EVOLUTIONARY BIOLOGY

Dosage analysis of the 7q11.23 Williams region identifies *BAZ1B* as a major human gene patterning the modern human face and underlying self-domestication

Matteo Zanella[1,2]*[†], Alessandro Vitriolo[1,2]*, Alejandro Andirko[3,4], Pedro Tiago Martins[3,4], Stefanie Sturm[3,4], Thomas O'Rourke[3,4], Magdalena Laugsch[5,6,7], Natascia Malerba[8], Adrianos Skaros[1,2], Sebastiano Trattaro[1,2], Pierre-Luc Germain[1,2,9], Marija Mihailovic[1,2], Giuseppe Merla[8], Alvaro Rada-Iglesias[5,10,11], Cedric Boeckx[3,4,12], Giuseppe Testa[1,2,13]‡

We undertook a functional dissection of chromatin remodeler BAZ1B in neural crest (NC) stem cells (NCSCs) from a uniquely informative cohort of typical and atypical patients harboring 7q11.23 copy number variants. Our results reveal a key contribution of BAZ1B to NCSC in vitro induction and migration, coupled with a crucial involvement in NC-specific transcriptional circuits and distal regulation. By intersecting our experimental data with new paleogenetic analyses comparing modern and archaic humans, we found a modern-specific enrichment for regulatory changes both in BAZ1B and its experimentally defined downstream targets, thereby providing the first empirical validation of the human self-domestication hypothesis and positioning BAZ1B as a master regulator of the modern human face. In so doing, we provide experimental evidence that the craniofacial and cognitive/behavioral phenotypes caused by alterations of the Williams-Beuren syndrome critical region can serve as a powerful entry point into the evolution of the modern human face and prosociality.

INTRODUCTION

Anatomically modern humans (AMHs) exhibit a suite of craniofacial and prosocial characteristics that are reminiscent of traits distinguishing domesticated species from their wild counterparts (*1–3*). This has led to the formulation of a self-domestication hypothesis according to which modern humans (*3*) went through a domestication process in the course of their evolution. Recent evidence, along with the well-warranted distinction between domestication and selective breeding (*4*), is also extending this notion to other species that might have undergone a self-domestication phase, such as cats, dogs, and bonobos (*3*). Thus, as self-domestication represents a special case of domestication, the most parsimonious hypothesis must posit the same core mechanisms to underlie both. For this reason, the self-domestication hypothesis also entails the prediction that key aspects of modern humans' anatomy and cognition can be illuminated by studies of the so-called "domestication syndrome," the

core set of domestication-related traits that was recently proposed to result from mild neural crest (NC) deficits (*5*). However, both the neurocristopathic basis of domestication and its extension to the evolution of AMHs remain to be tested experimentally.

Williams-Beuren syndrome [WBS; OMIM (Online Mendelian Inheritance in Man) 194050] and Williams-Beuren region duplication syndrome (7dupASD; OMIM 609757), caused respectively by the hemideletion or hemiduplication of 28 genes at the 7q11.23 region [WBS critical region (WBSCR)], represent a paradigmatic pair of neurodevelopmental conditions whose NC-related craniofacial dysmorphisms and cognitive/behavioral traits (*6*, *7*) bear directly on domestication-related traits relevant for AMHs (facial reduction and retraction, pronounced friendliness, and reduced reactive aggression) (fig. S1A). Structural variants in WBS genes, for example in the case of GTF2I and its paralogs, have been shown to underlie stereotypical hypersociability in domestic dogs and foxes (*8*, *9*).

Among the WBSCR genes, we focus here on the chromatin regulator *BAZ1B* (also known as Williams syndrome transcription factor, *WSTF*), on the basis of the following lines of evidence that implicate it in domestication-relevant craniofacial features: (i) its established role in NC maintenance and migration in *Xenopus laevis* and the craniofacial defects observed in knockout mice (*10*, *11*); (ii) the observation that its expression is affected by domestication-related events in canids (*12*); (iii) the first formulation of the neurocristopathic hypothesis of domestication, which included *BAZ1B* among the genes influencing NC development (*5*); (iv) the most comprehensive studies focusing on regions of the modern human genome associated with selective sweep signals compared to Neanderthals/Denisovans (hereafter "archaics") (*13*, *14*), one of which specifically included *BAZ1B* within the detected portions of the WBSCR; and (v) the thus far most detailed study systematically exploring high-frequency (HF) (>90%) changes in modern humans for which archaic humans carry the

[1]Department of Oncology and Hemato-Oncology, University of Milan, Milan, Italy. [2]Laboratory of Stem Cell Epigenetics, IEO, European Institute of Oncology, IRCCS, Milan, Italy. [3]University of Barcelona, Barcelona, Spain. [4]University of Barcelona Institute of Complex Systems (UBICS), Barcelona, Spain. [5]Center for Molecular Medicine Cologne (CMMC), University of Cologne, Cologne, Germany. [6]Institute of Human Genetics, University Hospital Cologne, Cologne, Germany. [7]Institute of Human Genetics, University Hospital Heidelberg, Heidelberg, Germany. [8]Division of Medical Genetics, Fondazione IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo, Foggia, Italy. [9]D-HEST Institute for Neuroscience, ETH Zürich, Switzerland. [10]Cluster of Excellence Cellular Stress Responses in Aging-Associated Diseases (CECAD), University of Cologne, Cologne, Germany. [11]Institute of Biomedicine and Biotechnology of Cantabria, University of Cantabria, Cantabria, Spain. [12]Catalan Institute for Advanced Studies and Research (ICREA), Barcelona, Spain. [13]Human Technopole, Center for Neurogenomics, Via Cristina Belgioioso 171, Milan, Italy.
*These authors contributed equally to this work.
†Present address: Evotec SE, Hamburg, Germany.
‡Corresponding author. Email: giuseppe.testa@unimi.it, giuseppe.testa@ieo.it, giuseppe.testa@htechnopole.it

ancestral state, which found *BAZ1B* enriched for mutations in modern humans (most of which fall in the regulatory regions of the gene) (*15*).

Our previous work had established the largest cohort of 7q11.23 patient-derived induced pluripotent stem cell (iPSC) lines and revealed major disease-relevant transcriptional dysregulation that was already apparent at the pluripotent state and became further exacerbated upon differentiation (*16*). Here, we first harness this resource to dissect the impact of *BAZ1B* dosage on the NC of patients with WBS and 7dupASD, both in terms of function (i.e., NC migration and induction) and of transcriptional and chromatin dysregulation, thereby defining the BAZ1B dosage–dependent circuits controlling the NC. Next, we apply these experimentally determined BAZ1B-dependent circuits underlying craniofacial morphogenesis to interrogate the evidence from paleogenomic analyses, which were thus far only of a correlative nature. We find major convergence between the BAZ1B control and the genes harboring regulatory changes in the modern human lineage. Together, the definition of the role of BAZ1B dosage in craniofacial neurocristopathy and its application to domestication-relevant paleogenomics demonstrate a major contribution of BAZ1B to the modern human face and offer experimental validation for the prediction at the heart of NC-based accounts of (self-) domestication: that the modern human face acquired its shape as an instance of mild neurocristopathy.

RESULTS

Establishment and validation of an extensive cohort of patient-specific BAZ1B-interfered NC stem cell lines

To dissect the role of BAZ1B in the craniofacial dysmorphisms that characterize WBS and 7dupASD, we started from our previous characterization of WBS patient– and 7dupASD patient–specific iPSC lines and differentiated derivatives (*16*) and selected a cohort of 11 NC stem cell (NCSC) lines (four from patients with WBS, three from patients with 7dupASD, and four from control individuals), which also represent the largest cohort of patient-specific NCSCs described so far. Given the centrality of the cranial NC for the development of the face, we first validated the cranial identity of our NCSC cohort by transcriptomic profiling through a manually curated gene expression signature (fig. S2A), confirming their suitability for the study of craniofacial dysregulations. We then knocked down BAZ1B via RNA interference in all lines across the three genetic conditions, including also NCSCs derived from a particularly informative patient with atypical WBS (hereafter atWBS) bearing a partial deletion of the region that spares *BAZ1B* and six additional genes (Fig. 1A) (*17*). To establish a high-resolution gradient of *BAZ1B* dosages, we selected two distinct short hairpin RNA (shRNA) against BAZ1B (i.e., sh1 and sh2) along with a scrambled shRNA sequence (hereafter scr) as negative control, for a total of 32 NCSC lines. Knockdown (KD) efficiency was evaluated at the RNA level by quantitative polymerase chain reaction (qPCR) (Fig. 1B and fig. S1C), confirming the attainment of the desired gradient with an overall reduction of about 40% for sh1 and 70% for sh2, as well as reduction at the protein level, as detected by Western blot (fig. S1E).

### BAZ1B dosage imbalance impairs NCSC migration and induction

NCSCs need to migrate to reach specific target regions in the developing embryo and give rise to distinct cell types and tissues, including craniofacial structures that are major areas of change in human

evolution. Since BAZ1B KD was shown to affect the migration of the NC in *X. laevis* and to promote cancer cell invasion in different lung cancer cell lines (*10, 18*), we hypothesized that the *BAZ1B* dosage imbalances entailed in the 7q11.23 syndromes could result in a defective regulation of NCSC migration and might underlie the NC-related alterations typical of patients with WBS and 7dupASD. To test this, we compared the migration properties of patient-specific BAZ1B KD NCSC lines (sh2) to their respective control NCSC line (scr) by the well-established wound-healing assay. The 7dupASD NCSC KD lines took longer to fill the wound when compared to the respective control lines (scr), as indicated by images taken at 8 and 16 hours after a gap was created on the plate surface (Fig. 1C and fig. S1F). We instead observed an opposite behavior for the WBS BAZ1B KD lines, which were faster than the respective scr lines in closing the gap (Fig. 1C and fig. S1F). In contrast to the previous observations from *X. laevis* (*10*), we also observed a minor delay in NC induction as a consequence of BAZ1B KD (Fig. 1D and fig. S1D), by means of a differentiation protocol based on NC delamination from adherent embryoid bodies (EBs), which recapitulates the initial steps of NC generation (*19*). In particular, starting from 2 to 3 days after attachment of EBs, we observed a lower number of outgrowing cells in the KD line (Fig. 1D, days 7 and 10), coupled with an evidently higher cell mortality. Cells were eventually able to acquire the typical NC morphology, although lower differentiation efficiency was evident, as shown by images taken at day 12. In addition, the delay in NC formation was associated with a down-regulation of well-established critical regulators of NC migration and maintenance, including *NR2F1*, *NR2F2*, *TFAP2A*, and *SOX9* (Fig. 1E). These results show that BAZ1B regulates the developing NC starting from its earliest migratory stages and that the symmetrically opposite 7q11.23 dosages alterations prime NCSCs to symmetrically opposite deficits upon BAZ1B interference. In turn, the central role of the NC in the development of facial morphology allows relating such findings to the symmetrically opposite craniofacial dysmorphisms of the two 7q11.23 syndromes.

BAZ1B interference disrupts key NC-specific transcriptional circuits

Having defined the functional impact of BAZ1B dosage on NC function, we predicted that a main molecular readout of its dosage imbalances would be at the level of transcriptional regulation, given its critical role as transcriptional regulator in different cell and animal models (*20–22*). To test this hypothesis and gain mechanistic insights into the specific BAZ1B dosage–dependent downstream circuits, we subjected 32 interfered NCSC lines to high-coverage RNA sequencing (RNA-seq) analysis. As shown in fig. S2A, a manually curated signature from an extensive literature review (*23–28*) validated the cranial identity of our NCSC lines, while clustering by Pearson correlation excluded the presence of any genotype- or hairpin-specific expression change. Confirming our previous observations in the two largest cohorts of iPSC lines (*29*), a principal component analysis (PCA) corroborated the significant impact of individual genetic backgrounds on transcriptional variability, with most "KD lines" clustering with their respective control "scr line." This was consistent with the narrow range of experimentally interfered BAZ1B dosages and pointed to a selective BAZ1B dosage–dependent transcriptional vulnerability (fig. S2B).

