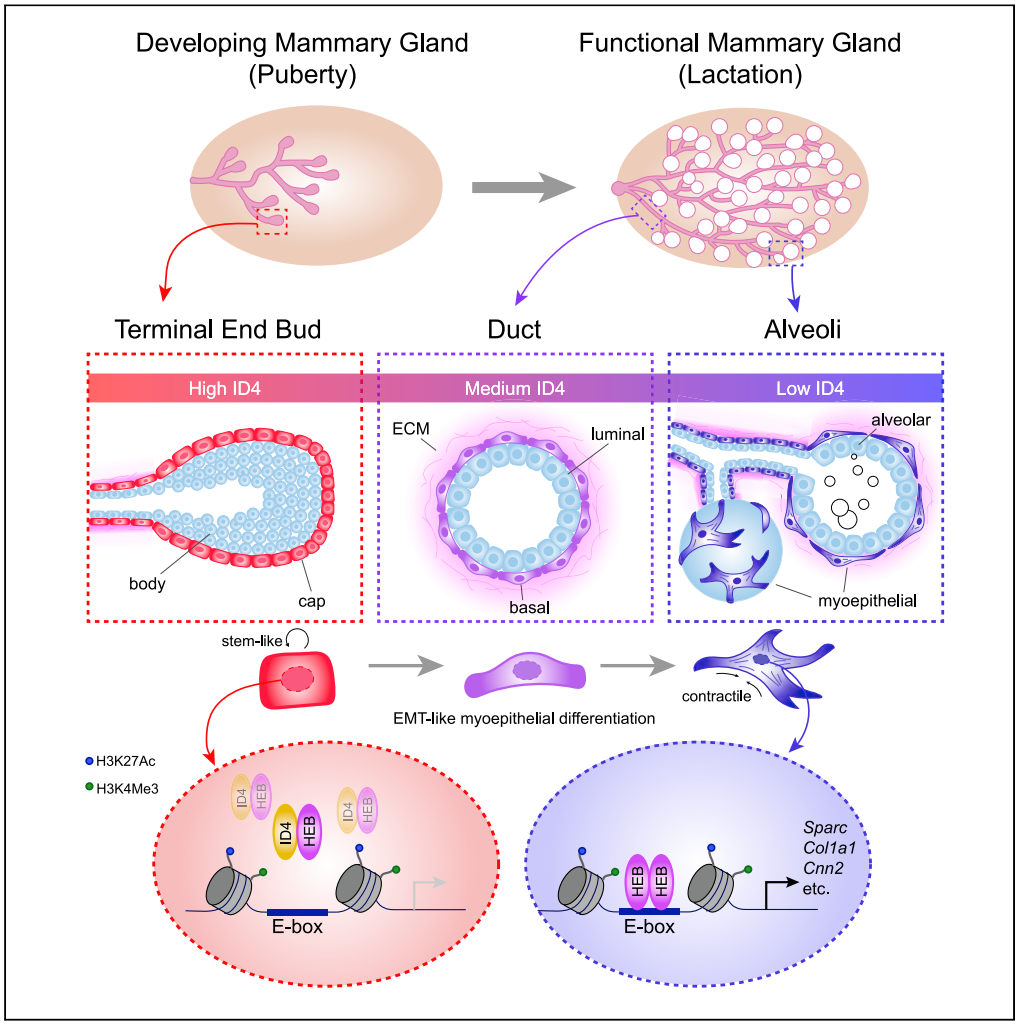
iScience

CellPress
OPEN ACCESS

Article

Inhibitor of Differentiation 4 (ID4) represses mammary myoepithelial differentiation via inhibition of HEB



Holly Holliday, Daniel Roden, Simon Junankar, ..., Jane Visvader, Mark P. Molloy, Alexander Swarbrick

a.swarbrick@garvan.org.au

HIGHLIGHTS

ID4 marks stem-like basal cells and is downregulated during pregnancy and lactation

ID4 interacts with and inhibits the bHLH transcription factor HEB

HEB binds to E-boxes in regulatory elements of ID4-regulated genes

ID4 represses functional myoepithelial genes involved in contraction, EMT, and ECM

Holliday et al., iScience 24, 102072
February 19, 2021
https://doi.org/10.1016/j.isci.2021.102072

iScience

CellPress
OPEN ACCESS

Article

Inhibitor of Differentiation 4 (ID4) represses mammary myoepithelial differentiation via inhibition of HEB

Holly Holliday,[1,2] Daniel Roden,[1,2] Simon Junankar,[1,2] Sunny Z. Wu,[1,2] Laura A. Baker,[1,2] Christoph Krisp,[3,4] Chia-Ling Chan,[1] Andrea McFarland,[1] Joanna N. Skhinas,[1] Thomas R. Cox,[1,2] Bhupinder Pal,[5,7] Nicholas D. Huntington,[8] Christopher J. Ormandy,[1,2] Jason S. Carroll,[9] Jane Visvader,[5,6] Mark P. Molloy,[3] and Alexander Swarbrick[1,2,10,*]

SUMMARY

Inhibitor of differentiation (ID) proteins dimerize with basic HLH (bHLH) transcription factors, repressing transcription of lineage-specification genes across diverse cellular lineages. ID4 is a key regulator of mammary stem cells; however, the mechanism by which it achieves this is unclear. Here, we show that ID4 has a cell autonomous role in preventing myoepithelial differentiation of basal cells in mammary organoids and *in vivo*. ID4 positively regulates proliferative genes and negatively regulates genes involved in myoepithelial function. Mass spectrometry reveals that ID4 interacts with the bHLH protein HEB, which binds to E-box motifs in regulatory elements of basal developmental genes involved in extracellular matrix and the contractile cytoskeleton. We conclude that high ID4 expression in mammary basal stem cells antagonizes HEB transcriptional activity, preventing myoepithelial differentiation and allowing for appropriate tissue morphogenesis. Downregulation of ID4 during pregnancy modulates gene regulated by HEB, promoting specialization of basal cells into myoepithelial cells.

INTRODUCTION

The mammary gland undergoes tissue remodeling throughout life. During murine pubertal development, terminal end buds (TEBs) located at the tips of the ducts invade into the surrounding stromal fat pad (Williams and Daniel, 1983; Macias and Hinck, 2012). This process is driven by collective migration and rapid proliferation of outer cap cells, which surround multiple layers of inner body cells (Williams and Daniel, 1983). As the ducts elongate, the cap cells differentiate into the basal cell layer, and the body cells adjacent to the basal cells give rise to the luminal cell layer (Williams and Daniel, 1983), whereas the innermost body cells undergo apoptosis to sculpt the bilayered ductal tree present in the adult gland (Paine and Lewis, 2017). During pregnancy the gland undergoes alveolar morphogenesis in preparation for production and secretion of milk at lactation. The luminal cells differentiate into milk-producing alveolar cells, and the milk is ejected from the gland by the contractile action of specialized myoepithelial cells, smooth muscle-like epithelial cells that differentiate from basal cells (Macias and Hinck, 2012). Here, "basal" refers to the basal lineage, encompassing cap cells, duct basal cells, and myoepithelial cells.

The basal compartment contains bipotent mammary stem cells (MaSCs), giving rise to basal and luminal lineages upon transplantation (Shackleton et al., 2006; Stingl et al., 2006). Lineage tracing studies have identified both bipotent (Rios et al., 2014; Wang et al., 2015) and unipotent myoepithelial-restricted (Van Keymeulen et al., 2011; van Amerongen et al., 2012; Prater et al., 2014; Wuidart et al., 2016, 2018; Davis et al., 2016; Scheele et al., 2017; Lilja et al., 2018; Lloyd-Lewis et al., 2018) stem cells in the basal compartment under physiological conditions.

Lineage-specifying transcription factors are responsible for directing luminal and myoepithelial differentiation and also for maintaining the self-renewal capacity of uncommitted stem cells upstream in the mammary epithelial hierarchy. Transcriptomic profiling of sorted epithelial subpopulations has identified lineage specifying transcription factors that regulate each step of luminal-alveolar differentiation (Carr et al.,

[1]The Kinghorn Cancer Centre, Garvan Institute of Medical Research, Sydney, NSW 2010, Australia

[2]St Vincent's Clinical School, Faculty of Medicine, UNSW Sydney, Sydney, NSW 2010, Australia

[3]Australian Proteome Analysis Facility, Macquarie University, Sydney, NSW 2109, Australia

[4]Institute of Clinical Chemistry and Laboratory Medicine, Mass Spectrometric Proteomics, University Medical Center Hamburg-Eppendorf, Hamburg 20251, Germany

[5]ACRF Cancer Biology and Stem Cells Division, The Walter and Eliza Hall Institute of Medical Research, Parkville, VIC 3052, Australia

[6]Department of Medical Biology, The University of Melbourne, Parkville, VIC 3010, Australia

[7]Olivia Newton-John Cancer Research Institute and School of Cancer Medicine, La Trobe University, Heidelberg, VIC 3084, Australia

[8]Biomedicine Discovery Institute, Department of Biochemistry and Molecular Biology, Monash University, Clayton, VIC 3168, Australia

[9]Cancer Research UK Cambridge Institute, University of Cambridge, Robinson Way, Cambridge CB2 0RE, UK

[10]Lead contact

*Correspondence: a.swarbrick@garvan.org.au

https://doi.org/10.1016/j.isci.2021.102072

2012; Kouros-Mehr et al., 2006; Asselin-Labat et al., 2007; Buchwalter et al., 2013; Bouras et al., 2008; Liu et al., 2008; Oakes et al., 2008; Chakrabarti et al., 2012; Yamaji et al., 2009). However, due to lack of specific cell markers that can resolve stem and myoepithelial populations, it has been challenging to dissect molecular regulators of basal differentiation. Although a number of basal-specific transcription factors have been identified, such as P63, SLUG, SOX9, SRF, and MRTFA (Mills et al., 1999; Yang et al., 1999; Guo et al., 2012; Li et al., 2006; Sun et al., 2006), their role in the basal compartment and myoepithelial specialization is poorly understood. ID proteins (ID1-4) are helix-loop-helix (HLH) transcriptional regulators that lack a DNA-binding domain. They function by dimerizing with basic HLH (bHLH) transcription factors and preventing them from binding to E-box DNA motifs and regulating transcription (Benezra et al., 1990). E-box motifs are found in regulatory regions of genes involved in lineage specification and as such ID proteins and bHLH transcription factors are critical regulators of stemness and differentiation across diverse cellular lineages (Massari and Murre, 2000). The expression of ID4 in mouse and human mammary epithelium is exclusive to the basal population (Lim et al., 2010). We and others have demonstrated that ID4 is a key regulator of mammary stem cells, required for ductal elongation during puberty (Best et al., 2014; Dong et al., 2011; Junankar et al., 2015). ID4 was also shown to have a role in blocking luminal differentiation (Best et al., 2014; Junankar et al., 2015). The precise molecular mechanisms by which ID4 functions in the mammary gland, including its full repertoire of transcriptional targets and interacting partners, have yet to be determined. Here we show that ID4 marks basal stem cells and demonstrate that ID4 also inhibits myoepithelial differentiation. Moreover, using unbiased interaction proteomics, we identify the bHLH transcription factor HEB as a factor in the mammary differentiation hierarchy. By mapping the genome-wide binding sites of HEB we show that it directly binds to regulatory elements of ID4 target genes involved in myoepithelial functions such as contraction and extracellular matrix (ECM) synthesis.

RESULTS

Loss of ID4 causes upregulation of myoepithelial genes in basal cells

In order to determine the genes regulated by ID4 in mammary epithelial cells, we FACS-enriched basal, luminal progenitor, and mature luminal cells from wild-type (WT) and ID4 knockout (KO) mice (Yun et al., 2004) and performed RNA-sequencing (RNA-seq) (Figure 1A). No significant differences were observed in the proportion of mammary epithelial subpopulations between WT and KO mice (Figure S1A). One hundred four genes were significantly (FDR<0.05) upregulated and 145 genes downregulated in ID4 KO basal cells. In contrast, only 1 gene was differentially expressed in each of the two luminal subpopulations (Figures 1B and 1C and Table S1), suggesting that ID4 predominantly regulates gene expression within basal cells *in vivo*.

In line with the known role of ID4 in promoting proliferation of mammary epithelial cells (Junankar et al., 2015; Dong et al., 2011), the genes downregulated in ID4 KO basal cells were enriched for pathways involved in cell growth such as translation and metabolism (Figures 1D, S1B, and S1C). Conversely, the genes upregulated in KO basal cells were enriched for pathways related to basal cells and smooth muscle function such as contraction, epithelial-mesenchymal transition (EMT), and serum response factor (SRF) targets (Figures 1E and S1B). These pathways shared common genes as indicated by the connecting edges in the enrichment map network (Figure 1D). SRF target genes are of relevance, as SRF is a master regulator of cytoskeletal contraction and is one of the only transcription factors implicated in myoepithelial differentiation (Miano et al., 2007; Li et al., 2006; Sun et al., 2006). Taken together, loss of ID4 causes basal cells to adopt a more differentiated myoepithelial and mesenchymal gene expression program, implicating ID4 in the repression of basal cell specialization.

ID4 expression decreases upon terminal myoepithelial differentiation of basal cells

ID4 is known to be heterogeneously expressed in basal cells, and ID4-positive cells have enhanced mammary reconstitution activity (Junankar et al., 2015). To further characterize the phenotype of ID4-positive basal cells, we used an ID4-GFP reporter mouse in which the ID4 promoter drives GFP expression (Best et al., 2014). Basal cells with high expression of the stem cell marker CD49f/ITGA6 and the epithelial marker EPCAM have been shown to be enriched for MaSC activity (Stingl et al., 2006; Prater et al., 2014). ID4-GFP expression was maximal in this $EPCAM^{hi}$ $CD49f^{hi}$ subset (Figure 2A). ID4-GFP expression within the basal gate was binned into three groups—bright (top 10%), intermediate (middle 80%), and dim (bottom 10%)—and the median fluorescent intensity (MFI) of EPCAM and CD49f was analyzed within these groups (Figures 2B and 2C). ID4-bright cells had significantly higher EPCAM and CD49f MFI than ID4-dim and -intermediate cells, suggesting ID4 marks basal stem cells.

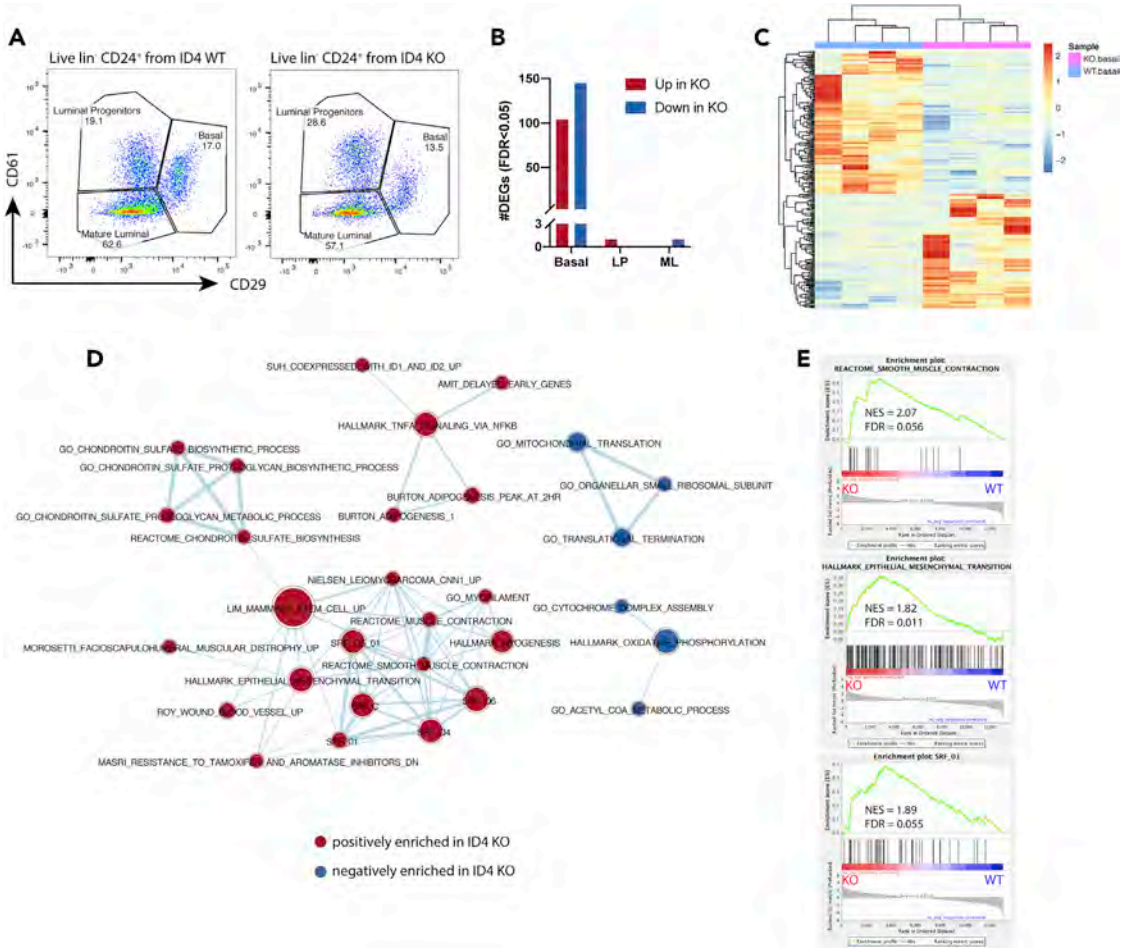


Figure 1. Loss of ID4 results in upregulation of myoepithelial genes in sorted mammary basal cells

(A) Live lineage negative $CD24^{+}CD29^{hi}CD61^{+}$ basal, $CD24^{+}CD29^{lo}CD61^{+}$ luminal progenitor, and $CD24^{+}CD29^{lo}CD61^{-}$ mature luminal cells were isolated by FACS from adult (10–12 weeks) ID4 wild-type (WT) and knockout (KO) mice at estrus for RNA-seq. Representative FACS plots shown from four experiments.

(B) Number of significantly (FDR<0.05) differentially expressed genes (DEGs) upregulated (red) or downregulated (blue) in ID4 KO epithelial subpopulations compared with ID4 WT. LP = luminal progenitor, ML = mature luminal.

(C) Heatmap displaying the significant differentially expressed genes between ID4 WT and ID4 KO basal cells.

(D) Genes were ranked based on the limma t-statistic comparing ID4 WT and KO basal cells, and GSEA was carried out using the C2all, C3TF, C5, and Hallmark gene sets. GSEA results were visualized using Cytoscape EnrichmentMap. Nodes represent gene sets, and edges represent overlap. Gene sets with an FDR<0.25 are shown.

(E) Representative GSEA enrichment plots displaying the profile of the running Enrichment Score (green) and positions of gene set members on the rank ordered list for pathways related to myoepithelial function. Normalized enrichment scores (NES) and FDR are indicated.

See also Figure S1 and Table S1.

