bioRxiv preprint doi: https://doi.org/10.1101/298950; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Genetic differences between humans and other hominins contribute to the "human condition"

Martin Kuhlwilm[1], Cedric Boeckx[2,3,4]
[1] Universitat Pompeu Fabra, Institute for Evolutionary Biology, Barcelona, Spain
[2] ICREA, Barcelona, Spain
[3] University of Barcelona, Barcelona, Spain
[4] UB Institute of Complex Systems, Barcelona, Spain

Abstract

Throughout the past decade, studying ancient genomes provided unique insights into human prehistory, and differences between modern humans and other branches like Neanderthals can enrich our understanding of the molecular basis of the human condition. Modern human variation and the interactions between different hominin lineages are now well studied, making it reasonable to explore changes that are observed at high frequency in present-day humans, but do not reach fixation. Here, we put forward interpretation of putative single nucleotide changes in recent modern human history, focusing on 571 genes with non-synonymous changes at high frequency. We suggest that molecular mechanisms in cell division and networks affecting cellular features of neurons were prominently modified by these changes. Complex phenotypes in brain growth trajectory and cognitive traits are likely influenced by these networks and other changes presented here. We propose that at least some of these changes contributed to uniquely human traits.

bioRxiv preprint doi: https://doi.org/10.1101/298950; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Homo sapiens appears to be a "very special primate"[1]. Our position among animal species stands out largely thanks to the composite complexity of our cultures, social structures and communication systems. It seems very reasonable that this "human condition" is rooted, at least in part, in the properties of our brain, and that these can be traced to changes in the genome on the modern human lineage. This phenotype in the population called "anatomically modern humans" emerged in Africa likely before the deepest divergence less than 100,000-200,000 years ago[2,3], although complex population structure may reach back up to 300,000 years ago[4–6]. Except of some early dispersals[7], humans most likely peopled other parts of the world than Africa and the Middle East permanently only after around 65,000 years ago. It has been claimed that the brain of modern humans adopted a specific, apomorphic growth trajectory early in life that gave rise to the skull shape difference between modern humans and extinct branches of the genus *Homo*[8], although the timing of this change is debated[9]. This ontogenic trajectory, termed the "globularization phase", might have contributed to our singular cognitive abilities[8,10,11].

We are now in a favorable position to examine the evolution of human biology with the help of the fossil record, in particular thanks to breakthroughs in paleogenomics: The recent reconstruction of the genomes of members of archaic *Homo* populations[12–14] has opened the door to new comparative genomic approaches and molecular analyses. The split of the lineages leading to modern humans and other archaic forms (Neanderthals and Denisovans) is estimated to around 600,000 years ago[2], setting the timeframe for truly modern human-specific changes after this split, but before the divergence of modern human populations (Fig. 1). Together with efforts to explore present-day human diversity[15], this progress has allowed to narrow down the number of candidate point mutations from ~35 million differences since the split from chimpanzee when comparing only reference genomes[16] to 31,389 fixed human-specific changes in a previous seminal study[1].

Some of these changes have been linked to putative functional consequences[1,13,17], and evidence is mounting that several molecular changes affecting gene expression in the brain were subject to selective pressures[18–22]. Furthermore, the genomic impact of interbreeding events is not evenly distributed across the genome. Genes expressed in regions of the brain regarded as critical for certain cognitive functions are depleted in introgressed archaic genetic material[23–26], and introgressed alleles are downregulated in some brain regions, suggesting natural selection acting on tissue-specific gene regulation[27]. Thus, it seems reasonable to conclude that there were differences between anatomically modern human and Neanderthal brains, and that these underlie at least some of the characteristics of our lineage[28]. We want to emphasize that such recent differences are likely to be subtle when compared to those after the split from our closest living relatives on a scale of 6-10 million years [29], where fundamental changes arose since the divergence from chimpanzees and bonobos[30]. The observation of recurrent gene flow between modern human and archaic populations also implies a broad overall similarity, yet, such subtle differences may still have contributed to the evolutionary outcome[31]. Obviously, not all human-specific changes are beneficial: While most mutations may be rather neutral and have little effect on the phenotype, some may have had deleterious effects or side-effects, possibly increasing the risks for neurodevelopmental or neurodegenerative disorders in humans[32–34].

The goal of this paper is to provide a set of recent single nucleotide changes in humans since their split from Neanderthals that could enrich our understanding of the molecular basis of the recent human condition. The previous focus on fixed alleles was reasonable given limited data[1], but having a better grasp of the magnitude of modern human variation and the interaction between different hominin lineages seems a good reason to cast a wider net, and take into account not only fixed differences but also high-frequency changes shared by more than 90% of present-day

bioRxiv preprint doi: https://doi.org/10.1101/298950; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

individuals. Here, we present a revised list of 36 genes that carry missense substitutions which are fixed across 1,000s of human individuals and for which all archaic hominin individuals sequenced so far carry the ancestral state. In total, 647 protein-altering changes in 571 genes reached a frequency of at least 90% in the present-day human population. We attempt to interpret this list, as well as some regulatory changes, since it seems very likely that some of these genes would have contributed to the human condition.

We will discuss some of their known functions, and how these relate to pathways that might have been modified during human evolution (Fig. 1), in a bottom-up fashion. Beginning at the molecular level, changes found in genes associated to the mitotic spindle complex might be relevant, as has been suggested in previous studies[1,13]. The cellular features of neurons (axons and synapses) have been considered important in light of their role in traits such as vocal learning[35–37], and possibly behavioral phenotypes[38]. Pathways influencing brain organization were modified during hominin evolution, and we suggest that this might have been extended further since the split from Neanderthals and other archaics[39]. Finally, we discuss implications for other complex phenotypic traits, with a focus on cognition and life history trajectory. We restrict our attention to genes where the literature may allow firm conclusions and predictions about functional effects, since many genes might have multiple different functions[40]. Obviously, experimental validation will be ultimately needed to confirm our hypotheses concerning alterations in specific functions.

bioRxiv preprint doi: https://doi.org/10.1101/298950; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

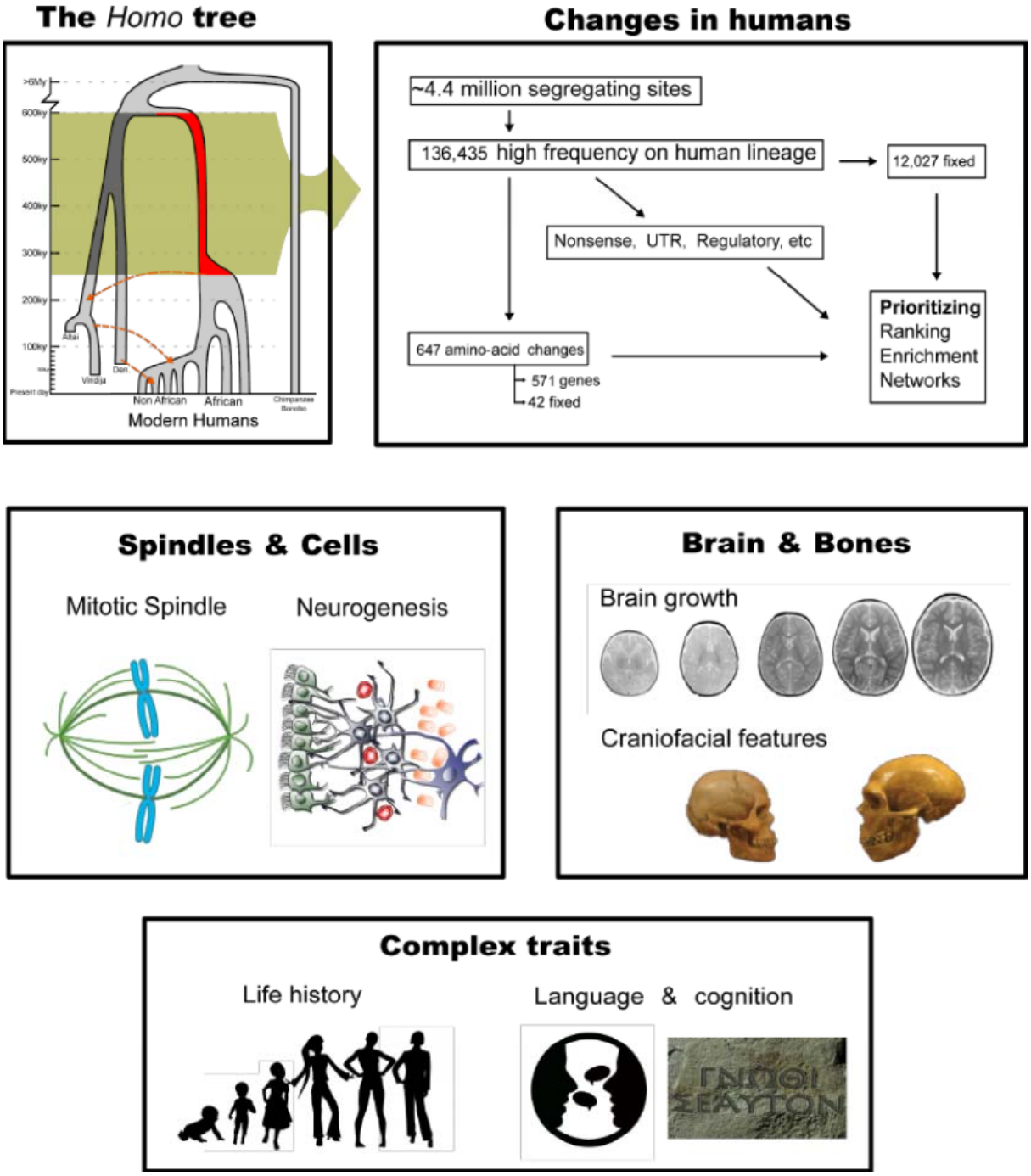

Figure 1: Conceptual summary of this study.

Results

Genetic differences between present-day humans and archaic hominins

Using publicly available data on one Denivosan and two Neanderthal individuals and present-day human variation (Methods), we calculated the numbers of single nucleotide changes (SNCs) which most likely arose recently on the respective lineages after their split from each other, and functional consequences as predicted by VEP (Table 1). Previously, a number of 31,389 sites has been reported as recently fixed derived in present-day humans, while being ancestral in archaics[1,13].

bioRxiv preprint doi: https://doi.org/10.1101/298950; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

We find a smaller number of only 12,027 positions in the genome, in part caused by including another archaic individual and different filters, but mainly by a richer picture of present-day human variation. The 1,000 Genomes Project as well as other sources contributing to the dbSNP database now provide data for thousands of individuals, which results in very high allele frequencies for many loci instead of fixation. Indeed, 29,358 positions show allele frequencies larger than 0.995, demonstrating that the level of near-fixation is similar to the level of previously presented fixation. The number of loci with high frequency (HF) changes of more than 90% in present-day humans is an order of magnitude larger than the number of fixed differences. The three archaic individuals carry more than twice as many changes than present-day humans; however, we emphasize that much of this difference is not due to more mutations in archaics, but rather the fact that data for only three individuals is available, compared to thousands of humans. The variation across the archaic population is not represented equally well, which makes these numbers not directly comparable.

Present-day humans carry 42 fixed amino acid-changes in 36 genes (Table 2, Fig. 2), while Neanderthals carry 159 such changes. Additionally, modern humans carry 605 amino acid-changes at high frequency (human-lineage high-frequency missense changes, referred to as HHMCs), amounting to a total of 647 such changes in 571 genes (Table S1). Together with 323 SNCs on the human lineage with low confidence (Methods, Table S2), almost 1,000 putative protein-altering changes were found across most present-day humans. Generally, synonymous changes are found at a similar magnitude as missense changes, but only few SNCs altering start and stop codons, and thousands of changes in putative regulatory and untranslated regions. We admit that some of the loci presented here are variable across the phylogenetic tree, or less reliable due to low coverage in the archaics, but we accept this since our intention is retrieve an inclusive picture of possibly functional recent changes. The 42 protein-altering changes for which the ancestral allele has not been observed in any present-day human, most of which have been presented before[1], constitute without doubt the strongest candidates for understanding the human condition. Only one gene, *SPAG5*, carries three such SNCs, and four genes (*ADAM18, CASC5, SSH2* and *ZNHIT2*) carry two fixed protein-coding changes in all modern humans. We identified 15 SNCs (in *AHR, BOD1L1, C1orf159, C3, DNHD1, DNMT3L, FRMD8, OTUD5, PROM2, SHROOM4, SIX5, SSH2, TBC1D3, ZNF106, ZNHIT2*) that have not been previously described as fixed differences between humans and archaics. We note that another 12 previously described[1] protein-altering substitutions were not found among the genotypes analyzed here (in *C21orf62, DHX29, FAM149B1, FRRS1L, GPT, GSR, HERC5, IFI44L, KLF14, PLAC1L, PTCD2, SCAF11*). These genotype calls are absent from the files provided for the three archaic genomes due to different genotype calling and filtering procedures compared to the original publication of the Altai Neanderthal genome[13,14]. Hence, some potentially relevant candidate changes were not included here, and future research is necessary to evaluate these as well. Despite attempting an extended interpretation, our data is thus still not fully exhaustive.

It is noteworthy that the number of fixed SNCs decreased substantially, and it is possible that single individuals will be found to carry some of the ancestral alleles. Hence, it is important to focus not only on fixed differences, but also consider variants at high frequency. When analyzing the 647 HHMCs, 68 genes carry more than one amino acid-altering change. Among these, *TSGA10IP* (Testis Specific 10 Interacting Protein) and *ABCC12* (ATP Binding Cassette Subfamily C Member 12) carry four such changes, and seven more genes (*MUC5B, NPAP1, OR10AG1, OR5M9, PIGZ, SLX4, VCAN*) carry three HHMCs. 1,542 genes carry at least one HF missense change on the archaic lineage (archaic-lineage high-frequency missense change, referred to as AHMC, Tables S3,

bioRxiv preprint doi: https://doi.org/10.1101/298950; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

S4). We find an overlap of 122 genes with HHMCs and AHMCs, which is more than expected considering that among 1,000 sets of random genes of a similar length distribution, no overlap of this extent was observed. The same genes seem to have acquired missense changes on both lineages since their divergence more often than expected. We find a high ratio of HHMCs over synonymous changes for chromosome 21 (1.75-fold), and a very small ratio (0.18-fold) for chromosome 13. We do not find such extreme ratios for AHMCs and corresponding synonymous changes, suggesting differences in the distribution of amino acid changes between both lineages (Fig. S1).

