Science Advances
AAAS

advances.sciencemag.org/cgi/content/full/5/12/eaaw7908/DC1

Supplementary Materials for

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, Alessandro Vitriolo, Alejandro Andirko, Pedro Tiago Martins, Stefanie Sturm, Thomas O'Rourke, Magdalena Laugsch, Natascia Malerba, Adrianos Skaros, Sebastiano Trattaro, Pierre-Luc Germain, Marija Mihailovic, Giuseppe Merla, Alvaro Rada-Iglesias, Cedric Boeckx, Giuseppe Testa*

*Corresponding author. Email: giuseppe.testa@unimi.it, giuseppe.testa@ieo.it, giuseppe.testa@htechnopole.it

Published 4 December 2019, *Sci. Adv.* **5**, eaaw7908 (2019)
DOI: 10.1126/sciadv.aaw7908

The PDF file includes:

Other Supplementary Material for this manuscript includes the following:

(available at advances.sciencemag.org/cgi/content/full/5/12/eaaw7908/DC1)

Table S2B (Microsoft Excel format). Regulatory excess in archaic humans, overlap with BAZ1B targets.
Table S2C (Microsoft Excel format). Mutation excess in archaic humans, overlap with BAZ1B targets.
Table S2D (Microsoft Excel format). Regulatory changes (exclusive) in archaic humans, overlap with BAZ1B targets.
Table S2E (Microsoft Excel format). Missense mutations in archaic humans, overlap with BAZ1B targets.
Table S2F (Microsoft Excel format). Mutation excess in archaic humans corrected for length, overlap with BAZ1B targets.
Table S2G (Microsoft Excel format). Regulatory excess in modern humans, overlap with BAZ1B targets.
Table S2H (Microsoft Excel format). Mutation excess in modern humans, overlap with BAZ1B targets.
Table S2I (Microsoft Excel format). Regulatory changes (exclusive) in modern humans, overlap with BAZ1B targets.
Table S2J (Microsoft Excel format). Missense mutations in modern humans, overlap with BAZ1B targets.
Table S2K (Microsoft Excel format). Mutation excess in modern humans corrected for length, overlap with BAZ1B targets.
Table S2L (Microsoft Excel format). Genes under positive selection in domesticated animals, overlap with BAZ1B targets.
Table S2M (Microsoft Excel format). Genes under positive selection from Peyrégne *et al.* (*13*) in modern humans, overlap with BAZ1B targets.
Table S2N (Microsoft Excel format). Genes under positive selection from Racimo (*14*) in modern humans, overlap with BAZ1B targets.

Supplementary Figure 1.