To dissect it, we thus resorted to a combination of classical pairwise comparative analysis, contrasting shBAZ1B-interfered NCSC lines
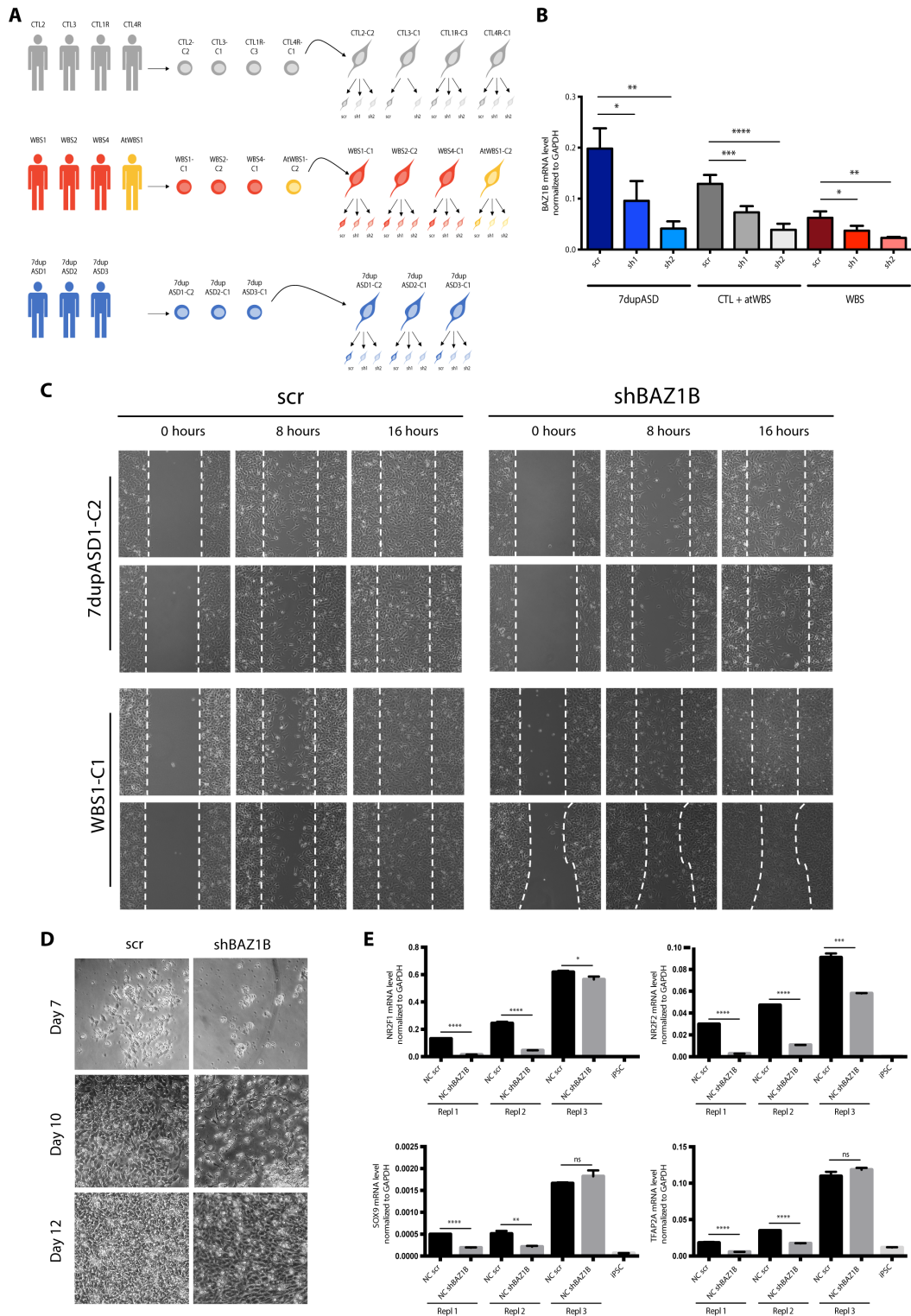
Fig. 1. *BAZ1B KD* impairs migration and induction of patient-specific iPSC-derived NCSCs. (**A**) Schematic representation of the KD strategy on our iPSC-derived NCSC cohort. (**B**) BAZ1B mRNA levels in all the interfered lines (scr, sh1, and sh2) as measured by qPCR. Data represent aggregates of samples with the same number of *BAZ1B* copies (7dup, CTL + atWBS, and WBS). Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) is used as a normalizer. (**C**) Eight- and 16-hour time points from the wound-healing assay analyses performed on a 7dupASD and a WBS NCSC line upon BAZ1B KD. Cells from the same line infected with the scr sh were used as references for the migration (*n* = 2). (**D**) Days 7, 10, and 12 of NC differentiation from embryoid bodies (EBs) of an scr-interfered iPSC line and its respective BAZ1B KD (*n* = 3). (**E**) mRNA levels of NC markers at day 12 of differentiation in three individual experimental replicates [bright-field images are reported in (D)]. An iPSC line is included as a negative control. Student's *t* test was used (ns, not significant; *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, and ****$P < 0.0001$).

(sh1 + sh2) with their respective controls (scr), with a complementary regression analysis using *BAZ1B* expression levels as independent variables, subtracting the contribution of individual genetic backgrounds. This design increases robustness and sensitivity in the identification of genes that, across multiple genetic backgrounds and target gene dosages, might have a different baseline (scr) across individuals while still being robustly dysregulated upon BAZ1B interference.

The two analyses identified a total of 448 genes with false discovery rate (FDR) < 0.1 (1192 with *P* < 0.01 and FDR < 0.25) whose transcriptional levels followed BAZ1B dosage, in either a direct (202; 539 with *P* < 0.01 and FDR < 0.25) or an inverse (246; 653 with *P* < 0.01 and FDR < 0.25) fashion. In addition, genes identified in the regression analysis included around 90% of the differentially expressed genes (DEGs) (27 of 29, FDR < 0.1) found in the comparative analysis (Fig. 2A). Consistent with the differential efficiency of the two short hairpins, we found a globally stronger transcriptional impact for the group of samples targeted by sh2 (fig. S2C) and a milder but nevertheless clearly distinguishable effect of sh1, resulting in particularly informative gradient of dosages over the scr control lines.

Particularly noteworthy among the genes that we found correlated with *BAZ1B* levels were (i) crucial regulators of cranial NC, further highlighting a convergent BAZ1B dosage–dependent dysregulation of the foundational *CUL3*-centered regulatory axis orchestrating NC-mediated craniofacial morphogenesis (*30*), and (ii) genes associated with variation of human facial shape or causative of dysmorphic facial features and mild intellectual disability when mutated (Fig. 2B and table S1).

Gene Ontology (GO) analysis performed on genes directly following BAZ1B levels suggested specific enrichments in biological processes such as histone phosphorylation, chromosome localization, RNA processing, and splicing. Genes inversely following BAZ1B levels were instead enriched in categories particularly relevant for NC and NC-derivative functions, such as cell migration and cardiovascular and skeletal development (Fig. 2C). By querying the OMIM database, we found that several DEGs were associated with genetic disorders whose phenotypes include "mental retardation," "intellectual disability," and/or "facial dysmorphisms" (Fig. 2D), underscoring the pertinence of BAZ1B-dependent dysregulation across both the neurocristopathic and cognitive axes.

Last, a master regulator analysis identified candidate regulators of BAZ1B DEGs, including factors involved in enhancer marking [CEBPB, p300, RBBP5, HDAC2 (histone deacetylase 2), KDM1A, and TCF12], promoter activation [TBP (TATA box–binding protein), TAF1 (TBP-associated factor 1), and POL2 (polymerase 2)], and chromatin remodeling (CTCF, RAD21, and YY1) (Fig. 2E and fig. S2D), several of which are themselves causative genes of intellectual disability syndromes with neurocristopathic involvement, as in the case of our recently identified Gabriele–de Vries syndrome caused by YY1 haploinsufficiency (*31*, *32*). Chromatin remodeling was indeed the most prominently enriched group within the overall domain of transcriptional regulation. Two master regulators are particularly noteworthy, as they are themselves regulated by BAZ1B dosage. The first is EGR1 (FDR < 0.1), which is itself among the genes inversely correlated with BAZ1B levels, which is implicated in cranial development (in animal models) (*33*, *34*) and whose promoter has been recently shown to feature a bivalent state in human embryonic cranial NC (*23*, *35*). The second is MXI1, identified as master regulator of genes directly following BAZ1B levels (FDR < 0.001), which is itself found among the genes inversely correlated with BAZ1B and is itself a regulator

of BAZ1B, pointing to a cross-talk between the two (fig. S2C). Notably, two differentially expressed targets of MXI1, *TGFB2* and *NFIB*, are also involved in intellectual disability and craniofacial defects (*30*, *36*, *37*).

## BAZ1B regulates the NC epigenome in a dosage-dependent manner

The transcriptional readout and functional impact of BAZ1B dosage (at the level of NC induction and migration) established its role as a master controller of the NC. We thus predicted, on the basis of its molecular function, that BAZ1B would directly bind to key NC target genes and that for some of these, the binding would be dosage sensitive. These genes would be, in turn, the most likely direct targets to mediate the dosage-dependent transcriptional and functional phenotypes described above. To test this prediction, we set out to both identify BAZ1B direct targets and characterize their promoter and enhancer states, so as to mechanistically link their transcriptional dysregulation with BAZ1B dosage–dependent chromatin binding. Given the absence of chromatin immunoprecipitation (ChIP)–grade BAZ1B antibodies, to carry out our ChIP coupled with sequencing (ChIP-seq) on scr and KD lines, we first designed a tagging strategy to establish, by CRISPR-Cas9 editing, a series of in-frame 3xFLAG endogenously tagged *BAZ1B* alleles in representative iPSCs of the four genotypes (Fig. 3A and fig. S3, A and B). These were then differentiated to NCSCs (fig. S3C) and subjected to ChIP-seq with anti-FLAG antibody, enabling a faithful characterization of BAZ1B genome-wide occupancy across dosages (one tagged allele in WBS, two tagged alleles in atWBS and CTL, and two tagged alleles in the context of 1.5-fold dosages in 7dupASD).

PCA shows a clear separation of the samples by *BAZ1B* copy number, with CTL and atWBS samples clustering more closely and WBS and 7dupASD samples clustering at opposing positions (Fig. 3B). To call NC-specific enhancer regions and promoter-enhancer associations, we exploited for chromatin annotation the unprecedented resolution afforded by the patients' cohort with its underlying variability and proceeded to (i) select chromosomal regions featuring H3K4me1 and H3K27ac in at least two individuals; (ii) exclude regions marked with H3K4me3 in at least two individuals;(iii) eliminate regions bearing a transcription start site (TSS); and (iv) associate each putative enhancer to the closest TSS, identifying a total of 30,8470 putative enhancer regions. Notably, BAZ1B binds 75% of its targets at their enhancer regions (6747 genes), with the remaining 2297 targets bound at promoters (Fig. 3C). In addition, 40% of genes expressed in NC are bound by BAZ1B, either exclusively at enhancers (27.4%) or exclusively at promoters (3.5%) or at both regions (9%). This highlights its pervasiveness within the NC epigenome (Fig. 3C) and is also reflected in the key functional enrichments observed for the BAZ1B direct targets that are also expressed and that include "axon guidance," "tube development," "dendrite development," "outflow tract morphogenesis," "odontogenesis," "wound healing," and "endochondral bone morphogenesis" (Fig. 3D). Many of the phenomena captured by these GO categories (e.g., odontogenesis and endochondral bone morphogenesis) are linked to recent changes in the bone structure of modern (versus archaic) humans, with *Homo sapiens* having characteristically smaller teeth than its extinct relatives.