	Fixed human	HF human	Extended human	Fixed archaic	HF archaic	Extended archaic
All	12,027	136,435	83,254	33,498	380,756	983
Non-synonymous	42	647	327	167	1,921	13
Synonymous	41	843	363	193	2,123	14
Start/stop	1	14	10	3	48	2
Splice site	4	23	8	4	54	0
TFBS	28	226	126	87	914	1
Upstream	1,935	19,599	11,235	4,920	55,188	289
5' UTR	180	1,853	1,012	195	2,016	7
3' UTR	77	702	334	506	5,303	19
Downstream	1,922	19,704	11,673	4,956	55,832	281
miRNA	0	1	2	0	4	0
Regulatory element	1,952	20,971	12,320	5,125	59,248	195

Table 1: Summary of single nucleotide changes. TFBS: Transcription factor binding sites. UTR: Untranslated Region. HF: High frequency. Fixed changes are a subset of HF changes.

Ranking and enrichment

We assessed the impact of mutations for different deleteriousness scores (Table 2), finding 12 genes with deleterious HHMCs according to SIFT, three according to PolyPhen, and 16 when using the Grantham score (>180), measuring the physical properties of amino acid changes. The C-score and GWAVA can be used to rank all mutation classes, and we present the top candidates.

Then, we attempted a ranking of genes by the density of lineage-specific changes in the dataset. As expected, the total number of segregating sites is correlated with gene length (Pearsons' R = 0.93). This correlation is weaker for HF human SNCs (R = 0.73) and fixed human-specific SNCs (R = 0.25), as well as for fixed (R = 0.37) and HF (R = 0.82) SNCs in archaics. We conclude that some genes with a large number of human-specific changes might carry these large numbers by chance, while others are depleted. Indeed, 17,453 (88.9%) of these genes do not carry any fixed human-specific change, and 80.5% do not carry fixed archaic-specific changes. Of note, genes that have attracted attention in the context of traits related to the "human condition" like *CNTNAP2* and *AUTS2* are among the longest genes in the genome, hence changes in these genes should be interpreted with caution as they are not unexpected. We ranked the genes by the number of HF changes in either modern humans or archaics, divided by their genomic lengths, and categorize the top 5% of this distribution as putatively enriched for changes on each lineage (Table S5). We note that 191 genes (30.9%) fall within this category for both human HF changes and archaic HF changes, as a result of differences in mutation density. In order to distinguish a truly lineage-specific enrichment, we calculated the ratios of HF changes for humans and archaics, defining the

bioRxiv preprint doi: https://doi.org/10.1101/298950; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

top 10% of genes in this distribution as putatively enriched (Table S5). Among the genes enriched for changes on the modern human lineage, 18 carry no HF changes on the archaic lineage, and ten of these also fall within the 5% of genes carrying many changes considering their length (*ARSJ, CLUAP1, COL20A1, EPPIN, KLHL31, MKNK1, PALMD, RIC3, TDRD7, UBE2H*). These might be candidates for an accumulation of changes, even though this is not identical to selective sweep signals. Among these, the collagen COL20A1 and the Epididymal Peptidase Inhibitor EPPIN carry HHMCs. *ACAD10, DST* and *TTC40*, which carry two HHMCs, might be other notable genes with a human-specific enrichment.

No Gene Ontology (GO) categories are enriched for HHMCs on the human lineage when using HF synonymous changes as background in a hypergeometric test. This is also true for genes carrying AHMCs, or HF changes in UTRs or transcription factor binding sites on either lineage. We applied a test for the ratio of the number of gene-wise HF changes on one lineage over the other lineage. For changes on the modern human lineage, this yields an enrichment for 12 GO categories (Table S6), with "soft palate development", "negative regulation of adenylate cyclase activity", "collagen catabolic process" and "cell adhesion" in the biological process category. Among the cellular components category, the "postsynaptic membrane", "spermatoproteasome complex", "collagen trimer", "dendrite" and "cell junction" show enrichment, as well as the molecular functions "calcium ion binding", "histone methyltransferase activity (H3-K27 specific)" and "metallopeptidase activity". We find no GO enrichment for genes with an excess of changes on the archaic lineage. In order to approach a deeper exploration of genes with associated complex traits in humans, we explored the NHGRI-EBI GWAS Catalog[41], containing 2,385 traits. We performed a systematic enrichment screen, finding 17 unique traits enriched for genes with HHMCs, and 11 for genes with AHMCs (Table S7). Changes in genes associated to "Cognitive decline (age-related)", "Rheumatoid arthritis" or "Major depressive disorder" might point to pathways that could have been influenced by protein-coding changes on the human lineage. In archaics, genes are enriched, among others, for associations to traits related to body mass index or cholesterol levels, which might reflect differences in their physiology. We also find an enrichment of genes associated to behavioral disorders on the archaic lineage.

We find a significant enrichment of protein-protein interactions ($P = 0.006$) among the gene products of HHMC genes (Fig. S2), meaning that these proteins interact with each other more than expected. Functional enrichment is found for the biological process "cellular component assembly involved in morphogenesis", most strongly for the cellular components cytoskeleton and microtubule, as well as the molecular function "cytoskeletal protein binding". Three proteins have at least 20 interactions in this network and might be considered important nodes: TOP2A, PRDM10 and AVPR2 (Table S8). However, proteins encoded by genes with synonymous changes on the modern human lineage seem to be enriched for interactions as well ($P = 0.003$), as are proteins encoded by genes with AHMCs ($P = 1.68 \times 10^{-14}$), with an enrichment in GO categories related to the extracellular matrix and the cytoskeleton, and the most interacting proteins with more than 40 interactions being GART, LRGUK, ARRB1, SPTAN1 and ATM (Table S8). We caution that these networks might be biased due to more mutations and possibly more interactions in longer, multi-domain genes.

Regulatory changes might have been important during our evolution[42], hence we tested for an overrepresentation of transcription factors (TFs). We find 78 known or putative TFs among the HHMC genes (Table S9) on the modern human lineage[43], which is not overrepresented among genes with HHMCs (with 49.2% of random genes sets containing fewer HHMCs). Despite lack of enrichment, single TFs on the modern human lineage might have been important, particularly those

bioRxiv preprint doi: https://doi.org/10.1101/298950; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

with an excess of modern human over archaic HF changes (*AHR*, *MACC1*, *PRDM2*, *TCF3*, *ZNF420*, *ZNF516*). Others TFs, like *RB1CC1*[13] or *PRDM10* and *NCOA6*[19] have been found in selective sweep screens, suggesting contributions of individual TFs, rather than TFs as a class. We also tested for an enrichment of gene expression in different brain regions and developmental stages[44,45], using the HF synonymous changes on each lineage as background sets. We find an enrichment of gene expression in the orbital frontal cortex at infant age (0-2 years) for genes with HHMCs, but no enrichment for genes with AHMCs. Furthermore, when testing the genes with HHMCs and using the set of genes with AHMCs as background, "gray matter of forebrain" at adolescent age (12-19 years) is enriched, while no enrichment was found for genes with AHMCs.

Fixed HHMCs	*ADAM18, ADSL, AHR, ANKMY1, ANKRD30A, BBIP1, BOD1L1, C1orf159, C3, CASC5, CDH16, DCHS1, DNHD1, DNMT3L, FRMD8, GBP5, GLDC, GREB1L, GRM6, KIF26B, LMNB2, NCOA6, NOTO, OTUD5, PRDM10, PROM2, RFNG, SCAP, SHROOM4, SIX5, SPAG5, SSH2, TBC1D3, ZNF106, ZNF185, ZNHIT2*
Selection 2014	*C11orf80, CKAP5, GREB1L, HMCN1, NLRX1, PDZD3, PRDM2, RB1CC1*
Selection 2015	*MSS51, NCOA6, OMD, SPAG17, SPAG5*
Selection 2016	*ACE, ADSL, ALMS1, ANKRD30A, BZRAP1, DNAH1, GREB1L, KMT2C, NWD1, PROM2, RASA1, STAB1, STARD9, ZNF106*
Selection 2017	*ADSL, AKAP8, BAP1, BBIP1, BCAR3, CAPN5, CR2, CSMD2, DNAH1, ENTHD1, FAAH, FRMD8, GBP5, GBP7, GPR157, GTF3C5, HERC5, HERC6, HMCN1, HRASLS5, KATNA1, KIF15, KIF18A, LYST, MKL1, MYH3, NAALADL1, NCOA6, PRDM10, PRDM2, PROM2, PTPRC, RNF44, SCAP, SLC12A8, SLC25A45, SLITRK1, TIGD3, TMEM235, TRGV4, TTC6, VOPP1, ZNF501, ZNF502, ZNHIT2*
Grantham	*ABHD14A-ACY1, ACY1, ABHD14A, CCDC158, CCDC30, DNHD1, EML2, ERI1, GBA3, GREB1, OR1K1, TTC6, UBQLN3, UIMC1, ZBP1, ZNF510, ZNHIT2*
SIFT	*BEND2, CCT6B, COPA, CUL4B, GBP7, KRTAP10-10, MEPE, NHEJ1, OR1K1, SLC6A15, TPO, ZNF510*
PolyPhen-2	*FSHR, NLN, TPO*
CADD	*C11orf80, C5orf66, CCT6B, CDH15, CEP128, CPM, FGF21, FMN2, FUT1, H2AFY, HERC6, KCNK5, KPNA4, KRT33A, KRT8P12, MUM1, NR1H2, OPRM1, PDSS2, ROCK1, RPS15P9, SLC22A31, SUCLG2P4, TMPRSS7, UNC5D*
GWAVA	*ANK2, COPA, CTRC, CYP2B6, MAPK10, MCTP1, SLC38A6, SYT1, YTHDC1*

Table 2: Genes with fixed non-synonymous changes on the human lineage, genes under positive selection with HHMCs, and deleterious candidate HHMCs. Selection 2014: Prüfer *et al.*, 2014. Selection 2015: Zhou *et al.*, 2015. Selection 2016: Racimo, 2016. Selection 2017: Peyrégne *et al.*, 2017.

bioRxiv preprint doi: https://doi.org/10.1101/298950; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

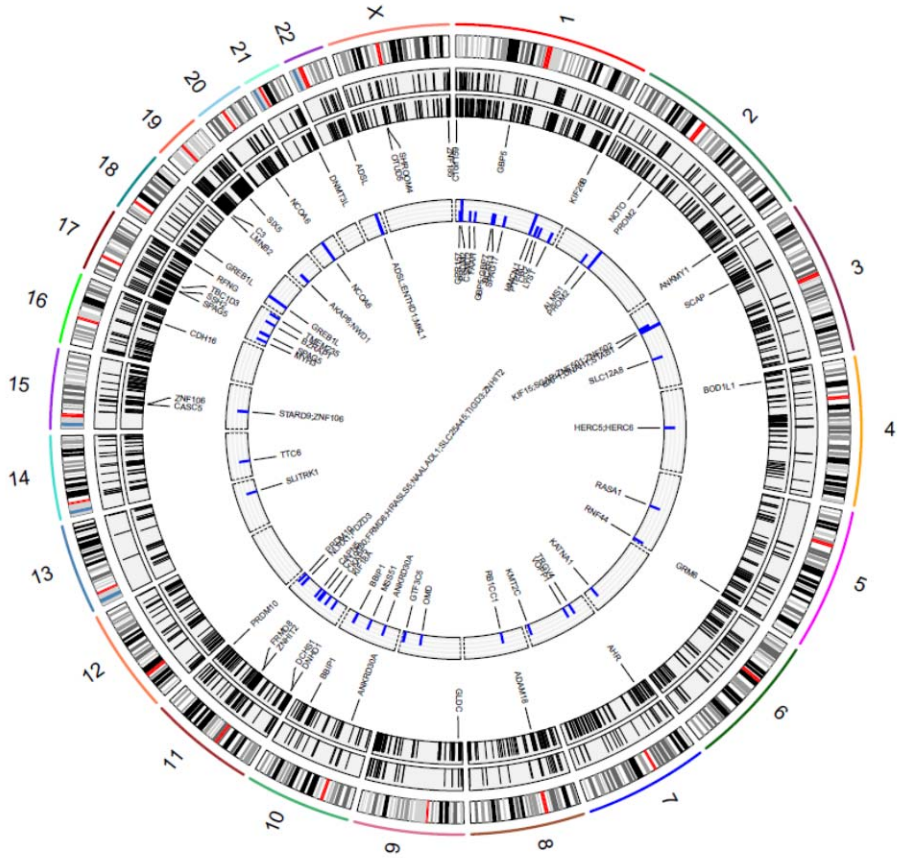

Figure 2: Features discussed in this study. From inside to outside: Genes with HHMCs and signatures of positive selection (compare Table 2), genes with fixed non-synonymous SNCs on the human lineage, HHMCs, AHMCs, karyogram of human chromosomes.

Discussion

The kinetochore and spindle complex

It has been proposed previously that protein-coding changes in cell cycle-related genes are highly relevant candidates for human-specific traits[1,13]. Indeed, three genes (*CASC5*, *SPAG5*, and *KIF18A*) have been singled out as involved in spindle pole assembly during mitosis[1]. Other genes with protein-coding SNCs (*NEK6* and *STARD9/KIF16A*) turn out to be implicated in the regulation of spindle pole assembly as well[46,47]. Furthermore, it has been claimed[13] that genes with fixed non-synonymous changes in humans are also more often expressed in the ventricular zone of the developing neocortex, compared to fixed synonymous changes. Since the kinetochore-associated genes *CASC5*, *KIF18A* and *SPAG5* are among these genes, it has been emphasized that this "may be relevant phenotypically as the orientation of the mitotic cleavage plane in neural precursor cells during cortex development is thought to influence the fate of the daughter cells and the number of neurons generated[48,,13]. Several fixed SNCs on the modern human lineage are observed for *CASC5* (two changes) and *SPAG5* (three changes), which is also among genes with a relatively high

bioRxiv preprint doi: https://doi.org/10.1101/298950; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

proportion of HF changes (Table S5). The changes in *KIF18A*, *KIF16A* and *NEK6* can no longer be considered as fixed, but occur at very high frequencies (>99.9%) in present-day humans.

