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], Alexander Swarbrick[1,2*]

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

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

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 antagonises HEB transcriptional activity, preventing myoepithelial differentiation and allowing for appropriate tissue morphogenesis. Downregulation of ID4 during pregnancy modulates gene regulated by HEB, promoting specialisation of basal cells into myoepithelial cells.

Keywords

Differentiation, Mammary gland, Myoepithelial, Stem cell, Transcription factor

Introduction

The mammary gland undergoes tissue remodelling 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), while 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 specialised myoepithelial cells, smooth muscle-like epithelial cells which 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, Davis et al., 2016, Scheele et al., 2017, Wuidart et al., 2018, 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., 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. While 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 specialisation 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 linages (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) (Fig. 1A). No significant differences were observed in the proportion of mammary epithelial subpopulations between WT and KO mice (Fig. S1A). 104 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 (Fig. 1B-C 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 (Fig. 1D and Fig. S1B-C). 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 (Fig. 1E and Fig. S1B). These pathways shared common genes as indicated by the connecting edges in the enrichment map network (Fig. 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 specialisation.

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 characterise 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 (Fig. 2A). ID4-GFP expression within the basal gate was binned into 3 groups: bright (top 10%), intermediate (middle 80%) and dim (bottom 10%) and the median fluorescent intensity (MFI) of EPCAM and CD49f was analysed within these groups (Fig. 2B-C). ID4-bright cells had significantly higher EPCAM and CD49f MFI than ID4-dim and intermediate cells, suggesting ID4 marks basal stem cells.

During ductal elongation at puberty, the cap cells differentiate at the neck of the TEBs and mature into myoepithelial cells which 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 (Fig. 2D and Fig. 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 to the more separated elongated morphology of the ID4-low duct cells (Fig. 2D and Fig. S2A). Quantification of fluorescence demonstrated that ID4 was more highly expressed in cap cells of TEBs than in basal cells of ducts, while the opposite was true for α-SMA expression (Fig. 2E-F). A negative correlation between α-SMA and ID4 was observed, with clear separation between cap (dark blue Fig. 2G) and duct basal cells (light blue Fig. 2G). As a negative control, α-SMA fluorescence was compared with nuclear stain Hoechst and no correlation or separation based on region was observed (Fig. 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) data-set (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 week), 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 (Fig. 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) (Fig. 2J-K), 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 (Fig. 2J-K). This result was validated on the protein level by immunohistochemical staining for ID4 on mammary gland sections at different developmental time points (Fig. 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 (Fig. 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, while basal cells at lactation were the most differentiated. *Id4* expression decreased, while several myoepithelial markers (*Acta2, Cnn1, Mylk, Myh11* and *Oxtr*) increased over pseuodotime (Fig. 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 (Fig 2L and Table S2). Basal cells with high *Id4* expression were enriched for genes involved in RNA binding, metabolic processes, translation and ribosome biogenesis (Fig. 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 (Fig. 2N). Myoepithelial genes such as *Oxtr,* encoding the oxytocin receptor, and contractile genes *Mylk, Cnn1, Cav1* and *Myh11,* were among the top differentially expressed genes in the cells with low *Id4* expression (Fig. 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 (Fig. 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 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) (Fig. 3A). Conditionally reprogrammed basal cells adopted an epithelial cobblestone morphology and expressed high levels of ID4 (Fig. S3A-B). Cells cultured in the absence of ROCK inhibitor or feeders adopted a flattened differentiated/senescent cell morphology and had reduced ID4 expression (Fig. S3a-b). Reprogrammed cells maintained basal marker expression of P63 and KRT14 (Fig.

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 (Fig. S3D).

ID4 expression was limited to the outer basal cells (Fig. 3B), recapitulating expression of ID4 in cap cells of TEBs (Fig. 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 (Fig. 3B-C). Loss of ID4 resulted in slightly smaller organoids with increased α-SMA fluorescent intensity (Fig. 3D-F). 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 hr after siRNA transfection (Fig. 4A). RNA-seq analysis of ID4 KD cells resulted in 471 and 421 (FDR<0.05) down and

upregulated genes, respectively, compared to the non-targeting siRNA control (Table S3).

Downregulated genes were predominantly involved in cell proliferation and growth pathways (Fig. 4B-C; blue) including hallmark gene sets such as E2F targets, MYC targets, and G2M checkpoint (Fig. S4A-B). Driving enrichment 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* (Fig. 1D). Genes acutely upregulated upon ID4 depletion were involved in development, morphogenesis, ECM remodelling, and immune signalling (Fig. 4B-C; red). Consistent with ID4 repressing myoepithelial specialisation, loss of ID4 resulted in upregulation of SRF targets, actomyosin cytoskeleton, EMT and myogenesis gene signatures (Fig. 4D and Fig. S4A-B). 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 (Fig. 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 (Fig. 4E i-ii), consistent with the upregulation of this marker in the primary organoid culture upon loss of ID4 (Fig. 3E-F). CNN2, another smooth muscle contractile protein, as well as classical EMT markers ZEB1 and SLUG, were also suppressed by ID4 (Fig. 4E i-ii). Finally, morphological inspection of ID4 overexpressing

cells revealed a cobblestone epithelial appearance compared to 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 (Fig. 4C-D 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 visualise fibrillar collagen (Fig. 4F). ID4 KO TEBs in pubertal mammary glands were surrounded by a thickened collagen-dense ECM when compared to ID4 WT TEBs (Fig. 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 fibres can be further distinguished using polarized light. Analysis of birefringence signal revealed an increase in thick fibres and a decrease in thin fibres in the ECM surrounding ID4 KO TEBs (Fig. 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 to the IgG negative control IPs in three independent RIME experiments (Fig. 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 (Fig. 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 dimerise 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 (Fig. 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 (Fig. 5B and S5B). E2A and HEB did not form heterodimers with each other in Comma-Dβ cells (Fig. 5B), implying that E-proteins either function as homodimers, or 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 (Fig. S5C). However, we did not identify any other bHLH transcription factors in this experiment (Fig. 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 (Fig. 5C-D).

