominin-specific and are preferentially enriched in oligodendrocytes in adults and deregulated in the brains of Autism Spectrum Disorder (ASD) patients. We propose that evolution of regulatory DNA in hominins may have helped set the stage for the emergence of the human brain and its susceptibility to disease.

Results

Annotation of hominin-specific regulatory change in brain

To track relevant genome changes in the brain across primate evolution, we annotated GREs in brain tissue from three healthy marmoset specimens (Supplementary Data 1). With a common ancestor living approximately 44 million years ago¹⁷, the divergence of marmosets and humans spans a major part of primate evolution (Fig. 1a). We analyzed prefrontal cortex (PFC) a key anatomical region involved in many of the executive processes that define our species and cerebellum (CB) a neuron dense structure harboring mainly granule neurons¹⁸. To identify active GREs, we used ChIP-Seq for histone H3 lysine 27 acetylation (H3K27ac), a robust assay to identify active GREs on a global scale¹⁹. In addition, we analyzed trimethylation of histone H3 lysine 4 (H3K4me3), which is specifically associated with active transcriptional start sites (TSS)²⁰ (Supplementary Data 3). Data were well within quality standards²¹ and were reproducible between biological replicates ($r > 0.9$, Supplementary Fig. 1a, b). We identified 33,754 marmoset GREs of which 16,086 were predicted promoter GREs and 17,668 predicted enhancer GREs (Supplementary Fig. 1c). Using human annotated promoter sequences, we observed that the majority (75%) of predicted promoters in marmoset overlapped with annotated promoters in humans. Consistent with previous observations, analyses of RNA-Seq data confirmed that active GREs were more often associated with active gene expression in marmoset brain (Supplementary Fig. 1d)^{22,23}.

To assess regulatory changes across primate evolution, we focused on H3K27ac enrichment and compared our data to active GREs identified in rhesus macaque, chimpanzee and human in PFC and CB (Supplementary Fig. 2a, b)¹⁴. Only GREs that could be consistently mapped onto all four genomes were included in the analyses (Supplementary Fig. 2c-e). While this excludes species-specific DNA, most GREs that were excluded were discarded due to poor genome annotation and/or ambiguous mapping of reads. This is consistent with the observation that regulatory changes primarily occur in conserved DNA as opposed to DNA that is evolutionary novel¹³. We retained a total of 37,308 GREs that could be mapped on all four species of which 25% overlapped a TSS in humans (Supplementary Data 4). Dimension reduction and visualization with t-SNE and hierarchical clustering revealed a clear separation of the two anatomical locations as well as the major primate clades, with high correlation between replicate samples (Fig. 1b, c, Supplementary Fig. 2f).

Fig. 1 Identification of hominin-specific regulatory changes in brain tissue. **a**, Schematic representation of the primate phylogenetic tree with time indication of the major branch points. **b**, t-Distributed Stochastic Neighbour Embedding (t-sne) analysis of all H3K27ac enriched GREs with orthologs on all four primate genomes ($n = 37,308$). Axes indicate semantic space (CB: cerebellum, PFC: prefrontal cortex). **c**, PCA analysis of the same regions as in (b), shown for the first two principal components. **d**, Heatmap showing H3K27ac enrichment on scaled hominin-specific regulatory changes in both cerebellum and prefrontal cortex. Heatmap colors indicate H3K27ac enrichment (rpm). CB gains $n = 713$; CB losses $n = 374$; PFC gains $n = 685$; PFC losses $n = 158$. **e**, Conservation scores (20 mammals) using PhastCon as defined by UCSC for hominin-specific gains and evolutionary stable GREs compared to all here identified GREs. Dissimilarities between distributions were calculated using a Student's t-test. **f**, Analysis as in (e) for hominin-specific nucleotide changes. Bottom and top of the box plots are the first and third quartile. The line within the boxes represents the median and whiskers denote interval within 1.5x the interquartile range from the median, outliers are depicted as points. Source data are provided in Source Data file.

While a prior analysis focused on identifying regulatory changes specific to the human brain¹⁴, significant differences in brain size as well as the emergence of complex behavior has also occurred prior to the separation of humans and chimpanzee in great apes¹⁶. To gain insight into the regulatory changes occurring prior to human evolution, we first selected elements that were differentially enriched between human and both marmoset and rhesus macaque using DESeq2 as demonstrated previously¹⁴ (Supplementary Fig. 3a). The same analysis was performed using chimpanzee data instead of human data and the resulting datasets were compared (Supplementary Fig. 3b). Similar to our prior analysis¹⁴, biological replicates were generated in separate batches to ameliorate batch related effects and no major batch effects were observed (Supplementary Fig. 3c-e). We found 1398 (713 CB, 685 PFC) regions that were designated as hominin (humans and chimpanzee)-specific gains and 532 (374 CB, 158 PFC) that were defined as hominin-specific losses (Fig. 1d, Supplementary Fig. 4a). For instance, several hominin-specific regulatory changes were found close to the *SORCS1* gene in CB (Supplementary Fig. 4b), which is a known regulator of synaptic trafficking and linked to aggression^{24,25}. Mutations in this gene have been linked to Alzheimer's disease (AD) as well as ASD²⁵. The observed imbalance between the hominin-specific gains and losses is likely due to hominin losses as defined here require conservation of activity across a longer evolutionary period (i.e. conservation between rhesus macaque and marmoset). As such, these conserved regions are more likely to be biologically relevant and thus their loss may be selected against. Hominin-specific gains were less frequently shared with other tissues (Supplementary Fig. 4c), consistent with previous observations suggesting that evolutionary changes occur preferentially in tissue-specific GREs¹⁴. They were also less frequently associated with promoters consistent with an enhanced evolutionary flexibility at distal enhancers (Supplementary Data 4)¹³. More regions were found differentially enriched in CB compared to PFC which is consistent with previous ChIP and gene expression analysis^{14,23} and likely due to CB being more homogeneous resulting in better resolved GREs that are easier to compare. Motif analysis did not yield a clear enrichment for transcription factor binding sites at these elements, suggesting their change was not directly linked to a single transcriptional program that was altered (Supplementary Data 5). This was further supported by an overall reduction in primate sequence conservation at these elements compared to elements that did not change activity during primate evolution (Fig. 1e). Furthermore, we observed an increase of hominin-specific nucleotide changes at these GREs, suggesting that the observed regulatory changes are sequence based (Fig. 1f). Thus, our analysis has uncovered a unique set of regulatory elements that were conserved in monkeys but changed activity in hominins.

Hominin PFC gains preferentially emerged in oligodendrocytes

As the cortex is highly heterogeneous, containing a variety of neural and glial cell types we asked whether GREs that recently evolved in hominins were spatially

Fig. 2 Hominin-specific gains are oligodendrocyte-specific. **a**, Bar plot showing cell-type-specific ATAC peaks in hominin-specific gains in PFC and CB and stable GREs. Difference in oligodendrocyte-specific open chromatin frequency was calculated using a Fisher's exact test. **b**, Heatmap showing H3K27ac enrichment for 685 scaled hominin-specific PFC gains, analyzed in different cell types as indicated. Heatmap colors indicate H3K27ac enrichment (rpm). **c**, Metaplot analysis showing the average H3K27ac enrichment profile in different cell types for hominin-specific PFC gains, stable and all GREs. **d**, Box plot showing normalized ATAC signal for hominin-specific gains and stable GREs in nuclei FACS sorted for NeuN. Dissimilarities between distributions were calculated using a Student's t-test. **e**, Box plots showing normalized H3K27ac enrichment for hominin-specific gains and stable GREs in prefrontal cortex in both human and rhesus macaque. Dissimilarities between distributions were calculated using a Student's t-test. **f**, Box plots showing normalized H3K27ac enrichment for hominin-specific PFC gains in white matter tissue per species. Dissimilarity between distributions was calculated using a Student's t-test. **g**, Box plots as in (f) but for normalized H3K27ac enrichment on hominin-specific CB gains. Bottom and top of all box plots are the first and third quartile. The line within the boxes represents the median and whiskers denote interval within 1.5x the interquartile range from the median, outlier are depicted as points. Source data are provided in Source Data file.

confined to the frontal cortex or rather restricted to a particular cell type. We therefore analyzed single cell ATAC-Seq data generated in human PFC²⁶, and found that hominin-specific gains were overwhelmingly overrepresented in oligodendrocyte-specific open chromatin domains in PFC (Fig. 2a). To independently confirm this observation, we analyzed H3K27ac data derived from FACS sorted NeuN+ glutamatergic neurons, Sox6+ GABAergic neurons and Sox10+ oligodendrocytes²⁷, FACS sorted CD11b+ CD45^{low} CD64+ CXCR1^{high} microglia²⁸ as well as primary astrocytes isolated from adult brain²⁹ (Supplementary Fig. 5a). We found that

most of the GREs that evolved in the PFC of hominins were selectively enriched in Sox10⁺ oligodendrocytes and not in other neural or glial cell types (Fig. 2b, c, Supplementary Fig. 5b). Hominin-specific gains in cerebellum did not show strong enrichment for cell-type-specific regions consistent with granule neurons not being represented by these data (Supplementary Fig. 5c). As oligodendrocytes represent the main constituent of glial cells in the brain, we used a third independent dataset³⁰ separating neurons and glial cells in PFC based on the expression of NeuN which selectively labels neural nuclei on the nuclear membrane and not glial nuclei. We confirmed that hominin-specific gains in PFC were selectively enriched in glial cells (Supplementary Fig. 5d, e).

As a modest increase in glia to neuron ratio, scaling with brain volume, was observed in larger primates^{31–33}, we wondered whether the link between regulatory gains and oligodendrocytes was the result of an altered glial content in the cortex. To address this, we generated ATAC-Seq data from human, chimpanzee, rhesus macaque and marmoset frontal cortex resolved between neuronal and glial cells by FANS sorting for NeuN (Supplementary Fig. 6a–c, Supplementary Data 3). Consistent with cell-type-specific H3K27ac enrichment and the single-cell ATAC-Seq data³⁴, hominin-specific PFC gains were enriched for open chromatin in NeuN[–] nuclei compared to NeuN⁺ nuclei in humans (Fig. 2d). The opposite was observed for hominin-specific CB gains which is consistent with an overwhelming neural content in CB (Fig. 2d). In addition, we found that ATAC-Seq signal at hominin-specific PFC gains was enhanced in humans and chimpanzee only in NeuN[–] nuclei and not in NeuN⁺ nuclei (Supplementary Fig. 6d). Moreover, we observed a slight increase in H3K27ac enrichment in regions specific to NeuN[–] nuclei in humans compared to rhesus macaque, however this was substantially less pronounced than the differences observed for the regions identified as hominin-specific gains (Supplementary Fig. 6e). This demonstrates that regions that gain H3K27ac enrichment in hominins are also preferentially in an open configuration in hominins. Furthermore, these data suggest that the skewing of hominin-specific gains towards oligodendrocytes is not solely due to an altered neural to glial ratio. To further exclude altered glial content as a factor in our analysis, we generated and analyzed H3K27ac ChIP-Seq data in white matter (WM) for all four primate species (Supplementary Fig. 6f–g, Supplementary Data 3). WM lacks neural content and consists primarily (~75%) of oligodendrocytes³². Hominin-specific gains in PFC also gained activity in human and chimpanzee WM compared to rhesus and marmoset, further excluding glial content as the main factor in our analysis (Fig. 2f). Hominin-specific gains identified in CB were not differentially enriched in WM (Fig. 2g). This strongly argues against these results emerging from an altered neural to glial ratio in PFC. Furthermore, it also suggests that the differences at these GREs are not spatially confined to the PFC. Thus, our data demonstrates that hominin-specific regulatory changes associate with hominin-specific sequence changes, preferentially affecting GREs in oligodendrocytes.

Hominin gains activate postnatally and link to myelination

To analyze whether there was temporal skewing of hominin-specific gains towards a particular developmental stage, we analyzed H3K27ac data from distinct developmental timepoints covering prenatal, postnatal and adult stages in humans³⁵. We found that hominin-specific PFC gains preferentially associate with regulatory DNA that activates between the postnatal and adult stages (Fig. 3a, Supplementary Fig. 7a) which, given their enrichment in oligodendrocytes, is consistent with increased myelination during this developmental stage¹. This was not due to our initial analysis being limited to adult tissue, as GREs identified here were overall not differentially enriched between the postnatal and adult stages (Fig. 3a).

To link hominin-specific gains to their target genes we used a combination of proximity-based gene linkage as well as HiC derived enhancer-promoters links from analysis in PFC³⁶. We used promoter locations based on the human genome to link GREs to genes for all primates as it the best annotated genome and promoter locations are largely conserved throughout primate evolution (Supplementary Fig. 7b, c). Using independent RNA-Seq data generated in PFC and CB for humans and rhesus macaque^{15,23}, we found that the expression changes of genes linked to hominin-specific gains correlated with H3K27ac gains in CB and PFC in humans (Supplementary Fig. 7d). In addition, we observed that these gene expression differences also predominantly occurred postnatally in both PFC and CB (Fig. 3b, Supplementary Fig. 7e) supporting the post-natal selectivity of hominin-specific gains. To verify that the expression changes for genes linked to hominin-specific PFC gains were consistent in glial cells, further excluding changes in glia to neural content as a factor in these observations, we generated WM RNA-Seq data from the four primate species (Supplementary Fig. 7f, g, Supplementary Data 3). In agreement with the changes at GREs, hominin-specific gains at regulatory DNA were mirrored by gene expression changes in WM (Fig. 3c, Supplementary Fig. 7h).

Functional analysis of genes linked to hominin-specific PFC changes revealed that they were enriched for genes involved Schwann cell differentiation, a myelinating cell of the nervous system further supporting oligodendrocyte identity (Fig. 3d, Supplementary Data 6). Furthermore, we found hominin-specific H3K27ac enrichment near genes involved in the regulation of axon guidance and sialylation while this was not observed for hominin-specific gains in cerebellum (Fig. 3d, Supplementary Fig. 7i). For instance, hominin-specific PFC gains were located near the sialyltransferase *ST3GAL5* (Supplementary Fig. 7j), which causes intellectual disability when mutated³⁷. Sialic acid is a key monosaccharide for the synthesis of brain gangliosides and sialylated glycoproteins (NCAMS), both of which are essential for cognitive stability and brain development³⁸. While sialic acid is especially enriched in brain tissue, it is also found in high concentrations in human milk and believed to play a role in postnatal synaptic plasticity and memory formation³⁹. Indeed, defects in these biosynthesis pathways have been linked to axon instability, altered nerve excitability and psychiatric disorders including schizophrenia and ASD³⁸. This may

Fig. 3 Hominin-specific gains emerge postnatally. **a**, Box plots showing normalized H3K27ac enrichment in PFC or CB samples across different developmental timepoints for hominin-specific gains, stable GREs and all primate GREs. Dissimilarities between distributions were calculated using a Student's t-test. **b**, Box plots showing normalized gene expression values of genes linked to hominin-specific PFC gains across developmental time, for both human and rhesus macaque. Dissimilarities between distributions were calculated using a Student's t-test. **c**, Box plots showing normalized gene expression values in the white matter tissue of different primates for genes linked to hominin-specific PFC changes. Dissimilarities between distributions were calculated using a Student's t-test. **d**, Gene-ontology analysis using GREAT for hominin-specific PFC gains. P-values represent significance of enrichment for stated biological process. Bottom and top of all box plots are the first and third quartile. The line within the boxes represents the median and whiskers denote interval within 1.5x the interquartile range from the median, outliers are depicted as points.

suggest that some regulatory features of these disorders link to evolutionary changes that emerged in great apes and predates the separation of humans and chimpanzees.

PFC hominin gains are deregulated in ASD patient brain

As a link between the evolution of the human brain and the emergence of several neural disorders was proposed previously²⁻⁵ we further explored the potential link between regulatory changes emerging in hominins and disorders of the brain. We first compared hominin-specific regulatory changes to genome-wide association (GWAS) data, to link nucleotide variants associated with AD⁴⁰, ASD⁴¹, Parkinson's disease⁴², bipolar disorder and schizophrenia⁴³ using stratified LD score regression on the summary statistics of these variants⁴³. While enrichment for heritability to schizophrenia was observed at GREs, particularly those that were evolutionary stable and promoters (Supplementary Fig. 8a), we did not observe a significant enrichment for disease associated variants in GREs that were gained or lost in

Fig. 4 Hominin-specific gains are selectively deregulated in autism patient brains. **a**, Heatmap showing enrichment of disease-associated GREs deregulated in ASD and AD patient brains correlating with hominin-specific changes found in cerebellum (CB) and prefrontal cortex (PFC). Color indicates odds ratio of enrichment compared to all identified GREs. P-values were determined using a Fisher's exact test. **b**, ChIP-seq track showing rpm normalized H3K27ac enrichment across regions containing the *CAMK2A* (left panel) and *DNMT3A* (right panel) gene in human PFC. Hominin-specific PFC gains are highlighted in green. Chromosomal interactions are indicated by the boxes below the panel. **c**, ChIP-seq tracks as in (a), for the *DNMT3A* gene in human PFC and different isolated cell types. A hominin-specific PFC gain is highlighted in green. (PFC: prefrontal cortex, Olig: oligodendrocytes, Glut: glutamatergic neurons, Gaba: GABAergic neurons, Astr: astrocytes, Micgl: microglia). **d**, ChIP-seq tracks as in (b) for a region containing the *c-MET* gene in human PFC and different isolated cell types. Two hominin-specific PFC gains

hominins (Supplementary Fig. 8a, b).

To contrast our data to more direct evidence for GRE deregulation in neural disease, we also analyzed H3K27ac ChIP-Seq data across 45 autism patient brains and 49 control samples, similarly covering PFC and CB⁴⁴ as well as H3K27ac ChIP-Seq data from the entorhinal cortex of 47 Alzheimer's disease patients⁴⁵. In agreement with previous analysis we found a correlation between regions that were deregulated in ASD patients and GWAS variants linked to schizophrenia (Supplementary Fig. 8a). Surprisingly, we also found strong evidence for a link between hominin-specific regulatory gains in the PFC and regions that lose H3K27ac enrichment specifically in ASD patient brains ($p = 1.5e^{-26}$, Fisher's exact test, Fig. 4a). These include changes in GREs linked to genes previously proposed to play a role in ASD such as *c-MET*⁴⁶, *CITED4*⁴⁷, *NLK*⁴⁷, *CAMK2A*⁴⁷, and *DNMT3A*^{47,48}. For example, two hominin-specific PFC gains that were reduced in ASD brains were linked to *DNMT3A* and *CAMK2A* by HiC data (Fig. 4d, c, Supplementary Fig. 8c, d), the latter of which was shown to regulate dendritic morphology and synaptic transmission⁴⁹ and is found mutated in autism patients⁵⁰. In addition, two oligodendrocyte-specific GREs emerged in the *c-MET* locus (Fig. 4d, Supplementary Fig. 8e), a region proposed to play a role in ASD⁴⁶, with both of these elements showing a significant reduction of H3K27ac enrichment in ASD patients⁴⁴. Enrichment of H3K27ac at these GREs positively correlated with both *c-MET* expression levels and H3K27ac enrichment on the promoter (Fig. 4e, f)⁴⁴ and both GREs contact the *c-MET* promoter when assessed in 4C experiments using white matter tissue (Fig. 4g).

To further explore the link between hominin-specific gains and ASD losses, we analyzed whether ASD losses could be assigned to a specific cell-type. Analyzing human NeuN- and NeuN+ specific open chromatin regions we observed a slight reduction in glia-specific open chromatin and a gain in neuron-specific signal in PFC samples from patients compared to controls (Supplementary Fig. 9a). In contrast AD brains showed a loss of NeuN+ signal which is consistent with a loss in neurons in the entorhinal cortex⁵¹. This was also observed using H3K27ac FANS data (Supplementary Fig. 9b). While the reduction in glia-specific signal in ASD patients could underlie a shift in cellular composition, they could also be related to cell-type-specific transcriptional programs that are changed. This is more likely

are highlighted in green. **e**, Correlation of H3K27ac enrichment in the 5' and 3' GRE with expression of the *c-MET* gene in autism patients. Black dots represent samples, blue line represents linear regression through the samples, gray area highlights the 95% confidence interval. **f**, As in (e), correlation of H3K27ac enrichment on the 5' and 3' GRE with H3K27ac enrichment on the promoter. Dots represent the different samples, colors indicate species. **g**, 4C-seq and rpkm normalized H3K27ac ChIP-seq tracks of human white matter in a region containing the *c-MET* gene. The *c-MET*-promoter was used as viewpoint for the 4C analysis. The green highlighted areas indicate the location of hominin-specific PFC gains as shown in (d). **h**, Bar plot showing the cell type specificity of regions based on single cell ATAC-Seq data, on regions identified in ASD patients and control brains, regions specifically downregulated in ASD patients and those that are down in ASD as well as hominin-specific gain in PFC. Difference in oligodendrocyte-specific frequency was calculated using a Fisher's exact test. Source data are provided in Source Data file.

as consistent changes in glial cell content in the PFC of adult autism patients were not observed previously^{52–54}. Independent data from single cell ATAC-Seq data confirmed enrichment for oligodendrocyte-specific open chromatin at regions that were lost in ASD (Fig. 4g). Nevertheless, regions that are specifically gained in hominins are substantially more often oligodendrocyte-specific compared to the regions lost in autism brains (Fig. 4g, Supplementary Fig. 9b). Moreover, LLM3D analysis⁵⁵ suggests that, while a part of the interaction between ASD and hominin changes may be driven by both of these preferentially affecting oligodendrocytes or a subtype thereof, the specificity of these regions is higher than expected based on a random interaction between the two sets ($p < 10e^{-16}$, LLM3D). Thus, oligodendrocyte overrepresentation in the two sets is not the sole explanation for the interaction between hominin gains and ASD.

To analyze whether the link between evolution of novel GREs and ASD was specific to hominin-specific gains we repeated the analysis for human and chimpanzee-specific regulatory changes. We defined these as consistent gains or losses of activity compared to all three other primate species (Supplementary Fig. 9c). We found that both human and chimpanzee-specific gains were also enriched for GREs depleted in ASD ($p = 4.0e^{-3}$ and $p = 1.2e^{-4}$ respectively, Fisher's exact test, Supplementary Fig. 9d). Cell type analysis of these regions showed that both preferentially originate from oligodendrocyte (Supplementary Fig. 9e). Nevertheless, we observed a reduction in oligodendrocyte specificity for chimpanzee-specific regions that are lost in ASD compared to human-specific GREs, the latter being similar to hominin-specific gains (Supplementary Fig. 9e, f). Thus, the skewing of regions that evolved in hominins towards oligodendrocytes and ASD persisted during human but not chimpanzee evolution. Therefore, we propose that a regulatory program affecting oligodendrocytes past the postnatal stage, which is preferentially disrupted in ADS, first emerged in hominins and continued to change in the human lineage.

Discussion

A connection between human evolution and the emergence of neural disease has long been suspected⁴, with several genetic elements that recently evolved in humans or in great apes showing evidence for disease related changes^{2,3,5,56}. For instance, several human accelerated regions (HARs) were linked to both ASD and schizophrenia supporting a role for human-specific regulatory changes near neural genes in these psychiatric disorders^{2,56}. As an increased susceptibility to neural disease is unlikely to have evolved in isolation without an added benefit, connecting human evolution to neural disorders may lead to unraveling of the key genetic changes that underlie the emergence of the human brain. However, data on great apes is rare and as such human evolution is often inferred based on comparison of human and rhesus macaque¹. Furthermore, it is far from clear what the defining features are that make the human brain special and to what extent

these may or may not be shared with larger primates⁸. Our data demonstrate that the regulatory changes that set the stage for the human condition as well as some of their associated disorders may have emerged earlier, prior to the separation of human and chimpanzee. These changes correlate with hominin-specific sequence changes and selectively affect GREs in oligodendrocytes, key regulators of synaptic plasticity throughout life and essential for higher executive function⁵⁷. Furthermore, these regulatory elements are specifically employed between postnatal life and adulthood. Indeed, protracted myelination during postnatal development has been proposed as one of the hallmarks of human brain development compared to other primates⁵⁸. Moreover, expression analysis suggests that regulatory programs affecting myelination recently changed in the human lineage¹⁵. Our data suggests that, while oligodendrocytes were selectively altered during hominin evolution, changes in oligodendrocytes continued to occur in the human lineage in elements that are relevant in ASD. This suggests that changes in this regulatory program may have been important both during the evolution of hominins as well as the emergence of humans.

Defects in oligodendrocyte function is gaining attention as a contributing factor to a variety of neurodegenerative and neuropsychiatric diseases⁵⁷. While the evolutionary changes in oligodendrocytes were linked to hominin-specific sequence changes, and not due to changes in oligodendrocyte content, it is unlikely that parallel mutations in sets of recently evolved regulatory elements in oligodendrocytes occur at the same time in ASD patients. Thus, the link between hominin evolution and ASD presented here may reflect a recently evolved oligodendrocyte-specific transcriptional program that is driven by the regulatory elements identified and that is preferentially affected in ASD or a specific oligodendrocyte subtype⁵⁹. As such, analysis of oligodendrocyte subpopulations as well as regulatory changes linked to ASD risk genes such as *CAMK2A* and *DNMT3A*^{47,48} may further our understanding of oligodendrocyte function in ASD. In addition, *c-MET*⁴⁶ a gene previously proposed to play a role in ASD, is of interest as it codes for a receptor tyrosine kinase that is targeted by hepatocyte growth factor (HGF), which can enhance the proliferation and migration of oligodendrocyte progenitors and delay their differentiation⁶⁰. A single nucleotide variation in the *c-MET* promoter reduces *c-MET* expression and protein levels and was associated with an increased ASD risk and altered connectivity in ASD patient brains^{61,62}. Furthermore *c-MET* protein levels were found reduced in autism brain samples⁶³. Interestingly, both HGF expression as well as *c-MET* expression were shown to also increase the migration and proliferation of oligodendrocyte progenitor cells in tissue culture⁶⁰ and expression of HGF by oligodendrocyte precursors was shown to enhance neural survival⁶⁴. In line with these observations, neural growth and dendrite and synapse maturation is dependent on MET receptor signaling⁶⁵. Nonetheless, levels of *c-MET* expression were not significantly altered in a recent study analyzing the PFC of autism patients⁶⁶. Thus, the potential role of this gene and its control of oligodendrocyte differentiation in ASD requires further

study. As such our data provides important insight into the regulatory changes that have contributed to the emergence of the human brain, while prioritizing a new set of regulatory elements that could underlie some of the pathology observed in ASD patients.

Methods

Sample and data collection

Marmoset (*Callithrix jacchus*; cj1, cj2, cj3) tissue samples were collected at the Biomedical Primate Research Centre (BPRC) in Rijswijk, the Netherlands (<http://www.bprc.nl>) and represent rest material involving no animal experimentation for the purpose of this work as determined by the Animal Experimental Committee (DEC) (Supplementary Data 1). Samples were flash frozen immediately after dissection and stored at -80 °C. Three brain regions, cerebellum, prefrontal cortex and white matter were used for analysis in this manuscript. Brain samples from human (*Homo sapiens*; hg1, hg2, hg3), chimpanzee (*Pan troglodytes*; pt1, pt2) and rhesus macaque (*Macaca mulatta*; rm1, rm2, rm3) have been generated by us previously^{14,67}, or were resampled from the original brains. The adult nondemented control female human hemispheres were obtained from the Netherlands Brain Bank (<http://www.brainbank.nl/>). Informed consent was acquired meeting all ethical and legal requirements for autopsy, tissue storage, and use of tissue and clinical data for research. Other datasets analyzed in this study were obtained from GEO (<https://www.ncbi.nlm.nih.gov/geo/>) or Synapse (<https://www.synapse.org>) (Supplementary Data 2).

Chromatin Immunoprecipitation followed by sequencing

Chromatin immunoprecipitation (ChIP) was performed as previously described by us^{14,67}. 60 mg of tissue was used per ChIP and homogenized in a glass douncer (Kontes Glass Co.) in 1 mL Dulbecco's Modified Eagle Medium with 0.2% Bovine Serum Albumine (BSA). Cells were crosslinked while rotating for 10' at RT in 10 ml fixation buffer (freshly made; 1% formaldehyde, 0.5 mM ethylenediamine-tetraacetic acid (EDTA), 0.05 mM ethylene glycol-bis(β-aminoethyl ether)-N,N,N',N'-tetraacetic acid (EGTA), 10 mM NaCl, 5 mM HEPES-KOH, pH 7.5). Samples were washed twice with PBS, centrifuged for 5' at 2095 x g and 4 °C, resuspended in 10 ml lysis buffer (50 mM HEPES-KOH pH 7.5, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% Igepal, 0.25% Triton X-100) and incubated for 10' at RT while rotating. Cells were then centrifuged for 5' at 2095 x g at 4 °C and resuspended in 10 ml wash buffer (200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 10 mM Tris-HCl pH 8.0) and incubated for 10' at RT while rotating. Cells were pelleted for 5' at 2095 x g at 4 °C and resuspended in 150 μl sonication buffer (1 mM EDTA, 0.5 mM EGTA, 10 mM Tris-HCl pH 8.0, 100 mM NaCl, 0.1% Na-Deoxycholate, 0.5% N-lauroyl sarcosine), divided over two microtubes (Covaris 520045) and sonicated using the Covaris

S series (12 cycles of 60 seconds: intensity 3, duty cycle 20%, 200 cycles/burst). After sonication, the multiple microtubes per sample were pooled and sonication buffer and Triton X-100 (final concentration 1%) was added to a total volume of 550 μ l. Immunoprecipitation with antibody (H3K27ac: ab4729 abcam, H3K4me3: ab8580 abcam) coated Dynabeads Protein G (Invitrogen 10003D) was performed overnight at 4 °C. Beads were subsequently washed 4 times with RIPA (50 mM HEPES-KOH pH 7.5, 1 mM EDTA, 0.7% DOC, 1% NP40 and 0.5 M LiCl) and once with 50 mM NaCl in TE. Elution of the DNA from the beads was done overnight in elution buffer (50 mM Tris pH 8.0, 10 mM EDTA, 1% SDS) at 65 °C. Beads were removed by short centrifugation and the supernatant 1:1 diluted with TE, followed by a 2 h incubation with RNase (final concentration 0.2 μ g/ μ l) at 37 °C and a 2 h incubation with proteinase K (final concentration 0.2 μ g/ μ l) at 55 °C. Finally, the DNA was extracted using phenol/chloroform with MaXtract High Density gel tubes (Qiagen) and purified using ethanol. Sequencing libraries were prepared according to the Illumina Truseq DNA library protocol and samples were sequenced at the MIT BioMicro Center (<http://openwetware.org/wiki/BioMicroCenter>) using the Illumina HiSeq 2000 sequencer.

RNA sequencing

White matter tissue for bulk RNA-sequencing was dissected from each primate species (100 mg per sample). Tissue was cut into small pieces of 1 mm length and collected in 1 ml Trizol. Samples were vortexed to dissolve the tissue followed by a 5' incubation at RT. Samples were centrifuged for 5' at 21,000 x g at 4 °C and the supernatant was transferred to a fresh Eppendorf tube. 400 μ l chloroform was added and samples were mixed by shaking, incubated on ice for 5' and centrifuged for 15' at 21,000 x g at 4 °C. The aqueous upper-layer was transferred to a fresh tube. 500 μ l isopropanol and 1 μ l glycoblue (Invitrogen) were added followed by mixing by shaking and incubation overnight at -20 °C. The following day, samples were centrifuged for 30' at 21,000 x g, 4 °C. RNA pellets were washed twice with 75% Ethanol and air-dried for 8' at RT. Pellets were resuspended in 11 μ l nuclease-free water and transferred to USEQ (www.useq.nl) for library preparation and sequencing. RNA libraries were prepared according to the Illumina TruSeq Stranded total RNA library protocol and samples were sequenced at USEQ on a high-output NextSeq500, 1 x 75bp.

ATAC-sequencing on FANS sorted nuclei

Prefrontal cortex tissue was dissected from the primate brains and homogenized in a glass douncer (Kontes Glass Co.) in 2 ml EZ buffer (Nuclei Isolation Kit, Sigma NUC101) and incubated for 5' on ice. Subsequently samples were centrifuged for 15' at 65 x g at 4 °C, resuspended in 2 ml EZ buffer and incubated on ice for 5'. Samples were again centrifuged for 10' at 65 x g at 4 °C, resuspended in 4 ml Nuclear suspension buffer (NSB; 0.01% BSA, 1x complete protease inhibitor cocktail in PBS) and filtered through a 40 μ m cell strainer. Samples were centrifuged for 15' at 65 x

g at 4 °C and resuspended in 30 ml NSB after which 10 ml 30% OptiPrep (Sigma D1556) was added on the bottom of the sample using a needle. Another layer of 5 ml 60% OptiPrep was then added underneath and the sample was centrifuged for 10' at 1000 x g at 4 °C. Purified nuclei were recovered from between the 30% OptiPrep and PBS layers and checked for quality under a bright field microscope. 100 μ l blocking solution (0.5% BSA, 10% FBS in PBS) was added containing anti-NeuN antibody conjugated with Alexa488 (1:1000, Merck Millipore MAB377X) and Hoechst 34580 (1 μ g/ml, Fisher Scientific H21486) and incubated for 1 h at 4 °C in the dark on a roller. Stained nuclei were then transferred to FACS tubes pre-coated with 5% BSA and sorted on NeuN signal using a FACS-Jazz (BD bioscience). 50,000 NeuN+ and NeuN-stained nuclei were collected in PBS and processes further for ATAC-sequencing⁶⁸. In short, 0.1% NP-40 was added to the sorted nuclei and samples were centrifuged for 15' at 150 x g at 4 °C. Pellets were then resuspended in 1 ml resuspension buffer (10 mM Tris-HCl pH 7.5, 10 mM NaCl, 3 mM MgCl₂, 0.1% Tween-20, 0.1% NP-40 in MQ) and incubated for 3' on ice. 1 ml resuspension buffer without NP-40 was then added to the sample and centrifuged for 10' at 500 x g at 4 °C. Nuclei were resuspended in 50 μ l transposition mix (1x TD buffer (20 mM Tris-HCl pH 7.6, 10 mM MgCl₂, 20% Dimethyl formamide in MQ), 100 nM Tn5 transposase, 33% PBS, 0.01% digitonin, 0.1% Tween-20 in MQ) and incubated for 30' at 37 °C while shaking at 1000 rotation per minute. Samples were purified using Qiagen MinElute PCR purification kit (Qiagen 28004) and eluted in 21 μ l MQ. Purified DNA was amplified with NEBnext High-Fidelity PCR master mix (NEB M0541S) and appropriate sequencing adapters for five PCR cycles. Library complexity was determined by qPCR on 5 μ l of the PCR sample and the number of extra PCR cycles determined⁶⁹. PCR samples were purified using AMPure XP beads (Agencount A63881), eluted in 12 μ l MQ and sequenced at USEQ on a high-output NextSeq500, 1 x 75 bp.