We attempted to determine whether an enrichment of genes with HHMCs on the human lineage can be observed in the ventricular zone for the same data[45], but instead find an enrichment in the intermediate zone, where less than 5% of random gene sets of the same size are expressed. However, synonymous HF changes also show an enrichment in this layer, as well as genes with AHMCs (Table S10), suggesting an overrepresentation of genes that carry mutations in the coding regions rather than lineage-specific effects. However, we were able to broadly recapitulate the observation of an enrichment of expression in the ventricular zone if restricting the test to genes with non-synonymous changes at a frequency greater than 99.9% in present-day humans, which is not observed for corresponding synonymous and archaic non-synonymous changes (Table S10). Expression of genes with AHMCs is enriched in the intermediate zone. Among the 28 genes expressed in the ventricular zone that carry almost fixed HHMCs, four might be enriched for HF changes in humans (*HERC5*, *LMNB2*, *SPAG5*, *VCAM1*), and one shows an excess of HF changes on the human compared to the archaic lineage (*AMKMY1*). Other notable genes discussed in this study include *ADSL*, *FAM178A*, *KIF26B*, *SLC38A10*, and *SPAG17*. We find 126 genes (Table S9) with 143 HHMCs that putatively interact with proteins at the centrosome-cilium interface[49], which is more than expected using 1,000 random gene sets of a similar length distribution, for which 98.9% contain fewer genes with HHMCs. However, 99.9% of random sets also contain fewer genes with AHMCs, suggesting that differences between humans and archaics might lie in the particular genes rather than their numbers. The centrosome-cilium interface is known to be critical for early brain development, and centrosome-related proteins are overrepresented in studies on the microcephaly phenotype in humans[50], which we will discuss below. Some of the genes listed here and discussed elsewhere in this study, such as *FMR1*, *KIF15*, *LMNB2*, *NCOA6*, *RB1CC1*, *SPAG5* and *TEX2*, harbor not only HHMCs, but an overall high proportion of HF changes on the human lineage.

Among the 15 fixed protein-coding changes identified here but absent from previous analyses[1,13], some might also contribute to complex modifications of pathways in cell division: The AHR protein is involved in cell cycle regulation[51] and shows an excess of HF changes on the human lineage, the dynein DNHD1 might be recruited to the kinetochore[52] and is overexpressed in fetal brain[53], and the SSH2 protein (two fixed changes, one of which is first described here; and one on the archaic lineage) might interact with spindle assembly checkpoint proteins[54]. SHROOM4, which is associated to a mental retardation syndrome with delayed speech and aggressive behavior[55], may also be relevant[56]. Other proteins that carry two HHMCs are involved in mitosis, for example the spindle checkpoint regulator CHEK1[57], the Dynein Axonemal Heavy Chain 1 (encoded by *DNAH1*), the mitotic regulator AZI1 (*CEP131*)[58], the Cyclin D2 (*CCND2*) and the Protein Tyrosine Phosphatase Receptor Type C (*PTPRC*)[59]. Other genes with HHMCs that could be part of the same functional network are *FOXM1*[60] and *FMR1*[61], which carry a putative enrichment of HF changes, and *TOP2A*[62]. The TOP2A protein shows the largest number of interactions (53) with other HHMC-carrying proteins, while CHEK1, KIF18A, KIF15 and PTPRC are among highly-interacting proteins with more than ten interactions, which suggests that these proteins might function as interaction hubs in modifications of the cell division complex. Furthermore, enrichment in cell-cycle related GO categories has been found for candidate regions for ancient positive selection[20], and *ANAPC10* has been highlighted, containing two potentially disruptive intronic changes that are fixed derived in modern humans and ancestral in both Neanderthals and Denisovans. This gene

bioRxiv preprint doi: https://doi.org/10.1101/298950; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

carries a total of 39 HF changes (11 of them fixed) specific to modern humans, but none for archaics.

All of this suggests that the cell cycle machinery might have been modified in a specific way in humans compared to other hominins. One particular example of specific consequences for a relevant SNC on the human lineage is *SPAG5*: One of the three fixed non-synonymous changes in the SPAG5 protein is a Proline-to-Serine substitution at position 43. This position is phosphorylated in humans[63] during the mitotic phase of the cell cycle, directly through the protein phosphatase 6 (PPP6C) at the Serine at this position[64], with the effect of a modification of the duration of the metaphase. PPP6C regulates the mitotic spindle formation[65], and the *PPP6C* gene itself carries five HF SNCs on the modern human lineage, one of which is a TF binding site (for HNF4A/HNF4G), and only one SNC on the archaic lineage. This specific substitution in *SPAG5* seems likely to influence the duration of the metaphase through phosphorylation, as a molecular consequence of this HHMC.

On the archaic lineage, we find an AHMC in the *ASPM* gene, along with 24 other HF changes, but none in modern humans, resulting in an excess of archaic SNCs. The proteins ASPM and CIT, which carries an AHMC that is listed among the most disruptive non-synonymous derived SNCs in archaics[17] (Table S31), are known to co-localize to the midbody ring during cytokinesis and regulate spindle orientation by affecting the dynamics of astral microtubules[66]. These proteins regulate astral microtubules and thus the orientation of cell division in archaics, whereas in modern humans we find proteins regulating kinetochore microtubules, thus the timing of cell division. This difference could indicate two alternative ways of modulating cell division on the different lineages.

Cellular features of neurons: Axons, the myelin sheath and synapses

Moving on from differences on the molecular level to cellular features of neurons, it is clear that wiring is key to cognitive ability, and in this context, axon guidance is relevant: To form critical networks during the early development of the brain, axonal extensions of the neurons in the cortical region must be sent and guided to eventually reach their synaptic targets. Studies conducted on avian vocal learners[37,67] have shown a convergent differential regulation of axon guidance genes of the *SLIT-ROBO* families in the pallial motor nucleus of the learning species, allowing for the formation of connections virtually absent in the brains of vocal non-learners. In modern humans, genes with axon-guidance-related functions such as *FOXP2*, *SLIT2* and *ROBO2* have been found to lie within deserts of archaic introgression[25,26,68], suggesting incompatibilities between modern humans and archaics for these regions. *SLITRK1* might have been under positive selection[19], and carries one HHMC, and *SLIT3* carries an excess of HF changes in modern humans. Several genes involved in wiring carry HHMCs, which we want to delineate here.

Some of the aforementioned microtubule-related genes, specifically those associated with axonal transport and known to play a role in post-mitotic neural wiring and plasticity[69], are associated with signals of positive selection, such as *KIF18A*[70] or *KATNA1*[71,72]. Furthermore, an interactor of *KIF18A*, *KIF15*[73], might have been under positive selection in modern humans[19], and contains two HHMCs (but also five AHMCs). Versican (*VCAN*), which promotes neurite outgrowth[74], carries three HHMCs, and *SSH2* (two HHMCs) might be involved in neurite outgrowth[75]. *PIEZO1*, which carries a non-synonymous change that is almost fixed in modern humans, is another factor in axon guidance[76], as well as *NOVA1*[77], which is an interactor of *ELAVL4*[78], a gene that codes for a neuronal-specific RNA-binding protein and might have been under positive selection in humans[19,22]. Furthermore, among genes with the most deleterious regulatory SNCs, we find the Netrin receptor UNC5D, which is critical for axon guidance[79].

bioRxiv preprint doi: https://doi.org/10.1101/298950; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The establishment of new connections requires protection, particularly as some of these connections reach long distance and are associated with enhanced activity following rewiring events, like for vocal motor neurons in songbirds[67]. The gene *MAL*, which is implicated in myelin biogenesis and function, shows up in selective sweep regions[19,20] and is enriched for HF changes on the human lineage, while its orthologue *MAL2* carries a HHMC. A gene with HHMCs that is associated with the organization of the axon initial segment and nodes of Ranvier during early development is *NFASC*[80]. The protein encoded by this gene is a L1 family immunoglobulin cell adhesion molecule, and we find that also the *L1CAM* gene carries an AHMC[81]. NFASC is also an interactor of DCX[82], which might have been under positive selection in humans[19] and is enriched for HF SNCs on the human lineage, but carries an AHMC as well. At least two genes associated with the process and timing of myelination, *PTEN*[83], and *NCMAP*[84] are among genes with an excess of HF SNCs in modern humans. Other genes carrying HHMCs in our dataset associated with myelination include *SCAP*[85], *RB1CC1*[86], *TENM4*[87], *CDKL1*[88] and *ADSL*[89], and genes with an excess of changes on the human lineage with similar functions include *FBXW7*[90], *KIFAP3*[91], and *AMPH*[92]. The AMPH protein interacts closely with the huntingtin protein HTT (which also carries a HHMC) and is involved in myelination processes[93].

Another interesting class that emerges from the set of genes is related to synaptic vesicle endocytosis, critical to sustain a high rate of synaptic transmission. We find a formal enrichment of genes with an excess of HF changes on the human compared to the archaic lineage with gene products located in the postsynaptic membrane and dendrites. *PACSIN1*[94] carries a HHMC, is among genes with an excess of HF changes, and has been highlighted as putatively under positive selection on the human lineage, along with other synaptic plasticity related genes such as *SIPA1L1*[19,20,22], *SH3GL2*[95] and *STX1A*[96]. Among genes harboring HHMCs and related to synaptic vesicle endocytosis, we find *LMNB2*[97] and *SV2C*[98]. Finally, *SYT1*, which is critical for synaptic vesicle formation[99], carries a deleterious HHMC (Table 2). Synaptic properties have been mentioned before in the context of human specific traits, for instance in postnatal brain development in humans, chimpanzees and macaques[100], with a focus on synaptogenesis and synaptic elimination in the prefrontal cortex. A period of high synaptic plasticity in humans has been related to a cluster of genes around a transcription factor encoded by the *MEF2A* gene. Even though this gene neither carries a protein-altering change nor shows a particular pattern in our analysis, any of the 26 HF SNCs it harbors on the modern human lineage could have had a functional impact not captured here. Apart from that, several of the genes with an excess of HF changes in modern humans do belong to this cluster: *CLSTN1*, *FBXW7*, *GABBR2*, *NRXN3*, *PTPRJ*, *PTPRN2*, *SLIT3*, and *STX1A*, three of which (*CLSTN1*, *FBXW7* and *STX1A*) are associated with signals of positive selection[19]. In addition, the above-mentioned AMPH interacts via CDKL5[101] with HDAC4[102]. The latter exhibits an excess of HF changes in modern humans, and is known to repress the transcriptional activation of *MEF2A*[103]. A putative signature of positive selection upstream of *MEF2A*[104] suggests that this may be part of a broader network which might be supported by our analysis. Finally, *ENTHD1/CACNA1I*, which contains a HHMC that can no longer be considered as fixed, but occurs at a very high frequency (>99.9%), lies in a selective sweep region[19]. The protein encoded by this gene is involved in synaptic vesicle endocytosis at nerve terminals[105] and is regulated by the *MEF2* gene family[106].

The brain growth trajectory

The number of neurons in the brain might be influenced by some of the changes in kinetochore-associated genes[13], and their organization and neuronal wiring clearly impose structural demands on

bioRxiv preprint doi: https://doi.org/10.1101/298950; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the organization of the brain. We have presented candidate genes and networks for these features above, but brain organization might involve broader networks. Brain growth factors identified by disease phenotypes in modern humans such as micro- and macrocephaly have been highlighted previously as potentially relevant for physiological differences between humans and archaics[18]. Although an early analysis suggested several candidate genes associated to microcephaly, not all of these could be confirmed by high-coverage data. Among eleven candidate genes[18], only two (*PCNT, UCP1*) are among the HHMC gene list presented here, while most of the other changes are not human-specific, and only *PCNT* has been related to microcephaly[107]. Nevertheless, more such changes are found on both lineages: For example, our data reveals that in archaics there are AHMCs in the microcephaly candidate genes *ASPM*[108] and *CIT*[109]. The ASPM-katanin complex controls microtubule disassembly at spindle poles and misregulation of this process can lead to microcephaly[110], which is of interest given the presence of a HHMC in *KATNA1* and a fixed non-coding change in *KATNB1*, while no such changes were observed in archaics[111]. Other genes associated with microcephaly that harbor non-synonymous SNCs are *CASC5* (two in humans, one in archaics)[112], *CDK5RAP2* (in humans), *MCPH1* (in archaics)[113], *ATRX* (one in humans and archaics each)[114], and *NHEJ1*[115] (a deleterious one in humans, and one in archaics). The SPAG5 protein, which carries three fixed HHMCs, has been claimed to interact with CDK5RAP2[116], is a direct target of PAX6[117], via which it affects cell division orientation, and therefore is critical in the course of brain development. Disease mutations in *SCAP* or *ADSL* have also been associated with microcephaly phenotypes as well[118,119], and Formin-2 (*FMN2*), which carries a deleterious regulatory change in modern humans, influences the development of the brain causing microcephaly in mice[120].