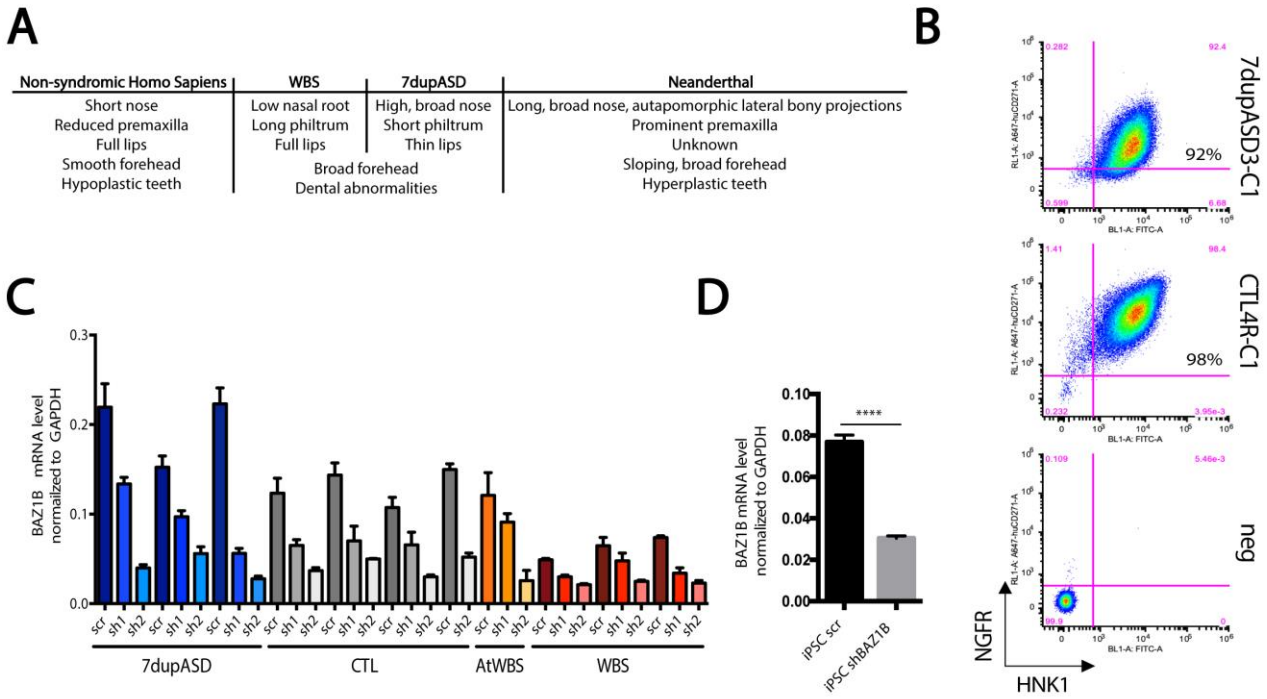
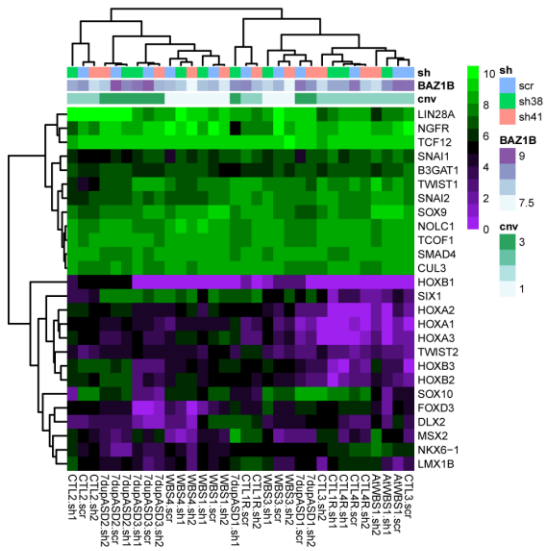


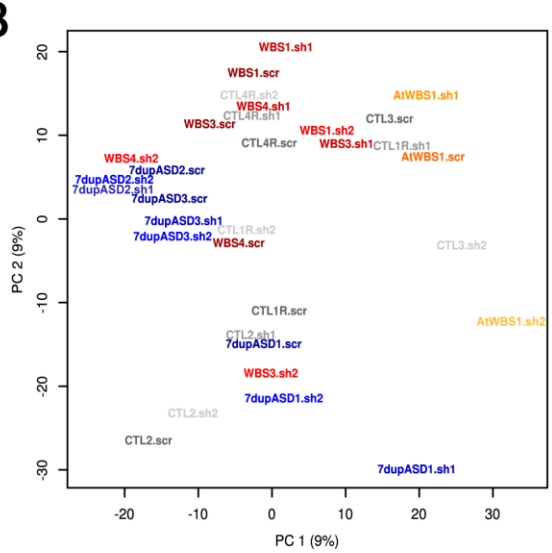
Fig. S1. BAZ1B KD validation in iPSC-derived NCSCs and evaluation of its impact on NCSC migration. (**A**) Symmetrically opposite and shared craniofacial phenotypes in WBS and 7dupASD patients compared to similar features in Non-syndromic Homo Sapiens and Neanderthal. (**B**) Flow cytometry analysis of $HNK1^{+}/NGFR^{+}$ NCSCs (7dupASD3-C1 and CTL4R-C1). Unstained cells are used as negative control**.** (**C**) BAZ1B mRNA levels in all the interfered lines (scr, sh1 and sh2) as measured by qPCR. Data represent individual samples with the same number of BAZ1B copies. GAPDH is used as normalizer. (**D**) qPCR validation of BAZ1B KD in the iPSC line used in the experiment reported in Fig. 1D, E. (**E**) Western blots showing BAZ1B levels upon KD in all NCSC lines. The arrow indicates the BAZ1B specific band. GAPDH is used as the normalizer. (**F**) 4 hours, 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).