Mapping and analysis of primate ChIP-Seq data

To ensure proper comparison of samples from different platforms and studies, all reads from primate cortical and cerebellar samples were trimmed to a 36 bp length using the Fastx-toolkit (http://hannonlab.cshl.edu/fastx_toolkit/index.html). Sequences were aligned using Bowtie 1.1.2 (<http://bowtie-bio.sourceforge.net/index.shtml>) excluding reads with more than one mismatch (seed length 36) or with multiple alignments, unless stated otherwise. Reads were mapped to the most recent reference genome available (Supplementary Data 3). Fraction of reads in peaks (FRiP) scores all exceed the 1% threshold used by the Encyclopedia of DNA Elements (ENCODE)²¹, and the percentage of unique reads was overall high (mean 89.25%, Supplementary Data 3). Genome-wide enriched regions were annotated per sample using MACS2 version 2.1.1 ($P = 10^{-5}$, extsize = 300, local lambda = 100,000). Whole-cell extract input controls were generated for each common marmoset brain region (Supplementary Data 3). We used the internal MACS2 lambda control to correct for local bias as whole-cell extract inputs often introduce sonication biases

at open chromatin regions. Identified H3K27ac or H3K4me3 enriched regions were extended to a minimum size of 2000 bp (peak center $\pm$ 1000 bp), the resolution typically observed for these regions^{13,67}. Lists of enriched regions per species and brain region were obtained by merging the identified regions of the replicates per tissue, with regions overlapping at least 1 bp being stitched together. This lenient cutoff was chosen to ensure that most enriched regions with multiple summits were merged. The average overlap was $\sim$ 700 bp, with only 3% of regions were a merger with an overlap of less than 100 bp. Reproducible enriched regions were defined as enriched regions present in at least two biological replicates of the same brain region per species. Non-redundant H3K27ac enriched region lists were obtained per primate species by merging the regions of both brain regions. Public data was downloaded from either the GEO or Synapse repositories (Supplementary Data 2) and reanalyzed to match our analysis using the most recent primate genome if required, using the settings described above.

GRE analysis and quality control across primate genomes

Primate regulatory elements were defined using the UCSC liftOver tool ($-\text{minMatch}=0.1$) (Supplementary Fig. 2c). H3K27ac-enriched regions of chimpanzee, rhesus macaque and marmosets were reciprocally mapped to the human genome (hg38). Regions that were mapped to multiple (non-unique) locations and regions that changed more than 50% in size were excluded ($n = 2043$). Following mapping to hg38, all four lists for the different primate species were merged. Average overlap was $\sim$ 2700 bp, with only 0.3% of regions were a merger with an overlap of less than 100 bp. The resulting list was again mapped to all three primate genomes using reciprocal liftOver to make sure that merged regions were mappable across all primates. To ensure equal mappability, $>90\%$ of the bases within a regulatory element had to be properly annotated in all reference genomes ($<10\%$ overlap with UCSC Table Browser's gap locations lists). This cut-off was chosen as unknown bases generally occur in stretches rather than as single nucleotides across the genome. To account for repetitive or duplicated genomic regions, which are especially susceptible to poor annotation in lower-quality genomes, enrichment scores were not allowed to change significantly in the target genome when allowing reads to map to multiple locations. These repetitive regions were defined per species by mapping reads from every sample to unique locations (bowtie: $-\text{best} -\text{strata} -\text{m } 1$) as well as to multiple locations (bowtie: $-\text{best} -\text{strata} -\text{M } 1$). Genomic regions that were enriched using the multimap settings but not with unique mapping are potential repetitive elements that are not annotated at similar depth across all the genomes. All regulatory elements that overlap a repetitive element on any of the primate species ($n = 11,432$) were therefore discarded. In total 37,308 regulatory elements could be identified with the above restrictions on all four genomes. GREs were assigned to their target gene based on their location on the human genome as it is the best annotated genome. GREs located within 1000 bp of an annotated TSS (USCS RefSeq hg38) were considered

promoters and assigned to the associated gene. Enriched regions located outside 1000 bp from a TSS were classified as putative enhancers and were assigned to a target gene based on PFC HiC data⁹. If no significant enhancer-promoter loop was annotated, the enhancer was assigned to its closest active TSS based on H3K4me3 enrichment (Supplementary Data 4).

Hierarchical clustering, PCA, t-sne analysis

Duplicate reads were removed from the bam files using Samtools 1.3.1 and read coverage within enriched regions was counted using Bedtools v2.26.0. Read counts were then normalized for the total number of uniquely mapped reads per sample and log₂ transformed using the rlog (blind) function in DESeq2. For hierarchical clustering (e.g. Supplementary Fig. 2b, f), Pearson correlations between the samples were calculated. Samples were clustered based on Pearson distance using average linkage and heatmaps were generated using the heatmap.2 function from the gplots R package. For principle component analysis (PCA, Fig. 1c), the prcomp function in R (<http://www.R-project.org>) was used. t-sne multidimensional scaling coordinates were determined using the t-sne R package (Fig. 1b). H3K27ac heatmaps and metaplots were generated using the ngs.plot.r (e.g. Supplementary Fig. 1c).

Annotation of hominin-specific regulatory changes

Hominin-specific regulatory changes were identified based on differential H3K27ac-enrichment between great ape (human and chimpanzee) and monkey (rhesus and marmoset) species. Pairwise comparison of human and chimpanzee with macaque and marmoset for both cerebellum and prefrontal cortex was performed using DESeq2 (Supplementary Fig. 3a, b). Per pairwise comparison we temporarily excluded all regions with zero original read count in all the replicates of a single species, in total 711 regions were excluded. Raw read counts were then used to define differentially enriched regions using the DESeq2 function in R. Regulatory elements with a twofold change in enrichment and an FDR < 0.01 were defined as significantly differentially enriched (DE). Hominin regulatory changes were defined as all regions that are DE for both human and chimpanzee compared to macaque and marmoset, or vice versa (Supplementary Fig. 4a).

Analysis of GRE cell type specificity

To assess the cell type specificity of GREs in our datasets, we leveraged ChIP-Seq data from FANS sorted PFC nuclei^{27,30} and single cell ATAC-Seq data from PFC³⁴. For FANS sorted nuclei, paired-end ChIP-Seq sequencing reads were obtained and mapped to the human genome using Bowtie 1.1.2, excluding reads with more than one mismatch, multiple alignments or where one of the pairs could not be mapped. Metaplots and heatmaps were obtained using the ngs.plot.r function.

For cell-type-specific ATAC-Seq regions we obtained lists of differentially enriched regions per cell type as provided³⁴. From these we selected regions that

were differentially enriched in a single cell type. GREs from our datasets were then overlapped with these cell-type-specific ATAC-Seq regions and GREs overlapping a single cell-type-specific ATAC-Seq region were annotated as specific to that cell type. GREs overlapping none or more than 1 cell-type-specific ATAC-Seq region were annotated as not cell-type-specific. The difference in oligodendrocyte-specific content between sets of GREs was calculated using a Fisher's exact test in R.

Confounder analysis and batch effect

Batch effect was controlled for as done previously by analyzing biological replicates in separate batches^{14,67}. This reduces consistency in batch related influences between replicates thus reducing the chance that these are picked up in DE analysis. This also allows for the reanalysis of data that has been generated at an earlier stage and has been handled similarly. To assess the effect of confounding variables, such as post-mortem delay (PMD) and sequencing batch between our datasets, we performed multivariate analysis on the H3K27ac enrichment of the here identified GREs. We defined several biological and technical covariates that could have influenced enrichment scores, including species, tissue type, gender, PMD, sequencing depth, sequencing platform, FRIP score and experimental batch. These covariates were used in a linear regression model on rpkm normalized read tables, after which we used ANOVA to extract the percentage of variance explained and significance per covariate for every GRE (Supplementary Fig. 3c-e). Only 27 out of 1930 GREs that changed in hominins (gains and losses) had a significant contribution of covariates (batch, gender, PMD, sequencing platform, $p < 0.01$, Supplementary Fig. 3d). One of these was also deregulated in ASD brain. To further exclude that the identified characteristics of hominin-specific PFC gains are not due to batch effect, we analyzed white matter ChIP-sequencing samples where the samples were equally distributed across pre-defined experimental batches with all of the different species occupying the same batch and batches containing biological replicates. The resulting regions were still enriched for oligodendrocytes and preferentially deregulated in ASD brain (Supplementary Fig. 9g, h), excluding batch as a basis for our observations.

Functional analysis of gene sets and motif analysis

Gene ontology analysis was done using the Genomic Regions Enrichment of Annotations Tool (GREAT version 3.0.0, <http://bejerano.stanford.edu/great/public/html>) with basal plus extension setting. Multiple genes can therefore be assigned to the supplied enriched regions (Fig. 3d, Supplementary Fig. 6f, Supplementary Data 6). Motif enrichment was performed on hominin-specific regulatory changes for cerebellum and prefrontal cortex separately, using HOMER version 4.9.1 with default parameters (Supplementary Data 5).

Stratified LD score regression of SNPs in hominin GREs

To study enrichment for disease heritability at hominin-specific GREs we obtained

the summary statistics of genome wide associate data for various brain related phenotypes (Alzheimer's disease, autism spectrum disorder, Parkinson's disease, Schizophrenia and bipolar disorder). To calculate disease enrichment, we added our sets of GREs (hominin-specific changes, all GREs, all enhancers, all promoters) and ASD associated GREs as defined previously⁴⁴ to the baseline model of this package. We also included a set of randomly shuffled GREs. Enrichment statistics were calculated as the ratio between explained heritability and the proportion of SNPs in H3K27ac enriched regions, using the LD score regression software (<https://github.com/bulik/ldsc>).

Autism spectrum disorder⁴⁴ or Alzheimer's disease⁴⁵-associated regulatory elements were categorized as more enriched or less enriched in patients brains over control samples using recent data. The resulting regions were compared with hominin gains and losses from cerebellum and prefrontal cortex. Significant enrichment or depletion compared to all annotated GREs per brain region was determined using a Fisher's exact test in R. The statistical significance cut-off was set at 0.01 and p-values were adjusted for multiple testing using the Benjamini-Hochber's FDR method.

Sequence conservation and hominin nucleotide changes

PhastCon scores were calculated using the UCSC phastCon 20 mammals track (<http://hgdownload.soe.ucsc.edu/goldenPath/hg38/phastCons20way/>). To increase the resolution at GREs for relevant nucleotides, we used only those nucleotides that were DNase enriched in the frontal cortex, as defined by ENCODE, to calculate conservation²⁹. Mean values of all scored nucleotides per GRE are plotted (Fig. 1e). Hominin-specific nucleotide changes were define based on the pairwise alignment files of chimpanzee, rhesus macaque and marmoset with human, as obtained from UCSC. All nucleotides that shared the same base between human and chimpanzee but not with macaque and marmoset were annotated as hominin-specific nucleotide changes. The percentage of hominin-specific nucleotide changes per GRE are plotted (Fig. 1f)

Analysis of gene expression data and open chromatin

Raw RNA sequencing reads were trimmed to a length of 60 bp, starting from base 12, using the Fastx-toolkit and aligned to the appropriate genome using hisat2 2.0.5 (<https://ccb.jhu.edu/software/hisat2/index.shtml>) using default settings (Supplementary Data 3). Duplicated reads which are likely PCR duplicates were removed. Reads mapping to multiple locations were also removed from further analysis. Expression values were counted for each gene using the featureCount function of the Rsubread packages in R. ATAC-sequencing reads were trimmed to remove Nextera adapter sequence 5'-CTGTCTCTTATA-3' using cutadapt (<https://cutadapt.readthedocs.io/en/stable/>) and processed similar to ChIP data as described above.

Analysis of interdependence

The interdependence between categorical variables (hominin-specific gain, deregulated in ASD and oligodendrocyte specificity) of the identified PFC GREs, was calculated using LLM3D⁵⁵. LLM3D fits a number of log-linear models to 3D contingency tables of GRE counts and selects the model that best fits the observed element characteristics. These models imply different (in)dependence relationships between the variables, with the null hypothesis assuming complete independence. The significant p-value indicates that the oligodendrocyte specificity of the hominin-specific gains deregulated in ASD is higher than expected for a random overlap between the two sets.

Chromosome Conformation Capture (4C) analysis

Frozen white matter tissue was dissected from the brain and pulverized in liquid nitrogen using a pestle and mortar. The frozen tissue was further homogenized in 1 ml PBS with 10% Fetal Bovine Serum (FBS) using a cold 2 ml dounce (Kontes Glass Co.). Cells were then cross-linked and processed, using DpnII and Csp6I as restriction enzymes. Cells were fixed for 10' on RT in fixation buffer (10% FBS, 2% Formaldehyde in PBS), after which glycine was added to a final concentration of 125 mM to quench the fixation. Samples were washed twice with PBS and centrifuged for 5' at 750 x g. Cells were lysed in 2 ml cold lyses buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 5 mM EDTA, 0.5% NP-40, 1% Triton X-100) and incubated for 5' at RT, 5' at 65 °C and 1' on ice. Samples were subsequently pelleted by centrifugation for 5' at 550 x g. The 3C template was obtained by addition restriction buffer (1X DpnII restriction buffer, 3% SDS in MQ) and incubated for 1 h at 37 °C. Triton X-100 was added to a 2.5% concentration and samples were again incubated for 1 h at 37 °C. Restriction of the samples was performed overnight at 37 °C after addition of 10U DpnII restriction enzyme, followed by an overnight ligation at 15 °C after addition of T4 ligation buffer (final concentration: 1x) and 1.5 µl T4 ligase. The following day, 100 µg proteinase K was added and samples incubated overnight at 65 °C, followed by a 45' incubated at 37 °C with 100 µg RNase A. The 3C template was purified using phenol-chloroform and processed for another round of restriction-ligation using Csp6I to create the 4C-template. For amplification on the MET promoter viewpoint the following primers were used: 5'- TACACGACGCTCTTCCGATCTCTAATGAA TTTTCTGTCATGAAGAT-3' as reading primer and 5'- ACTGGAGTTCAGACG TGTGCTCTTCCGATCTGTTACCAGCCCTAGACGTG-3' as non-reading primer. Sequencing was done on the Illumina MiniSeq platform. Sequencing reads were trimmed to remove primer sequences and mapped on hg19, removing all reads that mapped to multiple locations. 4C coverage was calculated by averaging mapped reads in running windows of 41 fragments ends.

Data Availability

The ChIP-Seq and RNA-Seq data reported in this study is available at Gene Expression Omnibus with accession code GSE130871. Other public datasets employed in our study can be found via accession numbers in Supplementary Data 2.

Code Availability

All analysis was done using R (<http://www.r-project.org>) or Python (<http://www.python.org>) employing the packages described above.

Acknowledgements

M.P.C. was supported by the Royal Academy of Sciences (KNAW) in the Netherlands and the Erasmus Medical Center. We thank Utrecht Sequencing Facility for providing sequencing service and data. Utrecht Sequencing Facility is subsidized by the University Medical Center Utrecht, Hubrecht Institute, Utrecht University and The Netherlands X-omics Initiative (NWO). We thank Kevin Kenna and Mark Bakker from UMC Utrecht Neurogenetics department for their help with the stratified LD score regression analysis.

Author Contribution

B.C. and M.P.C. conceived and designed the experiments. B.C., I.T, C.R.M.W., M.W.V and P.S. performed the experiments and were supervised by M.P.C. and N.G. I.K. collected and dissected primate brain specimens. B.C., M.L.B., G.G. and V.B. analyzed the data and were supervised by M.P.C. and W.d.L. B.C. and M.P.C. wrote the manuscript.

Competing Interest

The authors declare no competing interests

Author Information

Accession codes. Gene Expression Omnibus: GSE130871

Correspondence and requests for materials should be addressed to m.creyghton@hubrecht.eu

References

1. Sousa, A. M. M., Meyer, K. A., Santpere, G., Gulden, F. O. & Sestan, N. Evolution of the Human Nervous System Function, Structure, and Development. *Cell* 170, 226–247 (2017).
2. Doan, R. N. et al. Mutations in Human Accelerated Regions Disrupt Cognition and Social Behavior. *Cell* 167, 341–354.e12 (2016).
3. Cookson, M. R. Evolution of neurodegeneration. *Curr. Biol.* 22, R753–61 (2012).
4. Reardon, S. Geneticists are starting to unravel evolution's role in mental illness. *Nature* 551, 15–16 (2017).
5. Won, H. et al. Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature* 538, (2016).
6. Whiten, A. & van de Waal, E. Social learning, culture and the 'socio-cultural brain' of human and non-human primates. *Neurosci. Biobehav. Rev.* 82, 58–75 (2017).
7. Finch, C. E. & Austad, S. N. Commentary: is Alzheimer's disease uniquely human? *Neurobiol. Aging* 36, 553–555 (2015).
8. Waal, F. B. M. de (Frans B. M. . Are we smart enough to know how smart animals are?
9. Wang, D. et al. Comprehensive functional genomic resource and integrative model for the human brain. *Science* 362, eaat8464 (2018).
10. Visel, A. et al. A High-Resolution Enhancer Atlas of the Developing Telencephalon. *Cell* (2013). doi:10.1016/j.cell.2012.12.041
11. Haberle, V. & Stark, A. Eukaryotic core promoters and the functional basis of transcription initiation. *Nat. Rev. Mol. Cell Biol.* 19, 621–637 (2018).
12. Wittkopp, P. J. & Kalay, G. Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nat. Rev. Genet.* 13, 59–69 (2012).
13. Villar, D. et al. Enhancer Evolution across 20 Mammalian Species. *Cell* 160, 554–566 (2015).
14. Vermunt, M. W. et al. Epigenomic annotation of gene regulatory alterations during evolution of the primate brain. *Nat. Neurosci.* 19, (2016).
15. Zhu, Y. et al. Spatiotemporal transcriptomic divergence across human and macaque brain development. *Science* 362, eaat8077 (2018).
16. Reilly, S. K. et al. Evolutionary genomics. Evolutionary changes in promoter and enhancer activity during human corticogenesis. *Science* (80-.). 347, 1155–1159 (2015).
17. Perelman, P. et al. A molecular phylogeny of living primates. *PLoS Genet* 7, e1001342 (2011).
18. Andersen, B. B., Korbo, L. & Pakkenberg, B. A quantitative study of the human cerebellum with unbiased stereological techniques. *J. Comp. Neurol.* 326, 549–560 (1992).
19. Calo, E. & Wysocka, J. Modification of enhancer chromatin: what, how, and why? *Mol Cell* 49, 825–837 (2013).
20. Guenther, M. G., Levine, S. S., Boyer, L. A., Jaenisch, R. & Young, R. A. A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* 130, 77–88 (2007).
21. Landt, S. G. et al. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res.* 22, 1813–1831 (2012).
22. Cortez, D. et al. Origins and functional evolution of Y chromosomes across mammals. *Nature* 508, 488–493 (2014).
23. Brawand, D. et al. The evolution of gene expression levels in mammalian organs. *Nature* 478, 343–348 (2011).
24. Kukekova, A. V. et al. Red fox genome assembly identifies genomic regions associated with tame and aggressive behaviours. *Nat. Ecol. Evol.* 2, 1479–1491 (2018).
25. Savas, J. N. et al. The Sorting Receptor SorCS1 Regulates Trafficking of Neurexin and AMPA Receptors. *Neuron* 87, 764–780 (2015).
26. Lake, B. B. et al. Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nat. Biotechnol.* 36, 70–80 (2018).
27. Kozlenkov, A. et al. A unique role for DNA (hydroxy)methylation in epigenetic regulation of human inhibitory neurons. *Sci. Adv.* 4, eaau6190 (2018).
28. Gosselin, D. et al. An environment-dependent transcriptional network specifies human microglia

- p identity.
- Science*
- 356, (2017).
29. Bernstein, B. E. et al. An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74 (2012).
 30. Girdhar, K. et al. Cell-specific histone modification maps in the human frontal lobe link schizophrenia risk to the neuronal epigenome. *Nat. Neurosci.* 21, (2018).
 31. Gabi, M. et al. No relative expansion of the number of prefrontal neurons in primate and human evolution. *Proc. Natl. Acad. Sci.* 113, (2016).
 32. von Bartheld, C. S., Bahney, J. & Herculano-Houzel, S. The search for true numbers of neurons and glial cells in the human brain: A review of 150 years of cell counting. *J. Comp. Neurol.* 524, 3865–3895 (2016).
 33. Sherwood, C. C. et al. Evolution of increased glia-neuron ratios in the human frontal cortex. *Proc. Natl. Acad. Sci. U. S. A.* 103, 13606–11 (2006).
 34. Lake, B. B. et al. Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nat. Biotechnol.* 36, 70–80 (2017).
 35. Li, M. et al. Integrative functional genomic analysis of human brain development and neuropsychiatric risks. *Science* 362, eaat7615 (2018).
 36. Won, H. et al. Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature* 538, 523–527 (2016).
 37. Zhu, Y. et al. Spatiotemporal transcriptomic divergence across human and macaque brain development. *Science* 362, eaat8077 (2018).
 38. Boccuto, L. et al. A mutation in a ganglioside biosynthetic enzyme, ST3GAL5, results in salt & pepper syndrome, a neurocutaneous disorder with altered glycolipid and glycoprotein glycosylation. *Hum. Mol. Genet.* 23, 418–33 (2014).
 39. Schnaar, R. L., Gerardy-Schahn, R. & Hildebrandt, H. Sialic acids in the brain: gangliosides and polysialic acid in nervous system development, stability, disease, and regeneration. *Physiol. Rev.* 94, 461–518 (2014).
 40. Wang, B. Molecular mechanism underlying sialic acid as an essential nutrient for brain development and cognition. *Adv. Nutr.* 3, 465S–72S (2012).
 41. Lambert, J. C. et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer’s disease. *Nat. Genet.* 45, 1452–1458 (2013).
 42. Grove, J. et al. Identification of common genetic risk variants for autism spectrum disorder. *Nat. Genet.* 51, 431–444 (2019).
 43. Simon-Sanchez, J. et al. Genome-wide association study reveals genetic risk underlying Parkinson’s disease. *Nat. Genet.* 41, 1308–1312 (2009).
 44. Ruderfer, D. M. et al. Genomic Dissection of Bipolar Disorder and Schizophrenia, Including 28 Subphenotypes. *Cell* 173, 1705–1715.e16 (2018).
 45. Sun, W. et al. Histone Acetylome-wide Association Study of Autism Spectrum Disorder. *Cell* 167, 1385–1397.e11 (2016).
 46. Marzi, S. J. et al. A histone acetylome-wide association study of Alzheimer’s disease identifies disease-associated H3K27ac differences in the entorhinal cortex. *Nat. Neurosci.* 21, 1618–1627 (2018).
 47. Eagleson, K. L., Xie, Z. & Levitt, P. The Pleiotropic MET Receptor Network: Circuit Development and the Neural-Medical Interface of Autism. *Biol. Psychiatry* 81, 424–433 (2017).
 48. Ruzzo, E. K. et al. Inherited and De Novo Genetic Risk for Autism Impacts Shared Networks Article Inherited and De Novo Genetic Risk for Autism Impacts Shared Networks. *Cell* 178, 850–866 (2019).
 49. F. Kyle Satterstrom, Jack A. Kosmicki, Jiebiao Wang, Michael S. Breen, Silvia De Rubeis, Joon-Yong An, Minshi Peng, Ryan Collins, Jakob Grove, Lambertus Klei, Christine Stevens, Jennifer Reichert, Maureen S. Mulhern, Mykyta Artomov, Sherif Gerges, Brooke, M. J. D. Novel genes for autism implicate both excitatory and inhibitory cell lineages in risk. *bioRxiv*, 2018 - [biorxiv.org](https://www.biorxiv.org/content/10.1101/484113) (2018). doi:10.1101/484113
 50. Stephenson, J. R. et al. A Novel Human CAMK2A Mutation Disrupts Dendritic Morphology and Synaptic Transmission, and Causes ASD-Related Behaviors. *J. Neurosci.* 37, 2216–2233 (2017).
 51. Iossifov, I. et al. The contribution of de novo coding mutations to autism spectrum disorder. *Nature* 515, 216–221 (2014).

52. Saxena, S. & Caroni, P. Selective neuronal vulnerability in neurodegenerative diseases: from stressor thresholds to degeneration. *Neuron* 71, 35–48 (2011).
53. Zikopoulos, B. & Barbas, H. Changes in prefrontal axons may disrupt the network in autism. *J. Neurosci.* 30, 14595–609 (2010).
54. Donovan, A. P. A. & Basson, M. A. The neuroanatomy of autism - a developmental perspective. *J. Anat.* 230, 4–15 (2017).
55. Velmeshev, D. et al. Single-cell genomics identifies cell type-specific molecular changes in autism. *Science* (80-.). 364, 685–689 (2019).
56. Geeven, G. et al. LLM3D: a log-linear modeling-based method to predict functional gene regulatory interactions from genome-wide expression data. *Nucleic Acids Res.* 39, 5313–5327 (2011).
57. Xu, K., Schadt, E. E., Pollard, K. S., Roussos, P. & Dudley, J. T. Genomic and Network Patterns of Schizophrenia Genetic Variation in Human Evolutionary Accelerated Regions. *Mol. Biol. Evol.* 32, 1148–1160 (2015).
58. Bercury, K. K. & Macklin, W. B. Developmental Cell Review Dynamics and Mechanisms of CNS Myelination. *Dev. Cell* 32, 447–458 (2015).
59. Miller, D. J. et al. Prolonged myelination in human neocortical evolution. *Proc. Natl. Acad. Sci.* 109, 16480–16485 (2012).
60. Jäkel, S. et al. Altered human oligodendrocyte heterogeneity in multiple sclerosis. *Nature* 566, 543–547 (2019).
61. Yan, H. & Rivkees, S. A. Hepatocyte growth factor stimulates the proliferation and migration of oligodendrocyte precursor cells. *J. Neurosci. Res.* 69, 597–606 (2002).
62. Rudie, J. D. et al. Autism-associated promoter variant in MET impacts functional and structural brain networks. *Neuron* 75, 904–15 (2012).
63. Campbell, D. B. et al. A genetic variant that disrupts MET transcription is associated with autism. *Proc. Natl. Acad. Sci. U. S. A.* 103, 16834–9 (2006).
64. Campbell, D. B. et al. Disruption of cerebral cortex MET signaling in autism spectrum disorder. *Ann. Neurol.* 62, 243–250 (2007).
65. Nakano, M. et al. NG2 glial cells regulate neuroimmunological responses to maintain neuronal function and survival. *Sci. Rep.* 7, 42041 (2017).
66. Qiu, S., Lu, Z. & Levitt, P. MET receptor tyrosine kinase controls dendritic complexity, spine morphogenesis, and glutamatergic synapse maturation in the hippocampus. *J. Neurosci.* 34, 16166–79 (2014).
67. Parikshak, N. N. et al. Genome-wide changes in lncRNA, splicing, and regional gene expression patterns in autism. *Nat.* 2016 5407633 540, 423–427 (2016).
68. Vermunt, M. W. W. et al. Large-scale identification of coregulated enhancer networks in the adult human brain. *Cell Rep* 9, 767–779 (2014).
69. Corces, M. R. et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* 14, (2017).
70. Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr Protoc Mol Biol* 109, 21 29 1–9 (2015).

Supplementary Fig. 1 Annotation of marmoset regulatory elements. **a**, Scatterplots showing the correlation of H3K27ac enrichment for all annotated marmoset regulatory elements between the replicate samples. Red line depicts linear regression through the samples, Pearson correlation between the samples is shown in the right bottom corner. **b**, Similar as shown in **(a)** for H3K4me3 enrichment. **c**, Metaplot analysis showing average enrichment profiles across scaled enriched regions (left panels). Colored lines represent average ChIP-Seq enrichment for indicated sample, shaded area represents standard error of the mean. Heatmaps showing the H3K27ac (red) and H3K4me3 (green) enrichment in cerebellum and prefrontal cortex for putative enhancers and putative promoters across 5' to 3' scaled enriched regions. **d**, Box plots showing normalized gene expression values in marmoset brain for genes with a H3K27ac enriched promoter or near a H3K27ac enriched enhancer compared to genes not linked to H3K27ac enrichment, outliers not shown. Dissimilarities between distributions were calculated using a Student's t-test. Bottom and top of the box plots are the first and third quartile. The line within the boxes represents the median and whiskers denote interval within 1.5x the interquartile range from the median.

Supplementary Fig. 2 Integration of regulatory information across primate genomes. **a**, Venn diagram showing the overlap of regulatory elements identified in cerebellum (CB: blue) and prefrontal cortex (PFC: orange) for the different primate species as indicated. **b**, Hierarchical clustering of brain samples for each primate species, based on regulatory elements identified in the separate species. Correlation map colors indicate Pearson correlation coefficients between the samples. Side bars are color-coded as in (a). Pearson distances are represented by the tree on the left. **c**, Schematic overview of the liftover pipeline used to identify GREs with orthologous sequences on all primate genomes. **d**, Box plots showing the length distribution of identified primate GREs per primate species. **e**, Box plots showing the normalized H3K27ac enrichment of all identified primate GREs in cerebellum and prefrontal cortex for all primate species. Colors indicate tissue of origin as in (a). **f**, Hierarchical clustering of primate samples based on all identified primate GREs for cerebellum and prefrontal cortex. Heatmap colors represent the Pearson correlation coefficients between the samples. Pearson distances are represented by the tree on the left. Bottom and top of all box plots are the first and third quartile. The line within the boxes represents the median and whiskers denote interval within 1.5x the interquartile range from the median, outliers are depicted as points. Source data are provided as a Source Data file.

Supplementary Fig. 3 Differential enrichment analysis in PFC and CB. **a**, Box plots showing the average normalized H3K27ac enrichment per tissue as indicated for elements defined as human or monkey (macaque or marmoset) specific based on DESeq2 analysis. Colors indicate species (red: human, green: macaque, purple: marmoset). **b**, Box plots as in (a) using chimpanzee instead of human to compare to the monkey species. Colors indicated species (blue: chimpanzee, green: macaque, purple: marmoset). Bottom and top of the box plots are the first and third quartile. The line within the boxes represents the median and whiskers denote interval within 1.5x the interquartile range from the median, outliers are depicted as points. **c**, Violin plots showing contribution of multiple covariates on H3K27ac enrichment assessed by a multivariate linear regression using ANOVA on all GREs across the primate samples (FRIP: fraction of reads in peaks, PMD: postmortem delay). **d**, Similar as in (c) but showing the significance per covariate. Dashed line indicates the significance threshold $p < 0.01$. **e**, Similar is in (c) but for hominin-specific changes in prefrontal cortex and cerebellum.

Supplementary Fig. 4 Identification of hominin-specific regulatory changes in PFC and CB. a, Venn diagrams showing hominin-specific regulatory elements (highlighted in yellow) and species-specific gains and losses, as characterized based on differentially enrichment by DESeq2, for prefrontal cortex (left) and cerebellum (right). Numbers indicate the amount of regions in each category. **b**, ChIP-Seq tracks showing rpm normalized H3K27ac enrichment across a 310 kb region containing the *SORCS1* gene in human cerebellum for the different species as indicated. The region harbours hominin-specific losses (highlighted in green). **c**, Specificity of hominin-specific gains in other human tissues compared to regions that show no differentially enrichment between the primates. Difference in their frequency was calculated using a Fisher's exact test. Source data are provided as a Source Data file.

Supplementary Fig. 5 Analysis of hominin-specific regulatory elements at cell type resolution. **a**, ChIP-Seq tracks showing rpm normalized H3K27ac enrichment across several known marker genes specific for neuronal and glial cell types in the human brain. Markers are *VGLUT1* for glutamatergic neurons (Glut), *OLIG1*, *OLIG2* and *SOX10* for oligodendrocytes (Olig), *LHX6* for GABAergic neurons (Gaba), *NEUN* for neurons, *S100B* for Glia, *TMEM119* for microglia (Micgl) and *GFAP* for astrocytes (Astr). **b**, Heatmap showing H3K27ac enrichment across scaled regions for all GREs identified in PFC and CB across primates and analyzed in the different neuronal and glial cell types as indicated. Heatmap colors indicated H3K27ac enrichment (rpm). **c**, Metaplot analysis showing the average H3K27ac enrichment profile in different cell types for hominin-specific CB gains. **d**, Metaplot analysis showing the average rpm normalized H3K27ac enrichment profile of three NeuN- (green) and three NeuN+ (red) sorted nuclei samples on hominin-specific PFC gains. Shaded area depicts the standard deviation from the mean. **e**, Metaplot as shown in (**d**) but for all GREs.

Supplementary Fig. 6 Hominin-specific gains emerged in non-neuronal cells. **a**, ATAC-seq tracks showing the rpm normalized ATAC signal for NeuN-negative sorted nuclei (green) and NeuN-positive sorted nuclei (red) across several known marker genes specific for neuronal cells (*NEUN* & *GRIN1*) and glial cells (*S100B* and *SOX10*) in four primate species as indicated. **b**, Hierarchical clustering of NeuN sorted nuclei samples from prefrontal cortex of different primate species. Correlation map colors indicate Pearson correlation coefficients between samples. Side bars are color-coded based on sample origin and NeuN signal. Pearson distances are represented by the tree on the left. **c**, Box plots showing normalized ATAC enrichment for all primate GREs on nuclei sorted for NeuN in different primate species. Dissimilarities between the distributions were calculated using a Student's t-test. **d**, Box plots showing normalized ATAC enrichment in nuclei sorted for NeuN from different primate species as indicated on hominin-specific PFC gains. Dissimilarities between the distributions were calculated using a Student's t-test. **e**, Box plots showing normalized H3K27ac enrichment in the prefrontal cortex of human and macaque for NeuN-specific regions and hominin-specific PFC gains. Dissimilarities between the distributions were calculated using a Student's t-test. **f**, ChIP-seq track showing the rpm normalized H3K27ac enrichment across NeuN and Sox10 loci for white matter tissue of the different primates as indicated. **g**, t-Distributed Stochastic Neighbour Embedding (t-sne) analysis of H3K27ac enriched GREs in white matter tissue with orthologs on all four primate genomes ($n = 23,437$). Axes indicate semantic space. Bottom and top of all box plots are the first and third quartile. The line within the boxes represents the median and whiskers denote interval within 1.5x the interquartile range from the median, outliers are depicted as points. Source data are provided as a Source Data file.