Last and consistent with the enrichments in NC-defining categories uncovered above, the analysis of BAZ1B bound regions revealed major convergence with the binding motifs of critical NC regulators, including two motifs similar to those of TFAP2A and NEUROG2, and one equally associated to TAL1, TCF12, AP4, and ASCL1
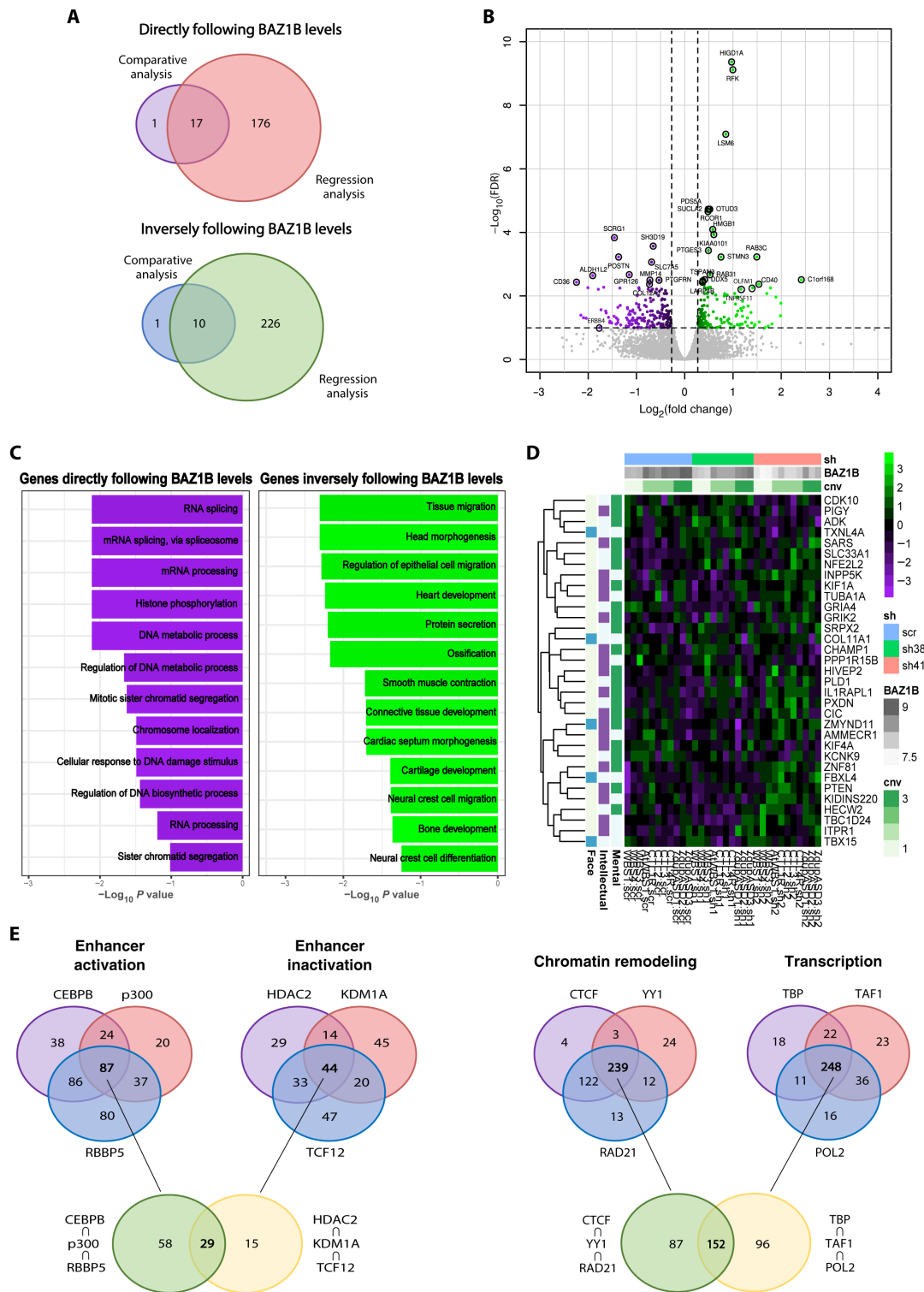
Fig. 2. BAZ1B KD is responsible for transcriptional alterations in NC-related pathways. (**A**) Overlap between genes directly or inversely following BAZ1B levels identified in the pairwise comparative analysis (scr versus shBAZ1B) and in the regression analysis on BAZ1B-level sensitive genes on iPSC-derived NCSCs (FDR < 0.1). (**B**) Volcano plot reporting DEGs identified in the RNA-seq analysis on iPSC-derived NCSCs [fold change (FC) > 1.25; FDR < 0.1]. (**C**) Top most specific enrichments for GO biological processes among the DEGs in the RNA-seq analysis on iPSC-derived NCSCs. (**D**) Heat map representing DEGs that are dysregulated in genetic disorders involving mental retardation ("Mental"), intellectual disability ("Intellectual"), and/or facial dysmorphisms ("Face") according to OMIM database classification. cnv, copy number variant. (**E**) Putative regulators of genes that follow BAZ1B levels identified by a master regulator analysis. Regulators were divided in four different groups based on their main functions.
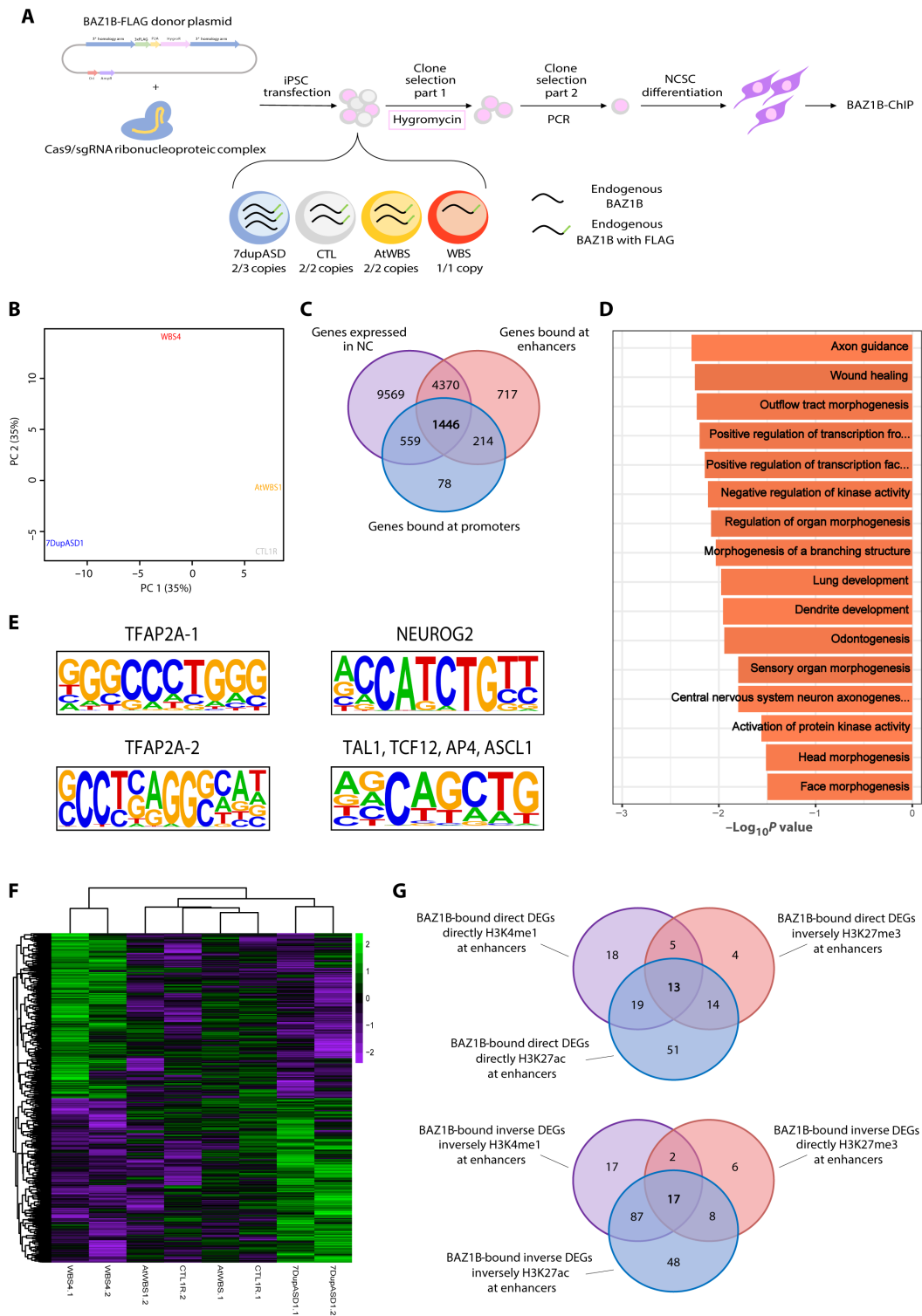
Fig. 3. BAZ1B preferentially binds its targets at their enhancer regions and its KD causes a redistribution of enhancer histone marks. (**A**) Schematic representation of the strategy for CRISPR-Cas9–mediated tagging of endogenous *BAZ1B*. Briefly, iPSCs from the four genotypes were electroporated with the donor plasmid and the Cas9/single-guide RNA ribonucleoprotein complex; clones were selected via hygromycin and PCR, differentiated to NCSCs, and then subjected to ChIP-seq. (**B**) PCA showing the distribution of the four *BAZ1B*-tagged NCSC lines according to their chromatin profiles. (**C**) Overlap between genes expressed in our NCSC lines (purple) and genes bound by BAZ1B at their enhancer (red) or promoter (blue) regions. (**D**) Top most specific enrichments for GO biological processes among the genes that are bound by BAZ1B and expressed in our NCSC cohort. (**E**) Most represented BAZ1B DNA binding motifs identified by HOMER show high similarity to neural and NCSC-specific transcription factors motifs. (**F**) BAZ1B differentially bound regions according to its copy number (FDR < 0.1; *n* = 2). (**G**) Overlap between genes that are differentially expressed have their enhancers differentially marked concordantly (H3K27ac, H3K4me1, and H3K27me3) and are bound by BAZ1B at enhancers.

(Fig. 3E and text S1A). Thus, BAZ1B binding regions are enriched for target sites of major regulators of NC and its neural derivatives (*38*, *39*), among which TFAP2A stands out given its core role in neural border formation and NC induction and differentiation (*40*) through the binding and stabilization of NC-specific enhancers, in concert with NR2F1, NR2F2, and EP300 (*41*).

Last, we identified 81 regions that are quantitatively bound by BAZ1B depending on its copy number (FDR < 0.1) (Fig. 3F), 153 regions differentially bound concordantly in WBS and 7dupASD compared to control and atWBS samples (FDR < 0.1) (fig. S4A), and 176 and 25 regions differentially bound preferentially in WBS (fig. S4B) and 7dupASD (FDR < 0.1) (fig. S4C), respectively.

Given the prominence of its binding to distal regulatory regions, we then set out to define the BAZ1B dosage–dependent impact on NCSC-specific enhancers by integrating H3K27ac, H3K4me1, H3K27me3, and H3K4me3 profiles. We thus performed a regression analysis on *BAZ1B* levels for the distribution of the three histone marks in the aforementioned regions and found H3K27ac to be the most affected, with 7254 genes differentially acetylated at their enhancers, followed by a differential distribution of the H3K4me1 (4048) and H3K27me3 (2136) marks (fig. S4D). This enabled the overlay of epigenomic and transcriptomic profiles, uncovering that among the 1192 DEGs identified in the regression RNA-seq analysis, 21.3% (257 of 1192) are associated to enhancers that are both bound by BAZ1B and differentially H3K27-acetylated in a manner concordant with *BAZ1B* levels (fig. S4E), with a stronger overlap for genes whose expression is inversely correlated with *BAZ1B* levels (160 versus 97). The same held for DEGs that have a concordant differential distribution of H3K4me1 mark at enhancers (123 versus 55), underscoring the consistency of the impact of BAZ1B dosage on distal regulation (fig. S4F). In contrast, a lower number of genes (*36*) showed a concordant differential distribution of the H3K27me3 mark and, at the same time, were bound by BAZ1B at enhancers (fig. S4G), indicating that BAZ1B preferentially affects active chromatin. From this integrative analysis, we could thus lastly identify a core set of 30 bona fide direct targets of BAZ1B, which are genes whose expression tightly follows BAZ1B levels and whose enhancers are bound by BAZ1B and clearly differentially modified (Fig. 3G, fig. S4H, and text S1B). Together, this first dosage-faithful analysis of BAZ1B occupancy in a diverse cohort of human NCSCs establishes its pervasive and mostly distal targeting of the NC-specific epigenome, with a preferential activator role on the critical transcriptional circuits that define NC fate and function.

Intersection with paleogenomic datasets uncovers a key evolutionary role for *BAZ1B*

Mild NC deficits have been put forth as a unifying explanatory framework for the defining features of the so-called domestication syndrome, with *BAZ1B* listed among the putative underlying genes because of its previously reported role in the NC of model organisms (*5*, *10*, *11*). The recent observation that its expression is affected by domestication-related mobile element insertion methylation in gray wolves (*12*) further supported its role in domestication, offering an intriguing parallel to the paleogenomic results that had detected *BAZ1B* within the regions of the modern genome reflective of selective sweeps and found it enriched for putatively regulatory mutations in AMHs (*15*).

Having defined the molecular circuits through which BAZ1B regulates NC, and since NC changes have been implicated in the

domestication syndrome (*5*), since craniofacial differences correlate with self-domestication (*1*), and since 7q11.23 dosage-related craniofacial differences in humans relate to the *H. sapiens* versus Neanderthal comparison (fig. S1A), we set out to test the role of BAZ1B dosage in the differences between modern and archaic humans. For this, we carried out a systematic integrative analysis of the overlaps between our empirically defined BAZ1B dosage–sensitive genes (blue Venn in Fig. 4B) and a combination of uniquely informative datasets highlighting differences between modern humans and archaics (Neanderthals/ Denisovans) (represented in Fig. 4A by skulls illustrating the more "gracile" and "juvenile" profile in AMH relative to Neanderthals visible in the overall shape of the neurocranium, reduced prognathism, brow ridges, and nasal projections) (*1*, *13*–*15*). Specifically, as shown in Fig. 4B, these datasets include (i) genes associated with signals of positive selection in the modern branch compared to archaic lineages (purple Venn) (*13*, *14*); (ii) genes harboring (nearly) fixed mutations in moderns versus archaics (pink Venn); and (iii) genes associated with signals of positive selection in the four paradigmatic domesticated species dog, cat, cattle, and horse (*1*) (orange Venn), to reveal statistically significant overlaps between them and genes associated with signals of positive selection in the modern branch compared to archaic lineages. In turn, the list of genes harboring (nearly) fixed mutations in moderns versus archaics contains three classes: (i) genes harboring high-frequency changes (*15*), (ii) genes harboring high-frequency missense mutations (red barplot), and (iii) genes enriched for high-frequency mutations in regulatory regions (green barplot) [data based on (*15*)] (Fig. 4C). As shown in the barplots, the obviously very limited number of high-quality coverage archaic genomes available results in a much higher number of nearly fixed changes identified in archaics (left/negative side of the plot) versus modern humans (right side) (Fig. 4C), setting a comparatively much higher threshold for the identification of nearly fixed modern changes.

