7q11.23 Syndromes Reveal BAZ1B as a Master Regulator of the Modern Human Face and Validate the Self-Domestication Hypothesis

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], Natascia Malerba[7], Adrianos Skaros[1,2], Sebastiano Trattaro[1,2], Pierre-Luc Germain[1,2,8], Giuseppe Merla[7], Alvaro Rada-Iglesias[5,9], Cedric Boeckx[3,4,10] and Giuseppe Testa[1,2,†]

[1] Department of Oncology and Hemato-Oncology, University of Milan, Italy.

[2] Department of Experimental Oncology, Laboratory of Stem Cell Epigenetics, European Institute of Oncology, Milan, Italy.

[3] University of Barcelona, Spain.

[4] University of Barcelona, Institute of Complex Systems, Barcelona, Spain.

[5] Center for Molecular Medicine Cologne (CMMC), University of Cologne, Germany.

[6] Institute of Human Genetics, CMMC, University Hospital, Cologne, Germany.

[7] Division of Medical Genetics, Fondazione IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo, Foggia, Italy.

[8] Current address: Brain Research Institute, University of Zürich, Switzerland.

[9] Institute of Biomedicine and Biotechnology of Cantabria (IBBTEC), University of Cantabria, Spain.

[10] Institute for Advanced Studies and Research (ICREA), Barcelona, Spain.

* These authors contributed equally to this work

† Corresponding author. Email: giuseppe.testa@unimi.it; giuseppe.testa@ieo.it

Abstract

Symmetrical 7q11.23 dosage alterations cause craniofacial and cognitive/behavioral phenotypes that provide a privileged entry point into the evolution of the modern human face and (pro-) sociality. We undertook a functional dissection of chromatin remodeler BAZ1B in neural crest stem cells (NCSCs) from a uniquely informative cohort of typical and atypical patients harboring 7q11.23 Copy Number Variants (CNVs). Our results reveal a key contribution of BAZ1B to NCSC *in vitro* induction and migration, coupled with a crucial involvement in neural crest (NC)-specific transcriptional circuits and distal regulation. By intersecting our experimental data with new paleogenetic analyses comparing modern and archaic humans, we uncover a modern-specific enrichment for regulatory changes both in BAZ1B and its experimentally defined downstream targets, thereby providing the first empirical validation of the self-domestication hypothesis and positioning BAZ1B as a master regulator of the modern human face.

One Sentence Summary: BAZ1B dosage shapes the modern human face.

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, 2*). Accordingly, the self-domestication hypothesis predicts 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 (*3*). Both the neurocristopathic basis of domestication and its extension to the evolution of AMHs remain however to be tested experimentally. Williams-Beuren syndrome (WBS, OMIM 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, hereafter WBSCR), represent a paradigmatic pair of neurodevelopmental conditions whose NC-related craniofacial dysmorphisms (Fig. S1A) and cognitive/behavioral traits (*4, 5*) are uniquely and symmetrically aligned to the domestication-related traits of AMHs. Structural variants in WBS genes have been shown to underlie stereotypical hypersociability in domestic dogs (*6*) and within the WBSCR the following lines of evidence point to the chromatin regulator *BAZ1B* as a key gene underlying domestication-relevant craniofacial features: i) its crucial role in NC maintenance and migration in *Xenopus Laevis* and the craniofacial defects observed in knock-out mice (*7, 8*); ii) the observation that its expression is impacted by domestication-related events in canines (*9*); iii) the first formulation of the neurocristopathic hypothesis of domestication, which included *BAZ1B* among the genes influencing NC development (*3*); iv) the most comprehensive studies focusing on regions of the modern human genome associated with selective sweep signals compared to Neanderthals/Denisovans (hereafter 'archaics') (*10, 11*), 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 (> 90%) changes in modern humans for which archaic humans carry the ancestral state, which found *BAZ1B* enriched for mutations in modern humans (most of which fall in the regulatory regions of the gene) (*12*). Through the thus far largest cohort of 7q11.23 patient-derived induced pluripotent stem cell (iPSC) lines, our previous work had

5   revealed major disease-relevant dysregulation that was already apparent at the pluripotent state and was further exacerbated upon differentiation (*13*). Here we build on this insight to specifically dissect the impact of BAZ1B dosage alterations on the NC of WBS and 7dupASD patients and to empirically test the self-domestication hypothesis by interrogating the correlative evidence from paleogenomic analyses with the experimentally determined BAZ1B-dependent circuits underlying

10 craniofacial morphogenesis. We demonstrate a major contribution of BAZ1B to the modern human face, validating the prediction that posits the evolution of the human face as an instance of mild neurocristopathy.

Results

15 ***Establishment and validation of an extensive cohort of patient-specific BAZ1B-interfered NCSC lines***

In order to dissect the role of BAZ1B in the paradigmatic craniofacial dysmorphisms that characterize WBS and 7dupASD, we started from our previous characterization of WBS and 7dupASD patient-specific iPSC lines and differentiated derivatives (*13*) and selected a cohort of

20 11 NCSC lines (4 from WBS patients, 3 from 7dupASD patients and 4 from control individuals). These represent the largest cohort of patient-specific NCSCs described so far and we used transcriptomic profiling to validate their cranial identity as a privileged entry point into disease-pathogenesis, given the centrality of the cranial NC in the development of the face. We then

knocked-down (KD) BAZ1B via RNA interference in all lines across the three genetic conditions, including also NCSCs derived from a particularly informative atypical WBS patient (hereafter atWBS) bearing a partial deletion of the region that spares *BAZ1B* and six additional genes (*14*) (Fig. 1A). In order to establish a high-resolution gradient of BAZ1B dosages, we selected two distinct 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. KD efficiency was evaluated both at the RNA level by quantitative PCR (qPCR) (Fig. 1B; 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 interference impairs NCSC migration and induction