Supplementary Figure 2.

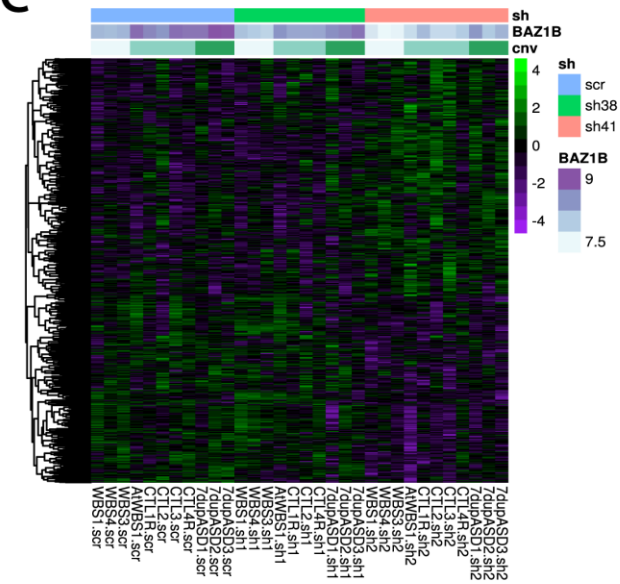
A



B



C



D

GENE	% of BAZ1B DEGs	GENE	% of BAZ1B DEGs
CEBPB	52,46	TBP	66,74
RAD21	86,16	TAF1	73,44
CTBP2	31,25	MAX	41,07
USF1	66,96	SP1	52,00
TEAD4	54,69	EGR1	37,50
RBBP5	64,73	KDM1A	27,46
MAFK	37,50	SIN3A	70,09
SRF	28,79	MXI1	31,03
CTCF	82,14	USF2	27,68
TCF12	32,10	CHD2	31,47
YY1	62,05	POLR2A	69,42
HDAC2	26,79	SP4	25,22
EP300	37,50	TAF7	40,63
GTF2I	31,25		