During ductal elongation at puberty, the cap cells differentiate at the neck of the TEBs and mature into myoepithelial cells that form the outer basal layer of the ducts (Paine and Lewis, 2017). To locate ID4-high and ID4-low populations in a tissue context and to further investigate the association between ID4 and markers of myoepithelial differentiation, mammary gland sections from pubertal mice, in which TEBs and ducts are both present, were stained for ID4 and the myoepithelial marker alpha-smooth muscle actin (α-SMA). ID4 expression was highest in the nuclei of cap cells at the extremity of the TEBs and expressed at lower levels in basal cells of ducts (Figures 2D and S2A). Conversely, α-SMA expression was higher in ductal basal cells and lower in cap cells. ID4-high cap cells had a compact cuboidal epithelial appearance compared with the more separated elongated morphology of the ID4-low duct cells (Figures 2D and S2A). Quantification of fluorescence demonstrated that ID4 was more highly expressed in cap cells of TEBs than in basal cells of ducts, whereas the opposite was true for α-SMA expression (Figures 2E and 2F). A negative correlation

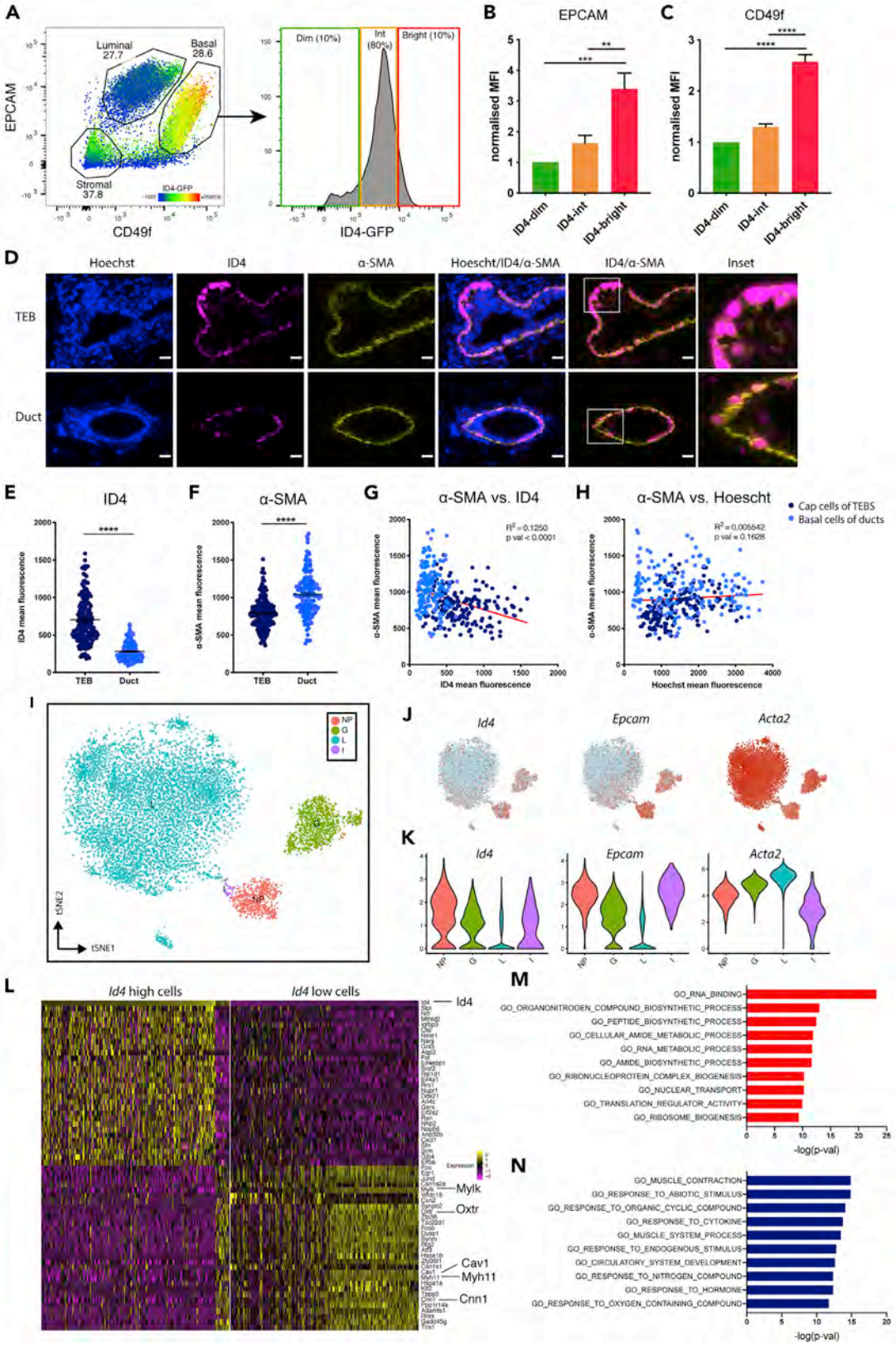

A
Luminal
27.7
Basal
28.6
Stromal
37.8
EPCAM
CD49f
ID4-GFP
Dim (10%)
Int
(80%)
Bright (10%)
ID4-GFP
B
EPCAM
normalised MFI
ID4-dim
ID4-int
ID4-bright
C
CD49f
normalised MFI
ID4-dim
ID4-int
ID4-bright
D
Hoechst
ID4
α-SMA
Hoescht/ID4/α-SMA
ID4/α-SMA
Inset
TEB
Duct
E
ID4
ID4 mean fluorescence
TEB
Duct
F
α-SMA
α-SMA mean fluorescence
TEB
Duct
G
α-SMA vs. ID4
α-SMA mean fluorescence
ID4 mean fluorescence
p val < 0.0001
H
α-SMA vs. Hoescht
α-SMA mean fluorescence
Hoechst mean fluorescence
p val = 0.1628
Cap cells of TEBS
Basal cells of ducts
I
NP
G
L
I
tSNE1
tSNE2
J
Id4
Epcam
Acta2
K
Id4
Epcam
Acta2
L
Id4 high cells
Id4 low cells
Id4
Mylk
Oxtr
Cav1
Myh11
Cnn1
M
GO_RNA_BINDING
GO_ORGANONITROGEN_COMPOUND_BIOSYNTHETIC_PROCESS
GO_PEPTIDE_BIOSYNTHETIC_PROCESS
GO_CELLULAR_AMIDE_METABOLIC_PROCESS
GO_RNA_METABOLIC_PROCESS
GO_AMIDE_BIOSYNTHETIC_PROCESS
GO_RIBONUCLEOPROTEIN_COMPLEX_BIOGENESIS
GO_NUCLEAR_TRANSPORT
GO_TRANSLATION_REGULATOR_ACTIVITY
GO_RIBOSOME_BIOGENESIS
-log(p-val)
N
GO_MUSCLE_CONTRACTION
GO_RESPONSE_TO_ABIOTIC_STIMULUS
GO_RESPONSE_TO_ORGANIC_CYCLIC_COMPOUND
GO_RESPONSE_TO_CYTOKINE
GO_MUSCLE_SYSTEM_PROCESS
GO_RESPONSE_TO_ENDOGENOUS_STIMULUS
GO_CIRCULATORY_SYSTEM_DEVELOPMENT
GO_RESPONSE_TO_NITROGEN_COMPOUND
GO_RESPONSE_TO_HORMONE
GO_RESPONSE_TO_OXYGEN_CONTAINING_COMPOUND
-log(p-val)

Figure 2. ID4 expression decreases in terminally differentiated myoepithelial cells

(A–C) (A) FACS analysis of EPCAM and CD49f in live lineage negative mammary cells from adult (10–14 weeks) Id4floxGFP reporter mice. ID4-GFP expression is indicated in the heatmap scale. Representative plots from five experiments shown. Basal cells were binned into three groups based on ID4-GFP expression, ID4-bright (red), ID4-intermediate (orange), and ID4-dim (green), and the median fluorescence intensity (MFI) of EPCAM (B) and CD49f (C) were compared between the three gates. MFI expressed as a fold change relative to the ID4-low basal cells. Ordinary one-way ANOVA test was used to test significance. n = 5. Error bars represent SEM. **p < 0.01, ***p < 0.001, ****p < 0.0001.

(D–H) (D) Representative co-immunofluorescent staining of ID4 and α-SMA in TEB and duct from a pubertal (6 weeks) mammary gland. Scale bar: 20 μm. Comparison of ID4 (E) and α-SMA (F) mean fluorescence between individual cap cells (dark blue) and basal duct cells (light blue). Unpaired two-tailed students t test. Error bars represent SEM. ****p < 0.0001. Correlation between α-SMA and ID4 (G) and Hoescht (H) mean fluorescence in individual cap cells (dark blue) and basal duct cells (light blue). R^2 and p values are displayed. Data are pooled from nine mice. Approximately 20 TEB cap cells and 20 ductal basal cells were analyzed per mouse.

(I–K) (I) tSNE plot of 9663 *Krt5+/Krt14* + basal cells from (Bach et al., 2017). Two mice were analyzed per developmental stage. NP = Nulliparous (8 weeks), G = Gestation (Day 14.5), L = Lactation (Day 6), I = Involution (Day 11). Feature plots (J) and Violin plots (K) displaying expression of *Id4*, *Epcam*, and *Acta2* in single cells in the different developmental stages.

(L–N) (L) Heatmap displaying top and bottom 30 differentially expressed genes between the top and bottom 200 *Id4* high and *Id4* low basal cells across all stages. Top 10 GO terms enriched in the top 50 genes upregulated in *Id4* high (M) and low (N) basal cells.

See also Figure S2 and Table S2.

between α-SMA and ID4 was observed, with clear separation between cap (dark blue Figure 2G) and duct basal cells (light blue Figure 2G). As a negative control, α-SMA fluorescence was compared with nuclear stain Hoechst and no correlation or separation based on region was observed (Figure 2H). Thus, based on marker expression, morphology, and spatial localization, ID4 expression is high in epithelial-like cap cells and is lower in more differentiated myoepithelial cells.

Terminal differentiation of basal cells into contractile myoepithelial cells occurs during lactation. We interrogated a published single cell RNA-seq (scRNA-seq) dataset (Bach et al., 2017) to examine *Id4* expression dynamics over postnatal murine mammary gland development. In this study, individual EPCAM+ mammary epithelial cells from four developmental stages—nulliparous (8 weeks), gestation (day 14.5), lactation (day 6), and involution (day 11)—were captured and profiled. We limited our analysis to basal cell clusters (9,663 cells), defined by expression of both *Krt5* and *Krt14*. Basal cells broadly clustered by developmental time point (Figure 2I). Increased differentiation of myoepithelial cells appears to proceed from nulliparous to gestation to lactation, associated with decreased expression of epithelial marker *Epcam* and increased expression of *Acta2* (encoding α-SMA) (Figures 2J and 2K), consistent with gradual acquisition of a smooth muscle phenotype and loss of adherent epithelial features (Deugnier et al., 1995). Like *Epcam*, *Id4* expression was highest in basal cells from nulliparous mice and decreased in basal cells of pregnant and lactating mice (Figures 2J and 2K). This result was validated on the protein level by immunohistochemical staining for ID4 on mammary gland sections at different developmental time points (Figure S2B). Using the Monocle 2 package (Trapnell et al., 2014), we performed pseudo-temporal ordering of all basal cells to form a myoepithelial differentiation trajectory (Figure S2C). The nulliparous and involution cells clustered together in pseudo-time space in the least differentiated part of the trajectory. Basal cells from gestating mice were dispersed between the nulliparous and lactation stages, whereas basal cells at lactation were the most differentiated. *Id4* expression decreased, whereas several myoepithelial markers (*Acta2*, *Cnn1*, *Mylk*, *Myh11*, and *Oxtr*) increased over pseudotime (Figure S2D).

We sought to identify transcriptional signatures associated with high and low *Id4*-expressing cells in the mammary basal epithelium across all developmental time points (Figure 2L and Table S2). Basal cells with high *Id4* expression were enriched for genes involved in RNA binding, metabolic processes, translation, and ribosome biogenesis (Figure 2M). Conversely, genes enriched in the *Id4*-low basal cells were involved in muscle contraction and response to cytokine and hormone stimuli and circulatory system development (Figure 2N). Myoepithelial genes such as *Oxtr*, encoding the oxytocin receptor, and contractile genes such as *Mylk*, *Cnn1*, *Cav1*, and *Myh11* were among the top differentially expressed genes in the cells with low *Id4* expression (Figure 2L and Table S2). Thus, ID4 downregulation during pregnancy and lactation is associated with terminal differentiation into functionally mature contractile myoepithelial cells consistent with its role as a basal stem cell marker.

Loss of ID4 results in myoepithelial differentiation of mammary organoids

To functionally validate the role of ID4 in suppressing the myoepithelial differentiation of basal cells we generated a primary basal cell organoid model (Figure 3A). The organoid model system complements and expands upon the findings from the KO mouse, as the acute consequence of ID4 loss on basal cell

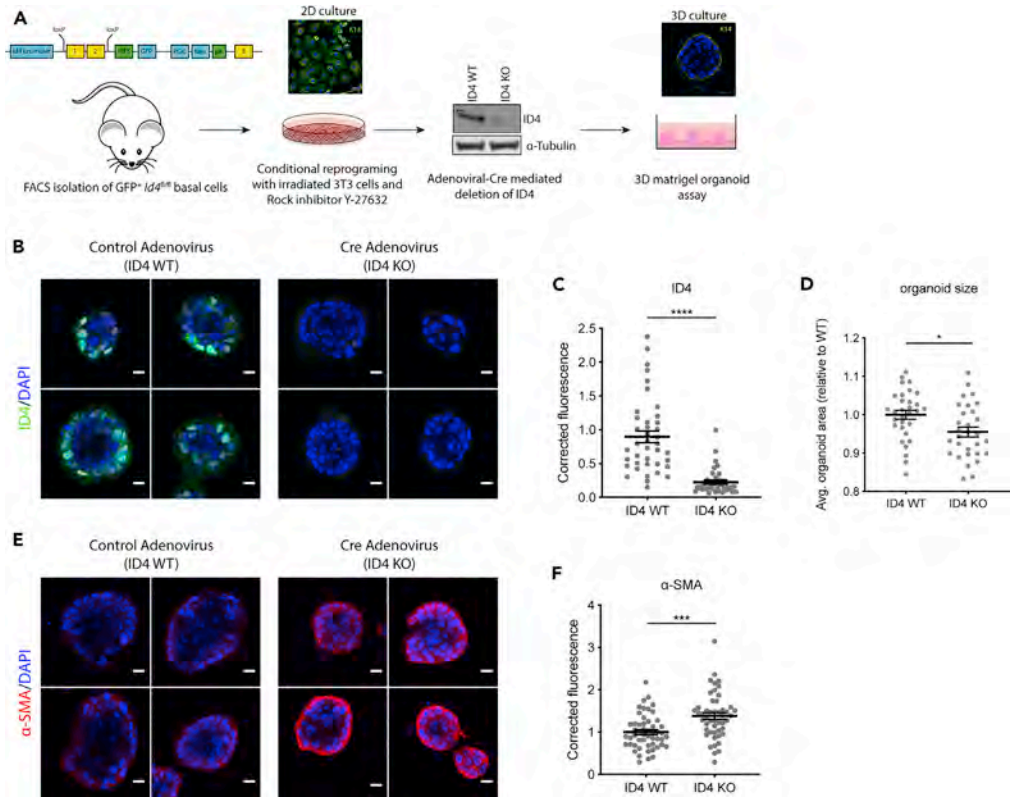


Figure 3. ID4 inhibits myoepithelial differentiation of organoids

(A) Schematic diagram of 3D Matrigel organoid assay. ID4-GFP + basal cells were FACS purified from adult (10–11 weeks) ID4-GFP reporter mice. Exon 1 and 2 of *Id4* are floxed and a GFP reporter cassette introduced. Basal cells are reprogrammed in culture using ROCK inhibitor Y-27632 and irradiated NIH-3T3 feeder cells. Adenoviral-Cre is used to knock out ID4 as shown by western blotting. Single cells are then seeded on top of a Matrigel plug and grown for 6 days followed by immunofluorescent staining and quantification. (B, C, and E–F) Organoids grown from conditionally reprogrammed basal cells were treated with control GFP adenovirus (ID4 WT) or with Cre adenovirus (ID4 KO). Organoids were stained for ID4 (B) and α-SMA (E). Scale bar: 10 μm. Fluorescence was quantified for ID4 (C) and α-SMA (F) in approximately 10 organoids per experiment. n = 4.

(D) The average organoid size was determined per chamber. Unpaired two-tailed students t test. Error bars represent SEM. *p < 0.05, ***p < 0.001, ****p < 0.0001.

See also Figure S3.

phenotype can be determined. Furthermore, organoids are less complex cellular systems than tissue, thus cell-autonomous effects can be isolated more precisely. ID4-positive basal cells from mice in which exons 1 and 2 of *Id4* are floxed ($Id4^{fl/fl}$) (Best et al., 2014) were isolated using cell sorting. To overcome culture-induced senescence, basal cells were conditionally reprogrammed into a proliferative stem/progenitor state using an established protocol utilizing irradiated 3T3 fibroblast feeders and Rho Kinase (ROCK) inhibition (Liu et al., 2012b; Prater et al., 2014) (Figure 3A). Conditionally reprogrammed basal cells adopted an epithelial cobblestone morphology and expressed high levels of ID4 (Figures S3A and S3B). Cells cultured in the absence of ROCK inhibitor or feeders adopted a flattened differentiated/senescent cell morphology and had reduced ID4 expression (Figures S3A and S3B). Reprogrammed cells maintained basal marker expression of P63 and KRT14 (Figure S3C) and could be grown as 3D organoids with basal marker KRT14 on the outer cell layer and luminal marker KRT8 on the inner cells of the organoids (Figure S3D).

