Zeb2 DNA-binding sites in ES cell derived neuroprogenitor cells reveal autoregulation and align with neurodevelopmental knockout mouse and disease phenotypes.

Judith C. Birkhoff[1], Anne L. Korporaal[1], Rutger W.W. Brouwer[2], Claudia Milazzo[1#], Lidia Mouratidou[1], Mirjam C.G.N. van den Hout[2], Wilfred F.J. van IJcken[1,2], Danny Huylebroeck[1,3*, $], Andrea Conidi[1*, $]

[1] Dept. Cell Biology, Erasmus University Medical Center, Rotterdam, 3015 CN, The Netherlands; [2] Center for Biomics-Genomics, Erasmus University Medical Center, Rotterdam 3015 CN, The Netherlands; [3] Dept. Development and Regeneration, KU Leuven, Leuven 3000, Belgium

* *shared senior authors*

[#] *Present address*: Dept. Clinical Genetics, Erasmus University Medical Center, Rotterdam, 3015 CN, The Netherlands

[$] *Corresponding author*:

Andrea Conidi, PhD, Dept. Cell Biology, Erasmus MC, Building Ee - room Ee-1040, Wytemaweg 80, 3015 CN Rotterdam, The Netherlands; Tel: +31-10-7043169; Fax: +31-10-7044743; E-mail: a.conidi@erasmusmc.nl

Author contributions

JB: Conception and design, collection and/or assembly of data, data analysis and interpretation, manuscript writing

AK: Collection and/or assembly of data

RB: Data analysis and interpretation

CM: Collection and/or assembly of data

LM: Collection and/or assembly of data

MvdH: Collection and/or assembly of data, data analysis and interpretation

WvIJ: Data analysis and interpretation, manuscript writing

DH: Conception and design, financial support, manuscript writing, final approval of manuscript

AC: Conception and design, collection and/or assembly of data, data analysis and interpretation, manuscript writing, final approval of manuscript

Key words

Embryonic stem cells; genome-wide binding sites; Mowat-Wilson syndrome; neuroprogenitor cells; Zeb2

Abstract

Perturbation and mechanistic studies have shown that the DNA-binding transcription factor Zeb2 controls cell fate decision, differentiation and/or maturation in multiple cell lineages in embryos and after birth. In cultured embryonic stem cells (ESCs) *Zeb2*'s strong upregulation is necessary for their exit from primed pluripotency and entering neural and general differentiation. We engineered mouse ESCs to produce epitope-tagged Zeb2 from one of its two endogenous alleles. Using crosslinking ChIP-sequencing, we mapped for the first time 2,432 DNA-binding sites of Zeb2 in ESC-derived neuroprogenitor cells (NPCs). A new, major site maps promoter-proximal to *Zeb2* itself, and its homozygous removal demonstrates that *Zeb2* autoregulation is necessary to elicit proper Zeb2 effects in ESC→NPC differentiation. We then cross-referenced all Zeb2 DNA-binding sites with transcriptome data from Zeb2 perturbations in ESCs, ventral forebrain in mouse embryos, and adult neurogenesis. While the characteristics of these neurodevelopmental systems differ, we find interesting overlaps. Collectively, these new results obtained in ESC-derived NPCs significantly add to Zeb2's role as neurodevelopmental regulator as well as the causal gene in Mowat-Wilson Syndrome. Also, Zeb2 was found to map to loci mutated in other human congenital syndromes, making variant or disturbed levels of ZEB2 a candidate modifier principle in these.

Significance statement

Zeb2 is needed for exit from primed pluripotency and differentiation of ESCs and being studied in various cell lineages and tissues/organs in knockouts and adult mice. Phenotyping has mainly documented transcriptomes of Zeb2-mutant affected cell types, but ChIP-seq data are still not available to document Zeb2's direct target loci as repressor or activator of transcription. This study maps for the first time the Zeb2 genome-wide binding sites in ESC-derived neuroprogenitors, and discovered *Zeb2* autoregulation. Cross-referencing these ChIP-seq data with transcriptome profiles of

Zeb2 perturbation further explains Mowat-Wilson Syndrome defects and proposes *ZEB2* as candidate modifier gene in other syndromes.

Introduction

Zeb2 (also named Sip1/Zfhx1b) and Zeb1 (δEF1/Zfhx1a), the two members of the small family of Zeb transcription factors (TFs) in vertebrates, bind to two separated E2-boxes CACCT(G), sometimes CACANNT(G), on DNA via two highly conserved, separated clusters of zinc fingers [1-4]. Mutations in *ZEB2* cause Mowat-Wilson Syndrome (MOWS, OMIM#235730), a rare congenital disease with intellectual disability, epilepsy/seizures and other defects [5-8]. Zeb2's action mechanisms, partner proteins, proven or candidate target genes, and genes whose normal expression depends on intact Zeb2 levels, have been studied in various cell types for explaining specific phenotypes caused by Zeb2 perturbation. Zeb2 DNA-binding around candidate direct target genes helped to explain *Zeb2* loss-of-function phenotypes in ESCs and cells of early and late embryos, and postnatal and adult mice. These genes are involved in ESC pluripotency (*Nanog, Sox2*), cell differentiation (*Id1*, *Smad7*), and maturation of various cell types, e.g., in embryonic cortical and adult neurogenesis (*Ntf3, Sox6*), as well as epithelial-to-mesenchymal transition (EMT) (*Cdh1*) [9-16]. In reverse, subtle mutagenesis of Zeb2 DNA-binding sites in demonstrated target genes has confirmed Zeb DNA-binding and its repressive activity (on the mesoderm TF X*Bra* and on epithelial *Cdh1*) [9, 17].

Despite its critical function in the precise spatial-temporal regulation of expression of system/process-specific relevant genes during embryogenesis and postnatal development, but recently also adult tissue homeostasis and stem cell based repair, and acute and chronic disease [18, 19], ChIP-sequencing (ChIP-seq) data for Zeb2 has been obtained in very few cases only. This is the case for high-Zeb2 hepatocellular carcinoma and leukemia cell lines, or cultured cells that overproduce tag-Zeb2 from episomal vectors or the safe *Rosa26* locus [15, 20-22], none of which represent normal, endogenous Zeb2 levels nor express *Zeb2* with normal dynamics. Furthermore, most if not all anti-Zeb2 antibodies cross-react with Zeb1, so do not discriminate between both proteins when their presence overlaps or succeeds to one another, and they compete for the same target genes, which for the individual proteins depends on cell identity/state, extrinsic stimulation or cellular context (e.g., in somitogenesis, in melanoma) [23,24].

In mouse (m) ESCs subjected to neural differentiation (ND), Zeb2 mRNA/protein is undetectable in undifferentiated cells, whereas its strong upregulation accompanies efficient

conversion of naïve ESCs into epiblast stem cell like cells (EpiLSCs), and is essential for subsequent exit from primed cells and progression to neuroprogenitor cells (NPCs) [15]. We have edited one *Zeb2* allele of mESCs by inserting a Flag-V5-epitope coding sequence just before the stop codon, in-frame with the last exon (ex9 of mouse *Zeb2*) [25]. These Zeb2-V5 mESCs were then differentiated to NPCs, and Zeb2 binding sites determined by V5-tag Chip-Seq. We identified ~2,430 binding sites, of which ~1,950 map to protein-encoding genes. We then cross-referenced the ChIP-positive (+) protein-encoding target genes with RNA-seq data of deregulated genes in cell-type specific neurodevelopment-relevant Zeb2 perturbations [13, 25, 26]. Although we compare non-identical systems, the ESC approach still revealed a number of interesting overlaps, as well as Zeb2's role in regulating critical targets in neurodevelopment. Taken together, we report for the first time the identification of Zeb2 genome-wide binding sites in ESC neural differentiation.

Materials and Methods

Tag-Zeb2 mouse ESCs

gRNAs (Table 1) targeting *Zeb2*-ex9, and tracrRNA (Integrated DNA Technologies, IDT), were diluted to 125 ng/µl in duplex buffer (IDT). gRNAs were annealed to tracrRNA at 1:1 ratio at 95°C for 5 min and cooling the samples to room temperature (RT, 24°C). 250 ng of these annealed gRNAs were transfected in 350,000 mESCs together with 2 µg pX459-Cas9-puro vector and 1 µg ssDNA oligo of the Donor Template containing the FlagV5-tag sequence (Table 1). Transfection was done in a gelatin-coated 6-well plate using DNA:Lipofectamine2000 (ratio of 1:2). Six hours after transfection the medium was refreshed, at 24 hours the cells were Puromycin-selected (2 µg/ml). After 2 days, the remaining cells were transferred to gelatin-coated 10-cm dishes and given fresh ESC-medium (see below). Per dish 1,000; 1,500; or 2,000 cells were plated and allowed to form colonies. Medium was changed every other day. Colonies were picked, expanded and genotyped by PCR (both outer and inner primer sets were used (Table S1; Fig. S1). All candidate clones were validated by Sanger-sequencing; correct clones were expanded and validated by Western blot.

CRISPR/Cas9-mediated deletion of Zeb2 binding site located at chr2:45109746-45110421

Oligonucleotides for gRNAs (Table 1) with target outside of this chr2-region were cloned into *Bbs*I-digested pX330-hspCas9-T2A-eGFP plasmid. All plasmids were sequenced. 4 µg of gRNA-plasmids (1 µg each) were transfected in 350,000 mESCs and selected (see above). After 24 hours these cells were sorted as GFP+ cells (Becton-Dickinson LSR Fortessa). Per well of a 6-well plate 1,000; 1,500; or 2,000 GFP+ cells were plated and colonies allowed to form, picked (see above) and genotyped by PCR using primers flanking this deletion, and within and outside of it. Clones showing a possible heterozygous or homozygous deletion, as concluded from the PCR analysis, were subjected to ND. At D8, they were harvested, RNA was isolated and cDNA synthesized (see below), and amplified (for the primers, see Table 1). All candidate clones were validated by Sanger-sequencing.

ChIP-seq

DNA libraries from input (i.e. control) and V5 ChIPs were prepared using ThruPLEX DNA protocol (TakaraBio) specific for low amounts of DNA, and sequenced on Illumina HiSeq-2500, and single reads of 50 bp were generated. Adapter sequences were trimmed from the 3'-end of the reads, after which the reads were aligned to the mm10/GRCm38 genome using HISAT2 [74]. From the alignments, secondary, supplementary, low quality and fragmented alignments (fragments > 150 bp) were filtered away. Peaks were called with MACS [75], and coverage was determined. 42 and 25 million reads were generated for input and V5 ChIP, respectively.

ChIP-seq data analysis

Peak calling was performed with MACS2 (Galaxy version 2.1.1.20160309.6) [75, 76], with default parameters (narrow peak calling, *Mm*1.87e9, FDR < 0.05) using the input sample as background. The No model parameter was used, and the extension size was set on 210 bp based on the predicted fragment lengths from the alignments (MACS2 predict-tool, Galaxy version 2.1.1.20160309.1) [75, 76]. The distance of the aligned reads from the TSS of the gene was analyzed using ComputeMatrix (Galaxy version 3.3.2.0.0) and PlotHeatmap Galaxy version 3.3.2.0.1; the used matrix is based on the

log2ratio of the aligned ChIP peaks over the input, calculated using BamCompare (Galaxy version 3.3.2.0.0) [77].

RNA-seq

The quality of total RNA (of biologically independent triplicates) of wild-type mESCs at D0, and at ND-D4, D6 and D8, was checked on Agilent Technologies-2100 Bioanalyzer, using a RNA nano-assay. All samples had RIN value of 9.8 or higher. Triplicate RNA-seq libraries were prepared (Illumina TruSeq stranded mRNA protocol; www.illumina.com). Briefly, 200 ng of total RNA was purified using polyT-oligo-attached magnetic beads for ending with polyA-RNA. The polyA-tailed RNA was fragmented, and cDNA synthesized (SuperScript II, random primers, in the presence of Actinomycin D). cDNA fragments were end-repaired, purified (AMPureXP beads), A-tailed using Klenow exo-enzyme and dATP. Paired-end adapters with dual index (Illumina) were ligated to the A-tailed cDNA fragments and purified (AMPureXP beads).

The resulting adapter-modified cDNAs were enriched by PCR (Phusion polymerase) as follows: 30 sec at 98°C, 15 cycles of (10 sec at 98°C, 30 sec at 60°C, 30 sec at 72°C), 5 min at 72°C. PCR products were purified (AMPureXP beads) and eluted in 30 µl resuspension buffer. One µl was loaded on an Agilent 2100 Bioanalyzer using a DNA-1000 assay to determine concentration and for quality check. Cluster generation was performed according to the Illumina TruSeq SR Rapid Cluster kit v2 Reagents Preparation Guide (www.illumina.com). After hybridization of the sequencing primer, sequencing-by-synthesis was performed using a HiSeq-2500 with a single-read 50-cycle protocol followed by dual index sequencing. Illumina adapter sequences have been trimmed off the reads, which were subsequently mapped against the GRCm38 mouse reference (using HiSat2 version 2.1.0) [74]. Gene expression values were called (using HTSeq-count version 0.9.1) [78] and Ensembl release 84 gene and transcript annotation. Sample QC and DEG analysis have been performed in the R environment for statistical computing (version 3.5.3, using DESeq2 version 1.22.1 and Tidyverse version 1.2.1 (https://github.com/tidyverse; https://www.r-project.org/) [79, 80].