Genes associated with brain growth trajectory changes lead not necessarily to a decrease but also an increase of brain size[121], suggesting that the disease phenotype of macrocephaly might point to genes relevant in the context of brain growth as well. One of the few genes with several HHMCs, *CASC5*, has been found to be associated with gray matter volume differences[122]. It has been claimed that mutations in *PTEN* alter the brain growth trajectory and allocation of cell types through elevated Beta-Catenin signaling[123]. This gene is also present among differentially expressed genes in human neurons compared to chimpanzee neural progenitor cells during cerebral cortex development, which may relate to a lengthening of the prometaphase-metaphase in humans compared to chimpanzees that is specific to proliferating progenitors and not observed in non-neural cells[124]. We find that *PTEN* falls among the genes with the highest number of HF SNCs on the human lineage per length, and also among the genes with an excess on the modern human over the archaic lineage, suggesting that regulatory changes in this gene might have contributed to human-specific traits. This is also the case for the HHMC-carrying transcription factor TCF3, which is known to repress Wnt-Beta-Catenin signaling and maintain the neural stem cell population during neocortical development[125]. Among other macrocephaly-related genes with HHMCs in *RNF135*[126], *CUL4B*[127] and *CCND2*[128], the latter also shows a large number of HF changes on the human lineage, and the HHMC in *CUL4B* is inferred to be deleterious (Table 2). Other macrocephaly candidates such as *NFIX*[129], *NSD1*[130] and *GLI3*[131] have been claimed to have played an important role in shaping the distinctly modern human head[132] and show numerous SNCs in non-coding regions. *GLI3* might have been under positive selection[19] and carries 20 HF SNCs on the human, but only one on the archaic lineage. Two of the very few genes hypothesized to regulate expansion and folding of the mammalian cerebral cortex by controlling radial glial cell number and fate, *TRNP1*[133] and *TMEM14B*[134], exhibit HF 3'-UTR changes in modern humans, and *TRNP1* shows an excess of changes on the modern human lineage. The expression of these two genes in the outer

bioRxiv preprint doi: https://doi.org/10.1101/298950; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

subventricular zone might be important[135], since this is an important region for complexification of neocortical growth in primates[136], and for which an enriched activation of mTOR signaling has been reported[137]. In addition to other genes in the mTOR-pathway, such as *PTEN*[138] or *CCND2*, two possibly interacting modulators[139] of the mTOR signaling pathway stand out in our dataset: *ZNHIT2* with one deleterious SNC (Table 2) might have been under positive selection[19], and *CCT6B* carries a deleterious change according to both SIFT and C-score. The TF encoded by *RB1CC1* is essential for maintaining adult neuronal stem cells in the subventricular zone of the cerebral cortex[140]. This gene carries a HHMC, a regulatory SNC that has been suggested to modify transcriptional activity[141], and a signature of positive selection[13].

Changes in the genes mentioned here could have contributed to the brain growth trajectory changes hypothesized to give rise to the modern human-specific globular braincase shape during the past several 100,000 years[8,10,39]. On the archaic side, an enrichment of genes with AHMCs associated to "Corneal structure" may relate to archaic-specific changes in brain growth-trajectories since the size and position of the frontal and temporal lobes might affect eye and orbit morphology[142], and the macrocephaly-associated gene *RIN2*[143] carries an AHMC. Finally, changes that might have affected the size of the cerebellum can be found in our dataset, such as the HF regulatory SNCs found in *ZIC1* and *ZIC4*[144], and the deleterious HHMC in *ABHD14A*, which is a target of ZIC1[145].

The craniofacial phenotype

Differences other than brain-related properties are likely to have emerged after the split from archaic humans, some of which may have had an impact on cognition more indirectly[146,147]. Among the genes harboring HHMCs and found in selective sweep regions, the gene encoding the TF PRDM10 stands out, since this is the second-most interacting protein within the HHMC dataset. Although little is known about *PRDM10*, it may be related dendrite growth[148] and to neural crest related changes that contributed to the formation of our distinct modern face[149]. Changes in genes related to craniofacial morphology would complement previous observations[17,132], and we find an enrichment of genes with an excess of HF SNCs on the modern human lineage for the GO term "soft palate development", which might relate to craniofacial properties that are relevant for language[150]. Among genes harboring an excess of HF SNCs associated with specific facial features, we find *RUNX2*, *EDAR*, and *GLI3*[151], *NFATC1*[152], *SPOP*[153], *DDR2*[154] and *NELL1*[155], possibly due to changes in regulatory regions, while mutations in the HHMC-carrying gene encoding for the TF ATRX cause facial dysmorphism[156]. In addition, genes with HHMCs such as *PLXNA2*[157], *EVC2*[158], *MEPE*[159] and *SPAG17*[160] are known to affect craniofacial bone and tooth morphologies. These genes appear to be important in determining bone density, mineralization and remodeling, hence they may underlie differences between archaic and modern human facial growth[161]. Some of these facial properties may have been present in the earliest fossils attributed to *H. sapiens*, like the Jebel Irhoud fossils[4], deviating from craniofacial features which emerged in earlier forms of *Homo*[162], and may have become established before some brain-related changes discussed here[39,163].

Other craniofacial morphology-related genes, such as *DCHS2*[151], *HIVEP2*[164], *HIVEP3*[165], *FREM1*[166], and *FRAS1*[167] harbor AHMCs, while another bone-related gene, *MEF2C*[168], shows an excess of HF changes on the archaic lineage. These changes may underlie some of the derived facial traits of Neanderthals[169].

bioRxiv preprint doi: https://doi.org/10.1101/298950; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The impact on cognition and language

It has long been hypothesized that language and its neurological foundation were important for the evolution of humans and uniquely human traits, closely related to hypotheses on the evolution of cognition and behavior. It is noteworthy that among traits associated with cognitive functions such as language or theory of mind, the timing of myelination appears to be a good predictor of computational abilities[170,171]. Computational processing might have been facilitated by some of the changes presented here, at least in some of the circuits that have expanded in our lineage[172], since subtle maturational differences early in development[173] may have had a considerable impact on the phenotype. This might be linked to the specific brain growth trajectory in modern humans[11], and reflected in the morphology of the parietal and temporal lobes[174,175], as well as in the size of the cerebellum[8]. Archaic hominins likely had certain language-like abilities[176,177], and hybrids of modern and archaic humans must have survived in their communities[178]. However, another important hint for human-specific features is that genes associated with axon guidance functions, which are important for the refinement of neural circuits including those relevant for speech and language, are found in introgression deserts[36,179], and especially the *FOXP2* region is depleted for archaic introgression, which seems to be a unidirectional and human-specific pattern[68]. This, together with putative positive selection after the split from Neanderthals[180] and regulatory changes affecting *FOXP2* expression[181], could indicate modifications of a complex network in cognition or learning, possibly related to other brain-related, vocal tract[132] or neural changes[182]. We suggest that some other genes with changes on the human lineage might have contributed more specifically to cognition-related changes, although we admit that the contribution of single SNCs to these functions is less straightforward than their contribution to molecular mechanisms, since disease mutations in many genes may have disruptive effects on cognitive abilities.

The basal ganglia are a brain region where *FOXP2* expression is critical for the establishment and maintenance of language-related functions[183,184]. Several genes carrying HHMCs have been described previously as important for basal ganglia functions related to language and cognition. The HTT protein has long been implicated in the development of Huntington's disease, which is associated with corticostriatal dysfunction, and is known to interact with FOXP2[185]. Mutations in *SLITRK1* have been linked to Tourette's syndrome, a disorder characterized by vocal and motor tics, resulting from a dysfunction in the corticostriatal-thalamocortical circuits[186]. NOVA1 regulates RNA splicing and metabolism in a specific subset of developing neurons, particularly in the striatum[187]. As pointed out above, *NOVA1* is an interactor of *ELAVL4*, which belongs to a family of genes known to promote the production of deep layer *FOXP2*-expressing neurons[188–190], and part of a neural network-related cluster that has been highlighted as putatively under positive selection in humans[22]. Within this network, α-synuclein (encoded by *SCNA*) might serve as a hub and is specifically expressed in brain regions important for vocal learning regions in songbirds[67]. *SCNA* and *SV2C*, which carries a HHMC, are involved in the regulation of dopamine release, with *SV2C* expression being disrupted in *SCNA*-deficient mice and in humans with Parkinson's disease[191]. Genes in the cluster of selection signals[22] are implicated in the pathogenesis of Alzheimer's disease, which (together with Huntingon's and Parkinson's diseases) is linked to a *FOXP2*-driven network[192]. Some introgressed archaic alleles are downregulated in specific brain regions[27], especially pronounced in the cerebellum and basal ganglia. One notable example is *NTRK2*, which shows an excess of HF changes on the human lineage and a signature of positive selection[19], and is also a FOXP2 target[35], a connection which has been highlighted for the vocal learning circuit in birds[193]. Other genes harboring HHMCs such as *ENTHD1*[106] and *STARD9*[194], as well as genes in introgression deserts[25], have been associated with language deficits. It may indeed have taken a

bioRxiv preprint doi: https://doi.org/10.1101/298950; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

complex composite of changes to make our brain fully language-ready[195], where not all changes needed to reach fixation.

In the broader context of cognition, we find an enrichment of HHMCs in genes associated to "Alzheimer's disease (cognitive decline)" and "Cognitive decline (age-related)", with seven associated genes (*COX7B2*, *BCAS3*, *DMXL1*, *LIPC*, *PLEKHG1*, *TTLL2* and *VIT*). Two other genes linked to Alzheimer's are *PTEN*[196], and *RB1CC1*[197]. Among genes with deleterious HHMCs, *SLC6A15* has been associated to emotional processing in the brain[198], and may be part of modifications in glutamatergic transmission[199], a category found in selective sweep regions[147]. *GPR153*, which carries one HHMC and two AHMCs, influences behavioral traits like decision making in rats, and is associated with various neuropsychiatric disorders in humans[200]. Another interesting candidate change in the context of cognitive abilities might affect the Adenylosuccinate Lyase (*ADSL*), for which the ancestral Neanderthal-like allele has not been observed in 1,000s of modern human genomes. This gene has been associated to autism[201], is part of behavioral traits like "aggressive behavior" which have been found to be enriched on the human lineage[17], and several studies detected a signal of positive selection in modern humans[19,20,202]. These observations make *ADSL* a strong candidate for human-specific features, particularly in light of the fact that the relevant HHMC is located in a region that is highly conserved and lies close to the most common disease mutation leading to severe adenylosuccinase deficiency[20]. Other relevant genes, similar to *ADSL* in carrying a fixed HHMC and being frequently found in selective sweep screens, are *NCOA6*, which might be related to autism as well[203], and *SCAP*. Downregulation of the cholesterol sensor encoded by this gene has been shown to cause microcephaly, impaired synaptic transmission and altered cognitive function in mice[119]. We want to emphasize that the networks presented in the previous sections influencing brain growth and neural wiring are likely to impact cognitive functions, since disruptions in these networks would impair the healthy human brain. Furthermore, we find an enrichment of AHMCs in genes associated to Parkinson's disease and "Attention deficit hyperactivity disorder and conduct disorder", suggesting that changes may have taken place in related networks on the archaic lineage as well.

Life history and other phenotypic traits

Apart from their consequences for cognitive functions, it has been suggested that changes involved in synaptic plasticity might be interpreted in a context of neoteny[19,100,204,205], with the implication of delayed maturation in humans[206] and a longer timeframe for brain development. However, given their similar brain sizes[207], humans and Neanderthals might both have needed a long overall maturation time[208,209]. Accordingly, notions like neoteny and heterochrony are unlikely to be fine-grained enough to capture differences between these populations, but early differences in infant brain growth between humans and Neanderthals[8,210] could have rendered our maturational profile distinct during limited developmental periods and within specific brain regions, imposing different metabolic requirements[211]. One of the brain regions where such differences are found is the orbitofrontal cortex (OFC)[174], and we find that the OFC at infant age (0-2 years) is enriched for the expression of genes that carry HHMCs compared to synonymous SNCs. We suggest that the development of the OFC in infants might have been subject to subtle changes since the split from Neanderthals rather than a general developmental delay, which is particularly interesting given that this brain region has been implied in social cognition[212] and learning[213].

Genes carrying HHMCs are enriched for expression in the gray matter of the forebrain at the adolescent age compared to AHMC-carrying genes, hence additional human-specific modifications during this period might have taken place, possibly linked to changes in myelination described

bioRxiv preprint doi: https://doi.org/10.1101/298950; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

above. It has been suggested that differences in childhood adolescence time existed between humans and Neanderthals, after a general developmental delay in the hominin lineage[214,215]. Dental evidence suggests an earlier maturation in Neanderthals than modern humans[216], and it has been claimed that Neanderthals might have reached adulthood earlier[217]. Furthermore, an introgressed indel from Neanderthals causes an earlier onset of menarche in present-day humans[218], supporting at least the existence of alleles for earlier maturation in the Neanderthal population. Among the genes carrying fixed HHMCs, *NCOA6* has also been linked to age at menarche and onset of puberty[219], as well as placental function[220]. This putative TF is enriched in HF changes and has been suggested to have been under positive selection on the modern human lineage[19,202]. The HHMC is located nearby and three 5'-UTR variants within a putatively selected region[22], with an estimated time of selection at around 150 kya (assuming a slow mutation rate). Even though this gene carries an AHMC as well, it remains possible that modern humans acquired subtle differences in their reproductive system through lineage-specific changes in this gene. A delay in reproductive age may influence overall longevity, another trait for which our data set yields an enrichment of genes with HHMCs (*SLC38A10*, *TBC1D22A* and *ZNF516*).

The male reproductive system might have been subject to changes as well, since we find that several proteins in spermatogenesis seem to carry two HHMCs: Sperm Specific Antigen 2 (*SSFA2*), Sperm Associated Antigen 17 (*SPAG17*), *ADAM18*[221] and *WDR52*[222], out of which *ADAM18* and *SPAG17* also carry AHMCs. Lineage-specific differences in genes related to sperm function or spermatogenesis might have been relevant for the genetic compatibility between humans and Neanderthals. Another gene harboring a HHMC with similar functions is *EPPIN*[223], which shows no HF changes on the archaic, but 27 such SNCs on the modern human lineage. The gene encoding for the Testis Expressed 2 protein (*TEX2*) is enriched for HF changes in both humans and archaics, with one HHMC and five AHMCs, but its function is not yet known. Another possible SNC that might be relevant in this context is a splice site change in *IZUMO4*, since proteins encoded by the IZUMO family form complexes on mammalian sperm[224]. The adjacent exon is not present in all transcripts of this gene, suggesting a functional role of this splice site SNC. Finally, genes in the GO category "spermatoproteasome complex" are enriched for an excess of HF changes on the human compared to the archaic lineage.

It has been found that Neanderthal alleles contribute to addiction and, possibly, pain sensitivity in modern humans[225,226]. In this context, an interesting protein-truncating SNC at high frequency in humans is the loss of a stop codon in the opioid receptor OPRM1 (6:154360569), potentially changing the structure of the protein encoded by this gene in some transcripts. Other mutations in this gene are associated to heroin addiction[227], and pain perception[228], but also sociality traits[229]. Interestingly, a recent study found a pain insensitivity disorder caused by a mutation in *ZFHX2*[230], which carries an AHMC, and three HHMCs are observed in *NPAP1*, which might be associated with the Prader-Willi syndrome, involving behavioral problems and a high pain threshold[231]. Such changes may point to differences in levels of resilience to pain between Neanderthals and modern humans.