ID4 expression was limited to the outer basal cells (Figure 3B), recapitulating expression of ID4 in cap cells of TEBs (Figure 2D). To test whether ID4 regulates differentiation of basal cells we deleted ID4 with adenoviral-Cre and compared these to cells treated with control adenoviral-GFP. ID4 protein was markedly downregulated in organoids infected with Cre adenovirus confirming successful gene deletion (Figures 3B and 3C). Loss of ID4 resulted in slightly smaller organoids with increased α-SMA fluorescent intensity

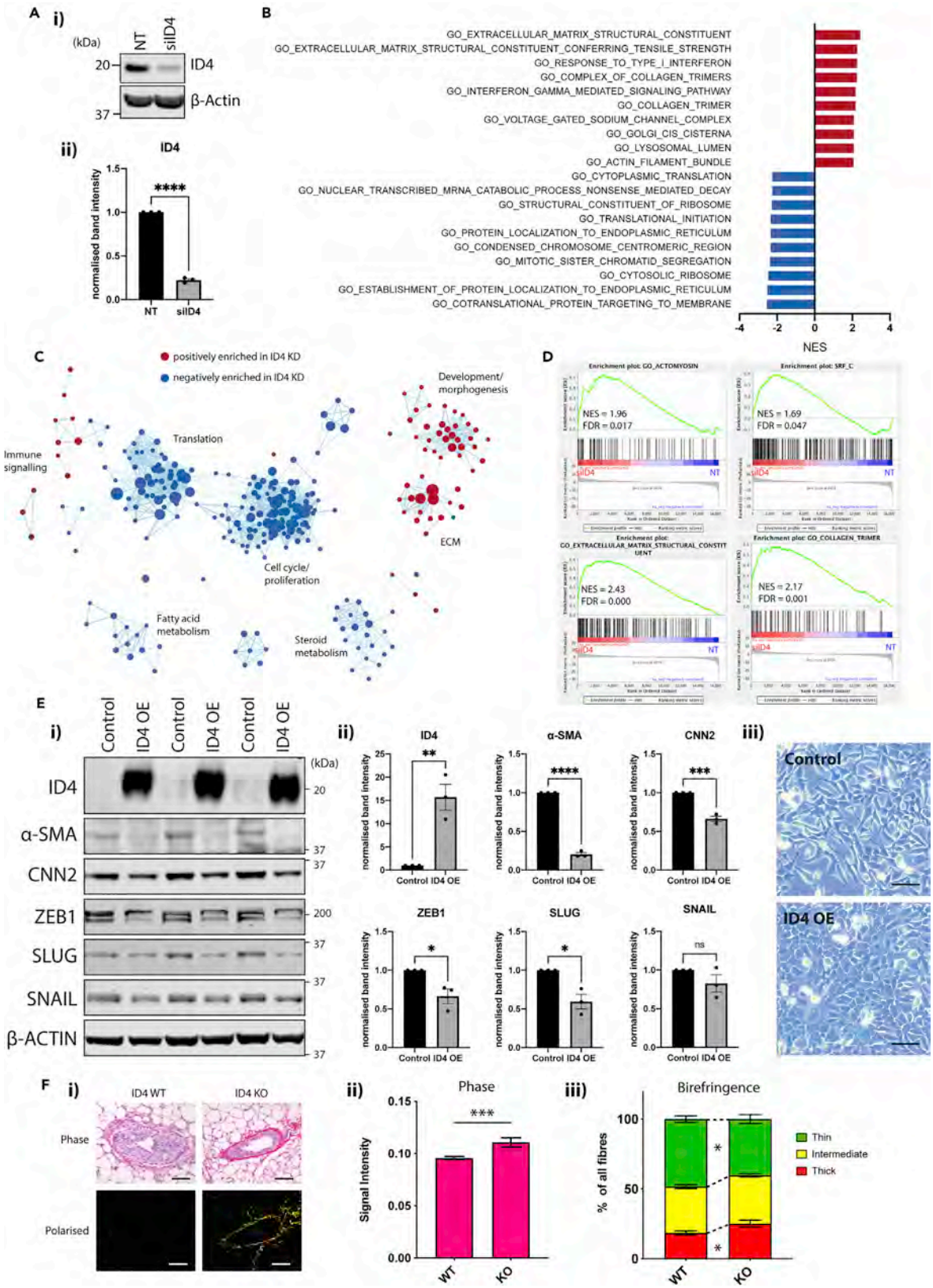

A
i)
(kDa)
NT
siID4
20
ID4
β-Actin
37
ii)
ID4
normalised band intensity
****
NT
siID4
B
GO_EXTRACELLULAR_MATRIX_STRUCTURAL_CONSTITUENT
GO_EXTRACELLULAR_MATRIX_STRUCTURAL_CONSTITUENT_CONFERRING_TENSILE_STRENGTH
GO_RESPONSE_TO_TYPE_I_INTERFERON
GO_COMPLEX_OF_COLLAGEN_TRIMERS
GO_INTERFERON_GAMMA_MEDIATED_SIGNALING_PATHWAY
GO_COLLAGEN_TRIMER
GO_VOLTAGE_GATED_SODIUM_CHANNEL_COMPLEX
GO_GOLGI_CIS_CISTERNA
GO_LYSOSOMAL_LUMEN
GO_ACTIN_FILAMENT_BUNDLE
GO_CYTOPLASMIC_TRANSLATION
GO_NUCLEAR_TRANSCRIBED_MRNA_CATABOLIC_PROCESS_NONSENSE_MEDIATED_DECAY
GO_STRUCTURAL_CONSTITUENT_OF_RIBOSOME
GO_TRANSLATIONAL_INITIATION
GO_PROTEIN_LOCALIZATION_TO_ENDOPLASMIC_RETICULUM
GO_CONDENSED_CHROMOSOME_CENTROMERIC_REGION
GO_MITOTIC_SISTER_CHROMATID_SEGREGATION
GO_CYTOSOLIC_RIBOSOME
GO_ESTABLISHMENT_OF_PROTEIN_LOCALIZATION_TO_ENDOPLASMIC_RETICULUM
GO_COTRANSLATIONAL_PROTEIN_TARGETING_TO_MEMBRANE
NES
C
positively enriched in ID4 KD
negatively enriched in ID4 KD
Development/ morphogenesis
Immune signalling
Translation
ECM
Cell cycle/ proliferation
Fatty acid metabolism
Steroid metabolism
D
Enrichment plot: GO_ACTOMYOSIN
NES = 1.96
FDR = 0.017
Enrichment plot: SRF_C
NES = 1.69
FDR = 0.047
Enrichment plot: GO_EXTRACELLULAR_MATRIX_STRUCTURAL_CONSTITUENT
NES = 2.43
FDR = 0.000
Enrichment plot: GO_COLLAGEN_TRIMER
NES = 2.17
FDR = 0.001
siID4
NT
E
i)
Control
ID4 OE
(kDa)
ID4
α-SMA
CNN2
ZEB1
SLUG
SNAIL
β-ACTIN
ii)
ID4
α-SMA
CNN2
ZEB1
SLUG
SNAIL
normalised band intensity
Control ID4 OE
iii)
Control
ID4 OE
F
i)
ID4 WT
ID4 KO
Phase
Polarised
ii)
Phase
Signal Intensity
***
WT
KO
iii)
Birefringence
% of all fibres
Thin
Intermediate
Thick

Figure 4. ID4 negatively regulates EMT and ECM production in mammary epithelial cells

(A) **(i)** Western blot analysis of ID4 expression in Comma-Dβ cells treated with non-targeting (NT) or ID4-targeting siRNA. Representative results from three western blots shown. **(ii)** Densitometry quantification of ID4 bands. Band intensity was normalized to β-Actin and expressed as fold change relative to NT control. N = 3. Unpaired two-tailed students t test.

(B) Genes were ranked based on the limma t-statistic comparing NT and siID4 cells, and GSEA was carried out using Gene Ontology (GO) gene sets. The top 10 positively (red) and negatively (blue) enriched pathways are displayed.

(C) GO GSEA results were visualized using Cytoscape EnrichmentMap. Nodes represent gene sets and edges represent overlap. Gene sets with an FDR<0.1 are shown.

(D) Representative GSEA enrichment plots displaying the profile of the running Enrichment Score (green) and positions of gene set members on the rank ordered list. NES and FDR are indicated.

(E) **(i)** Western blot analysis of ID4, α-SMA, CNN2, ZEB1, SLUG, and SNAIL in Comma-Dβ cells overexpressing ID4 (ID4 OE). **(ii)** Densitometry quantification of bands normalized to β-Actin as a fold-change relative to control cells. N = 3. Unpaired two-tailed students t test. Error bars represent SEM. **(iii)** Morphology of Comma-Dβ cells overexpressing ID4. Scale bar: 100 μm.

(F) **(i)** Collagen fibers were visualized by picrosirius red staining of TEBs from 6-week-old ID4 WT and KO mice. Scale bar: 50 μm. Total collagen staining **(ii)** and birefringence signal **(iii)** were quantified from approximately three TEBs from each mammary gland section. N = 9 for WT and N = 6 for KO mice. Unpaired two-tailed students t test. *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001. Error bars represent SEM.

See also Figure S4 and Table S3.

(Figures 3D–3F). This finding demonstrates that as well as marking undifferentiated basal cells, ID4 has a cell autonomous role in impeding maturation of basal cells into myoepithelial cells.

ID4 inhibits expression of contractile and ECM genes

Given the negative association between ID4 and the differentiated myoepithelial phenotype, we sought to determine direct ID4 target genes by performing RNA-seq following siRNA-mediated ID4 depletion in the spontaneously immortalized mouse mammary epithelial cell line Comma-Dβ (Danielson et al., 1984). This normal-like cell line expresses basal markers and is commonly used as a model of mammary stem/progenitor cells, as they retain the capacity of multi-lineage differentiation when transplanted into mammary fat pads (Deugnier et al., 2006; Idoux-Gillet et al., 2018; Danielson et al., 1984; Junankar et al., 2015; Best et al., 2014). Western blotting confirmed 70%–80% knockdown (KD) of ID4 protein 48 h after siRNA transfection (Figure 4A). RNA-seq analysis of ID4 KD cells resulted in 471 and 421 (FDR<0.05) down- and upregulated genes, respectively, compared with the non-targeting siRNA control (Table S3).

Downregulated genes were predominantly involved in cell proliferation and growth pathways (Figures 4B and 4C; blue), including hallmark gene sets such as E2F targets, MYC targets, and G2M checkpoint (Figures S4A and S4B). Driving the enrichment of these gene sets were several key cell cycle genes such as *Mki67*, *Cdk2*, *Cdk6*, and *Cdk17* (Table S3). This result complements the loss of cell growth gene expression programs in ID4 KO basal cells *in vivo* (Figure 1D).

Genes acutely upregulated upon ID4 depletion were involved in development, morphogenesis, ECM remodeling, and immune signaling (Figures 4B and 4C; red). Consistent with ID4 repressing myoepithelial specialization, loss of ID4 resulted in upregulation of SRF targets, actomyosin cytoskeleton, EMT, and myogenesis gene signatures (Figures 4D, S4A, and S4B). Several contractile genes were increased in ID4 depleted cells including *Cnn1*, *Cnn2*, *Tagln*, *Lmod1*, and *Acta2* (FDR = 0.06) (Table S3), many of which were inversely correlated with *Id4* expression in the scRNA-seq analysis (Figure 2L and Table S2).

To independently validate the transcriptomic results implicating ID4 in repression of contractile EMT genes, we overexpressed ID4 in Comma-Dβ cells and performed western blotting analysis for several EMT proteins. Overexpression of ID4 resulted in downregulation of α-SMA (Figure 4E i–ii), consistent with the upregulation of this marker in the primary organoid culture upon loss of ID4 (Figures 3E and 3F). CNN2, another smooth muscle contractile protein, as well as classical EMT markers ZEB1 and SLUG, were also suppressed by ID4 (Figure 4E i-ii). Finally, morphological inspection of ID4 overexpressing cells revealed a cobblestone epithelial appearance compared with the more mesenchymal control cells. Together, these results confirm that ID4 blocks expression of genes involved in myoepithelial contraction and EMT.

Several ECM genes encoding collagens (e.g. *Col1a1*, *Col1a2*, and *Col5a1*), basement membrane laminins (e.g. *Lamc1* and *Lama4*), and matricellular proteins (e.g. *Sparc*) were also upregulated upon ID4 depletion (Figures 4C and 4D and Table S3). The ECM provides physical support to the mammary gland and is a source of biochemical signals that coordinate morphogenesis (Muschler and Streuli, 2010). Changes in

ECM gene expression are associated with EMT and cellular contractility (Kiemer et al., 2001; Liu et al., 2012a). To functionally validate the role of ID4 in regulating ECM proteins we examined whether ID4 represses ECM deposition *in vivo* using picrosirius red staining of mammary glands from ID4 WT and KO mice to visualize fibrillar collagen (Figure 4F). ID4 KO TEBs in pubertal mammary glands were surrounded by a thickened collagen-dense ECM when compared with ID4 WT TEBs (Figure 4F i–ii), indicating that ID4 normally restrains collagen expression by cap cells during puberty. In addition to total abundance, the thickness of bundled collagen fibers can be further distinguished using polarized light. Analysis of birefringence signal revealed an increase in thick fibers and a decrease in thin fibers in the ECM surrounding ID4 KO TEBs (Figure 4F iii), signifying a redistribution of collagen composition, as well as an overall increase in collagen abundance, in the absence of ID4.

Hence, ID4 positively regulates proliferative genes and negatively regulates genes involved in myoepithelial functions such as contraction and ECM synthesis in mammary epithelial cells. These ID4-regulated functions are likely to be critical for morphogenesis of the ductal tree during pubertal development.

ID4 interacts with E-proteins in mammary epithelial cells

As ID4 lacks a DNA-binding domain, it influences transcription through its interaction with other DNA-binding proteins. We used Rapid Immunoprecipitation Mass spectrometry of Endogenous proteins (RIME) to discover the binding partners of ID4 in Comma-Dβ cells (Mohammed et al., 2013). We identified 48 proteins that were significantly ($p < 0.05$) more abundant in the ID4 IPs compared with the IgG-negative control IPs in three independent RIME experiments (Figure 5A and Table S4). ID4 was consistently identified in all replicates and was the top hit (Table S4), verifying the validity of the technique. Among the putative ID4-binding partners were many DNA- and RNA-interacting proteins (Figure 5A).

The E-proteins E2A and HEB were identified as ID4 interactors in our RIME analysis. There are three members of the E-protein family (E2A, HEB, and ITF-2), that dimerize with other E-proteins or tissue-specific bHLH transcription factors (e.g. MyoD and NeuroD) to regulate expression of lineage commitment genes (Wang and Baker, 2015). E2A and HEB expression was confirmed in the mammary gland epithelium by IHC (Figure S5A). E2A has been implicated in branching morphogenesis of mammary organoids (Lee et al., 2011); however, there are no previous studies implicating HEB in mammary gland development. Given this, and that E-proteins are the canonical binding partners of ID proteins in other lineages, we chose to pursue the ID4-HEB interaction further.

Reciprocal co-immunoprecipitation followed by western blotting (co-IP WB) experiments were performed to validate the ID4-E-protein interactions in the same Comma-Dβ cell line, as well as in normal human mammary epithelial cell lines PMC42 and MCF10A. IP of ID4 resulted in co-IP of E2A and HEB, and correspondingly, IP of E2A and HEB resulted in co-IP of ID4 (Figures 5B and S5B). E2A and HEB did not form heterodimers with each other in Comma-Dβ cells (Figure 5B), implying that E-proteins function either as homodimers or as heterodimers with other bHLH proteins in this context. To identify mammary-specific bHLH factors binding to HEB, we performed another RIME experiment by immunoprecipitating HEB protein. HEB (HTF4_MOUSE in Table S4) was identified as the top hit, and ID4 was also identified (Figure S5C). However, we did not identify any other bHLH transcription factors in this experiment (Figure S5C and Table S4). This suggests that HEB either binds DNA as a homodimer or binds another factor that was not detected by this assay. To independently validate the interaction between ID4 and HEB, the proximity ligation assay (PLA) was used to visualize protein-protein interactions *in situ*. Multiple PLA foci were detected in the nuclei of Comma-Dβ cells co-stained with ID4 and HEB antibodies (Figures 5C and 5D).

To test if ID4 and HEB interacted *in vivo*, we engineered a tagged ID4 mouse model in which a FlagV5 tag, a very efficient and specific target for immunoprecipitation, was integrated into the *Id4* locus downstream of the open reading frame. The result was FlagV5-tagged ID4 protein under the control of endogenous regulatory elements. We first ensured that ID4 and HEB were co-expressed by performing co-immunofluorescent staining for V5 and HEB. ID4-FlagV5 expression was tightly restricted to the cap and ductal basal cells, whereas HEB was more ubiquitous in its expression, detected in basal and luminal epithelial cells and surrounding stromal cells (Figure 5E). ID4 and HEB co-localized in the nuclei of cap and duct basal cells (insets; Figure 5E). Reciprocal co-IP was carried out from digested mammary glands of ID4-FlagV5 mice using antibodies raised against ID4, Flag, V5 (two different antibodies), and HEB (Figure 5F). Each antibody tested

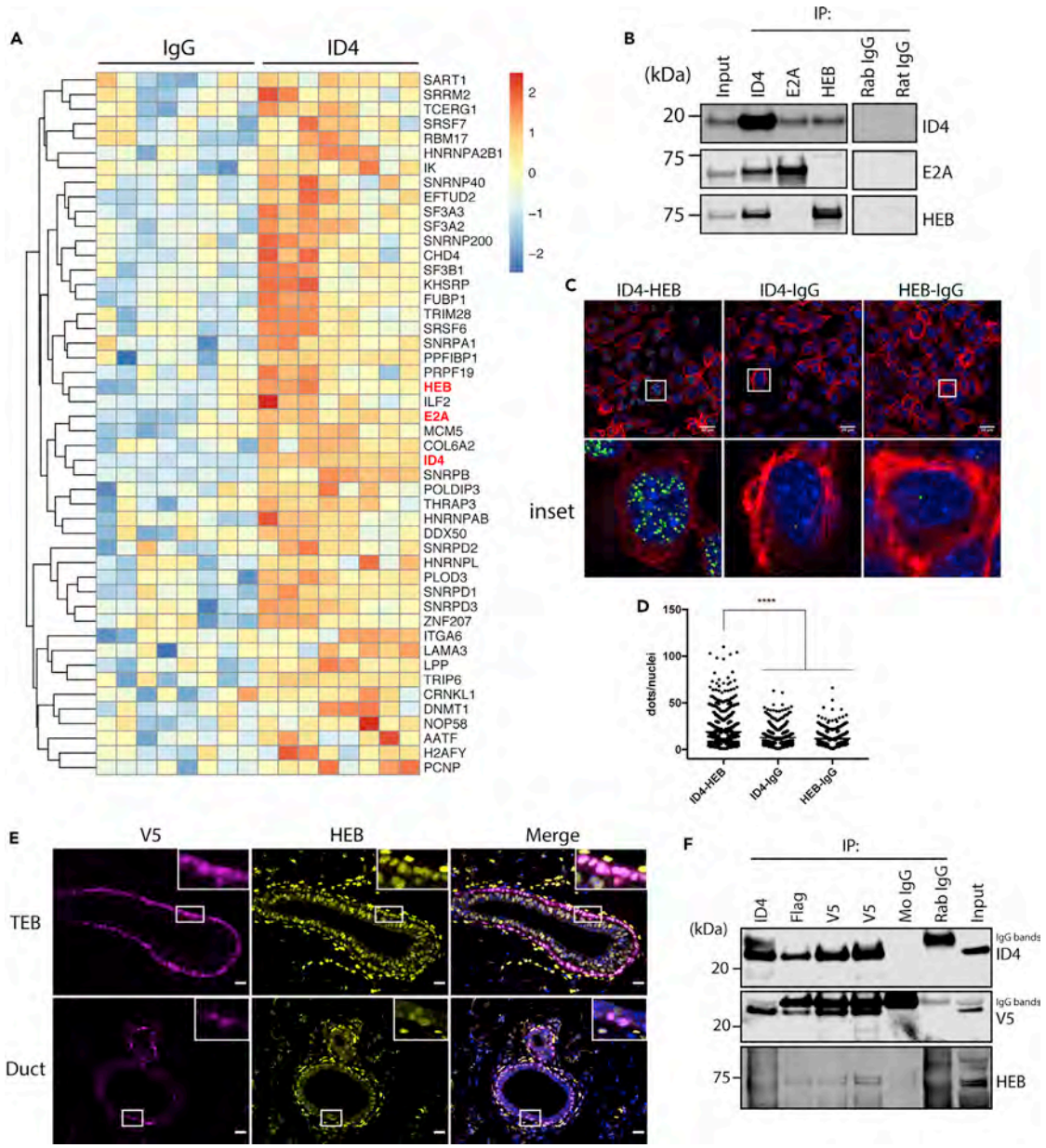


Figure 5. ID4 interacts with E-proteins E2A and HEB

(A) Unsupervised hierarchical clustering heatmap of SWATH RIME data from Comma-Dβ cells. Proteins with significantly higher abundance (p value<0.05) in the ID4 IPs compared with IgG IPs are displayed. Log2 protein area was used to generate the heatmap. Data from three independent experiments shown, each with 2–3 technical replicates.