Meta-analysis

RNA-seq datasets (as DEG tables; [13, 25]) were downloaded from GEO (https://www.ncbi.nlm.nih.gov/geo/, GSE35616 and GSE103003, respectively). Cross-referencing and visualization were performed in R, using Tidyverse, VennDiagram and pheatmap packages.

Pathways enrichment, GO, Function analysis and Gene-to-Disease association

These were performed with StringDB package for R, while for Gene-to-Disease association Disgenet2R for R was used [81, 82].

Results

Heterozygous Zeb2-Flag-V5 ESCs differentiate as wild-type cells

Addition of short epitope(s) at the N- or C-terminus, as well as activation/repression domains of heterologous TFs at the C-terminus, does not interfere with Zeb2's DNA-binding (in *Xenopus* embryos [27], heterologous cells [28], mouse forebrain [25], mESCs [15]). Here, we have successfully used a CRISPR/Cas9 approach (Materials and Methods) to insert a Flag-V5-tag sequence in *Zeb2*-ex9 of mESCs (clone 2BE3; Fig. S1). Allele-specific RT-qPCR using primers that amplify sequences between the ex9 and V5-tag showed mRNA expression from the tagged allele during ND at day (D) 6 and 8 (Fig. 1A). Western blot analysis confirmed in nuclear extracts of ND-ESCs at D8 (*i.e.*, NPCs [15]) the presence of Zeb2 of expected relative molecular mass, identified by anti-V5 and anti-Zeb2 antibodies (Fig. 1B). Both the Zeb2-V5 and wild-type ESCs were then verified for temporal expression of *Zeb2*, core pluripotency genes (*Pou5f1*, *Nanog*, downregulated upon ND, and *Sox2*, the latter also a neurogenesis TF) and an acknowledged NPC marker (*Pax6)* (Fig. 1C,D). Both cell lines displayed similar gross expression dynamics of these genes, indicating that Zeb2-V5 ESCs at D8 of ND can be used for ChIP-seq. Further confirmation in ND-ESCs came from selective pull-down of Zeb2 on the known target *Cdh1*, using ChIP-qPCR. Zeb2 binds to 2 of 3 E-boxes in the *Cdh1* promoter (Fig. 1E), which it represses during EMT [9. 15]. A 25-fold enrichment for Zeb2-V5 was obtained when probing this same *Cdh1* region, using anti-V5 antibody (αV5) conjugated beads (and

agarose beads as negative control; Fig. 1F). Hence, Zeb2-V5 binds to known Zeb2 target sites, and these NPCs are suitable for mapping the Zeb2 genome-wide binding sites (GWBS) at normal, endogenous levels of Zeb2.

One-third of the 2,432 Zeb2 DNA-binding sites in NPCs map close to the transcription start site of transcriptome-confirmed, system-relevant protein-encoding genes, including *Zeb2* itself

αV5-precipitated samples from upscaled Zeb2-V5 NPCs were used for cross-linking ChIP-seq, followed by analysis with Galaxy Software [29]. Of the total of 2,432 significant peaks (for threshold setting, Materials and Methods), 94% (2,294 peaks in total) mapped to loci (1,952 in total) that encode protein, while 5% (125 peaks) map to micro-RNA genes (miRNAs), and 1% to regions lacking annotation (ENSEMBL-GRCm38.99) (NA, Fig. 2A; File S1). About 37.5% of all binding sites are located within -10/+10 kb of the annotated transcription start site (TSS, Fig. 2B). Gene ontology (GO) pathway enrichment analysis of these 1,952 loci revealed binding of Zeb2 to classes of gene annotated to Wnt, integrin, chemokine/cytokine (predominantly as defined in inflammation) and cadherin signaling, respectively, and developmental signaling by EGF, VEGF, TGFβ and FGF family pathways (Fig. 2C). Among these 1,952 genes, those for transcription regulatory proteins, post-translational modification and metabolic enzymes, are well-represented (Fig. 2D).

In parallel we applied bulk temporal RNA-seq of wild-type mESCs at D0 (undifferentiated), D4 (neural induction), D6 (early NPCs) and D8 (NPCs) and checked the expression dynamics of the 1,952 Zeb2-bound genes (from the D8 sample); 1,244 of these were subject to significant transcriptional regulation between D4-8 as compared to D0 (Fig. 2E; log2FoldChange <-0.5 or >0.5 and p.value < 0.05; low-stringency analysis was opted for to assess also small differences in mRNA of Zeb2-bound genes). 335 of these genes are commonly expressed between D4-6-8, but at different levels (including *Zeb2* itself; for lists of all differentially expressed genes (DEGs), see File S2). Fig. S2A depicts the D4, D6 and D8 transcriptomes of ND-mESCs, each compared to D0, with indication whether the genes are bound or not by Zeb2, as determined by ChIP-seq. At each of the time points, hence at different Zeb2 mRNA level, about 12% of the DEGs are bound by Zeb2 (Fig. S2B).

Among the Zeb2-bound genes that are down regulated, *Dnmt3l* and *Esrrb* are present, suggesting that upregulation of *Zeb2* in ND-ESCs (D6 and D8) directly causes downregulation of these two genes (Fig. S2C; File S2). Zeb2 has been suggested as direct repressor of *Dnmt3l* and *Esrrb*, facilitating the switch from self-renewal of ESCs to their exit from pluripotency, and promotion of differentiation. Indeed, levels of all *Dnmt3* genes remained higher in *Zeb2*-knockout (KO) ND-ESCs [15]. However, these *Zeb2*-KO cells convert less efficiently into EpiLSCs, and fail to exit from primed pluripotency. Importantly, among the Zeb2-binding genes whose mRNA levels increased during ND, *Zeb2* itself is also present (yellow dot, Fig. S2C), indicating autoregulation. In fact, in this ND model the highest recruitment of Zeb2-V5 in ChIP-seq data was mapped upstream the TSS of *Zeb2* (File S1).

Out of the 1,244 Zeb2-bound genes that significantly changed steady-state mRNA levels in our D4 to D8 transcriptome data sets, 213 are exclusive DEGs in NPCs at D8 (Fig. 2E; Fig. S2D). Among these, *Tcf4* is bound by Zeb2 and becomes highly upregulated in NPCs (Fig. 2E; Fig. S2D). Tcf4 is a ubiquitous basic helix-loop-helix (bHLH) type TF that binds to E-boxes, and is active during central nervous system (CNS) development [30, 31]. In oligodendrocyte precursors, Tcf4 is essential for their subsequent differentiation; Tcf4 dimerizes with the lineage-specific bHLH-TF Olig2, further promoting their differentiation and maturation [32], while Zeb2 together with upstream Olig1/2 are essential for myelinogenesis in the embryonic CNS [11]. Here, Zeb2 generates anti-BMP(-Smad)/anti-Wnt(-β-catenin) activities, which is crucial for embryonic CNS oligodendrocyte differentiation. The regulatory action of ZEB2 on the *TCF4* target gene suggested by our ChIP-seq may underpin phenotypic similarities between MOWS and Pitt-Hopkins syndrome patients [33], the latter caused by mutations in *TCF4*, making us speculate that *TCF4* may be deregulated in MOWS neural cells.

Meta-analysis of RNA-seq data from selected neural-system *Zeb2* perturbations and the NPC ChIP-seq data reveal interesting overlaps

We performed a meta-analysis of three published transcriptome data sets from controls and *Zeb2*-KOs: sorted E14.5 mouse ventral forebrain interneurons (Nkx2.1-Cre driven *Zeb2*-KO; [25]) and sorted (day-2 after birth) progenitors of the V-SVZ, an adult neurogenic niche (Gsh2-Cre, see [25, 13]. In addition, we used high-throughput RT-qPCR data generated on a Fluidigm platform and obtained after

esiRNA-based knockdown (KD) of Zeb2, as part of a systems-biology study in ND-mESCs (the Zeb2 KD data subset were provided by R. Dries [26]).

The DEGs upon these respective Zeb2 perturbations (p.value <0.05; log2FoldChange<-1 and >1) were then filtered. This identified (i) genes that depend on normal Zeb2 levels for their downregulation/repression (if directly by Zeb2, then as repressor) or (ii) other genes that depend on Zeb2 for their upregulation/activation (if directly by Zeb2, as activator) (for Zeb2 as dual TF, see [19, 34]). In parallel, our 2,294 Zeb2-V5 sites in the 1,952 protein-encoding genes were filtered from the complete ChIP-seq dataset, and then used as reference for the RNA datasets (Fig. 3A). This cross-referencing identified 108 protein-encoding genes among the three transcriptomic and the ChIP-seq datasets (Fig. 3A). Fig. S3 shows a heatmap of the changes in mRNA levels of these 108 genes in ND of wild-type ESCs and their correlation with the analyzed datasets. Noteworthy, only *Cxcr4* was common in all RNA datasets. This is likely due to the two RNA-seq sets generated in different brain/neuron cell-type *in vivo* mouse models, while the other RNA data documented the effects of Zeb2-KD on mRNA levels of (only 96) TGFβ/BMP-system components [26]. Cxcr4 and its ligand Cxcl12/Sdf-1 are crucial for migration of interneurons from the ventral forebrain to the neocortex [35, 36], processes co-controlled by Zeb2 [25]. These 108 genes are further involved in regulation of stem cell pluripotency, signaling by TGFβ, FoxO and Hippo, and in axon guidance (Fig. 3B).

Zeb2 directly controls TGFβ/BMP-system component and neuronal differentiation/migration genes

We then validated 14 out of the 108 cross-referenced target genes, either already known as Zeb2-target (*Nanog*) or as TGFβ/BMP-system component (*Bmp7*, *Tgfbr2*, *Smad1*, *Smad2*, *Smad3*, *Id2*, *Cited2*) or with crucial role in neurogenesis and neuronal maturation (*Sema3f*, *Cxcr4*, *Lhx5*, *Ntng2*, *Pax6*, *Tcf4*; their mRNA levels in ND-ESCs are highlighted in the heatmap in Fig. S3). Because *Zeb2*-KO ESCs do neither exit from primed pluripotency nor enter differentiation (Stryjewska *et al.*, 2017), we chose to validate our findings using shRNA-mediated Zeb2-KD at ND-D8 (shZeb2; Materials and Methods), and analyzed these 14 genes two days later (D10) (Fig. 4A); at this read-out time point >50% reduction of *Zeb2* mRNA was obtained (Fig. 4B, top left panel).

Zeb2-KD resulted in reduced mRNA levels of *Cxcr4*, *Ntng2* and *Pax6* (Fig. 4B, bottom row), each involved in neuron specification and migration [37, 38]. Zeb2 -KD also caused down regulation of *Lhx5*, involved in differentiation of interneurons, including cytoskeletal rearrangements during dendritogenesis [39] and of *Tcf4*, involved in neurogenesis [30, 31]. *Sema3f* is an axon outgrowth and neuronal guided migration cue, and was slightly upregulated (Fig. 4B). These results confirm the regulation by Zeb2 of its direct targets in later phases of neuronal differentiation/migration; mutations either in protein-coding or promoter/regulatory regions of these targets cause neurodevelopmental disorders (see Discussion).

The expression of *Nanog*, the promoter of which binds Zeb2 repressor [15], was increased in Zeb2-KD cells (Fig. 4B, upper row). Zeb2-KD caused increased mRNA of *Bmp7*, *Tgfbr2*, *Smad1*, *Smad2*, *Smad3* and *Cited2* (Fig. 4B, upper/middle row), fitting with the normal levels of Zeb2 mounting anti-TGFβ/BMP family effects [19]. In contrast, *Id2* is significantly downregulated in shZeb2 ESCs (Fig. 4B, middle row). *Id2* is normally activated by BMP-Smads and, together with other Id proteins (Id1, Id3 and Id4), inhibits cell differentiation. e.g., Zeb2 represses *Id2* in immune cells to promote differentiation [12]. However, *Id2* as well as other *Id* genes [40-42] is, like *Zeb2* [10, 25], expressed in the developing forebrain. Taken together, these data suggest an active and direct role for Zeb2 in repressing genes regulating stem cell pluripotency as well as a number of TGFβ/BMP-system components (*Bmp7*, *Smad1/2/3*), but also in activating genes in neurogenesis (*Cxcr4*, *Ntng2*, *Lhx5*).

Zeb2 potentiates its own gene expression, which is also crucial for proper control of its direct target genes, but not for all

Strikingly, in our ChIP-seq dataset, the peak with the highest enrichment (~200-fold) mapped 232 bp upstream of the Zeb2 TSS (Fig. 5A, File S1). We deleted this region (chr2:45109746-45110421) using CRISPR/Cas9 in wild-type cells (Materials and Methods), thereby obtaining $Zeb2^{\Delta P/\Delta P}$ ESCs. Fig. 5B (top left panel) shows that the Zeb2 mRNA levels in the homozygous ΔP clone stayed significantly too low during ND, already from D8 onwards, compared to control cells. This classifies the $Zeb2^{\Delta P/\Delta P}$

clone as a very successful alternative Zeb2-KD. We used *Zeb2*$^{\Delta P/\Delta P}$ mESCs to read-out the same genes that depend on intact Zeb2 levels and are Zeb2 *ChIP+* (see Fig. 4B). Levels of *Zeb2* mRNA stayed abnormally very low at D10 in *Zeb2*$^{\Delta P/\Delta P}$ ND-mESCs, wheereas *Nanog* was still expressed and remained higher than in control cells (Fig. 5B, top row). Hence, Zeb2 levels, including those achieved by autoregulation, are critical, but to a different degree, for proper regulation of neuronal-relevant genes such as *Cxcr4*, *Lhx5*, *Ntng2*, *Pax6*, *Tcf4* (Fig. 5B, bottom row). Among the TGFβ/BMP-system components (see Fig. 4) we observed a slight reduction of *Bmp7*, *Smad1* and *Smad3*, whereas *Tgfbr2*, *Smad2*, *Cited2* and *Sema3f* were not affected in *Zeb2*$^{\Delta P/\Delta P}$ ND-mESCs.