NCSCs need to migrate in order to reach specific target regions in the developing embryo and give rise to distinct cell types and tissues, including craniofacial structures. BAZ1B KD was shown to affect the migration but not the induction of the NC in *Xenopus laevis,* and to promote cancer cell invasion in different lung cancer cell lines (*7, 15*). We thus hypothesized that the *BAZ1B* dosage imbalances entailed in the 7q11.23 syndromes could result in a defective regulation of NCSC migration and might hence underlie the NC-related alterations typical of WBS and 7dupASD patients. 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, 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, S1F). Interestingly,

in contrast to the previous observations from *Xenopus laevis* (*7*), we also observed a minor delay in NC induction as a consequence of BAZ1B KD (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 (*16*). In particular, starting from 2/3 days after attachment of EBs, we observed a lower number of outgrowing cells in the KD line (Fig. 1D, day 7 and 10), coupled with an evidently higher cell mortality. Cells were eventually able to acquire the typical NC morphology, even though lower differentiation efficiency was evident, as shown by images taken at day 12. In addition, the delay in NC formation was associated with a downregulation of well-established critical regulators of NC migration and maintenance, including *NR2F1*, *NR2F2*, *TFAP2A* and *SOX9* (Fig. 1E). These findings 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.

BAZ1B interference disrupts key neural crest-specific transcriptional circuits

A core insight from our previous work had been the discovery that already at the pluripotent state 7q11.23 CNVs cause transcriptional dysregulation in disease-relevant pathways, which are then selectively amplified along development in a tissue-specific manner (*13*). Given the critical role of BAZ1B as transcriptional regulator in different cell and animal models (*17–19*), as well as its functional involvement in human NC from the above described findings, we thus hypothesized that its dosage could significantly contribute to the transcriptional dysregulation in 7q11.23 CNV-NCSCs. 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 (*20–25*) 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 (*26*), 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).

In order to dissect it, we thus resorted to a combination of classical pairwise comparative analysis, contrasting shBAZ1B-interfered NCSC lines (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 FDR < 0.1 (1192 with p <0.01 and FDR < 0.25) whose transcriptional levels followed BAZ1B dosage, either in a direct (202; 539 with p < 0.01 and FDR < 0.25) or inverse (246; 653 with p < 0.01 and FDR < 0.25) fashion. In addition, genes identified only in the regression analysis included around 90% of the differentially expressed genes (DEGs) (27/29, 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

8

distinguishable effect of sh1, resulting in particularly informative gradient of dosages over the scr control lines.

Particularly noteworthy among the genes that we found inversely correlated with *BAZ1B* levels by highest fold changes (FC > 2) were the following crucial regulators of cranial NC: i) *POSTN*, which plays a key role in cranial NC-mediated soft palate development (*27*) and is responsible for heart valve defects and periodontal disease-like phenotypes in KO mice (*28*); and ii) *ERBB4* which is important for skeletal muscle development and NC migration (*29*) and causes heart defects and aberrant cranial NC migration in deficient mice (*30*) (Fig. 2B). Genes that instead directly follow BAZ1B levels include: i) *OLFM1,* whose overexpression promotes an increased and continued production of NC cells (*31*); ii) *TNFRSF11B,* involved in bone resorption and osteoclast activation (*32*) (Fig. 2B); iii) *CUL3*, which has been involved in craniofacial morphogenesis and ubiquitinates TCOF1 and NOLC1 as a master regulator of NC specification (*33*); and iv) the CUL3 adaptor *KLHL12* (*34*). As both *KLHL12* and *TCOF1*, acting up- and down-stream of *CUL3* respectively, are involved in craniofacial morphogenesis (*35*), these data highlight a convergent BAZ1B dosage-dependent dysregulation of the foundational *CUL3*-centered regulatory axis orchestrating NC-mediated craniofacial morphogenesis.

Additional particularly relevant genes inversely following BAZ1B dosage are: i) *TBX15*, one of the recently mapped loci associated to the variability of human facial shape (*36*); ii) *AMMECR1*, whose mutations cause midface hypoplasia and hearing impairment (*37*); iii) *ZMYND11*, a H3K36me3 reader that, when mutated, gives rise to dysmorphic facial features and mild intellectual disability (*38, 39*); iv) *SNAI2*, a critical regulator of NC-based organogenesis (*40*); and v) *BBC3*, which is repressed by *SNAI2* in response to DNA damage (*41*).

GO analysis performed on genes directly following BAZ1B levels suggested specific enrichments in biological process 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). Interestingly, by interrogating 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.

Finally, a Master Regulator Analysis identified candidate regulators of BAZ1B DEGs, including factors involved in enhancer marking (CEBPB, p300, RBBP5, HDAC2, KDM1A and TCF12), promoter activation (TBP, TAF1 and POL2) and chromatin remodeling (CTCF, RAD21 YY1) (Fig. 2E; S2C), several of which are themselves causative genes of intellectual disability (ID) syndromes with neurocristopathic involvement, as in the case of our recently identified Gabriele-de Vries syndrome caused by YY1 haploinsufficiency (42, 43). 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), itself among the genes inversely correlated with BAZ1B levels, which is implicated in cranial development (in animal models) (44, 45) and whose promoter has been recently shown to feature a bivalent state in human embryonic cranial NC (20, 33). 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 crosstalk between the two (Fig. S2C). Notably, two

differentially expressed targets of MXI1, *TGFB2* and *NFIB,* are also known to be involved in intellectual disability and craniofacial defects (*46–48*).