These analyses are visualized in Fig. 4D (and detailed in tables S2 and S3) through a matrix that intersects all *BAZ1B* dosage–dependent genes (partitioned in the two categories of directly and inversely correlated and ordered across the full range of biological significance and regulatory proximity, from simply DEGs to bona fide direct targets) with the evolutionary changes underlying domestication and self-domestication, yielding the following key insights (color coded for degree of overlap and marked for significance in hypergeometric tests). First, the most significant pattern was obtains at the intersection with the top 10% genes showing an excess of (nearly) fixed mutations in the regulatory regions of modern humans compared to archaics, across both directly and inversely BAZ1B level–dependent genes (table S2). This same category of nearly fixed modern regulatory changes is also the only one that returns a statistically significant overlap with the most stringent core of BAZ1B dosage–dependent targets (i.e., DEGs whose enhancers are both directly bound by BAZ1B and differentially marked upon its decrease), demonstrating that BAZ1B directly controls, in an exquisitely dosage-dependent manner, this coherent and particularly relevant set of genes that underwent regulatory changes in human evolution. Second, the overall strongest overlaps map to the class of genes that are inversely correlated to BAZ1B levels, which we found to be strongly and specifically enriched for head morphogenesis and NC categories (Fig. 2C), thereby confirming craniofacial morphogenesis as the key domain of functionally relevant overlap between BAZ1B dosage and (self-) domestication changes relevant to the evolution of AMHs. Third, despite the spuriously inflated number of apparently fixed mutations in archaics (*15*), the
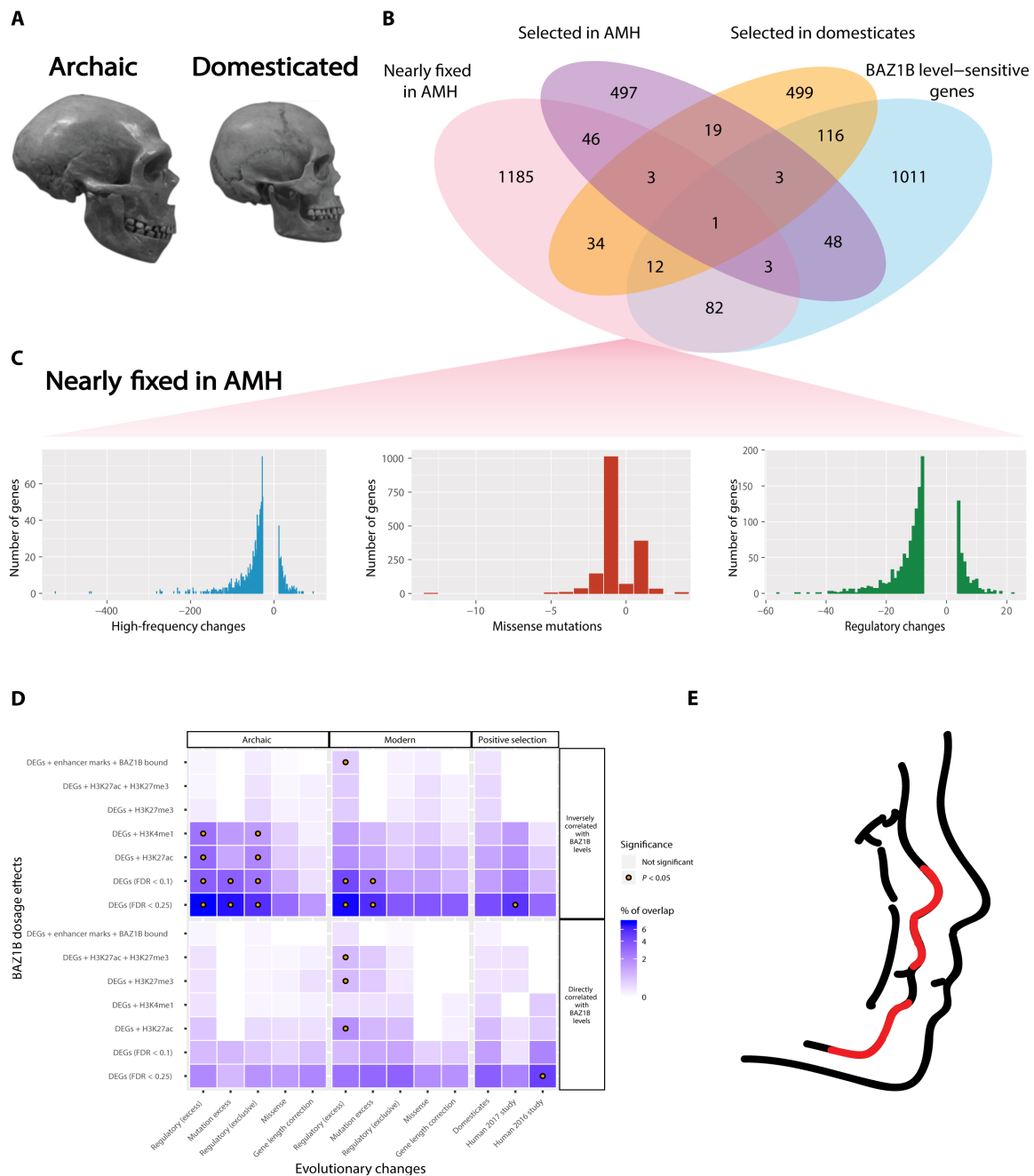
Fig. 4. Exploration of paleogenomic datasets supports a key evolutionary role for *BAZ1B* and validates the self-domestication hypothesis. (A) Archaic (Neanderthal) and modern skulls, illustrating the target domesticated phenotype that was captured by our analysis. Skull images were derived from work under a CC BY-SA 2.0 license (https://creativecommons.org/licenses/by-sa/2.0/deed.en) by hairymuseummatt. (B) Overlap between *BAZ1B* level–sensitive genes and datasets, which bring out differences between AMHs and archaics, as well as genes under positive selection in modern humans and domesticates. (C) Barplots showing the occurrence of high-frequency changes, missense mutations, and mutations in regulatory regions in genes from the AMH (nearly) fixed mutation dataset (pink Venn in B). (D) Heat map representing the amount of overlaps for each list selected from (B). Gene overlaps and detailed list descriptions are reported in table S2. (E) Rendering of a typical WBS face (left) against the background of a typical modern face (right). Red segments indicate areas of the lower face where the two faces most sharply depart (nose, philtrum, and lower front of the mandible). The lower midface region is most often associated with mutations in genes figuring prominently in our intersections, as discussed in the text and table S3.

overall extent of overlap between genes affected by BAZ1B dosage and our modern and archaic sets does not reveal significantly more hits for archaics. Globally, we found consistently more overlapping genes between the BAZ1B targets and the modern human data and even no statistically significant overlap for any list of the archaic-

specific mutations when crossed against genes directly correlated to *BAZ1B* level. We find this noteworthy, given the evidence that the Neanderthal face also displays derived characteristics (*42*) that could be the result of modifications of genes that could overlap with those highlighted in this work. Last, the (lower) midface emerges as a

particularly salient area of functionally relevant overlap (as illustrated in Fig. 4E and detailed in table S3), given the specific genes that our analysis unearthed: (i) *COL11A1*, one of the few craniofacial genes highlighted across domestication studies (dog, house sparrow, and pig breeds), which lies in a region of the human genome that resisted archaic introgression (*13*) and is associated with Marshall syndrome; (ii) *XYLT1*, one of the five genes (along with *ACAN*, *SOX9*, *COL2A1*, and *NFIX*) that affect lower and midfacial protrusion, are among the top differentially methylated genes compared to archaics and were also highlighted in a recent study on regulatory changes that shaped the modern human facial and vocal anatomy (tables S1 and S3) (*43*); and (iii) *NFIB*, which belongs to the same gene family as *NFIX* and shares some of its functions. In sum, the direct and dosage-sensitive control by BAZ1B of genes that underwent regulatory changes in human evolution and whose altered expression underlies neuro-cristopathic facial dysmorphisms is consistent with the hypothesis of mild neurocristopathy as the mechanistic core selected in the self-domestication of the modern human face.

DISCUSSION

As recently reconstructed (*3*), the idea of human self-domestication dates back, at least in terms of scientific record, to Johann Friedrich Blumenbach at the onset of the 19th century. Following on his seminal account of domestication systematized in *Variations of Animals and Plants under Domestication* (*44*), Charles Darwin also considered the analogy between modern humans and domesticated species in *The Descent of Man* (*45*), yet his emphasis on controlled breeding as a key aspect of domestication led him to frame domestication and self-domestication as distinct phenomena and thereby leave Blumenbach's intuition largely undeveloped (*46*). Since then, the possibility that the anatomical and cognitive-behavioral hallmarks of AMHs could result from an evolutionary process bearing such significant similarities to the domestication of animals as to share the same underlying cause has been refined into the full-fledged self-domestication hypothesis (*1*, *2*). As recently argued (*1*, *3*), convergent lines of evidence also indicate that self-domestication is temporally aligned with the emergence of AMH, although the process may have acquired further momentum with the gradual expansion of our species (*1*, *3*). However, despite spurring considerable interest, the self-domestication hypothesis has thus far failed to marshal conclusive evidence largely because of two factors: (i) the lack of a coherent explanation, even at a theoretical level, of what developmental and genetic mechanisms could underlie domestication in general and (ii) the absence of suitable experimental systems in which those mechanisms could be specifically tested in the case of human self-domestication. The first problem was tackled by the recent proposition of mild NC deficits as a central and unifying functional layer underlying domestication (*5*). This constituted a major conceptual advance, particularly because it generated the testable hypothesis of an altered NC gene expression program in domesticated species relative to their wild-type ancestors. For humans, given the obvious lack of gene expression data from archaic hominins, we reasoned that this hypothesis could be verified by examining the genetic changes between archaic and modern humans in light of the gene regulatory networks directly inferred from human neurocristopathies. We thus set out to test whether specific human neurodevelopmental disorders, carefully selected on the basis of both craniofacial and cognitive-behavioral traits relevant to domestication, could illuminate the regulatory circuits shaping the

modern human face and hence be harnessed for an empirical validation of the self-domestication hypothesis. Specifically, we reasoned that WBS and 7dupASD, through their uniquely informative set of symmetrically opposite phenotypes at the level of face morphology (fig. S1A) and sociality, constituted a paradigmatic test case to probe the heuristic potential of neurodevelopmental disease modeling for the experimental understanding of human evolution. The following key insights confirm the validity of this approach.

First, we identified the 7q11.23 region *BAZ1B* gene as a master regulator of the modern human face on the basis of a molecular and functional dissection in the thus far largest cohort of WBS patient– and 7dupASD patient–specific NCSCs and across an exhaustive range of BAZ1B dosages. Notably, our cohort also included NCSCs from a patient with rare WBS featuring a much milder WBS gestalt and harboring an atypical, *BAZ1B*-sparing deletion that served as a particularly informative control, as confirmed by the clustering of atypical NCSC lines with controls when probed for BAZ1B occupancy. In particular, exploiting the fine-grained resolution of BAZ1B dosages recapitulated in our cohort, we could couple classical pairwise comparisons with a more sophisticated regression analysis on BAZ1B levels, thereby revealing major BAZ1B dosage–dependent transcriptional alterations pivoting around clusters of pathways that are crucial for NC development and maintenance, as well as for its downstream skeletal and cardiac outputs.

Second, we repurposed the versatility of CRISPR-Cas9 to generate an allelic series of endogenously tagged *BAZ1B* across 7q11.23 dosages (including the *BAZ1B*-sparing atypical patient as uniquely relevant control) to define its dosage-dependent genome-wide occupancy. Taking advantage of previous extensive work on the NCSC chromatin landscape (*41*, *47–49*), we were able to define a pivotal role for BAZ1B in NCSC enhancer regulation, consistent with its preferential binding of distal regulatory regions, and to partition its dosage-dependent regulation into bona fide direct and indirect targets. The overall balance between the numbers of genes up- or down-regulated upon BAZ1B KD—together with the greater overlap, sheer size, and significance of enrichments in chromatin remodeling categories over other domains of transcription regulation—further corroborates the inclusion of BAZ1B among the factors acting upstream of enhancer and promoter modulations to enable or reinforce rather than specify their net outcome. Last, this molecular readout was translated to the functional level with the definition of an impairment in both NCSC migration and outgrowth from EBs upon decrease in BAZ1B, providing the first validation of BAZ1B involvement in key functions of the developing human NC.