We speculate that Zeb2, and now including its novel autoregulation, plays an active role in regulating a number of genes involved in neuron determination and, together with other TFs, TGFβ/BMP-system component genes. Thus, the precise amounts of Zeb2 and in a critical stage also its autoregulation are crucial in discriminating genes where Zeb2 plays the aforementioned primary, active role (as for *Cxcr4*, *Lhx5*, *Ntng*2 etc). For these genes, ~50% reduction (as observed in the Zeb2-KD) or the novel *ΔP/ΔP* "peak"mutation is sufficient to strongly deregulate them, but other genes are either not or slightly affected (*Sema3f*, *Smad2*, *Cited2* vs. *Bmp7*, *Tgfbr2*, *Smad1*, *Smad3*).

Because Zeb2 also binds phospho(p)-Smads [4, 11, 13, 19], we also scanned the Zeb2 ChIP+ direct target genes, and for which we saw strong deregulation upon Zeb2-KD and/or in *Zeb2*$^{\Delta P/\Delta P}$ cells during ND (wherein Smad-activation is not stimulated), for the presence of (i) the Zeb half-sites CACCT(G) (pragmatically neglecting the variable spacing between them) (ii) that combine with candidate p-Smad binding and responsive genes (using GTC($^{T}/_{G}$)CT($^{T}/_{G}$)($^{A}/_{C}$)GCC for p-Smad1/Smad5, GTCTAGAC for p-Smad2/3) and (iii) the co-Smad Smad4 (C($^{C}/_{T}$)AGAC) [43], using the Jaspar database. Fig. S4 shows the distribution of such identified Zeb- and Smad-binding motifs (threshold score >85%) in those genes strongly affected by Zeb2-KD and/or in *Zeb2*$^{\Delta P/\Delta P}$ cells. Interestingly, in the regions where Zeb2 binds close to the TSS (*Zeb2*, *Ntng2*, *Lhx5*, *Nanog*), the p-Smad and Smad4 binding elements are sometimes present in very close proximity of the *ChIP+* Zeb2-bound E-box.

Discussion

We report for the first time the endogenous GWBS for Zeb2 by ChIP-seq, in ESC-derived NPCs tailored for this. We have previously used *Zeb2*$^{\Delta ex7/\Delta ex7}$-KO mESCs and, for rescue purposes, such cells in which Flag$_3$-Strep-Zeb2 was produced from a Cre-controllable *Rosa26* locus [15, 44] . Conceptually, with regard to Zeb2 levels, the latter cells are different from the mESCs established now, for they had the drawback that *Zeb2* is not subjected to its normal temporal regulation during differentiation. However, precise dosage of Zeb2 is critical, as concluded from transgenic Zeb2 cDNA-based rescues in *Zeb2*-KO ESCs and similar genetic rescues in *Zeb2*-mutant cells in mice, which can via heterozygous/homozygous combinations create a large panel of Zeb2 levels (in interneurons [25], in NK cells [45], in ESCs [15]). Another illustration of fine-tuned control of Zeb2 levels are miRs (targeting Zeb2) and lncRNAs (that regulate these miRs) [46-49] , and Zeb2 on its turn also controlling some of its own miR-encoding genes or clusters [50-52]. In our ChIP-seq we find 125 peaks (~5% of the total) that correspond to the TSSs of 98 miR-genes (File S1). Among these miR-genes, Zeb2 binds to loci encoding miR-144, miR-148a, miR-9 and miR-153, known to target Zeb2 in the context of e.g., tumor progression [53-56]. We recently added the identification, in human iPSCs subjected to ND, of *Zeb2* distant (~600 kb upstream) enhancers, which act through DNA-looping to the *ZEB2* promoter-proximal region [57]. In addition, we have documented dynamic expression patterns of *Zeb2* in early embryos [17, 23, 58-60]. We have also shown that cDNA-based expression of various tag-Zeb2 proteins is compatible with functional embryology-type and action mechanism studies [15, 27, 28]. Importantly, our Zeb2-V5 allele enables normal production of tag-Zeb2 for the first time from its endogenous locus.

Only two studies present ZEB2 ChIP-seq data in human cells, i.e. SNU398 hepatocellular carcinoma and K562 erythroleukemia cells, respectively [21, 22]. In K562 ZEB2 binds to the promoters of *NR4A2*, *NEUROG2* and *PITX3*, expressed in midbrain dopaminergic neurons, wherein Zeb2 negatively regulates axon growth and target innervation [61]. In SNU398, ZEB2 represses the *GALNT3*, which is normally expressed in epithelial cells. This repression co-incides with acquisition of a mesenchymal phenotype, linking ZEB2 here again to an EMT-like process. These two valuable studies also present limitations. The use of cancer cell lines of genomic instable nature may create

possible bias in ChIP-seq, and they overproduce ZEB2. Our ChIP-seq identifies >2,400 peaks for Zeb2-V5 in mESCs at ND-D8. This is in cells not stimulated with TGFβ family ligands, with near-38% of Zeb2 sites mapping close to the TSS (-10/+10kb). Most of these genes function in growth factor/cytokine signaling and/or encode transcriptional regulators, the latter suggesting that Zeb2 orchestrates other (co)TFs driving the transcriptomic signature of NPCs. The regulation of Wnt signaling by Zeb2 is in line with observations that inhibition of the Wnt-βcatenin pathway suppresses ND *in vitro* and *in vivo*, and that Wnt (and Zeb2)-controlled *Tcf4* expression promotes neurogenesis and is required for normal brain development [62-66].

The 2,294 Zeb2 peaks map to 1,952 protein-coding genes, of which 1,244 are DEGs in ND-mESCs. Strikingly, the strongest enrichment of Zeb2 occurs on the *Zeb2* promoter itself, nicely correlating with highest *Zeb2* level during ND both in human iPSCs and mouse ESCs (*data not shown*). These high levels depend on DNA-looping bringing three enhancers, located ~600 kb upstream of the TSS, to the *Zeb2* promoter [57], at least at early NPC state, but is then rapidly lost. We have now identified a novel self-regulatory mechanism where Zeb2 binds upstream its TSS to maintain its levels sufficiently high, also and at least during ND. While this autoregulation and/or distant-enhancer mechanisms need further investigation in Zeb2-dependent differentiation and/or maturation of other cell types (e.g. in cKO mouse models or MOWS patient cells), we propose that lower Zeb2 levels might compromise this autoregulatory loop. When we deleted the Zeb2 autoregulatory site from both *Zeb2* alleles (the *ΔP/ΔP* cells), *Zeb2* mRNA levels drop significantly, but *Zeb2* is still expressed, and these cells can still exit from pluripotency and differentiate (*data not shown*). A number of genes, which are mainly linked to neuron maturation, are significantly affected in $Zeb2^{\Delta P/\Delta P}$ ESCs, whereas TGFβ/BMP-system component genes are not deregulated. Zeb2 dosage might thus underlie this difference in regulating its direct, *ChIP+* genes in our ESCs. Zeb2 might be key to maintaining expression of neuronal genes, while for TGFβ/BMP-system genes Zeb2 may co-operate with other TFs (including p-Smads) or DNA-modifying enzymes to regulate the expression of targets.

Zeb2 binds to p-Smads, and several studies indicate its negative regulation of BMP-Smad activation of specific target genes, although Zeb2 also has Smad-independent functions [13, 19].

BMP-pSmads bind to GGCGCC with high affinity [67]. Morikawa *et al.* [68] have confirmed these results using ChIP-seq, but identified a lower-affinity (higher BMP-doses required) BMP-Smad element (GGAGCC). For achieving full responsiveness it was proposed that the GG($^{A}/_{C}$)GCC element needs to be coupled with a Smad4 site, ideally located 5 bp away [68]. We find that in primary targets affected by varying levels of Zeb2, E-boxes are located close to Smad-binding motifs. However, whether Zeb2 and Smads are co-present in target regions requires further experiments, such as ChiP-on-ChIP assays, and (non-neural) differentiation protocols (adding BMP and/or Nodal). However, these may be further complicated because of post-translational modification status of Zeb2, nuclear p-Smads and/or Smad4 [69-72].

Striking are also the 1,093 Zeb2-binding DEGs at D8. When we performed a gene-to-disease association using the human orthologues of these D8-DEGs, we found a clear association with disorders (Fig. S6). These include neurodevelopmental, mental and eye defects, which occur in MOWS. Altogether, our data may provide novel insights in MOWS due to suboptimal ZEB2 amounts in humans, and includes an autoregulation aspect from now on, and also that *ZEB2* is a putative modifier gene for many other congenital disorders.

Several *Zeb2*-cKO mouse models have been generated, and for many bulk RNA-seq data are available. Here, we selected two such RNA-seq data sets [13, 25, 73]. In addition, similar data were obtained for cultured mESCs, either *Zeb2*-KO cells [15] or cells submitted to ND wherein e.g., Zeb2-KD was done [26]. Unfortunately, we cannot include *Zeb2*-KO mESCs, for they convert less efficiently to EpiLSCs and fail to differentiate whatsoever [15]. The meta-analysis of these three different data sets, overlaid with the 1,952 Zeb2-V5+ loci/genes, show therefore a limited number of common targets, *Cxcr4* being the only one. This narrows the Zeb2-bound gene collection to 108 in total. However, and interestingly, the latter enrich for GO terms pluripotency of stem cells, signaling by TGFβ and Wnt, cell fate commitment and neuron differentiation, all processes where Zeb2 plays a crucial role. Out of these 108, we selected 14 genes covering TGFβ/BMP, pluripotency, and neuron migration, differentiation/maturation, and checked their levels 2 days after Zeb2-KD at ND-D8. Most of these 14 genes relevant to NPC status critically depend on intact levels of Zeb2. They may help to explain why the defects caused by MOWS are observed later after birth and why (the few) missense

mutations in MOWS (besides the more abundant significant deletions etc) present with milder syndromic manifestation. Both the cross-reference of Zeb2-ChIP+ genes with the transcriptome of ND-ESCs and the meta-analysis identify a number of common genes, such as *Bmp7*, *Tgfbr2*, *Tcf4*, *Smad1/2/3*, and *Sema3f* (File S3; Fig. S3), making Zeb2 a likely direct regulator of these. It is also intriguing that Zeb2 is recruited to *Tcf4*, and controls it at D8, and that *Tcf4* is deregulated upon Zeb2-KD (esiRNA and shRNA). Mutations in *TCF4* cause Pitt-Hopkins syndrome (PTHS, OMIM#610954), a rare neurodevelopmental disorder with some defects overlapping with MOWS.

Conclusion and/or Summary

This ChIP-seq study in a suitable cell system identifies Zeb2 GWBS for the first time and discovers *Zeb2* autoregulation as a new principle, which may add to pathogenic mechanisms of MOWS, Zeb2-executed control in TGFβ/BMP-Smad signaling of target genes, and its role as candidate modifier gene in neurodevelopmental syndromes due to mutations in genes under direct ZEB2 control.

Acknowledgments

This work was supported mainly by Erasmus University Medical Center via departmental funds, extra funds coordinated with its Executive Board, and BIG project funding, a collaborative extra support between Erasmus University Rotterdam and its University Medical Center for promoting fundamental research at the Theme of Biomedical Sciences, and hence the Department of Cell Biology. We thank all co-workers of the Center for Biomics-Genomics at Erasmus University Medical Center for their expert technical assistance, and colleagues Frank Grosveld and Raymond Poot for constructive discussions.

Disclosure of Potential Conflicts of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data sets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

References

1 Funahashi J, Kamachi Y, Goto K *et al.* Identification of nuclear factor deltaEF1 and its binding site essential for lens-specific activity of the delta 1-crystallin enhancer. *Nucleic Acids Res* 1991;19:3543-3547.

2 Sekido R, Murai K, Funahashi J *et al.* The delta-crystallin enhancer-binding protein deltaEF1 is a repressor of E2-box-mediated gene activation. *Mol Cell Biol* 1194;14:5692-5700.

3 Remacle JE, Kraft H, Lerchner W *et al.* New mode of DNA binding of multi-zinc finger transcription factors: deltaEF1 family members bind with two hands to two target sites. *EMBO J* 1999;18:5073-5084.

4 Verschueren K, Remacle JE, Collart C *et al.* SIP1, a novel zinc finger/homeodomain repressor, interacts with Smad proteins and binds to 5'-CACCT sequences in candidate target genes. *J Biol Chem* 1999;274:20489-20498.

5 Garavelli L, Donadio A, Zanacca C *et al.* Hirschsprung disease, mental retardation, characteristic facial features, and mutation in the gene ZFHX1B (SIP1): confirmation of the Mowat-Wilson syndrome. *Am J Med Genet A* 2003;116A:385-388.

6 Zweier C, Thiel CT, Dufke A *et al.* Clinical and mutational spectrum of Mowat-Wilson syndrome. *Eur J Med Genet* 2005;48:97-111.

7 Garavelli L, Ivanovski I, Caraffi SG *et al.* Neuroimaging findings in Mowat-Wilson syndrome: a study of 54 patients. *Genet Med*. 2017;19:691-700.