Supplementary Fig. 7 Hominin-specific GREs link to postnatally active genes. **a**, Box plots showing normalized H3K27ac enrichment per developmental stage for all here identified GREs and hominin-specific PFC gains. Dissimilarities between the distributions were calculated using a Student's t-test. **b**, Pie chart showing the overlap in gene annotation of known human TSS on primate species. **c**, ChIP-Seq tracks showing rpm normalized H3K4me3 enrichment across a 824 kb region for macaque cerebellum and prefrontal cortex. UCSC-defined transcription start sites (RefSeq TSS) for macaque and human are depicted underneath. **d**, Box plot showing the log-fold differences in gene expression values between human and macaque for genes linked to hominin-specific changes and stable GREs in the prefrontal cortex and cerebellum. Dissimilarities between distributions were calculated using a Student's t-test. **e**, Box plots showing normalized gene expression values in cerebellum of genes linked to hominin-specific CB gains across developmental time, for both human and rhesus macaque. Dissimilarities between distributions were calculated using a Student's t-test. **f**, Hierarchical clustering of RNA-Seq samples from white matter tissue of different primates as indicated by the colored side bar. Correlation map colors indicate Pearson correlation coefficients between samples. Pearson distances are represented by the tree on the left. **g**, Box plots showing normalized gene expression values in white matter of different primate species for all genes annotated on the primate genome. Dissimilarities between distributions were calculated using a Student's t-test. **h**, Box plots as in (**g**) but for genes linked to hominin-specific CB gains and stable GREs. Dissimilarities between distributions were calculated using a Student's t-test. **i**, Gene-ontology analysis using GREAT for genes near hominin-specific CB changes. P-values represent significance of enrichment for stated biological process. **j**, ChIP-Seq tracks showing rpm normalized H3K27ac enrichment in prefrontal cortex samples for different primate species as indicated in a 118 kb region surrounding the *ST3GAL5* gene. Hominin-specific PFC gains are highlighted in green. Bottom and top of the box plots are the first and third quartile. The line within the boxes represents the median and whiskers denote interval within 1.5x the interquartile range from the median. Source data are provided as a Source Data file.

Supplementary Fig. 8 Hominin-specific gains are associated with neurological disease. **a**, Heritability enrichment estimates for selected GREs and neurological traits using stratified LD score regression. Heatmap color indicate the proportion of heritability divided by the proportion of SNPs per selected set of GREs and are represented with associated p-value if $p < 0.01$. **b**, Bar plots showing enrichment estimate for hominin-specific changes across different neurological traits. Error bars represent standard error around the estimate of enrichment. AD: Alzheimer's Disease, ASD: autism spectrum disorder, BIP: bipolar disorder, PD: Parkinson's disease, SCZ: schizophrenia. **c**, ChIP-Seq track showing rpm normalized H3K27ac enrichment across a 538 kb region containing the *DNMT3A* gene in prefrontal cortex for four primate species as indicated. Hominin-specific PFC gain highlighted in green. **d**, ChIP-Seq track showing rpm normalized H3K27ac enrichment across a 195 kb region containing the *CAMK2A* gene in prefrontal cortex for four primate species as indicated. Hominin-specific PFC gain highlighted in green. **e**, ChIP-Seq track showing rpm normalized H3K27ac enrichment across a 250 kb region containing the *c-MET* gene in prefrontal cortex for four primate species as indicated. Hominin-specific PFC gains are highlighted in green.

Supplementary Fig. 9 Cell type specificity for species-specific GREs deregulated in ASD. **a**, Box plots showing the log-fold differences in normalized H3K27ac enrichment between ASD and AD patients and healthy control brains for all identified GREs, regions either NeuN- or NeuN+ specific and regions identified as hominin-specific PFC gains deregulated in ASD. Dissimilarity between the distributions was calculated using a Student's t-test. Bottom and top of all box plots are the first and third quartile. The line within the boxes represents the median and whiskers denote interval within 1.5x the interquartile range from the median. **b**, Metaplot analysis showing the average H3K27ac enrichment profile in different cell types for regions differentially enriched in either autism spectrum disorder (ASD) or Alzheimer's disease (AD) brains compared to control samples and regions identified as hominin-specific PFC gain deregulated in ASD. **c**, Heatmap showing rpm normalized H3K27ac enrichment in the PFC of different primate species as indicated, on scaled human and chimpanzee-specific PFC regulatory elements. **d**, Heatmap showing the enrichment of disease associated GREs measured to be deregulated in autism spectrum disorder and Alzheimer's disease patient brains in correlation to human and chimpanzee-specific changes found in cerebellum (CB) and prefrontal cortex (PFC). Color coding indicates odds ratio of enrichment compared to all identified GRE. P-values were determined using a Fisher's exact test and represented in the heatmap if $p < 0.01$. **e**, Barplot showing cell type specificity, based on single cell ATAC, of regions identified to be down regulated in autism patients brains. Shown are: all ASD down regions and those that overlap either hominin, human or chimpanzee-specific gains. Difference in oligodendrocyte-specific frequency was calculated using a Fisher's exact test. **f**, Barplot as in (e) but for all GREs that are either hominin, human or chimpanzee-specific gain. **g**, Heatmap showing the enrichment of disease associated GREs measured to be deregulated in autism spectrum disorder (ASD) and Alzheimer's disease (AD) patient brains in correlation to hominin-specific changes found in white matter (WM). Color coding indicates odds ratio of enrichment compared to all identified GRE. P-values were determined using a Fisher's exact test and represented in the heatmap if $p < 0.01$. **h**, Metaplot analysis showing the average H3K27ac enrichment profile in different cell types for hominin-specific white matter gains. Source data are provided as a Source Data file.

Supplementary Data 1. Donor Information

Donor	Animal Code	Age	Sex	CoD	PMD	Source
Common Marmoset 1	M09057	5 years	Female	Euthanasia	NA	BPRC
Common Marmoset 2	M09101	5 years	Female	Euthanasia	NA	BPRC
Common Marmoset 3	M14041	8 Months	Female	Euthanasia	NA	BPRC

CoD: Cause of Death
PMD: Post Mortem Delay
BPRC: Biomedical Primate Research Center

Supplementary Data 2. Public data accessions

Data description	Portal	ID	Reference
H3K27ac ChIP in human, chimpanzee and macaque brains	Gene Expression Omnibus	GSE67978	Vermunt MW, et al., 2016
H3K27ac ChIP of human tissues	Gene Expression Omnibus	GSE16256	Roadmap Epigenomics Consortium RE, et al., 2015
Cell type specific H3K27ac ChIP	Synapse	syn12034263	Kozlenkov A, et al., 2018
H3K27ac ChIP in autism patients & healthy controls	Synapse	syn4587616	Sun W, et al., 2016
H3K27ac ChIP in Alzheimer patients & healthy controls	Gene Expression Omnibus	GSE102538	Marzi SJ, et al., 2018
Epigenome in human brain across developmental time	Synapse	syn17092080	Li M, et al., 2018
single-cell H3K27ac ChIP on (non)-neuronal cells in human brain	Synapse	syn5699145	Girdhar K, et al., 2017
H3K27ac ChIP of microglia	NCBI BioProject	PRJNA388724	Gosselin D, et al., 2017
H3K27ac ChIP of Astrocytes	ENCODE project	GSE29611	The ENCODE Project Consortium, et al 2012
Open Chromatin by DNaseI in Frontal Cortex	ENCODE project	GSE32970	The ENCODE Project Consortium, et al 2012
single-cell open chromatin by ACAT in Frontal Cortex	Gene Expression Omnibus	GSE97942	Lake BB, et al., 2018
RNA-seq across developmental time in human and macaque brains	NCBI BioProject	PRJNA448973	Zhu Y, et al, 2018
RNA-seq in adult brains of human, chimpanzee and macaque	Gene Expression Omnibus	GSE20352	Brawand D, et al., 2011

Supplementary Data 3. Sequence information

Sample Name	Species	Tissue	Assay	Antibody	Rep	Genome	Total reads	Unique reads	Mac2 Peaks	FRiP Score	GEO ID	Reanalysed form
Marmoset_Cerebellum_H3K27ac_1	Marmoset	CB	ChIP	H3K27ac	1	callac3	15777644	10343094	34070	20.38%	CSM3755466	
Marmoset_Cerebellum_H3K27ac_2	Marmoset	CB	ChIP	H3K27ac	2	callac3	14070187	11174561	29346	11.83%	CSM3755467	
Marmoset_Cerebellum_H3K27ac_3	Marmoset	CB	ChIP	H3K27ac	3	callac3	33948738	26626784	44723	22.82%	CSM3755468	
Marmoset_PrefrontalCortex_H3K27ac_1	Marmoset	PFC	ChIP	H3K27ac	1	callac3	36178201	20240326	37084	10.01%	CSM3755469	
Marmoset_PrefrontalCortex_H3K27ac_2	Marmoset	PFC	ChIP	H3K27ac	2	callac3	37292103	29385809	42138	11.44%	CSM3755470	
Marmoset_PrefrontalCortex_H3K27ac_3	Marmoset	PFC	ChIP	H3K27ac	3	callac3	39702004	29972436	21702	4.66%	CSM3755471	
Marmoset_Cerebellum_H3K4me3_1	Marmoset	CB	ChIP	H3K4me3	1	callac3	11709320	8819213	22482	31.82%	CSM3755472	
Marmoset_Cerebellum_H3K4me3_2	Marmoset	CB	ChIP	H3K4me3	2	callac3	11559319	6549852	16779	8.75%	CSM3755473	
Marmoset_PrefrontalCortex_H3K4me3_1	Marmoset	PFC	ChIP	H3K4me3	1	callac3	10485374	7781895	22491	27.82%	CSM3755474	
Marmoset_PrefrontalCortex_H3K4me3_2	Marmoset	PFC	ChIP	H3K4me3	2	callac3	16317000	12550519	22630	19.35%	CSM3755475	
Marmoset_Cerebellum_Input	Marmoset	CB	ChIP	Input	1	callac3	23580385	15664518	NA	NA	CSM3755476	
Marmoset_PrefrontalCortex_Input	Marmoset	PFC	ChIP	Input	1	callac3	13937158	9813827	NA	NA	CSM3755477	GSM1119154
Human_Cerebellum_H3K27ac_1	Human	CB	ChIP	H3K27ac	1	hg38	36465551	24779401	48249	41.06%	CSM3755478	
Human_Cerebellum_H3K27ac_2	Human	CB	ChIP	H3K27ac	2	hg38	53639765	36788418	67289	51.20%		GSM1119212
Human_Cerebellum_H3K27ac_3	Human	CB	ChIP	H3K27ac	3	hg38	28994680	16915886	45083	30.24%		GSM1119166
Human_PrefrontalCortex_H3K27ac_1	Human	PFC	ChIP	H3K27ac	1	hg38	22296757	17746904	44779	16.80%		GSM1660051
Human_PrefrontalCortex_H3K27ac_2	Human	PFC	ChIP	H3K27ac	2	hg38	28528907	23422310	59686	25.64%		GSM1119220
Human_PrefrontalCortex_H3K27ac_3	Human	PFC	ChIP	H3K27ac	3	hg38	23196905	8931133	17006	4.35%		GSM1821388
Human_PrefrontalCortex_H3K4me3_1	Human	PFC	ChIP	H3K4me3	1	hg38	12449161	8767170	27179	29.27%		GSM1660022
Macaque_Cerebellum_H3K27ac_1	Macaque	CB	ChIP	H3K27ac	1	rheMac8	13705331	10950425	37594	22.70%		GSM1660023
Macaque_Cerebellum_H3K27ac_2	Macaque	CB	ChIP	H3K27ac	2	rheMac8	11990753	9522577	36151	22.10%		GSM1660024
Macaque_Cerebellum_H3K27ac_3	Macaque	CB	ChIP	H3K27ac	3	rheMac8	26944025	20540341	50249	30.30%		

Supplementary Data 3 (continued). Sequence information

Macaque_PrefrontalCortex_H3K27ac_1	Macaque	PFC	ChIP	H3K27ac	1	rheMac8	19900219	15893436	41049	14.48%	GSM1660025
Macaque_PrefrontalCortex_H3K27ac_2	Macaque	PFC	ChIP	H3K27ac	2	rheMac8	15758242	11078753	22243	6.50%	GSM1660026
Macaque_PrefrontalCortex_H3K27ac_3	Macaque	PFC	ChIP	H3K27ac	3	rheMac8	33575320	25560363	57282	18.22%	GSM1660027
Chimpanzee_Cerebellum_H3K27ac_1	Chimpanzee	CB	ChIP	H3K27ac	1	panTro5	18920322	14297886	32485	15.35%	GSM1821400
Chimpanzee_Cerebellum_H3K27ac_2	Chimpanzee	CB	ChIP	H3K27ac	2	panTro5	20162613	14884584	28761	10.71%	GSM1821401
Chimpanzee_PrefrontalCortex_H3K27ac_1	Chimpanzee	PFC	ChIP	H3K27ac	1	panTro5	20681368	15399838	34007	18.21%	GSM1821406
Chimpanzee_PrefrontalCortex_H3K27ac_2	Chimpanzee	PFC	ChIP	H3K27ac	2	panTro5	18618647	13779378	32708	13.05%	GSM1821407
Human_WhiteMatter_RNAseq_1	Human	WM	RNA	NA	1	hg38	28732038	3767451	NA	NA	GSM3755479
Human_WhiteMatter_RNAseq_2	Human	WM	RNA	NA	2	hg38	30303680	6245283	NA	NA	GSM3755480
Chimpanzee_WhiteMatter_RNAseq_1	Chimpanzee	WM	RNA	NA	1	panTro5	28127951	4769441	NA	NA	GSM3755481
Chimpanzee_WhiteMatter_RNAseq_2	Chimpanzee	WM	RNA	NA	2	panTro5	24391066	2705467	NA	NA	GSM3755482
Macaque_WhiteMatter_RNAseq_1	Macaque	WM	RNA	NA	1	rheMac8	25830619	6449453	NA	NA	GSM3755483
Macaque_WhiteMatter_RNAseq_2	Macaque	WM	RNA	NA	2	rheMac8	31487620	8848574	NA	NA	GSM3755484
Marmoset_WhiteMatter_RNAseq_1	Marmoset	WM	RNA	NA	1	callJac3	40821726	11877598	NA	NA	GSM3755485
Marmoset_WhiteMatter_RNAseq_2	Marmoset	WM	RNA	NA	2	callJac3	32793030	12568247	NA	NA	GSM3755486
Human_WhiteMatter_H3K27ac_1	Human	WM	ChIP	H3K27ac	1	hg38	24175260	16939091	9023	1.62%	GSM4089120
Human_WhiteMatter_H3K27ac_2	Human	WM	ChIP	H3K27ac	2	hg38	29035691	20433617	15745	2.74%	GSM4089121
Human_WhiteMatter_H3K27ac_3	Human	WM	ChIP	H3K27ac	3	hg38	33993601	23967317	25707	4.80%	GSM4089122
Chimpanzee_WhiteMatter_H3K27ac_1	Chimpanzee	WM	ChIP	H3K27ac	4	panTro5	19312940	12817406	17487	4.43%	GSM4089123
Chimpanzee_WhiteMatter_H3K27ac_2	Chimpanzee	WM	ChIP	H3K27ac	5	panTro5	32475765	22551091	29678	5.80%	GSM4089124
Chimpanzee_WhiteMatter_H3K27ac_3	Chimpanzee	WM	ChIP	H3K27ac	6	panTro5	31771720	22187212	36734	10.44%	GSM4089125
Macaque_WhiteMatter_H3K27ac_1	Macaque	WM	ChIP	H3K27ac	7	rheMac8	28674986	19287171	12770	2.27%	GSM4089126
Macaque_WhiteMatter_H3K27ac_2	Macaque	WM	ChIP	H3K27ac	8	rheMac8	30919268	20450752	19870	3.42%	GSM4089127
Macaque_WhiteMatter_H3K27ac_3	Macaque	WM	ChIP	H3K27ac	9	rheMac8	28126109	18576681	19614	3.83%	GSM4089128

Supplementary Data 3 (continued). Sequence information

Marmoset_WhiteMatter_H3K27ac_1	Marmoset	WM	ChIP	H3K27ac	10	callac3	30562459	21812606	33424	7.98%	GSM4089129
Marmoset_WhiteMatter_H3K27ac_2	Marmoset	WM	ChIP	H3K27ac	11	callac3	36960153	26191125	41844	9.88%	GSM4089130
Marmoset_WhiteMatter_H3K27ac_3	Marmoset	WM	ChIP	H3K27ac	12	callac3	30362933	21069387	35328	10.21%	GSM4089131
Human_PrefrontalCortex_NeuNnegative_ATAC_1	Human	PFC	ATAC	NA	1	hg38	37184630	24752514	28125	7.43%	GSM4089132
Human_PrefrontalCortex_NeuNpositive_ATAC_1	Human	PFC	ATAC	NA	1	hg38	67188052	35797922	91492	19.21%	GSM4089133
Human_PrefrontalCortex_NeuNnegative_ATAC_2	Human	PFC	ATAC	NA	2	hg38	31182598	18312781	35805	10.65%	GSM4089134
Human_PrefrontalCortex_NeuNpositive_ATAC_2	Human	PFC	ATAC	NA	2	hg38	52090464	19037363	13754	2.52%	GSM4089135
Chimpanzee_PrefrontalCortex_NeuNnegative_ATAC_4	Chimpanzee	PFC	ATAC	NA	4	panTro5	38781601	22826404	47653	18.66%	GSM4089136
Chimpanzee_PrefrontalCortex_NeuNpositive_ATAC_4	Chimpanzee	PFC	ATAC	NA	4	panTro5	27450839	10708886	28157	6.84%	GSM4089137
Chimpanzee_PrefrontalCortex_NeuNnegative_ATAC_5	Chimpanzee	PFC	ATAC	NA	5	panTro5	41285772	19642867	32168	14.16%	GSM4089138
Chimpanzee_PrefrontalCortex_NeuNpositive_ATAC_5	Chimpanzee	PFC	ATAC	NA	5	panTro5	31190732	12059991	27347	9.11%	GSM4089139
Macaque_PrefrontalCortex_NeuNnegative_ATAC_2	Macaque	PFC	ATAC	NA	2	rheMac8	22572579	4773206	26011	17.32%	GSM4089140
Macaque_PrefrontalCortex_NeuNpositive_ATAC_2	Macaque	PFC	ATAC	NA	2	rheMac8	28217173	12091448	52525	24.19%	GSM4089141
Macaque_PrefrontalCortex_NeuNnegative_ATAC_3	Macaque	PFC	ATAC	NA	3	rheMac8	28151278	11809410	46906	32.67%	GSM4089142
Macaque_PrefrontalCortex_NeuNpositive_ATAC_3	Macaque	PFC	ATAC	NA	3	rheMac8	26071130	16795833	90754	30.51%	GSM4089143
Marmoset_PrefrontalCortex_NeuNnegative_ATAC_1	Marmoset	PFC	ATAC	NA	1	callac3	30612892	16319082	23090	6.86%	GSM4089144
Marmoset_PrefrontalCortex_NeuNpositive_ATAC_1	Marmoset	PFC	ATAC	NA	1	callac3	19006548	10349562	5844	1.27%	GSM4089145
Marmoset_PrefrontalCortex_NeuNnegative_ATAC_2	Marmoset	PFC	ATAC	NA	2	callac3	21928305	4786839	8057	3.83%	GSM4089146
Marmoset_PrefrontalCortex_NeuNpositive_ATAC_2	Marmoset	PFC	ATAC	NA	2	callac3	25673925	13881479	26079	5.98%	GSM4089147

CB: Cerebellum

PFC: Prefrontal Cortex

WM: White Matter

Rep: replicate

FRI: Fraction of Reads in Peaks

Supplementary Data 4. Genomic composition of identified GREs

	# regions	# TSS	# exonic	# intronic	# intergenic
All GREs	37308	9292	6857	11411	9748
PFC hominin gains	685	53	163	264	205
PFC hominin losses	158	15	17	52	74
CB hominin gains	713	101	216	187	209
CB hominin losses	274	26	68	147	133
Stable GREs	8441	3781	1027	2167	1466

CB: Cerebellum
PFC: Prefrontal Cortex
GRE: Gene Regulatory Element
TSS: Transcriptional Start Site

Supplementary Data 5. Motif analysis on hominin-specific prefrontal cortex changes

Motif	p-val	% targets	% background	Best match
Hominin gains prefrontal cortex
1e ⁻¹⁶	15.55	7.78	NeuroD1(bHLH)/Islet-NeuroD1-ChIP-Seq(GSE30298)/Homer(0.902)
1e ⁻¹⁵	0.64	0	POL004.1_CCAAT-box/Jaspar(0.591)
1e ⁻¹⁴	12.42	5.98	Cdx2(Homeobox)/mES-Cdx2-ChIP-Seq(GSE14586)/Homer(0.643)
1e ⁻¹³	23.64	15.06	FO XK2(Forkhead)/U2OS-FO XK2-ChIP-Seq(E-MTAB-2204)/Homer(0.753)
12e ⁻¹²	0.55	0	PB0110.1_Bcl6b_2/Jaspar(0.702)
12e ⁻¹²	2.67	0.48	PB0148.1_Mtf1_2/Jaspar(0.609)

Supplementary Data 5 (continued). Motif analysis on hominin-specific cerebellum changes

Motif	P-val	% Targets	% Background	Best Match
Hominin gains cerebellum
1e-14	7.71	2.52	MafA(bZIP)/Islet-MafA-ChIP-Seq(GSE30298)/Homer(0.754)
1e-13	3.08	0.43	RARg(NR)/ES-RARg-ChIP-Seq(GSE30538)/Homer(0.654)
1e-13	5.46	1.40	GATA3/MA0037.2/Jaspar(0.574)
1e-12	0.83	0.01	ISL2/MA0914.1/Jaspar(0.581)
1e-12	7.00	2.32	Pbx3(Homeobox)/GM12878-PBX3-ChIP-Seq(GSE32465)/Homer(0.711)
1e-12	1.19	0.03	NFIL3/MA0025.1/Jaspar(0.797)
1e-12	1.07	0.02	PH0026.1_Duxbl/Jaspar(0.670)

Supplementary Data 6. Gene ontology analysis of hominin-specific changes

Term	FDR	Category
Hominin-specific changes cerebellum		
cell-substrate adhesion	1.39e ⁻³	GO Biological Process
negative regulation of intracellular protein kinase cascade	1.58e ⁻³	GO Biological Process
regulation of blood pressure	1.57e ⁻³	GO Biological Process
positive regulation of Rho GTPase activity	1.89e ⁻³	GO Biological Process
nitric oxide metabolic process	3.89e ⁻³	GO Biological Process
glycogen catabolic process	1.96e ⁻²	GO Biological Process
detection of calcium ion	4.90e ⁻²	GO Biological Process
focal adhesion	1.58e ⁻²	GO Cellular Component
cell-substrate adherens junction	1.89e ⁻²	GO Cellular Component
laminin complex	2.28e ⁻²	GO Cellular Component
increased systemic arterial blood pressure	4.55e ⁻³	Mouse Phenotype
abnormal lung vasculature morphology	1.68e ⁻²	Mouse Phenotype
increased interleukin-12a secretion	3.38e ⁻²	Mouse Phenotype
cyclopia	4.79e ⁻²	Mouse Phenotype
abnormal circulating glucagon level	4.86e ⁻²	Mouse Phenotype
primary bacterial infectious disease	4.63e ⁻⁷	Disease Ontology
heart failure	6.46e ⁻⁵	Disease Ontology
colonic disease	1.72e ⁻⁴	Disease Ontology
aspergillosis	6.26e ⁻³	Disease Ontology
invasive pulmonary aspergillosis	1.88e ⁻²	Disease Ontology
JAK/STAT signaling pathway	7.36e ⁻³	PANTHER Pathway
Hominin-specific changes prefrontal cortex		
sialylation	3.52e ⁻⁴	GO Biological Process
Schwann cell differentiation	3.00e ⁻²	GO Biological Process
anterior/posterior axon guidance	3.13e ⁻²	GO Biological Process
netrin-activated signaling pathway	4.82e ⁻²	GO Biological Process
integral to Golgi membrane	1.25e ⁻⁶	GO Cellular Component
abnormal pulmonary alveolus morphology	1.11e ⁻²	Mouse Phenotype
demyelinating disease	7.77e ⁻³	Disease Ontology
demyelinating disease of central nervous system	6.13e ⁻³	Disease Ontology
multiple sclerosis	4.99e ⁻³	Disease Ontology
celiac disease	9.68e ⁻³	Disease Ontology
allergy	1.46e ⁻²	Disease Ontology
food allergy	2.10e ⁻²	Disease Ontology

Chapter 3

Recently evolved enhancers emerge with high interindividual variability and less frequently associate with disease

Bas Castelijns^{1*}, Mirna L. Baak^{1*}, Geert Geeven^{1*}, Marit W. Vermunt¹,
Caroline R.M. Wiggers^{1,2}, Ilia S. Timpanaro¹, Ivanela Kondova³, Wouter de Laat¹,
Menno P. Creyghton^{1,4}

*These authors contributed equally to this work

¹Hubrecht Institute-KNAW & University Medical Center Utrecht,
Utrecht, The Netherlands

²Division of Pediatrics, University Medical Center Utrecht,
Utrecht, The Netherlands

³Biomedical Primate Research Center,
Rijswijk, The Netherlands

⁴Erasmus University Medical Center, Department of Developmental Biology,
Rotterdam, The Netherlands

Abstract

Mutations in non-coding regulatory DNA such as enhancers underlie a wide variety of diseases including developmental disorders and cancer. As enhancers rapidly evolve, understanding their function and configuration in non-human disease models can have important clinical applications. Here, we analyze enhancer configurations in tissues isolated from the common marmoset, a widely used primate model for human disease. Integrating these data with human and mouse data, we find that enhancers containing trait-associated variants are preferentially conserved. In contrast, most human-specific enhancers are highly variable between individuals, with a subset failing to contact promoters. These are located further away from genes and more often reside in inactive B-compartments. Our data show that enhancers typically emerge as instable elements with minimal biological impact prior to their integration in a transcriptional program. Furthermore, our data provide insight into which trait variations in enhancers can be faithfully modeled using the common marmoset.

Introduction

The gene expression programs that dictate cellular behavior are controlled by non-coding gene regulatory elements (GREs) such as promoters and enhancers¹. Alterations in GREs lead to a panoply of diseases, including developmental disorders and cancer². Furthermore, many disease-associated variants reside in enhancer sequences suggesting that complex diseases may be associated with combinations of common variants that modulate enhancer activity³. As such, the contributions of several common enhancer variants to disease susceptibility have now been resolved biologically⁴. During evolution, coding sequences remain relatively stable, while enhancer elements that regulate the expression of genes in an often cell-type-specific manner show rapid evolutionary turnover^{5,6}. As this rapid enhancer turnover is a hallmark of mammalian evolution, it is unclear to what extent diseases that are driven in part by enhancer variants or mutations can be correctly recapitulated in non-human model systems. Therefore, elucidation of their configuration in model systems is required to be able to determine whether disease states can be faithfully captured.

A promising animal model that is gaining attention to study human disease is the common marmoset (*Callithrix jacchus*), a new-world primate originating from the northeast of Brazil^{7,8}. Due to its relatively small size, short gestation period, and early sexual maturation, it is increasingly being proposed as a more suitable model compared to mice or larger primates to study a variety of diseases such as age-related diseases, immunological disorders, and most prominently research into neuropsychiatric and neurodegenerative disorders^{7,8}. To facilitate these efforts high-quality genome assemblies as well as transcriptome data have recently been

generated^{9,10}. Nevertheless, insight into the gene regulatory network underlying these gene expression programs is still lacking. As many disease-associated variants are located in non-coding regulatory elements, it is thus crucial to understand the differences and similarities in gene regulation with humans to fully exploit the common marmoset as non-human primate model for biomedical studies.

Here we annotate putative promoters and enhancers in distinct marmoset tissues and compare these to human and mouse data. We provide evidence that repurposed enhancers are in part misclassified interindividual variabilities. Following this, we find that most enhancers that are new in the human lineage emerge with high interindividual variability and often fail to contact genes, providing insight into their attenuated effect on gene expression¹¹. Consequently, enhancers containing DNA variants that link to trait variation or disease susceptibility tend to be more often conserved between species. Thus, our analyses provide new insights into the evolution and function of gene regulatory networks and provide a framework for modeling disease variant containing regulatory elements in the common marmoset.

Results

Annotation of regulatory DNA in marmoset tissue

To analyze gene regulation in marmosets, we annotated GREs in eight different marmoset tissues (colon, heart, kidney, liver, pancreas, skeletal muscle, spleen, stomach) and two brain regions (cerebellum and prefrontal cortex¹² (Fig. 1a). We used chromatin immunoprecipitation followed by sequencing (ChIP-Seq) for histone 3 lysine 27 acetylation (H3K27ac), a robust mark with a good balance between sensitivity and specificity (both ~70%) to identify active promoters and enhancers^{6,13–16}. H3K27ac serves as a footprint for activity of its acetyltransferase, p300/CBP, which is sufficient to convey enhancer activity upon a genomic region¹⁷ and responsible for the acetylation of several histone residues¹⁸ as well as the polymerase complex itself thus affecting transcriptional output^{19,20}. In addition, we performed ChIP-Seq for histone 3 lysine 4 tri-methylation (H3K4me3), which is specifically found on active transcriptional start sites (TSSs)²¹, to distinguish between putative active promoters and putative active enhancers. Data were within quality standards²² and reproducible between biological replicates (H3K27ac: average $p = 0.86$; H3K4me3: average $p = 0.96$, Fig 1b, c, Supplementary data 1) and data published previously for marmoset liver⁶ (Supplementary Fig. 1a, b).

In total, we annotated 60,824 H3K27ac-enriched and 21,136 H3K4me3-enriched regions in the marmoset genome across all marmoset tissues combined (Supplementary Data 2, Supplementary Fig. 1c). Most (55%) of the H3K4me3-enriched GREs overlap an annotated marmoset TSS (Supplementary Fig. 1d). Moreover, as the human genome is more extensively annotated, mapping of all human TSSs on the marmoset genome showed that ~75% of the H3K4me3-enriched regions that were annotated in marmoset overlay a marmoset and/or human

Fig. 1 Conservation of regulatory elements across the primate lineage. **a**, Schematic representation of the common marmoset and all the tissues sampled. **b**, Hierarchical clustering of all sampled tissues based on H3K27ac enriched regions ($n = 60,824$). Correlation map colors indicate Pearson's correlation coefficients between samples. Side bars are color coded to indicate different tissues. Pearson distances are represented by the tree on the left. (CB: cerebellum, PFC: prefrontal cortex, Skel. Mucl: skeletal muscle). **c**, Hierarchical clustering as in (b) based on H3K4me3 enriched regions ($n = 21,136$). **d**, Heatmap depicting the H3K27ac enrichment in different tissues across 5' to 3' scaled tissue-specific GREs, centered around the H3K27ac peak. Heatmap colors indicate rpk normalized H3K27ac enrichment. **e**, Bar plot depicting the percentage of human GRE sequences that could be mapped to marmoset and mouse genomes. Dissimilarity between the ratio was calculated with a Fisher's exact test. **f**, Pie charts showing the degree of shared activity between marmoset and human for both H3K27ac and H3K4me3 enriched regions. **g**, Pie charts showing the degree of repurposed activity between human and marmoset for both H3K27ac and H3K4me3 enriched regions. **h**, ChIP-Seq tracks of rpm normalized H3K27ac enrichment in different tissues from both human and marmoset samples across an 8 kb region within the *TMEM175* gene. The repurposed GRE is highlighted in green. **i**, Bar plot depicting the percentage of GREs that have variable H3K27ac enrichment between different human LCL samples for both conserved as well as repurposed regions. Dissimilarity between the ratio was calculated with a Fisher's exact test.

TSS (Supplementary Fig. 1d). Of the remaining H3K4me3-enriched regions, the majority (67%) are located in non-coding DNA (Supplementary Fig. 1e), of which most (84%) are actively transcribed, as determined using RNA-sequencing (RNA-Seq) data derived from different marmoset tissues (Supplementary Fig. 1f)²³. These

could therefore represent marmoset-specific splice variants, novel genes or actively transcribed GREs in the marmoset genome. Mapping of these regions to the human genome revealed that these regions are enriched for species-specific DNA sequences not found in humans (4.7-fold enrichment, $p < 2.2e^{-16}$, fisher exact).

Of all 60,824 H3K27ac-enriched regions in the marmoset genome 42,128 did not co-localize with H3K4me3 enrichment, thus representing putative active enhancers^{6,14–16}, although additional analyses are required for each putative active enhancer or promoter separately to confirm its activity within the genome. For the purpose of simplicity throughout the manuscript, H3K27ac-enriched regions will be referred to as active GREs, or when specified, as active promoters and enhancers. Most enhancers were found active in only a single tissue, consistent with their known tissue specificity (Fig. 1d, Supplementary Fig. 1g). Functional analysis of genes in proximity of active enhancers, as defined by GREAT²⁴, reflected the tissues in which they were identified (Supplementary Fig. 1h).

A proportion of repurposed GREs are misclassified interindividual variabilities

To analyze to what extent human regulatory elements can be faithfully modelled in the common marmoset, we compared our data to human datasets of matching tissues^{6,25,26}. Using the same analysis as was done for marmoset samples, we identified 72,900 H3K27ac and 22,426 H3K4me3-enriched regions in human samples. Consistent with the evolutionary distance to their common ancestors, 39.7% of human GREs could not be mapped to the mouse genome, while only 21.4% of human GREs could not be mapped to the marmoset genome (Fig. 1e). We combined all marmoset and human GREs by selecting GREs that could be mapped on both species' genomes (39,153 regions) using reciprocal liftover (see methods) (Fig. 1f). Of these shared regions, 89.8% were enriched in the same tissue between the species (Fig. 1g). For example, GREs linked to *KLF5* were enriched in human colon and stomach, a tissue-specific pattern that was conserved in marmoset tissues (Supplementary Fig. 2a). Consistent with previous data^{27,28}, we found a fraction of H3K27ac enriched regions (10.2%) that were repurposed between tissues and species (Fig. 1g), with no major influence of confounder variables observed (Supplementary Fig. 2b). Repurposed GREs are elements that lose activity in one tissue while gaining it in another between two species. For instance, an enhancer in the *TMEM175* gene, a risk factor in Parkinson's disease²⁹ (Fig. 1h), switches its broad H3K27ac signal specifically to the human brain while being enriched in several other tissues in marmoset. Closer analysis of these regions revealed that most repurposing occurs between tissue-specific GREs (Supplementary Fig. 2c). To assess whether this repurposing was specific between human and marmoset, we compared our data to H3K27ac enrichment in matching mouse tissues²⁶. We observed that human-marmoset repurposed GREs are mainly specific to the primate lineage since these regions were often either depleted from H3K27ac signal in mouse tissues (53%) or their sequence could not be mapped onto the mouse genome (13.1%) ($p < 2.2e^{-16}$,

Fig. 2 Variability of enhancer is linked to evolutionary novelty. **a**, t-Stochastic Neighbor Embedding (t-sne) analysis of 10 human, 7 chimpanzee and 7 macaque samples, based on a non-redundant list of H3K27ac enriched regions mappable on all three species ($n = 54,793$). **b**, Heatmap of 911 evolutionary new enhancers and 2436 enhancer gains in human compared to chimpanzee and macaque. Read counts were normalized for library size and length of the region. Individual samples per species are aligned at the x-axis, H3K27ac enriched regions are depicted on the y-axis. Heatmap color represents normalized H3K27ac enrichment. **c**, Box plot depicting gene expression in human, chimpanzee and macaque for genes linked to evolutionary enhancer gains. Values are log₂ zero-mean normalized RNA-Seq counts. Enhancers were linked based on proximity. Dissimilarity between the distributions was calculated using a Student's t-test. **d**, Bar plot depicting the percentage of GREs that have variable H3K27ac enrichment in different human LCL samples for different categories of enhancers as indicated. Dissimilarities between the ratios was calculated using a Fisher's exact test. **e**, ChIP-Seq tracks of rpm normalized H3K27ac enrichment in LCL cells from different primate species as indicated across a 22 kb region containing the *ING1* gene. A human-specific GRE is highlighted in green. A conserved GRE is highlighted in grey. **f**, ChIP-Seq tracks showing the same region as in (e) for multiple human LCL samples from distinct individuals and distinct ancestries. The same human-specific GRE as in (e) is highlighted in green.