BAZ1B regulates the neural crest epigenome in a dosage-dependent manner

In order to mechanistically anchor transcriptional dysregulation to the dosage-dependent occupancy of BAZ1B, we set out to define its direct targets and investigate their chromatin state at both promoters and enhancers upon BAZ1B KD by chromatin immunoprecipitation coupled with sequencing (ChIP-seq). Given the absence of ChIP-grade antibodies, we 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 S3A, 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 the unprecedented resolution afforded by the patients' cohort with its underlying variability in chromatin annotation 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 308470 putative enhancer regions. Notably, BAZ1B binds 75% of its targets at their enhancer regions (6747 genes), with the

11

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 reflected also 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). Consistently, the analysis of BAZ1B bound regions uncovered major convergence with the binding motifs of critical NC regulators, including two motifs similar to that of TFAP2A: one *de novo* motif (reverse-engineered from BAZ1B ChIP-seq data and significantly aligning to HOMER motifs database) accounting for 21.49% and one *known* motif (already present in HOMER database and found significantly enriched in BAZ1B ChIP-seq data) accounting for 10.68 % of regions. Among *known* motifs we also found one for NEUROG2, accounting for 14%, and one accounting for 42% of BAZ1B bound regions equally associated to TAL1, TCF12, AP4 and ASCL1 (Fig. 3E). Thus, BAZ1B binding regions are enriched for target sites of major regulators of NC and its neural derivatives (*49*, *50*), among which TFAP2A stands out given its core role in neural borders formation and NC induction and differentiation (*51*) through the binding and stabilization of neural-crest specific enhancers, in concert with NR2F1, NR2F2 and EP300 (*52*).

Finally, 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), 176 and 25 regions differentially bound preferentially in, respectively, WBS (Fig. S4B) and 7dupASD (FDR < 0.1) (Fig. S4C).

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 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% (257/1192) 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 *vs* 97). The same held for DEGs that have a concordant differential distribution of H3K4me1 mark at enhancers (123 *vs* 55), underscoring the consistency of the impact of BAZ1B dosage on distal regulation (Fig. S4F). In contrast, a lower number of genes (about 30) 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 impacts active chromatin.

From this integrative analysis, we could thus finally identify a core set of 30 *bona fide* genes that: i) are DEGs in the RNAseq analysis (13 and 17, respectively following BAZ1B levels directly or inversely), ii) have their enhancers differentially marked in a concordant manner (*i.e.*, up in H3K27ac/H3K4me1 and down in H3K27me3 for the upregulated genes, and viceversa for the downregulated) and iii) are bound by BAZ1B at enhancers (Fig. 3G, S4H). This group includes key genes involved in cell migration and adhesion, such as *ARHGEF16*, *NRXN2*, *OLFM1* and *PLXNA4* (*31*, *53–55*), as well as cardinal NC markers, such as *HNK1* and *NR2F2* (*56*, *57*). Finally,

among the genes whose expression directly follows BAZ1B levels, and whose enhancers are bound by BAZ1B in a quantitatively dosage-dependent manner with a concomitant gain in H3K27ac, we found *PAX5*, which has been implicated in the regulation of neural stem cell proliferation and migration (*58*) and of the developmental balance between osteoblast and

5 osteoclast precursors (*59*). 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 preferentially activator role on the critical transcriptional circuits that define NC fate and function.

10 ***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 due to its previously reported role in the NC of model organisms (*3*, *7*, *8*). The recent observation that its expression is impacted by domestication-related mobile element insertion (MEI)

15 methylation in gray wolves (*9*) 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 (*12*). These convergent lines of enticing evidence notwithstanding, neither the neurocristopathic basis of domestication nor its extension to the evolution of the modern human

20 lineage (*i.e.*, self-domestication) have been empirically tested, hence keeping also BAZ1B-based domestication hypotheses to the theoretical ambit.

Having defined the molecular circuits through which BAZ1B shapes the modern human face, we set out to validate the self-domestication hypothesis and its neurocristopathic basis, carrying 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) and between wild/domesticated pairs of four paradigmatic species (*1*, *10–12*) (represented in Fig. 4A by skulls illustrating the more 'gracile' and 'juvenile' profile in AMH/dog relative to Neanderthal/wolf visible in the overall shape of the neurocranium, reduced proganthism, browridges and nasal projections). 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) (*10, 11*); ii) genes harboring (nearly) fixed mutations in moderns *vs* 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 *vs* archaics contains three classes: i) genes harboring high frequency changes (*12*) ii) genes harboring missense mutations (red barplot) and iii) genes enriched for mutations in regulatory regions (green barplot) (data based on (*12*)) (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 S1-15) 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 differentially expressed genes 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 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 (Tables S1-14). Importantly, 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 (ie. 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 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 BAZB 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 (*12*), the overall extent of overlap between genes impacted by BAZ1B dosage and our modern and archaic sets does not reveal significantly more hits for archaics. In fact, 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 BAZ1B level-directly correlated genes. Finally, the (lower) midface emerges as a particularly salient area of functionally relevant overlap (as illustrated in Fig. 4E and detailed in Table S15), given the specific genes that our analysis

16

unearthed: i) *COL11A1*, one of the few craniofacial genes highlighted across domestication studies (dog, house sparrow, pig breed) and which lies in a region of the human genome that resisted archaic introgression (*10*), is associated with Marshall syndrome and was also highlighted in a recent study on regulatory changes that shaped the modern human facial and vocal anatomy (*60*); ii) *XYLT1*, one of the 5 genes (along with *ACAN*, *SOX9*, *COL2A1*, and *NFIX*) that affect lower and midfacial protrusion and are among the top differentially methylated genes compared to archaics (Table S15); and iii) *NFIB*, which belongs to the same gene family as *NFIX* and shares some of its functions.

Discussion

Following on his seminal account of domestication systematized in *Variations of Animals and Plants under Domestication* (*61*), Charles Darwin first put forward the analogy between modern humans and domesticated species in *The Descent of Man* (*62*), although his emphasis on controlled breeding as a key aspect of domestication left his intuition largely unresolved (*63*). Since then the possibility that the anatomical and cognitive-behavioral hallmarks of AMHs could result from an evolutionary process bearing significant similarities to the domestication of animals has been refined into the full-fledged self-domestication hypothesis (*1, 2*), spurring considerable interest but failing to achieve epistemic closure largely due to 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 (*3*). This constituted a major conceptual advance, not least because it

generated for the first time 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 interrogating the genetic changes between archaic and modern humans with 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 modelling 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 and 7dupASD patient-specific NCSCs and across an exhaustive range of BAZ1B dosages. Notably, our cohort also included NCSCs from a rare WBS patient 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 (*36*, *52*, *56*, *64*), 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. Interestingly, 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. Finally, 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 NCSCs.