(B) Co-immunoprecipitation (co-IP) and western blotting of ID4, E2A, and HEB and IgG negative controls from uncrosslinked Comma-Dβ lysates. Irrelevant lanes were digitally removed as indicated by the space.

(C) Proximity ligation assay (PLA) in Comma-Dβ cells for ID4 and HEB or corresponding negative control IgG. The cytoskeleton is stained with phalloidin and nucleus with DAPI. Scale bar: 20 μm. High power insets are shown below.

(D) Quantification of PLA foci from six random fields of view for each condition, each with approximately 50 nuclei per image. Ordinary one-way ANOVA test was used to test significance. ****$p < 0.001$. N = 3.

(E) Co-immunofluorescent staining of V5 and HEB in TEB (upper) and duct (lower) from ID4-FlagV5 mouse mammary glands. High-power insets feature cells positive for both proteins. Scale bar: 20 μm.

(F) co-IP and western blotting from ID4-FlagV5 mammary gland protein extracts. ID4 was immunoprecipitated using antibodies raised against ID4 and Flag and two independent V5 antibodies. ID4, V5, and HEB were detected by western blotting.

See also Figure S5 and Table S4.

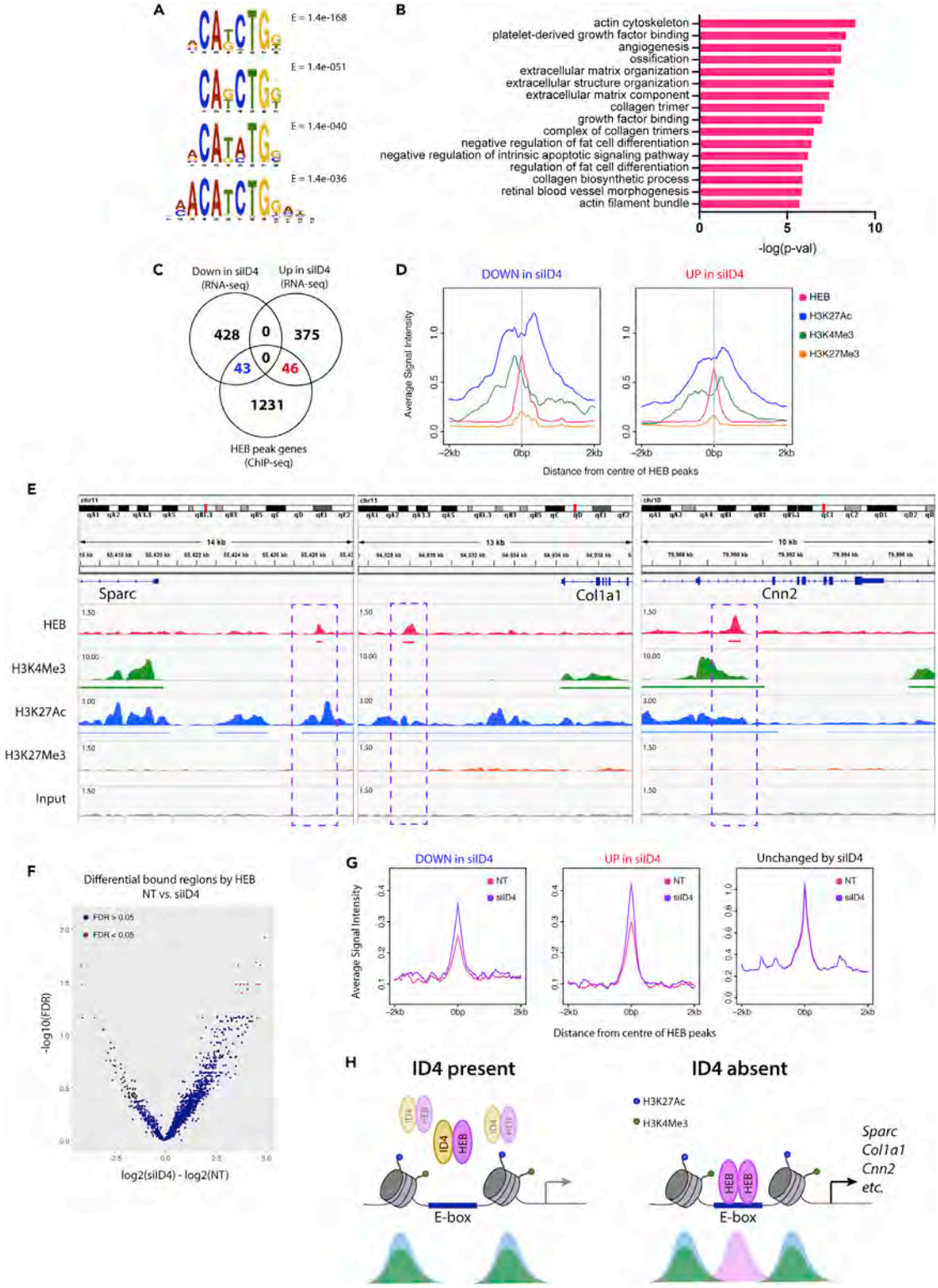


Figure 6. HEB directly binds to a subset of ID4 target genes

(A) Top four enriched transcription factor binding motifs determined using MEME-ChIP for consensus HEB ChIP-seq peaks in Comma-Dβ cells. E-values are displayed.

(B) GREAT pathway analysis of consensus HEB peaks. Top 16 Gene Ontologies (Biological process, cellular component, molecular function) are displayed.

Figure 6. *Continued*

(C) Venn diagram showing overlap between genes associated with an HEB peak and ID4 RNA-seq differentially expressed genes.

(D) Profile plots of average HEB, H3K4Me3, H3K27Ac, and H3K27Me3 signal intensity at regions associated with siID4 downregulated (left) and upregulated (right) RNA-seq differentially expressed genes.

(E) Examples of HEB and histone mark peaks occurring upstream of siID4 upregulated genes *Sparc*, *Col1a1*, and *Cnn2* from the Integrative Genomics Viewer (IGV). Bars beneath peaks represent consensus MACS call (FDR<0.05) in at least two of four biological replicates. Input was used as a negative control. Purple boxes highlight HEB-binding regions. Refseq genes shown in blue. Data scales for each track are indicated.

(F) Volcano plot of differential binding analysis. Analysis using edgeR of HEB binding in siID4 verses NT of three biological replicates. Regions with an FDR<0.05 are indicated in pink.

(G) Profile plots of average HEB signal intensity in NT control (pink) and siID4 (purple) conditions at regions associated with RNA-seq siID4 downregulated (left), upregulated (middle), and unchanged (right) genes. (H) Model of ID4 and HEB action in mammary epithelial cells. Left: when ID4 is expressed it interacts with HEB, antagonizing its transcriptional activity. Right: when ID4 is depleted, HEB dimerizes and binds to E-box motifs in the promoters and enhancers of developmental genes involved in contraction and ECM. Below are hypothetical ChIP signals for H3K27Ac (blue), H3K4Me3 (green), and HEB (pink).

See also Figure S6 and Table S5.

was able to precipitate both ID4 and HEB, confirming interaction between the two transcription factors *in vivo*.

HEB binds to regulatory elements of a subset of ID4 differentially expressed genes

To establish if ID4 regulates gene expression through HEB, we sought to determine if HEB directly binds to genes regulated by ID4 by performing ChIP-seq for HEB in Comma-Dβ cells. Across four biological replicates there were a total of 2,752 HEB peaks identified (FDR<0.05). This was narrowed down to 956 consensus peaks, which were present in at least two replicates. Transcription factor motif enrichment was carried out using MEME-ChIP (Machanick and Bailey, 2011), and the top enriched motifs were canonical E-box motifs (CANNTG), which are the binding sites for E-proteins (Figure 6A). The majority of peaks were mapped to intergenic and intronic regions, and approximately 5% of the peaks occurred at gene promoters (Figure S6A). We used the Genomics Regions Enrichment of Annotations Tool (GREAT) to analyze the functional significance of the regions bound by HEB (McLean et al., 2010). In this unbiased analysis, top enriched pathways were related to actin cytoskeleton and ECM organization (Figure 6B), resembling the pathways negatively regulated by ID4 in the gene expression profiling (Figure 4B). This suggests that ID4 mediates repression through its physical interaction with HEB.

The consensus peaks were annotated to 1,320 genes, using the default GREAT basal plus extension gene annotation rule (McLean et al., 2010). We overlapped these genes with those regulated following ID4 KD to determine the genes directly regulated by HEB (Figure 6C). Approximately 10% of these genes had an associated HEB peak, which is more than expected by chance ($p < 5.65 \times 10^{-13}$; hypergeometric test). The remainder of the genes regulated by ID4 KD are likely due to secondary effects of ID4 KD- or HEB-independent mechanisms. The fact that there was a similar number of genes overlapping in both upregulated (46) and downregulated (43) genes suggests that HEB can bind to sites both negatively and positively regulated by ID4. In line with this, E-proteins have previously been demonstrated to act as both activators and repressors of gene transcription by recruiting different co-factors (Bayly et al., 2004; Zhang et al., 2004).

In parallel, we performed ChIP-seq for three histone modifications—H3K4Me3 (active promoter mark), H3K27Ac (active enhancer mark), and H3K27Me3 (repressive chromatin mark)—to elucidate the chromatin context of HEB-bound regions. A number of HEB peaks associated with ID4-regulated genes demonstrated a bimodal distribution of H3K27Ac signal (Figures 6D and S6B), suggesting that HEB binds to enhancers. Some of the peaks were also localized to active promoters (Figures 6D and S6B). HEB peaks were observed at enhancer-marked chromatin upstream of ID4-repressed ECM genes expressed in myoepithelial cells such as *Sparc*, *Col1a1*, *Col1a2*, *Col3a1*, and *Col5a1* (Figures 6E and S6C) (Barsky and Karlin, 2005). HEB binding was also observed near the promoter of contractile gene *Cnn2* (Figure 6E), whose RNA and protein product CNN2 was suppressed by ID4 (Table S3 and Figure 4E). Of relevance, *Cnn2*, encoding Calponin 2, was recently discovered to be regulated by a super enhancer specifically accessible in myoepithelial cells (Pervolarakis et al., 2019). This further indicates that ID4 inhibits HEB's ability to activate transcription of genes that define the myoepithelial fate.

To determine whether HEB DNA binding is augmented when released from inhibition by ID4 we performed HEB ChIP-seq on Comma-Dβ cells in which ID4 had been depleted by siRNA. Western blotting revealed

that ID4 protein was reduced to approximately 20% of control levels, whereas HEB expression was unchanged (Figure S6D). E-box motifs were again enriched in both conditions (Figure S6E). Differential binding analysis revealed a total of 290 regions changing upon ID4 depletion ($p < 0.05$) (Table S5). More peaks were increased than decreased (263 compared with 27), suggesting that depletion of ID4 increased HEB's DNA-binding activity (Figures 6F and S6F). GREAT analysis revealed that the peaks that increased were involved in processes such as gland morphogenesis, skeletal development, and branching morphogenesis (Figure S6G). No pathways were enriched in the regions that were decreased when ID4 was knocked down. Finally, we observed an increase in HEB binding in cells depleted of ID4, specifically at genes that were differentially expressed by ID4 KD (Figure 6G). Together, our ChIP-seq analysis suggests that HEB binds to E-box motifs in regulatory elements of basal developmental genes involved in ECM and the contractile cytoskeleton, and this is antagonized by its interaction with ID4 (Figure 6H).

DISCUSSION

We show that ID4 represses genes associated with myoepithelial differentiation in mammary basal stem cells, in part through the inhibition of the E-protein HEB. ID4 has previously been demonstrated to block luminal commitment of basal cells via inhibition of key luminal driver genes including *Elf5*, *Notch*, *Brca1*, *Esr1*, *PR*, and *FoxA1* (Junankar et al., 2015; Best et al., 2014). The dual inhibition of both luminal and myoepithelial differentiation by ID4 likely protects the stem-cell phenotype of uncommitted basal cells during development. Subsequent downregulation of ID4, through a currently unknown mechanism, may then allow basal cells to adopt a luminal or myoepithelial fate depending on the cellular context.

HEB has not been associated with lineage commitment of epithelial tissues. It is, however, known to be involved in the specification of lymphocyte (Braunstein and Anderson, 2012), hematopoietic (Li et al., 2017), mesodermal (Yoon et al., 2015), neuronal (Mesman and Smidt, 2017), and skeletal muscle (Conway et al., 2004) lineages. Our proteogenomic analyses support the model outlined in Figure 6H. When ID4 is highly expressed, such as in cap cells, it sequesters HEB off chromatin, preventing expression of differentiation genes, thus determining a stem-like state (Figure 6H; left). When ID4 expression is low, such as in differentiating cells, HEB is able to bind to E-box DNA motifs at promoters/enhancers to activate transcription of developmental genes that specify functional myoepithelial cells (Figure 6H; right). Further functional studies are needed to demonstrate HEB's requirement in promoting myoepithelial differentiation.

Compared with the luminal lineage, the molecular regulators controlling the basal lineage remain poorly understood. One of the few transcription factors known to promote myoepithelial differentiation is MADS-box protein SRF and its associated co-activator MRTFA (Sun et al., 2006; Li et al., 2006). HEB and SRF cooperatively activate transcription of *Acta2* in cultured fibroblasts (Kumar et al., 2003). This occurred in an E-box-dependent manner and was inhibited by ID1 and ID2 overexpression (Kumar et al., 2003). Although we show that ID4 suppresses *Acta2*, we did not observe HEB binding to the *Acta2* promoter. However, it is possible that in the absence of ID4, HEB and SRF cooperate to drive expression of other myoepithelial genes. This is supported by the positive enrichment of SRF targets upon ID4 depletion. Subsequent studies should test whether HEB and SRF cooperate in mammary epithelial cells.

The roles of ID4 in regulating myoepithelial commitment and ECM deposition expand upon why ID4 is required for pubertal mammary gland morphogenesis (Junankar et al., 2015; Dong et al., 2011; Best et al., 2014). TEBs undergo collective migration, enabling the coordinated movement of adherent cells into the stromal fat pad (Ewald et al., 2008). We hypothesize that the high levels of ID4 in cap cells prevent epithelial cells from acquiring a mesenchymal/myoepithelial phenotype. Similar mechanisms have been observed for the transcriptional repressor OVOL2, which inhibits EMT to allow for collective migration (Watanabe et al., 2014), and for C/EBPα, which maintains epithelial homeostasis of human mammary epithelial cells (Lourenço et al., 2020). During ductal elongation, collagenous stromal ECM is absent directly in front of the invading TEBs (Silberstein et al., 1990; Sternlicht, 2006). We show that ID4 suppresses collagen synthesis and deposition around TEBs, which may otherwise act as a physical barrier to impede invasion. In support of this idea, ectopic deposition of collagen by mammary epithelial cells induced by exogeneous TGF-β, or forced expression of recombinant type I collagen that is resistant to collagenase attack, causes ensheathment of TEBs by collagen and retardation of ductal elongation (Silberstein et al., 1990; Feinberg et al., 2018).

Developmental transcription factors are often dysregulated in cancer. ID4 is highly expressed in ~50% of basal-like breast cancer (BLBC) cases and associates with poor prognosis (Junankar et al., 2015; Baker et al., 2016). Interestingly, ID4 has also been implicated in prostate development (Sharma et al., 2013) and acts as a tumor suppressor in this context (Carey et al., 2009). It is likely that the different repertoire of binding partners in different cell types gives rise to the organ-specific functions of ID4 in the breast and prostate. Given the cell-intrinsic role of ID4 in promoting growth/proliferation and inhibiting differentiation in the mammary gland, it is easy to envision how overexpression of ID4 could lead to an aggressive breast cancer phenotype. In addition, the suppression of ECM synthesis may allow tumor cells to easily invade into the surrounding stroma, akin to the invasion of cap cells during ductal elongation. Future work should test whether the mechanisms discovered here are conserved in breast cancer initiation and progression.

To conclude, these insights into ID4 and HEB function help unravel regulation within the basal differentiation hierarchy, with broader implications for the regulation of epithelial stem cells in general, and also in tumor progression.

Limitations of the study

A caveat of this study is that dissection of ID4's molecular mechanism by RIME and ChIP-seq was performed in one cell line. However, the Comma-D cell line is a well-accepted model for mammary development often used for the purpose of genomic and biochemical studies (Ibarra et al., 2007; Wellberg et al., 2010; Best et al., 2014), and results were validated in human cell lines and transgenic mouse models (Figures S5, 5E, and 5F).

Resource availability

Lead contact

Further information and requests for resources and reagents should be directed to the Lead Contact, Alexander Swarbrick (a.swarbrick@garvan.org.au).

Materials availability

The ID4-FlagV5 mouse model is available upon request with a Material Transfer Agreement.

Data and code availability

The accession number for the RNA-seq and ChIP-seq data reported in this paper isGEO: GSE149969. The accession number for the mass spectrometry proteomics data is PRIDE :: PXD017517.

METHODS

All methods can be found in the accompanying Transparent methods supplemental file.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.isci.2021.102072.

ACKNOWLEDGMENTS

We are very grateful for Amanda Khoury for her extensive advice on the ChIP-seq experiments, Samantha Oakes for providing the mammary gland FFPE blocks from different developmental stages, David Gallego Ortega for advice on tissue dissociation and sorting, and William Hughes for help with microscopy image analysis. We acknowledge the Mouse Engineering Garvan/ABR (MEGA) Facility for generating the ID4-FlagV5 mouse line. We acknowledge Guy Riddihough (Life Science Editors) for his thorough editing of this manuscript. This work was supported by funding from John and Deborah McMurtrie, the National Health and Medical Research Council (NHMRC) (1107671), and The Petre Foundation. A.S. is the recipient of a Senior Research Fellowship from the NHMRC. H.H was supported by an Australia Postgraduate Award. Aspects of this research were supported by access to the Australian Proteome Analysis Facility, funded by the Australian Government's National Collaborative Research Infrastructure Scheme. T.R.C and J.N.S were supported by NHMRC, Cancer Council NSW (CCNSW), Cancer Institute NSW (CINSW), and Love Your Sister in association with the National Breast Cancer Foundation (NBCF) and Susan G Komen.

AUTHOR CONTRIBUTIONS

Conceptualization: A.S. Investigation: H.H., L.A.B., B.P., C.K., C.C., and A.M. Formal Analysis: D.R., S.Z.W., H.H., C.C., C.K., M.P.M., J.N.S., and T.R.C. Writing—original draft: H.H. Writing—review and editing: A.S., S.J., S.Z.W., J.V., J.S.C., and C.J.O. Resources: J.V. and N.D.H. Funding Acquisition: A.S. and J.S.C. Supervision: A.S., S.J., and C.J.O.

DECLARATION OF INTERESTS

N.D.H. has ownership and stock options in oNKo-Innate Pty Ltd. The remaining authors declare no competing interests.

Received: September 10, 2020
Revised: November 24, 2020
Accepted: January 12, 2021
Published: February 19, 2021

REFERENCES

Asselin-Labat, M.L., Sutherland, K.D., Barker, H., Thomas, R., Shackleton, M., Forrest, N.C., Hartley, L., Robb, L., Grosveld, F.G., Van Der Wees, J., et al. (2007). Gata-3 is an essential regulator of mammary-gland morphogenesis and luminal-cell differentiation. Nat. Cell Biol. *9*, 201–209.