Third, our investigation provides the first experimental evidence for the neurocristopathic hypothesis that had been put forth to explain the domestication syndrome and had pointed to BAZ1B as one of the candidates underlying this syndrome (*5*). Among the key NC hubs affected by *BAZ1B* dosage, we uncovered three additional critical genes—*EDN3*, *MAGOH*, and *ZEB2*—that had also been predicted in the same model because they are associated with behavioral changes found in domesticates, thereby defining a regulatory hierarchy for this coherent set of genes underlying domestication.

Last, the empirical determination of BAZ1B dosage–sensitive genes in NC models from AMHs with accentuated domestication-relevant traits allowed us to expose, in a functionally relevant manner, the genetic differences between modern versus archaic. This brought to the fore the significant convergence between BAZ1B-dependent circuits and genes harboring regulatory changes in the human lineage,

reinforcing the notion that regulatory regions contain some of the most significant changes relevant for the modern lineage. This is also reinforced by the recent identification of AMH-specific hypermethylation in the regulatory region of *BAZ1B* itself (*43*).

Last, it is noteworthy that genes implicated in NC development also play significant roles in the establishment of brain circuits that are critical for cognitive processes like language or theory of mind prominently affected in 7q11.23 syndromes. Among the genes downstream of BAZ1B that we uncovered in this study, *FOXP2*, *ROBO1*, and *ROBO2* have long been implicated in brain wiring processes critical for vocal learning in several species (*50*, *51*), including humans, and will warrant further mechanistic dissection in light of the distinctive linguistic profile of WBS individuals. In conclusion, our findings establish the heuristic power of neurodevelopmental disease modeling for the study of human evolution.

MATERIALS AND METHODS

Human samples
Ethics approvals were reported in the study that established the original iPSC cohort (*16*) and also apply to the additional samples included in this study (7dupASD3 and CTL4R).

Fibroblast reprogramming and iPSC culture
WBS1, WBS2, WBS3, WBS4, 7dupASD1, atWBS1, and CTL2 fibroblasts were reprogrammed using the mRNA Reprogramming Kit (Stemgent), while the 7dupASD2 and CTL1R lines were reprogrammed with the microRNA Booster Kit (Stemgent). The CTL3 line was reprogrammed by transfection with the STEMCCA polycistronic lentiviral vector followed by Cre-mediated excision of the integrated polycistron. 7dupASD3 and CTL4R fibroblasts were reprogrammed using the Simplicon RNA Reprogramming Kit (Millipore).

Before differentiation, iPSC lines were cultured on Matrigel hESC-qualified Matrix (BD Biosciences)–coated plates, diluted 1:40 in Dulbecco's minimum essential medium/F-12, and grown in mTeSR 1 medium (STEMCELL Technologies). They were passaged upon treatment with Accutase (Sigma-Aldrich) and then plated in mTeSR 1 medium supplemented with 5 μM Y-27632 (Sigma-Aldrich).

Differentiation
Differentiation into NCSCs was performed as previously described (*52*), with the exception of NCSCs used in the experiment reported in Fig. 1 (D and E) (*19*).

Flow cytometry
NCSCs were detached using Accutase and counted, and 1×10^6 cells per experimental condition were fixed in 4% paraformaldehyde and then blocked in 10% bovine serum albumin. Cells were incubated for 1 hour with primary antibodies conjugated to fluorophores (HNK1–fluorescein isothiocyanate and nerve growth factor receptor–Alex Fluor 647; BD Biosciences). Analyses were performed on a FACSCalibur instrument (BD Biosciences), and data were analyzed with FCS express software (Tree Star). Fluorescence-activated cell sorting characterization for 7dupASD3 and CTL4R lines is reported in fig. S1B; for all the other lines, see (*16*).

Lentiviral vector production and NCSC transfection
BAZ1B KD was performed using validated pLKO.1 TRC vector TRCN0000013338 (referred to as sh1) and TRCN0000013341 (referred to as sh2). A pLKO.1 TRC vector containing a scrambled short hairpin sequence was used as a negative control.

Second generation lentiviral vectors were produced through calcium phosphate transfection of human embryonic kidney 293T cells and ultracentrifugation (2 hours, 20°C, 20,000 rpm).

NCSCs (3 to 4×10^5) were infected upon splitting and then selected by adding puromycin (1 μg/ml) to the medium.

RNA extraction, retrotranscription, and real-time qPCR
RNA was extracted using the RNeasy Micro Plus Kit (QIAGEN) according to the manufacturer's instructions. Retrotranscribed cDNA was obtained from 0.5 to 1 μg of total RNA using the SuperScript VILO retrotranscription kit (Thermo Fisher Scientific).

Real-time qPCR was performed on a 7500 Fast Real-Time PCR system (Applied Biosystems) using SYBR Green Master Mix (Applied Biosystems) as the detecting reagent. A total cDNA amount corresponding to 15 ng of starting RNA was used for each reaction. Each sample was analyzed in triplicate and normalized to *GAPDH*. Relative mRNA quantity was calculated by the comparative cycle threshold (Ct) method using the formula $2^{-\Delta Ct}$ [*BAZ1B*, CCTCGCAGTA-AGAAAGCAAAC (forward) and ACTCATCCAGCTCCTTTTGAC (reverse); *GAPDH*, GCACCGTCAAGGCTGAGAAC (forward) and AGGGATCTCGCTCCTGGAA (reverse); *NR2F1*, AGAAGCTCAAG-GCGCTACAC (forward) and GGGTACTGGCTCCTCACGTA (reverse); *NR2F2*, GCAAGTGGAGAAGCTCAAGG (forward) and GCTTTCCACATGGGCTACAT (reverse); *TFAP2A*, GCCTCTC-GCTCCTCAGCTCC (forward) and CGTTGGCAGCTTTACGTCTCCC (reverse); and *SOX9*, AGTACCCGCACTTGCACAAC (forward) and GTAATCCGGGTGGTCCTTCT (reverse)].

RNA-seq libraries preparation
Library preparation for RNA-seq was performed according to the TruSeq Total RNA sample preparation protocol (Illumina), starting from 250 ng to 1 μg of total RNA. cDNA library quality was assessed in an Agilent 2100 Bioanalyzer using the High Sensitivity DNA Kit. Libraries were sequenced with the Illumina HiSeq machine at a read length of 50–base pair (bp) paired end and a coverage of 35 million of reads per sample.

Protein extraction and Western blot
NCSCs were lysed in radioimmunoprecipitation assay buffer [10 mM tris (pH 8.0), 1% Triton X-100, 0.1% sodium deoxycholate, 0.1% SDS, 140 mM NaCl, and 1 mM EDTA] supplemented with protease inhibitor cocktail (Sigma-Aldrich) and 0.5 mM phenylmethylsulfonyl fluoride (Sigma-Aldrich) for 1 hour at 4°C.

Protein extracts (30 to 50 μg per sample) were supplemented with NuPAGE LDS sample buffer (Thermo Fisher Scientific) and 50 mM dithiothreitol (Thermo Fisher Scientific) and denatured at 95°C for 3 min. Then, extracts were run on a precast NuPAGE 4 to 12% bis-tris Gel (Thermo Fisher Scientific) in NuPAGE MOPS SDS Running Buffer (Thermo Fisher Scientific) and transferred to a 0.45-μm nitrocellulose membrane (GE Healthcare) for 1 hour at 100 V in a buffer containing 20% absolute ethanol and 10% 0.25 M tris base and 1.9 M glycine. The membranes were blocked in TBST [50 mM tris (pH 7.5), 150 mM NaCl, and 0.1% Tween 20] and 5% milk for 1 hour, incubated with primary antibodies overnight at 4°C and with secondary antibodies for 1 hour at room temperature. Primary [BAZ1B (Abcam) and glyceraldehyde-3-phosphate dehydrogenase (GAPDH; Millipore)] and secondary antibodies were diluted in TBST and 5% milk. Blots were detected

with the ECL Prime Western Blotting Detection Reagents (Sigma-Aldrich) and scanned using the ChemiDoc system (Bio-Rad).

Wound-healing assay

Cells (5×10^4 to 7×10^4) were plated in each of the two Matrigel-coated wells of silicone culture-inserts (Ibidi) attached to six-well culture plates. After 24 hours, the insert was removed, medium was changed to remove dead cells, and time lapse was performed for 24 hours at the rate of one image every 10 min at ×10 magnification; each condition was analyzed in duplicate. Images were acquired with the BX61 upright microscope equipped with a motorized stage from Olympus or the Nikon Eclipse Ti inverted microscope equipped with a motorized stage from Nikon and analyzed with ImageJ.

Endogenous *BAZ1B* tagging via CRISPR-Cas9

iPSCs were pretreated with 10 μM rho kinase inhibitor for 4 hours, and then 2×10^6 cells were electroporated using the Neon system with the Cas9/single-guide RNA ribonucleoprotein complex and the donor plasmid (synthesized by GeneArt). The donor plasmid contained three FLAG tags followed by a self-cleaving peptide (P2A) and a hygromycin resistance (HygroR). The 3xFLAG-P2A-HygroR cassette was flanked by *BAZ1B*-specific homology arms (5′ HA and 3′ HA) to promote homologous recombination and then subcloned into a bacterial backbone (Fig. 3A).

After 48 hours, iPSC medium was supplemented with hygromycin B (50 μg/μl), and selection medium was maintained for 15 days. Fifteen to 20 clones per iPSC line were then subjected to PCR to (i) evaluate the presence of the cassette and the insertion in the correct genomic locus and (ii) distinguish heterozygously tagged from homozygously tagged clones (fig. S3A). We could isolate a clone with a homozygous integration from the CTL, the atWBS, and the typical WBS but not from the 7dupASD line. In the 7dupASD clone, the FLAG tag was present in two of three copies, as shown by a digital PCR analysis (fig. S3B).

Digital PCR

DNA (60 ng) was amplified in a reaction volume containing the following reagents: QuantStudio 3D Digital PCR Master Mix v2 (Thermo Fisher Scientific), Custom TaqMan Copy Number Assays SM 20× FAM labeled (Thermo Fisher Scientific), and TaqMan Copy Number Reference Assay 20× (Thermo Fisher Scientific) VIC labeled (Thermo Fisher Scientific). The mix was loaded on a chip using the QuantStudio 3D Digital PCR Chip Loader. The chips were then loaded on the ProFlex PCR System (Thermo Fisher Scientific), and data were analyzed using the "QuantStudio 3D AnalysisSuite Cloud Software." The entire process was performed by the qPCR Service at Cogentech, Milano [Custom (FLAG) TaqMan Copy Number Assays: forward primer, TGGACAGTCCAGAGGACGAA; reverse primer, CACCCTTGTCGTCATCGTCTT; and probe, FAMACAGAAGA-AGGACTACAAAGACG and TaqMan Copy Number Reference Assay: TERT (VIC) (catalog number 4403316)].

ChIP coupled with sequencing

Approximately 2×10^5 cells were used (~100 μg of chromatin) for histone mark IP, and 1 mg of chromatin was used for BAZ1B-FLAG IP. Cells were fixed with phosphate-buffered saline, containing 1% formaldehyde (Sigma-Aldrich), for 10 min to cross-link proteins and DNA, when the reaction was then stopped by adding 125 mM glycine for 5 min. Cells were lysed with SDS buffer containing 100 mM NaCl,

50 mM tris-HCl (pH 8.0), 5 mM EDTA (pH 8.0), and 10% SDS, at which point chromatin pellets were resuspended in IP buffer containing 1 volume of SDS buffer and 0.5 volume of Triton dilution buffer [100 mM tris-HCl (pH 8.5), 5 mM EDTA (pH 8.0), and 5% Triton X-100]. Chromatin was then sonicated using the S220 Focused-ultrasonicator (Covaris) to generate <300 bp DNA fragments (for histone mark IPs) or the Branson Digital Sonifier to generate 500 to 800 bp DNA fragments (for BAZ1B-FLAG IP).

Sonicated chromatin was incubated overnight at 4°C with primary antibodies [H3K27ac (Abcam), H3K4me1 (Abcam), H3K4me3 (Abcam), H3K27me3 (Cell Signaling Technology), and FLAG (Sigma-Aldrich)] and then for 3 hours with Dynabeads Protein G (Thermo Fisher Scientific). Beads were washed three times with low-salt wash buffer [0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM tris-HCl (pH 8.0), and 150 mM NaCl] and once with high-salt wash buffer [0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM tris-HCl (pH 8.0), and 500 mM NaCl]. Immunocomplexes were eluted in decross-linking buffer (1% SDS and 100 mM NaHCO3) at 65°C for 2 hours. DNA was purified using QIAquick PCR columns (QIAGEN) and quantified with a Qubit dsDNA HS assay kit (Thermo Fisher Scientific). DNA libraries were prepared by the sequencing facility at European Institute of Oncology according to the protocol described by Blecher-Gonen and colleagues (*53*), and DNA was sequenced on the Illumina HiSeq 2000 platform. For the FLAG ChIP, samples were run in duplicate.

