A Mesenchymal-like Program of Dormancy controlled by ZFP281 Serves as a Barrier To Metastatic Progression of Early Disseminated Cancer Cells

Julio Aguirre-Ghiso (julio.aguirre-ghiso@mssm.edu)
Icahn School of Medicine at Mount Sinai
Ana Rita Nobre
Icahn School of Medicine at Mount Sinai
Erica Dalla
Icahn School of Medicine at Mount Sinai
Jihong Yang
Columbia University Irving Medical Center
Xin Huang
Columbia University Irving Medical Center
Ephraim Kenigsberg
Weizmann Institute of Science
Jianlong Wang
Columbia University Medical Center https://orcid.org/0000-0002-1317-6457

Biological Sciences - Article

Keywords: early disseminated cancer cells, metastatic progression, dormancy, metastasis

Posted Date: January 18th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-145308/v1

Version of Record: A version of this preprint was published at Nature Cancer on September 1st, 2022. See the published version at https://doi.org/10.1038/s43018-022-00424-8.

A Mesenchymal-like Program of Dormancy controlled by ZFP281 Serves as a Barrier To Metastatic Progression of Early Disseminated Cancer Cells

Ana Rita Nobre[1,2], Erica Dalla[1], Jihong Yang[3], Xin Huang[3], Ephraim Kenigsberg[4,5], Jianlong Wang[3] and Julio A. Aguirre-Ghiso[1*]

[1]Division of Hematology and Oncology, Department of Medicine and Department of Otolaryngology, Department of Oncological Sciences, Black Family Stem Cell Institute, Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA;
[2]Abel Salazar Biomedical Sciences Institute, University of Porto, Porto, Portugal;
[3]Department of Medicine, Columbia Center for Human Development, Columbia University Irving Medical Center, New York, USA;
[4]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA;
[5]The Precision Immunology Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA
Icahn Institute for Data Science and Genomic Technology, Icahn School of Medicine at Mount Sinai, New York, NY, USA.
[*]Correspondence: julio.aguirre-ghiso@mssm.edu

ABSTRACT

Increasing evidence shows that cancer cells can disseminate from early-evolved primary lesions much earlier than the classical metastasis models predicted. It is thought that a state of early disseminated cancer cell (early DCC) dormancy can precede genetic maturation of DCCs and metastasis initiation. Here we reveal at single cell resolution a previously unrecognized role of mesenchymal- and pluripotency-like programs in coordinating early cancer cell spread and a long-lived dormancy program in early DCCs. Using *in vitro* and *in vivo* models of invasion and metastasis, single cell RNA sequencing and human sample analysis, we provide unprecedented insight into how early DCC heterogeneity and plasticity control the timing of reactivation. We identify in early lesions and early DCCs the transcription factor ZFP281 as an inducer of mesenchymal- and primed pluripotency-like programs, which is absent in advanced primary tumors and overt metastasis. ZFP281 not only controls the early spread of cancer cells but also locks early DCCs in a prolonged dormancy state by preventing the acquisition of an epithelial-like proliferative program and consequent metastasis outgrowth. Thus, ZFP281-driven dormancy of early DCCs may be a rate-limiting step in metastatic progression functioning as a first barrier that DCCs must overcome to then undergo genetic maturation.

INTRODUCTION

The majority of cancer patients die of metastatic relapse, which frequently occurs years to decades after diagnosis and treatment. This happens because patients already carry numerous disseminated cancer cells (DCCs) that can remain dormant for long periods of time, only to later on give rise to metastasis[1]. Although cancer dormancy is a major concern in the clinic, our knowledge about the origin and nature of dormant DCCs and the mechanisms that allow these cells to remain quiescent, but still retain metastasis-initiating capacity, is still limited. Additionally, it was believed that the ability of cancer cells to disseminate and metastasize was exclusive to late stages of progression[2–4] when rare cells in primary tumors gained numerous genetic alterations considered necessary for spread and target organ colonization. However, increasing evidence has revealed that the barriers to activate invasion and motility programs may be enabled very early in cancer evolution resulting in early DCCs seeding organs over long periods of time[1].

Early dissemination or intra-organ dispersion was reported in human breast[5–9], pancreatic[10–12], lung[12], melanoma[13] and colorectal[14] cancer patients. In renal cell[15], ovarian[16], testicular[17] carcinomas and osteosarcoma[18], early spreading clones were also reported as metastasis founders across different organs. However, the exact programs leading early spread and the persistence of early DCCs are not clear from these studies. In mouse models of breast[7,19], pancreatic[10,20] and melanoma cancers[21,22] we and others identified early lesion cells with the ability to spread to secondary organs; however, while early DCCs can indeed found metastasis[7,19], information on their post-dissemination phenotypes had not been explored. This is important because the field has not resolved whether the time it takes for early DCCs to manifest as metastasis is due simply to a slow genetic maturation process or if indeed there is a program that holds early DCCs in a dormant state before they can initiate slow or fast (explosive growth[1]) proliferation and “mature” genetically.