Conclusion

The long-term evolutionary processes that led to the human condition[1] is still subject to debate and investigation, and the high-quality genomes from archaic humans provide opportunities to explore the recent evolution of our species. We want to contribute to an attempt to unveil the

bioRxiv preprint doi: https://doi.org/10.1101/298950; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

genetic basis of specific molecular events in the time-window after the split from these archaic populations and before the emergence of most of the present-day diversity. We sought to combine different sources of information, from genome-wide enrichment analyses to functional information available for specific genes, to identify threads linking molecular needles in this expanded haystack. In doing so, we have mainly built on existing proposals concerning brain-related changes, but we have divided the observations into different biological levels, from cellular changes through brain organization differences to complex phenotypic traits. Only future experimental work will determine which of the changes highlighted here contributed significantly to making us "fully human". We hope that our characterization and presentation of some new candidate genes will help prioritize inquiry in this area.

Acknowledgments

We thank S. Han and T. Marques-Bonet for helpful discussions, and A. G. Andirkó and P .T. Martins for help with figures. M.K. is supported by a Deutsche Forschungsgemeinschaft (DFG) fellowship (KU 3467/1-1). C.B. acknowledges research funds from the Spanish Ministry of Economy and Competitiveness (grant FFI2016-78034-C2-1-P), Marie Curie International Reintegration Grant from the European Union (PIRG-GA-2009-256413), research funds from the Fundació Bosch i Gimpera, MEXT/JSPS Grant-in-Aid for Scientific Research on Innovative Areas 4903 (Evolinguistics: JP17H06379), and Generalitat de Catalunya (Government of Catalonia) – 2017-SGR-341.

Author contributions

M.K. and C.B. analyzed data and wrote the manuscript.

Competing interests statement

The authors declare no competing interests.

Supplementary Table & Figure legends

Table S1: List of HHMCs and genomic features

Table S2: List of low-confidence HHMCs and genomic features

Table S3: List of AHMCs and genomic features

Table S4: List of low-confidence AHMCs and genomic features

Table S5: Top 5% of genes by HF SNC density on the modern human and archaic lineages, and top 10% of genes by relative excess of HF SNCs on one lineage over the other.

Table S6: GO enrichment for genes with relative excess of HF SNCs on the human over the archaic lineage

bioRxiv preprint doi: https://doi.org/10.1101/298950; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S7: GWAS enrichment for genes with HHMCs or AHMCs

Table S8: Number of interactions among genes with HHMCs or AHMCs

Table S9: Genes with HHMCs that are TFs, or at the centrosome interface

Table S10: Enrichment in developing brain zones for genes with HHMCs or AHMCs, proportion of random gene sets with larger overlap (Methods).

Figure S1: Distribution of missense and non-synonymous HF SNCs across chromosomes.

Figure S2: STRING graph of interactions among genes with HHMCs.

Methods

We used the publicly available high-coverage genotypes for three archaic individuals: One Denisovan[12], one Neanderthal from the Denisova cave in Altai mountains[13], and another Neanderthal from Vindija cave, Croatia[14]. The data is publicly available under http://cdna.eva.mpg.de/neandertal/Vindija/VCF/, with the human genome version *hg19* as reference. We applied further filtering to remove sites with less than 5-fold coverage and more than 105-fold coverage in the Altai Neanderthal or 75-fold coverage in the other archaic individuals, if such cases occurred. We also removed sites with genotype quality smaller than 20, and heterozygous sites with strong allele imbalance (<0.2 minor allele frequency). Although these permissive filters increase power compared to previous studies, we caution that in some cases genotypes might be incorrect. We added the genotype and coverage for the exome and chromosome 21 sequences of the Vindija and El Sidrón Neanderthals from previous studies[2,17], with 75-fold and 50-fold coverage cutoffs, respectively. These studies provided data for the same Vindija individual[14].

We applied the Ensembl Variant Effect Predictor[232] in order to obtain inferences for protein-coding and regulatory mutations, scores for SIFT[233], PolyPhen[234], CADD[235] and GWAVA[236], and allele frequencies in the 1000 Genomes and ExAC human variation databases[15,237]. We used the inferred ancestral allele from[238], and at positions where this information was not available, the macaque reference allele, *rheMac3*[239]. We determined the allele frequencies in present-day humans using the dbSNP database build 147[240]. We retrieved the counts for each allele type, and summarized the counts of non-reference alleles at each position. Grantham scores[241] were calculated for missense mutations.

Data processing and database retrieval was performed using bcftools/samtools v1.0[242], bedtools v2.16.2[243], and R/Bioconductor[244], rtracklayer[245] and biomaRt[246], and plotting with RCircos[247]. We analyzed all positions where at least two alleles (human reference and alternative allele) were observed among the human reference and at least one out of three of the high-coverage archaic individuals in at least one chromosome. The 22 autosomal chromosomes and the X chromosome were analyzed, in the absence of Y chromosome data for the three female archaic individuals. The data for 4,409,518 segregating sites is available under [http:tbd.database]. The following subsets were created:

bioRxiv preprint doi: https://doi.org/10.1101/298950; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Fixed differences: Positions where all present-day humans carry a derived allele, while at least two out of three archaics carry the ancestral allele, accounting for potential human gene flow into Neanderthals.

High-frequency (HF) differences: Positions where more than 90% of present-day humans carry a derived allele, while at least the Denisovan and one Neanderthal carry the ancestral allele, accounting for different types of errors and bi-directional gene flow.

Extended high-frequency differences: Positions where more than 90% of present-day humans carry a derived allele, while one of the following conditions is true: a) Not all archaics have reliable genotypes, but those that have carry the ancestral allele. b) Some archaics carry an alternative genotype that is not identical to either the human or the ancestral allele. c) The Denisovan carries the ancestral allele, while one Neanderthal carries a derived allele, which allows for gene flow from humans into Neanderthals. d) The ancestral allele is missing in the EPO alignment, but the macaque reference sequence is identical to the allele in all three archaics.

We also created corresponding lists of archaic-specific changes. Fixed changes were defined as sites where the three archaics carry the derived allele, while humans carry the ancestral allele at more than 99.999%. High-frequency changes occur to less than 1% in present-day humans, while at least two archaic individuals carry the derived allele. An extended list presents high-frequency changes where the ancestral allele is unknown, but the macaque allele is identical to the present-day human allele.

A ranking of mutation density was performed for genes with protein-coding sequences and their genomic regions as retrieved from Ensembl. For each gene, unique associated changes as predicted by VEP were counted. A ranking on the number of HF changes per gene length was performed for all genes that span at least 5,000 bp in the genome and carry at least 25 segregating sites in the dataset (at any frequency in humans or in archaics), in order to remove genes which are very short or poor in mutations. The top 5% of the empirical distribution was defined as putatively enriched for changes on each lineage. The ratio of lineage-specific HF changes was calculated for the subset of genes where at least 20 lineage-specific HF changes were observed on the human and the archaic lineages combined. The top 10% of the empirical distribution was defined as putatively enriched for lineage-specific changes.

We performed enrichment tests using the R packages ABAEnrichment[44] and DescTools[248]. We used the NHGRI-EBI GWAS Catalog[41], and overlapped the associated genes with protein-coding changes on the human and archaic lineages, respectively. We counted the number of HF missense changes on each lineage and the subset of those associated to each trait ("Disease trait"), and performed a significance test (G-test) against the number of genes associated to each trait, and all genes in the genome, with a P value cutoff at 0.1. This suggests a genome-wide enrichment of changes for each trait. We then performed a G-test between the numbers of HF missense changes on each lineage, and the subset of each associated to each trait (P-value cutoff at 0.1), to determine a difference between the two lineages. We then performed an empirical test by creating 1,000 random sets of genes with similar length as the genes associated to each trait, and counting the overlap to the lineage-specific missense changes. At least 90% of these 1,000 random sets were required to contain fewer missense changes than the real set of associated genes. Only traits were considered for which at least 10 associated loci were annotated.

Gene Ontology (GO) enrichment was performed using the software FUNC[249], with a significance cutoff of the adjusted p-value < 0.05 and a family-wise error rate < 0.05. When testing missense changes, a background set of synonymous changes on the same lineage was used for the hypergeometric test. When testing genes with relative mutation enrichment, the Wilcoxon rank test

bioRxiv preprint doi: https://doi.org/10.1101/298950; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

was applied. Enrichment for sequence-specific DNA-binding RNA polymerase II transcription factors (TFs) and TF candidate genes from[43], and genes interacting at the centrosome-cilium interface[49] was tested with an empirical test in which 1,000 random sets of genes were created that matched the length distributions of the genes in the test list. The same strategy was applied for genes expressed in the developing brain (Table S10)[45]. Protein-protein interactions were analyzed using the STRING online interface v10.5[250] with standard settings (medium confidence, all sources, query proteins only) as of January 2018. The overlap with selective sweep screens considers HHMCs within 50,000 bp of the selected regions[13,19,22].

bioRxiv preprint doi: https://doi.org/10.1101/298950; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

1. Pääbo, S. The Human Condition—A Molecular Approach. *Cell* **157,** 216–226 (2014).
2. Kuhlwilm, M. *et al.* Ancient gene flow from early modern humans into Eastern Neanderthals. *Nature* **530,** 429–433 (2016).
3. Schlebusch, C. M. *et al.* Genomic Variation in Seven Khoe-San Groups Reveals Adaptation and Complex African History. *Science (80-.).* **338,** 374 LP-379 (2012).
4. Hublin, J.-J. *et al.* New fossils from Jebel Irhoud, Morocco and the pan-African origin of Homo sapiens. *Nature* **546,** 289 (2017).
5. Schlebusch, C. M. *et al.* Southern African ancient genomes estimate modern human divergence to 350,000 to 260,000 years ago. *Science (80-.).* (2017).
6. Skoglund, P. *et al.* Reconstructing Prehistoric African Population Structure. *Cell* **171,** 59–71.e21 (2017).
7. Rabett, R. J. The success of failed Homo sapiens dispersals out of Africa and into Asia. *Nat. Ecol. Evol.* **2,** 212–219 (2018).
8. Hublin, J.-J., Neubauer, S. & Gunz, P. Brain ontogeny and life history in Pleistocene hominins. *Philos. Trans. R. Soc. B Biol. Sci.* **370,** (2015).
9. Zollikofer, C. P. E., Bienvenu, T. & Ponce de León, M. S. Effects of cranial integration on hominid endocranial shape. *J. Anat.* **230,** 85–105 (2017).
10. Gunz, P. *et al.* A uniquely modern human pattern of endocranial development. Insights from a new cranial reconstruction of the Neandertal newborn from Mezmaiskaya. *J. Hum. Evol.* **62,** 300–313 (2012).
11. Boeckx, C. The language-ready head: Evolutionary considerations. *Psychon. Bull. Rev.* **24,** 194–199 (2017).
12. Meyer, M. *et al.* a High Coverage Genome Sequence From an Archaic Denisovan Individual. *Science* **338,** 222–226 (2012).
13. Prüfer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505,** 43–9 (2014).
14. Prüfer, K. *et al.* A high-coverage Neandertal genome from Vindija Cave in Croatia. *Science (80-.).* (2017).
15. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526,** 68–74 (2015).
16. Consortium, T. C. S. and A. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437,** 69 (2005).
17. Castellano, S. *et al.* Patterns of coding variation in the complete exomes of three Neandertals. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 6666–71 (2014).
18. Green, R. E. *et al.* A Draft Sequence of the Neandertal Genome. *Science (80-.).* **328,** 710–722 (2010).
19. Peyrégne, S., Boyle, M. J., Dannemann, M. & Prüfer, K. Detecting ancient positive selection in humans using extended lineage sorting. *Genome Res.* **27,** 1563–1572 (2017).
20. Racimo, F. Testing for Ancient Selection Using Cross-population Allele Frequency Differentiation. *Genetics* **202,** 733 LP-750 (2016).
21. Somel, M., Liu, X. & Khaitovich, P. Human brain evolution: transcripts, metabolites and their regulators. *Nat. Rev. Neurosci.* **14,** 1–16 (2013).
22. Zhou, H. *et al.* A Chronological Atlas of Natural Selection in the Human Genome during the Past Half-million Years. *bioRxiv* (2015).
23. Sankararaman, S. *et al.* The genomic landscape of Neanderthal ancestry in present-day humans. *Nature* **507,** 354–357 (2014).
24. Vernot, B. & Akey, J. M. Resurrecting Surviving Neandertal Lineages from Modern Human Genomes. *Science (80-.).* **343,** 1017–1021 (2014).
25. Vernot, B. *et al.* Excavating Neandertal and Denisovan DNA from the genomes of Melanesian individuals. *Science (80-.).* **352,** 235–239 (2016).

bioRxiv preprint doi: https://doi.org/10.1101/298950; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