RNA-seq analysis

RNA-seq data were quantified using Salmon 0.91 to calculate read counts and transcripts per million in a transcript- and gene-wise fashion, using the quasi-mapping offline algorithm (*54*) on the GRCh38 (National Center for Biotechnology Information) database. edgeR was used for differential gene expression analysis (DEA), using generalized linear regression methods, to identify pattern of differential expression following two different schemes:

1) A factorial analysis based on the definition of one group of scrambled and one group of KD samples to identify genes dysregulated similarly across short hairpins characterized by different efficiencies.

2) A numerical analysis in which log-normalized [Trimmed Mean of M-values (TMM)] BAZ1B levels, as quantified by RNA-seq, was used as independent variable.

All analyses were performed dropping individual variations (~individual+KD or ~individual+BAZ1B) to account for the genetic background of each individual. In particular, this design is expected to permit the identification of genes, which change expression level upon KD even in situations in which genotype-specific makeups would lead BAZ1B-dependent genes to have unique expression levels in scramble lines. In the factorial analysis, DEGs were identified and characterized by filtering for fold change (FC) > 1.25 and FDR < 0.05 unless explicitly indicated.

To our knowledge, performing a regression analysis at a gene-specific level has never been performed. We were able to do this because of the availability of a large set of samples (11 individuals) and because of the two short hairpins robustly respectively reducing BAZ1B expression levels, respectively by ~40 and ~70% in all individuals lines. To validate the quality of our numerical differential expression analysis, we took advantage of HipSci data (*55, 56*) and iPSCpoweR tools (*29*). We took 50 of 105 possible combinations of 13 random individual RNA-seq data from the healthy HipSci cohort, representing both sexes and having at least two technical replicates

per individual. Unfortunately, HipSci does not contain at least 13 individuals with three clones per individual. Thus, we performed four alternative DEAs with edgeR (table S4) on the 50 different random combinations of 13 individuals identified (200 DEAs in total, on 22 samples, two clones per individual), using the same model matrix used for the regression analysis (~individual+BAZ1B) and using BAZ1B levels of scramble and sh2 lines. All analyses identified very low number of spurious DEGs (fig. S2E). Thus, we used the "Edg2" pipeline (table S4) because it does not discard genes with higher variability (Edg2 and Edg4 versus Edg1 and Edg3), and it is based on a better suited algorithm (Edg2 versus Edg4). With our model matrix, filtering by $P < 0.01$ (and FDR < 0.25), using Edg2 on a random HipSci data, we obtained an average of 93.32 DEGs (on average) with a median equal to 43 (table S5). GO enrichments were performed using topGO R package version 2.28.0.

Master regulatory analysis was performed via hypergeometric test by measuring gene set enrichments in lists of transcription factor targets provided by the TFBS tools database (57). Both GO and transcription factor enrichment analyses were performed considering background genes expressed in at least two samples in our NCSC cohort.

ChIP-seq analysis

ChIP-seq experiments were analyzed both qualitatively and quantitatively. Reads were trimmed with the FASTX-Toolkit (-Q33 -t 20 -l 22), aligned with Bowtie 1.0 (-v 2 -m 1) on the Human hg38 reference genome, and peaks were called using MACS 2.1.1. H3K4me1, H3K27ac, H3K4me3, and H3K27me3 peaks were called with --broad using default parameters and $q < 0.05$.

Qualitative analysis, including intersection and comparison of bed files, was performed using BedTools version 2.23.

To define enhancer regions, we intersected those marked by H3K4me1 and H3K27ac in at least two samples, discarded regions with H3K4me3 in at least two samples, and discarded regions overlapping with TSS. Motif enrichment was performed by using HOMER v4.10.

Quantification of reads per region was performed with DeepTools 3.0.2. Differential mark deposition was conducted by means of edgeR 3.24.1 inside R 3.3.3. To define mark deposition following BAZ1B levels, we used the same design as for RNA-seq data (~individual+BAZ1B).

To identify BAZ1B bound regions and to avoid losing identification of lowly covered regions, we resorted to (i) aggregation of all sample aligned reads and (ii) peak calling with MACS2 using −extsize 800 and $q < 0.25$. BAZ1B binding coverage was calculated with DeepTools, with the same parameters used for histone marks, on the identified peak regions. Differentially bound regions were identified with edgeR.

Assembly of archaic and modern human lists

The archaic/modern lists were generated from the material presented in (15). We used high-coverage genotypes for three archaic individuals: one Denisovan (58), one Neanderthal from the Denisova cave in Altai mountains (59), and another Neanderthal from Vindija cave, Croatia (60). The data are publicly available at http://cdna.eva.mpg.de/neandertal/Vindija/VCF/, with the human genome version hg19 as reference. High-frequency (HF) differences were defined as 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. High-frequency changes in archaics were defined as occurring at less than 1% in present-day humans, while at least two

archaic individuals carry the derived allele. The HF lists used here were examined as presented in (15), with the exception of the HF lists in regulatory regions, which were extracted from the same dataset but not presented as such in the original paper.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/5/12/eaaw7908/DC1

Fig. S1. BAZ1B KD validation in iPSC-derived NCSCs and evaluation of its impact on NCSC migration.

Fig. S2. BAZ1B KD affects the transcriptome of iPSC-derived NCSCs.

Fig. S3. Generation of BAZ1B-FLAG iPSC lines and differentiation to NCSCs.

Fig. S4. BAZ1B KD induces a significant chromatin remodeling at distal regions.

Text S1A. Detailed description of HOMER motif enrichments performed on BAZ1B ChIP-seq data.

Text S1B. List of key direct targets of BAZ1B involved in neural- and NC-related development and relevant associated literature.

Table S1. Genes relevant for NC and NC-derived features whose expression follows BAZ1B levels.

Table S2A. Significant genes in human evolution.

Table S2B. Regulatory excess in archaic humans, overlap with BAZ1B targets.

Table S2C. Mutation excess in archaic humans, overlap with BAZ1B targets.

Table S2D. Regulatory changes (exclusive) in archaic humans, overlap with BAZ1B targets.

Table S2E. Missense mutations in archaic humans, overlap with BAZ1B targets.

Table S2F. Mutation excess in archaic humans corrected for length, overlap with BAZ1B targets.

Table S2G. Regulatory excess in modern humans, overlap with BAZ1B targets.

Table S2H. Mutation excess in modern humans, overlap with BAZ1B targets.

Table S2I. Regulatory changes (exclusive) in modern humans, overlap with BAZ1B targets.

Table S2J. Missense mutations in modern humans, overlap with BAZ1B targets.

Table S2K. Mutation excess in modern humans corrected for length, overlap with BAZ1B targets.

Table S2L. Genes under positive selection in domesticated animals, overlap with BAZ1B targets.

Table S2M. Genes under positive selection from Peyrégne *et al.* (13) in modern humans, overlap with BAZ1B targets.

Table S2N. Genes under positive selection from Racimo (14) in modern humans, overlap with BAZ1B targets.

Table S3. Crucial genes identified in the overlap between BAZ1B datasets and archaic versus modern human datasets reported in this study.

Table S4. Alternative differential expression analysis functions tested with iPSCpower to assess the efficacy of our design matrix (~individual+BAZ1B). R code provided.

Table S5. Number of genes differentially expressed following BAZ1B data in our numerical analysis compared to an analysis conducted on randomized HipSci data, using Edg2 function (see table S4).

References (61–89)

View/request a protocol for this paper from *Bio-protocol*.

REFERENCES AND NOTES

1. C. Theofanopoulou, S. Gastaldon, T. O'Rourke, B. D. Samuels, A. Messner, P. T. Martins, F. Delogu, S. Alamri, C. Boeckx, Self-domestication in *Homo sapiens*: Insights from comparative genomics. *PLOS ONE* **12**, e0185306 (2017).

2. B. Hare, Survival of the friendliest: *Homo sapiens* evolved via selection for prosociality. *Annu. Rev. Psychol.* **68**, 155–186 (2017).

3. R. W. Wrangham, *The Goodness Paradox: How Evolution Made Us Both More and Less Violent* (Profile Books Ltd., 2019).

4. M. R. Sánchez-Villagra, C. P. van Schaik, Evaluating the self-domestication hypothesis of human evolution. *Evol. Anthropol.* **28**, 133–143 (2019).

5. A. S. Wilkins, R. W. Wrangham, W. T. Fitch, The "domestication syndrome" in mammals: A unified explanation based on neural crest cell behavior and genetics. *Genetics* **197**, 795–808 (2014).

6. B. R. Pober, Williams-Beuren syndrome. *N. Engl. J. Med.* **362**, 239–252 (2010).

7. S. J. Sanders, A. G. Ercan-Sencicek, V. Hus, R. Luo, M. T. Murtha, D. Moreno-de-Luca, S. H. Chu, M. P. Moreau, A. R. Gupta, S. A. Thomson, C. E. Mason, K. Bilguvar, P. B. S. Celestino-Soper, M. Choi, E. L. Crawford, L. Davis, N. R. Davis Wright, R. M. Dhodapkar, M. DiCola, N. M. DiLullo, T. V. Fernandez, V. Fielding-Singh, D. O. Fishman, S. Frahm, R. Garagaloyan, G. S. Goh, S. Kammela, L. Klei, J. K. Lowe, S. C. Lund, A. D. McGrew, K. A. Meyer, W. J. Moffat, J. D. Murdoch, B. J. O'Roak, G. T. Ober, R. S. Pottenger, M. J. Raubeson, Y. Song, Q. Wang, B. L. Yaspan, T. W. Yu, I. R. Yurkiewicz, A. L. Beaudet, R. M. Cantor, M. Curland, D. E. Grice, M. Günel, R. P. Lifton, S. M. Mane, D. M. Martin, C. A. Shaw, M. Sheldon, J. A. Tischfield, C. A. Walsh, E. M. Morrow, D. H. Ledbetter, E. Fombonne, C. Lord, C. L. Martin, A. I. Brooks, J. S. Sutcliffe, E. H. Cook Jr., D. Geschwind, K. Roeder, B. Devlin, M. W. State, Multiple

recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. *Neuron* **70**, 863–885 (2011).

8. B. M. vonHoldt, E. Shuldiner, I. J. Koch, R. Y. Kartzinel, A. Hogan, L. Brubaker, S. Wanser, D. Stahler, C. D. L. Wynne, E. A. Ostrander, J. S. Sinsheimer, M. A. R. Udell, Structural variants in genes associated with human Williams-Beuren syndrome underlie stereotypical hypersociability in domestic dogs. *Sci. Adv.* **3**, e1700398 (2017).

9. A. V. Kukekova, L. J. Johnson, X. Xiang, S. Feng, S. Liu, H. M. Rando, A. V. Kharlamova, Y. Herbeck, N. A. Serdyukova, Z. Xiong, V. Beklemischeva, K. P. Koepfli, R. G. Gulevich, A. V. Vladimirova, J. P. Hekman, P. L. Perelman, A. S. Graphodatsky, S. J. O'Brien, X. Wang, A. G. Clark, G. M. Acland, L. N. Trut, G. Zhang, Red fox genome assembly identifies genomic regions associated with tame and aggressive behaviours. *Nat. Ecol. Evol.* **2**, 1479–1491 (2018).

10. C. Barnett, O. Yazgan, H. C. Kuo, S. Malakar, T. Thomas, A. Fitzgerald, W. Harbour, J. J. Henry, J. E. Krebs, Williams syndrome transcription factor is critical for neural crest cell function in *Xenopus laevis*. *Mech. Dev.* **129**, 324–338 (2012).

11. A. Ashe, D. K. Morgan, N. C. Whitelaw, T. J. Bruxner, N. K. Vickaryous, L. L. Cox, N. C. Butterfield, C. Wicking, M. E. Blewitt, S. J. Wilkins, G. J. Anderson, T. C. Cox, E. Whitelaw, A genome-wide screen for modifiers of transgene variegation identifies genes with critical roles in development. *Genome Biol.* **9**, R182 (2008).

12. B. M. vonHoldt, S. S. Ji, M. L. Aardema, D. R. Stahler, M. A. R. Udell, J. S. Sinsheimer, Activity of genes with functions in Human Williams-Beuren syndrome is impacted by mobile element insertions in the gray wolf genome. *Genome Biol. Evol.* **10**, 1546–1553 (2018).

13. S. Peyrégne, M. J. Boyle, M. Dannemann, K. Prüfer, Detecting ancient positive selection in humans using extended lineage sorting. *Genome Res.* **2017**, 1563–1572 (2017).

14. F. Racimo, Testing for ancient selection using cross-population allele frequency differentiation. *Genetics* **202**, 733–750 (2016).

15. M. Kuhlwilm, C. Boeckx, A catalog of single nucleotide changes distinguishing modern humans from archaic hominins. *Sci. Rep.* **9**, 8463 (2019).

16. A. Adamo, S. Atashpaz, P. L. Germain, M. Zanella, G. D'Agostino, V. Albertin, J. Chenoweth, L. Micale, C. Fusco, C. Unger, B. Augello, O. Palumbo, B. Hamilton, M. Carella, E. Donti, G. Pruneri, A. Selicorni, E. Biamino, P. Prontera, R. McKay, G. Merla, G. Testa, 7q11.23 dosage-dependent dysregulation in human pluripotent stem cells affects transcriptional programs in disease-relevant lineages. *Nat. Genet.* **47**, 132–141 (2015).