Supplementary Fig. 2d, e). However, a proportion of human-marmoset repurposed elements (5.8%) showed conservation of H3K27ac enrichment between human and mouse as well as between marmoset and mouse, e.g. the mouse enhancer was enriched in both tissues in mouse (called 'ambiguous', Supplementary Fig. 2d). For example, an enhancer in the *FLNB* gene was annotated as repurposed from cerebellum and liver to heart tissue between human and marmoset (Supplementary Fig. 2f). Instead, analysis of the mouse data showed that this enhancer was enriched

Fig. 3 Cell type specificity of novel GREs. **a**, Heatmap showing rpkm normalized H3K27ac enrichment in different cell types on all GREs defined as LCL cell-type-specific ($n = 11,227$). Regions are scaled 5' to 3' and centered around the H3K27ac peak. Heatmap colors indicate RPKM normalized H3K27ac enrichment. **b**, Heatmap as in **(a)** but for all non-LCL cell-type-specific GREs ($n = 9685$). **c**, Bar plot depicting the percentage of GREs, classified by their evolutionary status as indicated on the axis, that are LCL cell-type-specific or shared with one of the other cell types. Dissimilarities between the ratios were calculated with a Fisher's exact test. **d**, ChIP-Seq tracks of rpm normalized H3K27ac enrichment on LCL samples of different primates as shown, across a 246 kb genomic regions containing several genes and a human-specific GRE (highlighted in green). **e**, ChIP-Seq tracks as in **(d)** for the different human cell types. The same human-specific GRE is highlighted in green. **f**, Bar plot depicting the percentage of cell-type-specific GREs for regions that are variable between human individuals and those that are not variable. Dissimilarity between the ratios was calculated with a Fisher's exact test.

in all of these tissues, suggesting that some repurposing events may rather represent species-specific losses of H3K27ac enrichment.

Surprisingly, we found that elements that were repurposed between human and marmoset and enriched for H3K27ac in any mouse tissue were also more likely to be repurposed to a different tissue in mouse ($p < 2.2e^{-16}$, Supplementary Fig. 2d, e). For example, a GRE within the *GLT8D2* gene is repurposed between human heart and marmoset liver while being enriched in different tissues in mouse (Supplementary Fig. 2g). A similar pattern in H3K27ac enrichment was observed for the promoter of *GLT8D2*. As repurposing the same elements multiple times across species seemed unlikely, we wondered whether these elements were truly repurposed or merely shared between tissues but variable between individuals within a species. To

analyze this, we compared our GREs to data generated in lymphoblastoid cell lines (LCL), analyzing enhancer variability across 19 different human individuals using ANOVA (see methods, Supplementary Fig. 2h) as shown previously³⁰. By assessing variation between samples from different human individuals compared to the variation of replicate samples from the same individuals, we found that GREs that were repurposed were more likely to be variable between LCL samples derived from different individuals ($p = 0.017$, Fig. 1i), suggesting that a proportion of repurposed elements may be misclassified as such due to interindividual variability.

Recently evolved regulatory DNA is predominantly variable between individuals

To further analyze whether there was a relationship between recent evolutionary changes at regulatory DNA and individual variation, we reanalyzed H3K27ac data from ten human, seven chimpanzee, and seven rhesus macaque lymphoblastoid cell lines (LCL)³¹, using the same analysis as the marmoset tissues. This allowed us to analyze individual variation and evolution in the same cell type in a more quantitative manner, while also providing a more precise annotation of when these regulatory elements emerged as compared to directly contrasting human and marmoset. We therefore generated a non-redundant list of H3K27ac-enriched regions based on samples from all three primate species using reciprocal liftover. We annotated 54,793 regions that could be mapped on the genomes of all three primate species (see methods, Fig. 2a), of which 40,924 were classified as putative enhancers based on not overlaying a known TSS. Using DESeq, 3347 gains and 1666 losses of H3K27ac enrichment were identified at enhancers, which were specific to the human lineage (Fig. 2b, Supplementary Fig. 3a), with expression analysis of nearby genes showing the typical correlation of modest expression change with enhancer change (Fig. 2c, Supplementary Fig. 3b).

We sub-classified three distinct classes of putative enhancers. Enhancers that increased H3K27ac enrichment in humans compared to the other two primate species ("gains"), elements that were new in humans based on absence of H3K27ac enrichment in the other two primates using a stringent background model ("new"), and elements that were stable and not differentially enriched across primates ("stable"). For each region, we determined its variability using ANOVA in the panel of 19 human LCL lines as described above³⁰. Surprisingly more than half of the elements that recently evolved were classified as variable between human individuals ($p < 2.2 \times 10^{-16}$, Fig. 2d). Increased variability was observed for both enhancers classified as gains (44.8%) and those classified as new (65.3%). In contrast, elements that were classified as stable across primates were less frequently (25.2%) variable between humans ($p < 2.2 \times 10^{-16}$, Fig. 2d). The increase in variability for recently evolved enhancers was not due to major differences in H3K27ac enrichment at recently evolved elements, as the relationship between variability and H3K27ac enrichment for gains and new enhancers was not significantly different from stable enhancers or enhancers in general ($p = 0.2$, Supplementary Fig. 3c). We also found no indication

that line to line variability was a factor in our analysis as flagging enhancers that were variable between 5 independent LCL lines from the same individual³² did not affect our observations (Supplementary Fig. 3d). In addition, we found no indication that variability of enhancers that were evolutionary new, was linked to differences in ancestry between the individuals analyzed (see methods, Supplementary Fig. 3e-h). For example, a putative enhancer that recently evolved near *ING1* was highly variable across individuals, independent of their ancestry (Fig. 2e, f).

We and others recently established a link between recently evolved enhancers and cell type specificity suggesting that cell type specificity and variability may be linked^{28,33}. By comparing H3K27ac enrichment at putative enhancers in lymphoblastoid cells with H3K27ac datasets generated in 7 unrelated cell types (Fig. 3a, b)³⁴, we confirmed that recently evolved enhancers were more often cell-type-specific (Fig. 3c-e). In addition, we found that cell-type-specific enhancers were also more often variable in LCLs between individuals (28% increase, $p < 2.2 \times 10^{-16}$, Fig. 3f). To test whether there was a direct dependence between all three categorical characteristics; variability, cell type specificity and enhancer type (gain/new/stable/all), we used LLM3D³⁵, a method to assess the interdependences of the different variables of these enhancers. We found that evolutionarily new and variable enhancers differ significantly in their rate of cell type specificity ($p = 2.0 \times 10^{-218}$). This demonstrates that the degree of variability between individuals and cell type specificity are two independent characteristics of recently evolved regulatory elements. We conclude that high interindividual variability is a property of the majority of recently evolved enhancers.

Recently evolved enhancers less frequently engage in functional contacts with gene promoters

As the gain or loss of an enhancer can be buffered by the presence of other enhancers that regulate the same gene^{11,28,36,37}, we next assessed whether enhancer variability was also linked to the number of other enhancers engaging their target gene. As gene proximity is an imperfect measure to couple enhancers to genes, we increased the specificity of our analysis by integration of HiC data from a lymphoblastoid cell line using a new method to determine distal contacts at kilobase resolution^{38,39}. We analyzed 422,144 reciprocal contacts between 155,084 viewpoints (GREs, TSSs and CTCF sites) and their anchors, including 84,855 reciprocal contacts with the H3K27ac-enriched putative enhancer elements. Using direct interactions over proximity-based analysis, we found that enhancers that are evolutionarily new and variable between individuals are not more likely to contact gene promoters with multiple GREs suggesting variability and buffering by alternative enhancers are not directly linked ($p = 0.31$, Supplementary Fig. 4a).

Unexpectedly, we found that new enhancers were less frequently in contact with any other anchor ($p = 2.8 \times 10^{-6}$, Fig. 4a). This modest drop in contact frequency was specific for enhancers that were classified as new and not for existing enhancers

Fig. 4 New enhancers are more frequently unengaged. **a**, Bar plot depicting the percentage of enhancers that have a reciprocal HiC interaction with any anchor for the different evolutionary categories as indicated on the axis. Dissimilarities between the ratios were calculated using a Fisher's exact test. **b**, Bar plot as in (a) showing the percentage of enhancers reciprocally contacting a TSS. **c**, Virtual 4C profile using an enhancer that is evolutionary new and unengaged, located within the *MROH1* gene, as a viewpoint (highlighted in blue). Normalized ChIP-Seq enrichment for H3K27ac (green) and CTCF (red)

that were classified as gains ($p = 0.047$). As both classes of enhancers are more often variable, the specificity for new enhancers to be less frequently engaged is unlikely the result of increased variability at new enhancers. Analyzing this reduction in contact frequency for new enhancers further, we found a much stronger decrease in contact frequency when assessing contacts with TSSs ($p < 2.2 \times 10^{-16}$, Fig. 4b). A similar reduction in TSS contacts was observed for new enhancers when assessing only those enhancers that were significantly enriched for H3K27ac in the GM12878 cell line, in which the HiC data was generated in (Supplementary Fig. 4b). This suggests that the reduction in promoter contacts at new enhancers is not due to the HiC analysis being limited to a single cell line. The opposite effect (increased TSS contact frequency) was observed for enhancers that were classified as stable ($p < 2.2 \times 10^{-16}$, Fig. 4b). Putative enhancers that failed to contact other regions were classified as unengaged enhancers and those that failed to contact promoters as TSS unengaged enhancers. For instance, a putative enhancer in the *MROH1* gene is new in humans and is not engaging any of its surrounding elements above the detection limit despite a variety of active GREs being present in the region (Fig. 4c, Supplementary Fig. 4c).

Further analysis of these unengaged enhancers revealed that they showed attenuated H3K27ac enrichment compared to enhancers that do contact other GREs (Supplementary Fig. 4d). This was not a specific property of enhancers that recently evolved as enrichment of H3K27ac was also lower for unengaged enhancers regardless of their evolutionary status ($p < 2.2 \times 10^{-16}$, Supplementary Fig. 4d). While reduced H3K27ac enrichment could be the result of detection issues at these sites, it could also be the consequence of not engaging a promoter, as promoters are typically highly enriched for this modification. We therefore analyzed enhancers that did engage in detectable contacts with other GREs but not with promoters (TSS-unengaged). We found that TSS-unengaged enhancers that were able to contact other GREs were also less enriched for H3K27ac compared to enhancers that did contact promoters ($p < 2.2 \times 10^{-16}$, Fig. 4d) regardless of their evolutionary status. This argues against detection issues at new and unengaged elements and for a lack of promoter engagement as a basis for the reduction in H3K27ac enrichment seen at

in GM12878 cells are shown in the top two tracks. **d**, Box plots depicting normalized H3K27ac enrichment on GREs that are either contacting a TSS (TSS-engaged) or that are TSS-unengaged. Data is shown for both evolutionary new and for all enhancers. Dissimilarity between the distributions was calculated using a Student's t-test. **e**, Box plots depicting CTCF enrichment (foldchange over input as defined by ENCODE) for enhancers that are evolutionary new in humans and are either unengaged or engaged as indicated on the axis. Dissimilarity between the distributions was calculated using a Student's t-test. **f**, Box plots depicting the maximum STARR-activity of enhancers that are evolutionary new in humans and either unengaged or engaged as indicated on the axis. Dissimilarity between the distributions was calculated using a Student's t-test. **g**, Histograms depicting the absolute distance of human-specific elements to their closest TSS for both engaged (upper) and unengaged (lower) GREs. Dissimilarity between the distributions was calculated using a Student's t-test. **h**, Bar plot depicting the percentage of both engaged and unengaged GREs that are located in either the A- or B-compartments as defined in the GM12878 cell line. Dissimilarity between the ratios was calculated with a Fisher's exact test. **i**, Bar plot as in (**h**) for the different evolutionary enhancer categories as indicated on the x-axis. Dissimilarities between the ratios were calculated with a Fisher's exact test.

these elements. Moreover, we observed only a minor depletion of CTCF binding at unengaged enhancers (Fig. 4e), again suggesting that the lack of engagement and lower H3K27ac enrichment is not due to an inability to detect signal at these regions. Finally, further analysis of enhancer activity of new enhancer regions in humans using mass parallel reporter assays⁴⁰ revealed no difference in intrinsic enhancer activity between engaged and unengaged enhancers (Fig. 4f, Supplementary Fig. 4e). Thus, the lower H3K27ac enrichment at unengaged enhancers can at least in part be attributed to their failure to interact with an H3K27ac enriched promoter and not due to a lack of signal or intrinsic activity.

Interestingly, analyzing the genomic positions of unengaged enhancers revealed that they are typically located further away from promoters compared to regular enhancers ($p = 3.2 \times 10^{-6}$, Fig. 4g). In addition, they are more often found in inactive B compartments, as identified in GM12878 cells³⁹, that are associated with closed and repressive chromatin⁴¹ (Fig. 4h, Supplementary Fig. 4f, g). Enrichment in B-compartments was also observed for new enhancers, regardless of whether they were specific to the GM12878 cell line (Fig. 4i, Supplementary Fig. 4h). As expected, gene expression of proximal genes was not impacted by the emergence of new enhancers that are unengaged (Supplementary Fig. 4i, j). Therefore, we propose that recently evolved enhancers more often emerge further away from transcriptional start sites and are less likely to engage in productive contacts. Thus, even though these regions are intrinsically active they are less likely to impact gene expression.

Enhancers containing trait associated variants are more often conserved

As enhancers that recently evolved within a species are less frequently integrated in a regulatory network and may therefore be more often biologically irrelevant, we asked what that would mean for the utility of marmosets as non-human model organism for human disease. Using the tissue datasets described above, we assessed which human GREs contained known variants (single nucleotide polymorphisms [SNPs]) that are coupled to known phenotypic traits including various diseases (see methods, Fig. 5a) and assessed their conservation in both marmoset and mouse tissues. We observed that H3K27ac enrichment of trait variant containing GREs was more likely to be conserved between species across evolution compared to GREs without trait variants ($p < 2.2 \times 10^{-16}$, Fig. 5b). In contrast, enhancers that are new in humans are also depleted for trait variants (Fig. 5c). For example, the *BSN* gene, a regulator of neurotransmission, contains several brain-specific GREs that are conserved between human, marmoset and mouse (Fig. 5d). These contain several SNPs associated with neurodegenerative diseases such as Parkinson's disease and Alzheimer's disease⁴².

In agreement with trait variants occurring preferentially at conserved enhancers, we also observed that these variant containing enhancers were less frequently variable within the human lineage compared to other GREs (Fig. 5e). This may be counterintuitive as variation is expected in trait variant containing GREs³⁰. However,

Fig. 5 DNA trait variant containing GREs are conserved across evolution. **a**, Pie chart depicting the number of human GREs that contain a known trait variants (SNP). **b**, Bar plots depicting the percentage of GREs with conserved activity in marmoset and mouse. For both SNP containing GREs and those that do not. Dissimilarity between the ratio of human-specific activity was calculated with a Fisher's exact test. **c**, Bar plot depicting the percentage of GRE that contain a known SNP for both evolutionary conserved as well as human-specific GREs. Dissimilarity between the ratios was calculated with a Fisher's exact test. **d**, ChIP-Seq tracks of rpm normalized H3K27ac enrichment in different tissues of human, marmoset and mouse as indicated, across a 222 kb region containing the *BSN* gene. Known SNPs are depicted as blue bars underneath the human panel. **e**, Bar plot depicting the percentage of variable GREs, for GREs that are both SNP containing and those that are not as indicated on the axis. Dissimilarity between the ratios was calculated with a Fisher's exact test. **f**, Bar plot depicting the percentage of GREs that are unengaged or engaged with a promoter or other anchor. Data is shown for elements containing SNPs or those that do not. Dissimilarity in TSS engagement was calculated with a Fisher's exact test. **g**, Box plots depicting the maximum STARR-activity of both SNP containing enhancers and those that do not as indicated on the axis. Dissimilarity between the distributions was calculated using a Student's t-test.

the strong interindividual variation that characterizes recently evolved enhancers may not reflect the subtle effects of common trait variation⁴³. Furthermore, common variants in GREs are expected to alter GREs that are of impact on their target gene, which would disfavor their appearance in enhancers that recently evolved. In agreement with trait variant containing enhancers being more often conserved and functional, we observed that these were less often unengaged enhancers and more frequently contacted a gene promoter ($p = 1.4 \times 10^{-15}$, Fig. 5f). Moreover, trait variant containing enhancers were intrinsically more active in reporter assays than GREs without trait variants ($p < 2.2 \times 10^{-16}$, Fig. 5g). Overall these results demonstrate that, while enhancers that recently evolved are often species-specific, variable and unengaged, trait variants typically occur in regulatory DNA that is conserved and likely functional.

Discussion

While rapid turnover of regulatory DNA has put non-coding elements in the spotlight of evolutionary research and has raised questions about the validity of other species as model organisms, the impact of these novel elements on phenotypical evolution may be more modest than initially anticipated. This is reflected in a general paucity in identifying evolutionary changes in non-coding DNA with large phenotypical consequences. Several mechanisms may contribute to this. For instance, regulatory buffering may occur when multiple non-coding elements are modulating expression control over the same gene^{11,36,37,44}, allowing for non-detrimental changes to occur at elements without major gene expression disturbances. In addition, the loss of an enhancer may be offset against the gain of a different enhancer, which is known as compensation²⁸. Furthermore, our current work shows that a subset of enhancers emerges without an apparent integration in the transcriptional network providing additional mechanisms that can in part explain the paucity of finding enhancers that affected phenotypical evolution. In our data, this is also reflected by an altered genomic distribution of unengaged enhancers, which are found to be located further away from genes and more often in inactive B-compartments. These unengaged enhancers may be ‘evolutionarily poised’ and serve as a breeding pool from which future evolutionary novelty may be selected but may be dispensable at the current evolutionary stage. Nevertheless, these effects are only seen for small subsets of enhancers and do not fully explain the inability to link phenotypical change to regulatory innovations.

Instead, our data show a surprisingly high variability of H3K27ac enrichment on enhancers that are evolutionarily new within the human population. This high interindividual variability affects over 50% of enhancers that changed H3K27ac enrichment during recent human evolution. This could suggest that over half of the enhancers that are affected by evolution have not yet stabilized within the population and are thus unlikely to support key phenotypical species changes. Combining variable enhancers with compensated enhancers, unengaged enhancers and buffered enhancers suggest that many enhancers that are evolutionary new may lack strong functional impact. This complicates our understanding of species-specific gene expression control in relationship to phenotypical change. Supporting this notion, we find that common trait variants more often reside in conserved enhancers, which makes sense if they are to affect function. Thus, our data suggest that there is an inverse relationship between the observed evolutionary flexibility at enhancers and their biological impact. In addition, we provide new insight into how enhancers evolve, and where their lack of impact originates from. Finally, as conserved enhancers are more relevant to understanding the control of a human biological process in a marmoset model, our data demonstrates that the species specificity of enhancers as affecting the suitability of other species as disease model organisms is likely overestimated.

Methods

Common marmoset tissue collection

Common marmoset (*Callithrix jacchus*, cj1, cj2, cj3) tissue samples were collected at the Biomedical Primate Research Centre (BPRC) in Rijswijk, the Netherlands (<http://www.bprc.nl>) and represent rest material involving no animal experimentation for the purpose of this work as determined by the Animal Experimental Committee (DEC) (Supplementary data 1). Samples were frozen down as fast as possible after death and stored at -80 °C. Ten different tissue samples, two brain regions and eight organs, were collected from cj1 and cj2. These samples include cerebellum, prefrontal cortex, colon, heart, kidney, liver, pancreas, skeletal muscle, spleen and stomach. Two brain region, cerebellum and prefrontal cortex were collected from cj3.

Chromatin immunoprecipitation and sequencing

Chromatin immunoprecipitation (ChIP) was performed as previously described¹². In short, 60 mg of tissue was used per sample and homogenized in 1 mL Dulbecco's Modified Eagle Medium (DMEM) with 0.2% Bovine Serum Albumine (BSA) in a glass douncer (Kontes Glass Co.). Cells were crosslinked at Room Temperature (RT) in 10 ml fixation buffer (freshly made; 1% formaldehyde, 0.5 mM Ethylenediaminetetraacetic acid (EDTA), 0.05 mM ethylene glycol-bis(β-aminoethyl ether)-N,N,N',N'-tetraacetic acid (EGTA), 10 mM NaCl, 5 mM HEPES-KOH, pH 7.5) while rotating for 10'. Next, samples were washed twice with PBS, pelleted for 5' at 2095 x g and 4 °C, resuspended in 10 ml lysis buffer (50 mM HEPES-KOH pH 7.5, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% Igepal, 0.25% Triton X-100) and lysed for 10' at RT while rotating. Samples were pelleted for 5' at 2095 x g at 4 °C and resuspended in 10 ml wash buffer (200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 10 mM Tris-HCl pH 8.0) and incubated for 10' at RT while rotating. Cells were pelleted for 5' at 2095 x g at 4 °C and resuspended in 150 µl sonication buffer (1 mM EDTA, 0.5 mM EGTA, 10 mM Tris-HCl pH 8.0, 100 mM NaCl, 0.1% Na-Deoxycholate, 0.5% N-lauroyl sarcosine) and sonicated using the Covaris S series (12 cycles of 60 seconds: intensity 3, duty cycle 20%, 200 cycles/burst) in two microtubes (Covaris 520045) per sample. After sonication, the microtubes per sample were pooled and sonication buffer and Triton X-100 (final concentration 1%) was added to a total volume of 550 µl. Immunoprecipitation with antibody (H3K27ac: ab4729 abcam, H3K4me3: ab8580 abcam) coated Protein G Dynabeads (Invitrogen 10003D) was performed overnight at 4 °C. The following day beads were washed 4 times with RIPA (50 mM HEPES-KOH pH 7.5, 1 mM EDTA, 0.7% DOC, 1% NP40 and 0.5 M LiCl) and once with 50 mM NaCl in TE. DNA elution from the beads was done in elution buffer (50 mM Tris pH 8.0, 10 mM EDTA, 1% Sodium dodecyl sulfate (SDS)) overnight at 65 °C. Samples were centrifuged shortly to remove the beads and the supernatant was 1:1 diluted with TE, followed by a 2 h incubation with RNase (final concentration 0.2 µg/µl) at 37 °C and a 2 h incubation with proteinase K (final concentration 0.2 µg/µl) at 55

°C. Finally, the DNA was extracted using phenol/chloroform and MaXtract High Density gel tubes (Qiagen) followed by ethanol purification. Sequencing libraries were prepared according to the Illumina Truseq DNA library protocol and samples were sequenced at the MIT BioMicro Center (<http://openwetware.org/wiki/BioMicroCenter>) or at the Utrecht DNA Sequencing facility (<http://www.useq.nl>) using the Illumina HiSeq 2000 or NextSeq 500 genome sequencer.

ChIP-Seq enrichment analysis

To ensure comparable mapping between the samples, all reads were trimmed to a length of 36 bp using the Fastx-toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). Reads were mapped to the most recent genome available (human; hg38, chimpanzee; panTro5, rhesus macaque; rheMac8, marmoset; calJac3, mouse; mm10) using Bowtie version 1.1.2⁴⁵, allowing 1 mismatch and discarding reads that could not be uniquely mapped. Samtools version 1.3.1⁴⁶ was used to discard duplicate reads. Enriched regions were called with MACS2 version 2.1.1⁴⁷ using a p -value $< 10^{-5}$, extsize 300 and local lambda 100,000. Regions were extended to a minimum of 2000 bp in length, (+/- 1000 bp from peak center) consistent with typical ChIP-Seq resolution^{6,16}. Lists of GREs per tissue per species were obtained by merging identified regions of replicate samples, regions with a minimal 1 bp overlap were stitched together. This cutoff was chosen to ensure that enriched regions with multiple summits were analyzed as one. All enriched regions were classified as enhancers or promoters based on the hg38 RefSeq list. Regions overlapping a 1000 bp window around a known human transcriptional start site (TSS) were considered promoters, others were annotated as enhancers. Distance to the closest TSS was determined for each enhancer using the RefSeq gene list of hg38 with Bedtools version 2.26.0⁴⁸

Comparison with public marmoset ChIP-Seq data

Marmoset H3K27ac samples were compared to publicly available liver H3K27ac ChIP-Seq data⁶. Liver data from human, rhesus macaque, vervet monkey and marmoset was obtained and analyzed as described above. Integration of public marmoset data was done by merging enriched regions from public liver samples with our non-redundant H3K27ac marmoset enriched regions. Integration of the new liver data with the publicly available data sets was done by the generation of a non-redundant, cross-species list of H3K27ac-enriched liver regions. Therefore, H3K27ac regions from liver samples of different primate species were reciprocally mapped onto the human genomes using the UCSC liftOver tool (-minMatch=0.1) as described previously¹². Reciprocal mapping had to yield unique regions and all regions that changed more than 50% in size were excluded. To ensure equal mappability, > 90% of the bases within a regulatory element had to be properly annotated in all reference genomes (< 10% overlap with UCSC Table Browser's gap locations lists). This cut-off was chosen as unknown bases generally occur in stretches rather than single nucleotides scattered across the genome. Moreover,

to account for repetitive or duplicated genomic regions, which are susceptible to poor annotation in lower-quality genomes, enrichment scores were not allowed to change significantly in the target genome when allowing reads to map to multiple locations. These repetitive regions were defined per species by mapping reads from every sample to unique locations (bowtie: -best -strata -m 1) as well as to multiple locations (bowtie: -best -strata -M 1). Genomic regions that were enriched using the multimapping settings but not with unique mapping are potential repetitive elements that are not annotated at similar depth across all the genomes.

Identification of repurposed elements

A non-redundant list of regions mappable on the human and marmoset genome was created using reciprocal liftover as described above. For every region its activity across the different tissues was determined based on peak calling. Repurposed regions were defined as being active in different tissues between human and marmoset based on significant enrichment calls using MACS2. These were then further refined by only considering regions where H3K27ac enrichment differed at least 2-fold. Similar repurposing analysis was used to compare human and marmoset regions with mouse data. The use of presence/absence calls omits the need to control for batch effects which is not feasible when the batches align with the different species. No strong contribution of confounders was detected (Supplementary Fig. 2b).

Evolutionary classification of GREs in LCLs

Non-redundant (NR) lists per species were created using LCL data from a single source³¹ and overlapping regions were merged as described above. Only regions called at least twice in a single species were kept for analysis. To compare enriched regions between species, the NR-list of rhesus macaque and chimpanzee were reciprocally mapped on the human genome as described above. Next, enriched regions of rhesus macaque and chimpanzee were combined with human enriched regions, merging all overlapping regions and discarding all regions that overlapped ENCODE blacklisted regions²², resulting in a total list of 54,793 H3K27ac enriched regions that could be assessed across all three species. The number of reads per region was counted using Bedtools version 2.26.0⁴⁸ and normalized for library size and length of the region (RPKM). To determine differentially enriched (DE) regions between species, seven samples per species were used in a DESeq2⁴⁹ analysis. Regions with a $\log_2\text{FC} > 1$ and $\text{FDR} < 0.01$ were considered as DE between species. Human gained elements were defined as regions that were DE between both human and chimpanzee, as well as human and rhesus macaque. Losses were defined as significantly lower in human versus chimpanzee and human versus rhesus macaque. One explanation for this is that to be able to assign differential enrichment in human, target elements require a time period of evolutionary stability (between chimpanzee and rhesus macaque). Stable elements were determined as not differentially enriched between all the three species.

Classification of evolutionary new elements

To define regions as new in human, GREs were compared to the genome-wide background enrichment of H3K27ac in each sample of rhesus macaque and chimpanzee. To calculate the background enrichment, reads were counted using a sliding window of 3000 bp with 500 bp steps across the genome as shown previously²⁸. Read count values for GREs were normalized to a 3000 bp size and compared with the background windows. Human gain elements were defined as evolutionary new when the enrichment was below the 90th percentile of the background windows in all seven samples of chimpanzee and rhesus.

Analysis of variability and human ancestry

As we defined “human-specific” changes in GREs using samples from Yoruba ancestry only, we verified that variation in ancestry was not causing misclassification of GREs as human-specific⁵⁰. Therefore, 4 samples from individuals from San, Caucasian and Yoruba descent were selected based on high FRIP scores and exclusion of direct family members. A NR-list was created based on enriched regions identified using MACS2 in at least 2 technical replicates per individual and 3 out of 4 individuals of a single ancestry (n = 39,495). As a control, an additional set was generated in a similar manner but composed of 3 independent groups of different individuals of Yoruba descent (n = 31,360). For both sets the elements were selected that were previously classified as human gains. To investigate the effect of ancestry on the classification of human gains; the number of human gains present in one ancestry lineage (Yoruba, San, or Caucasian) was plotted on the x-axis = 1. The number of human gain enhancers present in two groups was plotted on x-axis = 2 and number of enhancers present in all groups was on x-axis = 3 (Supplementary Fig. 3e). This analysis was done for all possible combinations and a regression line was plotted using the ggplot R package. Comparison of the slopes of the two linear regression lines was performed with ANOVA. Venn diagrams of the overlap between analyzed groups were created using the venn package in R. The same analysis was performed for all enhancers without intersection with human gains. Close to 25% of human-specific gains that were based on the Yoruba samples were not found enriched in individuals from both San and Caucasian descent (Supplementary Fig. 3e). However, when comparing the human gains to two control sets containing different Yoruba individuals, we observed a comparable reduction in the number of GREs classified as human-specific (Supplementary Fig. 3e). Indicating that these regions are sample rather than ancestry-specific. Similar results were obtained when analyzing all human GREs, including those that were not gained in the human lineage (Supplementary Fig. 3f) suggesting that this effect was not related to recent evolution. Ancestry specific gains and losses were defined using DESeq2 ($\log_2FC > 1$ and $FDR < 0.01$) between a single ancestry and both other ancestries using the NR-list generated above (n = 39,495 of which 28,113 putative enhancers). DE analysis was performed on read counts of one technical replicate per

individual, with four individuals per ancestry. Next, ancestry-specific regions were overlaid with human-specific gains and evolutionary novel elements using bedtools. Differences in overlap frequencies were calculated using a Fisher's exact test. While some examples of evolutionary miss-classifications were found (Supplementary Fig. 3g), i.e., 6 San-specific gains, their numbers were negligible and could be explained by random overlap ($p = 0.474$, Supplementary Fig. 3h). Thus, human ancestry plays no measurable role in the classification of human-specific regulatory DNA.

Gene expression analysis

Gene expression in LCL cells was determined using the exon read count tables derived from RNA-Seq data generated from 3 humans, chimpanzees and rhesus macaque samples⁵¹. Expression was calculated as the sum of reads per exon divided by gene length and the total number of reads followed by log₂ zero-mean transformation. Only genes with orthologous exons present in all three species were used for analysis. Significant differences in expression between human, chimpanzee and rhesus macaque gene sets were determined using a Student's t-test. Gene expression in common marmoset tissues was assessed using a previously published list of normalized FPKM values for which the female tissue samples were plotted.

Determination of interindividual variability

Variation between human individuals was defined using H3K27ac ChIP-Seq data generated from human LCL samples covering individuals ($n = 19$) of Caucasian, San, Yoruba and Asian descent³⁰, which were analyzed as described above. Aligned reads that were not paired were discarded. A read count table across all individuals and input data was generated for the above-mentioned NR-list, created on human, chimpanzee and rhesus macaque ($n = 54,793$). Variability of enhancers was calculated as described previously³⁰, using only those regions that could also be identified in at least 2 individuals by MACS2. In short, read counts of the ChIP-Seq samples and input samples, were normalized for library size and region length (rpkm) followed by asinh-transformation and quantile normalization using the R package preprocessCore. Fold change over input was calculated as signal/input followed by calculating the F-value (inter-replicate variation to inter-individual variation) using the ANOVA R package. Obtained F-values were log₁₀ transformed and elements with a log₁₀ F-value larger than 1 were considered variable. To refine the variable set, a pairwise differentially enrichment analysis of the 19 individuals was performed using DESeq2. A region with log₂FC > 3 and q-value < $1e^{-5}$ was defined as DE. Only GREs called DE between any pairwise comparison and the log₁₀ F-value > 1 was annotated as variable.

Analysis of line to line variability

Line to line variability was determined using H3K27ac data generated in 5 independent cell lines that were obtained from a single individual³². H3K27ac data

was obtained and mapped to the human genome (hg38) as described above. H3K27ac enrichment for the here identified GREs was counted and rpkm normalized. Only GREs that were significantly enriched for H3K27ac using MACS2 in at least 2 samples were selected and further processed using an asinh-transformation and quantile normalization. Variability was then calculated per GRE on the normalized readcount using ANOVA as described above.

t-SNE analysis and hierarchical clustering

Read counts were normalized for library size and length of the region. Pearson correlations between the samples were calculated and t-Distributed Stochastic Neighbor Embedding (t-sne) multiple scaling coordinates were defined with the t-SNE R package⁵² on the distance matrix. For hierarchical clustering, Pearson correlations between samples were calculated and samples were clustered based on Pearson distance with average linkage. Heatmaps were generated using the heatmap.2 function from the gplots R package.

Cell type specificity of human gains

To assess cell type specificity for human-specific enhancers, seven H3K27ac ChIP-Seq data sets of unrelated cell types were analyzed; including H1, HepG2, HMEC, HSMC, Huvec, NHEK, and NHLF³⁴. Enriched regions were identified using MACS2 as described above. NR-lists of all cell types were compared to the LCL enriched regions defined using the primate samples. Regions not found in any of the other cell types were classified as cell-type-specific for LCL. Difference in the ratio of cell-type-specific regions and regions active in multiple cell types within the human gains was performed using a Fisher's exact test.

Enhancer gene associations using HiC data

A high-resolution map of 3D contacts was obtained from a GM12878 HiC dataset³⁹, using the here identified H3K27ac enriched GREs, known TSS and ENCODE CTCF-sites as viewpoints⁵³. Loops were defined from every viewpoint's virtual 4C profile using PeakC³⁸, with a maximum distance of 1 Mb between viewpoint and anchor. Viewpoints and anchors were then resized to ± 5 kb from the center and only those loops retained that showed reciprocal interactions. Viewpoints were considered to interact with a TSS when their resized anchors overlapped a known promoter region. To analyze enhancer redundancy, the number of enhancers looping to each gene were counted from the HiC data. Enhancers that looped toward a TSS that was also contacting other enhancers, were considered redundant. Enhancers were considered as unengaged when they had no reciprocal contacts. Differences in contact frequencies between sets of enhancers with distinct evolutionary properties were calculated using a Fisher's exact test.