Bach, K., Pensa, S., Grzelak, M., Hadfield, J., Adams, D.J., Marioni, J.C., and Khaled, W.T. (2017). Differentiation dynamics of mammary epithelial cells revealed by single-cell RNA sequencing. Nat. Commun. *8*, 2128.

Baker, L.A., Holliday, H., and Swarbrick, A. (2016). ID4 controls mammary lineage commitment and inhibits BRCA1 in basal breast cancer. Endocr. Relat. Cancer *23*, R381–R392.

Barsky, S.H., and Karlin, N.J. (2005). Myoepithelial cells: autocrine and paracrine suppressors of breast cancer progression. J. Mammary Gland Biol. Neoplasia *10*, 249–260.

Bayly, R., Chuen, L., Currie, R.A., Hyndman, B.D., Casselman, R., Blobel, G.A., and Lebrun, D.P. (2004). E2A-PBX1 interacts directly with the KIX domain of CBP/p300 in the induction of proliferation in primary hematopoietic cells. J. Biol. Chem. *279*, 55362–55371.

Benezra, R., Davis, R.L., Lockshon, D., Turner, D.L., and Weintraub, H. (1990). The protein Id: a negative regulator of helix-loop-helix DNA binding proteins. Cell *61*, 49–59.

Best, S.A., Hutt, K.J., Fu, N.Y., Vaillant, F., Liew, S.H., Hartley, L., Scott, C.L., Lindeman, G.J., and Visvader, J.E. (2014). Dual roles for Id4 in the regulation of estrogen signaling in the mammary gland and ovary. Development *141*, 3159–3164.

Bouras, T., Pal, B., Vaillant, F., Harburg, G., Asselin-Labat, M.L., Oakes, S.R., Lindeman, G.J., and Visvader, J.E. (2008). Notch signaling regulates mammary stem cell function and luminal cell-fate commitment. Cell Stem Cell *3*, 429–441.

Braunstein, M., and Anderson, M.K. (2012). HEB in the spotlight: transcriptional regulation of T-cell specification, commitment, and developmental plasticity. Clin. Dev. Immunol. *2012*, 678705.

Buchwalter, G., Hickey, M.M., Cromer, A., Selfors, L.M., Gunawardane, R.N., Frishman, J., Jeselsohn, R., Lim, E., Chi, D., Fu, X., et al. (2013). PDEF promotes luminal differentiation and acts as a survival factor for ER-positive breast cancer cells. Cancer Cell *23*, 753–767.

Carey, J.P., Asirvatham, A.J., Galm, O., Ghogomu, T.A., and Chaudhary, J. (2009). Inhibitor of differentiation 4 (Id4) is a potential tumor suppressor in prostate cancer. BMC Cancer *9*, 173.

Carr, J.R., Kiefer, M.M., Park, H.J., Li, J., Wang, Z., Fontanarosa, J., Dewaal, D., Kopanja, D., Benevolenskaya, E.V., Guzman, G., and Raychaudhuri, P. (2012). FoxM1 regulates mammary luminal cell fate. Cell Rep. *1*, 715–729.

Chakrabarti, R., Wei, Y., Romano, R.A., Decoste, C., Kang, Y., and Sinha, S. (2012). Elf5 regulates mammary gland stem/progenitor cell fate by influencing notch signaling. Stem Cells *30*, 1496–1508.

Conway, K., Pin, C., Kiernan, J.A., and Merrifield, P. (2004). The E protein HEB is preferentially expressed in developing muscle. Differ. Res. Biol. Divers. *72*, 327–340.

Danielson, K.G., Oborn, C.J., Durban, E.M., Butel, J.S., and Medina, D. (1984). Epithelial mouse mammary cell line exhibiting normal morphogenesis in vivo and functional differentiation in vitro. Proc. Natl. Acad. Sci. U S A *81*, 3756–3760.

Davis, F.M., Lloyd-Lewis, B., Harris, O.B., Kozar, S., Winton, D.J., Muresan, L., and Watson, C.J. (2016). Single-cell lineage tracing in the mammary gland reveals stochastic clonal dispersion of stem/progenitor cell progeny. Nat. Commun. *7*, 13053.

Deugnier, M.A., Faraldo, M.M., Teuliere, J., Thiery, J.P., Medina, D., and Glukhova, M.A. (2006). Isolation of mouse mammary epithelial progenitor cells with basal characteristics from the Comma-Dbeta cell line. Dev. Biol. *293*, 414–425.

Deugnier, M.A., Moiseyeva, E.P., Thiery, J.P., and Glukhova, M. (1995). Myoepithelial cell differentiation in the developing mammary gland: progressive acquisition of smooth muscle phenotype. Dev. Dyn. *204*, 107–117.

Dong, J., Huang, S., Caikovski, M., Ji, S., Mcgrath, A., Custorio, M.G., Creighton, C.J., Maliakkal, P., Bogoslovskaia, E., Du, Z., et al. (2011). ID4 regulates mammary gland development by suppressing p38MAPK activity. Development *138*, 5247–5256.

Ewald, A.J., Brenot, A., Duong, M., Chan, B.S., and Werb, Z. (2008). Collective epithelial migration and cell rearrangements drive mammary branching morphogenesis. Dev. Cell *14*, 570–581.

Feinberg, T.Y., Zheng, H., Liu, R., Wicha, M.S., Yu, S.M., and Weiss, S.J. (2018). Divergent matrix-remodeling strategies distinguish developmental from neoplastic mammary epithelial cell invasion programs. Dev. Cell *47*, 145–160.e6.

Guo, W., Keckesova, Z., Donaher, J.L., Shibue, T., Tischler, V., Reinhardt, F., Itzkovitz, S., Noske, A., Zurrer-Hardi, U., Bell, G., et al. (2012). Slug and Sox9 cooperatively determine the mammary stem cell state. Cell *148*, 1015–1028.

Ibarra, I., Erlich, Y., Muthuswamy, S.K., Sachidanandam, R., and Hannon, G.J. (2007). A role for microRNAs in maintenance of mouse mammary epithelial progenitor cells. Genes Dev. *21*, 3238–3243.

Idoux-Gillet, Y., Nassour, M., Lakis, E., Bonini, F., Theillet, C., Du Manoir, S., and Savagner, P. (2018). Slug/Pcad pathway controls epithelial cell dynamics in mammary gland and breast carcinoma. Oncogene *37*, 578–588.

Junankar, S., Baker, L.A., Roden, D.L., Nair, R., Elsworth, B., Gallego-Ortega, D., Lacaze, P., Cazet, A., Nikolic, I., Teo, W.S., et al. (2015). ID4 controls mammary stem cells and marks breast cancers with a stem cell-like phenotype. Nat. Commun. *6*, 6548.

Kiemer, A.K., Takeuchi, K., and Quinlan, M.P. (2001). Identification of genes involved in epithelial-mesenchymal transition and tumor progression. Oncogene *20*, 6679–6688.

Kouros-Mehr, H., Slorach, E.M., Sternlicht, M.D., and Werb, Z. (2006). GATA-3 maintains the

differentiation of the luminal cell fate in the mammary gland. Cell *127*, 1041–1055.

Kumar, M.S., Hendrix, J.A., Johnson, A.D., and Owens, G.K. (2003). Smooth muscle alpha-actin gene requires two E-boxes for proper expression in vivo and is a target of class I basic helix-loop-helix proteins. Circ. Res. *92*, 840–847.

Lee, K., Gjorevski, N., Boghaert, E., Radisky, D.C., and Nelson, C.M. (2011). Snail1, Snail2, and E47 promote mammary epithelial branching morphogenesis. EMBO J *30*, 2662–2674.

Li, S., Chang, S., Qi, X., Richardson, J.A., and Olson, E.N. (2006). Requirement of a myocardin-related transcription factor for development of mammary myoepithelial cells. Mol. Cell Biol. *26*, 5797–5808.

Li, Y., Brauer, P.M., Singh, J., Xhiku, S., Yoganathan, K., Zúñiga-Pflücker, J.C., and Anderson, M.K. (2017). Targeted disruption of TCF12 reveals HEB as essential in human mesodermal specification and hematopoiesis. Stem Cell Rep. *9*, 779–795.

Lilja, A.M., Rodilla, V., Huyghe, M., Hannezo, E., Landragin, C., Renaud, O., Leroy, O., Rulands, S., Simons, B.D., and Fre, S. (2018). Clonal analysis of Notch1-expressing cells reveals the existence of unipotent stem cells that retain long-term plasticity in the embryonic mammary gland. Nat. Cell Biol. *20*, 677–687.

Lim, E., Wu, D., Pal, B., Bouras, T., Asselin-Labat, M.L., Vaillant, F., Yagita, H., Lindeman, G.J., Smyth, G.K., and Visvader, J.E. (2010). Transcriptome analyses of mouse and human mammary cell subpopulations reveal multiple conserved genes and pathways. Breast Cancer Res. *12*, R21.

Liu, J., Eischeid, A.N., and Chen, X.M. (2012a). Col1A1 production and apoptotic resistance in TGF-β1-induced epithelial-to-mesenchymal transition-like phenotype of 603B cells. PLoS One *7*, e51371.

Liu, S., Ginestier, C., Charafe-Jauffret, E., Foco, H., Kleer, C.G., Merajver, S.D., Dontu, G., and Wicha, M.S. (2008). BRCA1 regulates human mammary stem/progenitor cell fate. Proc. Natl. Acad. Sci. U S A *105*, 1680–1685.

Liu, X., Ory, V., Chapman, S., Yuan, H., Albanese, C., Kallakury, B., Timofeeva, O.A., Nealon, C., Dakic, A., Simic, V., et al. (2012b). ROCK inhibitor and feeder cells induce the conditional reprogramming of epithelial cells. Am. J. Pathol. *180*, 599–607.

Lloyd-Lewis, B., Davis, F.M., Harris, O.B., Hitchcock, J.R., and Watson, C.J. (2018). Neutral lineage tracing of proliferative embryonic and adult mammary stem/progenitor cells. Development *145*, dev164079.

Lourenço, A.R., Roukens, M.G., Seinstra, D., Frederiks, C.L., Pals, C.E., Vervoort, S.J., Margarido, A.S., Van Rheenen, J., and Coffer, P.J. (2020). C/EBPα is crucial determinant of epithelial maintenance by preventing epithelial-to-mesenchymal transition. Nat. Commun. *11*, 785.

Machanick, P., and Bailey, T.L. (2011). MEME-ChIP: motif analysis of large DNA datasets. Bioinformatics *27*, 1696–1697.

Macias, H., and Hinck, L. (2012). Mammary gland development. Wiley Interdiscip. Rev. Dev. Biol. *1*, 533–557.

Massari, M.E., and Murre, C. (2000). Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms. Mol. Cell Biol. *20*, 429–440.

Mclean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nat. Biotechnol. *28*, 495–501.

Mesman, S., and Smidt, M.P. (2017). Tcf12 is involved in early cell-fate determination and subset specification of midbrain dopamine neurons. Front. Mol. Neurosci. *10*, 353.

Miano, J.M., Long, X., and Fujiwara, K. (2007). Serum response factor: master regulator of the actin cytoskeleton and contractile apparatus. Am. J. physiol. Cell Physiol. *292*, C70–C81.

Mills, A.A., Zheng, B., Wang, X.J., Vogel, H., Roop, D.R., and Bradley, A. (1999). p63 is a p53 homologue required for limb and epidermal morphogenesis. Nature *398*, 708–713.

Mohammed, H., D'santos, C., Serandour, A.A., Ali, H.R., Brown, G.D., Atkins, A., Rueda, O.M., Holmes, K.A., Theodorou, V., Robinson, J.L., et al. (2013). Endogenous purification reveals GREB1 as a key estrogen receptor regulatory factor. Cell Rep. *3*, 342–349.

Muschler, J., and Streuli, C.H. (2010). Cell-matrix interactions in mammary gland development and breast cancer. Cold Spring Harb. Perspect. Biol. *2*, a003202.

Oakes, S.R., Naylor, M.J., Asselin-Labat, M.L., Blazek, K.D., Gardiner-Garden, M., Hilton, H.N., Kazlauskas, M., Pritchard, M.A., Chodosh, L.A., Pfeffer, P.L., et al. (2008). The Ets transcription factor Elf5 specifies mammary alveolar cell fate. Genes Dev. *22*, 581–586.

Paine, I.S., and Lewis, M.T. (2017). The terminal end bud: the little engine that could. J. Mammary Gland Biol. Neoplasia *22*, 93–108.

Pervolarakis, N., Nguyen, Q., Gutierrez, G., Sun, P., Jhutty, D., Zheng, G.X., Nemec, C.M., Dai, X., Watanabe, K., and Kessenbrock, K. (2019). Integrated single-cell transcriptomics and chromatin accessibility analysis reveals novel regulators of mammary epithelial cell identity. bioRxiv, 740746. https://doi.org/10.1101/740746.

Prater, M.D., Petit, V., Alasdair Russell, I., Giraddi, R.R., Shehata, M., Menon, S., Schulte, R., Kalajzic, I., Rath, N., Olson, M.F., et al. (2014). Mammary stem cells have myoepithelial cell properties. Nat. Cell Biol. *16*, 1–7.

Rios, A.C., Fu, N.Y., Lindeman, G.J., and Visvader, J.E. (2014). In situ identification of bipotent stem cells in the mammary gland. Nature *506*, 322–327.

Scheele, C.L., Hannezo, E., Muraro, M.J., Zomer, A., Langedijk, N.S., Van Oudenaarden, A., Simons, B.D., and Van Rheenen, J. (2017). Identity and dynamics of mammary stem cells during branching morphogenesis. Nature *542*, 313–317.

Shackleton, M., Vaillant, F., Simpson, K.J., Stingl, J., Smyth, G.K., Asselin-Labat, M.L., Wu, L., Lindeman, G.J., and Visvader, J.E. (2006). Generation of a functional mammary gland from a single stem cell. Nature *439*, 84–88.

Sharma, P., Knowell, A.E., Chinaranagari, S., Komaragiri, S., Nagappan, P., Patel, D., Havrda, M.C., and Chaudhary, J. (2013). Id4 deficiency attenuates prostate development and promotes PIN-like lesions by regulating androgen receptor activity and expression of NKX3.1 and PTEN. Mol. Cancer *12*, 67.

Silberstein, G.B., Strickland, P., Coleman, S., and Daniel, C.W. (1990). Epithelium-dependent extracellular matrix synthesis in transforming growth factor-beta 1-growth-inhibited mouse mammary gland. J. Cell Biol. *110*, 2209–2219.

Sternlicht, M.D. (2006). Key stages in mammary gland development: the cues that regulate ductal branching morphogenesis. Breast Cancer Res. *8*, 201.

Stingl, J., Eirew, P., Ricketson, I., Shackleton, M., Vaillant, F., Choi, D., Li, H.I., and Eaves, C.J. (2006). Purification and unique properties of mammary epithelial stem cells. Nature *439*, 993–997.

Sun, Y., Boyd, K., Xu, W., Ma, J., Jackson, C.W., Fu, A., Shillingford, J.M., Robinson, G.W., Hennighausen, L., Hitzler, J.K., et al. (2006). Acute myeloid leukemia-associated Mkl1 (Mrtf-a) is a key regulator of mammary gland function. Mol. Cell Biol. *26*, 5809–5826.

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S., and Rinn, J.L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat. Biotechnol. *32*, 381–386.

Van Amerongen, R., Bowman, A.N., and Nusse, R. (2012). Developmental stage and time dictate the fate of Wnt/beta-catenin-responsive stem cells in the mammary gland. Cell Stem Cell *11*, 387–400.

Van Keymeulen, A., Rocha, A.S., Ousset, M., Beck, B., Bouvencourt, G., Rock, J., Sharma, N., Dekoninck, S., and Blanpain, C. (2011). Distinct stem cells contribute to mammary gland development and maintenance. Nature *479*, 189–193.

Wang, D., Cai, C., Dong, X., Yu, Q.C., Zhang, X.O., Yang, L., and Zeng, Y.A. (2015). Identification of multipotent mammary stem cells by protein C receptor expression. Nature *517*, 81–84.

Wang, L.H., and Baker, N.E. (2015). E proteins and ID proteins: helix-loop-helix partners in development and disease. Dev. Cell *35*, 269–280.

Watanabe, K., Villarreal-Ponce, A., Sun, P., Salmans, M.L., Fallahi, M., Andersen, B., and Dai, X. (2014). Mammary morphogenesis and regeneration require the inhibition of EMT at terminal end buds by Ovol2 transcriptional repressor. Dev. Cell *29*, 59–74.

Wellberg, E., Metz, R.P., Parker, C., and Porter, W.W. (2010). The bHLH/PAS transcription factor singleminded 2s promotes mammary gland lactogenic differentiation. Development *137*, 945–952.

Williams, J.M., and Daniel, C.W. (1983). Mammary ductal elongation: differentiation of

myoepithelium and basal lamina during branching morphogenesis. Dev. Biol. *97*, 274–290.

Wuidart, A., Ousset, M., Rulands, S., Simons, B.D., Van Keymeulen, A., and Blanpain, C. (2016). Quantitative lineage tracing strategies to resolve multipotency in tissue-specific stem cells. Genes Dev. *30*, 1261–1277.

Wuidart, A., Sifrim, A., Fioramonti, M., Matsumura, S., Brisebarre, A., Brown, D., Centonze, A., Dannau, A., Dubois, C., Van Keymeulen, A., et al. (2018). Early lineage segregation of multipotent embryonic mammary gland progenitors. Nat. Cell Biol. *20*, 666–676.

Yamaji, D., Na, R., Feuermann, Y., Pechhold, S., Chen, W., Robinson, G.W., and Hennighausen, L. (2009). Development of mammary luminal progenitor cells is controlled by the transcription factor STAT5A. Genes Dev. *23*, 2382–2387.

Yang, A., Schweitzer, R., Sun, D., Kaghad, M., Walker, N., Bronson, R.T., Tabin, C., Sharpe, A., Caput, D., Crum, C., and Mckeon, F. (1999). p63 is essential for regenerative proliferation in limb, craniofacial and epithelial development. Nature *398*, 714–718.

Yoon, S.J., Foley, J.W., and Baker, J.C. (2015). HEB associates with PRC2 and SMAD2/3 to regulate developmental fates. Nat. Commun. *6*, 6546.

Yun, K., Mantani, A., Garel, S., Rubenstein, J., and Israel, M.A. (2004). Id4 regulates neural progenitor proliferation and differentiation in vivo. Development *131*, 5441–5448.