17. C. Fusco, L. Micale, B. Augello, M. Teresa Pellico, D. Menghini, P. Alfieri, M. Cristina Digilio, B. Mandriani, M. Carella, O. Palumbo, S. Vicari, G. Merla, Smaller and larger deletions of the Williams Beuren syndrome region implicate genes involved in mild facial phenotype, epilepsy and autistic traits. *Eur. J. Hum. Genet.* **22**, 64–70 (2014).

18. J. Meng, X. T. Zhang, X. L. Liu, L. Fan, C. Li, Y. Sun, X. H. Liang, J. B. Wang, Q. B. Mei, F. Zhang, T. Zhang, WSTF promotes proliferation and invasion of lung cancer cells by inducing EMT via PI3K/Akt and IL-6/STAT3 signaling pathways. *Cell. Signal.* **28**, 1673–1682 (2016).

19. R. Bajpai, D. A. Chen, A. Rada-Iglesias, J. Zhang, Y. Xiong, J. Helms, C. P. Chang, Y. Zhao, T. Swigut, J. Wysocka, CHD7 cooperates with PBAF to control multipotent neural crest formation. *Nature* **463**, 958–962 (2010).

20. C. Barnett, J. E. Krebs, WSTF does it all: A multifunctional protein in transcription, repair, and replication. *Biochem. Cell Biol.* **89**, 12–23 (2011).

21. R. A. Poot, L. Bozhenok, D. L. C. van den Berg, S. Steffensen, F. Ferreira, M. Grimaldi, N. Gilbert, J. Ferreira, P. D. Varga-Weisz, The Williams syndrome transcription factor interacts with PCNA to target chromatin remodelling by ISWI to replication foci. *Nat. Cell Biol.* **6**, 1236–1244 (2004).

22. A. E. Culver-Cochran, B. P. Chadwick, Loss of WSTF results in spontaneous fluctuations of heterochromatin formation and resolution, combined with substantial changes to gene expression. *BMC Genomics* **14**, 740 (2013).

23. A. Wilderman, J. VanOudenhove, J. Kron, J. P. Noonan, J. Cotney, High-resolution epigenomic atlas of human embryonic craniofacial development. *Cell Rep.* **23**, 1581–1597 (2018).

24. R. F. Spokony, Y. Aoki, N. Saint-Germain, E. Magner-Fink, J.-P. Saint-Jeannet, The transcription factor Sox9 is required for cranial neural crest development in *Xenopus*. *Development* **129**, 421–432 (2002).

25. F. Santagati, F. M. Rijli, Cranial neural crest and the building of the vertebrate head. *Nat. Rev. Neurosci.* **4**, 806–818 (2003).

26. S. Bhatt, R. Diaz, P. A. Trainor, Signals and switches in mammalian neural crest cell differentiation. *Cold Spring Harb. Perspect. Biol.* **5**, a008326 (2013).

27. S. O. Ko, I. H. Chung, X. Xu, S. Oka, H. Zhao, E. S. Cho, C. Deng, Y. Chai, Smad4 is required to regulate the fate of cranial neural crest cells. *Dev. Biol.* **312**, 435–447 (2007).

28. Y. Mishina, T. N. Snider, Neural crest cell signaling pathways critical to cranial bone development and pathology. *Exp. Cell Res.* **325**, 138–147 (2014).

29. P.-L. Germain, G. Testa, Taming human genetic variability: Transcriptomic meta-analysis guides the experimental design and interpretation of iPSC-based disease modeling. *Stem Cell Rep.* **8**, 1784–1796 (2017).

30. I. Schanze, J. Bunt, J. W. C. Lim, D. Schanze, R. J. Dean, M. Alders, P. Blanchet, T. Attié-Bitach, S. Berland, S. Boogert, S. Boppudi, C. J. Bridges, M. T. Cho, W. B. Dobyns, D. Donnai, J. Douglas, D. L. Earl, T. J. Edwards, L. Faivre, B. Fregeau, D. Genevieve, M. Gérard,

V. Gatinois, M. Holder-Espinasse, S. F. Huth, K. Izumi, B. Kerr, E. Lacaze, P. Lakeman, S. Mahida, G. M. Mirzaa, S. M. Morgan, C. Nowak, H. Peeters, F. Petit, D. T. Pilz, J. Puechberty, E. Reinstein, J. B. Rivière, A. B. Santani, A. Schneider, E. H. Sherr, C. Smith-Hicks, I. Wieland, E. Zackai, X. Zhao, R. M. Gronostajski, M. Zenker, L. J. Richards, *NFIB* haploinsufficiency is associated with intellectual disability and macrocephaly. *Am. J. Hum. Genet.* **103**, 752–768 (2018).

31. M. Gabriele, A. T. Vulto-van Silfhout, P. L. Germain, A. Vitriolo, R. Kumar, E. Douglas, E. Haan, K. Kosaki, T. Takenouchi, A. Rauch, K. Steindl, E. Frengen, D. Misceo, C. R. J. Pedurupillay, P. Stromme, J. A. Rosenfeld, Y. Shao, W. J. Craigen, C. P. Schaaf, D. Rodriguez-Buritica, L. Farach, J. Friedman, P. Thulin, S. D. McLean, K. M. Nugent, J. Morton, J. Nicholl, J. Andrieux, A. Stray-Pedersen, P. Chambon, S. Patrier, S. A. Lynch, S. Kjaergaard, P. M. Tørring, C. Brasch-Andersen, A. Ronan, A. van Haeringen, P. J. Anderson, Z. Powis, H. G. Brunner, R. Pfundt, J. H. M. Schuurs-Hoeijmakers, B. W. M. van Bon, S. Lelieveld, C. Gilissen, W. M. Nillesen, L. E. L. M. Vissers, J. Gecz, D. A. Koolen, G. Testa, B. B. A. de Vries, *YY1* haploinsufficiency causes an intellectual disability syndrome featuring transcriptional and chromatin dysfunction. *Am. J. Hum. Genet.* **100**, 907–925 (2017).

32. M. Gabriele, A. Lopez Tobon, G. D'Agostino, G. Testa, The chromatin basis of neurodevelopmental disorders: Rethinking dysfunction along the molecular and temporal axes. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **84**, 306–327 (2018).

33. A. P. McMahon, J. E. Champion, J. A. McMahon, V. P. Sukhatme, Developmental expression of the putative transcription factor Egr-1 suggests that Egr-1 and c-fos are coregulated in some tissues. *Development* **108**, 281–287 (1990).

34. J. Dalcq, V. Pasque, A. Ghaye, A. Larbuisson, P. Motte, J. A. Martial, M. Muller, RUNX3, EGR1 and SOX9B form a regulatory cascade required to modulate BMP-signaling during cranial cartilage development in zebrafish. *PLOS ONE* **7**, e50140 (2012).

35. A. Werner, S. Iwasaki, C. A. McGourty, S. Medina-Ruiz, N. Teerikorpi, I. Fedrigo, N. T. Ingolia, M. Rape, Cell-fate determination by ubiquitin-dependent regulation of translation. *Nature* **525**, 523–527 (2015).

36. L. P. Sanford, I. Ormsby, A. C. Gittenberger-de Groot, H. Sariola, R. Friedman, G. P. Boivin, E. L. Cardell, T. Doetschman, TGFβ2 knockout mice have multiple developmental defects that are non-overlapping with other TGFβ knockout phenotypes. *Development* **124**, 2659–2670 (1997).

37. C. Boileau, D. C. Guo, N. Hanna, E. S. Regalado, D. Detaint, L. Gong, M. Varret, S. K. Prakash, A. H. Li, H. d'indy, A. C. Braverman, B. Grandchamp, C. S. Kwartler, L. Gouya, R. L. Santos-Cortez, M. Abifadel, S. M. Leal, C. Muti, J. Shendure, M. S. Gross, M. J. Rieder, A. Vahanian, D. A. Nickerson, J. B. Michel; National, Lung Heart Project Blood Institute (NHLBI) Go Exome Sequencing, G. Jondeau, D. M. Milewicz, TGFβ2 mutations cause familial thoracic aortic aneurysms and dissections associated with mild systemic features of Marfan syndrome. *Nat. Genet.* **44**, 916–921 (2012).

38. O. L. Wapinski, T. Vierbuchen, K. Qu, Q. Y. Lee, S. Chanda, D. R. Fuentes, P. G. Giresi, Y. H. Ng, S. Marro, N. F. Neff, D. Drechsel, B. Martynoga, D. S. Castro, A. E. Webb, T. C. Südhof, A. Brunet, F. Guillemot, H. Y. Chang, M. Wernig, Hierarchical mechanisms for direct reprogramming of fibroblasts to neurons. *Cell* **155**, 621–635 (2013).

39. S. J. Hong, H. J. Choi, S. Hong, Y. Huh, H. Chae, K. S. Kim, Transcription factor GATA-3 regulates the transcriptional activity of dopamine beta-hydroxylase by interacting with Sp1 and AP4. *Neurochem. Res.* **33**, 1821–1831 (2008).

40. N. de Crozé, F. Maczkowiak, A. H. Monsoro-Burq, Reiterative AP2a activity controls sequential steps in the neural crest gene regulatory network. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 155–160 (2011).

41. A. Rada-Iglesias, S. L. Prescott, J. Wysocka, Human genetic variation within neural crest enhancers: molecular and phenotypic implications. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **368**, 20120360 (2013).

42. R. S. Lacruz, C. B. Stringer, W. H. Kimbel, B. Wood, K. Harvati, P. O'Higgins, T. G. Bromage, J.-L. Arsuaga, The evolutionary history of the human face. *Nat. Ecol. Evol.* **3**, 726–736 (2019).

43. D. Gokhman, L. Agranat-Tamir, G. Housman, M. Nissim-Rafinia, M. Nieves-Colón, H. Gu, Recent regulatory changes shaped human facial and vocal anatomy. bioRxiv 106955 [**Preprint**]. https://doi.org/10.1101/106955.

44. C. Darwin, *The Variation of Animals and Plants under Domestication* (J. Murray, 1868).

45. C. Darwin, *The Descent of Man: And Selection in Relation to Sex* (J. Murray, 1871).

46. M. Brüne, On human self-domestication, psychiatry, and eugenics. *Philos. Ethics Humanit. Med.* **2**, 21 (2007).

47. P. Claes, J. Roosenboom, J. D. White, T. Swigut, D. Sero, J. Li, M. K. Lee, A. Zaidi, B. C. Mattern, C. Liebowitz, L. Pearson, T. González, E. J. Leslie, J. C. Carlson, E. Orlova, P. Suetens, D. Vandermeulen, E. Feingold, M. L. Marazita, J. R. Shaffer, J. Wysocka, M. D. Shriver, S. M. Weinberg, Genome-wide mapping of global-to-local genetic effects on human facial shape. *Nat. Genet.* **50**, 414–423 (2018).

48. A. Rada-Iglesias, R. Bajpai, S. Prescott, S. A. Brugmann, T. Swigut, J. Wysocka, Epigenomic annotation of enhancers predicts transcriptional regulators of human neural crest. *Cell Stem Cell* **11**, 633–648 (2012).

49. C. Buecker, J. Wysocka, Enhancers as information integration hubs in development: Lessons from genomics. *Trends Genet.* **28**, 276–284 (2012).

50. S. C. Vernes, P. L. Oliver, E. Spiteri, H. E. Lockstone, R. Puliyadi, J. M. Taylor, J. Ho, C. Mombereau, A. Brewer, E. Lowy, J. Nicod, M. Groszer, D. Baban, N. Sahgal, J. B. Cazier, J. Ragoussis, K. E. Davies, D. H. Geschwind, S. E. Fisher, Foxp2 regulates gene networks implicated in neurite outgrowth in the developing brain. *PLOS Genet.* **7**, e1002145 (2011).

51. R. Wang, C. C. Chen, E. Hara, M. V. Rivas, P. L. Roulhac, J. T. Howard, M. Chakraborty, J. N. Audet, E. D. Jarvis, Convergent differential regulation of SLIT-ROBO axon guidance genes in the brains of vocal learners. *J. Comp. Neurol.* **523**, 892–906 (2015).

52. L. Menendez, M. J. Kulik, A. T. Page, S. S. Park, J. D. Lauderdale, M. L. Cunningham, S. Dalton, Directed differentiation of human pluripotent cells to neural crest stem cells. *Nat. Protoc.* **8**, 203–212 (2013).

53. R. Blecher-Gonen, Z. Barnett-Itzhaki, D. Jaitin, D. Amann-Zalcenstein, D. Lara-Astiaso, I. Amit, High-throughput chromatin immunoprecipitation for genome-wide mapping of in vivo protein-DNA interactions and epigenomic states. *Nat. Protoc.* **8**, 539–554 (2013).