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, while HEB was more ubiquitous in its expression; detected in basal and luminal epithelial cells and surrounding stromal cells (Fig. 5E). ID4 and HEB co-localized in the nuclei of cap and duct basal cells (insets; Fig. 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 (Fig. 5F). Each antibody tested 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 2752 HEB peaks identified (FDR<0.05). This was narrowed down to 956 consensus peaks which were present in at least 2 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 (Fig. 6A). The majority of peaks were mapped to intergenic and intronic regions, and approximately 5% of the peaks occurred at gene promoters (Fig. S6A). We used the Genomics Regions Enrichment of Annotations Tool (GREAT) to analyse 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 (Fig. 6B), resembling the pathways negatively regulated by ID4 in the gene expression profiling (Fig. 4B). This suggests that ID4 mediates repression though its physical interaction with HEB.

The consensus peaks were annotated to 1320 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 (Fig. 6C). Approximately 10% of these genes had an associated HEB peak, which is more than expected by chance ($p<5.65E-13$; 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 (Fig. 6D and Fig. S6B) suggesting that HEB binds to enhancers. Some of the peaks were also localized to active promoters (Fig. 6D and Fig. 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* (Fig. 6E and Fig. S6C) (Barsky and Karlin, 2005). HEB binding was also observed near the promoter of contractile gene *Cnn2* (Fig. 6E), whose RNA and

protein product CNN2 was suppressed ID4 (Table S3 and Fig. 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, while HEB expression was unchanged (Fig. S6D). E-box motifs were again enriched in both conditions (Fig. 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 to 27) suggesting that depletion of ID4 increased HEB's DNA binding activity (Fig. 6F and Fig. S6F). GREAT analysis revealed that the peaks that increased were involved in processes such as gland morphogenesis, skeletal development, and branching morphogenesis (Fig. 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 (Fig. 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 antagonised by its interaction with ID4 (Fig. 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), haematopoietic (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 Fig. 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 (Fig. 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 (Fig. 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). While 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 hypothesise 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 which 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 tumour 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. Additionally, the suppression of ECM synthesis may allow tumour 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 tumour progression.

Limitations of 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 (Fig. S5 and Fig. 5E-F).

Resource Availability

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 RNA-seq and ChIP-seq data have been deposited to GEO repository with the accession number GSE149969. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD017517.

Acknowledgements

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), Love Your Sister in association with the National Breast Cancer Foundation (NBCF) and Susan G Komen.

Author Contributions

Conceptualisation: 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.

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., Lindeman, G. J. & Visvader, J. E. 2007. Gata-3 is an essential regulator of mammary-gland morphogenesis and luminal-cell differentiation. *Nat Cell Biol,* 9**,** 201-9.

Bach, K., Pensa, S., Grzelak, M., Hadfield, J., Adams, D. J., Marioni, J. C. & 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. & Swarbrick, A. 2016. ID4 Controls Mammary Lineage Commitment and Inhibits BRCA1 in Basal Breast Cancer. *Endocr Relat Cancer,* 23**,** R381-92.

Barsky, S. H. & Karlin, N. J. 2005. Myoepithelial cells: autocrine and paracrine suppressors of breast cancer progression. *Journal of mammary gland biology and neoplasia,* 10**,** 249-260.

Bayly, R., Chuen, L., Currie, R. A., Hyndman, B. D., Casselman, R., Blobel, G. A. & 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**,** 55362-71.

Benezra, R., Davis, R. L., Lockshon, D., Turner, D. L. & 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. & Visvader, J. E. 2014. Dual roles for Id4 in the regulation of estrogen signaling in the mammary gland and ovary. *Development,* 141**,** 3159-64.

Bouras, T., Pal, B., Vaillant, F., Harburg, G., Asselin-Labat, M. L., Oakes, S. R., Lindeman, G. J. & Visvader, J. E. 2008. Notch signaling regulates mammary stem cell function and luminal cell-fate commitment. *Cell Stem Cell,* 3**,** 429-41.

Braunstein, M. & 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., Schiff, R., Brown, M. & Brugge, J. S. 2013. PDEF promotes luminal differentiation and acts as a survival factor for ER-positive breast cancer cells. *Cancer Cell,* 23**,** 753-67.

Carey, J. P., Asirvatham, A. J., Galm, O., Ghogomu, T. A. & 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. & Raychaudhuri, P. 2012. FoxM1 regulates mammary luminal cell fate. *Cell Rep,* 1**,** 715-29.

Chakrabarti, R., Wei, Y., Romano, R. A., Decoste, C., Kang, Y. & Sinha, S. 2012. Elf5 regulates mammary gland stem/progenitor cell fate by influencing notch signaling. *Stem Cells,* 30**,** 1496-508.

Conway, K., Pin, C., Kiernan, J. A. & Merrifield, P. 2004. The E protein HEB is preferentially expressed in developing muscle. *Differentiation; research in biological diversity,* 72**,** 327-340.

Danielson, K. G., Oborn, C. J., Durban, E. M., Butel, J. S. & 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**,** 3756-60.

Davis, F. M., Lloyd-Lewis, B., Harris, O. B., Kozar, S., Winton, D. J., Muresan, L. & 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. & Glukhova, M. A. 2006. Isolation of mouse mammary epithelial progenitor cells with basal characteristics from the Comma-Dbeta cell line. *Dev Biol,* 293**,** 414-25.