Zhang, J., Kalkum, M., Yamamura, S., Chait, B.T., and Roeder, R.G. (2004). E protein silencing by the leukemogenic AML1-ETO fusion protein. Science *305*, 1286–1289.

iScience, Volume *24*

Supplemental Information

Inhibitor of Differentiation 4 (ID4) represses mammary myoepithelial differentiation via inhibition of HEB

Holly Holliday, Daniel Roden, Simon Junankar, Sunny Z. Wu, Laura A. Baker, Christoph Krisp, Chia-Ling Chan, Andrea McFarland, Joanna N. Skhinas, Thomas R. Cox, Bhupinder Pal, Nicholas D. Huntington, Christopher J. Ormandy, Jason S. Carroll, Jane Visvader, Mark P. Molloy, and Alexander Swarbrick

Supplemental Figures and Legends

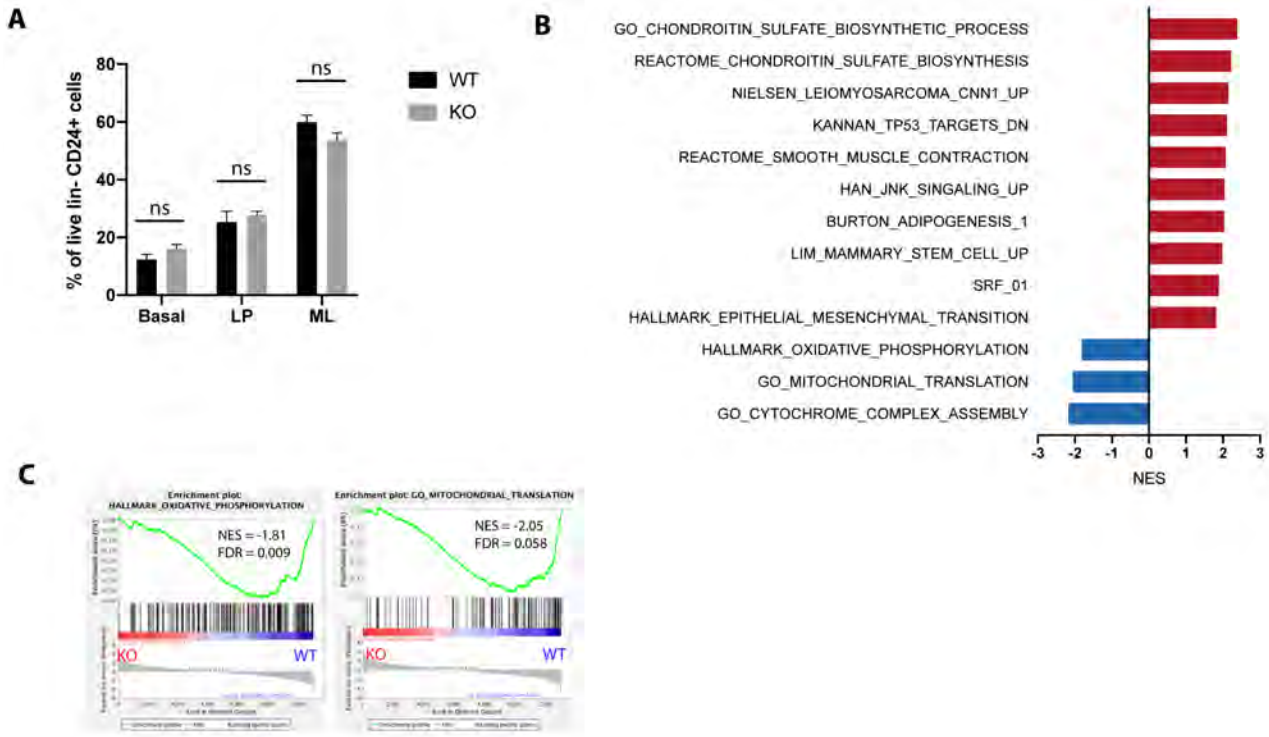


Figure S1. GSEA of ID4 knockout sorted basal cells, Related to Figure 1.

A) Proportions of basal, luminal progenitor (LP), and mature luminal (ML) subpopulations between ID4 WT and KO mammary epithelial cells. Unpaired two-tailed students t-test. ns = not significant. Error bars represent SEM. N=4. **B)** The top 10 positive and negatively enriched pathways with an FDR<0.1 are displayed. Only 3 pathways were negatively enriched with an FDR<0.1. **C)** Representative GSEA enrichment plots displaying the profile of the running Enrichment Score (green) and positions of gene set members on the rank ordered list for pathways related to cell growth**.** NES and FDR are indicated on the plots.

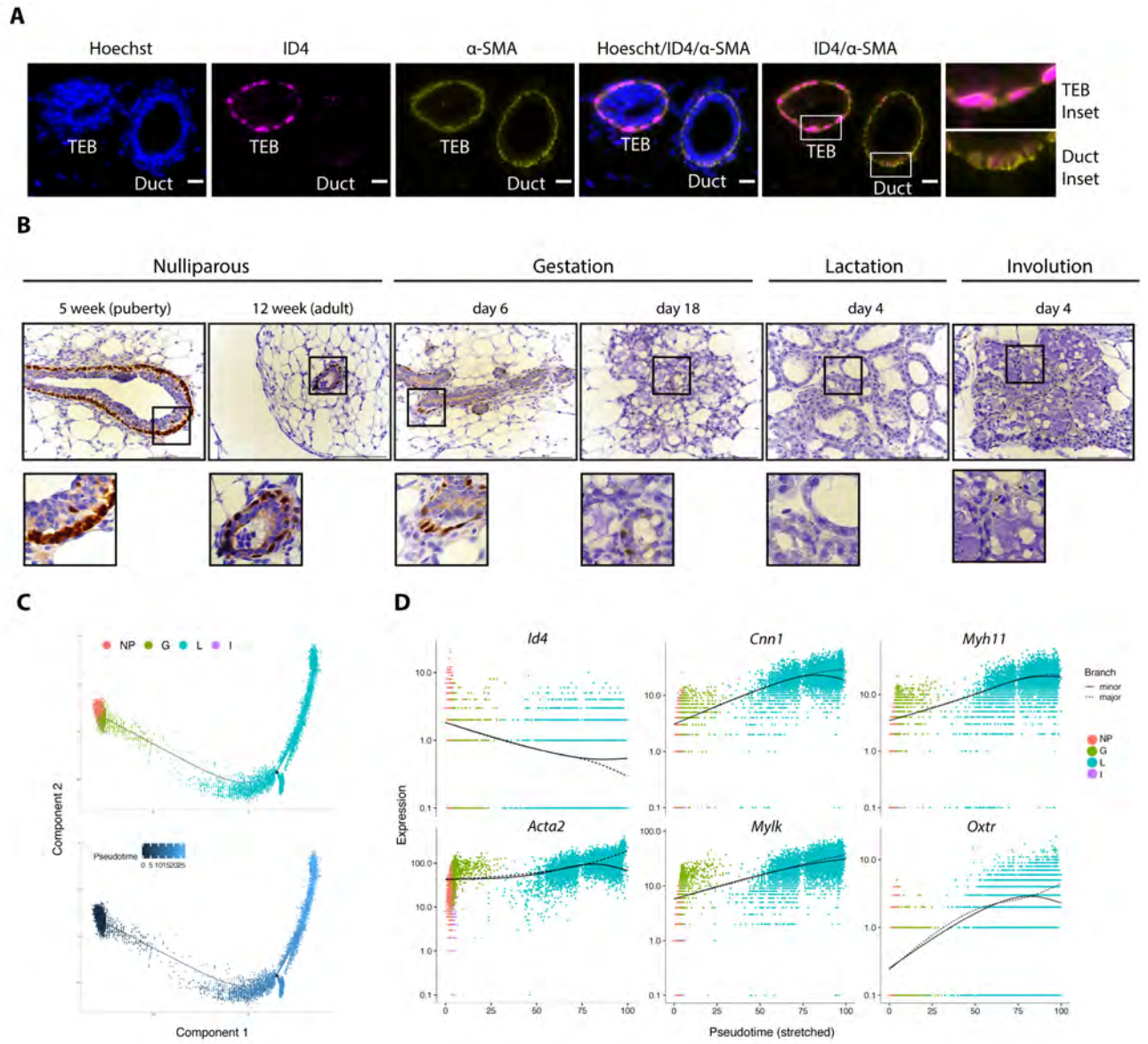


Figure S2. ID4 expression decreases in terminally differentiated myoepithelial cells, Related to Figure 2.

A) Co-immunofluorescent staining of ID4 and α-SMA in a section containing both a TEB and a duct from a pubertal (6 week) mammary gland. Scale bar = 20 µm**.** High power insets are shown. **B)** Mouse mammary glands at different developmental stages were stained by IHC for ID4. Scale bar = 100 µm. Representative images from 3 animals per stage shown. High power insets are displayed below images. **C)** Differentiation trajectory of basal cells from (Bach et al., 2017) coloured by developmental stage (upper) and pseudotime (lower). Low values (dark blue) represent undifferentiated cells. NP = Nulliparous (8 week), G= Gestation (Day 14.5), L = Lactation (Day 6), I = Involution (Day 11). **D)** Expression of *Id4, Cnn1, Myh11, Acta2, Mylk* and *Oxtr* as a function of pseudotime**.**

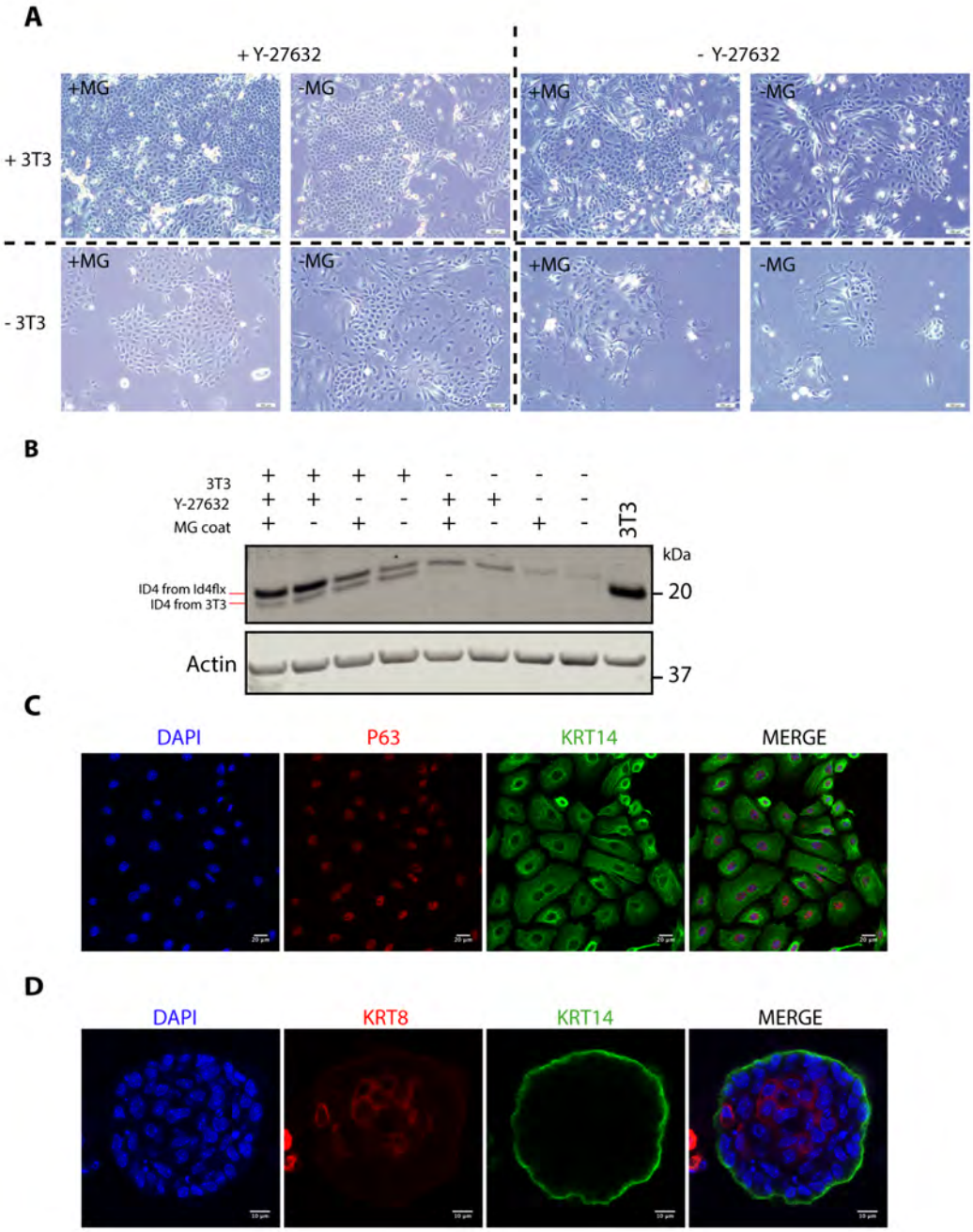


Figure S3. Establishment of 2D cultures and 3D organoids from primary basal cells, Related to Figure 3.

A) Phase contrast images of passage 11 conditionally reprogrammed cells grown in the presence and absence of ROCK inhibitor Y-27632, irradiated NIH-3T3 feeder cells, and Matrigel (MG) coated tissue culture flasks. Scale bar = 100 µm. **B)** ID4 western blot in cell lysates collected from cells shown in panel A. ID4 from basal cells runs at a higher molecular weight to ID4 expressed by NIH-3T3 cells. **C)** Co-immunofluorescent staining of primary basal cells with basal markers P63 and KRT14. Scale bar = 20 µm. **D)** Basal cells grown in 3D as organoids on top a plug of Matrigel. Confocal image of an organoid stained for luminal marker KRT8 and basal marker KRT14. Scale bar = 10 µm

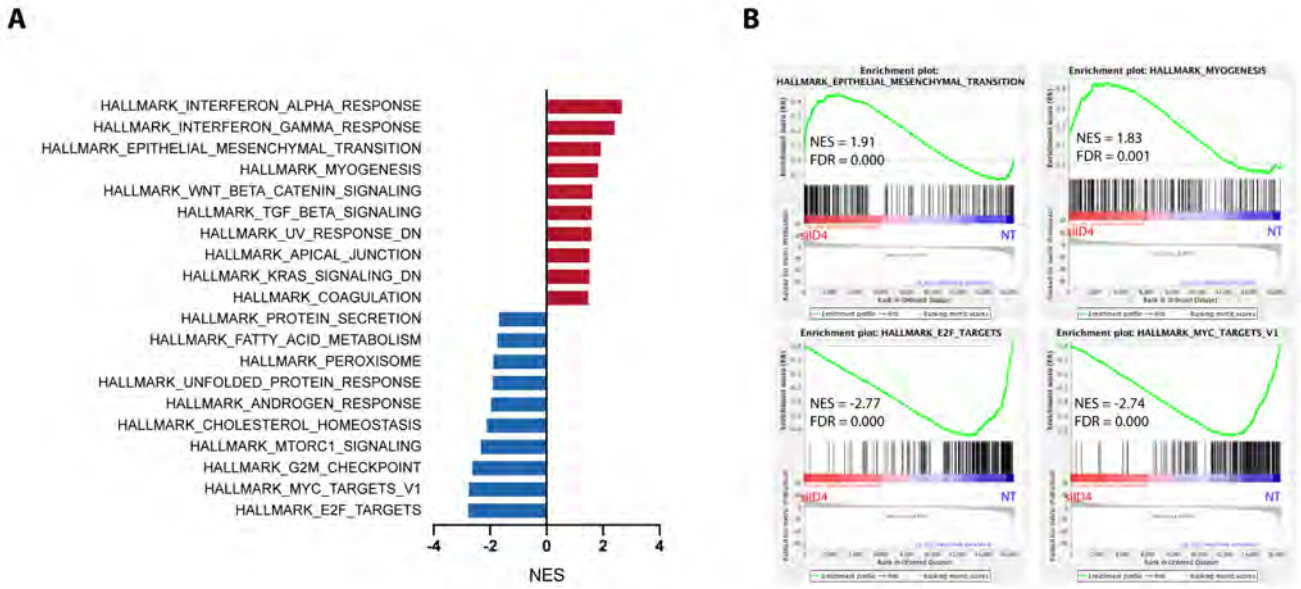


Figure S4. GSEA of ID4 knockdown Comma-Dβ cells, Related to Figure 4.

A) Genes were ranked based on the limma t-statistic comparing NT and siID4 cells and GSEA was carried out using the Hallmark gene sets. The top 10 positively (red) and negatively (blue) enriched pathways are displayed. **B)** Representative Hallmark GSEA enrichment plots displaying the profile of the running Enrichment Score (green) and positions of gene set members on the rank ordered list.

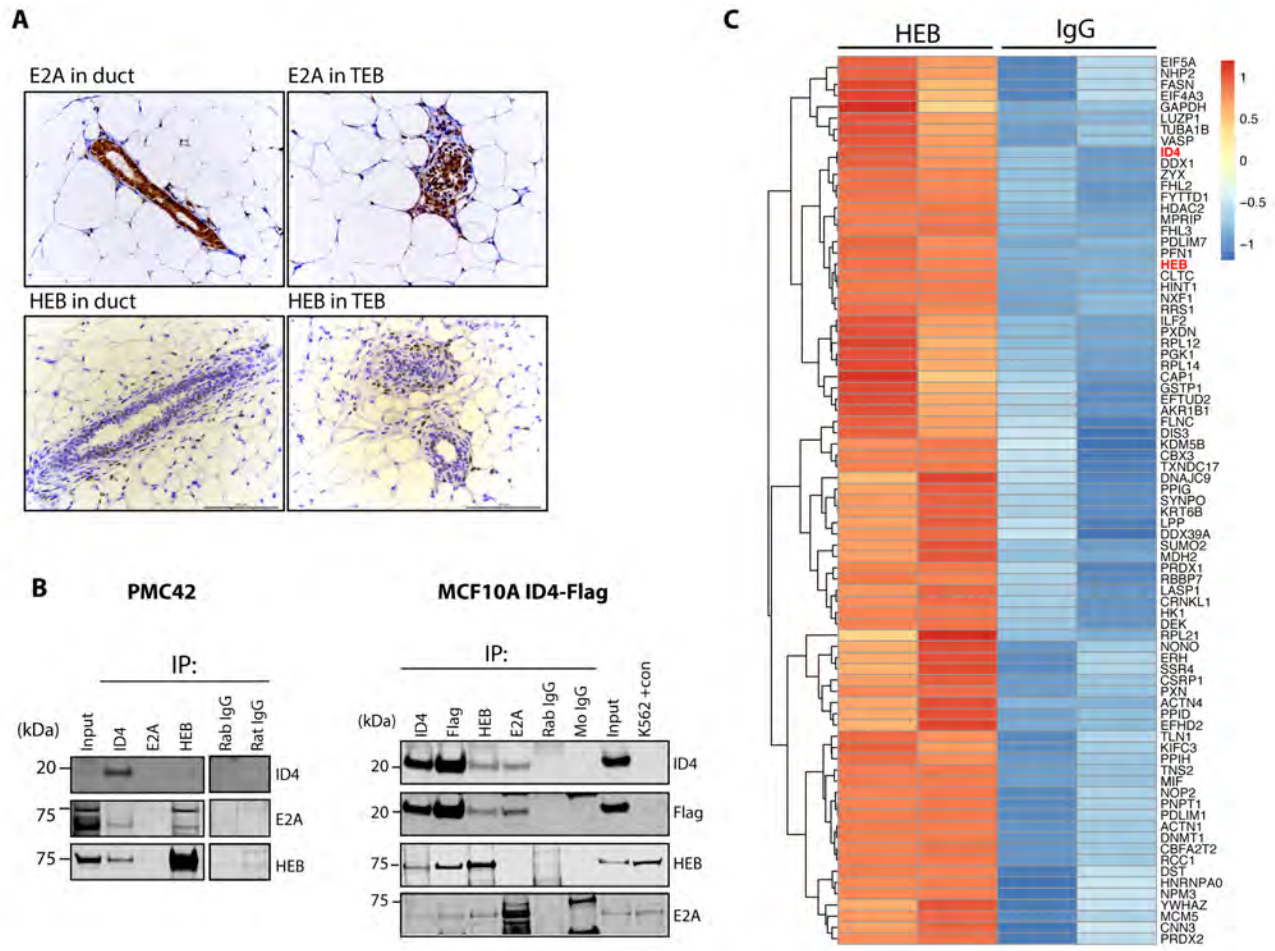


Figure S5. ID4 interacts with E-proteins E2A and HEB, Related to Figure 5.