Enhancer compartmentalization and activity

To determine whether GREs emerged in distinct chromatin domains, all enhancers were overlapped with A and B compartments of the GM12878 cell line, as defined previously³⁹. Enhancers that were located on a boundary were assigned to the domain with the largest overlap. Differences in the distribution between A and B compartments were calculated using a Fisher's exact test. To determine whether evolutionary novel enhancers harbored intrinsic activity, we calculated the activity of each element by determining the maximum reporter-activity within each region as defined using a mass parallel reporter assay⁴⁰. Differences in enhancer activity between enhancer groups was calculated using a Student's t-test.

Gene ontology analysis and SNP enrichment

Gene ontology analysis was done using the Genomic Regions Enrichment of Annotations Tool (GREAT version 3.0.0, <http://bejerano.stanford.edu/great/public/html>) with basal plus extension setting²⁴. Multiple genes are can therefore by assigned to the supplied enriched regions. For the marmoset tissue-specific H3K27ac enriched regions sets, regions that were specifically enriched in a single tissue type were analyzed (based on MACS2 peak calling). SNP regions were obtained using the annotated SNPs from the NIH GRASP database (<https://graps.nhlbi.nih.gov/Overview.aspx>)⁴². SNPs with a reported p-value < $1e^{-6}$ were extracted and these were extended to a +2.5 kb, merging any resulting overlapping regions. These were then overlapped with the here identified regulatory elements to obtain both SNP containing GREs and those that do not.

Analysis of confounder variable importance

To test the contribution of confounding variables across different tissue and species samples in the ChIP-Seq datasets, we analyzed the importance of various potential confounder variables including: species, tissue, gender, sequencing depth, lab of origin, GC-content, conservation score and fraction of reads in peaks on the observed read counts using the Boruta package in R⁵⁴ (Supplementary Fig. 2b). Sequence conservation was determined by calculating the average PhastCon score (20 mammals, UCSC) per GRE. The Boruta algorithm defines the importance of confounder variables on the observed read count by using an iterative random forest classification. Every iteration, the observed variables are randomly shuffled to create shadow variables and the importance of the original variables on the observed read count is compared to these shadows. Variables that perform significantly better are confirmed confounders (green boxplot Supplementary Fig. 2b), while variables that perform significantly worse are considered as having no confounding effect on the observed read counts. (red boxplots Supplementary Fig. 2b).

Analysis of interdependence

The interdependence between categorical variables (evolutionary timing, variability and cell-type specificity) of the identified evolutionary novel enhancers, was calculated using LLM3D as described previously³⁵. LLM3D fits a number of log-linear models to 3D contingency tables of enhancer counts and selects the model that best fits the observed enhancer characteristics. These models imply different (in) dependence relationships between the variables, with the null hypothesis assuming complete independence between the variables. Model selection assigned a model in which no (conditional) independence between any pair of the three variables was implied, meaning all levels of every attribute have different probabilities of occurring jointly. The significant p-value indicates that enhancers that are new and variable differ in their rate of cell type specificity.

Data and software availability

The ChIP-Seq data generated for this study is available at Gene Expression Omnibus with accession code GSE141563. Public sequencing datasets analyzed during this study were previously deposited in the NCBI Gene Expression Omnibus and available under accessions GSE5074 (Marmoset gene expression)²³, GSE16256 and E-MTAB-2633 (Human tissues H3K27ac ChIP-Seq)^{6,25}, GSE29184 (Mouse tissues H3K27ac ChIP-Seq)²⁶, GSE47991 (H3K27ac in human LCL from Yoruba descent)³⁰, GSE60269 (H3K27ac in chimpanzee and rhesus macaque LCL)³¹, GSE50893 (H3K27ac in human LCL from Caucasian, San, Yoruba and Asian descent)³⁰, GSE121926 (H3K27ac in five human LCL from the same individual)³², GSE24111 (RNA-Seq in human, chimpanzee and rhesus macaque LCL)⁵¹, GSE26386 (H3K27ac for 7 non-LCL cell types human)³⁴, GSE63525 (HiC data in human LCL GM12878)³⁹ and GSE29611 (CTCF in human LCL GM12878)⁵⁵.

Lead contact and materials availability

Further information and requests for resources and reagents should be directed to Lead Contact, Menno P. Creyghton (m.creyghton@erasmusmc.nl). This study did not generate any unique reagents. The datasets generated during this study are available at GEO (GSE141563).

Acknowledgements

M.P.C., I.S.T., B.C. and M.L.B. were supported by the Royal Academy of Sciences in The Netherlands (KNAW), The Erasmus University Medical Center and by funding acquired by the Friends of the Hubrecht Institute.

Author Contribution

B.C, M.L.B and M.P.C conceived and designed the study. B.C performed the experiments. M.L.B, B.C, M.W.V and G.G, W.d.L and M.P.C. designed the analysis. M.L.B, B.C and G.G performed the analysis and were supervised by W.d.L and M.P.C. I.S.T, I.K. and C.R.M.W collected data. M.W.V. contributed data analysis tools. M.L.B, B.C, M.W.V, C.R.M.W, I.S.T and M.P.C wrote the manuscript.

Declaration of interest

The authors declare no competing financial interests

References

1. Schoenfelder, S. & Fraser, P. Long-range enhancer-promoter contacts in gene expression control. *Nature reviews. Genetics* (2019). doi:10.1038/s41576-019-0128-0
2. Rickels, R. & Shilatifard, A. Enhancer Logic and Mechanics in Development and Disease. *Trends in cell biology* 28, 608–630 (2018).
3. Maurano, M. et al. Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. *Science* 337, 1190–1195 (2012).
4. Soldner, F. et al. Parkinson-associated risk variant in distal enhancer of α -synuclein modulates target gene expression. *Nature* 533, 95–99 (2016).
5. Villar, D., Flicek, P. & Odom, D. T. Evolution of transcription factor binding in metazoans - mechanisms and functional implications. *Nature reviews. Genetics* 15, 221–33 (2014).
6. Villar, D. et al. Enhancer evolution across 20 mammalian species. *Cell* 160, 554–566 (2015).
7. Carrion, R. & Patterson, J. L. An animal model that reflects human disease: the common marmoset (*Callithrix jacchus*). *Current opinion in virology* 2, 357–62 (2012).
8. Cyranoski, D. Marmosets are stars of Japan's ambitious brain project. *Nature* 514, 151–152 (2014).
9. Shimogori, T. et al. Digital gene atlas of neonate common marmoset brain. *Neuroscience Research* 128, 1–13 (2018).
10. Worley, K. C. et al. The common marmoset genome provides insight into primate biology and evolution. *Nature Genetics* 46, 850–857 (2014).
11. Berthelot, C., Villar, D., Horvath, J. E., Odom, D. T. & Flicek, P. Complexity and conservation of regulatory landscapes underlie evolutionary resilience of mammalian gene expression. *Nature Ecology and Evolution* 2, 152–163 (2018).
12. Castelijns, B. et al. Hominin-specific regulatory elements selectively emerged in oligodendrocytes and are disrupted in autism patients. *Nature communications* 11, 301 (2020).
13. Arnold, C. D. et al. Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science* (New York, N.Y.) 339, 1074–7 (2013).
14. Bonn, S. et al. Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. *Nature Genetics* 44, 148–156 (2012).
15. Nord, A. S. et al. Rapid and pervasive changes in genome-wide enhancer usage during mammalian development. *Cell* 155, 1521–1531 (2013).
16. Vermunt, M. W. et al. Large-scale identification of coregulated enhancer networks in the adult human brain. *Cell Reports* 9, 767–779 (2014).
17. Hilton, I. B. et al. Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nature biotechnology* 33, 510–7 (2015).
18. Calo, E. & Wysocka, J. Modification of enhancer chromatin: what, how, and why? *Molecular cell* 49, 825–37 (2013).
19. Schröder, S. et al. Acetylation of RNA polymerase II regulates growth-factor-induced gene transcription in mammalian cells. *Molecular cell* 52, 314–24 (2013).
20. Boija, A. et al. CBP Regulates Recruitment and Release of Promoter-Proximal RNA Polymerase II. *Molecular cell* 68, 491–503.e5 (2017).
21. Guenther, M. G., Levine, S. S., Boyer, L. A., Jaenisch, R. & Young, R. A. A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* 130, 77–88 (2007).
22. Landt, S. & Marinov, G. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome ...* 1813–1831 (2012). doi:10.1101/gr.136184.111.
23. Cortez, D. et al. Origins and functional evolution of Y chromosomes across mammals. *Nature* 508, 488–93 (2014).
24. McLean, C. Y. et al. GREAT improves functional interpretation of cis-regulatory regions. *Nature biotechnology* 28, 495–501 (2010).
25. Kundaje, A. et al. Integrative analysis of 111 reference human epigenomes. *Nature* 518, 317–330 (2015).
26. Shen, Y. et al. A map of the cis-regulatory sequences in the mouse genome. *Nature* 488, 116–20 (2012).
27. Vierstra, J. et al. Mouse regulatory DNA landscapes reveal global principles of cis-regulatory evolution. *Science* 346, 1007–1012 (2014).
28. Vermunt, M. W. et al. Epigenomic annotation of gene regulatory alterations during evolution of the primate brain. *Nature Neuroscience* 19, 494–503 (2016).
29. Jinn, S. et al. TMEM175 deficiency impairs lysosomal and mitochondrial function and increases

- α -synuclein aggregation. *Proceedings of the National Academy of Sciences* 114, 2389–2394 (2017).
30. Kasowski, M. et al. Extensive variation in chromatin states across humans. *Science* 342, 750–752 (2013).
31. Zhou, X. et al. Epigenetic modifications are associated with inter-species gene expression variation in primates. *Genome Biology* 15, 1–19 (2014).
32. Ozgryn, L., Horvath, A., Hevessy, Z. & Balint, B. L. Extensive epigenetic and transcriptomic variability between genetically identical human B-lymphoblastoid cells with implications in pharmacogenomics research. *Scientific Reports* 9, 1–16 (2019).
33. Fish, A., Chen, L. & Capra, J. A. Gene regulatory enhancers with evolutionarily conserved activity are more pleiotropic than those with species-specific activity. *Genome Biology and Evolution* 9, 2615–2625 (2017).
34. Ernst, J. et al. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* 473, 43–9 (2011).
35. Geerven, G. et al. LLM3D: A log-linear modeling-based method to predict functional gene regulatory interactions from genome-wide expression data. *Nucleic Acids Research* 39, 5313–5327 (2011).
36. Hong, J.-W., Hendrix, D. A. & Levine, M. S. Shadow enhancers as a source of evolutionary novelty. *Science (New York, N.Y.)* 321, 1314 (2008).
37. Perry, M. W., Boettiger, A. N., Bothma, J. P. & Levine, M. Shadow enhancers foster robustness of *Drosophila* gastrulation. *Current biology : CB* 20, 1562–7 (2010).
38. Geerven, G., Teunissen, H., de Laat, W. & de Wit, E. peakC: a flexible, non-parametric peak calling package for 4C and Capture-C data. *Nucleic acids research* 46, e91 (2018).
39. Rao, S. S. P. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159, 1665–80 (2014).
40. Wang, X. et al. High-resolution genome-wide functional dissection of transcriptional regulatory regions and nucleotides in human. *Nature Communications* 9, (2018).
41. Rowley, M. J. & Corces, V. G. Organizational principles of 3D genome architecture. *Nature reviews. Genetics* 19, 789–800 (2018).
42. Leslie, R., Donnell, C. J. O. & Johnson, A. D. GRASP : analysis of genotype – phenotype results from 1390 genome-wide association studies and corresponding open access database. 30, 185–194 (2014).
43. Tam, V. et al. Benefits and limitations of genome-wide association studies. *Nature reviews. Genetics* 20, 467–484 (2019).
44. Osterwalder, M. et al. Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature* 554, 239–243 (2018).
45. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9, 357–359 (2012).
46. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079 (2009).
47. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 9, (2008).
48. Quinlan, A. R. & Hall, I. M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842 (2010).
49. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* 15, 550 (2014).
50. Sirugo, G., Williams, S. M. & Tishkoff, S. A. The Missing Diversity in Human Genetic Studies. *Cell* 177, 26–31 (2019).
51. Cain, C. E., Blekman, R., Marioni, J. C. & Gilad, Y. Gene expression differences among primates are associated with changes in a histone epigenetic modification. *Genetics* 187, 1225–34 (2011).
52. Maaten, L. van der & Hinton, G. Visualizing Data using t-SNE. *Journal of Machine Learning Research* 1 620, 267–84 (2008).
53. Sanyal, A., Lajoie, B. R., Jain, G. & Dekker, J. The long-range interaction landscape of gene promoters. *Nature* 489, 109–113 (2012).
54. Kursu, M. B. & Rudnicki, W. R. Feature selection with the boruta package. *Journal of Statistical Software* 36, 1–13 (2010).
55. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74 (2012).

Supplementary Fig. 1 Classification of marmoset regulatory elements. **a**, Hierarchical clustering of all marmoset tissue samples including 3 liver samples obtained from⁶, based on H3K27ac enriched regions (n = 60,824). Correlation map colors indicate Pearson correlation coefficients between the samples. Side bars are color coded to indicate different tissues. Pearson distances are represented by the tree on the left. **b**, Hierarchical clustering as in (a), with H3K27ac enriched regions of different primate liver samples obtained from⁶ in combination with the marmoset liver samples generated here (black). **c**, Venn diagram depicting the overlap between H3K27ac and H3K4me3 enriched regions. **d**, Pie chart depicting the number of H3K4me3 enriched regions that overlap either a marmoset or human TSS. **e**, Pie chart depicting the genomic distribution of H3K4me3 enriched regions that do not overlap an annotated TSS. **f**, Box plots depicting the normalized expression values in different marmoset tissues for H3K4me3 regions not overlapping an annotated TSS. **g**, Bar plot depicting the number of tissues in which the same H3K4me3 or H3K27ac enriched regions are active. **h**, Go-terms related to tissue-specific H3K27ac enriched GREs. Colors indicate different tissues as in (a)

Supplementary Fig. 2. Enhancer repurposing across evolution. **a**, ChIP-Seq tracks of rpm normalized H3K27ac enrichment in human and marmoset tissues across a 206 kb regions containing the *KLF5* gene. **b**, Box plots depicting the importance of confounding variables on the observed H3K27ac enrichment for human and marmoset. GREs importance was determined using the Boruta package in R (FRIP: Fraction of Reads in Peaks). A green color indicates a significance contribution of the variable to the observed H3K27ac enrichment. **c**, Bar plot depicting the percentage of tissue-specific GREs for both conserved and repurposed elements. Dissimilarities between the ratios was calculated with a Fisher's exact test. **d**, Pie charts depicting the number of repurposed elements that are either primate-specific or shared with mouse (left) and the number of elements that are repurposed in mouse (right). **e**, Pie charts as in (d) but for elements conserved between human and marmoset. **f**, ChIP-Seq tracks of rpm normalized H3K27ac enrichment in different tissues from human, marmoset and mouse, across a 35 kb regions containing the *FLNB* gene. An ambiguous GRE is highlighted in green. **g**, ChIP-Seq tracks of rpm normalized H3K27ac enrichment in different tissues of human, marmoset and mouse, across a 100 kb regions containing the *GLT8D2* gene. A GRE repurposed between all three species highlighted in green, the *GLT8D2* promoter highlighted in blue. **h**, t-Stochastic Neighbor Embedding (t-sne) analysis of LCL samples from 19 human individuals, on the NR-list of human, chimpanzee and rhesus macaque enriched regions (n = 54,793).

Supplementary Fig. 3 Enhancer activity across evolution. **a**, Heatmaps depicting rpkms normalized H3K-27ac enrichment for GREs defined as human losses ($n = 1666$), stable ($n = 7098$) and all GREs ($n = 40,924$). Heatmap color indicates normalized H3K27ac enrichment. **b**, Box plots depicting the $\log_2$, zero-mean normalized gene expression in human, chimpanzee and macaque for genes linked to evolutionary losses

at GREs (left) and linked to stable GREs (right). **c**, Correlation between rpkm normalized H3K27ac enrichment and variability scores ($\log_{10}(\text{F-value})$) used to determine interindividual variability ($n = 20,912$). **d**, Bar plot depicting the percentage of enhancers that have interindividual variable H3K27ac enrichment in different human LCL samples for enhancers that have stable line-to-line H3K27ac enrichment. Different evolutionary categories of enhancers are indicated on the x-axis. Dissimilarities between the ratios was calculated using a Fisher's exact test. **e**, Venn diagrams showing the overlap of human gains (determined using Yoruba samples) and GREs that are active in individuals from different ancestries (upper) or different control sets that are also of Yoruba decent (lower). The graphs show the relationship between individual variation and ancestral variation on human enhancer classification. Putative enhancers based on 4 individuals of Caucasian, Yoruba and San ancestries (CYS), intersected with the human gains ($n = 2930$) and a control NR-list consisting of 3 sets of 4 different individuals of Yoruba ancestry ($n = 3004$). Points in the graph are the number of enhancers, in one group ($x\text{-axis} = 1$), number of enhancers present in two groups ($x\text{-axis} = 2$) and enhancers present in all three groups ($x\text{-axis} = 3$). All combinations are plotted, starting the analysis with a different starting group. The plotted log regression lines for the CYS (Caucasian, Yoruba and San), are compared to the control set using ANOVA, $p\text{-value} = 0.596$. **f**, Graphs as in (e) but for all GREs. **g**, Pie chart and bar plot showing differential enrichment analysis using DESeq2 on NR-list of 4 samples Caucasian, Yoruba and San ($n = 28,113$) between sets of individuals selected by ancestry ($\log_2\text{FC} > 1$ and $\text{FDR} < 0.01$). Distribution of ancestry-specific elements is shown in the right bar. **h**, Bar plots showing the overlap between ancestry-specific elements and evolutionary human gains and new elements, with the absolute numbers indicated in the bars. Differences in ratio were calculated with a Fisher's exact test.

Supplementary Fig. 4 Unengaged GREs are depleted from H3K27ac and emerge in genomic inactive domains. **a**, Bar plot depicting the percentage of all GREs sharing TSS contacts with one or more other GREs shown for enhancers that are variable and non-variable as indicated on the axis. Right plot shows the same analysis for GREs that are evolutionary new in humans. Dissimilarities between the ratios were calculated using a Fisher's exact test. **b**, Bar plot depicting the percentage of GREs significantly enriched for H3K27ac in GM12878 cells that have a reciprocal HiC interaction with a TSS, for the different evolutionary categories as indicated on the axis. Dissimilarities between the ratios were calculated using a Fisher's exact test. **c**, ChIP-Seq track of rpm normalized H3K27ac enrichment in a LCL sample from human, chimpanzee and macaque across a 188 kb region surrounding an evolutionary novel enhancer that

is also classified as unengaged within the *MROH1* gene (highlighted in blue). **d**, Box plot depicting rpkm normalized H3K27ac enrichment in GM12878 cells for either engaged or unengaged enhancers that are evolutionary new and for all enhancers as shown on the axis. Dissimilarities between the distributions were calculated with a Student's t-test. **e**, Box plots depicting the differences in the maximum STARR-activity between engaged and unengaged enhancers for all GREs. Dissimilarity between the distribution was calculated using a Student's t-test. **f**, Bar plots depicting the percentage of GREs in both A- and B-compartments defined in the GM12878 cell line for enhancers that are evolutionary new. Dissimilarity between ratios was calculated with a Fisher's exact test. **g**, Virtual 4C profile using an unengaged enhancer that is evolutionary new in humans, located on chromosome 22, as viewpoint (highlighted in blue). Normalized ChIP-Seq enrichment for H3K27ac (green) and CTCF (red) in GM12878 cells is shown in the top two tracks. On top a zoom-out of the 14 Mb region on chromosome 22 containing the enhancer of interest (highlighted in blue), with the corresponding H3K27ac enrichment ChIP-Seq track from GM12878 cells and A-, B-compartments shown in grey underneath. **h**, Bar plot as in (f) for enhancers that are significantly enriched in GM12878 cells in the different evolutionary enhancer categories as indicated on the x-axis. Dissimilarities between the ratios were calculated with a Fisher's exact test. **i**, Box plots depicting log₂, zero-mean normalized gene expression of genes linked to new unengaged elements based on proximity. Dissimilarities between the distributions were calculated with a Student's t-test. **j**, Box plot as in (i) but for genes linked to new elements by HiC.

Supplementary Data 1. Sequence information

Sample Name	Tissue	Antibody	Rep	Age	Gender	Total reads	Unique reads	Mac2 Peaks	FRIP Score	Geo accession number	Reanalysed form
Marmoset_Cerebellum_H3K27ac_1	Cerebellum	H3K27ac	1	5 Years	Female	15777644	10343094	34070	20.38%		GSM3755466
Marmoset_Cerebellum_H3K27ac_2	Cerebellum	H3K27ac	2	5 years	Female	14070187	11174561	29346	11.83%		GSM3755467
Marmoset_Cerebellum_H3K27ac_3	Cerebellum	H3K27ac	3	8 Months	Male	33948738	26626784	44723	22.82%		GSM3755468
Marmoset_Cerebellum_H3K4me3_1	Cerebellum	H3K4me3	1	5 Years	Female	11709320	8819213	22482	31.82%		GSM3755472
Marmoset_Cerebellum_H3K4me3_2	Cerebellum	H3K4me3	2	5 years	Female	11559319	6549852	16779	8.75%		GSM3755473
Marmoset_Colon_H3K27ac_1	Colon	H3K27ac	1	5 Years	Female	12914180	8580410	40507	32.68%	GSM4207030	
Marmoset_Colon_H3K27ac_2	Colon	H3K27ac	2	5 years	Female	19343954	15300792	35220	15.08%	GSM4207031	
Marmoset_Colon_H3K4me3_1	Colon	H3K4me3	1	5 Years	Female	9970760	6000146	19355	18.53%	GSM4207032	
Marmoset_Colon_H3K4me3_2	Colon	H3K4me3	2	5 years	Female	16953666	13105063	22730	20.63%	GSM4207033	
Marmoset_Heart_H3K27ac_1	Heart	H3K27ac	1	5 Years	Female	37952379	18853051	26689	6.56%	GSM4207034	
Marmoset_Heart_H3K27ac_2	Heart	H3K27ac	2	5 years	Female	25344185	17842688	24590	7.06%	GSM4207035	
Marmoset_Heart_H3K4me3_2	Heart	H3K4me3	2	5 Years	Female	10613539	7463973	18231	10.98%	GSM4207036	
Marmoset_Kidney_H3K27ac_1	Kidney	H3K27ac	1	5 years	Female	16322860	11683420	32794	11.64%	GSM4207037	
Marmoset_Kidney_H3K27ac_2	Kidney	H3K27ac	2	5 Years	Female	41154373	31649247	31674	7.94%	GSM4207038	
Marmoset_Kidney_H3K4me3_1	Kidney	H3K4me3	1	5 years	Female	9850187	7271551	21008	19.19%	GSM4207039	
Marmoset_Kidney_H3K4me3_2	Kidney	H3K4me3	2	5 Years	Female	14070961	9808156	21046	15.78%	GSM4207040	
Marmoset_Liver_H3K27ac_1	Liver	H3K27ac	1	5 years	Female	34382761	23132801	46317	20.37%	GSM4207041	
Marmoset_Liver_H3K27ac_2	Liver	H3K27ac	2	5 Years	Female	37199191	28423706	51554	26.08%	GSM4207042	
Marmoset_Liver_H3K4me3_1	Liver	H3K4me3	1	5 Years	Female	15147762	11262057	20826	19.94%	GSM4207043	
Marmoset_Liver_H3K4me3_2	Liver	H3K4me3	2	5 Years	Female	16793521	12452723	23787	25.74%	GSM4207044	
Marmoset_Pancreas_H3K27ac_1	Pancreas	H3K27ac	1	5 Years	Female	35881832	12720175	38827	17.20%	GSM4207045	
Marmoset_Pancreas_H3K27ac_2	Pancreas	H3K27ac	2	5 Years	Female	40242798	31465778	37291	10.98%	GSM4207046	

Supplementary Data 1 (continued). Sequence information

Marmoset_Pancreas_H3K4me3_1	Pancreas	H3K4me3	1	5 Years	Female	13036949	9923232	16386	7.30%	GSM4207047	GSM3755469 GSM3755470 GSM3755471 GSM3755474 GSM3755475
Marmoset_Pancreas_H3K4me3_2	Pancreas	H3K4me3	2	5 Years	Female	19320302	14836465	22107	16.60%	GSM4207048	
Marmoset_PrefrontalCortex_H3K27ac_1	PFC	H3K27ac	1	5 Years	Female	36178201	20260326	37084	10.01%		
Marmoset_PrefrontalCortex_H3K27ac_2	PFC	H3K27ac	2	5 Years	Female	37292103	29385809	42138	11.44%		
Marmoset_PrefrontalCortex_H3K27ac_3	PFC	H3K27ac	3	8 Months	Male	39702004	29972436	21702	4.66%		
Marmoset_PrefrontalCortex_H3K4me3_1	PFC	H3K4me3	1	5 years	Female	10485374	7781895	22491	27.82%		GSM3755469 GSM3755470 GSM3755471 GSM3755474 GSM3755475
Marmoset_PrefrontalCortex_H3K4me3_2	PFC	H3K4me3	2	5 Years	Female	16317000	12550519	22630	19.35%	GSM4207049	
Marmoset_SkeletalMuscle_H3K27ac_1	Skel. Muscle	H3K27ac	1	5 years	Female	23946498	17875564	37118	12.64%	GSM4207050	
Marmoset_SkeletalMuscle_H3K27ac_2	Skel. Muscle	H3K27ac	2	5 Years	Female	19361271	13870882	34890	15.57%	GSM4207051	
Marmoset_SkeletalMuscle_H3K4me3_1	Skel. Muscle	H3K4me3	1	5 years	Female	9628333	7011388	16850	8.63%	GSM4207052	
Marmoset_SkeletalMuscle_H3K4me3_2	Skel. Muscle	H3K4me3	2	5 Years	Female	20167534	14415440	24424	15.12%	GSM4207053	GSM3755469 GSM3755470 GSM3755471 GSM3755474 GSM3755475
Marmoset_Spleen_H3K27ac_1	Spleen	H3K27ac	1	5 Years	Female	41958268	32368777	35819	9.28%	GSM4207054	
Marmoset_Spleen_H3K27ac_2	Spleen	H3K27ac	2	5 Years	Female	31104114	23948638	30663	9.19%	GSM4207055	
Marmoset_Spleen_H3K4me3_1	Spleen	H3K4me3	1	5 Years	Female	17457291	13114936	16946	7.35%	GSM4207056	
Marmoset_Spleen_H3K4me3_2	Spleen	H3K4me3	2	5 Years	Female	14140872	10704744	18658	11.31%	GSM4207057	
Marmoset_Stomach_H3K27ac_1	Stomach	H3K27ac	1	5 Years	Female	33273885	23033562	30803	9.38%	GSM4207058	GSM3755469 GSM3755470 GSM3755471 GSM3755474 GSM3755475
Marmoset_Stomach_H3K27ac_2	Stomach	H3K27ac	2	5 Years	Female	34870645	27215630	31573	8.93%	GSM4207059	
Marmoset_Stomach_H3K4me3_1	Stomach	H3K4me3	1	5 Years	Female	12248101	8935215	21217	16.39%	GSM4207060	
Marmoset_Stomach_H3K4me3_2	Stomach	H3K4me3	2	5 Years	Female	18115256	13563953	22504	15.26%		
Marmoset_Cerebellum_Input	Cerebellum	Input	1	5 Years	Female	23580385	15664518				
Marmoset_Colon_Input	Colon	Input	1	5 Years	Female	44798515	24613374			GSM4207061	GSM3755469 GSM3755470 GSM3755471 GSM3755474 GSM3755475
Marmoset_Heart_Input	Heart	Input	1	5 Years	Female	80162881	57619308			GSM4207062	
Marmoset_Kidney_Input	Kidney	Input	1	5 Years	Female	14791159	10582851			GSM4207063	
Marmoset_Liver_Input	Liver	Input	1	5 Years	Female	13713437	9250742			GSM4207064	

Supplementary Data 1 (continued). Sequence information

Marmoset_Pancreas_Input	Pancreas	Input	1	5 Years	Female	29612925	19273617	GSM4207065	GSM3755477
Marmoset_PrefrontalCortex_Input	PFC	Input	1	5 Years	Female	13937158	9813827	GSM4207066	
Marmoset_SkeletalMuscle_Input	Skel. Muscle	Input	1	5 Years	Female	39563959	30521543		
Marmoset_Spleen_Input	Spleen	Input	1	5 Years	Female	16867029	8697690	GSM4207067	
Marmoset_Stomach_Input	Stomach	Input	1	5 Years	Female	40873539	30976654	GSM4207068	

PFC: Prefrontal Cortex
WM: White Matter
Skel. Muscle: Skeletal Muscle
Rep: replicate
FRIP: Fration of Reads in Peaks

Chapter 4

Cell type resolved gene expression reveals no distinct oligodendrocyte subpopulation affected by hominin-specific regulatory rewiring

Bas Castelijns¹, Ilia S. Timpanaro¹ and Menno P. Creyghton^{1,2}

¹Hubrecht Institute-KNAW & University Medical Center Utrecht,
Utrecht, The Netherlands

²Erasmus University Medical Center, Department of Developmental Biology,
Rotterdam, The Netherlands

Abstract

Recent evolution of the primate brain is associated with regulatory changes that preferentially occur in oligodendrocytes. However, whether these changes are associated with a particular subtype of oligodendrocytes is currently unclear. Here we annotated differences in gene expression of both neuronal and non-neuronal cells in the prefrontal cortex of several primate species and generated gene expression profiles at a single cell resolution. Combined with previously identified regulatory changes, we confirmed the oligodendrocyte specificity of a recently evolved genetic program. Moreover, we observed that this program was enriched in all oligodendrocyte subtypes, suggesting that the observed gain of activity is not due to an altered cell type abundance. Overall, our data provide an extended view on regulatory rewiring during primate evolution and offers a framework from which cell-specific gene expression can be linked to regulatory changes.

Introduction

The prefrontal cortex (PFC) is one of the last brain regions to mature and involved in many complex cognitive functions such as decision making¹ and social behaviour². However, to what extent the human brain is unique among primates and how the PFC is involved is still not clear. Several studies have shown that the human PFC went through a round of accelerated cortical expansion as the amount of gray and white matter volume in the PFC is larger than expected based on size of the rest of the cortex³⁻⁵. However, others have shown that the expansion of several PFC areas is not specific to humans but rather shared with the other great apes⁶. While recent studies did not find any significant increase in either PFC size or neuronal content relative to the increase in size of the rest of the brain^{7,8}. Nevertheless, all these studies have mainly relied on size modelling and cell counting and do not take the underlying genetic processes into account.

Cellular processes such as development are controlled by gene expression programs, which are in turn regulated by gene regulatory elements (GREs), including enhancers and promoters^{9,10}. Changes in gene expression have been proposed as a driving force of evolution¹¹⁻¹³ and many evolutionary alterations in their regulatory networks have been identified that could account for some of the differences between species¹⁴⁻¹⁷. Therefore, to fully understand which genetic programs underlie the evolution of the PFC and how these explain differences between humans and other primates, both gene expression as well as regulatory control needs to be assessed. Recently we have identified a set of hominin-specific regulatory elements in the PFC, that emerged selectively in oligodendrocytes and could be linked to autism spectrum disorders¹⁸. While we have shown that these changes are not due to a hominin-specific shift in oligodendrocyte abundance, recent data has shown that the human brain contains multiple subpopulations of oligodendrocytes that are

differentially implicated in the development and progression of multiple sclerosis^{19,20}. It is therefore possible that enrichment of oligodendrocyte regulatory elements stems from an altered distribution of oligodendrocyte subtypes.

In the current analysis we generated cell type resolved gene expression profiles of the prefrontal cortex from different primate species. Single cell analysis of the previously identified hominin-specific PFC gains demonstrates that the observed differences during primate evolution are not due to alterations in a specific subpopulation of oligodendrocytes, suggesting that during evolution of the prefrontal cortex, oligodendrocyte-specific regulatory rewiring occurred rather than a change in cell type content.

Results

Nuclei isolation from prefrontal cortex

To obtain cell type resolved expression profiles, we isolated nuclei from the prefrontal cortex of four primate species (human, chimpanzee, macaque and marmoset). As isolating intact cells from frozen brain tissue is technically not possible, we optimized a two-step nuclei isolation protocol to achieve higher purity of nuclei and decrease sorting time²¹. We therefore performed an additional purification step using OptiPrep as density gradient medium, which significantly improved both the quality of the nuclei as well as the required sorting time (Supplementary Fig. 1a). Isolated nuclei were stained with NeuN, a pan-neuronal nuclear membrane marker²², to separate between neuronal (NeuN+) and non-neuronal (NeuN-) nuclei and FACS sorted for all primates (Supplementary Fig. 1b). Visual inspection after sorting confirmed that this approach successfully separated nuclei based on the membrane marker (NeuN) (Supplementary Fig. 1c), and thus enabled us to process these further for RNA-sequencing to obtain their expression profiles.

Cell-type-specific GRE usage is correlated with expression profiles

To obtain gene expression profiles of neuronal and non-neuronal cells across primate evolution, isolated nuclei were bulk sorted in trizol and processed using the CEL-Seq1 protocol²³. Comparison of gene expression per species revealed that most genes are actively transcribed in both neuronal and non-neuronal cells with no overall differences in expression levels (Fig. 1a). Only a minor fraction (average: 10.2%) of genes were expressed in a single cell type (Supplementary Fig. 1d) and analysis of known marker genes showed that their expression is relatively higher in the corresponding cell type. For example, *GRIN1* and *NEUN* are higher expressed in NeuN+ neurons, while the opposite is true for *SOX10* and *S100B* (Supplementary Fig. 1e). To address whether gene expression patterns and GRE usage can be linked per cell type, we obtained chromatin accessibility profiles of both neuronal and non-neuronal cells by leveraging previously published ATAC-sequencing data from similar FACS sorted nuclei¹⁸. The identified accessible regions were merged to create

Fig. 1 Cell type resolved gene expression of primate PFC. **a**, Box plots depicting expression levels of genes identified in NeuN+ and NeuN- sorted nuclei for all primate species as indicated. Dissimilarities between distributions were calculated using a Student's t-test. **b**, Heatmap showing the ATAC enrichment on scaled NeuN+ and NeuN- specific GREs for both NeuN+ and NeuN- sorted nuclei, per primate species as indicated. Heatmap colors indicate ATAC enrichment (rpm). **c**, Heatmap as in (b) for NeuN+ and NeuN- specific GREs shared across the primate lineage. **d**, Box plots depicting fold change in gene expression between NeuN+ and NeuN- samples for genes linked to NeuN+ and NeuN- specific GREs shared across evolution, per primate as indicated. Dissimilarities between distributions were calculated using a Student's t-test. **e**, Box plots depicting normalized gene expression levels in NeuN+ sorted nuclei for genes linked to hominin-specific regulatory changes as defined by¹⁸, per primate as indicated. Dissimilarities between distributions were calculated using a Student's t-test. **f**, Box plots depicting the fold change in gene expression of both NeuN+ and Olig2+ nuclei between human and macaque as defined by²⁹, for genes linked to hominin-specific gains and evolutionary stable GREs. Dissimilarities between distributions were calculated using a Student's t-test. Bottom and top of the box plots are the first and third quartile. The line within the boxes represents the median and whiskers denote interval within 1.5x the interquartile range from the median, outliers are depicted as points.

a single non-redundant list ($n = 49,322$) of GREs mappable on all assessed primate genomes. Comparison of the samples revealed a clear separation of neuronal and non-neuronal samples as well as the different primate species (Supplementary Fig. 2a). This is in line with previous research showing that cell type differences within species are more dominant than differences between species within the same tissue^{14,18,24,25}.