E

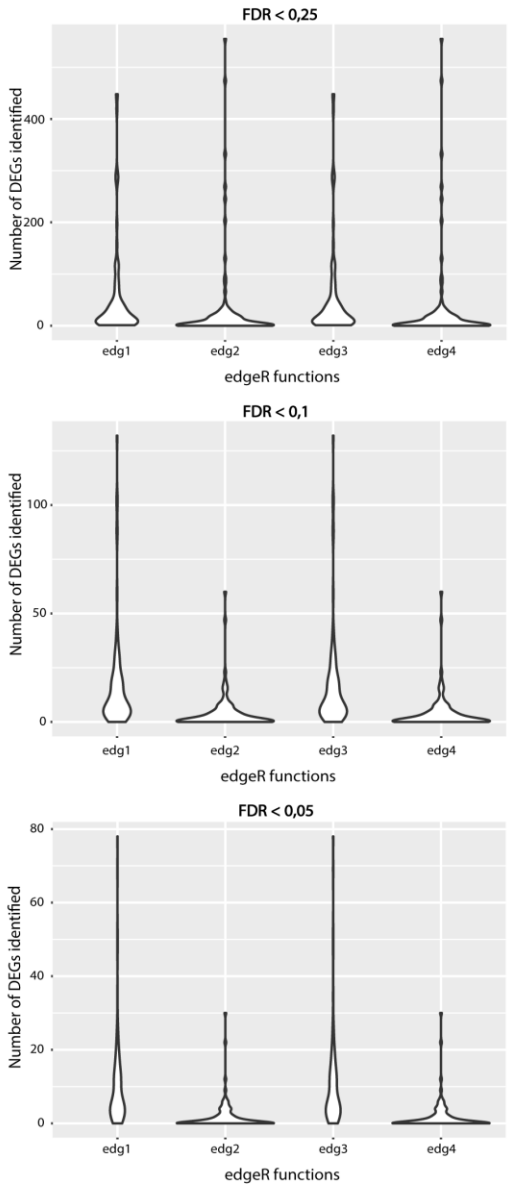
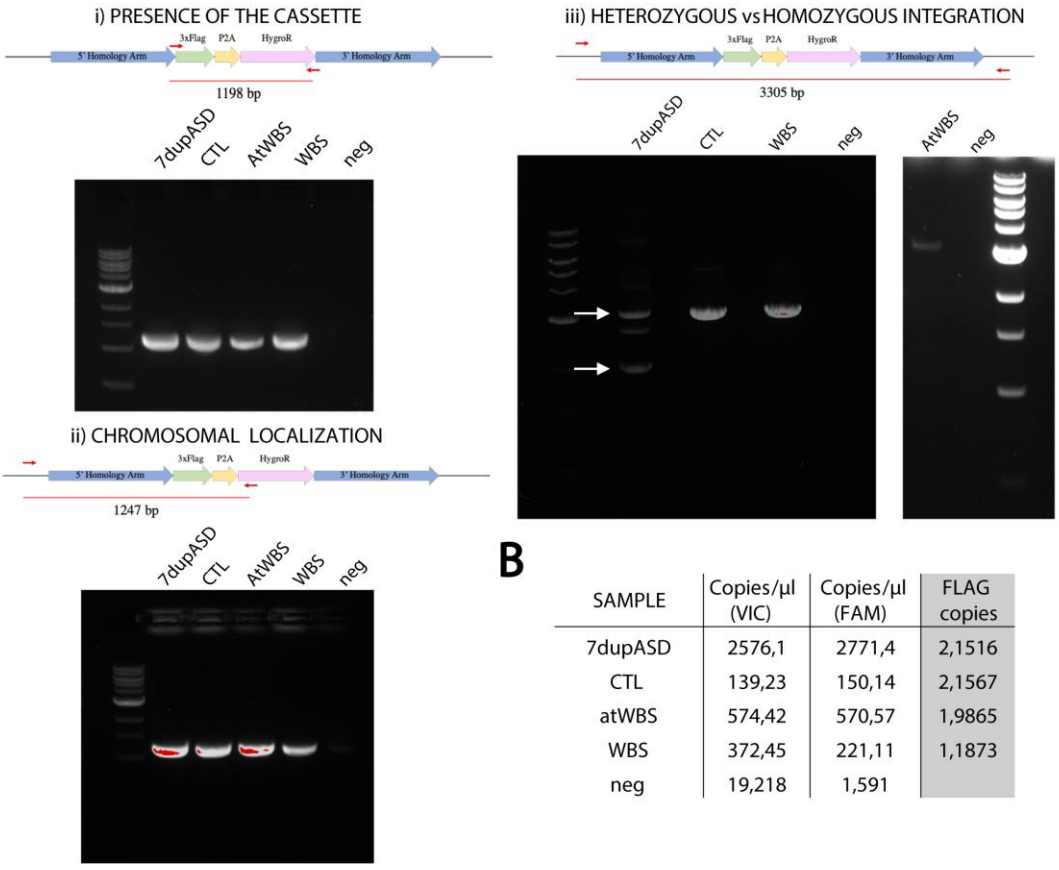