26. Sankararaman, S., Mallick, S., Patterson, N. & Reich, D. The Combined Landscape of Denisovan and Neanderthal Ancestry in Present-Day Humans. *Curr. Biol.* **26,** 1241–1247 (2016).
27. McCoy, R. C., Wakefield, J. & Akey, J. M. Impacts of Neanderthal-Introgressed Sequences on the Landscape of Human Gene Expression. *Cell* **168,** 916–927.e12 (2017).
28. Wynn, T. & Coolidge, F. L. The expert Neandertal mind. *J. Hum. Evol.* **46,** 467–487 (2004).
29. Langergraber, K. & Prüfer, K. Generation times in wild chimpanzees and gorillas suggest earlier divergence times in great ape and human evolution. *Proc. …* **109,** 15716–15721 (2012).
30. Varki, A. & Altheide, T. K. Comparing the human and chimpanzee genomes: Searching for needles in a haystack. *Genome Res.* **15,** 1746–1758 (2005).
31. Wynn, T., Overmann, K. & Coolidge, F. The false dichotomy: a refutation of the Neandertal indistinguishability claim. *J. Anthropol. Sci. = Riv. di Antropol. JASS* **94,** 201–221 (2016).
32. Bufill, E., Blesa, R. & Augusti, J. Alzheimer's disease: an evolutionary approach. *J. Anthropol. Sci. = Riv. di Antropol. JASS* **91,** 135–157 (2013).
33. Bufill, E., Agustí, J. & Blesa, R. Human neoteny revisited: The case of synaptic plasticity. *Am. J. Hum. Biol.* **23,** 729–739 (2011).
34. Bruner, E. & Jacobs, H. I. L. Alzheimer's disease: the downside of a highly evolved parietal lobe? *J. Alzheimers. Dis.* **35,** 227–240 (2013).
35. Vernes, S. C. *et al.* Foxp2 Regulates Gene Networks Implicated in Neurite Outgrowth in the Developing Brain. *PLOS Genet.* **7,** 1–17 (2011).
36. Lei, H. *et al.* Axon guidance pathways served as common targets for human speech/language evolution and related disorders. *Brain Lang.* **174,** 1–8 (2017).
37. Wang, R. *et al.* Convergent Differential Regulation of SLIT-ROBO Axon Guidance Genes in the Brains of Vocal Learners. *J. Comp. Neurol.* **523,** 892–906 (2015).
38. Gilbert, J. & Man, H.-Y. Fundamental Elements in Autism: From Neurogenesis and Neurite Growth to Synaptic Plasticity. *Front. Cell. Neurosci.* **11,** 359 (2017).
39. Neubauer, S., Hublin, J.-J. & Gunz, P. The evolution of modern human brain shape. *Sci. Adv.* **4,** (2018).
40. Gratten, J. & Visscher, P. M. Genetic pleiotropy in complex traits and diseases: implications for genomic medicine. *Genome Med.* **8,** 78 (2016).
41. MacArthur, J. *et al.* The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.* **45,** D896–D901 (2017).
42. Wray, G. A. The evolutionary significance of cis-regulatory mutations. *Nat. Rev. Genet.* **8,** 206 (2007).
43. Chawla, K., Tripathi, S., Thommesen, L., Lægreid, A. & Kuiper, M. TFcheckpoint: a curated compendium of specific DNA-binding RNA polymerase II transcription factors. *Bioinformatics* **29,** 2519–2520 (2013).
44. Grote, S., Prüfer, K., Kelso, J. & Dannemann, M. ABAEnrichment: an R package to test for gene set expression enrichment in the adult and developing human brain. *Bioinformatics* **32,** 3201–3203 (2016).
45. Miller, J. A. *et al.* Transcriptional Landscape of the Prenatal Human Brain. *Nature* **508,** 199–206 (2014).
46. O'Regan, L. & Fry, A. M. The Nek6 and Nek7 Protein Kinases Are Required for Robust Mitotic Spindle Formation and Cytokinesis. *Mol. Cell. Biol.* **29,** 3975–3990 (2009).
47. Torres, J. Z. *et al.* The STARD9/Kif16a Kinesin Associates with Mitotic Microtubules and Regulates Spindle Pole Assembly. *Cell* **147,** 1309–1323 (2017).
48. Fietz, S. A. & Huttner, W. B. Cortical progenitor expansion, self-renewal and neurogenesis—a polarized perspective. *Curr. Opin. Neurobiol.* **21,** 23–35 (2011).
49. Gupta, G. D. *et al.* A Dynamic Protein Interaction Landscape of the Human Centrosome-Cilium Interface. *Cell* **163,** 1484–1499 (2015).

bioRxiv preprint doi: https://doi.org/10.1101/298950; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

50. Megraw, T. L., Sharkey, J. T. & Nowakowski, R. S. Cdk5rap2 exposes the centrosomal root of microcephaly syndromes. *Trends Cell Biol.* **21,** 470–480 (2011).
51. Puga, A., Xia, Y. & Elferink, C. Role of the aryl hydrocarbon receptor in cell cycle regulation. *Chem. Biol. Interact.* **141,** 117–130 (2002).
52. Bader, J. R. *et al.* Polo-like Kinase1 Is Required for Recruitment of Dynein to Kinetochores during Mitosis. *J. Biol. Chem.* **286,** 20769–20777 (2011).
53. Safran, M. *et al.* GeneCards Version 3: the human gene integrator. *Database* **2010,** baq020-baq020 (2010).
54. Bailey, M. L., Singh, T., Mero, P., Moffat, J. & Hieter, P. Dependence of Human Colorectal Cells Lacking the FBW7 Tumor Suppressor on the Spindle Assembly Checkpoint. *Genetics* **201,** 885 LP-895 (2015).
55. Stocco dos Santos, R. C. *et al.* Stocco dos Santos X-linked mental retardation syndrome: Clinical elucidation and localization to Xp11.3–Xq21.3. *Am. J. Med. Genet.* **118A,** 255–259 (2003).
56. Yoder, M. & Hildebrand, J. D. Shroom4 (Kiaa1202) is an actin-associated protein implicated in cytoskeletal organization. *Cell Motil. Cytoskeleton* **64,** 49–63 (2007).
57. Zachos, G. *et al.* Chk1 Is Required for Spindle Checkpoint Function. *Dev. Cell* **12,** 247–260 (2017).
58. St-Denis, N. *et al.* Phenotypic and Interaction Profiling of the Human Phosphatases Identifies Diverse Mitotic Regulators. *Cell Rep.* **17,** 2488–2501 (2017).
59. Melkerson-Watson, L. J., Waldmann, M. E., Gunter, A. D., Zaroukian, M. H. & Esselman, W. J. Elevation of lymphocyte CD45 protein tyrosine phosphatase activity during mitosis. *J. Immunol.* **153,** 2004 LP-2013 (1994).
60. Thiru, P. *et al.* Kinetochore genes are coordinately up-regulated in human tumors as part of a FoxM1-related cell division program. *Mol. Biol. Cell* **25,** 1983–1994 (2014).
61. Callan, M. A. *et al.* Fragile X protein controls neural stem cell proliferation in the Drosophila brain . *Hum. Mol. Genet.* **19,** 3068–3079 (2010).
62. Yoshida, M. M. & Azuma, Y. Mechanisms behind Topoisomerase II SUMOylation in chromosome segregation. *Cell Cycle* **15,** 3151–3152 (2016).
63. Dephoure, N. *et al.* A quantitative atlas of mitotic phosphorylation. *Proc. Natl. Acad. Sci.* **105,** 10762–10767 (2008).
64. Rusin, S. F., Schlosser, K. A., Adamo, M. E. & Kettenbach, A. N. Quantitative phosphoproteomics reveals new roles for the protein phosphatase PP6 in mitotic cells. *Sci. Signal.* **8,** rs12 LP-rs12 (2015).
65. Zeng, K., Bastos, R. N., Barr, F. A. & Gruneberg, U. Protein phosphatase 6 regulates mitotic spindle formation by controlling the T-loop phosphorylation state of Aurora A bound to its activator TPX2. *J. Cell Biol.* **191,** 1315 LP-1332 (2010).
66. Gai, M. *et al.* ASPM and CITK regulate spindle orientation by affecting the dynamics of astral microtubules. *EMBO Rep.* **17,** 1396–1409 (2016).
67. Pfenning, A. R. *et al.* Convergent transcriptional specializations in the brains of humans and song-learning birds. *Science (80-.).* **346,** (2014).
68. Kuhlwilm, M. The evolution of FOXP2 in the light of admixture *(manuscript under review).*
69. Lüders, J. *The Microtubule Cytoskeleton.* (Springer Vienna, 2016). doi:10.1007/978-3-7091-1903-7
70. McVicker, D. P. *et al.* Transport of a kinesin-cargo pair along microtubules into dendritic spines undergoing synaptic plasticity. *Nat. Commun.* **7,** 12741 (2016).
71. Ahmad, F. J., Yu, W., McNally, F. J. & Baas, P. W. An Essential Role for Katanin in Severing Microtubules in the Neuron. *J. Cell Biol.* **145,** 305 LP-315 (1999).
72. Karabay, A., Yu, W., Solowska, J. M., Baird, D. H. & Baas, P. W. Axonal Growth Is Sensitive to the Levels of Katanin, a Protein That Severs Microtubules. *J. Neurosci.* **24,** 5778 LP-5788 (2004).

bioRxiv preprint doi: https://doi.org/10.1101/298950; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

73. Kevenaar, J. T. *et al.* Kinesin-Binding Protein Controls Microtubule Dynamics and Cargo Trafficking by Regulating Kinesin Motor Activity. *Curr. Biol.* **26,** 849–861 (2017).
74. Wu, Y. *et al.* Versican V1 Isoform Induces Neuronal Differentiation and Promotes Neurite Outgrowth. *Mol. Biol. Cell* **15,** 2093–2104 (2004).
75. Cuberos, H. *et al.* Roles of LIM kinases in central nervous system function and dysfunction. *FEBS Lett.* **589,** 3795–3806 (2015).
76. Koser, D. E. *et al.* Mechanosensing is critical for axon growth in the developing brain. *Nat. Neurosci.* **19,** 1592 (2016).
77. Jensen, K. B. *et al.* Nova-1 Regulates Neuron-Specific Alternative Splicing and Is Essential for Neuronal Viability. *Neuron* **25,** 359–371 (2000).
78. Ratti, A. *et al.* Post-transcriptional regulation of neuro-oncological ventral antigen 1 by the neuronal RNA-binding proteins ELAV. *J. Biol. Chem.* (2008). doi:10.1074/jbc.M706082200
79. Takemoto, M. *et al.* Laminar and Areal Expression of Unc5d and Its Role in Cortical Cell Survival. *Cereb. Cortex* **21,** 1925–1934 (2011).
80. Ango, F. *et al.* Ankyrin-Based Subcellular Gradient of Neurofascin, an Immunoglobulin Family Protein, Directs GABAergic Innervation at Purkinje Axon Initial Segment. *Cell* **119,** 257–272 (2004).
81. Pollerberg, G. E., Thelen, K., Theiss, M. O. & Hochlehnert, B. C. The role of cell adhesion molecules for navigating axons: Density matters. *Mech. Dev.* **130,** 359–372 (2013).
82. Yap, C. C. *et al.* Doublecortin (DCX) Mediates Endocytosis of Neurofascin Independently of Microtubule Binding. *J. Neurosci.* **32,** 7439 LP-7453 (2012).
83. Harrington, E. P. *et al.* Oligodendrocyte PTEN is required for myelin and axonal integrity, not remyelination. *Ann. Neurol.* **68,** 703–716 (2010).
84. Ryu, E. J. *et al.* Analysis of Peripheral Nerve Expression Profiles Identifies a Novel Myelin Glycoprotein, MP11. *J. Neurosci.* **28,** 7563 LP-7573 (2008).
85. Verheijen, M. H. G. *et al.* SCAP is required for timely and proper myelin membrane synthesis. *Proc. Natl. Acad. Sci.* **106,** 21383–21388 (2009).
86. Menzies, F. M., Fleming, A. & Rubinsztein, D. C. Compromised autophagy and neurodegenerative diseases. *Nat. Rev. Neurosci.* **16,** 345 (2015).
87. Hor, H. *et al.* Missense mutations in TENM4, a regulator of axon guidance and central myelination, cause essential tremor. *Hum. Mol. Genet.* **24,** 5677–5686 (2015).
88. Hsu, L.-S., Liang, C.-J., Tseng, C.-Y., Yeh, C.-W. & Tsai, J.-N. Zebrafish Cyclin-Dependent Protein Kinase–Like 1 (zcdkl1): Identification and Functional Characterization. *Int. J. Mol. Sci.* **12,** 3606–3617 (2011).
89. Jurecka, A., Jurkiewicz, E. & Tylki-Szymanska, A. Magnetic resonance imaging of the brain in adenylosuccinate lyase deficiency: a report of seven cases and a review of the literature. *Eur. J. Pediatr.* **171,** 131–138 (2012).
90. Kearns, C. A., Ravanelli, A. M., Cooper, K. & Appel, B. Fbxw7 Limits Myelination by Inhibiting mTOR Signaling. *J. Neurosci.* **35,** 14861 LP-14871 (2015).
91. Morfini, G. A. *et al.* Axonal Transport Defects in Neurodegenerative Diseases. *J. Neurosci.* **29,** 12776 LP-12786 (2009).
92. Butler, M. H. *et al.* Amphiphysin II (SH3P9; BIN1), a Member of the Amphiphysin/Rvs Family, Is Concentrated in the Cortical Cytomatrix of Axon Initial Segments and Nodes of Ranvier in Brain and around T Tubules in Skeletal Muscle. *J. Cell Biol.* **137,** 1355–1367 (1997).
93. Huang, B. *et al.* Mutant Huntingtin Downregulates Myelin Regulatory Factor-Mediated Myelin Gene Expression and Affects Mature Oligodendrocytes. *Neuron* **85,** 1212–1226 (2015).
94. Widagdo, J., Fang, H., Jang, S. E. & Anggono, V. PACSIN1 regulates the dynamics of AMPA receptor trafficking. *Sci. Rep.* **6,** 31070 (2016).
95. Arranz, A. M. *et al.* LRRK2 functions in synaptic vesicle endocytosis through a kinase-dependent mechanism. *J. Cell Sci.* **128,** 541 LP-552 (2015).

bioRxiv preprint doi: https://doi.org/10.1101/298950; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