8 Ivanovski I, Djuric O, Caraffi SG *et al.* Phenotype and genotype of 87 patients with Mowat-Wilson syndrome and recommendations for care. *Genet Med* 2018;20:965-975.

9 Comijn J, Berx G, Vermassen P *et al.* The two-handed E box binding zinc finger protein SIP1 downregulates E-cadherin and induces invasion. *Mol Cell* 2001;7:1267-1278.

10 Seuntjens E, Nityanandam A, Miquelajauregui A *et al.* Sip1 regulates sequential fate decisions by feedback signaling from postmitotic neurons to progenitors. *Nat Neurosci* 2009; 12: 1373-1380.

11 Weng Q, Chen Y, Wang H *et al.* Dual-mode modulation of Smad signaling by Smad-interacting protein Sip1 is required for myelination in the central nervous system. *Neuron* 2012;73:713-728.

12 Scott CL, Soen B, Martens L *et al.* The transcription factor Zeb2 regulates development of conventional and plasmacytoid DCs by repressing Id2. *J Exp Med* 2016;213:897-911.

13 Deryckere A, Stappers E, Dries R *et al.* Multifaceted actions of Zeb2 in postnatal neurogenesis from the ventricular-subventricular zone to the olfactory bulb. *Development* 2020;147:dev184861.

14 Wu LM, Wang J, Conidi A *et al.* Zeb2 recruits HDAC-NuRD to inhibit Notch and controls Schwann cell differentiation and remyelination. *Nat Neurosci* 2016;19:1060-1072.

15 Stryjewska A, Dries R, Pieters T *et al.* Zeb2 Regulates Cell Fate at the Exit from Epiblast State in Mouse Embryonic Stem Cells. *Stem Cells* 2017;35:611-625.

16 Menuchin-Lasowski Y, Dagan B, Conidi A *et al.* Zeb2 regulates the balance between retinal interneurons and Muller glia by inhibition of BMP-Smad signaling. *Dev Biol* 2020;468: 80-92.

17 Lerchner W, Latinkic BV, Remacle JE et al. Region-specific activation of the Xenopus brachyury promoter involves active repression in ectoderm and endoderm: a study using transgenic frog embryos. *Development* 2000;127:2729-2739.

18 van Grunsven LA, Taelman V, Michiels C *et al.* deltaEF1 and SIP1 are differentially expressed and have overlapping activities during Xenopus embryogenesis. *Dev Dyn* 2006;235:1491-1500.

19 Conidi A, Cazzola S, Beets K *et al.* Few Smad proteins and many Smad-interacting proteins yield multiple functions and action modes in TGFbeta/BMP signaling in vivo. *Cytokine Growth Factor Rev* 2011;22:287-300.

20 ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012;489:57-74.

21 Balcik-Ercin P, Cetin M, Yalim-Camci I *et al.* Genome-wide analysis of endogenously expressed ZEB2 binding sites reveals inverse correlations between ZEB2 and GalNAc-transferase GALNT3 in human tumors. *Cell Oncol (Dordr)* 2018;41:379-393.

22 Yang S, Toledo EM, Rosmaninho P *et al.* A Zeb2-miR-200c loop controls midbrain dopaminergic neuron neurogenesis and migration. *Commun Biol* 2018;1:75.

23 Miyoshi T, Maruhashi M, Van De Putte T *et al.* Complementary expression pattern of Zfhx1 genes Sip1 and deltaEF1 in the mouse embryo and their genetic interaction revealed by compound mutants. *Dev Dyn* 2006;235:1941-1952.

24 Vandamme N, Denecker G, Bruneel K *et al.* The EMT Transcription factor ZEB2 promotes proliferation of primary and metastatic melanoma while suppressing an invasive, mesenchymal-like phenotype. *Cancer Res* 2020;80: 2983-2995.

25 Nelles L, Van de Putte T, van Grunsven L *et al.* Organization of the mouse Zfhx1b gene encoding the two-handed zinc finger repressor Smad-interacting protein-1. *Genomics* 2003;82:460-469.

25 van den Berghe V, Stappers E, Vandesande B *et al.* Directed migration of cortical interneurons depends on the cell-autonomous action of Sip1. *Neuron* 2013;77:70-82.

26 Dries R, Stryjewska A, Coddens K *et al.* Integrative and perturbation-based analysis of the transcriptional dynamics of TGFbeta/BMP system components in transition from embryonic stem cells to neural progenitors. *Stem Cells* 2020;38:202-217.

27 Papin C, van Grunsven LA, Verschueren K *et al.* Dynamic regulation of Brachyury expression in the amphibian embryo by XSIP1. *Mech Dev* 2002;111:37-46.

28 Verstappen G, van Grunsven LA, Michiels C *et al.* Atypical Mowat-Wilson patient confirms the importance of the novel association between ZFHX1B/SIP1 and NuRD corepressor complex. *Hum Mol Genet* 2008;17:1175-1183.

29 Afgan E, Baker D, Batut B *et al.* The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res* 2018;46:W537-W544.

30 Corneliussen B, Thornell A, Hallberg B *et al.* Helix-loop-helix transcriptional activators bind to a sequence in glucocorticoid response elements of retrovirus enhancers. *J Virol* 1991;65:6084-6093.

31 Forrest MP, Hill MJ, Quantock AJ *et al.* The emerging roles of TCF4 in disease and development. *Trends Mol Med* 2014;20:322-331.

32 Fu H, Cai J, Clevers H *et al.* A genome-wide screen for spatially restricted expression patterns identifies transcription factors that regulate glial development. *J Neurosci* 2009;29:11399-11408.

33 Brockschmidt A, Filippi A, Charbel Issa P *et al.* Neurologic and ocular phenotype in Pitt-Hopkins syndrome and a zebrafish model. *Hum Genet* 2011;130:645-655.

34 Hegarty SV, Sullivan AM, O'Keeffe GW. Zeb2: A multifunctional regulator of nervous system development. *Prog Neurobiol* 2015;132:81-95.

35 Stumm RK, Zhou C, Ara T *et al.* CXCR4 regulates interneuron migration in the developing neocortex. *J Neurosci* 2003;23:5123-5130.

36 Nash B, Meucci O. Functions of the chemokine receptor CXCR4 in the central nervous system and its regulation by mu-opioid receptors. *Int Rev Neurobiol* 2014;118:105-128.

37 Georgala PA, Carr CB, Price DJ. The role of Pax6 in forebrain development. *Dev Neurobiol* 2011;71:690-709.

38 Dias CM, Punetha J, Zheng C *et al.* Homozygous missense variants in NTNG2, encoding a presynaptic Netrin-G2 adhesion protein, lead to a distinct neurodevelopmental disorder. *Am J Hum Genet* 2019;105:1048-1056.

39 Lui NC, Tam WY, Gao C *et al.* Lhx1/5 control dendritogenesis and spine morphogenesis of Purkinje cells via regulation of Espin. *Nat Commun* 2017;8:15079.

40 Tzeng SF, de Vellis J. Id1, Id2, and Id3 gene expression in neural cells during development. *Glia* 1998;24: 372-381.

41 Blomfield IM, Rocamonde B, Masdeu MDM *et al.* Id4 promotes the elimination of the pro-activation factor Ascl1 to maintain quiescence of adult hippocampal stem cells. *Elife* 2019;8:e48561.

42 Havrda MC, Harris BT, Mantani A *et al.* Id2 is required for specification of dopaminergic neurons during adult olfactory neurogenesis. *J Neurosci* 2008;28:14074-14086.

43 Hill CS. Transcriptional Control by the SMADs. *Cold Spring Harb Perspect Biol* 2016;8:a022079.

44 Higashi Y, Maruhashi M, Nelles L *et al.* Generation of the floxed allele of the SIP1 (Smad-interacting protein 1) gene for Cre-mediated conditional knockout in the mouse. *Genesis* 2002;32:82-84.

45 van Helden MJ, Goossens S, Daussy C *et al.* Terminal NK cell maturation is controlled by concerted actions of T-bet and Zeb2 and is essential for melanoma rejection. *J Exp Med* 2015;212:2015-2025.

46 Guan J, Liu P, Wang A *et al.* Long noncoding RNA ZEB2AS1 affects cell proliferation and apoptosis via the miR1225p/PLK1 axis in acute myeloid leukemia. *Int J Mol Med* 2020;46:1490-1500.

47 Jiang Y, Wu K, Cao W *et al.* Long noncoding RNA KTN1-AS1 promotes head and neck squamous cell carcinoma cell epithelial-mesenchymal transition by targeting miR-153-3p. *Epigenomics* 2020;12:487-505.

48 Yao H, Hou G, Wang QY *et al.* LncRNA SPRY4IT1 promotes progression of osteosarcoma by regulating ZEB1 and ZEB2 expression through sponging of miR101 activity. *Int J Oncol* 2020;56:85-100.

49 Cheng H, Zhao H, Xiao X *et al.* Long non-coding RNA MALAT1 upregulates ZEB2 expression to promote malignant progression of glioma by attenuating miR-124. *Mol Neurobiol* 2021;58:1006-1016.

50 Brabletz S, Brabletz T. The ZEB/miR-200 feedback loop--a motor of cellular plasticity in development and cancer? *EMBO Rep* 2010;11:670-677.

51 Exposito-Villen A, A EA, Franco D. Functional role of non-coding RNAs during epithelial-to-mesenchymal transition. *Noncoding RNA* 2018;4:14.

52 Gregory PA, Bracken CP, Bert AG *et al.* MicroRNAs as regulators of epithelial-mesenchymal transition. *Cell Cycle* 2008;7:3112-3118.

53 Guan H, Liang W, Xie Z *et al.* Down-regulation of miR-144 promotes thyroid cancer cell invasion by targeting ZEB1 and ZEB2. *Endocrine* 2015;48:566-574.

54 Pan Y, Zhang J, Fu H *et al.* miR-144 functions as a tumor suppressor in breast cancer through inhibiting ZEB1/2-mediated epithelial mesenchymal transition process. *Onco Targets Ther* 2016;9:6247-6255.

55 Nourmohammadi B, Tafsiri E, Rahimi A *et al.* Expression of miR-9 and miR-200c, ZEB1, ZEB2 and E-cadherin in non-small cell lung cancers in Iran. *Asian Pac J Cancer Prev* 2019;20:1633-1639.

56 Wahab NA, Othman Z, Nasri NWM *et al.* Inhibition of miR-141 and miR-200a increase DLC-1 and ZEB2 expression, enhance migration and invasion in metastatic serous ovarian cancer. *Int J Environ Res Public Health* 2020;17:2766.

57 Birkhoff JC, Brouwer RWW, Kolovos P *et al.* Targeted chromatin conformation analysis identifies novel distal neural enhancers of ZEB2 in pluripotent stem cell differentiation. *Hum Mol Genet* 2020;29:2535-2550.

58 van Grunsven LA, Papin C, Avalosse B *et al.* XSIP1, a Xenopus zinc finger/homeodomain encoding gene highly expressed during early neural development. *Mech Dev* 2000;94:189-193.

59 Van de Putte T, Maruhashi M, Francis A *et al.* Mice lacking ZFHX1B, the gene that codes for Smad-interacting protein-1, reveal a role for multiple neural crest cell defects in the etiology of Hirschsprung disease-mental retardation syndrome. *Am J Hum Genet* 2003;72:465-470.

60 Takagi T, Nishizaki Y, Matsui F *et al.* De novo inbred heterozygous Zeb2/Sip1 mutant mice uniquely generated by germ-line conditional knockout exhibit craniofacial, callosal and behavioral defects associated with Mowat-Wilson syndrome. *Hum Mol Genet* 2015;24:6390-6402.

61 Hegarty SV, Wyatt SL, Howard L *et al.* Zeb2 is a negative regulator of midbrain dopaminergic axon growth and target innervation. *Sci Rep* 2017;7:8568.

62 Hirabayashi Y, Itoh Y, Tabata H et al. The Wnt/beta-catenin pathway directs neuronal differentiation of cortical neural precursor cells. *Development* 2004;131:2791-2801.

63 Slawny NA, O'Shea KS. Dynamic changes in Wnt signaling are required for neuronal differentiation of mouse embryonic stem cells. *Mol Cell Neurosci* 2011;48:205-216.

64 Li M, Santpere G, Imamura Kawasawa Y *et al.* Integrative functional genomic analysis of human brain development and neuropsychiatric risks. *Science* 2018;362:eaat7615.

65 Li H, Zhu Y, Morozov YM *et al.* Disruption of TCF4 regulatory networks leads to abnormal cortical development and mental disabilities. *Mol Psychiatry* 2019;24:1235-1246.

66 Mesman S, Bakker R, Smidt MP. Tcf4 is required for correct brain development during embryogenesis. *Mol Cell Neurosci* 2020;106:103502.

67 BabuRajendran N, Palasingam P, Narasimhan K *et al.* Structure of Smad1 MH1/DNA complex reveals distinctive rearrangements of BMP and TGF-beta effectors. *Nucleic Acids Res* 2010;38:3477-3488.

68 Morikawa M, Koinuma D, Tsutsumi S *et al.* ChIP-seq reveals cell type-specific binding patterns of BMP-specific Smads and a novel binding motif. *Nucleic Acids Res* 2011;39:8712-8727.

69 Liu S, Long J, Yuan B *et al.* SUMO Modification Reverses Inhibitory Effects of Smad Nuclear Interacting Protein-1 in TGF-beta Responses. *J Biol Chem* 2016;291:24418-24430.