Fig. S2. BAZ1B KD affects the transcriptome of iPSC-derived NCSCs. (**A**) Expression levels (log-TMM) of cranial- and NC-specific genes constituting the signature used to assess the quality of our NCSC lines. (**B**) Principal component analysis showing the distribution of the 32 NCSC lines according to their transcriptional profiles. (**C**) Expression profile of the 448 genes that follows BAZ1B levels. The samples are ordered according to shRNAs. In each shRNA group samples are further ordered based on *BAZ1B* copies (CNVs) (FDR < 0.1). (**D**) List of genes identified as regulators of BAZ1B levels-sensitive genes in the Master regulator analysis. For each gene the percentage of BAZ1B-regulated DEGs is reported. (**E**) Violin Plot representation of the number of spurious DEGs generated by the four implemented edgeR pipelines on HipSci RNA-seq data, given the model matrix applied for the regression on BAZ1B levels in NCSCs.

Supplementary Figure 3.

A



B

SAMPLE	Copies/µl (VIC)	Copies/µl (FAM)	FLAG copies
7dupASD	2576,1	2771,4	2,1516
CTL	139,23	150,14	2,1567
atWBS	574,42	570,57	1,9865
WBS	372,45	221,11	1,1873
neg	19,218	1,591	

C

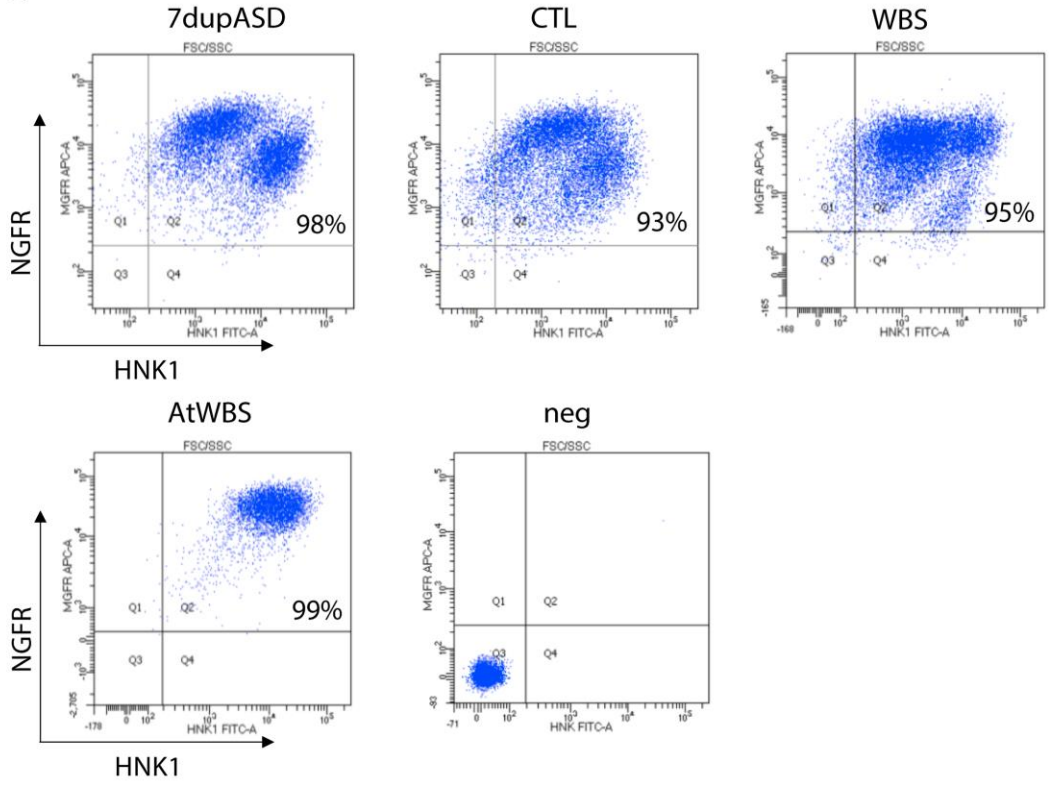
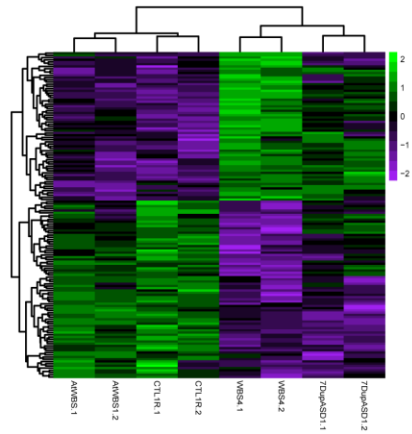