96. Craig, T. J., Anderson, D., Evans, A. J., Girach, F. & Henley, J. M. SUMOylation of Syntaxin1A regulates presynaptic endocytosis. *Sci. Rep.* **5,** 17669 (2015).
97. Razafsky, D. *et al.* Lamin B1 and lamin B2 are long-lived proteins with distinct functions in retinal development. *Mol. Biol. Cell* **27,** 1928–1937 (2016).
98. Janz, R. & Südhof, T. C. SV2C is a synaptic vesicle protein with an unusually restricted localization: anatomy of a synaptic vesicle protein family. *Neuroscience* **94,** 1279–1290 (1999).
99. Lee, J. & Littleton, J. T. Transmembrane tethering of synaptotagmin to synaptic vesicles controls multiple modes of neurotransmitter release. *Proc. Natl. Acad. Sci.* **112,** 3793–3798 (2015).
100. Liu, X. *et al.* Extension of cortical synaptic development distinguishes humans from chimpanzees and macaques. *Genome Res.* **22,** 611–622 (2012).
101. Sekiguchi, M. *et al.* Identification of amphiphysin 1 as an endogenous substrate for CDKL5, a protein kinase associated with X-linked neurodevelopmental disorder. *Arch. Biochem. Biophys.* **535,** 257–267 (2013).
102. Trazzi, S. *et al.* HDAC4: a key factor underlying brain developmental alterations in CDKL5 disorder. *Hum. Mol. Genet.* **25,** 3887–3907 (2016).
103. Miska, E. A. *et al.* HDAC4 deacetylase associates with and represses the MEF2 transcription factor. *EMBO J.* **18,** 5099 LP-5107 (1999).
104. Somel, M., Rohlfs, R. & Liu, X. Transcriptomic insights into human brain evolution: acceleration, neutrality, heterochrony. *Curr. Opin. Genet. Dev.* **29,** 110–119 (2014).
105. Ryan, T. A. A pre-synaptic to-do list for coupling exocytosis to endocytosis. *Curr. Opin. Cell Biol.* **18,** 416–421 (2006).
106. Kornilov, S. A. *et al.* Genome-Wide Association and Exome Sequencing Study of Language Disorder in an Isolated Population. *Pediatrics* (2016).
107. Li, F.-F. *et al.* Identification of two novel critical mutations in PCNT gene resulting in microcephalic osteodysplastic primordial dwarfism type II associated with multiple intracranial aneurysms. *Metab. Brain Dis.* **30,** 1387–1394 (2015).
108. Tungadi, E. A., Ito, A., Kiyomitsu, T. & Goshima, G. Human microcephaly ASPM protein is a spindle pole-focusing factor that functions redundantly with CDK5RAP2. *J. Cell Sci.* **130,** 3676 LP-3684 (2017).
109. Bianchi, F. T. *et al.* Citron Kinase Deficiency Leads to Chromosomal Instability and TP53-Sensitive Microcephaly. *Cell Rep.* **18,** 1674–1686 (2017).
110. Jiang, K. *et al.* Microtubule minus-end regulation at spindle poles by an ASPM–katanin complex. *Nat. Cell Biol.* **19,** 480 (2017).
111. Yigit, G. *et al.* A syndrome of microcephaly, short stature, polysyndactyly, and dental anomalies caused by a homozygous KATNB1 mutation. *Am. J. Med. Genet. Part A* **170,** 728–733 (2016).
112. Genin, A. *et al.* Kinetochore KMN network gene CASC5 mutated in primary microcephaly. *Hum. Mol. Genet.* **21,** 5306–5317 (2012).
113. Arroyo, M. *et al.* MCPH1, mutated in primary microcephaly, is required for efficient chromosome alignment during mitosis. *Sci. Rep.* **7,** 13019 (2017).
114. Ritchie, K., Watson, L. A., Davidson, B., Jiang, Y. & Bérubé, N. G. ATRX is required for maintenance of the neuroprogenitor cell pool in the embryonic mouse brain. *Biol. Open* **3,** 1158 LP-1163 (2014).
115. El Waly, B., Buhler, E., Haddad, M.-R. & Villard, L. Nhej1 Deficiency Causes Abnormal Development of the Cerebral Cortex. *Mol. Neurobiol.* **52,** 771–782 (2015).
116. Kodani, A. *et al.* Centriolar satellites assemble centrosomal microcephaly proteins to recruit CDK2 and promote centriole duplication. *Elife* **4,** e07519 (2015).
117. Asami, M. *et al.* The role of Pax6 in regulating the orientation and mode of cell division of progenitors in the mouse cerebral cortex. *Development* **138,** 5067 LP-5078 (2011).

bioRxiv preprint doi: https://doi.org/10.1101/298950; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

118. Jurecka, A., Zikanova, M., Kmoch, S. & Tylki-Szymańska, A. Adenylosuccinate lyase deficiency. *J. Inherit. Metab. Dis.* **38,** 231–242 (2015).
119. Suzuki, R., Ferris, H. A., Chee, M. J., Maratos-Flier, E. & Kahn, C. R. Reduction of the Cholesterol Sensor SCAP in the Brains of Mice Causes Impaired Synaptic Transmission and Altered Cognitive Function. *PLOS Biol.* **11,** e1001532 (2013).
120. Lian, G. *et al.* Filamin A- and formin 2-dependent endocytosis regulates proliferation via the canonical Wnt pathway. *Development* **143,** 4509 LP-4520 (2016).
121. Montgomery, S. H., Capellini, I., Venditti, C., Barton, R. A. & Mundy, N. I. Adaptive Evolution of Four Microcephaly Genes and the Evolution of Brain Size in Anthropoid Primates. *Mol. Biol. Evol.* **28,** 625–638 (2011).
122. Shi, L. *et al.* Regional selection of the brain size regulating gene CASC5 provides new insight into human brain evolution. *Hum. Genet.* **136,** 193–204 (2017).
123. Chen, Y., Huang, W.-C., Séjourné, J., Clipperton-Allen, A. E. & Page, D. T. Pten Mutations Alter Brain Growth Trajectory and Allocation of Cell Types through Elevated β-Catenin Signaling. *J. Neurosci.* **35,** 10252 LP-10267 (2015).
124. Mora-Bermúdez, F. *et al.* Differences and similarities between human and chimpanzee neural progenitors during cerebral cortex development. *Elife* **5,** e18683 (2016).
125. Kuwahara, A. *et al.* Tcf3 Represses Wnt–β-Catenin Signaling and Maintains Neural Stem Cell Population during Neocortical Development. *PLoS One* **9,** e94408 (2014).
126. Douglas, J. *et al.* Mutations in RNF135, a gene within the NF1 microdeletion region, cause phenotypic abnormalities including overgrowth. *Nat. Genet.* **39,** 963 (2007).
127. Tarpey, P. S. *et al.* Mutations in CUL4B, Which Encodes a Ubiquitin E3 Ligase Subunit, Cause an X-linked Mental Retardation Syndrome Associated with Aggressive Outbursts, Seizures, Relative Macrocephaly, Central Obesity, Hypogonadism, Pes Cavus, and Tremor. *Am. J. Hum. Genet.* **80,** 345–352 (2007).
128. Mirzaa, G. M. *et al.* De novo CCND2 mutations leading to stabilization of cyclin D2 cause megalencephaly-polymicrogyria-polydactyly-hydrocephalus syndrome. *Nat. Genet.* **46,** 510 (2014).
129. Klaassens, M. *et al.* Malan syndrome: Sotos-like overgrowth with de novo NFIX sequence variants and deletions in six new patients and a review of the literature. *Eur. J. Hum. Genet.* **23,** 610 (2014).
130. Buxbaum, J. D. *et al.* Mutation analysis of the NSD1 gene in patients with autism spectrum disorders and macrocephaly. *BMC Med. Genet.* **8,** 68 (2007).
131. Jamsheer, A. *et al.* Expanded mutational spectrum of the GLI3 gene substantiates genotype–phenotype correlations. *J. Appl. Genet.* **53,** 415–422 (2012).
132. Gokhman, D. *et al.* Extensive Regulatory Changes in Genes Affecting Vocal and Facial Anatomy Separate Modern from Archaic Humans. *bioRxiv* (2017).
133. Stahl, R. *et al.* Trnp1 Regulates Expansion and Folding of the Mammalian Cerebral Cortex by Control of Radial Glial Fate. *Cell* **153,** 535–549 (2013).
134. Liu, J. *et al.* The Primate-Specific Gene TMEM14B Marks Outer Radial Glia Cells and Promotes Cortical Expansion and Folding. *Cell Stem Cell* **21,** 635–649.e8 (2017).
135. Martínez-Martínez, M. Á. *et al.* A restricted period for formation of outer subventricular zone defined by Cdh1 and Trnp1 levels. *Nat. Commun.* **7,** 11812 (2016).
136. Dehay, C., Kennedy, H. & Kosik, K. S. The Outer Subventricular Zone and Primate-Specific Cortical Complexification. *Neuron* **85,** 683–694 (2015).
137. Nowakowski, T. J. *et al.* Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. *Science (80-.).* **358,** 1318 LP-1323 (2017).
138. Li, Y. *et al.* Induction of Expansion and Folding in Human Cerebral Organoids. *Cell Stem Cell* **20,** 385–396.e3 (2017).
139. Cloutier, P. *et al.* R2TP/Prefoldin-like component RUVBL1/RUVBL2 directly interacts with ZNHIT2 to regulate assembly of U5 small nuclear ribonucleoprotein. *Nat. Commun.* **8,** 15615 (2017).

bioRxiv preprint doi: https://doi.org/10.1101/298950; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

140. Wang, C., Liang, C.-C., Bian, Z. C., Zhu, Y. & Guan, J.-L. FIP200 is required for maintenance and differentiation of postnatal neural stem cells. *Nat. Neurosci.* **16,** 532 (2013).
141. Weyer, S. & Pääbo, S. Functional Analyses of Transcription Factor Binding Sites that Differ between Present-Day and Archaic Humans. *Mol. Biol. Evol.* **33,** 316–322 (2016).
142. Pereira-Pedro, A. S., Masters, M. & Bruner, E. Shape analysis of spatial relationships between orbito-ocular and endocranial structures in modern humans and fossil hominids. *J. Anat.* **231,** 947–960 (2017).
143. Basel-Vanagaite, L. *et al.* RIN2 Deficiency Results in Macrocephaly, Alopecia, Cutis Laxa, and Scoliosis: MACS Syndrome. *Am. J. Hum. Genet.* **85,** 254–263 (2009).
144. Blank, M. C. *et al.* Multiple developmental programs are altered by loss of Zic1 and Zic4 to cause Dandy-Walker malformation cerebellar pathogenesis. *Development* **138,** 1207 LP-1216 (2011).
145. Hoshino, J., Aruga, J., Ishiguro, A. & Mikoshiba, K. Dorz1, a novel gene expressed in differentiating cerebellar granule neurons, is down-regulated in Zic1-deficient mouse. *Mol. Brain Res.* **120,** 57–64 (2003).
146. Wilkins, A. S., Wrangham, R. W. & Fitch, W. T. The 'Domestication Syndrome' in Mammals: A Unified Explanation Based on Neural Crest Cell Behavior and Genetics. *Genetics* **197,** 795 LP-808 (2014).
147. Theofanopoulou, C. *et al.* Self-domestication in Homo sapiens: Insights from comparative genomics. *PLoS One* **12,** e0185306 (2017).
148. Siegel, D. A., Huang, M. K. & Becker, S. F. Ectopic dendrite initiation: CNS pathogenesis as a model of CNS development. *Int. J. Dev. Neurosci.* **20,** 373–389 (2002).
149. Park, J.-A. & Kim, K. C. Expression patterns of PRDM10 during mouse embryonic development. *BMB Rep.* **43,** 29–33 (2010).
150. Lieberman, P. Primate Vocalizations and Human Linguistic Ability. *J. Acoust. Soc. Am.* **44,** 1574–1584 (1968).
151. Adhikari, K. *et al.* A genome-wide association scan implicates DCHS2, RUNX2, GLI3, PAX1 and EDAR in human facial variation. *Nat. Commun.* **7,** 11616 (2016).
152. Kim, J. H. & Kim, N. Regulation of NFATc1 in Osteoclast Differentiation. *J Bone Metab* **21,** 233–241 (2014).
153. Cai, H. & Liu, A. Spop promotes skeletal development and homeostasis by positively regulating Ihh signaling. *Proc. Natl. Acad. Sci.* **113,** 14751–14756 (2016).
154. Zhang, Y. *et al.* An essential role of discoidin domain receptor 2 (DDR2) in osteoblast differentiation and chondrocyte maturation via modulation of Runx2 activation. *J. Bone Miner. Res.* **26,** 604–617 (2011).
155. Zhang, X. *et al.* Calvarial Cleidocraniodysplasia-Like Defects With ENU-Induced Nell-1 Deficiency. *J. Craniofac. Surg.* **23,** (2012).
156. Moncini, S. *et al.* ATRX mutation in two adult brothers with non-specific moderate intellectual disability identified by exome sequencing. *Meta Gene* **1,** 102–108 (2013).
157. Oh, J.-E. *et al.* PlexinA2 mediates osteoblast differentiation via regulation of Runx2. *J. Bone Miner. Res.* **27,** 552–562 (2012).
158. Kwon, E. K. *et al.* The Role of Ellis-Van Creveld 2(*EVC2*) in Mice During Cranial Bone Development. *Anat. Rec.* **301,** 46–55 (2018).
159. Gullard, A. *et al.* MEPE Localization in the Craniofacial Complex and Function in Tooth Dentin Formation. *J. Histochem. Cytochem.* **64,** 224–236 (2016).
160. Teves, M. E. *et al.* Spag17 Deficiency Results in Skeletal Malformations and Bone Abnormalities. *PLoS One* **10,** e0125936 (2015).
161. Lacruz, R. S. *et al.* Ontogeny of the maxilla in Neanderthals and their ancestors. *Nat. Commun.* **6,** 8996 (2015).
162. Lacruz, R. S. *et al.* Facial Morphogenesis of the Earliest Europeans. *PLoS One* **8,** e65199 (2013).

bioRxiv preprint doi: https://doi.org/10.1101/298950; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