Deugnier, M. A., Moiseyeva, E. P., Thiery, J. P. & Glukhova, M. 1995. Myoepithelial cell differentiation in the developing mammary gland: progressive acquisition of smooth muscle phenotype. *Dev Dyn,* 204**,** 107-17.

Dong, J., Huang, S., Caikovski, M., Ji, S., Mcgrath, A., Custorio, M. G., Creighton, C. J., Maliakkal, P., Bogoslovskaia, E., Du, Z., Zhang, X., Lewis, M. T., Sablitzky, F., Brisken, C. & Li, Y. 2011. ID4 regulates mammary gland development by suppressing p38MAPK activity. *Development,* 138**,** 5247-56.

Ewald, A. J., Brenot, A., Duong, M., Chan, B. S. & Werb, Z. 2008. Collective epithelial migration and cell rearrangements drive mammary branching morphogenesis. *Developmental cell,* 14**,** 570-581.

Feinberg, T. Y., Zheng, H., Liu, R., Wicha, M. S., Yu, S. M. & Weiss, S. J. 2018. Divergent Matrix-Remodeling Strategies Distinguish Developmental from Neoplastic Mammary Epithelial Cell Invasion Programs. *Developmental 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., Tam, W. L., Mani, S. A., Van Oudenaarden, A. & Weinberg, R. A. 2012. Slug and Sox9 cooperatively determine the mammary stem cell state. *Cell,* 148**,** 1015-28.

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

Idoux-Gillet, Y., Nassour, M., Lakis, E., Bonini, F., Theillet, C., Du Manoir, S. & 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., Yang, J., Mcfarland, A., Harvey, K., Naylor, M. J., Lakhani, S. R., Simpson, P. T., Raghavendra, A., Saunus, J., Madore, J., Kaplan, W., Ormandy, C., Millar, E. K., O'toole, S., Yun, K. & Swarbrick, A. 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. & 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. & Werb, Z. 2006. GATA-3 maintains the differentiation of the luminal cell fate in the mammary gland. *Cell,* 127**,** 1041-55.

Kumar, M. S., Hendrix, J. A., Johnson, A. D. & 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**,** 840-7.

Lee, K., Gjorevski, N., Boghaert, E., Radisky, D. C. & Nelson, C. M. 2011. Snail1, Snail2, and E47 promote mammary epithelial branching morphogenesis. *Embo j,* 30**,** 2662-74.

Li, S., Chang, S., Qi, X., Richardson, J. A. & Olson, E. N. 2006. Requirement of a myocardin-related transcription factor for development of mammary myoepithelial cells. *Mol Cell Biol,* 26**,** 5797-808.

Li, Y., Brauer, P. M., Singh, J., Xhiku, S., Yoganathan, K., Zúñiga-Pflücker, J. C. & Anderson, M. K. 2017. Targeted Disruption of TCF12 Reveals HEB as Essential in Human Mesodermal Specification and Hematopoiesis. *Stem cell reports,* 9**,** 779-795.

Lilja, A. M., Rodilla, V., Huyghe, M., Hannezo, E., Landragin, C., Renaud, O., Leroy, O., Rulands, S., Simons, B. D. & 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. & 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. & 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. & Wicha, M. S. 2008. BRCA1 regulates human mammary stem/progenitor cell fate. *Proc Natl Acad Sci U S A,* 105**,** 1680-5.

Liu, X., Ory, V., Chapman, S., Yuan, H., Albanese, C., Kallakury, B., Timofeeva, O. A., Nealon, C., Dakic, A., Simic, V., Haddad, B. R., Rhim, J. S., Dritschilo, A., Riegel, A., Mcbride, A. & Schlegel, R. 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. & Watson, C. J. 2018. Neutral lineage tracing of proliferative embryonic and adult mammary stem/progenitor cells. *Development,* 145.

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

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

Macias, H. & Hinck, L. 2012. Mammary gland development. *Wiley Interdiscip Rev Dev Biol,* 1**,** 533-57.

Massari, M. E. & Murre, C. 2000. Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms. *Mol Cell Biol,* 20**,** 429-40.

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

Mesman, S. & Smidt, M. P. 2017. Tcf12 Is Involved in Early Cell-Fate Determination and Subset Specification of Midbrain Dopamine Neurons. *Frontiers in Molecular Neuroscience,* 10.

Miano, J. M., Long, X. & Fujiwara, K. 2007. Serum response factor: master regulator of the actin cytoskeleton and contractile apparatus. *American journal of physiology. Cell physiology,* 292**,** C70-C81.

Mills, A. A., Zheng, B., Wang, X. J., Vogel, H., Roop, D. R. & Bradley, A. 1999. p63 is a p53 homologue required for limb and epidermal morphogenesis. *Nature,* 398**,** 708-13.

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., Zwart, W., Saadi, A., Ross-Innes, C. S., Chin, S. F., Menon, S., Stingl, J., Palmieri, C., Caldas, C. & Carroll, J. S. 2013. Endogenous purification reveals GREB1 as a key estrogen receptor regulatory factor. *Cell Rep,* 3**,** 342-9.

Muschler, J. & 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., Lindeman, G. J., Visvader, J. E. & Ormandy, C. J. 2008. The Ets transcription factor Elf5 specifies mammary alveolar cell fate. *Genes Dev,* 22**,** 581-6.

Paine, I. S. & 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. & Kessenbrock, K. 2019. Integrated single-cell transcriptomics and chromatin accessibility analysis reveals novel regulators of mammary epithelial cell identity. *bioRxiv***,** 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., Metzger, D., Faraldo, M. M., Deugnier, M. A., Glukhova, M. A. & Stingl, J. 2014. Mammary stem cells have myoepithelial cell properties. *Nat Cell Biol,* 16**,** 942-50, 1-7.