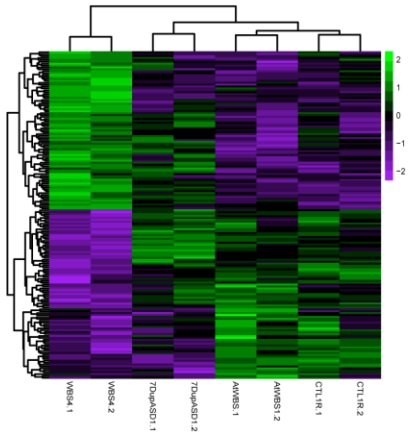
Fig. S3. Generation of BAZ1B-FLAG iPSC lines and differentiation to NCSCs. (**A**) PCR reporting FLAG-tagged selected clones. Clones were screened for i) the presence of the cassette (band at 1198 bp) and ii) its proper chromosomal localization (band at 1247 bp) and iii) to distinguish a heterozygous (two bands: one at 2138 bp and one at 3305 bp) from a homozygous integration (single band at 3305 bp). (**B**) Digital PCR analysis to evaluate the exact number of FLAG integrated copies. The FLAG probe was labeled with FAM, while a TERT probe labeled with VIC was used as an internal reference (two known copies). (**C**) FACS analysis of $HNK1^{+}/NGFR^{+}$ iPSC-derived NCSC tagged clones. Unstained cells were used as negative control.

Supplementary Figure 4.

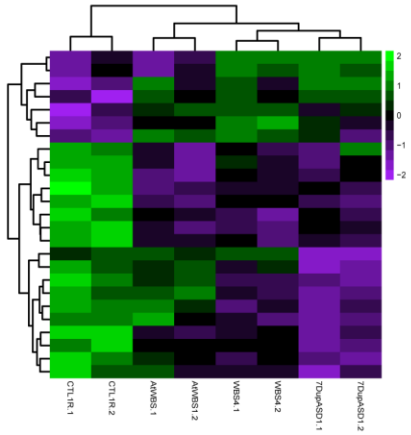
A



B



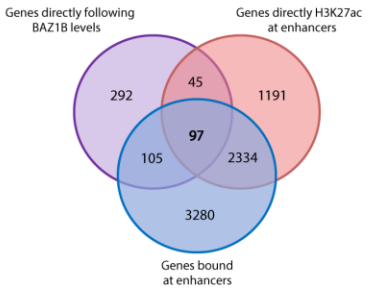
C

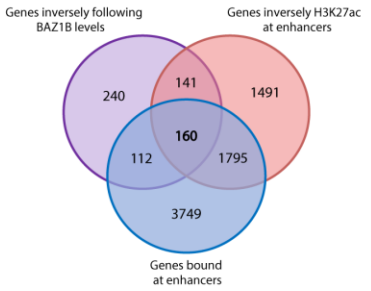


D

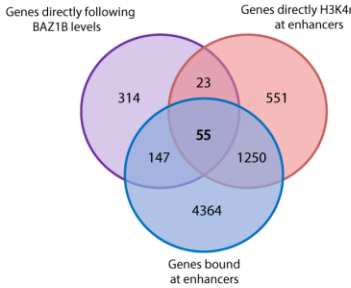
HISTONE MARK	DIRECT	INVERSE
H3K27ac	3667	3587
H3K4me1	1879	2169
H3K27me3	1039	1097