54. P.-L. Germain, A. Vitriolo, A. Adamo, P. Laise, V. Das, G. Testa, RNAontheBENCH: Computational and empirical resources for benchmarking RNAseq quantification and differential expression methods. *Nucleic Acids Res.* **44**, 5054–5067 (2016).

55. I. Streeter, P. W. Harrison, A. Faulconbridge; The HipSci Consortium, P. Flicek, H. Parkinson, L. Clarke, The human-induced pluripotent stem cell initiative-data resources for cellular genetics. *Nucleic Acids Res.* **45**, D691–D697 (2017).

56. H. Kilpinen, A. Goncalves, A. Leha, V. Afzal, K. Alasoo, S. Ashford, S. Bala, D. Bensaddek, F. P. Casale, O. J. Culley, P. Danecek, A. Faulconbridge, P. W. Harrison, A. Kathuria, D. McCarthy, S. A. McCarthy, R. Meleckyte, Y. Memari, N. Moens, F. Soares, A. Mann, I. Streeter, C. A. Agu, A. Alderton, R. Nelson, S. Harper, M. Patel, A. White, S. R. Patel, L. Clarke, R. Halai, C. M. Kirton, A. Kolb-Kokocinski, P. Beales, E. Birney, D. Danovi, A. I. Lamond, W. H. Ouwehand, L. Vallier, F. M. Watt, R. Durbin, O. Stegle, D. J. Gaffney, Common genetic variation drives molecular heterogeneity in human iPSCs. *Nature* **546**, 370–375 (2017).

57. G. Tan, B. Lenhard, TFBSTools: An R/Bioconductor package for transcription factor binding site analysis. *Bioinformatics* **32**, 1555–1556 (2016).

58. M. Meyer, M. Kircher, M. T. Gansauge, H. Li, F. Racimo, S. Mallick, J. G. Schraiber, F. Jay, K. Prufer, C. de Filippo, P. H. Sudmant, C. Alkan, Q. Fu, R. Do, N. Rohland, A. Tandon, M. Siebauer, R. E. Green, K. Bryc, A. W. Briggs, U. Stenzel, J. Dabney, J. Shendure, J. Kitzman, M. F. Hammer, M. V. Shunkov, A. P. Derevianko, N. Patterson, A. M. Andres, E. E. Eichler, M. Slatkin, D. Reich, J. Kelso, S. Paabo, A high-coverage genome sequence from an archaic Denisovan individual. *Science* **338**, 222–226 (2012).

59. K. Prüfer, F. Racimo, N. Patterson, F. Jay, S. Sankararaman, S. Sawyer, A. Heinze, G. Renaud, P. H. Sudmant, C. de Filippo, H. Li, S. Mallick, M. Dannemann, Q. Fu, M. Kircher, M. Kuhlwilm, M. Lachmann, M. Meyer, M. Ongyerth, M. Siebauer, C. Theunert, A. Tandon, P. Moorjani, J. Pickrell, J. C. Mullikin, S. H. Vohr, R. E. Green, I. Hellmann, P. L. Johnson, H. Blanche, H. Cann, J. O. Kitzman, J. Shendure, E. E. Eichler, E. S. Lein, T. E. Bakken, L. V. Golovanova, V. B. Doronichev, M. V. Shunkov, A. P. Derevianko, B. Viola, M. Slatkin, D. Reich, J. Kelso, S. Pääbo, The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43–49 (2014).

60. K. Prüfer, C. de Filippo, S. Grote, F. Mafessoni, P. Korlević, M. Hajdinjak, B. Vernot, L. Skov, P. Hsieh, S. Peyrégne, D. Reher, C. Hopfe, S. Nagel, T. Maricic, Q. Fu, C. Theunert, R. Rogers, P. Skoglund, M. Chintalapati, M. Dannemann, B. J. Nelson, F. M. Key, P. Rudan, Ž. Kućan, I. Gušić, L. V. Golovanova, V. B. Doronichev, N. Patterson, D. Reich, E. E. Eichler, M. Slatkin, M. H. Schierup, A. M. Andrés, J. Kelso, M. Meyer, S. Pääbo, A high-coverage Neandertal genome from Vindija Cave in Croatia. *Science* **358**, 655–658 (2017).

61. M. Barembaum, T. A. Moreno, C. LaBonne, J. Sechrist, M. Bronner-Fraser, Noelin-1 is a secreted glycoprotein involved in generation of the neural crest. *Nat. Cell Biol.* **2**, 219–225 (2000).

62. D. Huang, Y. Wang, L. Xu, L. Chen, M. Cheng, W. Shi, H. Xiong, D. Zalli, S. Luo, GLI2 promotes cell proliferation and migration through transcriptional activation of *ARHGEF16* in human glioma cells. *J. Exp. Clin. Cancer Res.* **37**, 247 (2018).

63. T. C. Südhof, Neuroligins and neurexins link synaptic function to cognitive disease. *Nature* **455**, 903–911 (2008).

64. G. Chen, J. Sima, M. Jin, K. Y. Wang, X. J. Xue, W. Zheng, Y. Q. Ding, X. B. Yuan, Semaphorin-3A guides radial migration of cortical neurons during development. *Nat. Neurosci.* **11**, 36–44 (2008).

65. E. Betters, Y. Liu, A. Kjaeldgaard, E. Sundström, M. I. García-Castro, Analysis of early human neural crest development. *Dev. Biol.* **344**, 578–592 (2010).

66. Q. Wu, W. Tang, Z. Luo, Y. Li, Y. Shu, Z. Yue, B. Xiao, L. Feng, DISC1 regulates the proliferation and migration of mouse neural stem/progenitor cells through Pax5, Sox2, Dll1 and Neurog2. *Front. Cell. Neurosci.* **11**, 261 (2017).

67. M. C. Horowitz, Y. Xi, D. L. Pflugh, D. G. T. Hesslein, D. G. Schatz, J. A. Lorenzo, A. L. M. Bothwell, Pax5-deficient mice exhibit early onset osteopenia with increased osteoclast progenitors. *J. Immunol.* **173**, 6583–6591 (2004).

68. H. Hsu, D. L. Lacey, C. R. Dunstan, I. Solovyev, A. Colombero, E. Timms, H. L. Tan, G. Elliott, M. J. Kelley, I. Sarosi, L. Wang, X. Z. Xia, R. Elliott, L. Chiu, T. Black, S. Scully, C. Capparelli, S. Morony, G. Shimamoto, M. B. Bass, W. J. Boyle, Tumor necrosis factor receptor family member RANK mediates osteoclast differentiation

69. and activation induced by osteoprotegerin ligand. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 3540–3545 (1999).

69. C. A. McGourty, C. Akopian, C. Walsh, A. Gorur, A. Werner, R. Schekman, D. Bautista, M. Rape, Regulation of the CUL3 ubiquitin ligase by a calcium-dependent co-adaptor. *Cell* **167**, 525–538.e14 (2016).

70. K. Oka, M. J. Honda, E. Tsuruga, Y. Hatakeyama, K. Isokawa, Y. Sawa, Roles of collagen and periostin expression by cranial neural crest cells during soft palate development. *J. Histochem. Cytochem.* **60**, 57–68 (2011).

71. H. Rios, S. V. Koushik, H. Wang, J. Wang, H. M. Zhou, A. Lindsley, R. Rogers, Z. Chen, M. Maeda, A. Kruzynska-Frejtag, J. Q. Feng, S. J. Conway, Periostin null mice exhibit dwarfism, incisor enamel defects, and an early-onset periodontal disease-like phenotype. *Mol. Cell. Biol.* **25**, 11131–11144 (2005).

72. J. P. Golding, P. Trainor, R. Krumlauf, M. Gassmann, Defects in pathfinding by cranial neural crest cells in mice lacking the neuregulin receptor ErbB4. *Nat. Cell Biol.* **2**, 103–109 (2000).

73. S. Burden, Y. Yarden, Neuregulins and their receptors: A versatile signaling module in organogenesis and oncogenesis. *Neuron* **18**, 847–855 (1997).

74. G. Andreoletti, E. G. Seaby, J. M. Dewing, I. O'Kelly, K. Lachlan, R. D. Gilbert, S. Ennis, *AMMECR1*: Single point mutation causes developmental delAay, midface hypoplasia and elliptocytosis. *J. Med. Genet.* **54**, 269–277 (2017).

75. B. Tumiene, Ž. Čiuladaitė, E. Preikšaitienė, R. Mameniškienė, A. Utkus, V. Kučinskas, Phenotype comparison confirms *ZMYND11* as a critical gene for 10p15.3 microdeletion syndrome. *J. Appl. Genet.* **58**, 467–474 (2017).

76. J. M. Cobben, M. M. Weiss, F. S. van Dijk, R. de Reuver, C. de Kruiff, W. Pondaag, R. C. Hennekam, H. G. Yntema, A *de novo* mutation in *ZMYND11*, a candidate gene for 10p15.3 deletion syndrome, is associated with syndromic intellectual disability. *Eur. J. Med. Genet.* **57**, 636–638 (2014).

77. K. F. Oram, E. A. Carver, T. Gridley, Slug expression during organogenesis in mice. *Anat. Rec. A Discov. Mol. Cell. Evol. Biol.* **271**, 189–191 (2003).

78. W. S. Wu, S. Heinrichs, D. Xu, S. P. Garrison, G. P. Zambetti, J. M. Adams, A. T. Look, Slug antagonizes p53-mediated z progenitors by repressing *puma*. *Cell* **123**, 641–653 (2005).

79. L. Garbes, K. Kim, A. Rieß, H. Hoyer-Kuhn, F. Beleggia, A. Bevot, M. J. Kim, Y. H. Huh, H. S. Kweon, R. Savarirayan, D. Amor, P. M. Kakadia, T. Lindig, K. O. Kagan, J. Becker, S. A. Boyadjiev, B. Wollnik, O. Semler, S. K. Bohlander, J. Kim, C. Netzer, Mutations in *SEC24D*, encoding a component of the COPII machinery, cause a syndromic form of osteogenesis imperfecta. *Am. J. Hum. Genet.* **96**, 432–439 (2015).

80. A. J. Griffith, L. K. Sprunger, D. A. Sirko-Osadsa, G. E. Tiller, M. H. Meisler, M. L. Warman, Marshall syndrome associated with a splicing defect at the *COL11A1* locus. *Am. J. Hum. Genet.* **62**, 816–823 (1998).

81. E. Roessler, Y. Ma, M. V. Ouspenskaia, F. Lacbawan, C. Bendavid, C. Dubourg, P. A. Beachy, M. Muenke, Truncating loss-of-function mutations of *DISP1* contribute to holoprosencephaly-like microform features in humans. *Hum. Genet.* **125**, 393–400 (2009).

82. J. Punetha, A. Kesari, E. P. Hoffman, M. Gos, A. Kamińska, A. Kostera-Pruszczyk, I. Hausmanowa-Petrusewicz, Y. Hu, Y. Zou, C. G. Bönnemann, M. Jędrzejowska, NovelCol12A1variant expands the clinical picture of congenital myopathies with extracellular matrix defects. *Muscle Nerve* **55**, 277–281 (2017).

83. T. O'Rourke, C. Boeckx, Converging roles of glutamate receptors in domestication and prosociality. bioRxiv 439869 [**Preprint**]. https://doi.org/10.1101/439869.

84. S. Srivastava, H. Engels, I. Schanze, K. Cremer, T. Wieland, M. Menzel, M. Schubach, S. Biskup, M. Kreiß, S. Endele, T. M. Strom, D. Wieczorek, M. Zenker, S. Gupta, J. Cohen, A. M. Zink, S. B. Naidu, Loss-of-function variants in *HIVEP2* are a cause of intellectual disability. *Eur. J. Hum. Genet.* **24**, 556–561 (2016).

85. H. Lei, Z. Yan, X. Sun, Y. Zhang, J. Wang, C. Ma, Q. Xu, R. Wang, E. D. Jarvis, Z. Sun, Axon guidance pathways served as common targets for human speech/language evolution and related disorders. *Brain Lang.* **174**, 1–8 (2017).

86. E. Lausch, P. Hermanns, H. F. Farin, Y. Alanay, S. Unger, S. Nikkel, C. Steinwender, G. Scherer, J. Spranger, B. Zabel, A. Kispert, A. Superti-Furga, TBX15 mutations cause craniofacial dysmorphism, hypoplasia of scapula and pelvis, and short stature in Cousin syndrome. *Am. J. Hum. Genet.* **83**, 649–655 (2008).

87. M. K. Lee, J. R. Shaffer, E. J. Leslie, E. Orlova, J. C. Carlson, E. Feingold, M. L. Marazita, S. M. Weinberg, Genome-wide association study of facial morphology reveals novel associations with *FREM1* and *PARK2*. *PLOS ONE* **12**, e0176566 (2017).

88. J. Iwata, C. Parada, Y. Chai, The mechanism of TGF-β signaling during palate development. *Oral Dis.* **17**, 733–744 (2011).

89. Y. Zhang, J. Su, J. Yu, X. Bu, T. Ren, X. Liu, L. Yao, 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).