Rios, A. C., Fu, N. Y., Lindeman, G. J. & Visvader, J. E. 2014. In situ identification of bipotent stem cells in the mammary gland. *Nature,* 506**,** 322-7.

Scheele, C. L., Hannezo, E., Muraro, M. J., Zomer, A., Langedijk, N. S., Van Oudenaarden, A., Simons, B. D. & 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. & Visvader, J. E. 2006. Generation of a functional mammary gland from a single stem cell. *Nature,* 439**,** 84-8.

Sharma, P., Knowell, A. E., Chinaranagari, S., Komaragiri, S., Nagappan, P., Patel, D., Havrda, M. C. & 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. *Molecular Cancer,* 12**,** 67.

Silberstein, G. B., Strickland, P., Coleman, S. & Daniel, C. W. 1990. Epithelium-dependent extracellular matrix synthesis in transforming growth factor-beta 1-growth-inhibited mouse mammary gland. *J Cell Biol,* 110**,** 2209-19.

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. & Eaves, C. J. 2006. Purification and unique properties of mammary epithelial stem cells. *Nature,* 439**,** 993-7.

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

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N. J., Livak, K. J., Mikkelsen, T. S. & 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. & 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. & Blanpain, C. 2011. Distinct stem cells contribute to mammary gland development and maintenance. *Nature,* 479**,** 189-93.

Wang, D., Cai, C., Dong, X., Yu, Q. C., Zhang, X. O., Yang, L. & Zeng, Y. A. 2015. Identification of multipotent mammary stem cells by protein C receptor expression. *Nature,* 517**,** 81-4.

Wang, L. H. & Baker, N. E. 2015. E Proteins and ID Proteins: Helix-Loop-Helix Partners in Development and Disease. *Dev Cell,* 35**,** 269-80.

Watanabe, K., Villarreal-Ponce, A., Sun, P., Salmans, M. L., Fallahi, M., Andersen, B. & 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. & Porter, W. W. 2010. The bHLH/PAS transcription factor singleminded 2s promotes mammary gland lactogenic differentiation. *Development,* 137**,** 945-52.

Williams, J. M. & Daniel, C. W. 1983. Mammary ductal elongation: differentiation of myoepithelium and basal lamina during branching morphogenesis. *Dev Biol,* 97**,** 274-90.

Wuidart, A., Ousset, M., Rulands, S., Simons, B. D., Van Keymeulen, A. & Blanpain, C. 2016. Quantitative lineage tracing strategies to resolve multipotency in tissue-specific stem cells. *Genes Dev,* 30**,** 1261-77.

Wuidart, A., Sifrim, A., Fioramonti, M., Matsumura, S., Brisebarre, A., Brown, D., Centonze, A., Dannau, A., Dubois, C., Van Keymeulen, A., Voet, T. & Blanpain, C. 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. & Hennighausen, L. 2009. Development of mammary luminal progenitor cells is controlled by the transcription factor STAT5A. *Genes Dev,* 23**,** 2382-7.

Yang, A., Schweitzer, R., Sun, D., Kaghad, M., Walker, N., Bronson, R. T., Tabin, C., Sharpe, A., Caput, D., Crum, C. & Mckeon, F. 1999. p63 is essential for regenerative proliferation in limb, craniofacial and epithelial development. *Nature,* 398**,** 714-8.

Yoon, S. J., Foley, J. W. & 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. & Israel, M. A. 2004. Id4 regulates neural progenitor proliferation and differentiation in vivo. *Development,* 131**,** 5441-8.

Zhang, J., Kalkum, M., Yamamura, S., Chait, B. T. & Roeder, R. G. 2004. E protein silencing by the leukemogenic AML1-ETO fusion protein. *Science,* 305**,** 1286-9.