70 Xu P, Lin X, Feng XH. Posttranslational Regulation of Smads. *Cold Spring Harb Perspect Biol* 2016;8.

71 Kim YJ, Kang MJ, Kim E *et al.* O-GlcNAc stabilizes SMAD4 by inhibiting GSK-3beta-mediated proteasomal degradation. *Sci Rep* 2020;10:19908.

72 Lin X, Wang Y, Jiang Y *et al.* Sumoylation enhances the activity of the TGF-beta/SMAD and HIF-1 signaling pathways in keloids. *Life Sci* 2020;255:117859.

73 McKinsey GL, Lindtner S, Trzcinski B *et al.* Dlx1&2-dependent expression of Zfhx1b (Sip1, Zeb2) regulates the fate switch between cortical and striatal interneurons. *Neuron* 2013;77:83-98.

74 Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* 2015;12:357-360.

75 Zhang Y, Liu T, Meyer CA *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 2008;9: R137.

76 Feng J, Liu T, Qin B *et al.* Identifying ChIP-seq enrichment using MACS. *Nat Protoc* 2012;7:1728-1740.

77 Ramirez F, Ryan DP, Gruning B *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* 2016;44:W160-165.

78 Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* 2015;31:166-169.

79 R Core Team, 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.r-project.org/

80 Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014;15:550.

81 Szklarczyk D, Gable AL, Lyon D *et al.* STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* 2019;47:D607-D613.

82 Pinero J, Ramirez-Anguita JM, Sauch-Pitarch J *et al.* The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Res* 2020;48:D845-D855.

Figure legends

Figure 1. Characterization of the heterozygous Zeb2-Flag-V5 mESC line and ChIP-qPCR validation.

A) Allele-specific RT-qPCR using two sets of primers located either in exon7 (and therefore able to detect the whole Zeb2 mRNA produced by both alleles; orange bar) or located in exon9 and the V5-tag (thus recognizing specifically the knocked-in tagged allele; light blue bar).

B) Western blot analysis showing V5 epitope containing Zeb2 in ES cell derived NPCs (at D8 of neural differentiation, ND) in nuclear extracts (NE), but not in cytoplasmic extracts (CE). Membranes were blotted with anti-V5 antibody (left panel, αV5) or anti-Zeb2 antibody (right panel, αZeb2 [10]. As control, a fraction of Zeb2-rich extract obtained from HeLa cells transfected with pcDNA3-V5Zeb2 vector was also separated in the same gel.

C) Zeb2 mRNA levels in wild-type (WT, orange bar) and Zeb2-V5 (grey bar, clone 2BE3, indicated as Zeb2-V5) mESCs during ND.

D) Pluripotency markers *Nanog*, *Pou5f1* (*Oct4*) and *Sox2* are down regulated in Zeb2-V5 mESCs similarly to WT. The neuronal marker *Pax6* is also significantly upregulated during differentiation, like in WT mESCs.

E) Scheme of the mouse *Cdh1* promoter showing the three E-boxes located upstream the ATG start codon. Zeb2 binds specifically to only two of these, indicated as R1 [15].

F) ChIP-qPCR showing enrichment for Zeb2-V5 binding to the R1 region of the *Cdh1* promoter. Agarose beads were used as negative control (in grey).

Figure 2. Zeb2V5 protein is mainly recruited at the TSS of transcriptional regulator encoding genes, predominantly those classified in Wnt signaling.

A) 2,432 peaks were selected from our ChIP-seq data set (see Materials and Methods). Of these, 94% are associated with protein coding loci, 5% with miRNAs and the remaining 1% map to regions without functional annotation (NA).

B) Frequency plot showing the binding of Zeb2V5 at and around (-10 to +10 kb) the TSS.

C) The 2,294 peaks map to 1,952 protein encoding genes, many of which operate in Wnt signaling (File S1) or are (shown in D) transcriptional regulators.

E) Of the 1,952 protein encoding genes, 1,244 are differentially expressed during ND, when compared to the undifferentiated state (at D0). Of these 1,244 genes, 335 are differentially expressed at all three time points of ND. A few examples are listed of DEGs uniquely expressed at one time point, as well as these that are shared among three time points; a full list is provided in File S2.

Figure 3. Schematic representation of the meta-analysis Zeb2-bound genes versus RNA-seq datasets.

A) 108 genes bound by Zeb2 are also differentially expressed in the three datasets from other studies in mouse models and ESCs (see text for details). *Cxcr4* is the only DEG bound by Zeb2 and common among the three datasets.

B) These 108 genes mainly map to signaling pathways regulating stem cell pluripotency, and effects of TGFβ family, FoxO, and Hippo signaling/activity.

Figure 4. shRNA-mediated KD of Zeb2 discriminate between primary and secondary target genes.

A) Schematic overview of the shRNA transfection targeting Zeb2 (shZeb2) and read-out of the effect. Cellular aggregates at D8 of ND are dissociated and transfected with shZeb2 or against a scrambled, control sequence (shCTRL). Read-out is done two days after the start of shRNA addition. The list of shRNAs is given in Table S2.

B) *Zeb2* levels after KD were reduced to 40-50% of their normal level (shZeb2, orange bars) compared to shCTRL (blue bars). *Bmp7*, *Cited2*, *Nanog*, *Sema3f*, *Smad1*, *Smad2*, *Smad3*, and *Tgfbr2* were significantly upregulated following Zeb2 KD, whereas genes encoding for neuronal specification and migration (*Cxcr4*, *Lhx5*, *Ntng2*, *Pax6* and *Tcf4*) were downregulated.

Figure 5. Deletion of the Zeb2-binding, candidate autoregulatory site (chr2:45109746-45110421) impairs Zeb2 mRNA levels and neuronal markers.

A) Schematic overview of the log(FoldEnrichment) of the peaks identified by ChIP-seq located in the mouse *Zeb2* locus, and localization on top of *Zeb2* intron/exon structure. The highest peak is located 232 bp upstream of the first translated exon. Grey arrow indicates the TSS of *Zeb2*.

B) Zeb2 mRNA levels are strongly reduced in the Zeb2$^{\Delta P/\Delta P}$ clone. Expression levels (mRNA) of the target genes validated with shRNA (Figure 4). Most, but not all of the genes found to be affected following Zeb2 KD are also deregulated in Zeb2$^{\Delta P/\Delta P}$ mESCs in particular *Id2* and the neuronal markers *Cxcr4*, *Lhx5*, *Ntng2*, *Pax6* and *Tcf4*.

Tables

Table 1: **gRNAs and donor template used for CRISPR/Cas9-mediated *Zeb2* editing.**

Name	**Sequence**	**CRISPR/Cas9**
FlagV5 Donor Template	aaaatggaaaccaaatcagaccacgaagaagacaatatggaa gatggcatcgaaGACTACAAAGACGATGACG ACAAG***gatatc***GGTAAGCCTATCCCTAAC CCTCTCCTCGGTCTCGATTCTACG**TAA**a ctactgcattttaagcttcctattttttttccagtagtattgtt	In-frame knock-in of Flag-V5 tag
gRNA_ex9_1	GGAAACCAAATCAGACCACGAGG	
gRNA_ΔZP1	CCCGCGCGCGTTTCAATGGGCGC	Zeb2 peak deletion
gRNA_ΔZP2	CCCTCGCGAGTGCAACACACCAA	
gRNA_ΔZP3	GGGCTCGGAGCGCTGCCGATCGG	
gRNA_ΔZP4	CCGCTGGACCGGGGGGGAGTTGA	

Donor template: lowercase: homology arms located in exon 9 and 3'-UTR of *Zeb2*, underlined lowercase: mutated PAM sequences, uppercase: *FLAG* coding sequence, lowercase italics bold: *Eco*RI restriction site, underlined uppercase: *V5* coding sequence, bold uppercase: STOP codon.

Supplemental figures

Figure S1. Characterization of Zeb2-V5 mESC line.

A) Schematic overview of the design strategy, including showing also primers used for genotyping and detection of tagged allele.

B) Genotyping results (selected part) showing the heterozygous band (red arrow) present in mouse (m) ESC clone 2BE3. Lower band represents the wild-type (WT) allele.

C) Zeb2-V5 specific PCR showing the presence of the tagged allele only in clone 2BE3, and not in WT genomic DNA material. Clone 1BD4 was also used as negative control.

D) Sanger DNA-sequencing results of clone 2BE3 showing the alignment with the Zeb2V5 sequence designed *in silico*. The last three amino acids (i.e. GME) of WT Zeb2 have been modified in DK to remove a PAM sequence and mutate the remaining PAM sequence, target of the gRNA, to avoid multiple cutting of the CAS9. An *Eco*RI restriction site was inserted between the Flag and the V5 coding sequences to facilitate for screening and an artificial STOP codon was added after the V5-coding sequence.

Figure S2. Cross-reference of Zeb2Flagv5-bound protein-coding genes and transcriptome of differentiating mESCs.

A) Volcano plots showing the distribution of Zeb2 bound genes in the whole transcriptomes of D4, D6 and D8 neural differentiated mESCs. Red dots depict the genes bound by Zeb2 and grey crosses those not bound.

B) About 12% of the up- or down-regulated genes are bound by Zeb2 (red bars) at D8 and the percentage does not change when expanding our analysis to the other points of differentiation.

C) Of the 1,952 genes bound by Zeb2 and found being differentially expressed during mESC differentiation, 335 are in common among the three considered time points. *Zeb2* itself is among these common genes and its expression increases during differentiation (yellow dot). The volcano plots show the distribution of the common genes during differentiation.

D) Timepoint specific DEGs depicted as volcano plots. The grey arrow in D8 panel indicates *Tcf4*.

Figure S3. Expression of genes resulting from the meta-analysis of Zeb2-bound genes vs three independent RNA-seq datasets.

Heatmap visualizes the log2FoldChange of the resulting genes during mESC differentiation. RNA-seq datasets where the genes have been found to be differentially expressed are annotated as green (Nkx2.1|Zeb2), purple (Zeb2 KD in mESCs) or yellow (Gsh2|Zeb2) squares.

Figure S4. Distribution of Zeb binding sites (performed as single E-boxes; for details, see text) **and TGFβ/BMP activated (phospho-)Smads and Smad4 binding elements within the identified Zeb2-bound regions for primary target genes.**

A) Grey triangles represent distance from the TSS, while blue triangles depict the TSS of the individual genes. Direction of the arrow shows also whether the genes is transcribed on the + or – strand.

B) Motif analysis identifies the binding motif for Zeb2 as CACCTG.

Figure S5. Genotyping of *Zeb2*$^{\Delta P/\Delta P}$ mESCs.

A) Primers (listed in Table 1) used for genotyping.

B) PCR confirming the proper deletion of the selected region in clone E9.

C) Further PCR showing the difference in molecular weight for regions amplified with different set of primers and discriminating between WT, heterozygous (*Zeb2*$^{\Delta P/+}$) and homozygous (*Zeb2*$^{\Delta P/\Delta P}$**)** deletion clones.

D) Sanger sequencing showing the correct deletion of the selected region in *Zeb2*$^{\Delta P/\Delta P}$ clone.

Figure S6. Gene-to-Disease association of the D8 DEGs bound by Zeb2.

The gene names of the DEGs expressed at D8 of neural differentiation were initially subjected to a mouse to human gene conversion. After that, gene to disease association analysis was performed showing that a number of those genes, could be associated with neurodevelopmental disorders, mental disorders, eye defects, seizures and speech impairment.

Supplementary files

File S1. Narrow peaks obtained from the ChIP-seq.

File S2. Transcriptomic data of Zeb2 bound genes in differentiating mESCs.

File S3. Target genes bound by Zeb2 and deregulated in different already published RNA-seq datasets.

Supplementary tables

Table S1: List of primers used in the study.