A) IHC staining for E2A and HEB in murine mammary gland sections. **B)** Co-immunoprecipitation and western blotting of ID4, E2A and HEB from human normal-like mammary epithelial cell lines PMC42 (left) and MCF10A cells overexpressing Flag-tagged ID4 protein (right). **C)** Unsupervised hierarchical clustering heat map of SWATH RIME data from Comma-Dβ cells. Proteins with significantly higher abundance (p-value<0.05) in the HEB IPs compared to IgG IPs are displayed. Log2 protein area was used to generate the heatmap.

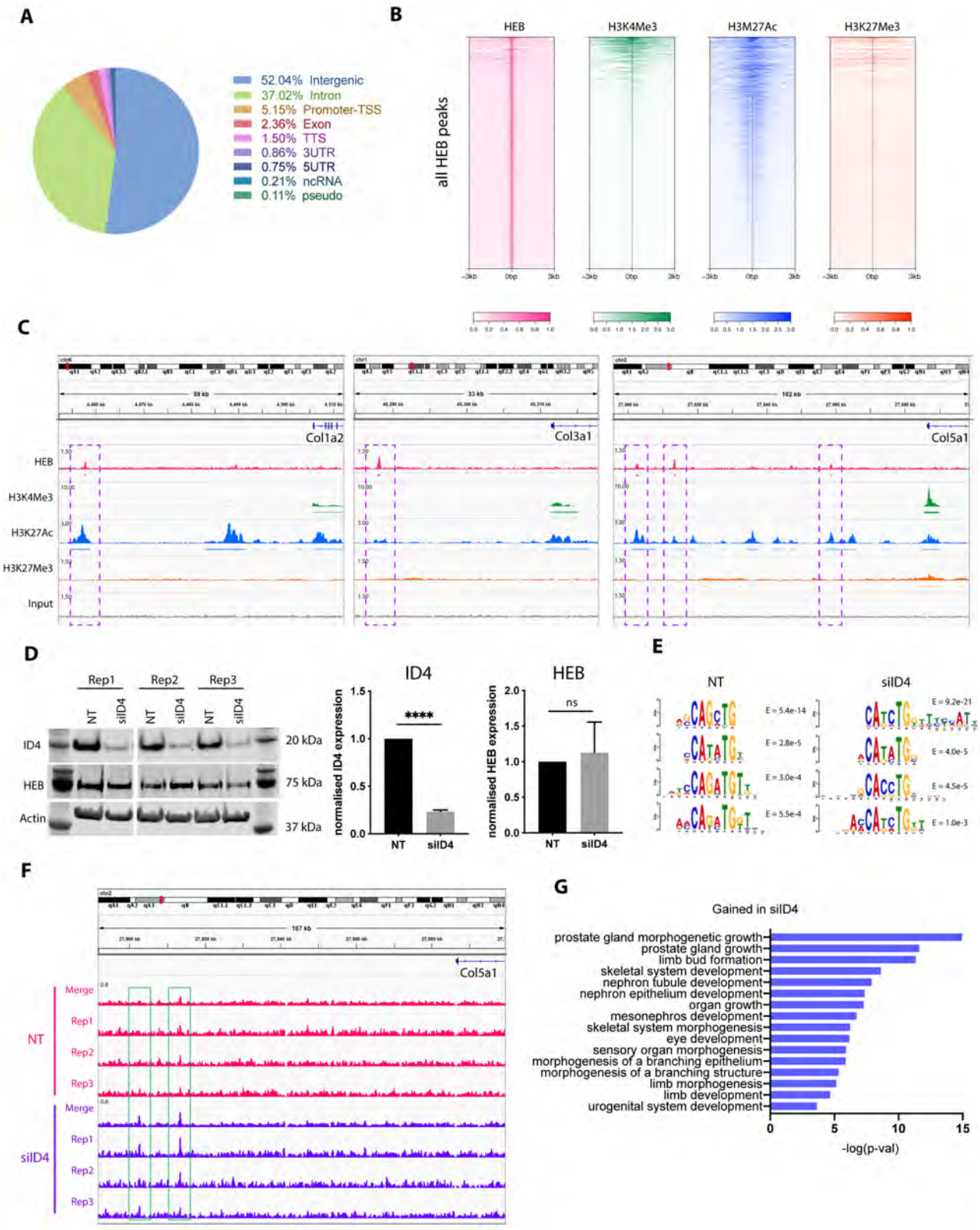


Figure S6. HEB directly binds to a subset of ID4 regulated genes and binding increases upon ID4 knockdown, Related to Figure 6.

A) Genomic distribution of HEB consensus peaks. TSS = transcriptional start site. TTS = transcriptional termination site. **B)** Heatmaps of HEB, H3K4Me3, H3K27Ac, and H3K27Me3 ChIP-seq signal at HEB-

bound regions. **C)** Examples of HEB and histone mark peaks occurring upstream of ID4 repressed genes *Col1a2, Col3a1*, and *Col5a1* (multiple HEB peaks) from the Integrative Genomics Viewer (IGV). Bars beneath peaks represent consensus MACS call (FDR<0.05) in at least 2 of 4 biological replicates. Input was used as a negative control. Purple boxes highlight HEB binding regions. Refseq genes shown in blue. Data scales for each track are indicated. **D)** Western blot analysis of ID4 and HEB expression in Comma-Dβ cells treated with NT or ID4-targeting siRNA. Irrelevant lanes were digitally removed indicated by the gap. Densitometry quantification of ID4 and HEB bands. Band intensity was normalised to β-Actin and expressed as fold change relative to NT control. N=3. Unpaired two-tailed students t-test. Error bars represent SEM. **** p<0.0001. ns = not significant. **E)** Top 4 enriched transcription factor binding motifs determined using MEME-ChIP for consensus HEB ChIP-seq peaks in Comma-Dβ cells treated with NT or ID4-targeting siRNA. E-values are displayed**. F)** Example of HEB peaks in control NT cells and siID4 cells upstream of *Col5a1* from the Integrative Genomics Viewer (IGV). Green boxes highlight HEB bound regions. Individual replicates and merged tracks are displayed. Refseq genes shown in blue. Data scales indicated. **G)** GREAT pathway analysis of 263 peaks increased in siID4 compared to NT from Fig. 6F. Top 16 Gene Ontologies (Biological process, cellular component, molecular function) are displayed.

Transparent Methods

Mice

All mice experiments were performed in accordance with the ethical regulations of the Garvan Institute Animal Experimentation Committee.

ID4 KO mice were generated as previously described (Yun et al. 2004).

ID4-FlagV5 mice were produced by the Mouse Engineering Garvan/ABR (MEGA) Facility using CRISPR/Cas9 gene targeting in C57BL/6J mouse embryos following established molecular and animal husbandry techniques (Yang et al. 2014). A single guide RNA (sgRNA) was produced based on a Cas9 target site that contained the TGA stop codon in exon 2 of *Id4* (TCTCTGCCGC**TGA**GCTGCGA*TGG*) (stop codon, bold underlined; protospacer-associated motif = PAM, italics underlined). sgRNA was microinjected into the nucleus and cytoplasm of C57BL/6J zygotes together with polyadenylated *S.pyogenes* Cas9 mRNA and a 150 base single-stranded, sense, deoxy-oligonucleotide homologous recombination substrate encoding for the insertion of a FLAG-V5 tag (DYKDDDDKGKPIPNPLLGLDST) immediately prior to the stop codon. A founder mouse heterozygous for the desired 66 bp insertion was identified by PCR amplification and Sanger DNA sequencing and the line maintained by backcrossing with inbred C57BL/6J mice.

The ID4floxGFP mice were generated as previously described (Best et al. 2014). All mice used were on the FVB/N background. For the gene expression profiling experiment, mice were synchronised in estrus to reduce hormone-induced gene expression variation (Dalal et al. 2001), and checked by vaginal swab cytology.

Mammary epithelial cell preparations

Mammary epithelial cells were prepared from freshly harvested 3rd and 4th mammary glands pooled from 4-8 female mice at the indicated ages. Glands were mechanically disrupted using a McIlwain tissue chopper then digested with 15,000 U collagenase (Sigma-Aldrich C9891) and 500 U hyaluronidase (Sigma-Aldrich H3506) in FV media (DMEM/F12 (Gibco), 5% (v/v) FBS (HyClone) 10 mM HEPES (Gibco), 0.14 IU/mL Insulin (Novo Nordisk), 500 ng/mL Hydrocortisone (Sigma-Aldrich), 20 ng/mL Cholera Toxin (Sigma-Aldrich), 2 ng/mL mEGF (Life Technologies)). Digestion was carried out in a shaking incubator for 1 hr at 37°C 225 rpm. The resulting organoids were further digested with warm 0.25% trypsin for 1 min with constant pipetting followed by treatment with 5 mg/mL Dispase (Roche 165859) for 5 min in at 37°C. Cells were incubated with 1X red blood cell lysis buffer (BD Biosciences) for 5 min at room temperature. Cells were passed through a 70 µm strainer followed by a 40 µm strainer to remove clumps and debris. Live cells were counted using trypan blue (Thermo Fisher Scientific) and a haemocytometer prior to use in downstream applications.

Flow cytometry and FACS

Single-cell suspensions of mouse mammary epithelial cells were blocked in Fc block cocktail (FACS buffer, 6.25 µg/mL Mouse BD Fc block (BD Bioscience) and 200 µg/mL Rat Gamma Globulin (Jackson ImmunoResearch)) for 10 min on ice. Fluorophore- or biotin-conjugated antibodies were diluted in Fc block cocktail and incubated on the cells on ice for 20 min in the dark. Antibodies used were CD24-PE (1:200, BD Biosciences, Clone M1/69), EPCAM-PerCP/Cy5.5 (1:200, Biolegend, Clone G8.8), CD29-APC/Cy7 (BioLegend, Clone HMβ1-1), CD61-APC (1:50, Thermo Fisher Scientific, Clone HMβ3-1), CD49f-APC (1:100, BioLegend, Clone GoH3), Ter119-biotin (1:80, BD Biosciences, Clone TER119), CD45-biotin (1:100, BD Biosciences, Clone 30-F11), CD31-biotin (1:40, Biolegend, Clone 390) and BP1-biotin (1:500, Thermo Fisher Scientific, Clone 6C3). Cells were washed and stained with streptavidin-BV421 (1:100, Biolegend). Following washing, cells were resuspended in FACS buffer at a density of

$1x10^7$ cells/mL and DAPI was added (1:1000, Invitrogen). FACS was performed on a FACS Aria III 4 laser 15 colour sorter with BD FACS DIVA software. For analytical flow cytometry, a CytoFLEX 3 laser 13 colour flow cytometer (Beckman Coulter) was used. All flow cytometry data was analysed using FlowJo software version 10 (Tree Star Inc).

RNA-sequencing

For the FACS-enriched mammary subpopulation RNA-seq experiment, cells were sorted directly into 700 Qiazol lysis reagent to minimise cell loss, with a maximum of $5x10^4$ cells per 1.5 mL tube and RNA was extracted using the Qiagen miRNeasy micro kit (Qiagen). Four biological replicates were performed. RNA was extracted from Comma-Dβ cells using the Qiagen miRNeasy mini kit (Qiagen) in biological triplicate. The Qubit RNA BR Assay kit (Thermo Fisher Scientific) was used to measure RNA concentration and RNA integrity was determined using the Agilent Bioanalyser 2100 with the 6000 Nano Assay (Agilent Technologies).

For low input RNA extracted from FACS sorted cells, the Ovation RNA-seq System V2 kit (NuGEN) was used to synthesise cDNA with RNA inputs ranging from 0.5-2 ng. The Ovation Ultralow System V2 kit was then used to prepare libraries from the cDNA. For the Comma-Dβ experiment the Illumina TruSeq Stranded mRNA Library Prep Kit (Illumina) was used with 1 μg of input RNA.

cDNA libraries from were sequenced on a HiSeq 2500 system (high output mode) (Illumina), with 125 bp paired-end reads for the FACS experiment or the NextSeq system (Illumina), with 75 bp paired-end reads for the Comma-Dβ experiment.

Quality control was checked using FastQC (Andrews 2010) to remove poor quality reads. Illumina sequencing adapters were then trimmed using Cutadapt (Martin 2011). Reads were then aligned to the mouse reference genome mm10 using STAR ultrafast universal RNA-seq aligner (Dobin et al. 2013). RSEM accurate transcript quantification for RNA-seq data was used to generate a gene read

count table and to filter genes with read counts of 0 (Li and Dewey 2011).

Differential gene expression analysis was performed using edgeR (Robinson et al. 2010; McCarthy et al. 2012) and voom/limma (Ritchie et al. 2015) R packages. Genes were ranked based on the limma moderated t-statistic and this was used as input GSEA pre-ranked (Subramanian et al. 2005) using Molecular Signature Database (MSigDB) collections (v6.2). The EnrichmentMap Cytoscape plugin (Merico et al. 2010) was used to visualise the GSEA results.

Analysis of public single-cell RNA-seq data

Raw Unique Molecular Identifier (UMI) expression data was taken from Bach *et al.* and re-processed using Seurat v2 using default parameters, as recommended by the developers (Satija et al. 2015). All basal cell clusters, defined by *Krt5* and *Krt14* expression, were extracted for subsequent analysis. Developmental trajectories were inferred using Monocle 2 (Qiu et al. 2017). Clustering was first performed using the 'densityPeak' and 'DDRTree' methods, with a delta local distance threshold of 2, and a rho local density threshold of 3. For inferring pseudotime, we first selected all genes with a mean expression greater than 0.5, and an empirical dispersion greater than 1, and proceeded with the top 1000 significant genes for ordering. Differential gene expression between the two branch states were performed using the MAST algorithm through the Seurat package (Finak et al. 2015; Satija et al. 2015). For comparisons between *Id4*-high and *Id4*-low states, we first filtered for cells with the detection of *Id4* to avoid technical biases from gene drop out. Differential gene expression was performed (as described above) between the top and bottom 200 cells grouped based on log normalised gene expression of *Id4*.

Immunohistochemistry and immunofluorescence on mammary tissue sections

IHC staining was performed on 4 μm sections of formalin-fixed paraffin-embedded (FFPE) tissue.

Slides were dewaxed with xylene and hydrated through graded alcohols. Antigen retrieval was performed for 1 min in a pressure cooker in DAKO target retrieval reagent s1699. A DAKO Autostainer was used for subsequent steps. Briefly, slides were incubated with DAKO peroxide block for 5 min and blocked with DAKO protein block for 30 min. Primary antibody (ID4 1:400 Biocheck BCH9/82-12, E2A 1:100 abcam 69999, HEB ProteinTech 21073-1-AP 1:250) was incubated on the slides for 60 min. Following washing, slides were incubated with Envision rabbit secondary antibody (Agilent Technologies, Santa Clara, CA, USA) for 30 min. Slides were washed then incubated with DAKO DAB+ (Agilent Technologies) reagent for 10 min. Slides were then rinsed and counterstained with haematoxylin for 20-30 sec and dehydrated through graded alcohols, cleared using xylene and mounted using Ultramount #4 (Fronine). Bright-field images were captured using a Leica DM 4000 microscope with high-resolution colour camera (DFC450).

Immunofluorescence was performed manually on antigen retrieved tissue sections prepared as described for IHC. Slides were blocked with Mouse on Mouse (MOM) blocking buffer (Vector Biolabs) for 1 hour. Primary antibody was diluted in MOM diluent and incubated overnight at 4°C. Antibodies used were ID4 (1:400, Biocheck BCH9/82-12), α-SMA (1:200, Sigma-Aldrich A5228), HEB (1:200, Protein Tech 14419-1-AP), V5 (1:200, Santa Cruz sc-58052). Following washing, slides were incubated with fluorescent secondary antibody (1:500, Jackson ImmunoResearch) for 1 hr at room temperature. Nuclei were stained using Hoechst 33342 (Sigma-Aldrich). Slides were mounted using Prolong Diamond mounting media (Thermo Fisher Scientific). Fluorescent images were captured using a Leica DM 5500 microscope.

For quantification of fluorescence of ID4 and α-SMA in cap and ductal basal cells, 3 representative ducts and TEBs were imaged from each mouse (n=9). Using FIJI software, a region of interest was drawn around 10 randomly selected basal cells from each image. The mean fluorescence of ID4 and α-SMA was determined within each cell.

Picrosirius red staining

Picrosirius red staining was performed following as per (Vennin et al. 2017; Vennin et al. 2019). Briefly, dewaxed FFPE sections were stained with haematoxylin. Sections were then treated with 0.2% Phosphomolybdic acid (Sigma-Aldrich) followed by 0.1% Sirius red (Sigma-Aldrich) Picric Acid-Saturated Solution (Sigma-Aldrich). Slides were then rinsed with acidified water (90 mM glacial acetic acid) then with 70% ethanol. A Leica DM 4000 microscope was used to image total collagen staining in the tissue sections by phase contrast. To image birefringence of collagen fibres, 2 polarised filters were used. Matched phase-contrast and polarised images were taken for each region of interest. Image analysis was performed using FIJI image analysis software.

Proximity Ligation Assay

Comma-Dβ cells were grown on glass coverslips in 6 well plates until ~80% confluent. Cells were fixed in 4% Paraformaldehyde (PFA) (ProSciTech) for 15 min then permeablised in 0.2% Triton-X-100 for 15 min. The Duolink Proximity Ligation Assay (PLA) (Sigma-Aldrich) was performed as per the manufacturer's protocol and using the following antibodies: ID4 (1:50, Santa Cruz sc-365656) and HEB (1:200, Protein Tech 14419-1-AP), and equivalent concentrations of IgG negative controls (sc-2027 and sc-2025). DAPI (Life Technologies) and Phalloidin (Life Technologies) were added to the final wash step. Coverslips were mounted onto glass slides with Prolong Diamond mountant (Thermo Fisher Scientific).

Cells were imaged using a Leica DMI Sp8 confocal microscope (63X oil objective). Six random fields of view per image were captured, each with approximately 50 cells, and images were quantified using Andy's algorithms FIJI package (Law et al. 2017) to enumerate the number of PLA foci per nuclei.

Cell lines

The mouse mammary epithelial cell line Comma-Dβ was a gift was a gift from Joseph Jeffery (University of Massachusetts, Amherst, MA, USA). Comma-Dβ cells were maintained in DMEM/F12 media (Gibco) supplemented with 2% FBS (HyClone), 10 mM HEPES (Gibco), 0.125 IU/mL Insulin (Novo Nordisk) and 5 ng/mL mEGF (Life Technologies). The human mammary epithelial cell line PMC42 was a gift from Professor Leigh Ackland (Deakin University, Melbourne, Victoria, Australia). PMC42 cells were maintained in RPMI 1640 (Gibco) supplemented with 10% FBS (HyClone). The MCF10A cell line was obtained from the American Type Culture Collection and were maintained in DMEM/F12 (Gibco), 5% Horse Serum (Thermo Fisher Scientific), 20 ng/mL hEGF (In Vitro Technologies), 0.5 mg/mL Hydrocortison (Sigma-Aldrich), 100 ng/mL Cholera Toxin (Sigma-Aldrich) and 0.125 IU/mL Insulin (Novo Nordisk).