Main figure titles and legends

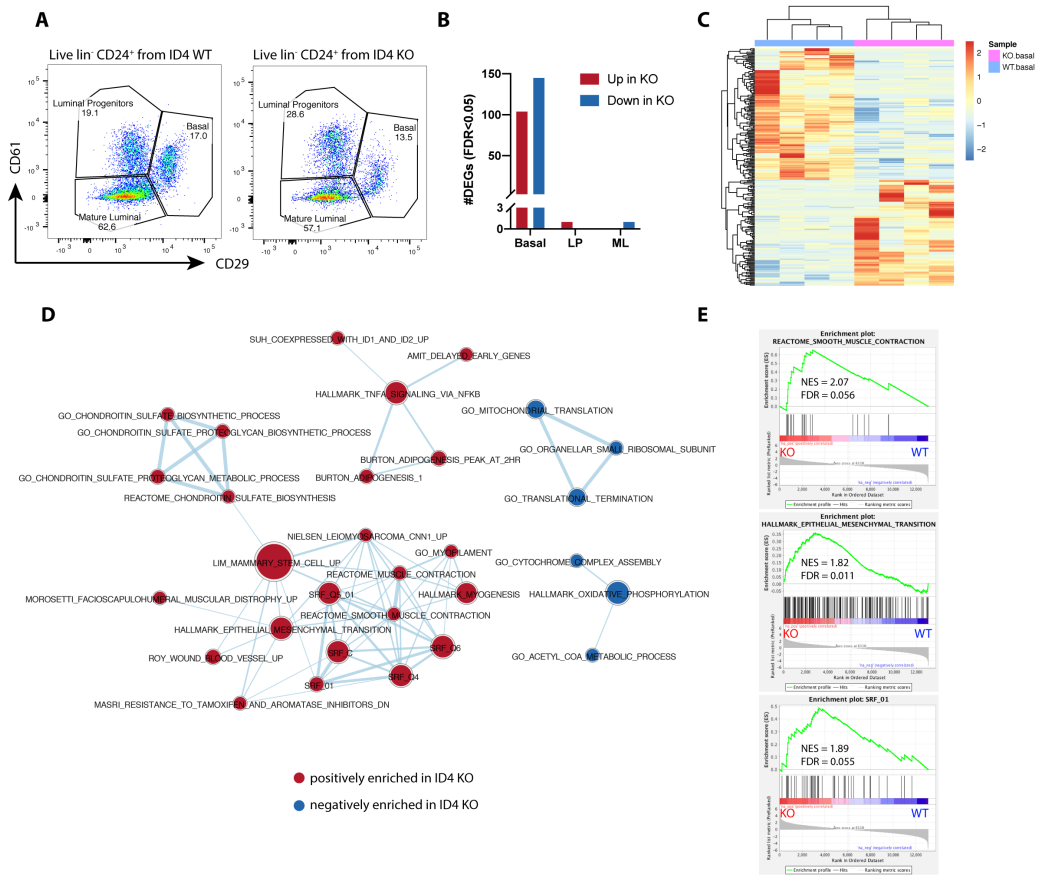

Figure 1. Loss of ID4 results in upregulation of myoepithelial genes in sorted mammary basal cells.

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 4 experiments. **B)** Number of significantly (FDR<0.05) differentially expressed genes (DEGs) upregulated (red) or downregulated (blue) in ID4 KO epithelial subpopulations compared to ID4 WT. LP = luminal progenitor, ML = mature luminal. **C)** Heat map 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 visualised 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. Normalised enrichment scores (NES) and FDR are indicated. See also Figure S1 and Table S1.

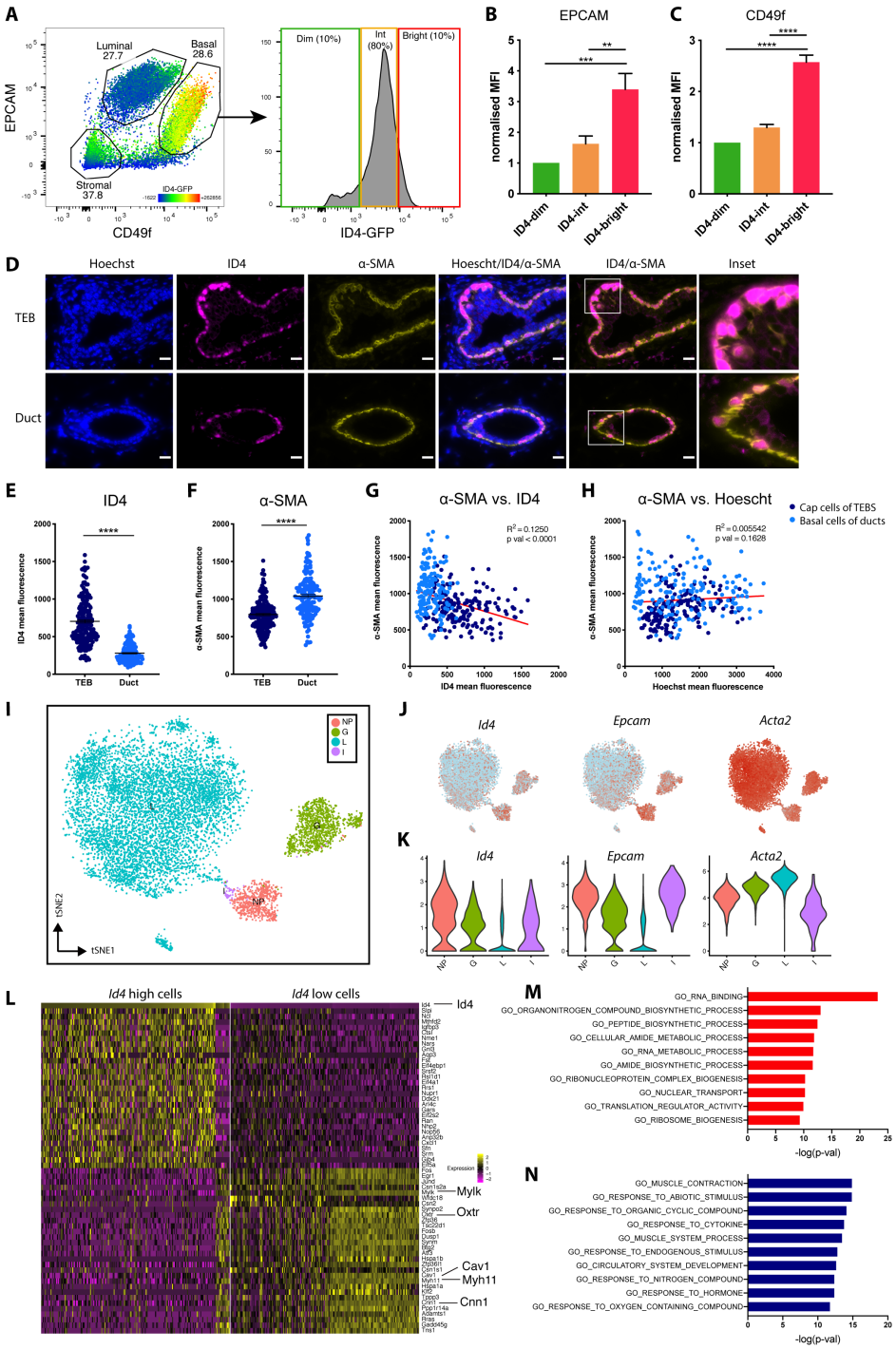

Figure 2. ID4 expression decreases in terminally differentiated myoepithelial cells.