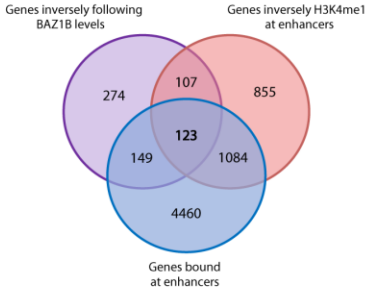
E



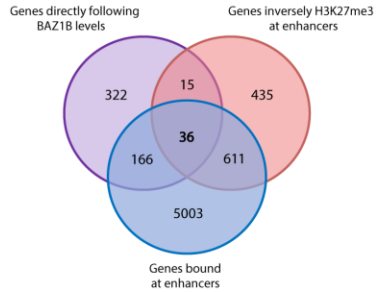


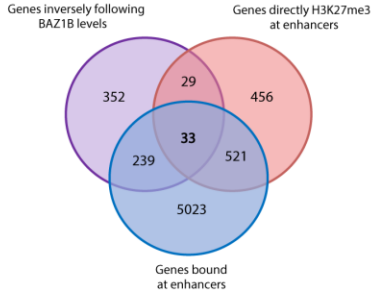
F





G





H

DIRECT	INVERSE
ABLIM2	ADORA2B
ARHGEF16	CHST15
B3GAT1	COL12A1
DOK7	DOCK9
IL4R	GMPR
KHDRBS3	KLF7
MGAT5B	MAFB
NRXN2	MRC2
OLFM1	NR2F2
PLXNA4	PHACTR1
PRKCQ	PXDC1
SH3TC1	RAB11FIP1
SLC6A6	RASL11B
	SEMA6A
	SLC1A4
	TGFBI
	TIAM2

Fig. S4. BAZ1B KD induces a significant chromatin remodeling at distal regions. (**A**) Regions differentially bound by BAZ1B in a WBS and 7dupASD *vs* CTL and atWBS comparison (FDR < 0.1, n=2). (**B**) Regions differentially bound by BAZ1B in a WBS *vs* 7dup, CTL and atWBS comparison (FDR < 0.1, n=2). (**C**) Regions differentially bound by BAZ1B in a 7dupASD *vs* CTL, WBS and atWBS comparison (FDR < 0.1, n=2). (**D**) Total number of genes that have their enhancers differentially marked following BAZ1B levels. (**E**) Overlap between genes whose expression follows BAZ1B levels (purple), genes whose enhancer-associated H3K27 acetylation follows BAZ1B levels (red) and genes bound by BAZ1B at their enhancers (blue). (**F**) Overlap between genes whose expression follows BAZ1B levels (purple), genes whose enhancer-associated H3K4 mono-methylation follows BAZ1B levels (red) and genes bound by BAZ1B at their enhancers (blue). (**G**) Overlap between genes whose expression follows BAZ1B levels (purple), genes whose enhancer-associated H3K27 trimethylation follows BAZ1B levels (red) and genes bound by BAZ1B at their enhancers (blue). (**H**) List of genes that, at the same time, follow BAZ1B levels, have their enhancers differentially marked concordantly (H3K27ac, H3K4me1 and H3K27me3) and are bound by BAZ1B at enhancers (overlap in Fig. 3G).

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

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).

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

Through the intersection of genes following BAZ1B regions, and genes whose regulatory regions were bound by BAZ1B and differentially marked, we could 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 *PLXNA (61-64)*, as well as cardinal NC markers, such as *HNK1* and *NR2F2 (48,65)*. 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 (*66*) and of the developmental balance between osteoblast and osteoclast precursors (*67*).