Overexpression of ID4

Comma-Dβ cells ($1.1x10^5$) were seeded into a 6-well plate. 16-24 hours later the cells were infected with pMSCV-Id4-DSred or pMSCV-DSred retrovirus diluted 1:10 in Comma-D media with 8 µg/ml polybrene. 24 hours later the media was changed. DSred positive cells were then FACS enriched using the BD FACSAria fluorescence activated cell sorter and BD FACSDIVA software.

Conditional reprogramming of primary mouse basal cells

Viable basal cells were purified by FACS from 10-12 week-old female mice as described above. Cells were collected into FAD media (DMEM/F12 3:1 (Gibco) supplemented with 10% FBS, 0.18 mM Adenine (Sigma-Aldrich), 500 ng/mL Hydrocortisone (Sigma-Aldrich), 8.5 ng/mL Cholera Toxin (Sigma-Aldrich), 10 ng/mL mEGF (Life Technologies), 0.14 IU/mL Insulin (Novo Nordisk), 5 µM Y-27632 (Seleckchem), 1X AB/AM (Gibco) and 50 µg/mL Gentamicin (Life Technologies). Basal cells

were maintained in culture using as per (Prater et al. 2014). Tissue culture flasks coated with Growth Factor-reduced Matrigel (Corning) diluted 1:60 in PBS for 30 min to 1 hr at 37°C. Excess Matrigel/PBS solution was aspirated from the flasks and basal cells were seeded at a density of ~5000 cells/cm^2. Basal cells were co-cultured with irradiated (50 Gy) NIH-3T3 feeder cells at a density of 10,000 cells/cm^2. Cells were maintained in a 37°C low 5% oxygen 5% CO_2 incubator. Differential trypsinisation was used to first remove the less-adherent 3T3 cells when passaging.

Cre mediated deletion of ID4 from primary cells

1.0-1.5x10^5 ID4floxGFP primary basal cells were seeded into a T75 with 7.5x10^5 irradiated 3T3 cells in 20 mL FAD media. Adenovirus was added to the media at a multiplicity of infection (MOI) of 100. Adenoviruses' used were codon optimised Cre (iCre) and GFP adenovirus (Vector Biolabs 1772) and control eGFP adenovirus (Vector Biolabs 1060). Infection efficiency was assessed the following day using a fluorescence microscope to check GFP expression within the cells. Cells were harvested after 72 hr for downstream experiments and analysis.

Organoid culture

A 40 µL plug of Matrigel (Corning) was added to each well of 8-well chamber slides (Corning) and allowed to set for 30 min in a 37°C incubator. Primary cells were resuspended in Epicult-B (Stemcell Technologies) media containing 2% Matrigel and 12,000 cells were seeded into each chamber in a total volume of 400 µL. Chambers were observed using a light microscope to ensure that cells were not aggregated. Organoids were allowed to form over 1 week and media was changed every 3 days. Organoids from each chamber were photographed using an inverted epifluorescence microscope (4x objective) and the average organoid area per image was determined using Andy's algorithms FIJI package (Law et al. 2017).

Immunofluorescent staining of organoids

Organoids were stained within the chamber slides. Media was aspirated and organoids were fixed with 2% PFA diluted in PBS for 20 min at room temperature. Organoids rinsed with PBS for 5 min following permeabilisation with 0.5% Triton X-100 in PBS for 10 min at 4°C then rinsed 3 times with 100 mM Glycine (Astral Scientific) in PBS for 10 min each. Organoids were blocked for 1 hr in IF buffer with 10% goat serum (Vector Laboratories). Primary antibody made up in blocking buffer was incubated in the chambers overnight at 4°C in a humidified chamber. Antibodies used were ID4 (1:200, Biocheck BCH9/82-12), α-SMA (1:100, Abcam ab5694), KRT14 (1:1000, Covance PRB-155P), KRT8 (1:500, DSHB TROMA1) and P63 (1:100, Novus NB100-691). The following day, slides were equilibrated to room temperature for 1-2 hr. Organoids were rinsed 2 times with IF buffer (0.1% BSA, 0.2% Triton-X-100, 0.05% Tween-20 in PBS) for 20 min each. Fluorescent secondary antibody (Jackson ImmunoResearch) diluted in blocking buffer (1:500) was added to the chambers and incubated for 45 min in the dark. Secondary antibody was rinsed off for 20 min in IF buffer, then 2 times with PBS for 10 min each. DAPI (Life Technologies) was incubated in the chambers for 15 min followed by a 10 min PBS rinse. The walls of the chamber slides were removed and one drop of Prolong Diamond mounting media (Thermo Fisher Scientific) was added to each well. Slides were coverslipped and edges sealed with clear nail varnish. Slides were allowed to dry for 24 hr at room temperature protected from light.

Organoids were imaged using a Leica DMI Sp8 confocal microscope using the 40X oil objective and 3X optical zoom. The FIJI software package was used to quantify the fluorescence within the organoids. For α-SMA, a circle was drawn around the entire organoid and fluorescence measured. For ID4, a mask was made using the DAPI channel and signal was measured within the nuclear mask. Multiple representative regions of background staining were quantified. Corrected fluorescence (CF)

was calculated as described in (McCloy et al. 2014). CF = integrated density – (area of selected organoid * mean fluorescence of all background readings)

siRNA transfections

Comma-Dβ cells were seeded at a density of $1.5x10^4$ cells/cm^2 into 6 well plates (Corning) for RNA-seq or 100 mm dishes (Corning) for ChIP-seq in antibiotic-free media. The following day siRNA constructs (Dharmacon) were transfected into the cells using Dharmafect-4 transfection reagent (Dharmacon) as per the manufacturer's instructions at 20 nM. siRNA used were siGENOME Mouse Id4 SMARTpool (M-043687) and ON-TARGETplus Non-targeting Control Pool (D-001810). Media was changed the following day and cells were harvested 48 hr post-transfection.

Western blotting

Cells were lysed in RIPA buffer (50 mM Tris-HCl pH 7.4, 1% NP-40, 0.5% Sodium Deoxycholate, 0.1% SDS, 1% Glycerol, 137.5 mM NaCl, 100 µM Sodium Orthovanadate, 20 µM MG132, 1 mM DTT and 1x cOmplete ULTRA Tablet (Roche)) and protein concentration was quantified using the Pierce BCA Protein Assay Kit (Thermo Fisher Scientific) according to the manufacturer's instructions. Protein lysates (15-30 µg protein) were mixed with 1X NuPage loading buffer (Life Technologies) and 1X NuPage reducing agent (Life Technologies) and denatured by heating at 85°C for 5 min. Samples were run on a 4-12% Bis/Tris gels (Life Technologies) in MES or MOPS running buffer (Life Technologies). Protein was transferred to a 0.45 µm PVDF membrane (Merck Millipore) using BioRad transfer modules in transfer buffer (25 mM Tris, 192 mM Glycine, 20% Methanol). Membranes were blocked in Odyssey blocking buffer (LiCOR) for 1 hr at room temperature then incubated in primary antibody diluted in 5% BSA/TBS overnight at 4°C. The following antibodies were used for western blotting: ID4 (1:20,000, Biocheck BCH9/82-12), E2A (1:1000, from Dr Nicolas Huntington from the Walter Eliza Hall

Institute (WEHI)), HEB (1:1000, from Dr Nicolas Huntington, WEHI or 1:1000 ProteinTech 144191-1-AP), Flag (1:5000, Sigma-Aldrich F1804), V5 (1:200, sc-58052), α-SMA (1:1000, Abcam ab5694), SNAIL (Cell Signalling Technology 3879), SLUG (Cell Signalling Technology 9585), CNN2 (ProteinTech 21073-1-AP), ZEB1 (Cell Signalling Technology 3396), β-Actin (1:1000, Sigma-Aldrich A5441) and α-Tubulin (1:1000, Santa Cruz sc-5286). Fluorescent secondary antibody conjugated to IRDye680 or IRDye800 (LiCOR) diluted in Odyssey blocking buffer (1:15,000 - 1:20,000) were used for detection. An Odyssey CLx Infrared Imaging System (LiCOR) was used to image and quantify the western blots.

Co-immunoprecipitation

Pierce Protein A/G magnetic breads (Thermo Fisher Scientific) were incubated with antibody for a minimum of 4 hr at 4°C on a rotating platform. For each IP 15 µL of beads and 2.5 µg antibody were used. Immunoprecipitating antibodies used were ID4 (pool of sc-491 and sc-13047), E2A (from Dr Nicolas Huntington, WEHI), HEB (Santa Cruz sc-357 and antibody from Dr Nicolas Huntington, WEHI), Flag (Sigma-Aldrich F1804), V5 (Santa Cruz sc-58052 and Thermo Fisher R960-25), species matched IgG negative controls (sc-2027, sc-2025 and BioLegend 400602). Protein was extracted in IP lysis buffer (10% Glycerol, 0.03% $MgCl_2$, 1.2% HEPES, 1% Sodium acid pyrophosphate, 1% Triton-X, 0.8% NaCl, 0.4% NaF, 0.04% EGTA, 1x cOmplete ULTRA Tablet (Roche), 100 µM Sodium Orthovanadate, 20 µM MG132, 1 mM DTT and cells were passed 5 times through a 23 g needle to aid in lysis. Protein lysates (0.5-2 mg per IP) were added to washed beads and incubated overnight at 4°C on a rotating platform. Following washing, beads were resuspended in 2X NuPage loading buffer (Life Technologies) and 2X reducing agent (Life Technologies) in IP lysis buffer. Samples were incubated at 85°C for 5 min. Beads were separated on a magnetic rack and supernatant loaded onto NuPage gel for SDS-PAGE and western blotting. For detection of protein, fluorescent TrueBlot secondary antibodies were used (Jomar Life Research).

RIME

The RIME protocol was adapted from the protocol developed by Mohammed *et al.* (Mohammed et al. 2013). Comma-Dβ cells were grown in 150 mm tissue culture dishes (Corning) until 80-90% confluent. A total of 8x 150 mm dishes were used per RIME sample. Cells were fixed with 1% PFA (ProSciTech) in DMEM/F12 (Gibco) for 7 min at room temperature on a rocking platform. Cross-linking was quenched by addition of molecular grade glycine (Astral Scientific) to a final concentration of 125 mM for 2 min on a rocking platform. Cells were washed twice with ice-cold PBS (Gibco) and scraped in 1 mL of ice-cold PBS containing magnesium and calcium salts (Gibco). Cells were centrifuged at 4°C for 5 min at 1200 rpm. Cell nuclei were enriched through a series of lysis buffers. Cells were first resuspended in 10 mL lysis buffer 1 (50 mM HEPES-KOH pH 7.5, 140 mM Sodium, 1 mM EDTA, 10% Glycerol, 0.5% NP-40 or Igepal CA-630, 0.25% Triton X-100, 100 µM Sodium Orthovanadate, 20 µM MG132, 1 mM DTT and 1x cOmplete ULTRA Tablet (Roche)) and incubated for 30 min at 4°C on a rotating platform. Cells were then centrifuged at 3000 g for 3 min at 4°C and supernatant removed. The pellet was then resuspended in 10 mL lysis buffer 2 (10 mM Tris-HCl pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 100 µM Sodium Orthovanadate, 20 µM MG132, 1 mM DTT and 1x cOmplete ULTRA Tablet (Roche)) and incubated for a further 30 min at 4°C on a rotating platform. Following centrifugation, the resulting nuclei pellet was lysed in 2.5 mL lysis buffer 3 (10 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% Sodium Deoxycholate, 0.5% N-lauroylsarcosine, 100 µM Sodium Orthovanadate, 20 µM MG132, 1 mM DTT and 1x cOmplete ULTRA Tablet (Roche)). The nuclear lysate was sonicated using a Bioruptor sonicator (Diagenode) for 20 cycles of 30 sec on/30 sec off. Triton-X-100 was added to a final concentration of 1% and the sample was then centrifuged at maximum speed for 5 min at 4°C to remove cellular debris. Immunoprecipitation was carried out from the sheared nuclear supernatant described above with

20 µg of antibody (ID4 sc-491 and sc-13047 pool, HEB sc-357) and 100 µL of magnetic beads per sample. The following day, the beads were washed 10 times with RIPA buffer then 5 times with 100 mM ammonium hydrogen carbonate (AMBIC) (Sigma-Aldrich) solution to remove salts and detergents, resuspended in 50 µL AMBIC solution and transferred into clean tubes.

Mass Spectrometry

Samples were processed as described in (Huang et al. 2015) for liquid chromatography coupled mass spectrometry (LC-MS/MS) using Sequential Windowed Acquisition of all THeoretical fragment-ion spectra (SWATH) acquisition. Briefly, samples were denatured in 100 mM triethylammonium bicarbonate and 1% sodium deoxycholate, disulfide bonds were reduced in 10 mM DTT, alkylated in 20 mM iodo acetamide, and proteins digested on the magnetic beads using trypsin. After C18 reversed phase (RP) StageTip sample clean up, peptides were analysed using a TripleToF 6600 mass spectrometer (SCIEX, MA, USA) coupled to a nanoLC Ultra 2D HPLC system (SCIEX). Peptides were separated for 60 min using a 15 cm chip column (ChromXP C18, 3 µm, 120 Å) (SCIEX) with an acetonitrile gradient from 3-35%. The MS was operated in positive ion mode using either a data dependent acquisition method (DDA) or SWATH acquisition mode. DDA was performed of the top 20 most intense precursors with charge stages from 2+ - 4+ with a dynamic exclusion of 30 s. SWATH-MS was acquired using 100 variable size precursor windows. DDA files were searched using ProteinPilot 5.0 (SCIEX) against the reviewed UniProt *Mus musculus* protein database (release February 2016) using an unused score of 0.05 with decoy search strategy enabled. These search outputs were used to generate a spectral library for targeted information extraction from SWATH-MS data files using PeakView v2.1 with SWATH MicroApp v2.0 (SCIEX) importing only peptides with < 1% FDR. Protein areas, summed chromatographic area under the curve of peptides with extraction FDR ≤ 1%, were calculated and used to compare protein abundances between bait and control IPs.

ChIP-seq

ChIP-seq was adapted from (Khoury et al. 2020). Comma-Dβ cells were grown in 150 mm or 100 mm dishes (Corning) until 80-90% confluent. For the ChIP-seq experiment on unperturbed cells 2x 150 mm dishes were used for HEB ChIP and 1x 150 mm dishes were used for histone marks. Four independent replicates were conducted. For the ID4 siRNA experiment 3x 100 mm dishes per condition were used and the experiment was repeated 3 times. Cells were scraped in ice-cold PBS (Gibco) and resuspended by passing 10 times through a 19 g syringe. Cells were fixed in 1% PFA (ProSciTech) in PBS (Gibco) for 15 min at room temperature followed by quenching with glycine to a final concentration of 125 mM for 5 minutes followed by two PBS washes. Nuclei was extracted by resuspending cells in nuclei extraction buffer (10 mM Tris-HCl pH 7.5, 10 mM NaCl, 3 mM $MgCl_2$, 0.1 mM EDTA, 0.5% IGEPAL, 1x cOmplete ULTRA Tablet (Roche)) and incubated on ice for 10 min followed by 20 passes on a tight Dounce homogeniser, or until nuclei were extracted. Nuclei were visually inspected using trypan blue and a haemocytometer. Nuclei were pelleted by centrifugation and resuspended in sonication buffer (50 mM Tris-HCl pH 8, 1% SDS, 10 mM EDTA, 1x cOmplete ULTRA Tablet (Roche)) and sonicated using a Bioruptor sonicator (Diagenode) to achieve DNA fragments between 100-500 bp with a mean fragment size of approximately 300 bp. In the siRNA experiment, protein was quantified using the Pierce BCA assay to load equal amounts of input material for NT and ID4 siRNA for immunoprecipitation. Sonicated sample was diluted with IP dilution buffer to 1 mL (16.7 mM Tris-HCl pH 8, 0.01% SDS, 1% Triton X-100, 167 mM NaCl, 1.2 mM EDTA). Samples were cleared with protein A/G magnetic beads for 1.5 hr at 4°C. 1% of the cleared nuclear lysate was removed for the input control and immunoprecipitation was performed as described above. For HEB ChIP 100 µL of beads and 20 µg of HEB antibody (sc-58052) was used with chromatin from approximately 20-30x10^6 cells. For histone mark ChIP, 50 µL beads with 10 µg of the following

antibodies – H3K4Me3 (Active Motif 28431), H3K27Me3 (Merck Millipore 07-449), was used with chromatin from approximately 10-15x10^6 cells. The following day, beads were washed for 5 min each with 1 mL of the following buffers: Low salt buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl, pH 8, 150 mM NaCl), High salt buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl, pH 8, 500 mM NaCl), LiCl buffer (0.25 M LiCl, 1% IGEPAL, 1% deoxycholic acid (sodium salt), 1 mM EDTA, 10 mM Tris, pH 8) followed by 2 washes with TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8). DNA was eluted twice in 100 uL ChIP elution buffer (1% SDS, 0.1 M Sodium Biocarbonate) for 15 minutes at room temperature each. Cross-linking was reversed overnight by treating samples with 200 mM NaCl and 250 µg/mL Proteinase K (New England Biolabs) and incubating overnight at 65°C. The following day samples were treated with 100 µg/mL RNAse A (Qiagen) for 1 hr at 37°C. DNA was purified using Phase Lock Gel Light tubes (Quantabio 5Prime) according to the manufacturer's instructions. DNA was eluted in 20 uL nuclease-free TE buffer, pH8 (Qiagen).

DNA concentration was measured using the Qubit HS Assay Kit (Thermo Fisher Scientific). Libraries were prepared using the Illumina TruSeq ChIP library prep kit (Illumina) following the manufacturer's instructions except the gel purification step was replaced with a two-sided AMPure XP bead (Beckman Coulter) size selection to obtain libraries between 200-500 bp. Library sizes were verified using the 4200 TapeStation System (Agilent) with a D1000 ScreenTape (Agilent) and concentration was determined using the Qubit HS Assay Kit (Thermo Fisher Scientific). ChIP libraries were sequenced on the NextSeq system (Illumina), with 75 bp paired-end reads.

Reads were aligned with BWA (Li and Durbin 2009) and all reads with a MAPQ<15 were removed. Alignment statistics were generated using Samtools Flagstat (Li et al. 2009). ChIP-seq peaks were called using the peak calling algorithm MACS (Zhang et al. 2008) and ENCODE blacklist regions were removed. For unperturbed cells, consensus peaks present in at least 2 of 4 replicates were used for downstream analysis. Due to lower amount of input material available from the siRNA experiment,

less consensus peaks were called and for this reason analysis was conducted on merged peaks from the 3 replicates. Motif enrichment analysis was performed using MEME-ChIP (Machanick and Bailey 2011). Peaks were annotated to genomic features using HOMER (Heinz et al. 2010). GREAT was used for functional enrichment analysis (v4.0.4) and gene annotation using the default parameters (McLean et al. 2010). SeqPlots (Stempor and Ahringer 2016) and IGV (Robinson et al. 2011) software were used for data visualisation. Differential binding analysis was performed using the DiffBind package (Stark 2011). For overlapping of siID4 RNA-seq and HEB ChIP-seq genes, a hypergeometric test was used to determine if overlap was significant, assuming 30,000 genes in the mouse genome.