163. Stringer, C. The origin and evolution of Homo sapiens *Philos. Trans. R. Soc. B Biol. Sci.* **371,** (2016).
164. Jones, D. C. *et al.* Uncoupling of growth plate maturation and bone formation in mice lacking both Schnurri-2 and Schnurri-3. *Proc. Natl. Acad. Sci.* **107,** 8254–8258 (2010).
165. Imamura, K. *et al.* Human Immunodeficiency Virus Type 1 Enhancer-binding Protein 3 Is Essential for the Expression of Asparagine-linked Glycosylation 2 in the Regulation of Osteoblast and Chondrocyte Differentiation. *J. Biol. Chem.* **289,** 9865–9879 (2014).
166. Lee, M. K. *et al.* Genome-wide association study of facial morphology reveals novel associations with FREM1 and PARK2. *PLoS One* **12,** e0176566 (2017).
167. Talbot, J. C. *et al.* Pharyngeal morphogenesis requires fras1-itga8-dependent epithelial-mesenchymal interaction. *Dev. Biol.* **416,** 136–148 (2016).
168. Verzi, M. P. *et al.* The Transcription Factor MEF2C Is Required for Craniofacial Development. *Dev. Cell* **12,** 645–652 (2007).
169. Rak, Y. The Neanderthal: A new look at an old face. *J. Hum. Evol.* **15,** 151–164 (1986).
170. Skeide, M. A. & Friederici, A. D. The ontogeny of the cortical language network. *Nat. Rev. Neurosci.* **17,** 323 (2016).
171. Grosse Wiesmann, C., Schreiber, J., Singer, T., Steinbeis, N. & Friederici, A. D. White matter maturation is associated with the emergence of Theory of Mind in early childhood. *Nat. Commun.* **8,** 14692 (2017).
172. Mars, R. B., Eichert, N., Jbabdi, S., Verhagen, L. & Rushworth, M. F. S. Connectivity and the search for specializations in the language-capable brain. *Curr. Opin. Behav. Sci.* **21,** 19–26 (2018).
173. Dubois, J. *et al.* Exploring the Early Organization and Maturation of Linguistic Pathways in the Human Infant Brain. *Cereb. Cortex* **26,** 2283–2298 (2016).
174. Bastir, M. *et al.* Evolution of the base of the brain in highly encephalized human species. *Nat. Commun.* **2,** 588 (2011).
175. Bryant, K. L. & Preuss, T. M. A Comparative Perspective on the Human Temporal Lobe BT - Digital Endocasts: From Skulls to Brains. in (eds. Bruner, E., Ogihara, N. & Tanabe, H. C.) 239–258 (Springer Japan, 2018). doi:10.1007/978-4-431-56582-6_16
176. Dediu, D. & Levinson, S. C. On the antiquity of language: The reinterpretation of neandertal linguistic capacities and its consequences. *Front. Psychol.* **4,** 1–17 (2013).
177. Dediu, D. & Levinson, S. C. Neanderthal language revisited: not only us. *Curr. Opin. Behav. Sci.* **21,** 49–55 (2018).
178. Fu, Q. *et al.* An early modern human from Romania with a recent Neanderthal ancestor. *Nature* **524,** 216–219 (2015).
179. Jeong, J.-W., Sundaram, S., Behen, M. E. & Chugani, H. T. Relationship between genotype and arcuate fasciculus morphology in six young children with global developmental delay: Preliminary DTI stuy. *J. Magn. Reson. Imaging* **44,** 1504–1512 (2016).
180. Ptak, S. E. *et al.* Linkage Disequilibrium Extends Across Putative Selected Sites in FOXP2. *Mol. Biol. Evol.* **26,** 2181–2184 (2009).
181. Maricic, T. *et al.* A Recent Evolutionary Change Affects a Regulatory Element in the Human FOXP2 Gene. *Mol. Biol. Evol.* **30,** 844–852 (2013).
182. Belyk, M. & Brown, S. The origins of the vocal brain in humans. *Neurosci. Biobehav. Rev.* **77,** 177–193 (2017).
183. Vargha-Khadem, F., Gadian, D. G., Copp, A. & Mishkin, M. FOXP2 and the neuroanatomy of speech and language. *Nat. Rev. Neurosci.* **6,** 131 (2005).
184. Enard, W. *et al.* A Humanized Version of Foxp2 Affects Cortico-Basal Ganglia Circuits in Mice. *Cell* **137,** 961–971 (2009).
185. Hachigian, L. J. *et al.* Control of Huntington’s Disease-Associated Phenotypes by the Striatum-Enriched Transcription Factor Foxp2. *Cell Rep.* **21,** 2688–2695 (2017).
186. Abelson, J. F. *et al.* Sequence Variants in SLITRK1 Are Associated with Tourette's Syndrome. *Science (80-.).* **310,** 317 LP-320 (2005).

bioRxiv preprint doi: https://doi.org/10.1101/298950; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

187. Jelen, N., Ule, J. & Živin, M. Cholinergic regulation of striatal Nova mRNAs. *Neuroscience* **169,** 619–627 (2010).
188. Alsiö, J. M., Tarchini, B., Cayouette, M. & Livesey, F. J. Ikaros promotes early-born neuronal fates in the cerebral cortex. *Proc. Natl. Acad. Sci.* **110,** E716–E725 (2013).
189. Popovitchenko, T. *et al.* The RNA binding protein HuR determines the differential translation of autism-associated FoxP subfamily members in the developing neocortex. *Sci. Rep.* **6,** 28998 (2016).
190. Konopka, G. *et al.* Human-Specific Transcriptional Networks in the Brain. *Neuron* **75,** 601–617 (2012).
191. Dunn, A. R. *et al.* Synaptic vesicle glycoprotein 2C (SV2C) modulates dopamine release and is disrupted in Parkinson disease. *Proc. Natl. Acad. Sci.* **114,** E2253–E2262 (2017).
192. Oswald, F. *et al.* The FOXP2-Driven Network in Developmental Disorders and Neurodegeneration . *Frontiers in Cellular Neuroscience* **11,** 212 (2017).
193. Hilliard, A. T., Miller, J. E., Fraley, E. R., Horvath, S. & White, S. A. Molecular Microcircuitry Underlies Functional Specification in a Basal Ganglia Circuit Dedicated to Vocal Learning. *Neuron* **73,** 537–552 (2012).
194. Chen, X. S. *et al.* Next-generation DNA sequencing identifies novel gene variants and pathways involved in specific language impairment. *Sci. Rep.* **7,** 46105 (2017).
195. Boeckx, C. & Benítez-Burraco, A. Globularity and language-readiness: generating new predictions by expanding the set of genes of interest . *Frontiers in Psychology* **5,** 1324 (2014).
196. Ferrarelli, L. K. PTEN contributes to Alzheimer's disease. *Sci. Signal.* **9,** ec45 LP-ec45 (2016).
197. Chano, T., Okabe, H. & Hulette, C. M. RB1CC1 insufficiency causes neuronal atrophy through mTOR signaling alteration and involved in the pathology of Alzheimer's diseases. *Brain Res.* **1168,** 97–105 (2007).
198. Choi, S. *et al.* Effects of a Polymorphism of the Neuronal Amino Acid Transporter SLC6A15 Gene on Structural Integrity of White Matter Tracts in Major Depressive Disorder. *PLoS One* **11,** e0164301 (2016).
199. Santarelli, S. *et al.* SLC6A15, a novel stress vulnerability candidate, modulates anxiety and depressive-like behavior: involvement of the glutamatergic system. *Stress* **19,** 83–90 (2016).
200. Sreedharan, S. *et al.* The G protein coupled receptor Gpr153 shares common evolutionary origin with Gpr162 and is highly expressed in central regions including the thalamus, cerebellum and the arcuate nucleus. *FEBS J.* **278,** 4881–4894 (2011).
201. Fon, E. A. *et al.* Adenylosuccinate lyase (ADSL) and infantile autism: Absence of previously reported point mutation. *Am. J. Med. Genet.* **60,** 554–557 (1995).
202. Racimo, F., Kuhlwilm, M. & Slatkin, M. A test for ancient selective sweeps and an application to candidate sites in modern humans. *Mol. Biol. Evol.* **31,** 3344–3358 (2014).
203. Takata, A. *et al.* Integrative Analyses of De Novo Mutations Provide Deeper Biological Insights into Autism Spectrum Disorder. *Cell Rep.* **22,** 734–747 (2018).
204. Somel, M. *et al.* Transcriptional neoteny in the human brain. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 5743–8 (2009).
205. Sherwood, C. C. & Gómez-Robles, A. Brain Plasticity and Human Evolution. *Annu. Rev. Anthropol.* **46,** 399–419 (2017).
206. Bednarik, G. R. From Human Past to Human Future. *Humanities* **2,** (2013).
207. Hofman, M. a. Encephalization in hominids: evidence for the model of punctuationalism. *Brain. Behav. Evol.* (1983). doi:10.1159/000121511
208. Ponce de León, M. S., Bienvenu, T., Akazawa, T. & Zollikofer, C. P. E. Brain development is similar in Neanderthals and modern humans. *Curr. Biol.* **26,** R665–R666 (2017).
209. Rosas, A. *et al.* The growth pattern of Neandertals, reconstructed from a juvenile skeleton from El Sidrón (Spain). *Science (80-.).* **357,** 1282 LP-1287 (2017).

bioRxiv preprint doi: https://doi.org/10.1101/298950; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

210. Gunz, P., Neubauer, S., Maureille, B. & Hublin, J. J. Brain development after birth differs between Neandertals and modern humans. *Curr. Biol.* **20,** 921–922 (2010).
211. Bruner, E., de la Cuétara, J. M., Masters, M., Amano, H. & Ogihara, N. Functional craniology and brain evolution: from paleontology to biomedicine. *Front. Neuroanat.* **8,** 19 (2014).
212. Beer, J. S., John, O. P., Scabini, D. & Knight, R. T. Orbitofrontal Cortex and Social Behavior: Integrating Self-monitoring and Emotion-Cognition Interactions. *J. Cogn. Neurosci.* **18,** 871–879 (2006).
213. Miller, K. J., Botvinick, M. M. & Brody, C. D. Value Representations in Orbitofrontal Cortex Drive Learning, but not Choice. *bioRxiv* (2018).
214. Smith, B. H. & Tompkins, R. L. Toward A Life History of the Hominidae. *Annu. Rev. Anthropol.* **24,** 257–279 (1995).
215. Bock, J. & Sellen, D. W. Childhood and the evolution of the human life course. *Hum. Nat.* **13,** 153–159 (2002).
216. Smith, T. M. *et al.* Dental evidence for ontogenetic differences between modern humans and Neanderthals. *Proc Natl Acad Sci U S A* **107,** 20923–20928 (2010).
217. Ramirez Rozzi, F. V & de Castro, J. M. surprisingly rapid growth in Neanderthals. *Nature* **428,** (2004).
218. Chintalapati, M., Dannemann, M. & Prüfer, K. Using the Neandertal genome to study the evolution of small insertions and deletions in modern humans. *BMC Evol. Biol.* **17,** 179 (2017).
219. Day, F. R. *et al.* Genomic analyses identify hundreds of variants associated with age at menarche and support a role for puberty timing in cancer risk. *Nat. Genet.* **49,** 834–841 (2017).
220. Antonson, P. *et al.* Inactivation of the Nuclear Receptor Coactivator RAP250 in Mice Results in Placental Vascular Dysfunction. *Mol. Cell. Biol.* **23,** 1260–1268 (2003).
221. Zhu, G.-Z., Lin, Y., Myles, D. G. & Primakoff, P. Identification of four novel ADAMs with potential roles in spermatogenesis and fertilization. *Gene* **234,** 227–237 (1999).
222. Tang, S. *et al.* Biallelic Mutations in CFAP43 and CFAP44 Cause Male Infertility with Multiple Morphological Abnormalities of the Sperm Flagella. *Am. J. Hum. Genet.* **100,** 854–864 (2017).
223. Wang, Z., Widgren, E. E., Sivashanmugam, P., O'Rand, M. G. & Richardson, R. T. Association of Eppin with Semenogelin on Human Spermatozoa1. *Biol. Reprod.* **72,** 1064–1070 (2005).
224. Ellerman, D. A. *et al.* Izumo is part of a multiprotein family whose members form large complexes on mammalian sperm. *Mol. Reprod. Dev.* **76,** 1188–1199 (2009).
225. Simonti, C. N. *et al.* The phenotypic legacy of admixture between modern humans and Neandertals. *Science (80-.).* **351,** 737 LP-741 (2016).
226. Dannemann, M., Prüfer, K. & Kelso, J. Functional implications of Neandertal introgression in modern humans. *Genome Biol.* **18,** 61 (2017).
227. Shi, J. *et al.* Sequence variations in the mu-opioid receptor gene (OPRM1) associated with human addiction to heroin. *Hum. Mutat.* **19,** 459–460 (2002).
228. Tan, E. *et al.* Ethnicity and OPRM variant independently predict pain perception and patient-controlled analgesia usage for post-operative pain. *Mol. Pain* **5,** 32 (2009).
229. Pearce, E., Wlodarski, R., Machin, A. & Dunbar, R. I. M. Variation in the β-endorphin, oxytocin, and dopamine receptor genes is associated with different dimensions of human sociality. *Proc. Natl. Acad. Sci.* **114,** 5300 LP-5305 (2017).
230. Habib, A. M. *et al.* A novel human pain insensitivity disorder caused by a point mutation in ZFHX2. *Brain* awx326-awx326 (2017).
231. Buiting, K. *et al.* C15orf2 and a novel noncoding transcript from the Prader–Willi/Angelman syndrome region show monoallelic expression in fetal brain. *Genomics* **89,** 588–595 (2007).
232. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol.* **17,** 122 (2016).

bioRxiv preprint doi: https://doi.org/10.1101/298950; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

233. Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* **4,** 1073–1081 (2009).
234. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7,** 248–9 (2010).
235. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* **46,** 310–315 (2014).
236. Ritchie, G. R. S., Dunham, I., Zeggini, E. & Flicek, P. Functional annotation of noncoding sequence variants. *Nat. Methods* **11,** 294–6 (2014).
237. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536,** 285–291 (2016).
238. Paten, B. *et al.* Genome-wide nucleotide-level mammalian ancestor reconstruction. *Genome Res.* **18,** 1829–1843 (2008).
239. Yan, G. *et al.* Genome sequencing and comparison of two nonhuman primate animal models, the cynomolgus and Chinese rhesus macaques. *Nat. Biotechnol.* **29,** 1019 (2011).
240. Sherry, S. T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **29,** 308–311 (2001).
241. Grantham, R. Amino Acid Difference Formula to Help Explain Protein Evolution. *Science (80-.).* **185,** (1974).
242. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27,** 2987–2993 (2011).
243. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26,** 841–842 (2010).
244. Huber, W. *et al.* Orchestrating high-throughput genomic analysis with Bioconductor. *Nat Meth* **12,** 115–121 (2015).
245. Lawrence, M., Gentleman, R. & Carey, V. rtracklayer: an R package for interfacing with genome browsers. *Bioinformatics* **25,** 1841–1842 (2009).
246. Durinck, S. *et al.* BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics* **21,** 3439–3440 (2005).
247. Zhang, H., Meltzer, P. & Davis, S. RCircos: an R package for Circos 2D track plots. *BMC Bioinformatics* **14,** 244 (2013).
248. Signorell, A. DescTools: Tools for Descriptive Statistics. (2017).
249. Prüfer, K. *et al.* FUNC: a package for detecting significant associations between gene sets and ontological annotations. *BMC Bioinformatics* **8,** 41 (2007).
250. Szklarczyk, D. *et al.* The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic Acids Res.* **45,** D362–D368 (2017).