Using HER2 and PyMT oncogene-driven mouse models, we modeled early cancer cell dissemination and found that HER2 signaling activates a partial epithelial-to-mesenchymal transition (EMT) program, leading to dissemination[19]. Activation of progesterone and Wnt signaling and inhibition of the p38 pathway[7,19] fuels early spread through a process that resembles mammary tree branching morphogenesis[23]. Similarly, early dissemination is further fueled by early lesion infiltrating CD206+/Tie2+ macrophages in these same models, a similar population of macrophages that regulates mammary tree development[24]. Using a binary set of mesenchymal vs. epithelial markers (TWIST1 and CDH1), we found that HER2+ early DCCs in secondary organs maintain a mesenchymal and long-lived dormant phenotype that preceded metastasis initiation[19]. Here we expanded this analysis and used single cell RNA sequencing (scRNAseq) to reveal the DCC heterogeneity and plasticity in lungs across the evolutionary spectrum of the disease. Surprisingly, we

found that early DCCs activate a novel program with the dual function of fostering early dissemination and initiating dormancy of early DCCs in lungs. We reveal that the primed pluripotency transcription factor *ZFP281* is a key regulator of early DCC spread and dormancy. Using both organoid and *in vivo* models we show that ZFP281 induces mesenchymal-like and primed pluripotency programs, which while not enabling proliferation in the primary or secondary site, allows for efficient dissemination to the lungs. After dissemination, if not downregulated, ZFP281 maintains early DCCs in a prolonged dormant state. Importantly, we show that even aggressive late cancer cells can be reprogrammed into dormancy and prevented from metastasizing by regaining ZFP281 expression.

Our efforts in understanding early and late DCC heterogeneity have yielded a novel mechanism and regulator of metastatic dormancy that would have been missed if we had only focused on advanced primary tumor biology and the classical view of the metastatic cascade. Our work opens the way for understanding how early DCC dormancy is likely a first barrier for genetically immature DCCs to overcome to further evolve. These data may also enable exploiting these mechanisms to eliminate DCCs or force them into an indolent and harmless dormant phenotype.

RESULTS

Early lesions activate a mesenchymal-like program that persists in early and late lung DCCs.

To investigate the mechanisms of early dissemination, dormancy and metastasis awakening, we used the MMTV-ErbB2/HER2/Neu mouse model[25], a spontaneous breast cancer model of HER2+ cancer with luminal characteristics[26]. This mouse model shows slow tumor progression, providing a significant temporal window to study early stages of tumorigenesis and metastatic progression prior to the occurrence of primary tumors (**Figure S1A**,[19]). To understand the gene programs present in early *versus* late MMTV-HER2 lesions, we performed RNA sequencing (bulk RNAseq) of early lesion (EL) and primary tumor (PT) spheres, which recapitulate the *in vivo* behavior of EL and PT lesions[19]. We identified 4290 differentially expressed genes (DEGs, adj. p-value<0.05 and FC>2 or <0.5), 2873 upregulated and 1417 downregulated, in EL vs. PT spheres (**Figure 1A and STable 1**). Among the upregulated gene ontology programs enriched in EL cells, we found terms associated with TGFβ, ECM, collagen, focal adhesion, PI3K and β1 integrin signaling, pathways associated with EMT, adhesion, cellular morphogenesis and ECM remodeling[27,28]. In contrast, the top downregulated gene ontology term was tight junction regulating genes (**Figure S1B and STable 2**, Enrichr analysis[29,30]). Further supporting a gain of a more mesenchymal program, GSEA analysis[31,32] revealed epithelial-to-mesenchymal transition (EMT) as the most enriched hallmark of EL over PT cells (**Figure 1B and STable 3**). EL cells were also enriched in 'mammary luminal down' and 'mammary stem cell' signatures (**Figure 1B and STable 3**,[33]. Together these results suggest that $HER2^+$ EL cells activate

mesenchymal-like (M-like) and basal/stem-like programs, which are subsequently silenced in advanced PT cells that appear to gain an epithelial-like (Ep-like) program.