Third, our investigation provides the first experimental evidence for the neurocristopathic hypothesis that had been put forth to explain domestication and had pointed to BAZ1B as one of the candidates underlying the domestication syndrome (*3*). Indeed, among the key NC hubs impacted 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.

19

Finally, the empirical determination of BAZ1B dosage-sensitive genes, in NC models from AMHs with accentuated domestication-relevant traits, allowed us to functionally interrogate datasets exposing the genetic differences between modern *vs* archaic. This brought into relief the significant convergence between BAZ1B-dependent circuits and genes harboring regulatory changes in the

5   human lineage, reinforcing the notion that regulatory regions contain some of the most significant changes relevant for the modern lineage. In this context, 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. Indeed, among the genes downstream of BAZ1B that we uncovered in this study,

10   *FOXP2*, *ROBO1* and *ROBO2*, have long been implicated in brain wiring processes critical for vocal learning in several species (*65*, *66*), including humans, and will warrant further mechanistic dissection in light of the distinctive linguistic profile of WBS individuals.

Thus, our findings establish the heuristic power of neurodevelopmental disease modelling for the study of human evolution.

15

References

1. C. Theofanopoulou *et al.*, 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. 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).

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

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

6. B. M. vonHoldt *et al.*, Structural variants in genes associated with human Williams-Beuren syndrome underlie stereotypical hypersociability in domestic dogs. *Sci. Adv.* **3**, e1700398 (2017).

7. C. Barnett *et al.*, Williams Syndrome Transcription Factor is critical for neural crest cell function in Xenopus laevis. *Mech. Dev.* **129**, 324–338 (2012).

8. A. Ashe *et al.*, A genome-wide screen for modifiers of transgene variegation identifies genes with critical roles in development. *Genome Biol.* **9**, R182 (2008).

9. B. M. vonHoldt *et al.*, 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).

10. S. Peyrégne, M. J. Boyle, M. Dannemann, K. Prüfer, Detecting ancient positive selection in

humans using extended lineage sorting. *Genome Res.* (2017), doi:10.1101/gr.219493.116.

11. F. Racimo, Testing for Ancient Selection Using Cross-population Allele Frequency Differentiation. *Genetics*. **202**, 733–750 (2016).

12. M. Kuhlwilm, C. Boeckx, Genetic differences between humans and other hominins contribute to the "human condition." *BioRxiv* (2018), doi:10.1101/298950.

13. A. Adamo *et al.*, 7q11.23 dosage-dependent dysregulation in human pluripotent stem cells affects transcriptional programs in disease-relevant lineages. *Nat. Genet.* **47**, 132–141 (2015).

14. C. Fusco *et al.*, 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).

15. J. Meng *et al.*, 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).

16. R. Bajpai *et al.*, CHD7 cooperates with PBAF to control multipotent neural crest formation. *Nature*. **463**, 958–962 (2010).

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

18. R. A. Poot *et al.*, The Williams syndrome transcription factor interacts with PCNA to target chromatin remodelling by ISWI to replication foci. *Nat. Cell Biol.* **6**, 1236–1244 (2004).

19. 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).

20. 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).

21. 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).

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

23. S. Bhatt, R. Diaz, P. A. Trainor, Signals and switches in Mammalian neural crest cell differentiation. *Cold Spring Harb. Perspect. Biol.* **5** (2013), doi:10.1101/cshperspect.a008326.

24. S. O. Ko *et al.*, Smad4 is required to regulate the fate of cranial neural crest cells. *Dev. Biol.* **312**, 435–447 (2007).

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

26. 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).

27. K. Oka *et al.*, Roles of collagen and periostin expression by cranial neural crest cells during soft palate development. *J. Histochem. Cytochem.* **60**, 57–68 (2012).

28. H. Rios *et al.*, periostin null mice exhibit dwarfism, incisor enamel defects, and an early-onset periodontal disease-like phenotype. *Mol. Cell. Biol.* **25**, 11131–11144 (2005).

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

30. 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).

31. 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).

32. H. Hsu *et al.*, Tumor necrosis factor receptor family member RANK mediates osteoclast differentiation and activation induced by osteoprotegerin ligand. *Proc. Natl. Acad. Sci. USA*. **96**, 3540–3545 (1999).

33. A. Werner *et al.*, Cell-fate determination by ubiquitin-dependent regulation of translation. *Nature*. **525**, 523–527 (2015).

34. C. A. McGourty *et al.*, Regulation of the CUL3 Ubiquitin Ligase by a Calcium-Dependent Co-adaptor. *Cell*. **167**, 525–538.e14 (2016).

35. J. Dixon *et al.*, Tcof1/Treacle is required for neural crest cell formation and proliferation deficiencies that cause craniofacial abnormalities. *Proc. Natl. Acad. Sci. USA*. **103**, 13403–13408 (2006).

36. P. Claes *et al.*, Genome-wide mapping of global-to-local genetic effects on human facial shape. *Nat. Genet.* **50**, 414–423 (2018).

37. G. Andreoletti *et al.*, AMMECR1: a single point mutation causes developmental delay, midface hypoplasia and elliptocytosis. *J. Med. Genet.* **54**, 269–277 (2017).

38. B. Tumiene *et al.*, Phenotype comparison confirms ZMYND11 as a critical gene for 10p15.3 microdeletion syndrome. *J. Appl. Genet.* **58**, 467–474 (2017).

39. J. M. Cobben *et al.*, 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).

40. 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).

41. W.-S. Wu *et al.*, Slug antagonizes p53-mediated apoptosis of hematopoietic progenitors by repressing puma. *Cell*. **123**, 641–653 (2005).

42. M. Gabriele *et al.*, YY1 haploinsufficiency causes an intellectual disability syndrome featuring transcriptional and chromatin dysfunction. *Am. J. Hum. Genet.* **100**, 907–925 (2017).

43. 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).

44. 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).