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

Gene	Function	Reference	Dysregulation
OLFM1	overexpression promotes an increased and prolonged production of NC cells	*(61)*	**Directly correlated to BAZ1B levels**
TNFRSF11B	bone resorption and osteoclast activation	*(68)*	
CUL3	craniofacial morphogenesis: it ubiquitinates TCOF1 and NOLC1 as a master regulator of NC specification	*(34)*	
KLHL12	CUL3 adaptor protein, craniofacial morphogenesis	*(69)*	
POSTN	cranial NC-mediated soft palate development	*(70, 71)*	**Inverseley correlated to BAZ1B levels**
ERBB4	Skeletal muscle development and NC migration; causes heart defect and aberrant cranial NC migration in deficient mice	*(72, 73)*	
TBX15	one of recently mapped loci associated to the variability of human facial shape is located inside its gene body	*(47)*	
AMMECR1	Its mutations cause midface hypoplasia and hearing impairment	*(74)*	
ZMYND11	H3K36me3 reader that gives rise to dysmorphic facial features when mutated; causes mild intellectual disability when mutated	*(75, 76)*	
SNAI2	critical regulator of NC-based organogenesis	*(77)*	
BBC3	repressed by SNAI2 in response to DNA damage	*(78)*	

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

Gene	Phenotype	Reference
AMMECR1, SEC24D, COL11A1	Midface Hypoplasia	*(74, 79, 80)*
DISP1	Holoprosencephaly	*(81)*
COL12A1	Micrognathia	*(82)*
GRIK2, GRID2, DLGAP1	Glutamate signaling, tameness and domestication	*(83)*
PTEN, FMR1, NRXN2, CBX4	Autism candidates	SFARI Gene Autism Database
HIVEP2	Intellectual Disability Phenotypes	*(84)*
PLXNA4, GAP43, DUSP5, ROBO2, ROBO1, SRGAP1, FOXP2	Axon Guidance, Brain Wiring Processes	*(50, 85)*
TBX15, FREM1	Facial Development	*(47, 86, 87)*
TGFB2, TGFBR2, POSTN	Palate Development	*(70, 88)*
DDR2	Osteoblast differentiation, Chondrocyte maturation	*(89)*
OLFM1	Continued production of NC cells when overexpressed	*(61)*
ERBB4	Positive selection in modern humans and domesticated species, skeletal muscle development and NC	*(1, 72)*

	migration	
EDN3, MAGOH, ZEB2	Associated with behavioral changes found in domesticates	*(5)*

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

Edg1	**Edg2**	Edg3	Edg4
edg1 <- function(e,mm){	**edg2 <- function(e,mm){**	**edg3<- function(e,mm){**	**edg4 <- function(e,mm){**
(mm) <- colnames(e)	(mm) <- colnames(e)	(mm) <- colnames(e)	(mm) <- colnames(e)
dds <- calcNormFactors(DGEList(e))	dds <- calcNormFactors(DGEList(e))	dds <- calcNormFactors(DGEList(e))	dds <- calcNormFactors(DGEList(e))
dds <- estimateDisp(dds,mm,robust=T)	dds <- estimateGLMRobustDisp(dds,mm)	dds <- estimateDisp(dds,mm,robust=T)	dds <- estimateGLMRobustDisp(dds,mm)
dds <- glmFit(dds, mm)	dds <- glmFit(dds, mm)	dds <- glmQLFit(dds,mm,robust=T)	dds <- glmQLFit(dds,mm,robust=T)
res <- as.data.frame(topTags(glmLRT(dds, "baz1b"), nrow(e)))	res <- as.data.frame(topTags(glmLRT(dds, "baz1b"), nrow(e)))	res <- as.data.frame(topTags(glmLRT(dds, "baz1b"), nrow(e)))	res <- as.data.frame(topTags(glmLRT(dds, "baz1b"), nrow(e)))
res	res	res	res
}	}	}	}

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).

	BAZ1B DEGs	**Mean (randomized HipSci data)**	**Median (randomized HipSci data)**
FDR < 0.25	830	43	93.32
FDR < 0.1	306	19	27.92
FDR < 0.05	178	13.5	17.38