We next sought to better understand the heterogeneity of globally M-like EL and Ep-like PT cells, as well as early and late lung (eL and LL) DCCs. To this end, we sorted HER2+ tumor cells from EL, PT, eL DCCs and LL DCCs (**Figure 1F and S1C**) and performed single cell RNA sequencing (scRNAseq). Gene expression profiles from 3686 cells were obtained and compared with the trends we observed in EL vs. PT by bulk RNAseq. As expected, in single cells sorted from EL cells, the expression of genes which were upregulated (containing mesenchymal and stem genes) in the bulk RNAseq of 7-day sphere cultures of EL over PT spheres was higher and reciprocally, in single cells sorted from PTs, the expression of downregulated genes (containing luminal genes) was higher (**Figure 1C**). Interestingly, EL cells showed more heterogeneity (expanded cloud range) than PT cells. Additionally, early and late lung (eL and LL) DCCs clustered further away from PT cells but closer to EL cells (**Figure 1C**), suggesting that perhaps DCCs represent a subpopulation of EL cells with a more mesenchymal and stem signature. Further unsupervised clustering on the DEGs, using a previously described batch-aware algorithm[34], showed that although EL and PT cells clustered almost independently, clusters 3 and 5 contain both EL and PT cells (**Figure 1D**), arguing that some PT cells maintain an EL program or regain such program. Interestingly, lung DCCs clustered separately from EL and PT cells; however, cluster 9 is uniquely composed by EL cells, eL DCCs and LL DCCs, no PT cells (**Figure 1D**), suggesting that ELs contain a subpopulation of cells that already carry a signature that allows them to disseminate and persist in lung DCCs. Our analysis also showed that DCCs from early and late lungs, while heterogeneous and distinct from EL and PT cells, were always contained in the same signature clusters (7-11) (**Figure 1D**) suggesting that DCCs with early lesion signatures may persist in the late stages. This prevented distinguishing early DCCs from those DCCs populating late lungs, most likely because late lungs carry early DCCs (derived from ELs), DCCs derived from primary tumors and growing metastasis, all co-existing in the same lungs. Nonetheless, Ep- and M-like signatures found in the primary early and late lesions (**Figure 1A** and [19]) were also found in lung DCCs and these cells could be broadly grouped into Ep- (7-8) and M-like (9-11) clusters (**Figure 1E**). Interestingly, the Ep-like clusters 7 and 8 shared epithelial signatures more homogeneously, while the M-Like clusters 9-11 showed more non-overlapping mesenchymal transcriptional signatures and all populations co-exist in early or late stage lungs.

Analyses by both FACS (**Figure 1F and S1D**) and immunofluorescence (**Figure 1G**) revealed that PT cells show high levels of HER2 and a predominant epithelial phenotype, characterized by strong EpCAM expression and non-invasive organoids. In contrast, EL cells showed a broader spectrum of HER2 expression and a mixture of epithelial $EpCAM^+$ cells and also mesenchymal $Eng/CD105^+$ cell

populations (**Figure 1F and S1D**) which correlated with a more invasive phenotype when in culture on top of matrigel (**Figure 1G** and [19]). Endoglin (Eng/CD105) was a mesenchymal marker[35] selected from the scRNAseq data (**Figure 1E**) due to its selective upregulation in M-like DCCs enriched in early lungs. FACS analysis confirmed that the M-like/Eng$^+$ and invasive phenotype found in EL cells, persists and increases in frequency in early lung DCCs (**Figure 1F, S1D and 1G**). In contrast, late lungs presented a smaller population of M-like DCCs (Eng$^+$ and invasive) and enrichment in Ep-like DCCs (EpCAM$^+$ and non-invasive), resembling PT cells. We conclude that a subpopulation of EL cells carries a M-like signature that is found in M-like clusters in lung DCCs, which revealed a previously unrecognized heterogeneity of cellular states of DCCs in early and late stage lungs. Remarkably, late lungs are still populated by a significant fraction of DCCs with transcriptional programs found in early DCCs.

Early DCCs gain multiple M-like cellular states and display a dormant phenotype.