45. J. Dalcq *et al.*, RUNX3, EGR1 and SOX9B form a regulatory cascade required to modulate BMP-signaling during cranial cartilage development in zebrafish. *PLoS One*. **7**, e50140 (2012).

46. I. Schanze *et al.*, NFIB Haploinsufficiency Is Associated with Intellectual Disability and Macrocephaly. *Am. J. Hum. Genet.* **103**, 752–768 (2018).

47. L. P. Sanford *et al.*, TGFbeta2 knockout mice have multiple developmental defects that are non-overlapping with other TGFbeta knockout phenotypes. *Development*. **124**, 2659–2670 (1997).

48. C. Boileau *et al.*, TGFB2 mutations cause familial thoracic aortic aneurysms and dissections associated with mild systemic features of Marfan syndrome. *Nat. Genet.* **44**, 916–921 (2012).

49. O. L. Wapinski *et al.*, Hierarchical mechanisms for direct reprogramming of fibroblasts to neurons. *Cell.* **155**, 621–635 (2013).

50. S. J. Hong *et al.*, Transcription factor GATA-3 regulates the transcriptional activity of dopamine beta-hydroxylase by interacting with Sp1 and AP4. *Neurochem. Res.* **33**, 1821–1831 (2008).

51. 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. USA.* **108**, 155–160 (2011).

52. 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).

53. D. Huang *et al.*, GLI2 promotes cell proliferation and migration through transcriptional activation of ARHGEF16 in human glioma cells. *J Exp Clin Cancer Res.* **37**, 247 (2018).

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

55. G. Chen *et al.*, Semaphorin-3A guides radial migration of cortical neurons during development. *Nat. Neurosci.* **11**, 36–44 (2008).

56. A. Rada-Iglesias *et al.*, Epigenomic annotation of enhancers predicts transcriptional regulators of human neural crest. *Cell Stem Cell.* **11**, 633–648 (2012).

57. 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).

58.   Q. Wu *et al.*, DISC1 Regulates the Proliferation and Migration of Mouse Neural Stem/Progenitor Cells through Pax5, Sox2, Dll1 and Neurog2. *Front. Cell Neurosci.* **11**, 261 (2017).

59.   M. C. Horowitz *et al.*, Pax5-deficient mice exhibit early onset osteopenia with increased osteoclast progenitors. *J. Immunol.* **173**, 6583–6591 (2004).

60.   D. Gokhman *et al.*, Recent regulatory changes shaped human facial and vocal anatomy. *BioRxiv* (2017), doi:10.1101/106955.

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

62.   C. Darwin, *The Descent of Man: and selection in relation to sex* (J. Murray, London, 1871).

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

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

65.   S. C. Vernes *et al.*, Foxp2 regulates gene networks implicated in neurite outgrowth in the developing brain. *PLoS Genet.* **7**, e1002145 (2011).

66.   H. Lei *et al.*, Axon guidance pathways served as common targets for human speech/language evolution and related disorders. *Brain Lang.* **174**, 1–8 (2017).

67.   L. Menendez *et al.*, Directed differentiation of human pluripotent cells to neural crest stem cells. *Nat. Protoc.* **8**, 203–212 (2013).

68.   R. Blecher-Gonen *et al.*, High-throughput chromatin immunoprecipitation for genome-wide mapping of in vivo protein-DNA interactions and epigenomic states. *Nat. Protoc.* **8**, 539–554 (2013).

69. P.-L. Germain *et al.*, RNAontheBENCH: computational and empirical resources for benchmarking RNAseq quantification and differential expression methods. *Nucleic Acids Res.* **44**, 5054–5067 (2016).

70. I. Streeter *et al.*, The human-induced pluripotent stem cell initiative-data resources for cellular genetics. *Nucleic Acids Res.* **45**, D691–D697 (2017).

71. H. Kilpinen *et al.*, Common genetic variation drives molecular heterogeneity in human iPSCs. *Nature*. **546**, 370–375 (2017).

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

73. M. Meyer *et al.*, A high-coverage genome sequence from an archaic Denisovan individual. *Science*. **338**, 222–226 (2012).

74. K. Prüfer *et al.*, The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature*. **505**, 43–49 (2014).

75. K. Prüfer *et al.*, A high-coverage Neandertal genome from Vindija Cave in Croatia. *Science*. **358**, 655–658 (2017).

76. L. Garbes *et al.*, 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).

77. A. J. Griffith *et al.*, Marshall syndrome associated with a splicing defect at the COL11A1 locus. *Am. J. Hum. Genet.* **62**, 816–823 (1998).

78. E. Roessler *et al.*, Truncating loss-of-function mutations of DISP1 contribute to holoprosencephaly-like microform features in humans. *Hum. Genet.* **125**, 393–400 (2009).

79. J. Punetha *et al.*, Novel Col12A1 variant expands the clinical picture of congenital myopathies with extracellular matrix defects. *Muscle Nerve*. **55**, 277–281 (2017).

80.   T. O'Rourke, C. Boeckx, Converging roles of glutamate receptors in domestication and prosociality. *BioRxiv* (2018), doi:10.1101/439869.

81.   S. Srivastava *et al.*, Loss-of-function variants in HIVEP2 are a cause of intellectual disability. *Eur. J. Hum. Genet.* **24**, 556–561 (2016).

82.   E. Lausch *et al.*, TBX15 mutations cause craniofacial dysmorphism, hypoplasia of scapula and pelvis, and short stature in Cousin syndrome. *Am. J. Hum. Genet.* **83**, 649–655 (2008).

83.   M. K. Lee *et al.*, Genome-wide association study of facial morphology reveals novel associations with FREM1 and PARK2. *PLoS One*. **12**, e0176566 (2017).