Primer name	Sense/Antisense	Sequence (5’ -> 3’)	Application
FlV5mZeb2Ex9_Fwd	Sense	GGCTTACCTGCAGAGCATCA	genotyping
FlV5mZeb2Ex9_Rev	Antisense	CTCCATCTAACTCTGTCTTGGC	genotyping
FlV5_Fwd	Sense	CTACTCGCAGCACATGAATC	genotyping
FlV5_Rev	Antisense	GAGAGGGTTAGGGATAGGC	genotyping
ΔZP_P1_Fwd	Sense	GTCAGTCCGTCCCCAGGTTT	genotyping
ΔZP_P2_Rev	Antisense	GGCATGCTAGCTGGGCTGGT	genotyping
LN249_Fwd	Sense	GGAGCAAACTGAACAAAACCTCGCC	genotyping
LN249_Rev	Antisense	GGCGAGGTTTTGTTCAGTTTGCTCC	genotyping
LN209_Fwd	Sense	AGCGGATCAGATGGCAGTTCGCATG	genotyping
LN209_Rev	Antisense	CATGCGAACTGCCATCTGATCCGCT	genotyping
Zeb2_Fwd	Sense	CAATGCAGCACTTAGGTGTA	qPCR
Zeb2_Rev	Antisense	TTGCCTAGAAACCGTATTGT	qPCR
Zeb2V5_Fwd	Sense	GAAACGATACGGGATGAGGA	qPCR
Zeb2V5_Rev	Antisense	AGGAGAGGGTTAGGGATAGG	qPCR
Nanog_Fwd	Sense	TCT TCC TGG TCC CCA CAG TTT	qPCR
Nanog_Rev	Antisense	GCA AGA ATA GTT CTC GGG ATG AA	qPCR
Pou5f1_Fwd	Sense	AGA GGA TCA CCT TGG GGT ACA	qPCR
Pou5f1_Rev	Antisense	CGA AGC GAC AGA TGG TGG TC	qPCR
Sox2_Fwd	Sense	GCGGAGTGGAAACTTTTGTCC	qPCR
Sox2_Rev	Antisense	CGGGAAGCGTGTACTTATCCTT	qPCR
Pax6_Fwd	Sense	ACATCTTTTACCCAAGAGCA	qPCR
Pax6_Rev	Antisense	GGCAAACACATCTGGATAAT	qPCR
Acrv1b_Fwd	Sense	CTGCCTACAGACCAACTACACC	qPCR
Acrv1b_Rev	Antisense	CCACGCCATCCAGGTTAAAGA	qPCR
Lhx5_Fwd	Sense	AGAACCGAAGGTCCAAAGAA	qPCR
Lhx5_Rev	Antisense	TCACTTTGGTAGTCTCCGTA	qPCR
Ntng2_Fwd	Sense	CAAGGACTCTACGCTTTTCG	qPCR
Ntng2_Rev	Antisense	AGCACTCGCAGTCTTGAAAT	qPCR
Sema3f_Fwd	Sense	CTACACAGCATCCTCCAAGA	qPCR
Sema3f_Rev	Antisense	ACGGCATTCTTGTTTGCATT	qPCR
Smad1_Fwd	Sense	TACTATGAGCTCAACAACCG	qPCR
Smad1_Rev	Antisense	GAAGCGGTTCTTATTGTTGG	qPCR
Smad3_Fwd	Sense	CACGCAGAACGTGAACACC	qPCR
Smad3_Rev	Antisense	GGCAGTAGATAACGTGAGGGA	qPCR

Sox13_Fwd	Sense	CTTACAGGAGGTTGTGCCA	qPCR
Sox13_Rev	Antisense	TCCTTAGCTTCCACATTGCT	qPCR
Stat3_Fwd	Sense	CAATACCATTGACCTGCCGAT	qPCR
Stat3_Rev	Antisense	GAGCGACTCAAACTGCCCT	qPCR
Tcf4_Fwd	Sense	TTGAAGATGTTTTCGCCTCC	qPCR
Tcf4_Rev	Antisense	CCTGCTAGTCATGTGGTCAT	qPCR
Tgfbr2_Fwd	Sense	GAAGGAAAAGAAAAGGGCGG	qPCR
Tgfbr2_Rev	Antisense	TGCTGGTGGTGTATTCTTCC	qPCR
Amylase_Fwd	Sense	GGCTGAGTGTTCTGGGAT	ChIP-qPCR
Amylase_Rev	Antisense	CACGGTGCTCTGGTAGAT	ChIP-qPCR
Cdh1_R1_Fwd	Sense	GCTAGGCTAGGATTCGAACGAC	ChIP-qPCR
Cdh1_R1_Rev	Antisense	TGCAGGGCCCTCAACTT	ChIP-qPCR

Table S2: shRNAs used.

Name	**Sequence**
shZeb2_1	CCGG**CCGAATGAGAAACAATATCAA**CTCGAGTTGATATTGTTTCTCATTCGGTTTTTG
shZeb2_2	CCGG**CCTCAGGAATTTGTGAAGGAA**CTCGAGTTCCTTCACAAATTCCTGAGGTTTTTG
shZeb2_3	CCGG**CCAGTGTCAGATTTGTAAGAA**CTCGAGTTCTTACAAATCTGACACTGGTTTTTG
shZeb2_4	CCGG**CCCATTTAGTGCCAAGCCTTT**CTCGAGAAAGGCTTGGCACTAAATGGGTTTTTG
shCTRL	CCGGCAACAAGATGAAGAGCACCAACTCGAGTTGGTGCTCTTCATCTTGTTGTTTTT

Supplementary Materials and Methods

ESC culture and differentiation

CGR8 (strain 129) wild-type and Zeb2-Flag-V5+ mESCs were cultured and differentiated towards the neural lineage (Bibel *et al.*, 2007, with few modifications). Briefly, mESCs were cultured on 0.1% Gelatin-coated plates in ESC-medium: DMEM supplemented with 15% heat-inactivated (HI) FBS, 2mM L-Glutamine, 1x Non-Essential Amino Acids (NEAA), 143 µM β-Mercapto-Ethanol (β-EtSH) (all ThermoFisher Scientific, TFS) and LIF (10^3 U/ml).

For ND, $4x10^6$ cells were plated on non-adherent 10-cm dishes (Greiner) and allowed to form cellular aggregates (CAs) in 10 ml CA-medium (DMEM, 10% HI-FBS, 2 mM L-Glutamine, 1x NEAA, and 143 µM β-EtSH). From D4 of ND, cells were grown in CA-medium supplemented with 5 µM Retinoic Acid (RA). During the aggregation stages of ND, medium was changed every other day by carefully collecting the aggregates with a 10-ml pipet and transferring them to a 15-ml conical tube. The CAs were allowed to sink to the bottom of the tube where, after the previous medium was carefully discarded, the CAs were then resuspended in fresh medium and transferred back to the dishes. At D8 of ND, the aggregates were harvested and dissociated by resuspension in 1 ml Accutase (TFS) and pipetting them up-and-down using a 1-ml pipet, after shaking them in a 37 C water bath for 5 min. The Accutase was deactivated by adding 9 ml of fresh N2-medium (DMEM with 2 mM L-Glutamine, 50 µg BSA/ml and 1x N2-supplement) to the dissociated cells and pelleting the cells gently for 5 min at 200 g. The cells were resuspended in fresh N2. To ensure single-cell suspension, the cells were filtered by passing them through a 40-mm nylon cell strainer (Corning). 2.5 x 10^5 cells/cm^2 were plated on Poly-DL-Ornithine hydrobromide (Sigma) / Laminin (Sigma) coated plates. Cells were harvested at D8 or D10 of ND.

Bibel M, Richter J, Lacroix E *et al.* Generation of a defined and uniform population of CNS progenitors and neurons from mouse embryonic stem cells. *Nat Protoc* 2007;2:1034-1043.

Western blots

To check Zeb2-V5 protein, the 2BE3-clone ESCs were subjected to ND till D8. Cytoplasmic and nuclear fractions were extracted using NePer-kit® (TFS). Protein concentrations were measured using the Bradfort BCA (TFS) and equal quantities of protein lysates were loaded on 6% SDS-PAGs and thereafter cut based on protein relative molecular mass. Gels were then transferred onto nitrocellulose membranes (Amersham Bioscience), which were incubated overnight with anti-Zeb2 (Seuntjens *et al.*, 2009) and anti-V5 (Life Technologies) antibody, followed by incubation at RT with HorseRadish Peroxidase (HRP) conjugated secondary anti-rabbit and anti-mouse antibodies, respectively (Jackson ImmunoResearch). Protein bands corresponding to Zeb2 or Zeb2-V5 were visualized on an AI-600 digital imager (Amersham|). As loading control, we used Valosin-containing Protein (VCP) and anti-VCP antibody (Santa Cruz sc-57492, mouse).

RNA extraction and RT-qPCR analysis

Total RNA was extracted from ESCs using TRI Reagent (Sigma), and used for cDNA synthesis with RevertAid RT Kit (TFS) with oligodT-primers. RT-qPCR was performed using SybrGreen dye (BioRad) on a CFX96 T1000 thermal cycler (BioRad). All data shown are averages of 3 independent biological replicates and 3 technical replicates, normalized to β-Actin mRNA levels. Primers are listed in Table 2. Analysis and data visualization was performed in R environment for statistical computing version 3.5.3, implemented with the *tidyverse* version 1.3 package (https://github.com/tidyverse).

ChIP

For ChIP $2x10^8$ cells were harvested at ND-D8 in 10 ml of PBS, and cross-linked at 1% formaldehyde (Sigma Aldrich) for 15 min, rotating at RT. Quenching followed with 125 mM glycine during 5 min, again rotating at RT. Cross-linked cells were washed twice with ice-cold PBS (5 min, 1,500 rpm (240 rcf), 4˚C), the pelleted cells snap-frozen, and stored at -80 □C. For sonication, the cell pellets were thawed on ice, resuspended in 1 ml sonication buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA, 0.5 mM EGTA) supplemented with protease and phosphatase inhibitors (PPI, Roche) and incubated on ice for 10 min. DNA was sheared by sonicating the cells using a probe sonicator (32 cycles, 30 sec-on

amplitude 9, and 30 sec-off). These samples were centrifuged at 13,200 rpm (17,000 rcf) for 10 min at 4 °C. Chromatin pellets were snap-frozen and stored at -80 °C.

To check sonication efficiency, 50 µl of sample was de-crosslinked overnight by adding 5 mM NaCl at 65 °C, shaking (950 rpm, Eppendorf ThermoMixer C). The next morning 5, 10 and 20 µl of sample were loaded on a 2% agarose gel, revealing ideally a DNA-smear around 300 bp. A 50-µl sample was used as control input. For immunoprecipitation, chromatin of 10^7 cells was diluted in ChIP-dilution buffer (17 mM Tris-HCl pH8, 170 mM NaCl, 1.2 mM EDTA, 0.01% SDS, 1.1% Triton X-100, with 1xPPI) to a final volume of 1 ml. Samples were pre-cleared by adding pre-washed Protein A/G agarose beads (Santa Cruz) and further incubation for 1 hour, rotating at 4 °C. Then, samples were centrifuged for 1 min (1,000 rpm; 106 rcf) at 4 °C, and the pre-cleared chromatin (supernatant) was transferred to a new low-binding 1.5-ml tube and incubated with 50 µl of pelleted V5-Agarose beads (Sigma Aldrich), rotating at 4 °C overnight. Before addition of V5-agarose beads, they were washed 5 times (5 min each) in PBS by rotating them. As negative control, half of the sample was incubated with Protein A/G beads (Santa Cruz, sc-2003).

The following day the beads were pelleted (1,000 rpm; 1 min) and washed as follows: once with lower-salt buffer (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 2 mM EDTA, 0.1% SDS, 1% Triton X-100), transferred to non-stick low-binding 1.5-ml tubes and then washed once with high-salt buffer (i.e. lower-salt buffer, but now 500 mM NaCl), once washed with LiCl buffer (10 mM Tris-HCl pH 8.0, 250 mM LiCl, 1 mM EDTA, 1% NP-40, 1% sodium deoxycholate (DOC), and twice washed with 10 mM Tris-HCl pH 8.0, 1 mM EDTA (each incubation for 5 min, rotating at 4 °C, followed by gently spinning down.

The protein-chromatin was then eluted from the beads by adding 250 µl of elution buffer (1% SDS + 100 mM $NaHCO_3$), rotating for 1 hour at RT twice, and combining the eluates from both steps. To the input sample, 450 µl of elution buffer was also added, and all samples were de-crosslinked by addition of 5 mM NaCl at 65 °C overnight, shaking 950 rpm. The day-after, 2µl Proteinase-K (10 mg/ml), 20 mM (final concentration) Tris-HCl pH6.5, 5mM (final concentration) EDTA pH8.0 and 10 mg/ml RNase-A (Sigma) were added to each sample and incubated for 1 hour at 45°C while shaking (700 rpm). DNA was extracted from the samples using the PCI method and diluted in water. Five

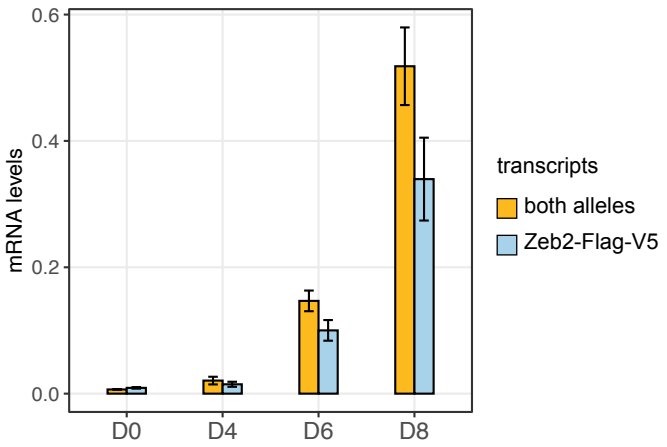
independent ChIPs were performed, for a total of 10^8 mESCs used per condition, and pulled-down chromatin was pooled. ChIP efficiency was assessed by qPCR using primers amplifying *Cdh1* promoter sequences bound by Zeb2 (Stryjewska *et al.*, 2017). All primers used are listed in Table S1.

Zeb2 shRNA-mediated KD

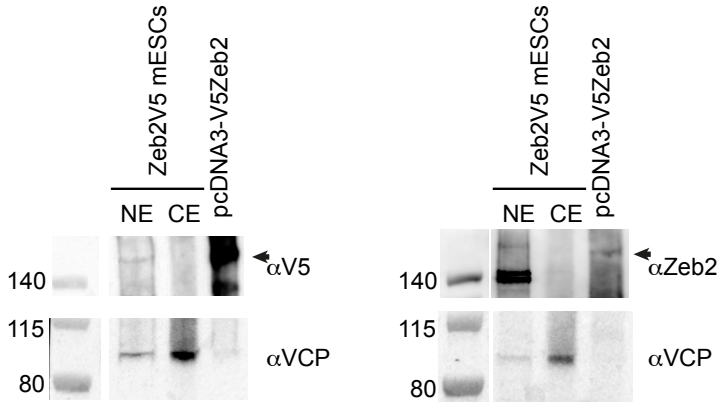
Zeb2-KD in was done by transfecting Zeb2-shRNAs into mESCs at ND-D8. For this, the CAs were dissociated as described above, and single-cell suspensions were transfected using Amaxa Nucleofector II (using kit V, program A-33). In total 4 µg of shRNA was used for transfection of $4.5x10^6$ cells. After transfection, ells were plated in 5 ml N2-medium on a poly-ornithine/laminin-coated 6-cm cell culture dish. Two hours post-transfection, the medium was refreshed, and 24 hours after transfection was changed to N2-medium+Puromycin (see above) for 48 hours. The cells were then harvested, and KD efficiencies examined by RT-qPCR. As a control, scrambled shRNA was used.