To gain further insight into the heterogeneity of M- and Ep-like phenotypes of lung DCCs, we performed additional scRNAseq profiling focusing exclusively on lung DCCs and increasing the number of sampled cells. HER2-negative non-cancer lung cells and HER2+ DCCs from early and late stage mice were analyzed, and a comprehensive analysis and clustering of 15,287 additional cells was performed (**Figure S2A**). We identified 25 distinct clusters, 10 were excluded due to their high prevalence in normal lung cells resulting in 15 DCC clusters (with less than 16% of normal lung cells). These DCCs were further sub-grouped in M-like (1 to 4), hybrid (5 to 8) and Ep-like (9 to 15) clusters based on canonical mesenchymal and epithelial signatures (**Figure 2A and S2A**). Ep-like scores are variable but it seems the vast majority of DCCs keep an epithelial identity, gaining or losing mesenchymal traits (**Figure 2A and S2A**). Of note, DCCs show a high degree of cellular plasticity but few cells undergo full EMT or MET, which made us use the terms M- and Ep-like, not strict categorizations. We also found that late stage mice with advanced disease carried more Ep-like DCCs, while early stage mice had more frequently DCCs with M-like and hybrid phenotypes (**Figure 2A**). Interestingly, only one cluster, cluster 15, was enriched exclusively in LL DCCs and had a strong Ep-like signature. All other clusters had some representation of early DCCs and DCCs from late-stage animals with Ep- (more frequent in late lungs) and M-like (more frequently in early DCCs) phenotypes (**Figure 2A-B, S2B and S2F**).

Analysis of gene-to-gene correlation of highly variable genes identified gene modules with strong co-expression patterns. The enrichment of TF targets that correlated with the expression of the modules (Enrichr analysis[29,30]) revealed multiple programs activated in DCCs that are associated with pluripotency, mixed-lineage differentiation and EMT (**Figure 2B and STable 5**). M-like DCCs from

cluster 1, enriched in gene module A, express brain- and osteoblast-lineage genes, and this A signature revealed genes commonly regulated by the TFs Neurod1 (neurogenic differentiation 1), SOX8, SOX9 and SOX10 (embryonic development regulators) and Vim, Col4a1 and Col4a2 (EMT-associated genes). M-like DCCs from cluster 2, enriched in gene module B, share some of the above mentioned genes and also genes commonly regulated by the TFs and chromatin remodelers SUZ12, SOX17, SOX18, POU5F1/OCT4, well-known pluripotency regulators. M-like DCCs from cluster 3 and 4, still carried genes controlled by the above transcriptional regulators but also gained EMT genes (Zeb2 and Col3a1) and genes commonly regulated by the TFs Snai2, Twist1, Prrx1, Fbn1 (also EMT inducers) and SMADs. These data support that an M-like program initiated in EL cells (**Figures 1 and S1**) is carried by early DCCs in the lung and even persists in DCCs in late lungs. In hybrid DCCs (clusters 5 to 8), we noted a shift to expression of genes that are commonly regulated by the TFs GATA6, Tp63, Tp73 and KLF4, typical basal and luminal epithelium switch regulators, and epithelial markers such as Krt7 and Krt8 (**Figure 2B and STable 5**). These changes suggest that DCCs in hybrid clusters might be in transit between M- and Ep-like states. Supporting this, cluster 8 (hybrid), which expresses several signatures, is comprised of distinct cell populations that express gene modules H (B, C and D) and I (**Figures 2C, S2D and S2E**). Overall, these DCCs spread between intermediate Ep and M-like states, starting to lose gene modules B and D while gaining gene module I (**Figure 2C**). When analyzing the Ep-like clusters (9-15), we found that gene module I is homogeneously expressed by almost all clusters (9 to 14). Interestingly, cluster 15 is distinct and, as mentioned above, composed only by DCCs from late lungs (LL). These DCCs express luminal epithelial genes (EpCAM and Krt18), Ovol1, Ovol2, Grhl2 TFs (also epithelial genes), as well as mammary gland/lactation genes (Csn1s1, Csn1s2, and Csn3) (**Figure 2B and STable 5**). These results suggest that cluster 15 corresponds to more luminal differentiated Ep-like proliferative DCCs/metastasis, as evidenced by an increase in CCND1 gene, which is almost absent in M-like and hybrid clusters (**Figure 2D**).

This led us to better define which clusters contain dormant DCCs. We found that Cdkn1c/p57^{Kip2}, NR2F1 and TGFβ2, all genes previously linked to quiescence and dormancy of DCCs, are more frequently expressed by M-like DCCs (**Figure 2D**), which as mentioned above show higher frequency of early DCCs (**Figure 2A-C**, negative for p-histone H3 and p-Rb[19]). These data suggest that early DCCs activate gene programs linked to cell plasticity, progenitor-like, M-like and dormant programs and that the transition from these programs to an epithelial and more “differentiated program” associates with their ability to proliferate and form metastasis.

ZFP281 is a marker of the M-like programs in EL cells and early DCCs.