We identified cell-type-specific GREs per primate species using DESeq2²⁶ (Fig. 1b),

assigned a target gene to every GRE based on available HiC data from the PFC²⁷ and assessed gene expression levels of linked genes in both neuronal and non-neuronal nuclei. Comparison of genes linked to either neuronal or non-neuronal-specific GREs revealed that these genes have an elevated expression in that cell type compared to genes linked to shared GREs (Supplementary Fig. 2b). This was observed for all primates and more pronounced for genes with a differentially accessible promoter than genes linked to an enhancer that is differentially accessible (Supplementary Fig. 2b). Moreover, we identified 693 non-neuronal and 135 neuronal-specific GREs that were conserved across the primate lineage (Fig. 1c, Supplementary Fig. 2c). For example, a GRE located within the *KALRN* gene, which is involved in neuronal growth and synaptic density²⁸, was neuron-specific in all primates (Supplementary Fig. 2c).

In accordance with the expression patterns observed in the separate species, genes linked to conserved cell-type-specific GREs displayed significant upregulation of expression in that cell type, in all primates (Fig. 1d). While genes linked to shared elements displayed no difference in expression between neuronal and non-neuronal nuclei (Supplementary Fig. 2d). Furthermore, analysis of the biological function of these evolutionary conserved cell-type-specific GREs, showed that they were linked to processes relevant to their specific cell type (Supplementary Fig. 2e, f). Neuronal-specific GREs were linked to processes such as ion transport and synaptic signalling, while non-neuronal GREs were enriched for axon ensheathment and gliogenesis.

Likewise, assessment of genes linked to the previously identified hominin-specific H3K27ac gains showed that they were upregulated in human and chimpanzee non-neuronal cells specifically, compared to the other primates (Fig. 1e). Moreover, leveraging recently published gene expression data from different primate cell types²⁹, we observed that these expression changes are specific to oligodendrocytes (Fig. 1f). Overall these results show that GRE usage is linked to relevant gene expression profiles and that cell-type-specific alterations in GRE activity is coupled to gene expression changes within the same cells

Evolutionary novel enhancers acquire an open configuration prior to activation

In previous research, we identified hominin-specific regulatory changes in the prefrontal cortex based on differences in H3K27ac enrichment and showed that enhancer activity primarily emerged in non-neuronal cells¹⁸. We wondered whether the cell-type-specific regulatory changes could be linked to cell-type-specific chromatin accessibility difference. We therefore focused on the hominin-specific prefrontal cortex gains, which were shown to be oligodendrocyte-specific. Comparison with ATAC-Seq data confirmed that many hominin-specific PFC gains (49%) are also differentially enriched for open chromatin in human non-neuronal cells compared to neuronal cells (Fig. 2a). Surprisingly, only a subset of hominin-specific H3K27ac GREs showed hominin-specific gain of accessibility within non-neuronal cells. (Supplementary Fig. 3a). For example, a hominin-specific GRE at

the promoter of the *CLCA4* gene, which has been implicated in the development and homeostasis of glial cells³⁰, shows distinct H3K27ac enrichment and chromatin accessibility in humans and chimpanzees (Supplementary Fig. 3b). For most of the GREs however, the regions seemed to be accessible prior to activation. For example, hominin-specific activation was observed for a GRE in the intron of *ATP2C2*, a calcium transporting ATPase implicated in language impairment³¹, while the region was already configured to an open chromatin state in rhesus macaque (Fig. 2b). This suggests that some regions annotated as hominin-specific might already have been in a poised configuration prior to the separation of the hominin lineage. Alternatively, these could represent regulatory repurposing of elements with a different regulatory function in rhesus macaque³².

Single nuclei analysis reveals no difference in oligodendrocyte subpopulation

Recent studies have shown that brain cell types are highly heterogeneous and constitute of multiple subtypes^{19,20}. Moreover, loss of certain subpopulations of cells has been implicated in disease¹⁹. We therefore wondered whether the gain of hominin-specific GREs in oligodendrocytes was due to alterations of a specific oligodendrocyte subpopulation. Nuclei from the prefrontal cortex of all primates were therefore single cell sorted and processed with CEL-Seq2 to obtain their individual transcriptomes³³. To obtain an unbiased sample, nuclei were not stained for NeuN. Sequencing reads were mapped to the corresponding reference transcriptome, using both exon and intron sequences. This was done as the nucleus contains unspliced transcripts and we observed a substantial increase in mappability when allowing reads to map to intron sequences (Supplementary Fig. 3c). Using RaceID these nuclei were clustered based on their transcriptome and separated in distinct groups (Supplementary Fig. 3d)³⁴. To assign cell types to these clusters, we leveraged the cell-type-specific genes obtained using bulk data (see methods). Mapping of both the NeuN-positive and NeuN-negative-specific genes based on both bulk ATAC- and RNA-sequencing data revealed distinct clusters of neuronal and non-neuronal nuclei for every primate (Supplementary Fig. 3e). Except for chimpanzee, where the separation of neuronal and non-neuronal genes was less pronounced, possibly due to the relative low gene count per nuclei (Supplementary Fig. 3c, e).

To investigate whether we could identify a subpopulation of cells affected by the hominin-specific regulatory gains, we combined the nuclei data from human and macaque as these displayed the highest average gene count. Merging of the human and macaque nuclei revealed distinct clusters (Fig. 2c), with both species distributed across all clusters (Supplementary Fig. 3f). Cell type analysis based on cell-type-specific genes defined previously²⁰ revealed clear clusters of both non-neuronal (astrocytes, microglia and oligodendrocytes) and neuronal cells (excitatory neurons and inhibitory neurons) (Fig. 2c, Supplementary Fig. 3g). As hominin-specific GREs were preferentially enriched in oligodendrocytes, we wondered whether we could identify a subpopulation of cells that was specifically altered during evolution.

Fig. 2 Hominin-specific GREs are linked to oligodendrocyte-specific gene expression. **a**, Heatmap showing the ATAC enrichment on scaled hominin-specific regulatory changes as defined by¹⁸ for both human NeuN+ and NeuN- sorted nuclei. Heatmap colors indicate ATAC enrichment (rpm). **b**, ChIP-Seq and ATAC-Seq tracks showing rpm normalized H3K27ac (top panel) and ATAC (middle and lower panel) enrichment across a region containing the *ATP2C2* gene, for all primates as indicated. Hominin-specific H3K27ac GRE is highlighted in green. **c**, t-Distributed Stochastic Neighbour Embedding (t-sne) map of all human and macaque nuclei combined. Different colors represent different clusters as defined by RaceID and corresponding cell types were defined based on cell-type-specific genes annotated by²⁰. **d**, t-sne map as in (c), showing expression levels of genes linked to hominin-specific regulatory changes. Transcript counts are given in log scale. **e**, t-sne map of all human and macaque nuclei identified as oligodendrocyte in (c), colors represent nuclei origin as indicated. **f**, t-sne map as in (e), showing expression levels of genes linked to hominin-specific regulatory changes. Transcript counts are given in log scale. **g**, Box plot depicting the average log normalized expression count for genes linked to hominin-specific gains in different oligodendrocyte subtypes as defined by¹⁹. Dissimilarities between OPC and the other clusters were calculated using a Student's t-test. (OPCs: oligodendrocyte precursor cells, COPs: committed oligodendrocyte precursors). Bottom and top of the box plots are the first and third quartile. The line within the boxes represents the median and whiskers denote interval within 1.5x the interquartile range from the median, outliers are depicted as points.

Analysis of genes linked to the previously identified hominin-specific regulatory gains confirmed that these were preferentially expressed in oligodendrocytes, with higher expression levels in human cells (Fig. 2d, Supplementary Fig. 3h). Although we observed less macaque oligodendrocytes in total, we did not observe a significant difference in the ratio between human and macaque cells within any of the oligodendrocyte subclusters (Fig. 2e, Supplementary Fig. 3i). Moreover, expression of genes linked to the hominin-specific GREs was not confined to a specific set of adult oligodendrocytes (Fig. 2f, Supplementary Fig. 3j). We did observe an enrichment for these genes in mature oligodendrocytes compared to oligodendrocyte precursor cells, which fits the observation that hominin gains are preferentially enriched in

adult tissue (Fig. 2g). Overall these results show that expression of genes linked to hominin-specific GREs is also oligodendrocyte-specific but not restricted to a subpopulation of cells.

Discussion

While glial cells have long been considered passive support cells for neurons, it has become clear that they play an active role in both brain development as well as function^{35,36}. Moreover, recent evidence suggest that they are important in both disease susceptibility as well as evolution^{15,18,19,37}. Here we have further elaborated on the function of glial cells in primate brain evolution, especially oligodendrocytes, by focussing on a set of regulatory elements that were shown by H3K27ac enrichment to have emerged during recent primate evolution¹⁸. Our data shows that while H3K27ac enrichment was specific to hominins, some GREs already displayed an open configuration which was shared with rhesus macaque suggesting this open configuration may have preceded the hominin branch during primate evolution. As expression of genes linked to these GREs was not yet affected in macaque, the open configuration might indicate that these regions were in a primed state^{38,39}. It will therefore be interesting to assess H3K4me1 enrichment on these regions to further verify whether these regions truly acquired a poised configuration prior to activation during primate evolution^{9,40}. Alternatively, these regions may harbor other functions in rhesus macaque such as insulator sites, in which case the hominin enhancers may represent repurposing of function³². Finally, there are many non-functional transcription factors binding events in our genome⁴¹, which may lead to increased accessibility. These regions may therefore represent non-functional precursor elements that can more easily become an enhancer^{42,43}.

The brain is a complex organ, consisting of many different cell types with a diverse set of functions^{19,20,44–47}. As regulatory elements act in a cell-type-specific manner, identifying which cell types or states are affected by regulatory changes is crucial to understand their precise role in human evolution. We previously showed that hominin-specific regulatory elements preferentially emerged in oligodendrocytes and were not due to an increase in oligodendrocyte content within the hominin-lineage¹⁸. However, it was recently shown that oligodendrocytes consist of multiple subtypes and that alterations in the abundance of these subtypes are linked to disease^{19,20}. We observed that genes linked to the hominin-specific gains were expressed at a higher level in human oligodendrocytes compared to those from rhesus macaque. Moreover, these genes were expressed at a relative higher level in mature oligodendrocytes compared to precursor cells, which reflects the adult tissue in which these elements were identified¹⁸. We did not observe selective expression in subsets of adult oligodendrocytes as defined previously¹⁹, nor an altered abundance of cells between human and macaque. This supports the hypothesis that the gain in activity is not due to alterations of a oligodendrocytes subtype^{8,18}.

In summary, we show that the hominin-specific regulatory elements, that emerged during the evolution of the human brain, are coupled to gene expression changes in oligodendrocytes and although regulatory differences between the subtypes remain to be assessed, these do not result in gene expression alterations in a specific subpopulation.

Methods

Sample collection

Prefrontal cortex samples from human, chimpanzee, rhesus macaque and common marmoset brains were resampled from the brains described before^{14,18,48}. Human brain samples were obtained from the Netherlands brain bank (<http://www.brainbank.nl/>) and informed consent was obtained for all donors. Primate brain samples represent rest material involving no animal experimentation for the purpose of this work as determined by the Animal Experimental Committee (DEC) and were obtained from the Biomedical Primate Research Centre (BPRC) in Rijswijk, the Netherlands (<http://www.bprc.nl>).

Nuclei isolation

Prefrontal cortex tissue was obtained from the primate brains (~200mg) and pulverized with pestle and mortar on dry ice. Tissue powder was then homogenized in a glass douncer (Kontes Glass Co.) in 2 mL EZ buffer (Nuclei isolation kit, Sigma NUC101), followed by a 5' incubation on ice. Homogenates were centrifuged for 15' at 65 x g at 4 °C and the resulting pellets were resuspended in 2 mL EZ buffer and incubated for 5' on ice. Samples were centrifuged again for 15' at 65 x g at 4 °C, pellets were resuspended in 4 mL nuclear suspension buffer (NB; 0.01% BSA, 1x 'complete' protease inhibitor cocktail (Roche) in MQ) and filtered through a 40 µM cell strainer. Samples were centrifuged for 15' at 65 x g at 4 °C. To separate nuclei from other cell organelles, a density gradient was created with OptiPrep (Sigma D1556). Samples were resuspended in 30 mL NB, then 10 mL of 30% OptiPrep was carefully added at the bottom with a syringe and needle, followed by another 5 mL layer of 60% OptiPrep underneath. Samples were centrifuged for 10' at 1000 x g at 4 °C. Nuclei were retrieved from between the 30% OptiPrep and PBS layers with a syringe and needle and checked for quality under a bright field microscope.

FANS sorting and RNA-sequencing

Isolated nuclei were incubated with blocking solution (0.5% BSA, 10% FBS in PBS, 100 µL for 1 mL of nuclei) for 30' rolling at 4 °C. Anti-NeuN antibody conjugated with Alexa488 (1:1000, Merck Millipore MAB377X) and Hoechst 34580 (1 µg/mL, Fisher Scientific H21486) were added to the nuclei solution and incubated for 1 h at 4 °C in the dark on a roller. Stained nuclei were transferred to FACS tubes pre-coated with 5% BSA and sorted using a BD FACS-Jazz (BD Biosciences). Nuclei were first

gated based on Hoechst signal and of all Hoechst+ nuclei, G1 nuclei were selected. G1 nuclei were then gated for NeuN-AF488 signal. 50,000 NeuN+ and NeuN- nuclei were finally sorted into separate tubes, in 20 μ L EZ buffer.

Obtained nuclei were centrifuged at 9000 $\times$ g at 4 °C for 10' and resuspended in 500 μ l Trizol and incubated for 5' at RT. 200 μ l chloroform was added per sample and mixed by shaking, incubated for 5' on ice and centrifuged for 15' at 21,000 $\times$ g at 4 °C. The upper aqueous layer was transferred to a new eppendorf and 250 μ l iso-propanol and 1 μ l glycoblue (Invitrogen) was added. After which the samples were mixed by shaking and incubated at -20 °C overnight. The next day, samples were centrifuged at 21,000 $\times$ g at 4 °C for 30' and resuspended in 500 μ l 70% ethanol. Samples were handed over to single cell discoveries (<https://www.scdiscoveries.com>) for processing using the CEL-Seq1 protocol²³.

Single nuclei sequencing

Nuclei were obtained as described above, stained with Hoechst 34580 (1 μ g/ml, Fisher Scientific H21486) to select intact nuclei and sorted using a BD-influx Cell Sorter (BD bioscience). Nuclei were gated based on Hoechst signal and sorted in 384-well hard-shell plates (Biorad) with 5 μ l vapor-lock (QIAGEN) containing 100 nl of RT-primers (consisting of a 24 bp polyT stretch, a 6 bp random molecular barcode (UMI), a cell-specific barcode, the 5'-prime Illumina TruSeq small RNA kit adaptor and a T7 promoter), dNTPs and mRNA ERCC Spike-Ins (Agilent). After sorting, plates were briefly centrifuged and stored at -80 °C until further processing. Single cell RNA-sequencing libraries were prepared following the SORT-Seq protocol, an automated and improved version of CEL-Seq^{23,49}. In short, cells were lysed for 5' at 65 °C, and reverse transcriptase and second strand mixes were added using the Nanodrop II liquid handling platform (GC biotech). Single cell double stranded cDNAs were pooled to create 1 library per plate and *in vitro* transcribed for linear amplification. Illumina sequencing libraries were prepared using the TruSeq small RNA primers (Illumina) and paired-end sequenced at 75 bp read length on an Illumina NextSeq.

Gene expression analysis

For both bulk RNA and single-cell RNA sequences, paired-end sequencing reads were aligned to the reference transcriptome containing both exon and intron sequences using BWA 0.7.17-r1188⁵⁰. Read 1 contains the cell type or sample-specific barcode and was therefore used to assign reads to the correct cell or sample, while read 2 was mapped on the reference transcriptome, discarding all reads that mapped equally to multiple locations or duplicated reads. Read counts were corrected using the UMI barcode by removing duplicated reads with identical combinations of library, cell-specific, molecular barcodes and that mapped to the same gene. For every sample, the number of UMIs for each transcript was counted and transcript counts were adjusted to the expected number of molecules based on counts, number

of possible UMIs and poissonian counting statistics⁵¹. For bulk RNA analysis, NeuN-positive and NeuN-negative samples were combined to create a single list of genes expressed per species, adding a pseudo-count of 0.1 for any genes not expressed in one of the samples and rlog normalized. Cell-type-specific genes were identified based on having a differentially accessible promoter based on ATAC-Seq (see below) and at least 2-fold difference in expression. For the single cell analysis, only those cells were retained that had at least 155 transcripts, set by visual inspection of the transcript count distribution for all primates. Cells were clustered using RaceID3 from the RaceID package in R, which calculates transcriptome similarities and groups similar cells^{34,52}. Cell type interference was done by assessing expression levels of cell-type-specific genes as defined by²⁰.

Annotation of cell-type-specific open chromatin

ATAC-sequencing data was obtained from¹⁸ and re-analysed. In short, reads were trimmed using cutadapt (<https://cutadapt.readthedocs.io/en/stable/>) to remove the Nextera adapter sequence 5'-CTGTCTCTTATA-3'. Reads were then mapped to the most recent primate genome using Bowtie 1.1.2 (<http://bowtie-bio.sourceforge.net/index.shtml>). All reads with more than one mismatch, multiple alignments or duplicated reads were removed from further analysis. Accessible regions were defined using MACS2 version 2.1.1 (p-value: 10^{-5} , extsize: 300, local lambda: 100,000). Primate accessible regions were obtained by reciprocal liftover using the UCSC liftOver tool as described before¹⁸. In short, regions identified in chimpanzee, macaque and marmoset were reciprocally mapped to the human genome (hg38), discarding all regions that change more than 50% in size or mapped to multiple locations. Regions were subsequently merged to obtain a single non-redundant list of accessible regions. This list was reciprocally mapped to the different primate genomes and only those regions retained that could be mapped on all primate species, resulting in 49,322 primate GREs. Cell-type-specific accessible regions were defined per primate using DESeq2²⁶. Therefore, reads were counted for every GRE in both the NeuN-positive and NeuN-negative samples. Regions were considered cell-type-specific if they displayed a 2-fold enrichment with an FDR < 0.01 between the cell types.

Functional analysis of regulatory elements

Target genes were assigned to the GREs based on their genomic location. GREs located within 1 kb of a known human transcriptional start site (TSS, UCSC RefSeq hg38), were assigned to the associated gene and considered promoters. GREs located outside this 1 kb window were annotated as enhancer. Enhancers were linked to a target gene based on PFC HiC data²⁷. If no significant promoter loop was identified for that region, the enhancer was linked to its closest active TSS based on H3K4me3 enrichment as determined by¹⁸. Gene ontology analysis was performed using the GREAT tools version 3.0.0 (<http://bejerano.stanford.edu/great/public/html>) with

basal plus extension.

Code Availability

All analysis was done using R (<http://www.r-project.org>) or Python (<http://www.python.org>) employing the packages described above.

Acknowledgements

M.P.C., I.S.T and B.C. were supported by the Royal Academy of Sciences (KNAW) in the Netherlands and by funding acquired by the Friends of the Hubrecht Institute. We thank single cell discoveries for providing sequencing service and data.

Author Contribution

B.C., I.S.T. and M.P.C. conceived and designed the experiments. B.C. and I.S.T performed the experiments and were supervised by M.P.C. B.C analyzed the data. B.C. and M.P.C. wrote the manuscript.

Competing interest

The authors declare no competing interests

References

1. Miller, E. K. & Cohen, J. D. An integrative theory of prefrontal cortex function. *Annual review of neuroscience* 24, 167–202 (2001).
2. Hiser, J. & Koenigs, M. The Multifaceted Role of the Ventromedial Prefrontal Cortex in Emotion, Decision Making, Social Cognition, and Psychopathology. *Biological psychiatry* 83, 638–647 (2018).
3. Sherwood, C. C., Holloway, R. L., Semendeferi, K. & Hof, P. R. Is prefrontal white matter enlargement a human evolutionary specialization? *Nature neuroscience* 8, 537–8; author reply 538 (2005).
4. Schoenemann, P. T., Sheehan, M. J. & Glotzer, L. D. Prefrontal white matter volume is disproportionately larger in humans than in other primates. *Nature neuroscience* 8, 242–52 (2005).
5. Passingham, R. E. & Smaers, J. B. Is the prefrontal cortex especially enlarged in the human brain allometric relations and remapping factors. *Brain, behavior and evolution* 84, 156–66 (2014).
6. Smaers, J. B., Gómez-Robles, A., Parks, A. N. & Sherwood, C. C. Exceptional Evolutionary Expansion of Prefrontal Cortex in Great Apes and Humans. *Current biology : CB* 27, 714–720 (2017).
7. Barton, R. A. & Venditti, C. Human frontal lobes are not relatively large. *Proceedings of the National Academy of Sciences of the United States of America* 110, 9001–6 (2013).
8. Gabi, M. et al. No relative expansion of the number of prefrontal neurons in primate and human evolution. *Proceedings of the National Academy of Sciences of the United States of America* 113, 9617–22 (2016).
9. Creyghton, M. P. et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences of the United States of America* 107, 21931–21936 (2010).
10. Ernst, J. et al. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* 473, 43–9 (2011).
11. Khaitovich, P., Enard, W., Lachmann, M. & Pääbo, S. Evolution of primate gene expression. *Nature reviews. Genetics* 7, 693–702 (2006).
12. Brawand, D. et al. The evolution of gene expression levels in mammalian organs. *Nature* 478, 343–8 (2011).
13. Nowick, K., Gernat, T., Almaas, E. & Stubbs, L. Differences in human and chimpanzee gene expression patterns define an evolving network of transcription factors in brain. *Proceedings of the National Academy of Sciences of the United States of America* 106, 22358–63 (2009).
14. Vermunt, M. W. et al. Epigenomic annotation of gene regulatory alterations during evolution of the primate brain. *Nature neuroscience* 19, 494–503 (2016).
15. Zhu, Y. et al. Spatiotemporal transcriptomic divergence across human and macaque brain development. *Science* 362, eaat8077 (2018).
16. Reilly, S. K. et al. Evolutionary genomics. Evolutionary changes in promoter and enhancer activity during human corticogenesis. *Science (New York, N.Y.)* 347, 1155–9 (2015).
17. Villar, D. et al. Enhancer evolution across 20 mammalian species. *Cell* 160, 554–566 (2015).
18. Castelijns, B. et al. Hominin-specific regulatory elements selectively emerged in oligodendrocytes and are disrupted in autism patients. *Nature communications* 11, 301 (2020).
19. Jäkel, S. et al. Altered human oligodendrocyte heterogeneity in multiple sclerosis. *Nature* 566, 543–547 (2019).
20. Lake, B. B. et al. Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nature biotechnology* 36, 70–80 (2018).
21. Habib, N. et al. Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nature Methods* 14, 955–958 (2017).
22. Mullen, R. J., Buck, C. R. & Smith, A. M. NeuN, a neuronal specific nuclear protein in vertebrates. *Development (Cambridge, England)* 116, 201–11 (1992).
23. Hashimshony, T., Wagner, F., Sher, N. & Yanai, I. CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. *Cell reports* 2, 666–73 (2012).
24. Gilad, Y. & Mizrahi-Man, O. A reanalysis of mouse ENCODE comparative gene expression data. *F1000Research* 4, 121 (2015).
25. Vierstra, J. et al. Mouse regulatory DNA landscapes reveal global principles of cis-regulatory evolution. *Science* 346, 1007–1012 (2014).
26. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* 15, 550 (2014).

27. Won, H. et al. Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature* 538, 523–527 (2016).
28. Mandela, P. & Ma, X.-M. Kalirin, a key player in synapse formation, is implicated in human diseases. *Neural plasticity* 2012, 728161 (2012).
29. Berto, S. et al. Accelerated evolution of oligodendrocytes in the human brain. *Proceedings of the National Academy of Sciences of the United States of America* 116, 24334–24342 (2019).
30. Piirsoo, M., Meijer, D. & Timmusk, T. Expression analysis of the CLCA gene family in mouse and human with emphasis on the nervous system. *BMC developmental biology* 9, 10 (2009).
31. Newbury, D. F. et al. CMIP and ATP2C2 modulate phonological short-term memory in language impairment. *American journal of human genetics* 85, 264–72 (2009).
32. Klemm, S. L., Shipony, Z. & Greenleaf, W. J. Chromatin accessibility and the regulatory epigenome. *Nature reviews. Genetics* 20, 207–220 (2019).
33. Hashimshony, T. et al. CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. *Genome biology* 17, 77 (2016).
34. Grün, D. et al. Single-cell messenger RNA sequencing reveals rare intestinal cell types. *Nature* 525, 251–255 (2015).
35. Allen, N. J. & Lyons, D. A. Glia as architects of central nervous system formation and function. *Science (New York, N.Y.)* 362, 181–185 (2018).
36. He, F. & Sun, Y. E. Glial cells more than support cells? *The international journal of biochemistry & cell biology* 39, 661–5 (2007).
37. Sankowski, R. et al. Mapping microglia states in the human brain through the integration of high-dimensional techniques. *Nature Neuroscience* 22, (2019).
38. Bozek, M. et al. ATAC-seq reveals regional differences in enhancer accessibility during the establishment of spatial coordinates in the *Drosophila* blastoderm. *Genome research* 29, 771–783 (2019).
39. Miyamoto, K. et al. Chromatin Accessibility Impacts Transcriptional Reprogramming in Oocytes. *Cell reports* 24, 304–311 (2018).
40. Rada-Iglesias, A. et al. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* 470, 279–83 (2011).
41. Fisher, W. W. et al. DNA regions bound at low occupancy by transcription factors do not drive patterned reporter gene expression in *Drosophila*. *Proceedings of the National Academy of Sciences of the United States of America* 109, 21330–21335 (2012).
42. Long, H. K., Prescott, S. L. & Wysocka, J. Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. *Cell* 167, 1170–1187 (2016).
43. Emera, D., Yin, J., Reilly, S. K., Gockley, J. & Noonan, J. P. Origin and evolution of developmental enhancers in the mammalian neocortex. *Proceedings of the National Academy of Sciences* 113, E2617–E2626 (2016).
44. Hawrylycz, M. J. et al. An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* 489, 391–399 (2012).
45. Pollen, A. A. et al. Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex. *Nature biotechnology* 32, 1053–8 (2014).
46. Tasic, B. et al. Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nature Neuroscience* 19, 335–346 (2016).
47. Lake, B. B. et al. Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. 352, 352–357 (2016).
48. Vermunt, M. W. et al. Large-scale identification of coregulated enhancer networks in the adult human brain. *Cell Reports* 9, 767–779 (2014).
49. Muraro, M. J. et al. A Single-Cell Transcriptome Atlas of the Human Pancreas. *Cell systems* 3, 385–394.e3 (2016).
50. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)* 25, 1754–60 (2009).
51. Grün, D., Kester, L. & van Oudenaarden, A. Validation of noise models for single-cell transcriptomics. *Nature methods* 11, 637–40 (2014).
52. Herman, J. S., Sagar & Grün, D. FateID infers cell fate bias in multipotent progenitors from single-cell RNA-seq data. *Nature methods* 15, 379–386 (2018).

Supplementary Fig. 1 RNA-sequencing of nuclei isolated from frozen brain samples. **a**, FACS plots depicting the sorting purity of nuclei stained for NeuN, with and without OptiPrep. **b**, FACS plots depicting the sorting results using OptiPrep for nuclei isolated from the prefrontal cortex and stained for NeuN, for different primates as indicated. **c**, Microscope images depicting the stained nuclei after sorting. **d**, Venn diagrams depicting the overlap of genes expressed in NeuN+ and NeuN- sorted samples per primate as indicated. **e**, Scatter plot depicting the correlation of gene expression levels between human NeuN+ and NeuN- sorted nuclei. Known cell type marker genes are depicted as colored dots.

Supplementary Fig. 2 Cell-type-specific GREs correlate with gene expression. **a**, t-sne analysis of all ATAC enriched GREs with orthologs on all four primate genomes ($n = 49,322$). Axes indicate semantic space. **b**, Box plots depicting the fold change in gene expression between NeuN+ and NeuN- sorted nuclei for genes linked to NeuN+ and NeuN- specific GREs, per primate as indicated. Dissimilarities between distributions were calculated using a Student's t-test. **c**, ATAC-Seq tracks showing the rpm normalized ATAC signal in both NeuN+ and NeuN- sorted nuclei per primate as indicated. For regions containing a shared GRE or a GRE specific to either NeuN+ or NeuN- nuclei for all primates. **d**, Box plots depicting the fold change in gene expression between NeuN+ and NeuN- sorted nuclei for genes linked to GREs shared between NeuN+ and NeuN-. Dissimilarities between distributions were calculated using a Student's t-test. **e**, Gene-ontology analysis using GREAT for NeuN- specific GREs. P-values represent significance of enrichment for stated biological process. **f**, Gene-ontology analysis as in (e), for NeuN+ specific GREs. Bottom and top of the box plots are the first and third quartile. The line within the boxes represents the median and whiskers denote interval within 1.5x the interquartile range from the median, outliers are depicted as points.

No oligodendrocyte-specific subpopulation affected by hominin-specific rewiring

Supplementary Fig. 3 Hominin-specific GREs link to oligodendrocyte-specific genes. **a**, Bar plot depicting the number of hominin-specific regulatory changes as defined by¹⁸ that show a similar hominin-specific gain in accessibility in NeuN- sorted nuclei. **b**, ChIP-Seq and ATAC-Seq tracks showing rpm normalized H3K27ac (top panel) and ATAC (middle and lower panel) enrichment across a region containing the *CLCA4* gene, for all primates as indicated. Hominin-specific H3K27ac GRE is highlighted in green. **c**, Box plot depicting the number of mapped transcripts per cell using either only exon sequences or exon and intron sequences, per primate as indicated. **d**, t-sne maps of all nuclei isolated from the prefrontal cortex of different primates as indicated. Different colors represent different clusters as defined by RaceID. **e**, t-sne maps as in (d), showing expression levels of NeuN+ and NeuN- specific genes per primate as indicated. Transcript counts are given in log scale. **f**, t-sne map of all human and macaque nuclei combined, colors represent nuclei origin as indicated. **g**, t-sne maps as in (f), showing expression levels cell-type-specific genes as defined by²⁰. Transcript counts are given in log scale. **h**, t-sne map as in (f), showing expression levels of genes linked to evolutionary stable GREs. Transcript counts are given in log scale. **i**, t-sne map of all human and macaque nuclei identified as oligodendrocyte. Different colors represent different clusters as defined by RaceID. **j**, Box plot depicting the average log normalized expression count for genes linked to hominin-specific gains in different oligodendrocyte subtypes as defined by¹⁹. Dissimilarities between distributions were calculated using a Student's t-test. (*: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$, OPCs: oligodendrocyte precursor cells, COPs: committed oligodendrocyte precursors). Bottom and top of the box plots are the first and third quartile. The line within the boxes represents the median and whiskers denote interval within 1.5x the interquartile range from the median, outliers are depicted as points.

Chapter 5

General Discussion

B. Castelijns¹

¹Hubrecht Institute-KNAW & University Medical Center Utrecht,
Utrecht, The Netherlands

Introduction

‘Without a doubt, this is the most important, most wondrous map ever produced by humankind.’ With these words Bill Clinton announced the release of the first draft of the human genome sequence project in 2000¹. At that point, for the first time in history, we had an almost complete picture of the human sequence at every base pair. Almost 20 years later we realize that, although we have mapped our genome base-by-base, the biological function of many regions remains unknown, and for most of the genome we are able to read the letters but do not yet fully understand their role in biology.

Only 1-2% of our genome codes for actual proteins and many parts of the genome are presumed to contain regulatory function. Recent project such as ENCODE and Roadmap^{2,3} have indeed shown that many regions have active regulatory elements scattered throughout. In this thesis we have investigated how these elements emerged, and their impact on human evolution. In **chapter 2** and **chapter 3** we identified gene regulatory elements (GREs) in the brain and other organs of the common marmoset. Comparison of these elements across primate evolution revealed a link between hominin-specific regulatory gains and neurodevelopmental disorders, which is described in **chapter 2**. In **chapter 3** we used these elements to identify characteristics for evolutionary novel GREs, such as interindividual variability and lack of promoter interactions. In **chapter 4**, we used bulk RNA-sequencing and single cell sequencing to link the hominin-specific gains to gene expression changes across evolution and observed that the gain in activity is not due to alteration in a specific subset of oligodendrocytes. This chapter will further elaborate on the implications of these results and will lay out future perspectives.

The limitations of epigenetics

One of the major advances in genome biology research was the development of chromatin immunoprecipitation techniques (reviewed in^{4,5}), which has enabled the identification of distinct non-coding genetic regions that harbour regulatory function based on the presence of histone modifications^{6,7}. For example, H3K27ac marks active chromatin regions (e.g. enhancers and promoter) while trimethylation of the same residue (H3K27me3) is found on repressed chromatin⁸⁻¹¹. In the work described in this thesis we have primarily relied on the presence of H3K27ac to identify regulatory elements, as this histone modification is widely accepted as mark for active chromatin¹²⁻¹⁴. Additionally, we employed ChIP-sequencing for H3K4me3 to separate active enhancer and promoters¹⁵. Although these marks are commonly used to identify active genetic regions, it is still unclear whether these marks are required for activity or that they are simple passenger modifications that serve as a footprint for proteins involved in the regulatory machinery. So far, studies have been published that support both roles.