Figure 1

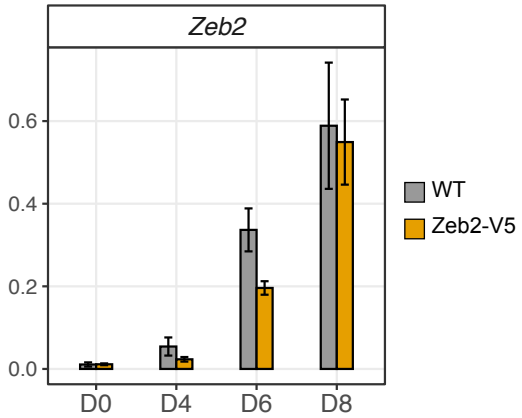
A



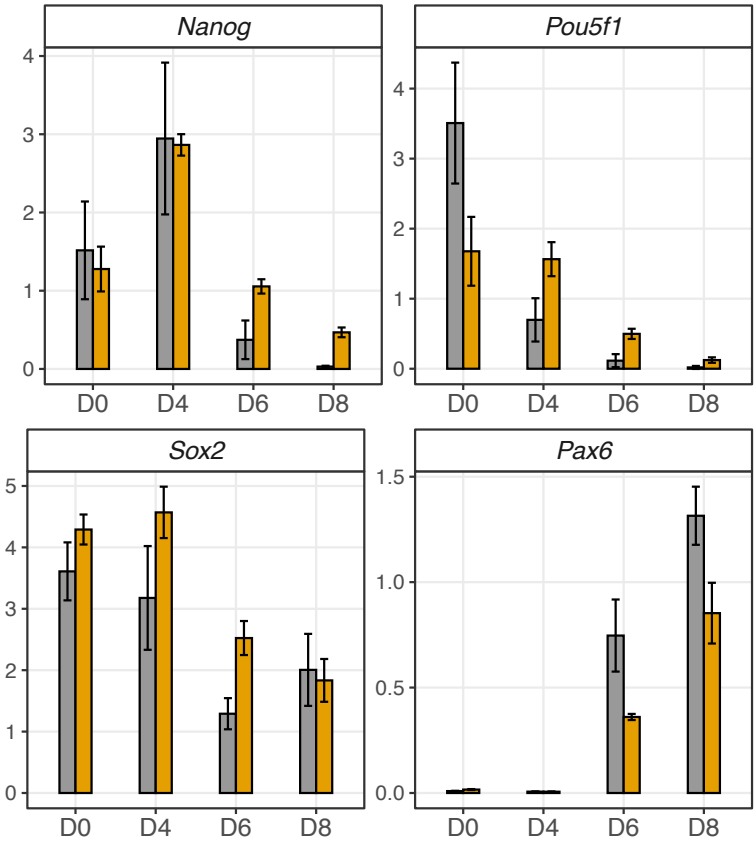
B



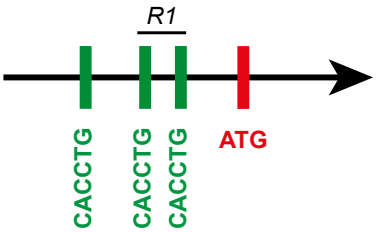
C



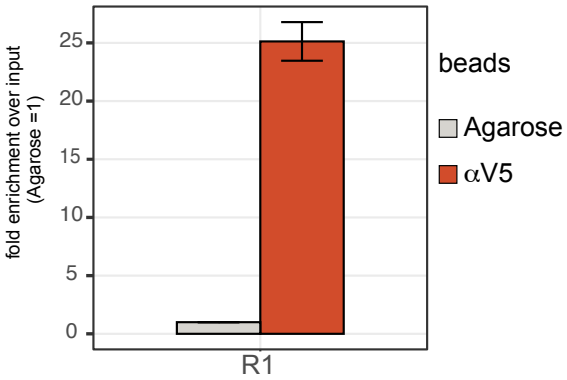
D



E



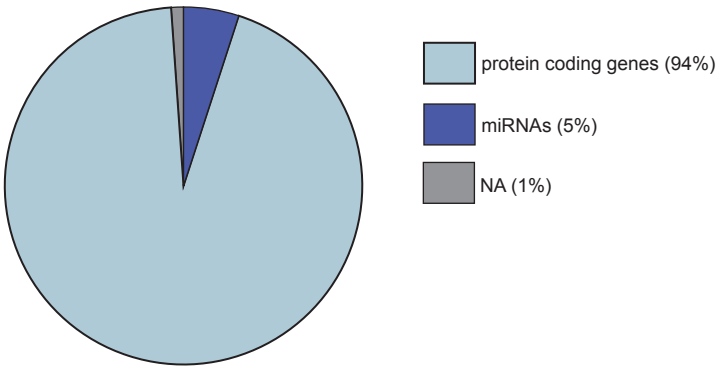
F



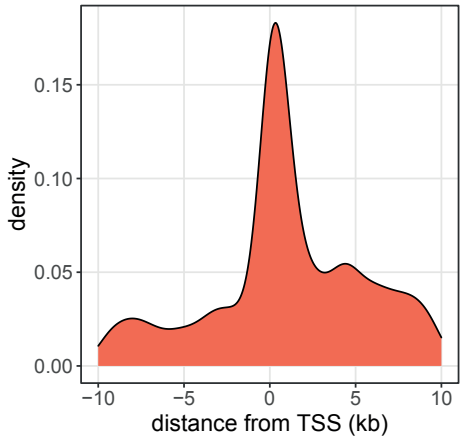
Birkhoff et al 2021_
Zeb2V5 ChIPseq

Figure 2

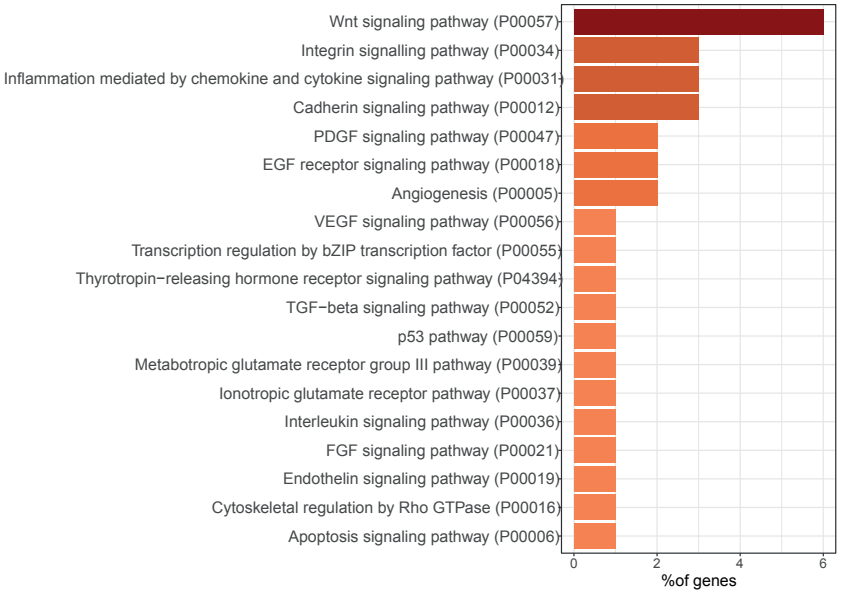
A



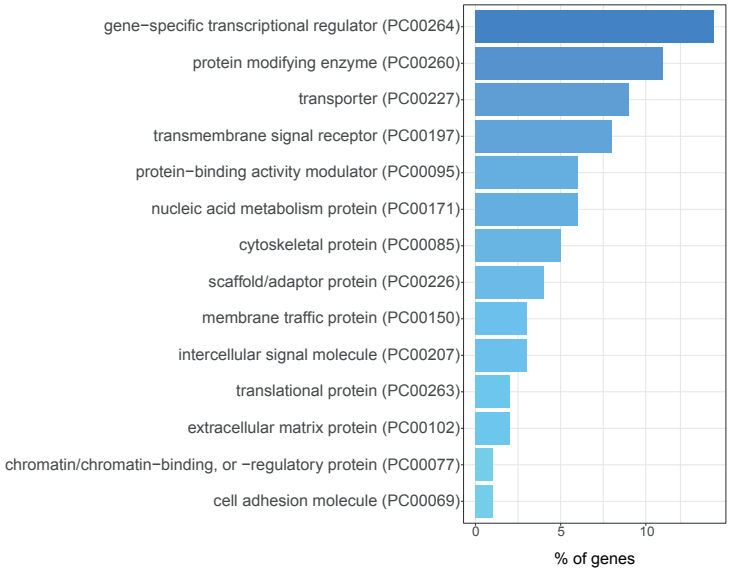
B



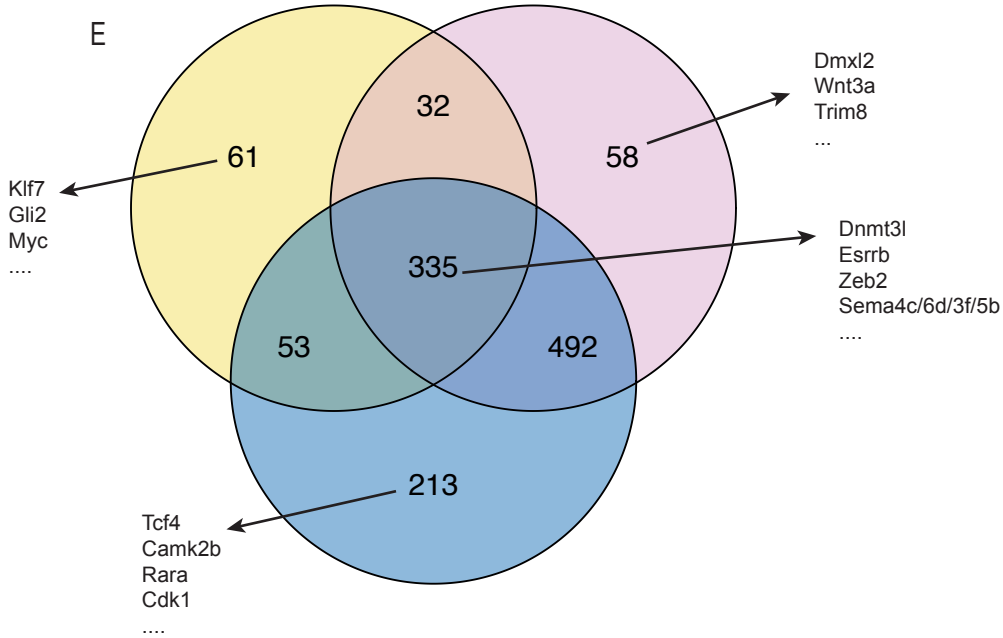
C



D



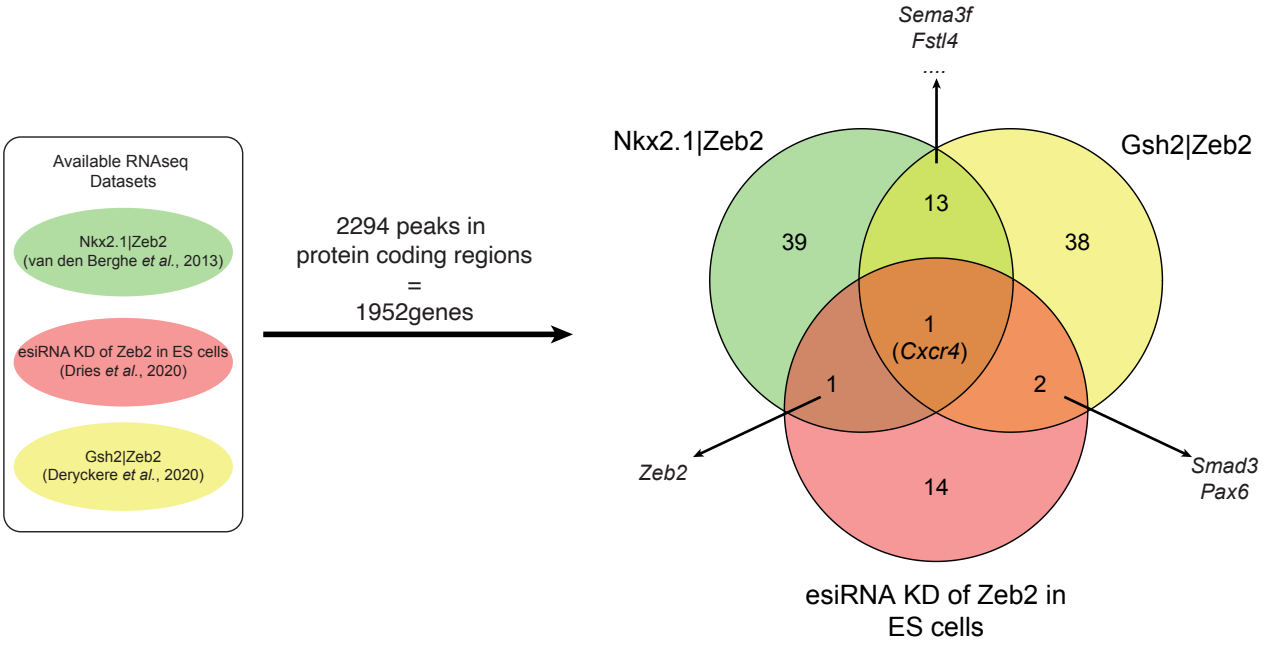
E