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

85.   Y. Zhang *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).

Acknowledgements

Author contributions: M.Z. carried out all the experiments including the generation and characterization of both BAZ1B-interfered lines and BAZ1B-tagged lines, the preparation of RNA-seq libraries, ChIPs and wound-healing assays; A.V. performed RNA-seq and ChIP-seq bioinformatic analyses with the supervision of P.-L.G.; A.A., P.T.M., S.S., T.O. and C.B. performed analyses in Fig. 4; M.L. and A.R.-I. characterized NC induction phenotype reported in Fig. 1D,E; N.M. reprogrammed 7dupASD3 and CTL4R fibroblasts and generated the corresponding NC lines; A.S. helped with cell culture; M.Z. and S.T. designed the tag strategy; G.M. provided 7dupASD3 and CTL4R fibroblasts; M.Z., C.B. and G.T. wrote the manuscript with contributions from A.V. and A.S.; G.T. conceived, designed and supervised the study.

Competing interests: None. **Data and materials availability**: All data are available in the manuscript or the supplementary materials.

Supplementary Materials:

Materials and Methods

Figures S1-S4
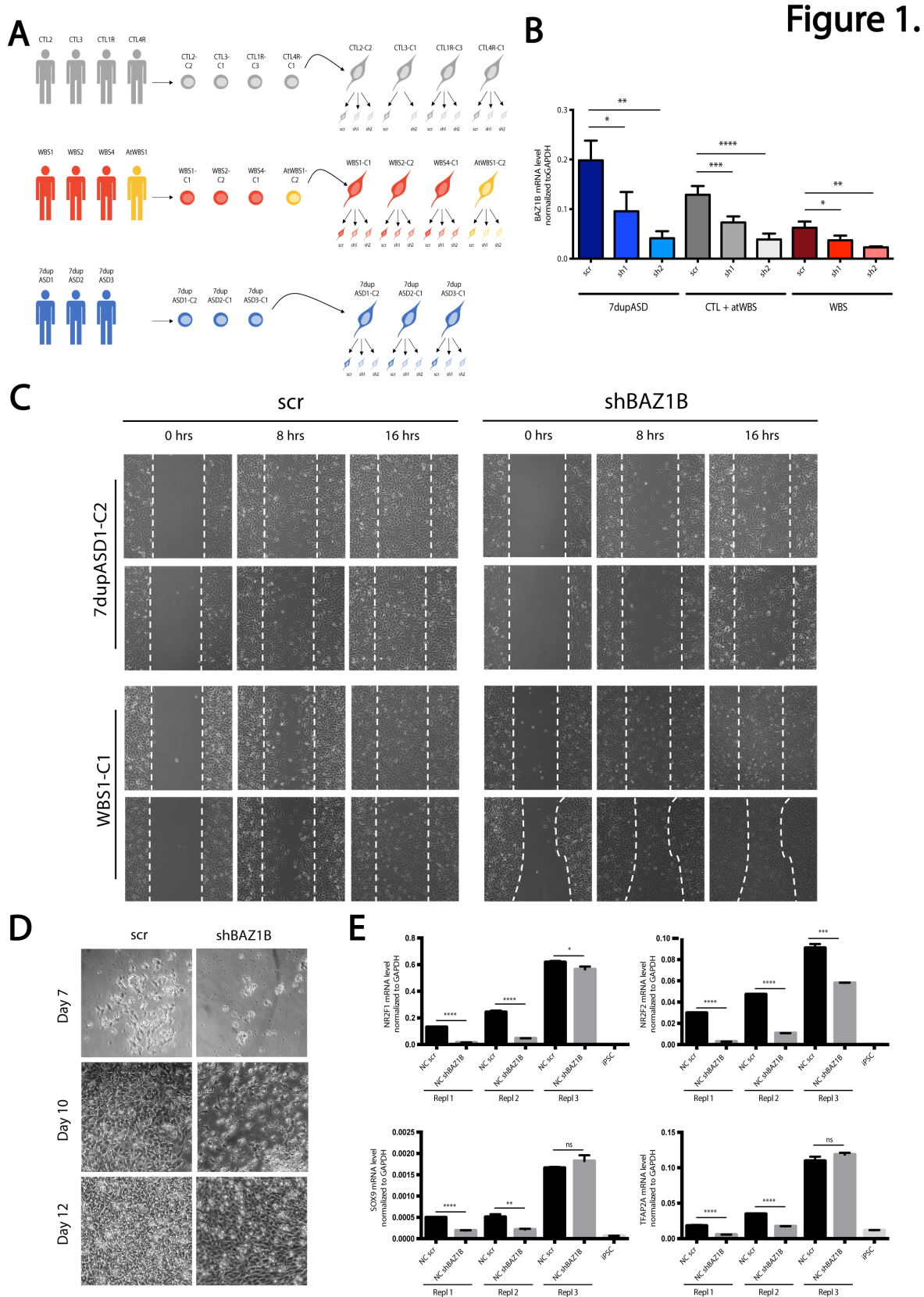
Tables S1-S17

References (67-85)

Figure 1.