For example, it has recently been shown that GREs can be activated by deposition of acetylation on H3K27 by p300 fused to a catalytic dead Cas9 protein¹⁶, while

recruitment of repressors led to the deactivation of transcription¹⁷. Nevertheless, selective knockdown studies have also shown that acetylation of lysine residues is a redundant mechanism, as loss of a specific lysine will only have a minor influence on gene transcription¹⁷. It is therefore likely the combined presence of these marks that correlate with GRE activity rather than the modification of a specific residue¹⁸. Similar results have also been shown for H3K4me3, while a recent study showed that depletion of this mark had no effect on either initiation or elongation of transcription¹⁹, others have shown that this histone mark mainly acts upon environmental stress and has only minor influence on steady-state transcription²⁰. Despite these conflicting results, the mere fact that the genome is laden with hundreds of modifications, with new ones still being discovered²¹, indicates that they have potential biological roles in chromatin dynamics and serve as docking stations for transcription factors^{22,23}, rather than just being convenient markers to identify regulatory elements (reviewed in^{6,24,25}). It will therefore be of interest to further investigate the precise role of these modifications on the regulation of gene expression. Moreover, due to the multitude of these marks and their yet unclear function, identification of GREs based on only two epigenetic marks, as done in **chapter 2** and **chapter 3**, may limit the number of GREs able to be identified. For example, the efficiency of finding regulatory elements using H3K27ac is estimated around 70%^{13,26}, suggesting that many elements still remain undiscovered using this approach. Recently, several active regulatory elements have been identified harbouring H3K122ac without the presence of H3K27ac²⁷. Future assessment of different histone modifications combined with other GRE characteristics such open chromatin and the production of enhancer-RNA^{28,29} will therefore not only lead to the identification of new GREs but will also help to get a better picture on the complexity of regulatory elements.

Cell type specificity of gene regulation

In both **chapter 2** and **chapter 3** of this thesis we rely on the identification of GREs by ChIP-sequencing of bulk tissue samples. However, many regulatory elements such as enhancers are known to act in a cell-type-specific manner^{10,30,31}. The use of bulk tissue as input material could therefore result in the identification of only the most enriched GREs, while cell-type-specific elements are missed. Moreover, because all cells within a tissue sample are treated together, it is impossible to, in hindsight, assign cell-types to the identified GREs. In **chapter 2** we have resolved this problem by leveraging publicly available cell-type-specific datasets^{32,33}. However, this requires these data to be available and is limited to GREs identified in both sets.

One of the major bottlenecks of ChIP-sequencing is the requirement of millions of cells in order to obtain a decent epigenetic profile. Generating cell-type-specific profiles therefore not only requires the use of cell-type-specific antibodies but also long sorting times in order to obtain a large enough population of pure cells if the cell type of interest is rare^{33,34}. Moreover, due to the use of antibodies, only cell types for which specific markers are known can be addressed. In frozen donor tissue, such

as the brain, intact cells cannot be extracted and the isolation relies on the use of nuclear markers. For instance NeuN, for the distinction between neuronal and non-neuronal cells in **chapter 4**. With the recent development of single cell techniques it is now possible to assess regulatory elements at a single cell resolution^{35,36}, although assigning cell types based on epigenomic profiles alone remains challenging due to the sparsity of the data and the still limited knowledge of cell type dynamics of GREs^{32,36}. Combinatory approaches are therefore required to fully exploit single cell epigenetic assays^{32,37,38}. One of the promising leads is the development of single cell multi-omics techniques, in which the transcriptome as well as epigenetic marks or DNA-protein interactions can be assessed within the same cell^{39,40}.

However, with the advance of better single cell sequencing techniques, the questions of what specifically defines a cell type, and how to differentiate between them arise. How much variation do we allow between two cells to consider them different cell types? With more precise assays we will be able to identify increasingly subtle differences in genes expression between cells. For example, it has recently been shown that mouse embryonic stem cells display fluctuating gene expression profiles over time^{41,42} and that this heterogeneity influences the response to environmental cues and differentiation potential of these cells^{43,44}. Should we therefore consider these cells as different cell types or as variants within the same type⁴⁵? Similarly, Jäkel and colleagues identified different oligodendrocyte subtypes and postulated that these represent different maturation states⁴⁶. Although they largely resemble the oligodendrocyte heterogeneity found in mice^{47,48}, to what extent these differences are biological relevant and how many subpopulations reside in the brain remains to be resolved. Nevertheless, the advance of more combinatorial assessments of gene expression as well as the underlying regulatory network at a single cell level, will greatly contribute to the dissection of cell-type-specific regulatory networks in the coming years.

Deciphering evolution from an enhancer point of view

Throughout this thesis, we have assessed regulatory networks through the lens of evolution, primarily focussing on the primate lineage. By comparing GRE usage between species, we were able to assign changes in regulatory networks during primate evolution. However, in order to make fair comparisons, we only assessed regulatory elements for which we could identify orthologous sequences in all assessed primates. This was primarily done because of the difference in genome quality. While the human genome sequence is almost completely determined, the quality of the primate assemblies differs greatly and many unresolved contigs remain^{49,50}. It is therefore not always possible to distinguish between species-specific sequences and unresolved sequence gaps. The number of GREs identified in this work are thus an underrepresentation of the actual number of GREs per species. In **chapter 2**, we identified 37,308 primate GREs, comparison with the original GREs revealed that roughly 10,000 GREs per species are not assessed as they could

Fig. 1 Evolutionary time of emergence for hominin prefrontal cortex gains. **a**, Schematic representation of the primate phylogenetic tree with time indication of the major branch points. Species and associated lineages are stated on the right. Branches leading from ancestral primates to humans are defined by their taxonomy. Branch lengths are not scaled. **b**, Heatmap showing H3K27ac enrichment on scaled hominin gains, as defined in chapter 2, in the prefrontal cortex of different primate species. Heatmap colors indicate H3K27ac enrichment (rpm). **c**, ChIP-Seq track showing rpm normalized H3K27ac enrichment in the PFC across a 245 kb region containing *c-MET* gene for different primates as indicated. Two hominin gains, as defined in chapter 2, are highlighted in green.

not be unambiguously mapped to all primates. Consequently, we were unable to investigate the role of these regulatory sequences on primate evolution, while it has been shown that these sequences have impact on human biology^{51–53}. For example, it has been shown that the human genome contains roughly 15,000 human-specific insertions⁵² and that most genomic variation between primates resides in structural changes rather than nucleotide variants⁵⁴. Better genome annotation and comparative analyses are therefore crucial to further dissect true species-specific sequence differences from genome-assembly errors and will contribute to a better understanding on the evolution of regulatory elements^{54,55}.

Besides differences in genome quality, annotation of evolutionary differences in GRE usage largely depends on the sampled species. Throughout the work described in this thesis, we have used several primates as representative species of their lineage. Chimpanzee for non-human apes, rhesus macaque for old world monkeys

and the common marmoset for new world monkeys. However, every lineage consists of multiple species and the use of a single representative might therefore lead to the misclassification of elements. For example, the ape lineage consists, next to humans and chimpanzees also of bonobos, gorillas, orangutans and gibbons (Fig. 1a). Therefore, by the use of human and chimpanzee as representative species, we are unable to exactly annotate the time of emergence of GREs such as the hominin-specific gains of **chapter 2** or assess whether they are shared across the whole ape lineage. Similarly, by the use of rhesus macaque as single representative species for old world monkeys, regions annotated as loss or gain in that lineage might be considered misclassified if more species are assessed.

To investigate whether we could refine the evolutionary timing of the in **chapter 2** identified set of GREs, we recently annotated GREs in the cerebellum and prefrontal cortex of orangutan and baboon, thereby obtaining an additional ape and old world monkey sample. Re-evaluation of the hominin-specific gains and losses, showed that most hominin-gains are also active in orangutan tissues (Fig. 1b). These elements are therefore likely not hominin-specific but emerged at an earlier evolutionary timepoint within the ape lineage, while those inactive in the orangutan samples (17% of hominin gains) are potential true hominin-specific regions. However, a subset of these regions also showed H3K27ac enrichment in baboon samples (Fig. 1b). These elements could therefore represent GREs that emerged before the split of apes from new world monkeys but were subsequently lost in macaques rather than being hominin-specific gains. For example, one of the hominin-specific GREs located 90 kb upstream of the *c-MET* promoter was shown to be inactive in macaque and marmoset PFC while enriched in baboon (Fig. 1c). This indicates that analysing additional species per lineage can help fine-tune the evolutionary timing and eliminate false-positive hits due to species-specific alterations.

Furthermore, we should keep in mind that evolution is not a linear process that has led from a common primate ancestor to humans, with species divergence at regular intervals (Fig. 1a). While we assume that regions shared between human and chimpanzee were also active in their common ancestor, we have no real method to assess this. Additionally, while primates have a shared evolutionary history, the chimpanzee has had 6 million years of evolution independent of humans since the species split. This is even more pronounced when comparing relative distant species such as macaque and marmoset, which had their own clade-specific evolution independent of the apes and each other. It will therefore be important to incorporate the phylogenetic relationship of species when analysing regulatory divergence between multiple species, as pairwise comparison of GRE activity might not capture the whole interdependence of evolution^{56,57}.

Assessment of enhancer function and their modeling

ChIP-sequencing for histone marks, as performed in **chapter 2** and **chapter 3**, identifies active regulatory elements on a genome-wide scale. However, H3K27ac

Fig. 2 Hominin-specific nucleotide change creates a SOX-binding site. **a**, Upper panel: ChIP-Seq track showing rpm normalized H3K27ac enrichment in human PFC across a 2 kb region containing the 5' *c-MET* enhancer. Hominin-specific nucleotide changes are depicted underneath, with those that create an *OLIG2* or *SOX10* binding site highlighted in red. Lower panel: alignment of the human, chimpanzee, macaque and marmoset sequence surrounding the created *SOX10* binding site. Hominin-specific nucleotide change highlighted in yellow, *SOX10* binding motif displayed above. Asterisks indicate conserved alignment between the primate species. **b**, Scatterplot depicting the binding of peptides to either the human or macaque-specific sequence surrounding the *SOX10* binding site. Each axis is the log₂ transformed light/heavy dimethyl ratio from a single replicate. Each data point is a peptide. Enriched peptides using a significance cutoff of 1.5 interquartile ranges are colored in red. Background proteins are colored in black.

enrichment alone does not indicate functionality, as seen in **chapter 3**, where many enhancers harbour H3K27ac but seem to be devoid of function as they do not engage in any contact with other genomic regions. Moreover, from ChIP-Seq experiments alone it is impossible to properly assign functionality to the identified GREs. Additional techniques are therefore required to identify target genes or functional effects of enhancers. Throughout this thesis we have assigned target genes based on chromatin conformation. For example, in **chapter 2** we have shown that two hominin-specific gains are in close conformation with the *c-MET* promoter, which indicates that these elements likely regulate expression of this gene. While in **chapter 3** we have leveraged publicly available HiC data to, on a genome-wide-scale, assign target genes to every identified GRE. While these methods can reliably point toward target genes of distal regulatory elements, they do not elucidate how these GREs act and what their true activating potential is. In order to assess enhancer functionality different assays are thus required.

A method to assess how enhancers function is by analysing the sequence for transcription factor binding sites that explain their activity^{58,59}. In **chapter 2** we performed motif analysis to identify transcription factors that could account for the observed changes in activity⁶⁰. Although this did not identify a common transcriptional program, analysis of individual elements did reveal some interesting candidates. For example, sequence analysis of a hominin-specific enhancer located upstream of the *c-MET* gene revealed the creation of both an *OLIG2* and *SOX10* binding site in

the hominin-lineage (Fig. 2a). As these factors are both essential for oligodendrocyte development^{61,62}, we postulated that the oligodendrocyte-specific gain of activity could be due to the recruitment of these factors. We therefore performed a protein pull down using both the human as well as the macaque sequence surrounding the *SOX10* binding-site (Fig. 2b). Although we did not observe differential enrichment for *SOX10*, other SOX factors were specifically recruited to the human sequence *in-vitro* (Fig. 2b), indicating that the hominin-specific DNA change created a functional SOX binding site. As this does not yet confirm a role for *SOX10* in the activity of this enhancer, assessment on whether SOX-proteins actually bind this enhancer *in vivo*, for example by ChIP-sequencing, can hopefully further elucidate the effect of the creation of a transcription factor binding site on the activation of this locus.

Another method to assess enhancer function is by performing a knock-out study, in which individual enhancers are deleted to test their effect on target gene expression^{63–65}. For example, a recent study showed that an enhancer linked to the *SOX9* gene is required for testis development and that deletion of the enhancer led to complete sex reversal in mouse embryos⁶³. As generating human enhancer knock-outs is not possible, these kinds of studies are limited to rodent species⁶⁶ or non-human primates⁶⁷, in which trait associated enhancers are largely conserved as shown in **chapter 3**. Nevertheless, this requires the generation of knock-out animals which is a laborious process, especially for non-human primates⁶⁸. A way to circumvent this is by studying the effect of enhancer deletion in cell lines^{69–71}. However, this is difficult for enhancers identified in the mature brain, which not only consists of many different cell types but for most of which no pure *in-vitro* culture system exists. Although some cell-lines have been generated that resemble neuronal and non-neuronal cells (e.g. Sy5Y⁷² or G144⁷³), these are all derived from tumors and do not fully resemble mature cells states. One approach to avoid the use of tumor cell lines is with the culture of primary cell, directly isolated from the brain^{74–76} or by differentiation of embryonic stem cells (ESC) into the desired cell type^{77,78}. The latter has the advantage of providing an unlimited supply of starting material, which can be easily modified using gene editing methods such as CRISPR-Cas9⁷⁹. Several protocols have been developed in recent years that allow the differentiation of human ESC (hESC) into both neuronal cells such as glutamatergic neurons⁷⁷, as well as glial cell such as oligodendrocytes⁷⁸.

As an example, we assessed the functionality of the hominin-specific GRE located upstream of the *c-MET* gene. To study its effect on *c-MET* expression we deleted the enhancer using CRISPR-Cas9 in hESC and created a wild-type (WT) reporter line (Fig. 3a, b). We then performed a co-culture with these lines and differentiated them to oligodendrocytes, in which we have shown that the enhancer is active (Fig. 3c). Single cell sequencing revealed a slight difference in differentiation potential as knock-out cells seemed to be depleted from the glial cell lineage (Fig. 3d, e), indicating that proper regulation of *c-MET* is required for differentiation into oligodendrocytes. This is in line with previous research which showed that increased *c-MET* expression

Fig. 3 *c-MET* enhancer knock-out alters differential potential of hESC. **a**, Schematic of the reported knock-in construct and its location within the AAVS1 locus. A construct containing a mScarlet and neonGreen fluorescent protein controlled by a EF1- α promoter was, combined with a puromycin cassette, inserted in the first intron of the *PPP1R12C* gene using CRISPR-Cas9. **b**, Schematic for the knock-out of the 5' *c-MET* enhancer. Two guides flanking the 5' enhancer were used simultaneously to generate the enhancer knock-out. Depicted primers were used to screen clones for homozygous knock-outs. **c**, Schematic overview of hESC differentiation protocol. hESC wild-type (WT) cells were engineered to create either the reporter knock-in or enhancer knock-out as shown in (**a**, **b**). Cells were then mixed in a 1:1 ratio to create a co-culture and differentiated into oligodendrocytes. After 50 days cell were harvested and sorted to generate single cell transcriptome profiles. **d**, t-Distributed Stochastic Neighbour Embedding (t-sne) map of hESC differentiated into oligodendrocytes. Different colors represent different clusters as defined by RaceID and corresponding cell lineages were defined based on cell type marker genes. **e**, t-sne map as in (**d**) where the colors represent the origin of the cell. KO: *c-MET* enhancer knock-out cell, WT: hESC containing the mScarlet/neonGreen reporter line. Differences in ratio of WT/KO per cluster was compared to the overall cell count. **f**, t-sne map as in (**d**), showing expression levels of *c-MET*. Transcript counts are given in log scale.

induces proliferation of oligodendrocyte precursor cells⁸⁰. However, we were unable to assess the direct effect on *c-MET* expression as the gene was lowly expressed in our culture system (Fig. 3f). It therefore remains to be addressed to what extent and in which cell type the knock-out of this GRE influences *c-MET* expression. Despite the limitations of differentiation assays, such as the ability to obtain a pure culture^{81,82}, their relative long differentiation timing⁸³ and the variability between cell lines^{81,83}. Neuronal cultures represent a powerful tool to study enhancer functionality and the effect of DNA variants *in-vitro*^{84,85}.

Neuronal culture systems can also be used to elucidate the effect of enhancers changes during primate evolution, for example by the use of brain organoids⁸⁶⁻⁸⁹. These organoids offer an improved model system for the brain compared to 2D cultures as they capture the complex cell type content of the human brain as well as its developmental progress^{86,87}. Moreover, organoids can be grown *in-vitro* for over 20 months and have been shown to recapitulate both the chromatin accessibility dynamics and the transcriptional landscape of the developing forebrain^{88,90}. Recent comparisons of neuronal differentiation between cells derived from humans and other primates have identified numerous differences in developmental processes⁹¹⁻⁹³. For example, several genes have been shown to be differentially expressed in human organoids compared to chimpanzee and other primates, which they could link to a human-specific increase in activation of the PI3K-AKT-mTOR signalling pathway specifically in radial glial cells⁹¹. While others identified remarkable differences in developmental timing between humans and other species^{91,93}. This difference in timing has also been observed in gene expression profiles directly sampled from brains of different age⁹⁴. Together these studies point towards a prolonged development of the human brain compared to non-human primates. This extended developmental period is one of the proposed mechanisms that has led to the increased size of the human brain, as it expands the number neural progenitor cells^{95,96}. However, this difference in timing is also one of the major difficulties of these comparisons, as identifying gene expression changes during development requires the correct linkage of developmental stages between species using pseudotime analysis^{92,94,97}. Nevertheless, brain organoids offer a promising tool to unravel the differences in brain development between humans and non-human primates.

Concluding remarks

In this thesis we have tried to shed light on how regulatory elements evolved during primate evolution and which changes have enabled the divergence of humans from its closely related species. In **chapter 2** we have shown that a large set of brain regulatory elements emerged after the split of great apes (human, chimp) from the rest of the primate lineage. These hominin-specific regulatory changes preferentially emerged in oligodendrocytes and activate between early postnatal life and adulthood. Moreover, when comparing them to disease associated GREs, we found a clear enrichment for regions that are evolutionary novel that seem to be deregulated in autism spectrum disorder, suggesting that regions required for the more complex brain are also the ones that are affected during disease. However, the question still remains as to whether deregulation of these regions has major consequences on the disease or that they are an indirect effect due to other factors.

In **chapter 3** we showed that evolutionary novel GREs are inherently variable between individuals without a clear functionality. This has major implication on other studies, as it became clear that absence of an enhancer might be caused by subsampling within the population. A lot of GREs that are identified as being

species, cell-type or temporal-specific might just be variable within the population and a sufficient sample size is therefore required to further validate these findings as truly specific enhancers. One of the explanations for this variability is that many of these enhancers seem to appear without any biological function as they do not contact any other genomic regions such as promoters. We termed these regions unengaged enhancers and further investigations will have to reveal whether they are really stand-alone or that the right environment or signals are needed to anchor them towards genes.

Finally, in **chapter 4**, we showed that hominin-specific gains identified in the prefrontal cortex link to gene expression changes in oligodendrocytes. This confirmed the observed oligodendrocyte specificity of the regulatory elements and hints towards a potential role of oligodendrocytes in the divergence of the human brain from the other non-human primates. However, the underlying genetic program of these hominin-specific elements remains elusive and further investigation is required to identify key factors that have contributed to this gain of activity. Moreover, whereas we observed that these changes are not specific for a subtype of oligodendrocytes in the adult brain, the emergence of organoid systems now makes it possible to study these changes during development. Further investigation of these elements in both human and non-human primate brain organoids will hopefully elucidate the role of these elements in development and, in combination with single cell sequencing techniques, point toward the cell-type-specific effects of these regulatory gains.

To conclude, while we have identified numerous regulatory alterations that could account for the phenotypic differences between humans and other species, there is still a lot of work to be done to elucidate what the functional outcome of these changes is. It will therefore be of great interest to combine the recently developed single cell techniques with targeted enhancer knock-outs in human and primate brain organoids.

References

1. Secretary, Office of the President. On the completion of the first survey of the entire human genome project. (2000). https://web.ornl.gov/sci/techresources/Human_Genome/project/clinton2.shtml
2. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
3. Kundaje, A. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
4. Mundade, R., Ozer, H. G., Wei, H., Prabhu, L. & Lu, T. Role of ChIP-seq in the discovery of transcription factor binding sites, differential gene regulation mechanism, epigenetic marks and beyond. *Cell cycle (Georgetown, Tex.)* **13**, 2847–52 (2014).
5. Kim, T. H. & Ren, B. Genome-wide analysis of protein-DNA interactions. *Annual review of genomics and human genetics* **7**, 81–102 (2006).
6. Calo, E. & Wysocka, J. Modification of enhancer chromatin: what, how, and why? *Molecular cell* **49**, 825–37 (2013).
7. Zhou, V. W., Goren, A. & Bernstein, B. E. Charting histone modifications and the functional organization of mammalian genomes. *Nature reviews. Genetics* **12**, 7–18 (2011).
8. Pan, G. *et al.* Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells. *Cell stem cell* **1**, 299–312 (2007).
9. Mikkelsen, T. S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553–60 (2007).
10. Creighton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 21931–21936 (2010).
11. Heintzman, N. D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature genetics* **39**, 311–8 (2007).
12. Nord, A. S. *et al.* Rapid and pervasive changes in genome-wide enhancer usage during mammalian development. *Cell* **155**, 1521–1531 (2013).
13. Bonn, S. *et al.* Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. *Nature Genetics* **44**, 148–156 (2012).
14. Villar, D. *et al.* Enhancer evolution across 20 mammalian species. *Cell* **160**, 554–566 (2015).
15. Guenther, M. G., Levine, S. S., Boyer, L. A., Jaenisch, R. & Young, R. A. A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* **130**, 77–88 (2007).
16. Hilton, I. B. *et al.* Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nature biotechnology* **33**, 510–7 (2015).
17. Martin, A. M., Pouchnik, D. J., Walker, J. L. & Wyrick, J. J. Redundant roles for histone H3 N-terminal lysine residues in subtelomeric gene repression in *Saccharomyces cerevisiae*. *Genetics* **167**, 1123–32 (2004).
18. Zentner, G. E. & Henikoff, S. Regulation of nucleosome dynamics by histone modifications. *Nature structural & molecular biology* **20**, 259–66 (2013).
19. Murray, S. C. *et al.* H3K4me3 is neither instructive for, nor informed by, transcription. *bioRxiv* 709014 (2019). doi:10.1101/709014
20. Weiner, A. *et al.* Systematic dissection of roles for chromatin regulators in a yeast stress response. *PLoS biology* **10**, e1001369 (2012).
21. Farrelly, L. A. *et al.* Histone serotonylation is a permissive modification that enhances TFIID binding to H3K4me3. *Nature* **567**, 535–539 (2019).
22. Wang, L. *et al.* Histone Modifications Regulate Chromatin Compartmentalization by Contributing to a Phase Separation Mechanism. *Molecular cell* **76**, 646–659.e6 (2019).
23. Laugesen, A., Højfeldt, J. W. & Helin, K. Molecular Mechanisms Directing PRC2 Recruitment and H3K27 Methylation. *Molecular cell* **74**, 8–18 (2019).
24. Stricker, S. H., Köferle, A. & Beck, S. From profiles to function in epigenomics. *Nature reviews. Genetics* **18**, 51–66 (2017).
25. Zhao, Y. & Garcia, B. A. Comprehensive Catalog of Currently Documented Histone Modifications.

- Cold Spring Harbor perspectives in biology* 7, a025064 (2015).
26. Arnold, C. D. *et al.* Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science (New York, N.Y.)* **339**, 1074–7 (2013).
 27. Pradeepa, M. M. *et al.* Histone H3 globular domain acetylation identifies a new class of enhancers. *Nature genetics* **48**, 681–6 (2016).
 28. Andersson, R. *et al.* An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455–461 (2014).
 29. Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Current protocols in molecular biology* **109**, 21.29.1–9 (2015).
 30. FANTOM Consortium and the RIKEN PMI and CLST (DGT) *et al.* A promoter-level mammalian expression atlas. *Nature* **507**, 462–70 (2014).
 31. Ernst, J. *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43–9 (2011).
 32. Lake, B. B. *et al.* Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nature biotechnology* **36**, 70–80 (2018).
 33. Kozlenkov, A. *et al.* A unique role for DNA (hydroxy)methylation in epigenetic regulation of human inhibitory neurons. *Science Advances* **4**, 1–12 (2018).
 34. Hu, R. *et al.* Brain cell type – specific enhancer – promoter interactome maps and disease-risk association. **1139**, 1134–1139 (2019).
 35. Rotem, A. *et al.* Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. *Nature biotechnology* **33**, 1165–72 (2015).
 36. Cusanovich, D. A. *et al.* A Single-Cell Atlas of In Vivo Mammalian Chromatin Accessibility. *Cell* 1309–1324 (2018). doi:10.1016/j.cell.2018.06.052
 37. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* 1888–1902 (2019). doi:10.1016/j.cell.2019.05.031
 38. Welch, J. D. *et al.* Single-Cell Multi-omic Integration Compares and Contrasts Features of Brain Cell Identity. *Cell* 1873–1887 (2019). doi:10.1016/j.cell.2019.05.006
 39. Hou, Y. *et al.* Single-cell triple omics sequencing reveals genetic, epigenetic, and transcriptomic heterogeneity in hepatocellular carcinomas. *Cell research* **26**, 304–19 (2016).
 40. Rooijers, K. *et al.* Simultaneous quantification of protein-DNA contacts and transcriptomes in single cells. *Nature biotechnology* **37**, 766–772 (2019).
 41. Kolodziejczyk, A. A. *et al.* Single Cell RNA-Sequencing of Pluripotent States Unlocks Modular Transcriptional Variation. *Cell stem cell* **17**, 471–85 (2015).
 42. Karwacki-Neisius, V. *et al.* Reduced Oct4 expression directs a robust pluripotent state with distinct signaling activity and increased enhancer occupancy by Oct4 and Nanog. *Cell stem cell* **12**, 531–45 (2013).
 43. Kalmar, T. *et al.* Regulated fluctuations in nanog expression mediate cell fate decisions in embryonic stem cells. *PLoS biology* **7**, e1000149 (2009).
 44. Karwacki-Neisius, V. *et al.* Reduced Oct4 expression directs a robust pluripotent state with distinct signaling activity and increased enhancer occupancy by Oct4 and Nanog. *Cell Stem Cell* **12**, 531–545 (2013).
 45. Fu, X., He, F., Li, Y., Shahveranov, A. & Hutchins, A. P. Genomic and molecular control of cell type and cell type conversions. *Cell regeneration (London, England)* **6**, 1–7 (2017).
 46. Jäkel, S. *et al.* Altered human oligodendrocyte heterogeneity in multiple sclerosis. *Nature* **566**, 543–547 (2019).
 47. Marques, S. *et al.* Oligodendrocyte heterogeneity in the mouse juvenile and adult central nervous system. *Science* **352**, 1326–1329 (2016).
 48. Falcão, A. M. *et al.* Disease-specific oligodendrocyte lineage cells arise in multiple sclerosis. *Nature medicine* **24**, 1837–1844 (2018).
 49. Kronenberg, Z. N. *et al.* High-resolution comparative analysis of great ape genomes. *Science* **360**, eaar6343 (2018).
 50. Sato, K. *et al.* Resequencing of the common marmoset genome improves genome assemblies and gene-coding sequence analysis. *Scientific Reports* **5**, 1–8 (2015).
 51. Jacques, P.-É., Jeyakani, J. & Bourque, G. The majority of primate-specific regulatory sequences

- are derived from transposable elements. *PLoS genetics* **9**, e1003504 (2013).
52. Tang, W., Mun, S., Joshi, A., Han, K. & Liang, P. Mobile elements contribute to the uniqueness of human genome with 15,000 human-specific insertions and 14 Mbp sequence increase. *DNA research : an international journal for rapid publication of reports on genes and genomes* **25**, 521–533 (2018).
 53. McLean, C. Y. *et al.* Human-specific loss of regulatory DNA and the evolution of human-specific traits. *Nature* **471**, 216–9 (2011).
 54. Marques-Bonet, T., Ryder, O. A. & Eichler, E. E. Sequencing primate genomes: what have we learned? *Annual review of genomics and human genetics* **10**, 355–86 (2009).
 55. Franchini, L. F. & Pollard, K. S. Human evolution: The non-coding revolution. *BMC Biology* **15**, 1–12 (2017).
 56. Dunn, C. W., Zapata, F., Munro, C., Siebert, S. & Hejnal, A. Pairwise comparisons across species are problematic when analyzing functional genomic data. *Proceedings of the National Academy of Sciences of the United States of America* **115**, E409–E417 (2018).
 57. Rohlf, R. v & Nielsen, R. Phylogenetic ANOVA: The Expression Variance and Evolution Model for Quantitative Trait Evolution. *Systematic biology* **64**, 695–708 (2015).
 58. Vandel, J., Cassan, O., Lèbre, S., Lecellier, C.-H. & Bréhélin, L. Probing transcription factor combinatorics in different promoter classes and in enhancers. *BMC genomics* **20**, 103 (2019).
 59. Grossman, S. R. *et al.* Systematic dissection of genomic features determining transcription factor binding and enhancer function. *Proceedings of the National Academy of Sciences of the United States of America* **114**, E1291–E1300 (2017).
 60. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular cell* **38**, 576–89 (2010).
 61. Pozniak, C. D. *et al.* Sox10 directs neural stem cells toward the oligodendrocyte lineage by decreasing Suppressor of Fused expression. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 21795–800 (2010).
 62. Takebayashi, H. *et al.* The basic helix-loop-helix factor olig2 is essential for the development of motoneuron and oligodendrocyte lineages. *Current biology : CB* **12**, 1157–63 (2002).
 63. Gonen, N. *et al.* Sex reversal following deletion of a single distal enhancer of Sox9. *Science (New York, N.Y.)* **360**, 1469–1473 (2018).
 64. Nott, A. *et al.* Cell type-specific enhancer-promoter connectivity maps in the human brain and disease risk association. *bioRxiv* 778183 (2019). doi:10.1101/778183
 65. Cunningham, T. J., Lancman, J. J., Berenguer, M., Dong, P. D. S. & Duester, G. Genomic Knockout of Two Presumed Forelimb Tbx5 Enhancers Reveals They Are Nonessential for Limb Development. *Cell reports* **23**, 3146–3151 (2018).
 66. Dickel, D. E. *et al.* Ultraconserved Enhancers Are Required for Normal Development. *Cell* **172**, 491–499.e15 (2018).
 67. Kumita, W. *et al.* Efficient generation of Knock-in/Knock-out marmoset embryo via CRISPR/Cas9 gene editing. *Scientific reports* **9**, 12719 (2019).
 68. Izpisua Belmonte, J. C. *et al.* Brains, genes, and primates. *Neuron* **86**, 617–31 (2015).
 69. Tak, Y. G. *et al.* Effects on the transcriptome upon deletion of a distal element cannot be predicted by the size of the H3K27Ac peak in human cells. *Nucleic acids research* **44**, 4123–33 (2016).
 70. Bahr, C. *et al.* A Myc enhancer cluster regulates normal and leukaemic haematopoietic stem cell hierarchies. *Nature* **553**, 515–520 (2018).
 71. Diao, Y. *et al.* A tiling-deletion-based genetic screen for cis-regulatory element identification in mammalian cells. *Nature methods* **14**, 629–635 (2017).
 72. Biedler, J. L., Helson, L. & Spengler, B. A. Morphology and growth, tumorigenicity, and cytogenetics of human neuroblastoma cells in continuous culture. *Cancer research* **33**, 2643–52 (1973).
 73. Pollard, S. M. *et al.* Glioma Stem Cell Lines Expanded in Adherent Culture Have Tumor-Specific Phenotypes and Are Suitable for Chemical and Genetic Screens. *Cell Stem Cell* **4**, 568–580 (2009).
 74. Chen, Y. *et al.* Isolation and culture of rat and mouse oligodendrocyte precursor cells. *Nature Protocols* **2**, 1044–1051 (2007).
 75. Malik, A. N. *et al.* Genome-wide identification and characterization of functional neuronal activity-dependent enhancers. *Nature neuroscience* **17**, 1330–9 (2014).

76. Ray, B., Chopra, N., Long, J. M. & Lahiri, D. K. Human primary mixed brain cultures: preparation, differentiation, characterization and application to neuroscience research. *Molecular brain* **7**, 63 (2014).
77. Shi, Y., Kirwan, P. & Livesey, F. J. Directed differentiation of human pluripotent stem cells to cerebral cortex neurons and neural networks. *Nature Protocols* **7**, 1836–1846 (2012).
78. Douvaras, P. & Fossati, V. Generation and isolation of oligodendrocyte progenitor cells from human pluripotent stem cells. *Nature Protocols* **10**, 1143–1154 (2015).
79. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nature Protocols* **8**, 2281–2308 (2013).
80. Yan, H. & Rivkees, S. A. Hepatocyte growth factor stimulates the proliferation and migration of oligodendrocyte precursor cells. *Journal of neuroscience research* **69**, 597–606 (2002).
81. Zeng, X. & Couture, L. A. Pluripotent stem cells for Parkinson's disease: progress and challenges. *Stem cell research & therapy* **4**, 25 (2013).
82. Burke, E. E. *et al.* Dissecting transcriptomic signatures of neuronal differentiation and maturation using iPSCs. *Nature communications* **11**, 462 (2020).
83. Tao, Y. & Zhang, S.-C. Neural Subtype Specification from Human Pluripotent Stem Cells. *Cell stem cell* **19**, 573–586 (2016).
84. Soldner, F. *et al.* Parkinson-associated risk variant in distal enhancer of α -synuclein modulates target gene expression. *Nature* **533**, 95–99 (2016).
85. Prescott, S. L. *et al.* Enhancer divergence and cis-regulatory evolution in the human and chimp neural crest. *Cell* **163**, 68–83 (2015).
86. Lancaster, M. A. *et al.* Cerebral organoids model human brain development and microcephaly. *Nature* **501**, 373–379 (2013).
87. Velasco, S. *et al.* Individual brain organoids reproducibly form cell diversity of the human cerebral cortex. *Nature* (2019). doi:10.1038/s41586-019-1289-x
88. Trevino, A. E. *et al.* Chromatin accessibility dynamics in a model of human forebrain development. *Science (New York, N.Y.)* **367**, (2020).
89. Pasca, S. P. The rise of three-dimensional human brain cultures. *Nature* **553**, 437–445 (2018).
90. Sloan, S. A. *et al.* Human Astrocyte Maturation Captured in 3D Cerebral Cortical Spheroids Derived from Pluripotent Stem Cells. *Neuron* **95**, 779–790.e6 (2017).
91. Pollen, A. A. *et al.* Establishing Cerebral Organoids as Models of Human-Specific Brain Evolution. *Cell* **176**, 743–756.e17 (2019).
92. Kanton, S. *et al.* Organoid single-cell genomic atlas uncovers human-specific features of brain development. *Nature* **574**, 418–422 (2019).
93. Otani, T., Marchetto, M. C., Gage, F. H., Simons, B. D. & Livesey, F. J. 2D and 3D Stem Cell Models of Primate Cortical Development Identify Species-Specific Differences in Progenitor Behavior Contributing to Brain Size. *Cell stem cell* **18**, 467–80 (2016).
94. Zhu, Y. *et al.* Spatiotemporal transcriptomic divergence across human and macaque brain development. *Science* **362**, eaat8077 (2018).
95. Kalebic, N. *et al.* Neocortical Expansion Due to Increased Proliferation of Basal Progenitors Is Linked to Changes in Their Morphology. *Cell stem cell* **24**, 535–550.e9 (2019).
96. Rakic, P. A small step for the cell, a giant leap for mankind: a hypothesis of neocortical expansion during evolution. *Trends in neurosciences* **18**, 383–8 (1995).
97. Boroviak, T. *et al.* Single cell transcriptome analysis of human, marmoset and mouse embryos reveals common and divergent features of preimplantation development. *Development* **145**, dev167833 (2018).