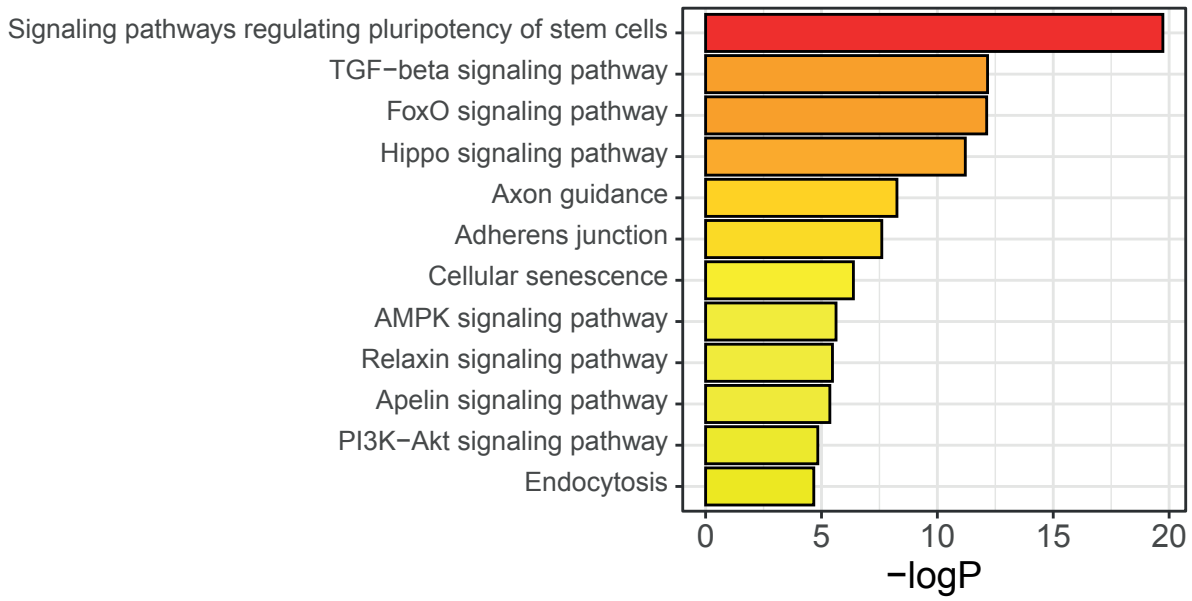
Birkhoff et al 2020_
Zeb2V5 ChIPseq

Figure 3

A



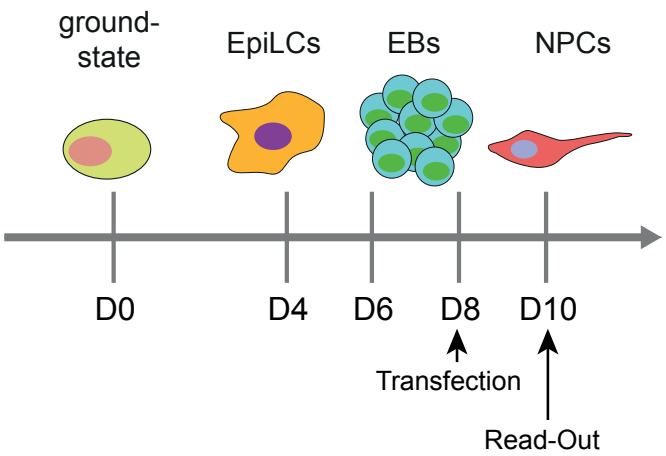
B



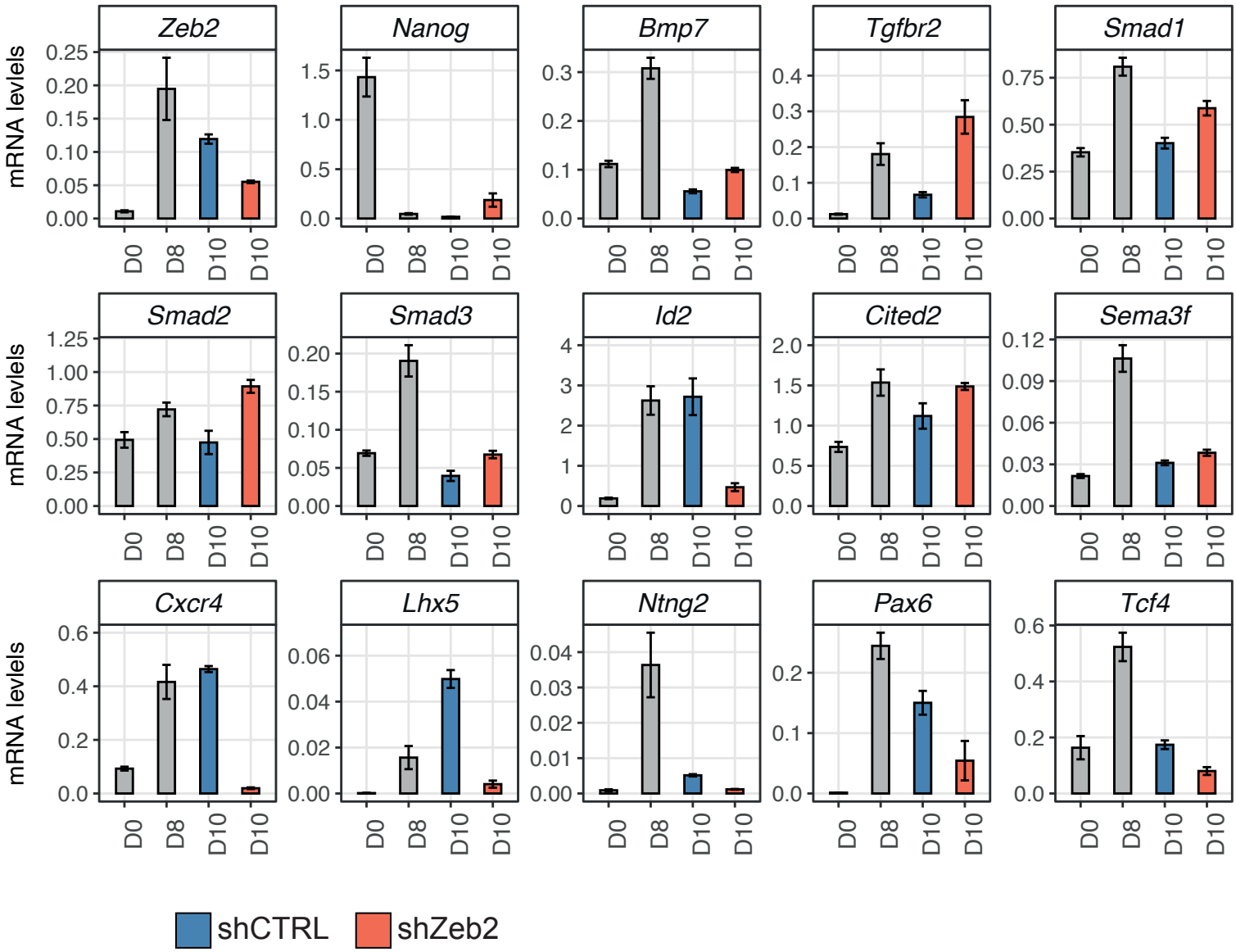
Birkhoff et al 2020_
Zeb2V5 ChIPseq

Figure 4

A

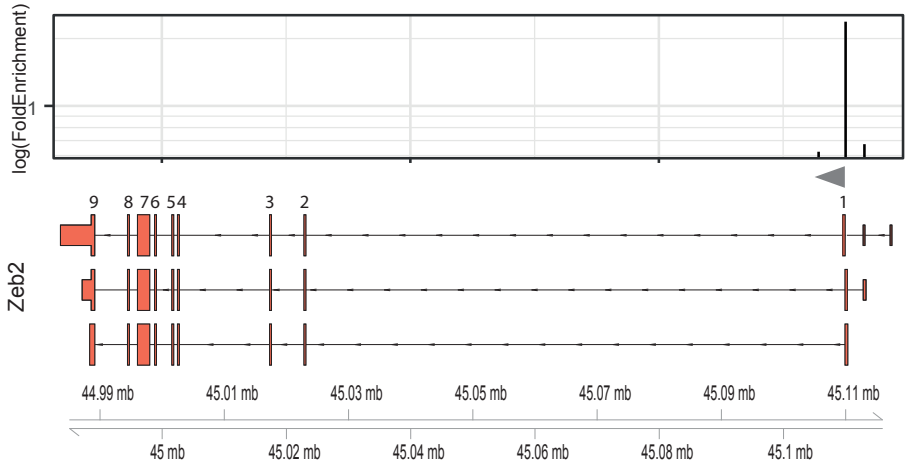

ground-state
EpiLCs
EBs
NPCs
D0
D4
D6
D8
D10
Transfection
Read-Out

B

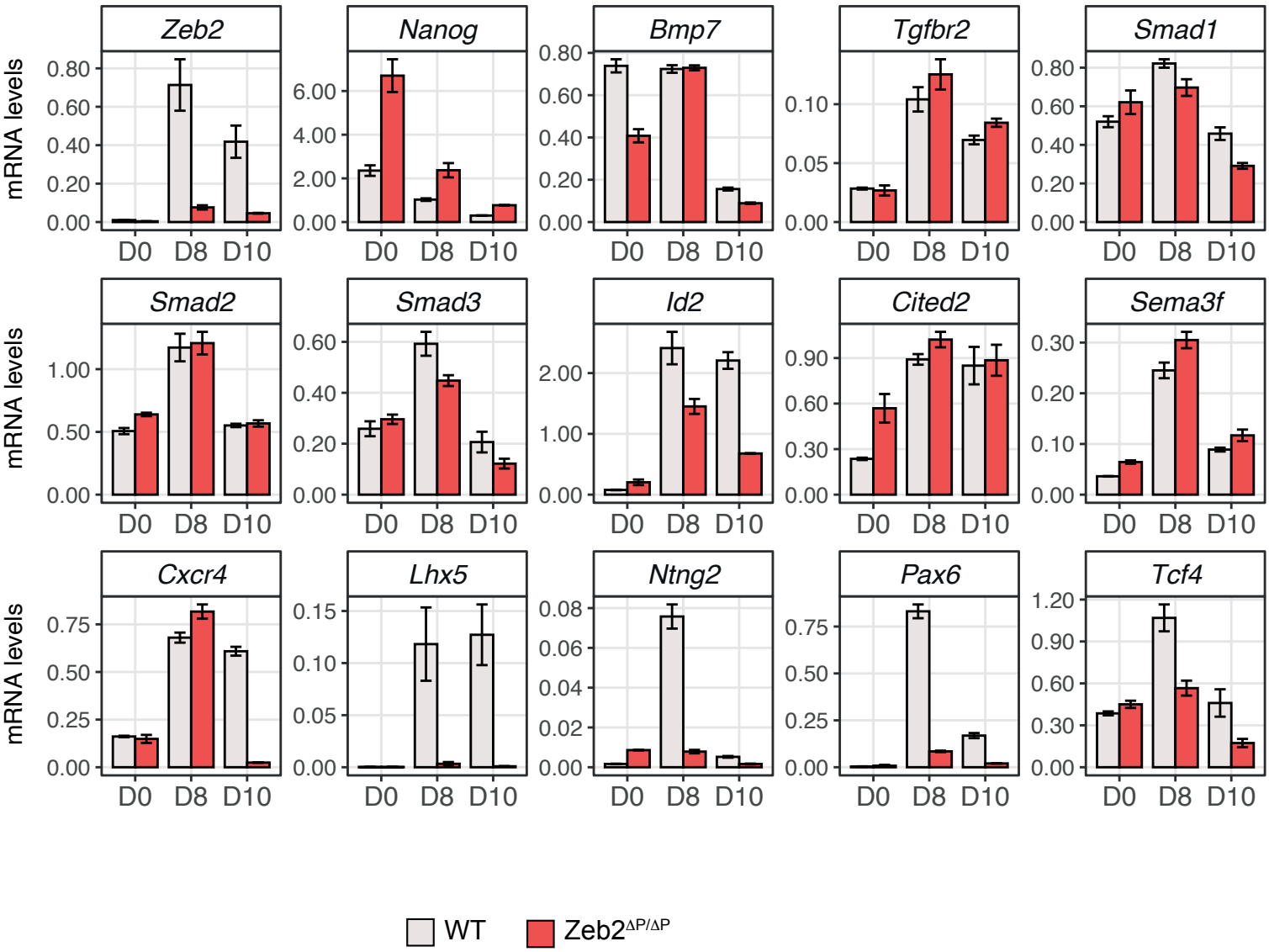

Zeb2
Nanog
Bmp7
Tgfbr2
Smad1
Smad2
Smad3
Id2
Cited2
Sema3f
Cxcr4
Lhx5
Ntng2
Pax6
Tcf4
mRNA levels
D0
D8
D10
D10
shCTRL
shZeb2

A

Figure 5



B



Birkhoff et al 2021