A) FACS analysis of EPCAM and CD49f in live lineage negative mammary cells from adult (10-14 week) Id4floxGFP reporter mice. ID4-GFP expression is indicated in the heatmap scale. Representative plots from 5 experiments shown. Basal cells were binned into 3 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 3 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)** Representative co-immunofluorescent staining of ID4 and α-SMA in TEB and duct from a pubertal (6 week) 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 is pooled from 9 mice. Approximately 20 TEB cap cells and 20 ductal basal cells were analysed per mouse. **I)** tSNE plot of 9663 *Krt5*+/*Krt14*+ basal cells from (Bach et al., 2017). 2 mice were analysed per developmental stage. NP = Nulliparous (8 week), 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)** 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.

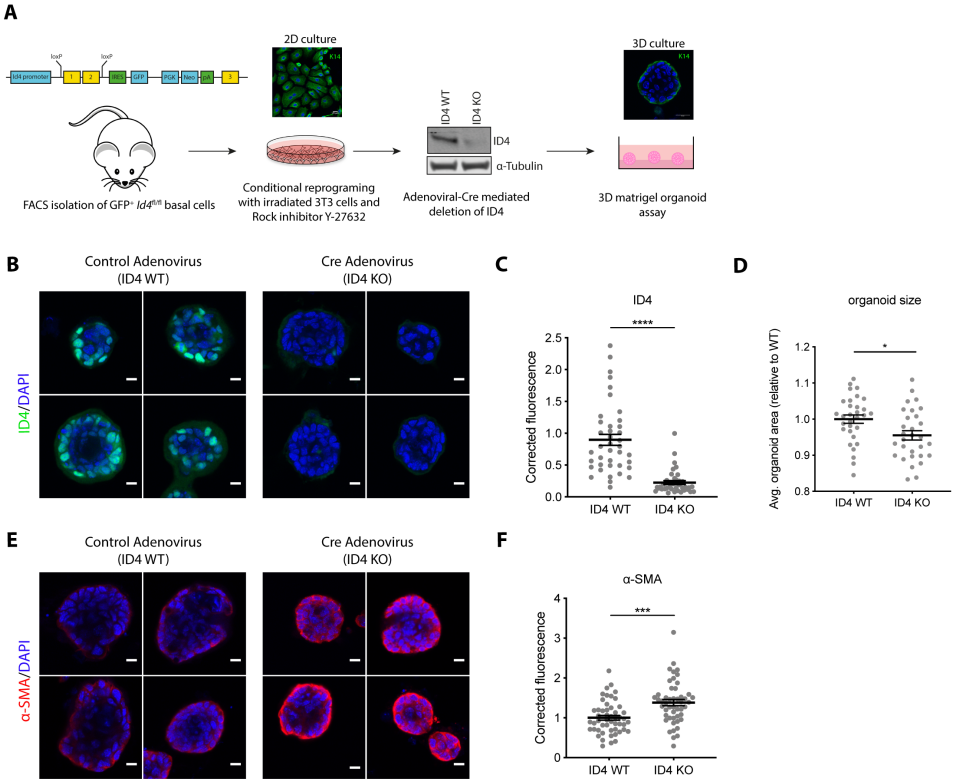

Figure 3. ID4 inhibits myoepithelial differentiation of organoids.

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. 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.