Addendum

Nederlandse samenvatting

Acknowledgements

List of publications

Curriculum Vitea

Nederlandse samenvatting

Het menselijke genoom bevat alle informatie die nodig is voor het maken van een compleet lichaam vanuit één enkele bevruchte eicel. Dit genoom bestaat uit lange ketens DNA, opgebouwd uit 4 verschillende nucleotiden (A, T, G, C) en is in totaal ongeveer 3 miljard letters lang. Het is daarom zeer compact en gestructureerd opgevouwen om in een celkern te passen. De basis bouwsteen van het genoom is de nucleosoom, een complex van 8 histon eiwitten omringd door ~150 nucleotiden. De vouwing van deze nucleosomen in zogenoemde 30nm-fibers en de welbekende chromosoom structuur zorgen ervoor dat het gehele genoom in de celkern past (**figuur 1a** van de introductie). In 2005 is de sequentie van het menselijke genoom voor het eerst volledig in kaart gebracht. Hierdoor ontdekten we dat dit genoom codeert voor ongeveer 20.000 genen. Dit is evenveel als in andere zoogdieren zoals chimpansees en muizen maar veel minder dan in bijvoorbeeld planten, wat betekent dat het aantal genen of de lengte van het genoom geen directe indicatie is voor de complexiteit van een organisme. De meest opmerkelijke vinding was echter dat het grootste gedeelte van het genoom niet codeert voor eiwitten en dus geen functie lijkt te hebben. Deze niet-coderende gedeeltes werden lange tijd gezien als zogenoemd “junk-DNA”, delen van het genoom die overgebleven waren vanuit evolutie maar geen functionaliteit meer hadden. Recent onderzoek door onder andere het ENCODE (Encyclopedia of DNA Elements) project in 2012 heeft echter laten zien dat er wel degelijk biologische activiteit te vinden is in deze regio's. Het is nu dan ook duidelijk dat, ondanks dat grootte delen van het genoom niet coderen voor eiwitten, deze bezaaid zijn met actieve genexpressie regulerende elementen.

Omdat elke cel in een menselijk lichaam in principe dezelfde DNA-sequentie heeft, maar er wel honderden verschillende celtypen zijn, zal elke cel andere delen van het genoom moeten activeren om zo de juiste genen tot expressie te brengen. Dit wordt geregeld door de zogenoemde gen-regulerende-elementen (GREs), deze elementen coderen niet voor eiwitten maar zorgen ervoor dat de juiste genen op het juiste moment tot expressie worden gebracht. GREs zijn grofweg te verdelen in twee typen: promotoren die direct aan het begin van een gen te vinden zijn en enhancers die tot op een afstand tot wel 1 miljoen nucleotiden van een gen kunnen liggen. Terwijl promotoren zorgen voor de basale activiteit van een gen zijn enhancers relatief specifiek, ze reguleren expressie vaak in één bepaald celtype of ontwikkelingsfase. Het totaal aantal GREs in het menselijke genoom is nog niet bekend, maar er wordt in het algemeen van uit gegaan dat onze ongeveer 20.000 genen gereguleerd worden door meer dan 400.000 verschillende GREs. Enhancers reguleren expressie door het binden van bepaalde eiwitten, zoals transcriptie factoren, en brengen deze in nabijheid van een gen met behulp van enhancer-promoter lussen (**figuur 1b** van de introductie). Deze dynamische verbindingen zorgen ervoor dat enhancers genexpressie kunnen reguleren onafhankelijk van de afstand of oriëntatie van het te reguleren gen.

Omdat GREs niet coderen voor eiwitten worden ze doorgaans geïdentificeerd door middel van epigenetische modificaties. Dit zijn chemische verbindingen die op het DNA of de nucleosomen zitten en de onderliggende sequentie niet veranderen. Hoewel het effect van deze modificaties nog niet volledig bekend is, lijken ze de activiteit van GREs te reguleren door óf transcriptie factoren te rekruteren óf door de interactie van het DNA met de onderliggende histonen te moduleren. In de laatste jaren zijn de locaties van veel van deze modificaties in kaart gebracht in verschillende celtypen en organismen. Hierdoor weten we nu dat verschillende elementen specifieke modificaties hebben afhankelijk van hun functie. Zo worden actieve enhancers gemarkeerd door acetylactie van lysine 27 van histon 3 (H3K27ac), terwijl methylering van datzelfde lysine inactieve delen markeert. Het lokaliseren van deze modificaties kan dus worden gebruikt om actieve GREs te identificeren in verschillende celtypen of weefsels.

Hoewel slechts 1-2% van ons genoom codeert voor eiwitten, is in de laatste jaren gebleken dat de sequenties van deze eiwitten relatief geconserveerd zijn gedurende evolutie. Zo komt de DNA-sequentie van genen in mensen voor ongeveer 98% overeen met chimpansees en zo'n 90% met muizen. Zelfs met zebravissen delen we ongeveer 50% van deze sequenties. De meeste variatie is juist te vinden in de niet-coderende gedeeltes van het genoom en deze veranderingen hebben dus potentieel effect op de functie van regulerende elementen zoals enhancers. Recent onderzoek heeft dan ook aangetoond dat er een grote diversiteit zit in enhancer activiteit gedurende evolutie. Zo heeft elke soort zo'n 10.000 specifieke enhancer elementen. Omdat een verandering in enhancer activiteit vaak een relatief klein effect heeft op genexpressie, door bijvoorbeeld compensatie van andere enhancers, en enhancers vaak celtype specifiek zijn, is de snelle omloop van enhancers een ideaal speelveld voor evolutie om genexpressie patronen aan te passen in een specifieke cel zonder de eiwitten zelf te veranderen.

Sinds de voltooiing van het human genome project, waarin voor het eerst het gehele menselijke genoom in kaart is gebracht, zijn steeds meer genomen gesequenced. Hierdoor hebben we een steeds beter inzicht in de genetische variatie tussen individuen en blijkt dat iedereen ongeveer 8000 unieke DNA-varianten heeft. Deze variaties variëren van een enkel-nucleotide polymorfisme, kleine inserties en deleties tot grotere structurele variaties zoals translocaties en kopie nummer variaties. Om in kaart te brengen hoe deze genetische variatie bijdraagt aan de ontvankelijkheid tot ziektes zijn er verschillende grote zogenaamde genoom-wijde associatie studies (GWAS) uitgevoerd, waarbij wordt geprobeerd om te achterhalen hoeveel invloed een DNA variant heeft op een bepaald fenotype. Vandaag de dag zijn er al meer dan 3000 verschillende GWAS-studies uitgevoerd, naar eenzelfde aantal fenotypes. Deze variëren van onderzoek naar de erfelijkheid van ziektes zoals Parkinson en Alzheimer tot meer generieke eigenschappen zoals intelligentie en lengte. Van de meeste op deze manier geïdentificeerde DNA variaties blijft de exacte invloed echter onduidelijk, daarnaast verklaren ze vaak maar een klein gedeelte

van de zichtbare fenotypische verschillen. Net als evolutionaire verschillen zien we hier dat de meeste variaties tussen individuen zich bevindt in de niet-coderende gedeeltes van het genoom en dat deze dus voornamelijk effect hebben op enhancers.

Dit betekent dat veranderingen in genregulatie waarschijnlijk van grote invloed zijn op zowel de ontwikkeling van ziektes als de evolutie van de mens. En dat het begrijpen van de wisselwerking tussen genetische variatie en gen regulerende elementen essentieel is om de zichtbare fenotypes te kunnen verklaren (**figuur 1c** van de introductie). In het onderzoek beschreven in deze thesis gaan we in op de verschillen in genregulatie tussen mensen en verschillende primaten en proberen we de invloed van deze veranderingen tijdens evolutie als mede de implicatie voor ziekte te doorgronden.

In **hoofdstuk 2** focussen we op de evolutie van GREs in het primaten brein, en dan met name de frontale cortex, een belangrijke brein regio betrokken bij cognitieve en emotionele functies zoals plannen, beslissingen nemen en sociaal gedrag, en het cerebellum, ook wel kleine hersenen genoemd en voornamelijk verantwoordelijk voor motoriek. Hiervoor hebben we actieve enhancers en promotoren in het brein van de gewone penseelaap (*Callithrix jacchus*) geïdentificeerd. Deze zijn vervolgens vergeleken met actieve GREs in de hersenen van mensen, chimpansees en resusapen. Hierdoor waren we in staat de veranderingen in genregulatie te identificeren die zijn ontstaan tijdens de evolutie van primaten. We ontdekten een groep GREs die specifiek actief zijn in mensen en chimpansees en in niet resusapen en penseelapen. Deze zogenoemde hominini-specifieke GREs zijn voornamelijk actief in oligodendrocyten, een gliacel-type die de axonen van neuronen ondersteund en beschermt, en komen pas na de geboorte op. Dit impliceert dat veranderingen in de cellen die de neuronen ondersteunen en niet zo zeer in de neuronen zelf een rol hebben gespeeld bij de evolutie van het menselijk brein. Verrassend genoeg zagen we dat een gedeelte van deze evolutionair nieuwe GREs gedereguleerd zijn in de hersenen van patiënten met autisme en dat veel van deze elementen expressie reguleren van genen die al in eerder onderzoek geïmpliceerd waren met autisme. Deze resultaten suggereren dat er een link is tussen veranderingen in regulatie tijdens evolutie van het menselijke brein en de ontwikkeling voor neuronale aandoeningen. Echter, verder onderzoek is nodig om te achterhalen of de deregulatie van deze GREs een direct effect heeft of dat ze het indirecte gevolg zijn van andere factoren.

In **hoofdstuk 3** hebben we actieve enhancer en promoter elementen in kaart gebracht in verschillende organen van de penseelaap. Door deze te vergelijken met actieve GREs in dezelfde organen van zowel mensen als muizen, ontdekten we dat de meeste elementen geconserveerd zijn tussen de verschillende soorten. Echter, we identificeerde een kleine groep enhancers waarvan de activiteit leek te variëren tussen verschillende organen. Door verder onderzoek te doen op deze elementen, ontdekten we dat evolutionair nieuwe GREs vaker specifiek waren voor één enkel celtype en dat hun activiteit vaak variabel was tussen verschillende individuen. Deze observatie heeft implicaties voor andere onderzoeken naar GREs,

omdat afwezigheid van enhancer activiteit verklaard zou kunnen worden doordat deze variabel is binnen een populatie. Bovendien zagen we dat deze nieuwe GREs vaker in de meer inactieve regio's van het genoom lagen en dat ze minder vaak in contact staan met promotoren of andere GREs. De variabiliteit van deze enhancers zou daarom verklaard kunnen worden doordat ze nog niet geïntegreerd zijn in het genregulatie netwerk. Verder onderzoek is echter nodig om te achterhalen of deze elementen daadwerkelijk alleenstaand zijn of dat ze in andere omstandigheden of bij specifieke signalen wel genexpressie reguleren. Daarnaast observeerden we dat DNA-mutaties, geassocieerd met ziekte of eigenschappen zoals lengte en intelligentie, vaker voorkomen in evolutionair geconserveerde GREs. Dit suggereert dat de veranderingen in enhancer activiteit tussen soorten geen invloed heeft op de bruikbaarheid van model organismes zoals de penseelaap voor onderzoeken naar het effect van DNA-mutaties op genregulatie en het geassocieerde fenotype.

In **hoofdstuk 4** analyseren we de eerder in hoofdstuk 2 geïdentificeerde hominini-specifieke enhancers verder. Door celkernen uit de prefrontale cortex van de diverse primaten te isoleren en deze te verdelen in neuronale en niet-neuronale celtypen waren we in staat om celtipe specifieke genexpressie profielen te verkrijgen. Analyse van de genen gelinkt aan de hominini-specifieke GREs liet zien dat ook deze genen specifiek tot expressie kwamen in oligodendrocyten. Verrassend genoeg vonden we dat het DNA van veel van deze GREs al in een open configuratie was in resusapen, zonder dat we activiteit op basis van H3K27ac zagen. Dit kan impliceren dat deze elementen al waren geprepareerd om een enhancer te worden zonder genexpressie te beïnvloeden of dat deze elementen een andere regulerende functie hadden in resusapen. Verdere genexpressie analyse in individuele cellen onthulde dat de genen gelinkt aan de hominini-specifieke GREs niet specifiek zijn voor een bepaalde oligodendrocyten subtype en dat er geen verschil in verhouding van de subtypes is tussen mensen en resusapen is. Dit suggereert dat de veranderingen in de geobserveerde enhancer activiteit niet het resultaat zijn van veranderingen in een specifieke set cellen of door variatie in de hoeveelheid subtypen. Desondanks zijn de onderliggende genetische veranderingen, die de activiteit van deze hominini-specifieke GREs hebben veroorzaakt, nog niet bekend. Verder onderzoek is dus nodig om te ontdekken welke factoren bij hebben gedragen aan het ontstaan van deze elementen.

Samengevat draagt het onderzoek beschreven in deze thesis bij aan het beter begrijpen van hoe de regulatie van genexpressie is geëvolueerd van primaten tot mensen. Daarbij laten we zien dat er een potentiële correlatie zit tussen de evolutie van onze hersenen en de bevattelijkheid voor neuronale aandoeningen. Hierin lijkt een onverwachte rol weggelegd voor gliacellen en dan oligodendrocyten in het bijzonder. Hoewel veel onderscheid in gen-regulerende-elementen in kaart zijn gebracht, blijft er nog een hoop werk te verrichten om te begrijpen hoe deze variaties precies hebben bijgedragen aan de evolutie van de mens en wat de functionele effecten van deze veranderingen precies zijn.

Acknowledgements

“Don’t it always seem to go, that you don’t know what you’ve got till it’s gone”. When I started writing the acknowledgements, we were all in home isolation due to the corona crisis. It is a different time than I had thought would be the last couple of months in the Hubrecht. One day everything is fine and you’re at work writing your thesis, the next day you are forced to stay home and miss the occasional chat at the coffee machine. Now more than ever, I realize what an amazing the place the Hubrecht is and that this thesis would not have seen the light of day without the help of many of its wonderful people. I would therefore like to take this opportunity to say a few words of thanks.

Allereerst, **Menno**, na een master stage in jouw lab mocht ik blijven hangen voor een promotie. Dank je wel voor de mogelijkheid om mijn PhD in jouw groep te doen, de vrijheid die je gaf in zowel de projecten als de tijd om me te richten op de bioinformatica kant. Hoewel de projecten niet altijd even soepel liepen bleef je enthousiast en hebben we er een paar mooie verhalen uit weten te slepen. Naast alles over de wetenschap heb ik afgelopen jaren ook veel kunnen leren over schrijven, presenteren en het publicatie spel. Bedankt voor de toffe projecten over evolutie, de discussies en de kunst om overal een goed verhaal in te zien!

Natuurlijk wil ik ook de rest van de **Creyghton** groep bedanken. **Maartje**, dit avontuur is bij jou begonnen toen ik als master student mocht komen helpen aan het project naar de evolutie van enhancers in macaques. Jouw enthousiasme en toewijding voor de wetenschap zijn een voorbeeld voor iedereen en ik heb dan ook grote bewondering over de manier waarop jij te werk gaat en je gedrevenheid. Ik heb echt een toptijd gehad, zowel als master student als later als collega PhD, bedankt voor alle lol in het kantoor maar ook tijdens borrels/weekenden weg en festivals. **Caro**, we zijn onze PhDs ongeveer tegelijk begonnen en hebben dan ook de ups en downs samen meegemaakt. Bedankt voor alle gezelligheid op kantoor en daarbuiten. Het was mooi om te zien hoe jouw gedrevenheid op zowel sportief als wetenschappelijk vlak soms omsloeg in blind fanatisme zodra er iets te winnen viel. Heel veel plezier in Boston! Ik ga onze gesprekken over sport, wie is de mol, expeditie robinson en alle andere onzin missen maar zal je avonturen op Strava zeker blijven volgen. **Ilia**, thank you for being part of the Creyghton gang and bringing some Italian culture to our office! It was great seeing you grow from a master student to a driven PhD and I’m glad to have been part of that. Last few months we were the sole survivors of the Creyghton branch at the Hubrecht and I really enjoyed working with you. I hope you will soon settle at your new spot in the Stratum, keep up the positive vibes and you will have a great PhD! **Mirna**, bedankt voor de 2 jaren nuchterheid die je ons lab als bio-informatica bracht, je was een aanwinst voor de groep. Ik hoop dat je blij bent met het resultaat van het enhancer variabiliteit verhaal,

dit was toch een beetje jouw kindje en zonder jouw start had ik hier nu niet gestaan. **Sander**, helaas maar eventjes samen gewerkt als PhD voordat je de overstap maakte naar The Hyve. Het waren alsnog mooie jaren waarin ik veel van je heb mogen leren op bio-infomatica vlak maar waarin we ook veel lol hebben gehad op het kantoor. Ben soms zeker wel jaloers geweest op jouw durf om na 2 jaar iets anders te gaan doen! Daarnaast wil ik natuurlijk de andere (oud)-groepsleden niet vergeten. **Peter, Charles, Djawad** en mijn andere studenten **Elise en Jessica**, ook zonder jullie had dit boekje het daglicht niet gezien.

Naast directe collega's had ik deze jaren niet overleefd zonder **Wouter** en **Wim**. Heren, tof dat jullie mijn paranifmen willen zijn tijdens mijn promotie. Ik voel me vereerd om jullie support te hebben die dag. **Wouter**, wij kennen elkaar zowat vanaf dag 1 van onze Hubrecht carrière en hebben dan ook al een hoop dingen meegemaakt. Niet alleen ben je een onmisbare lunchbuddy maar ook het jaar samen in de PV en de vele borrels, spelletjesavonden en Ardennen weekendjes waren fantastisch. Daarnaast ben jij degene (met hulp van **Caro**) die me heeft aangestoken met het wielrenvirus. Laten we binnenkort weer een keertje onze beloofde rondje samen gaan fietsen! Sorry voor alle grapjes als je weer een keer de vrijdagmiddagborrel of uitje miste vanwege je papadag of vanwege je keuze om naar Lent (waar???) te verhuizen, hoop dat je het me kan vergeven. **Wim**, wij gaan zelfs nog verder terug. Mijn allereerste stapjes in de Utrechtse wetenschapswereld waren samen met jou als studenten in het Scheres lab. Gelukkig hebben we sindsdien contact gehouden, niet alleen op congressen en retreats maar vooral buiten de wetenschap tijdens borrels, carnaval, feestjes, spelletjesavonden en natuurlijk niet te vergeten de biweekly pubquiz met burger vooraf. Van the flying dutchmen tot de koala-fight quizzers, we hebben heel wat uren doorgebracht in de florin. Daarnaast was het fantastisch om eens per jaar zelf uit te pakken met een quiz op het Ardennen weekend. Nog steeds is de eerste quiz die we met **Wouter** organiseerden de meest legendarische, al was het maar vanwege de gewilde loser-prize. Omdat het helaas toch niet gelukt is om op dezelfde dag te promoveren moeten we het behalen van onze PhDs maar dubbel vieren!

There are of course a lot more people that I cannot forget to mention. Allereerst, de **Knipscheertjes**, met wie we afgelopen jaren lief en leed hebben gedeeld tijdens de lunch en de vele vrijmibo's. **Roxanne, Alice, Merlijn, Aiko, Jamie**, en oud-leden **Daisy, Nerea** en **Rick**. Thank you all for the fun and laughter during my time at the Hubrecht. It was a joy to pick you up for lunch every day at 12pm sharp! I hope you will survive without us, as you will need to keep track of time by yourselves from now on, I will miss our talks about massaging frogs and chopping balls over a brie sandwich filled with lemon.

During my first year as PhD I had the pleasure of joining the PV. **Wouter**,

Annabel, Joep, Sander (tof dat jullie ook voor Houten gekozen hebben!), **Euclides, Kim** en **Daniël**. Thank you for the fun we had in organizing all those amazing borrels, BBQ with mechanical bull, LabStapDag, Christmas dinner and PV-weekend, they were all a great success! I think that the still strong tradition of a dedicated PV-suit is maybe our biggest contribution to the Hubrecht. **Annabel**, running-buddy, dank je wel voor alle vrolijke momenten op het lab en daarbuiten! Ik ben blij dat ik jou heb leren kennen. Naast de PV, pubquizen en spelletjesavonden hebben we een mooie traditie van ochtendloopjes voor werk. Deze zijn zowel ontspannend als therapeutisch en hebben vooral in onze laatste jaar geholpen om de moed erin te houden. Wat mij betreft houden we deze dan ook zeker in stand, ondanks dat we allebei ergens anders zijn gaan werken. Doe **Willem** de groetjes! **Joep**, squashmaatje, dank je wel voor al het bloed, zweet en tranen op de squashbaan. Hoewel we niet altijd even goed waren was het heerlijk om eens per week te ontladen. Sorry nog voor alle blauwe plekken door ongerichte ballen!

Zo in thuis isolatie moet ik misschien wel een van de belangrijkste rituelen op werk missen, de wandeling naar en gesprekken bij de koffieautomaat. Ik wil dan ook hoop mensen van de derde verdieping van het Hubrecht bedanken voor alle gezellige gesprekken hier. Allereerst de **Korswagen** en **Galli** groep. **Marco**, je bent onmisbaar voor het reilen en zeilen van de 3^e verdieping. Niet alleen bedankt voor alle hulp op het lab maar vooral voor alle goede gesprekken en je positieve kijk op het leven in science en de vrolijke “goedemorgen” groet elke vroege ochtend. Heel veel succes met je nieuwe baan in het UMC. **Erik** (Ardennen opperhoofd), **Jonas, Christa** (sorry nog voor de lading pipetpuntjes die ‘per ongeluk’ op jou bench terecht kwamen), **Annabel, Euclides, Lotte** (misschien woensdag een keertje pubquizen?), **Gaby** (Thanks for all the baked cakes that you didn’t want to eat yourself), **Lorenzo** en **Reinier**, ondanks de soms lange wachtrijen voor koffie na de lunch, dank jullie dat ik altijd bij jullie kon aansluiten voor een praatje. Ook kan ik de **Geijsen** groep natuurlijk niet vergeten; **Nune** (bedankt voor het leren van stamcel kweek. Veel succes met je eigen PhD in het AvO lab!), **Ator, Melissa, Peng, Jorik** en **Fanny**, bedankt voor de gezelligheid in het ML2 lab en jullie hulp bij al mijn celkweek pogingen.

A big thanks also goes out to all the amazing people I had the pleasure of meeting during my years in the Hubrecht. **Geert**, bedankt voor al je hulp in de wereld van bioinformatica, zonder jou en **Mirna** was het 3^e hoofdstuk nooit van de grond gekomen. Daarnaast denk ik met plezier terug aan alle borrels en Ardennen weekenden waarin we naast een biertje konden genieten van jouw ultieme karaoke skills. **Tim**, door jou heb ik leren genieten van Nina’s grote gyros schotel, mijn dank is groot! During my PhD I spend a lot of Wednesday nights at the Florin for the pubquiz, and while our name changes every year; from The Failing Dutchmen, through the Worst-Kaas-Scenario to the best Koala-Fight Quizzers, we have a strong

team that keeps coming back. **Javi**, it's really nice to see that you still try to make it every time. Thanks for all the fun over the years and your somewhat cynical view on our team results. **Kim** (your enthusiasm and dedication are unmatched by all!), **Lars** (our football encyclopedia, a shame we have to miss you when Ajax is playing), **Marit** (for all those top 40 song), **Jeff**, **Wim**, **Annabel**, **Margit**, **Irene**, **Eduard**, **Reinier**, **Freek**, **Phillip**, and all other people that occasionally joined in, thank you for all the fun during these impossible quizzes. Hopefully we can soon enjoy the Wednesday night quizzes again, I will definitely try to keep on joining for both the quiz and burger.

But it doesn't stop there! **Axel**, **Wim**, **Guy**, **Erik**, **Mike**, **Louk**, **Jens**, **Freek**, **Nico**, **Enric**, **Jan**, **Lucas**, **Margit** en **Lars**, it was a lot of fun to join in the football team last year, even with my clear lack of football skills! But also to all others at the Hubrecht, UMC and PMC, who made the Ardennes weekends, Retreats, Masterclasses, Super-Group-Meetings and of course the regular Friday afternoon borrels such a success! In any order and not limited to; rode-wijn-**Bas** (al overtuigd van onze pieken?), **AJ** (Beer commander in rest), **Charlotte** (quiz-girl!), **Pieterjan**, **Rob**, **Erica**, **Niels**, **Ajit**, **Hesther**, **Sasja**, **Mark**, **Sven**, **Deepak**, **Stijn**, **Clement**, **Wouter T**, **Wessel**, **Mauro**, **Chloe**, **Judith**, **Anne**, **Pim**, **Bas dW**, **Bram**, **Colinda**, **Silke**, **Colinda**, **Lotte**, **Arianna**, **Brian**, **Abel**, **Lennart**, **Koen** (passen jullie goed op ons oude kantoortje), **Jimmy**, **Peter**, **Evelyn**, **Michael**, **Rijndert**, **Saman** (good to see you again at PwC), **Spiros**, **Naomi**, **Roy**, **Elke** and many others. Without you the Hubrecht would have been a much more boring place, THANKS!

Hoewel ik vele uren op het Hubrecht heb rondgebracht, was er gelukkig tijd genoeg om te genieten van het leven buiten het werk. Daarom wil ik naast collega's graag ook een paar mensen bedanken die ervoor gezorgd hebben dat mijn leven niet alleen om wetenschap draaide. Ik ben zo'n 8 jaar geleden vanuit Brabant naar Utrecht verhuisd en ben in die tijd netjes opgevangen door de volgende mensen. Allereerst natuurlijk mijn lieve oud-huisgenootjes, **Marcel**, **Marcella**, **Nathalie**, **Arthur**, **Just**, **Marieke** en **Thomas** (hoewel we officieel nooit hebben samengewoond horen jullie hier uiteraard bij!), **Bas**, **Jasper**, **Lotte**, **Sanne**, **Dirk**, **Marije**, **Roos**. In totaal heb ik zo'n 3 jaar op de Warande 180 gewoond, dank jullie wel voor alle toffe huisfeesten, avondjes spelletjes spelen met bier drinken, BBQs op het balkon, huisuitjes en gewoon de gezelligheid. Waardoor ik telkens weer de stress en frustraties van een PhD-life kon vergeten. Daarnaast kan ik **Jasper**, **Niels**, **Lars**, **Tim**, **Lisa**, **Johanna**, **Mitch**, **Femke** en **Sterre** natuurlijk niet vergeten, ben blij dat ik jullie via **Margit** heb leren kennen. De weekendjes weg met jullie naar Dublin (of was het toch Cork), Krakau en Ljubljana waren onvergetelijk! En fantastisch dat we zoveel passies delen, zoals het bijhouden van Untapped onder het genot van een speciaal biertje en altijd de moeite willen doen om overal veel te slechte woordgrappen voor te verzinnen. Naast squash en voetbal heb ik de laatste jaren het plezier in hardlopen weer

teruggevonden en heb ik al vele malen een halve marathon tot een goed einde mogen brengen. **Jody, Eddy, Daphne, Joris, Irina, Jora, Daan, Jet**, dank jullie wel voor al het zweet tijdens **Dirks** donderdagavond trainingen, de (onderlinge) wedstrijden en de motivatie om steeds wat harder te lopen, zonder jullie had ik het niet gekund!

Deze hardloopsessie door de acknowledgement brengt me misschien wel tot de belangrijkste groep mensen die mij met beide benen op de grond houden. **Fabian, Ilse, Bart, Nienke, Freek, Laura, Jochem, Maartje** en natuurlijk **Ivar, Merle, Bodile, Pien** en **Isa**. Jullie zijn degene die me elke keer weer terug in Brabant brengen. Of het nu voor verjaardagen, weekendjes weg, rondjes fietsen of gewoon gezellig een drankje doen is bij jullie voel ik me altijd welkom. Wat begon met gezamenlijke trainingen bij Scorpio is uitgegroeid tot een hechte vriendschap en ik ben blij dat ik hier nog altijd deel van mag zijn, ondanks mijn verhuizing naar boven de rivieren. Ik hoop dat met dit boekje het ook een beetje duidelijk is geworden wat ik de afgelopen 6 jaar heb uitgespookt in Utrecht.

Han, Ans, Willemijn, Jeffrey, Anouk (Toek) en **Tomas**, bedankt dat ik deel uit mag maken van jullie familie en voor alle interesse in mijn werk, de steun en hulp (bij het verhuizen, als we weer in een keer in Frankrijk stranden op vakantie of wat dan ook) maar ook voorde leuke weekendjes weg en het feit dat de deur altijd open staat.

Dat brengt me bij mijn eigen familie, **Pap, Mam, Mark en Jeroen**, bedankt voor al het vertrouwen en het geduld. Hoewel mijn loopbaan als student zeker niet altijd even soepel liep is het uiteindelijk toch goed gekomen en ben ik blij dat ik alles met jullie support heb kunnen doen. Van stage lopen in Valencia tot het opnieuw beginnen aan een studie. En hoewel we allemaal uitgevlogen zijn ondertussen is het altijd fijn om weer terug op het nest in Oosterhout te zijn, voor een spelletje, hapje eten of borrelen tot in de kleine uurtjes. Zonder jullie had ik hier niet gestaan!

Voor het schrijven echt af is wil ik één iemand nog wel het meest bedanken. Liefste **Margit**, bedankt voor alles de afgelopen jaren! We hebben elkaar op het Hubrecht leren kennen, hebben al vele avonturen mee gemaakt en passen perfect bij elkaar. Van stedentrip in Amerika, relaxed kamperen in Luxemburg en rondjes toeren in de Kever tot kipjes houden in de eigen achtertuin, ik kan me geen beter maatje wensen om dit mee te doen. Ik ben blij dat ik met jou mag vieren dat mijn PhD is afgerond en ik ga vol vertrouwen mee op jouw eigen PhD avontuur. Ik hou van jou!

List of publications

Bas Castelijns*, Mirna L. Baak*, Geert Geeven*, Marit W. Vermunt, Caroline R.M. Wiggers, Ilia S. Timpanaro, Ivanela Kondova, Wouter de Laat, Menno P. Creyghton. Recently evolved enhancers emerge with high interindividual variability and less frequently associate with disease. *Cell Reports*. 2020 Jun 23;31(12):107799

Bas Castelijns, Mirna L. Baak, Ilia S. Timpanaro, Caroline R.M. Wiggers, Marit W. Vermunt, Peng Shang, Ivanela Kondova, Geert Geeven, Valerio Bianchi, Wouter de Laat, Niels Geijssen, Menno P. Creyghton. Hominin-specific regulatory elements selectively emerged in oligodendrocytes and are disrupted in autism patients. *Nature Communications*. 2020 Jan 16;11(1):301

Marit W Vermunt, Sander C. Tan, **Bas Castelijns**, Geert Geeven, Peter Reinink, Ewart de Bruijn, Ivanela Kondova, Stephan Persengiev, Netherlands Brain Bank, Ronald Bontrop, Edwin Cuppen, Wouter de Laat, Menno P. Creyghton. Epigenomic annotation of gene regulatory alterations during evolution of the primate brain. *Nature Neuroscience*. 2016 Mar;19(3):494-503

Yuchen Long, Wouter Smet, Alfredo Cruz-Ramírez, **Bas Castelijns**, Wim de Jonge, Ari Pekka Mähönen, Benjamin P. Bouchet, Gabino Sanchez Perez, Anna Akhmanova, Ben Scheres, Ikram Blilou. Arabidopsis BIRD Zinc Finger Proteins Jointly Stabilize Tissue Boundaries by Confining the Cell Fate Regulator SHORT-ROOT and Contributing to Fate Specification. *The Plant Cell*. 2015 Apr;27(4):1185-99

Roy Baas, Daphne Lelieveld, Hetty van Teeffelen, Philip Lijnzaad, **Bas Castelijns**, F.M. van Schaik, Michiel Vermeulen, David A. Egan, H. Th. Marc. Timmers, Petra de Graaf. A Novel Microscopy-Based High-Throughput Screening Method to Identify Proteins That Regulate Global Histone Modification Levels. *Journal of Biomolecular Screening*. 2014 Feb;19(2):287-69

Alejandro Sarrion-Perdigones, Marta Vazquez-Vilar, Jorge Palací, **Bas Castelijns**, Javier Forment, Peio Ziarsolo, José Blanca, Antonia Granell, Diego Orzaez. GoldenBraid 2.0: A Comprehensive DNA Assembly Framework for Plant Synthetic Biology. *Plant Physiology*. 2013 Jul;162(3):1618-31

José Manuel Julve, Antoni Gandía, Asun Fernández-del-Carmen, Alejandro Sarrion-Perdigones, **Bas Castelijns**, Antonio Granell, Diego Orzaez. A coat-independent superinfection exclusion rapidly imposed in *Nicotiana benthamiana* cells by tobacco mosaic virus is not prevented by depletion of the movement protein. *Plant Molecular Biology*. 2013 Apr;81(6):553-64

Curriculum Vitea

Bas Castelijns was born on July 2nd in 1989 in Alkmaar, the Netherlands. In 2007 he obtained his high school degree at the Nassau Scholengemeenschap te Breda, after which he started the Bachelor Chemical Engineering and Chemistry at the Eindhoven University of Technology. After two years he switched to the Bachelor Biology and Medical Laboratory Research at Avans University of Applied Sciences which degree he obtained cum laude in 2012. During his Bachelor he performed an internship abroad in the lab of dr. Diego Orzaez at the Instituto de Biología Molecular y Celular de Plantas in Valencia, Spain, where he studied viral infection rates in tobacco plants. For his second Bachelor internship he joined the lab on prof. dr. Ben Scheres at the Utrecht University, studying transcription factor dynamics in the root of *Arabidopsis thaliana* under the supervision of dr. Ikram Blilou. Bas stayed in Utrecht to obtain a Master's degree in Biomedical Sciences for which he performed an internship in the lab of prof. dr. Marc Timmers at the University Medical Center Utrecht, studying readers and writers of epigenetic modifications. His second internship in the lab of dr. Menno Creyghton at the Hubrecht Institute for Developmental Biology and Stem Cell Research involved research to the epigenetic difference between the brains of human and rhesus macaque. Following the writing of a Master's thesis in the lab of prof. dr. Elly Hol at the University Medical Center Utrecht, on the roles of DNA mutations on the onset of Parkinson's disease, he graduated in 2014. Bas returned to the lab of dr. Menno Creyghton to start a PhD in October 2014 on the role of gene regulatory elements in primate evolution and disease, of which the results are presented in this thesis. In July 2020, Bas joined the international financial service company PricewaterhouseCoopers as consultant.