Figure 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). GAPDH is used as normalizer. (**C**) 8 hours and 16 hours-time points from the wound-healing assay analysis 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**) Day 7, 10 and 12 of NC differentiation from EBs of a scr-interfered iPSC line and its respective BAZ1B KD (n=3). (**E**) mRNA levels of NC markers at day 12 of differentiation (experiment reported in panel D). An iPSC line is included as a negative control. Student's t- test (ns: non-significant; *: p-value < 0.05; **: p-value < 0.01; ***: p-value < 0.001 and ****: p-value < 0. 0001).
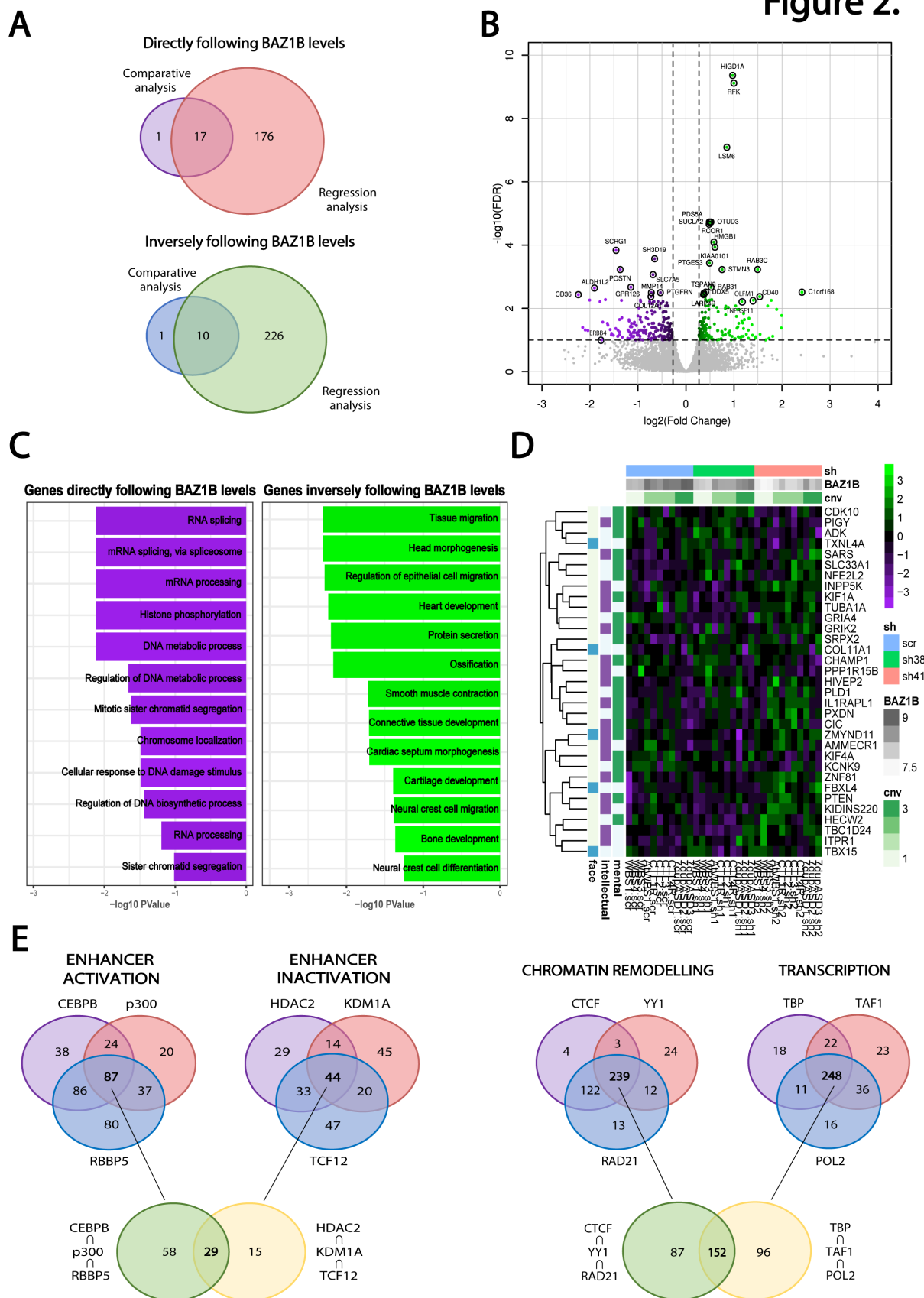
Figure 2.

Figure 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 *vs* 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 (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**) Heatmap representing DEGs that are dysregulated in genetic disorders involving mental retardation, intellectual disability and/or facial dysmorphisms according to OMIM database classification. (**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.
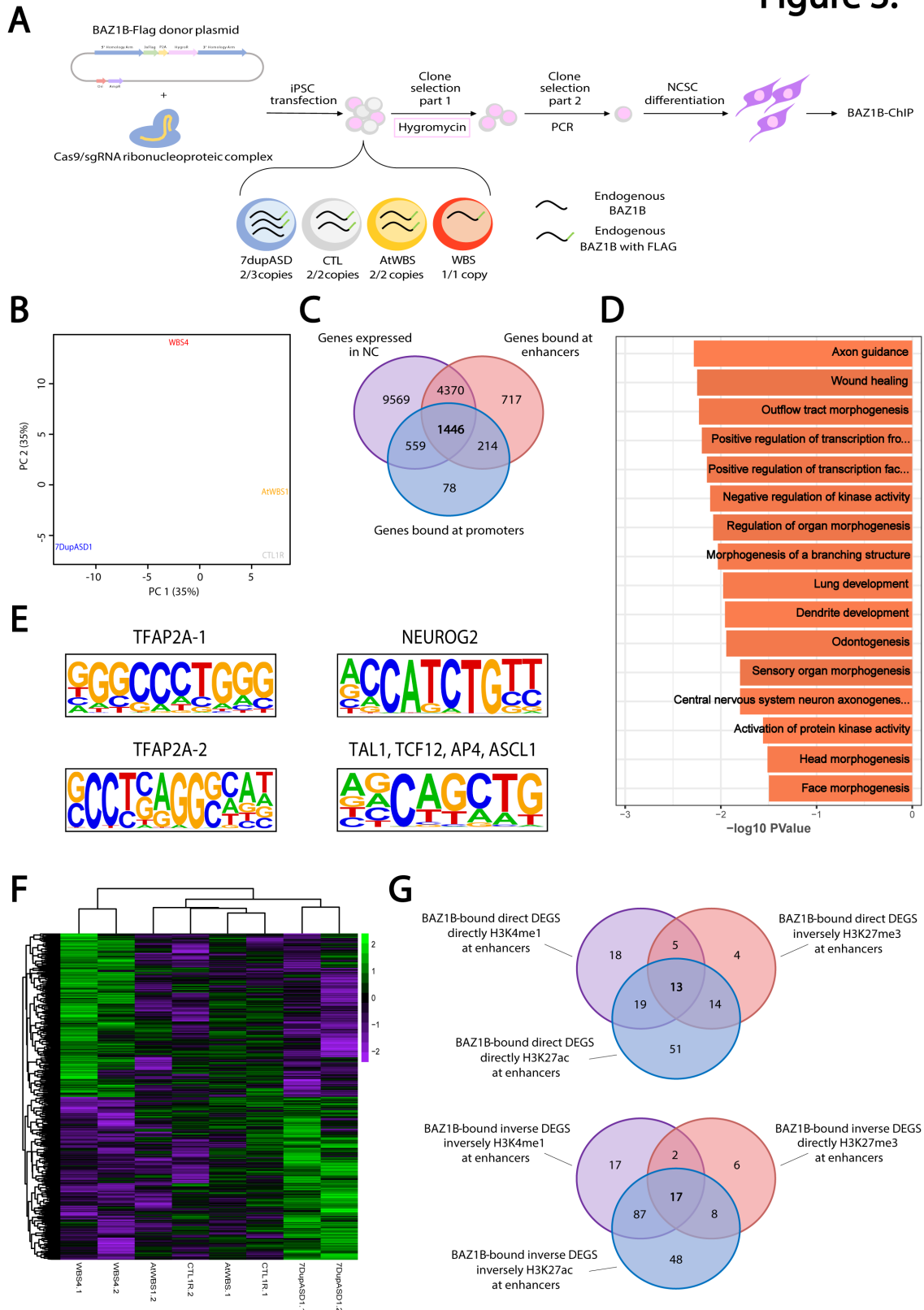
35