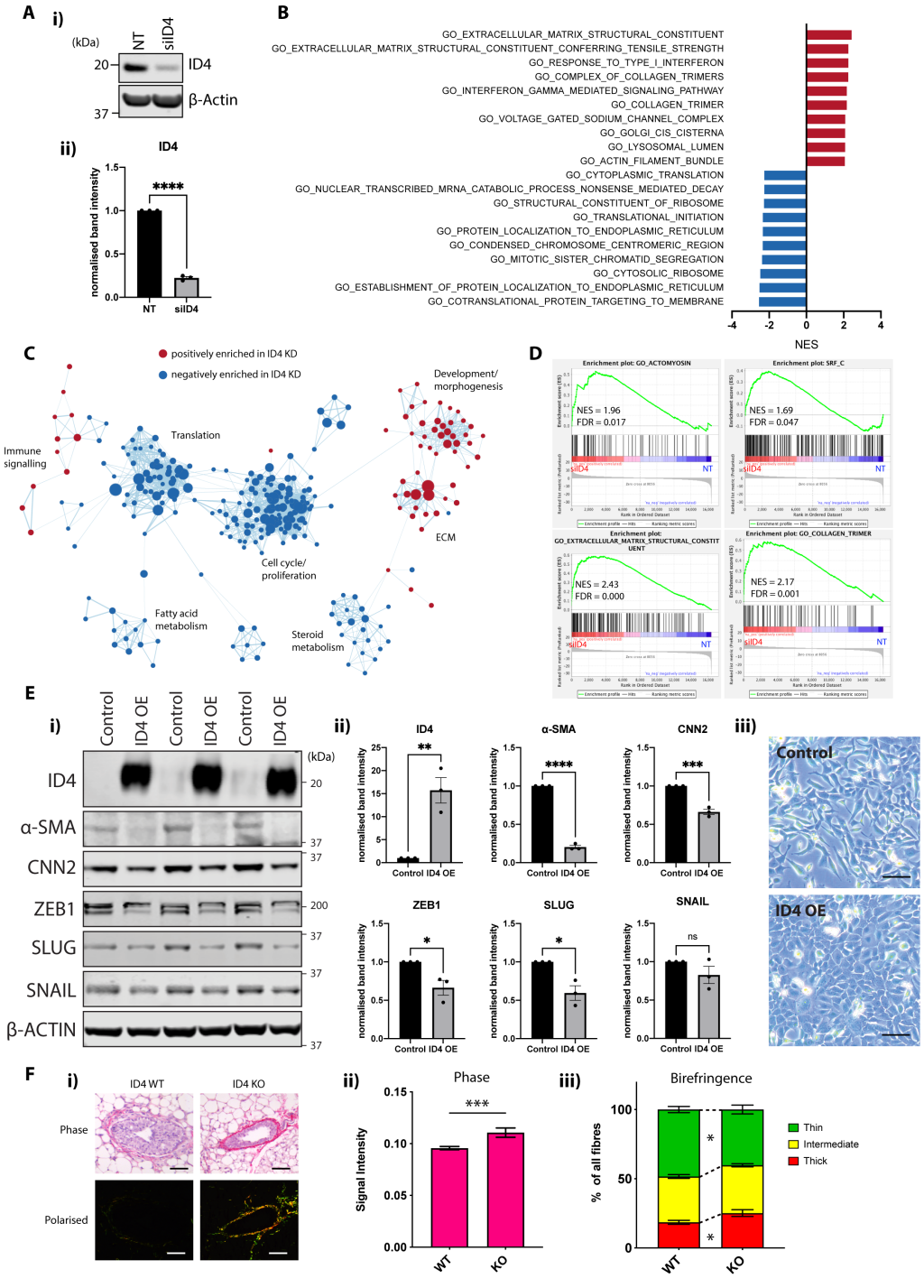

Figure 4. ID4 negatively regulates EMT and ECM production in mammary epithelial cells.

A) (i) Western blot analysis of ID4 expression in Comma-Dβ cells treated with non-targeting (NT) or ID4-targeting siRNA. Representative results from 3 western blots shown. **(ii)** Densitometry quantification of ID4 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. **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 visualised 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 normalised to β-Actin as a fold-change relative to control cells. N=3. Unpaired two-tailed students t-test. **(iii)** Morphology of Comma-Dβ cells overexpressing ID4. Scale bar = 100 μm**. F) (i)** Collagen fibres were visualised 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 3 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.

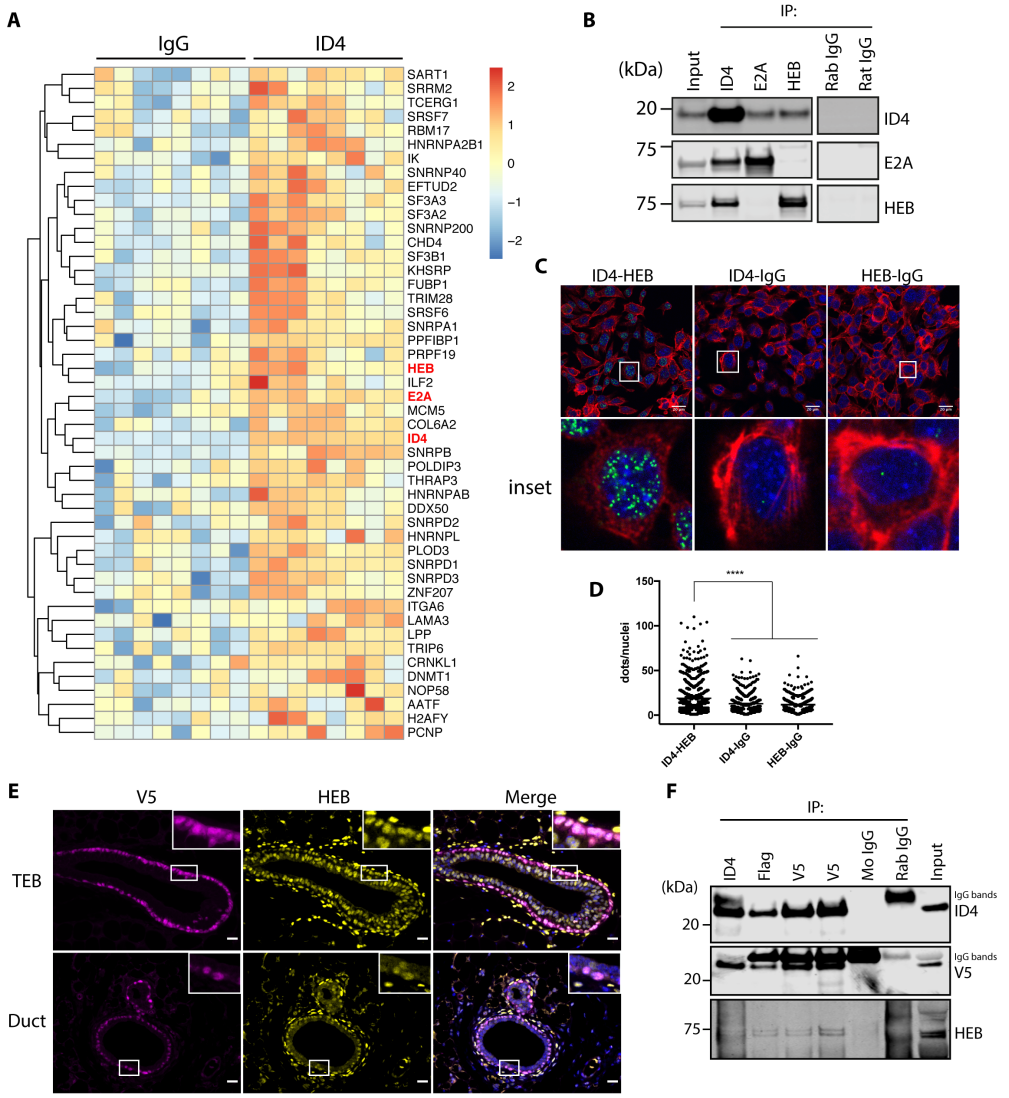

Figure 5. ID4 interacts with E-proteins E2A and HEB.

A) Unsupervised hierarchical clustering heat map of SWATH RIME data from Comma-Dβ cells. Proteins with significantly higher abundance (p-value<0.05) in the ID4 IPs compared to IgG IPs are displayed. Log2 protein area was used to generate the heatmap. Data from 3 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 6 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.

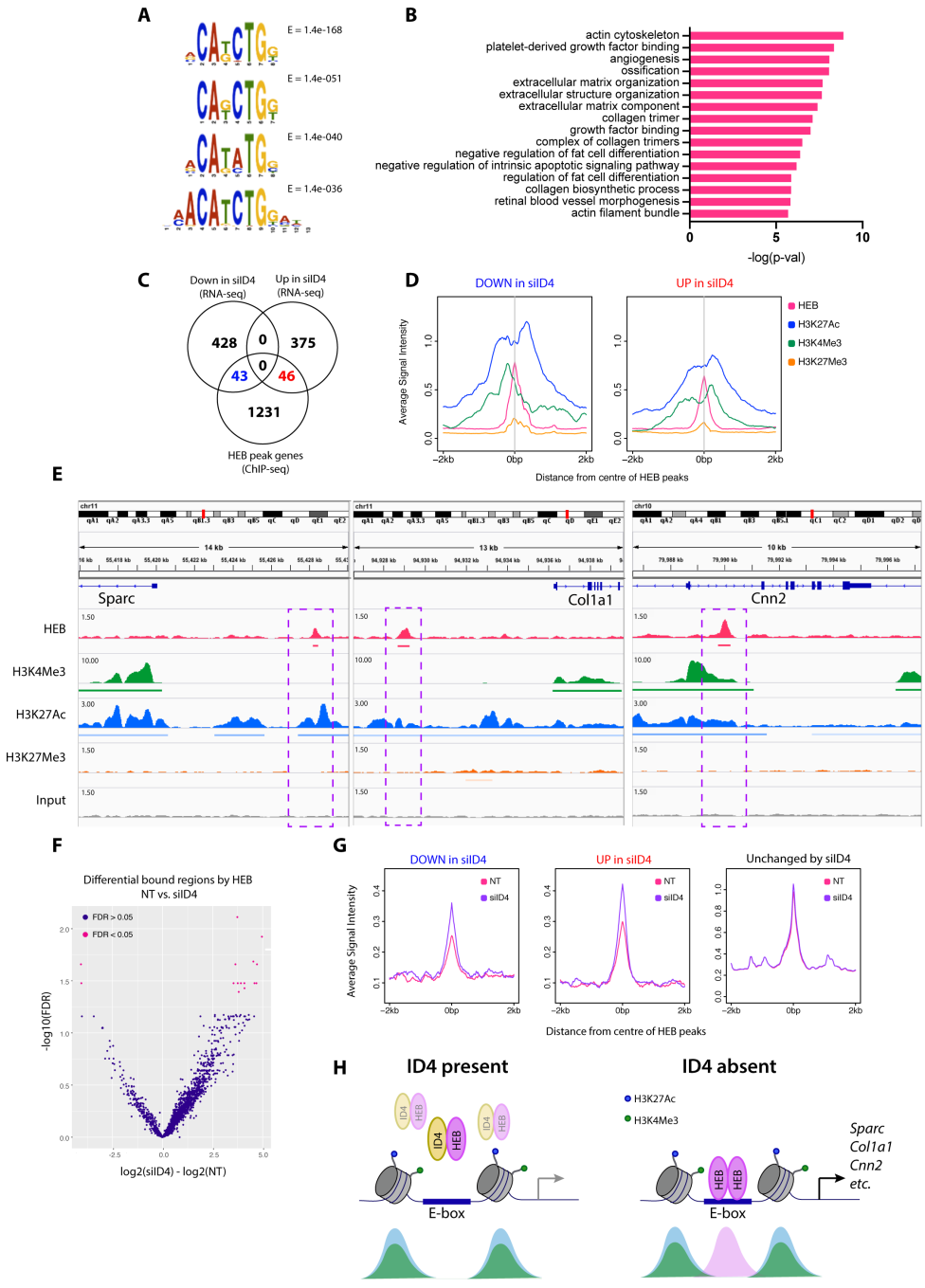

Figure 6. HEB directly binds to a subset of ID4 target genes

A) Top 4 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. **C)** Venn diagram

showing overlap between genes associated with a 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 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. **F)** Volcano plot of Differential Binding Analysis. Analysis using edgeR of HEB binding in siID4 verses NT of 3 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, antagonising its transcriptional activity. Right: when ID4 is depleted, HEB dimerises 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.

Supplemental tables titles and legends

Table S1. Differentially expressed genes between ID4 wild type and knockout sorted mammary populations (Basal, luminal progenitor and mature luminal), Related to Figure 1.

Table S2. Top 50 differentially expressed genes between *Id4*-high and *Id4*-lo basal cells from Bach et al. single cell RNA-seq analysis, Related to Figure 2.

Table S3. Differentially expressed genes comparing siID4 and Non-targeting control Comma-Dβ cells, Related to Figure 4.

Table S4. Proteins identified in ID4 and HEB RIME experiments, Related to Figure 5.

Table S5. Genomic regions of Differentially bound HEB ChIP-Seq peaks between siID4 and NT control Comma-Dβ cells, Related to Figure 6.