Figure 3.

Figure 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 (sgRNA) ribonucleoprotein complex, clones were selected via Hygromycin and PCR, differentiated to NCSCs and then subjected to ChIP-seq. (**B**) Principal component analysis 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 NC stem cell 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.
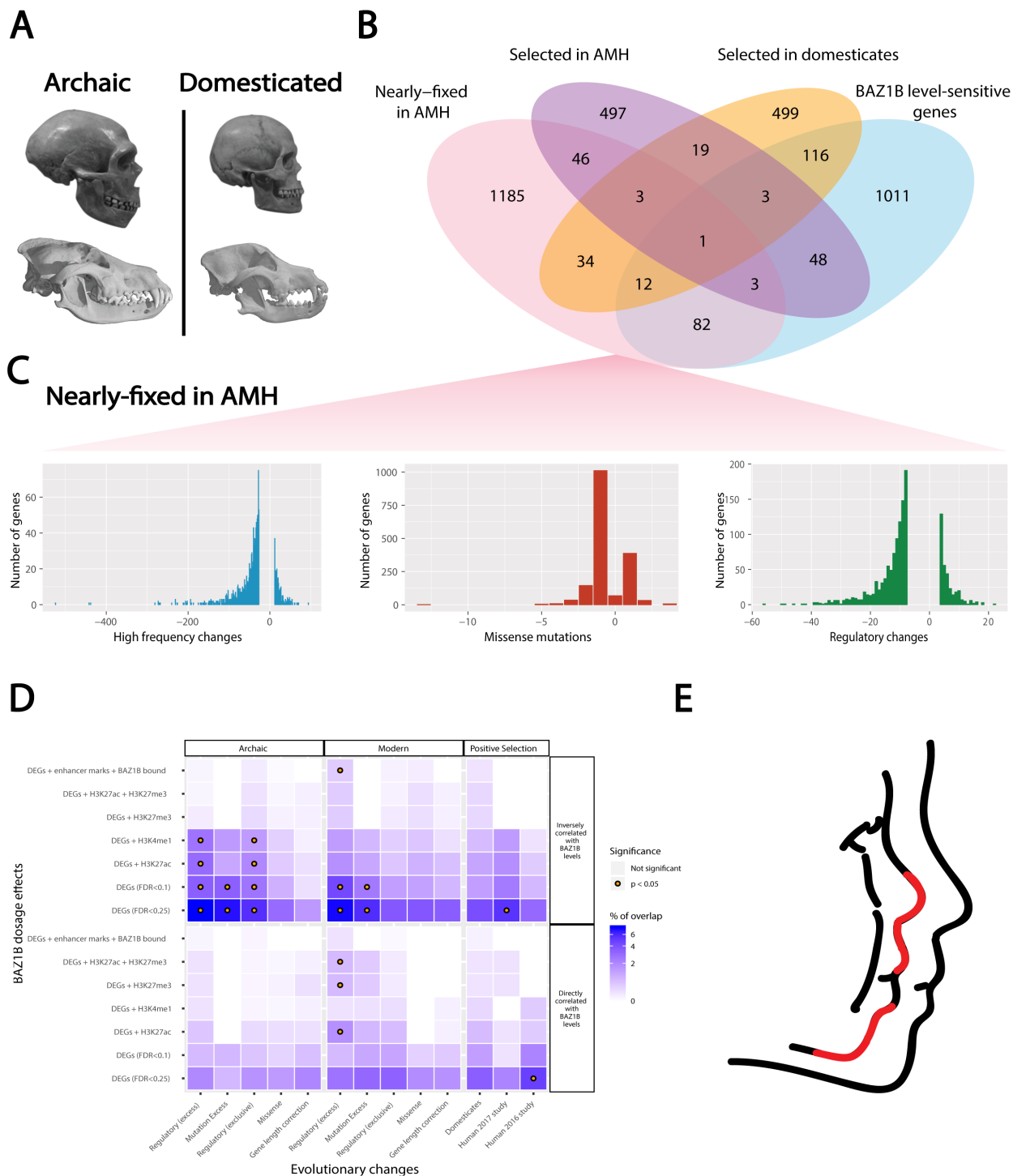
Figure 4.

Figure 4. Exploration of paleogenomic datasets supports a key evolutionary role for BAZ1B and validates the self-domestication hypothesis. (**A**) Skulls from archaic human (Neanderthal) and wolf compared to modern human and dog, capturing the parallelism between domestication and self-domestication. (**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 panel B). (**D**) Heatmap representing the amount of overlaps for each list selected from Panel B. Gene overlaps and detailed list descriptions are reported in Tables S1-S14. (**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, lower front of the mandible). The lower mid-face region is the most often associated with mutations in genes figuring prominently in our intersections, as discussed in the text and Table S15.