Since EL cells do not form tumors but disseminate efficiently and persist as DCCs in lungs[19], we hypothesized that the gain of an M-like program found in the DCCs may be transcriptionally encoded already in the early lesions. To address this hypothesis we performed a transcription factor (TF) network analysis mining the bulk RNAseq data derived from EL *versus* PT spheres. This analysis identified 8 interconnected nodes where ZFP281 was the TF node with the highest number of DEGs in EL cells **(Figure 3A and STable 6)**. ZFP281 is a key transcriptional regulator of primed pluripotency in both mouse and human embryonic stem cells and functions as a barrier toward achieving naive pluripotency[36]. ZFP281 is absent in terminally differentiated human tissues and it was shown to counteract osteogenic[37] and muscle differentiation[38], and we did not find it expressed in normal mammary gland cells (**Figure S3B**). Further, ZFP281 promotes EMT in colorectal cancer cells by upregulation of *SNAI1* and *CDH1*[39]. ZFP281 is upregulated during the naïve-to-primed pluripotent state transition[36] where EMT or partial EMT/epithelial plasticity was postulated to happen[40,41]. When overexpressed in mouse ESCs, ZFP281 also suppresses growth (unpublished data). The second largest node we identified, NR5A2 (also known as LRH-1) (**Figure 3A**), also plays an important role in maintaining stem cell pluripotency during embryonic development[42] but its link to EMT is still unclear[43,44]. Among other TFs, we found RARβ and RARγ, previously linked to dormancy[45], that may also play a role in early lesions and early DCCs. Thus, EL cells seem to upregulate a set of TFs that are involved in pluripotent stem cell plasticity and invasion programs.

We focused on ZFP281 as a potential paradigmatic EMT parallel between normal stem cell and cancer development owing to its roles in regulating stem cell pluripotency, growth arrest and invasion. We validated the increase in ZFP281 mRNA levels and its activity by measuring the changes of its predicted target genes (**Figure 3A**) by qPCR in EL over PT spheres **(Figure S3A)**. This analysis shows that M-like genes such as TGFBRII, CDH11 and Eng are induced in EL over PT cells, while Ep-like genes such as CDH1 and EpCAM were downregulated, arguing that ZFP281 represses an Ep-like identity. Analysis of the expression of the predicted ZFP281 target genes (**Figure 3A**) in the lung DCC clusters revealed that the predicted ZFP281 targets upregulated in EL cells are also frequently upregulated in M-like lung DCCs. In contrast, Ep-like lung DCCs do not show upregulation of these predicted ZFP281 targets or observed upregulated genes (scRNAseq, **Figure 3B**). At the protein level, we found even stronger differences in ZFP281 expression: in normal FvB mammary glands (fully differentiated tissue) ZFP281 is expressed only in 3% of the cells; whereas 30% of EL cells express ZFP281, which is then downregulated in PT cells (8% $ZFP281^{+}$; **Figure 3C and S3B**). Staining of EMT markers (E-cadherin and Twist1 as epithelial and mesenchymal markers, respectively) in sequential sections show that structures enriched in ZFP281 are $Ecad^{low}$ (less intense membrane staining) and $Twist1^{high}$ (**Figure S3B**). When monitoring ZFP281 expression in the early

DCCs, we also found that 42% of single early DCCs are ZFP281^{+}, while only 5% of cells within proliferative metastasis are ZFP281^{+} (**Figure 3D-E**). This further supports that ZFP281 and its regulated programs turned on in EL cells persist in early lung DCCs. Supporting the notion that ZFP281 may drive a dormant phenotype, Ki67 and ZFP281 expression were found to be mutually exclusive in early lung DCCs (**Figure 3D**). To determine if ZFP281 upregulation was also a property of early lesions in human breast cancer we stained ductal carcinoma in situ (DCIS) and invasive breast cancer (IBC) samples for ZFP281. Our analysis revealed that, among 14 human DCIS samples, 48% of the cells per lesion were ZFP281^{+}. In contrast, only 11% of the IBC cells were positive for ZFP281 (**Figure 3F and G**). These data strongly support that ZFP281 is a novel TF in cancer mainly associated with early breast cancer progression by controlling EMT programs while suppressing active cell proliferation. These further suggest that EL cells that turn on ZFP281 would be candidates for systemic spread followed by a dormant phenotype in target organs.

ZFP281 regulates components of an M-like program and primed pluripotency in early DCCs

ZFP281 tightly coordinates cell fate through regulation of primed pluripotency programs in development[36] with possible participation of EMT/partial EMT[40,41]. To determine whether ZFP281 also regulates such programs in early mammary cancer cells, we compared RNAseq data from naïve *versus* primed mouse pluripotent stem cells (a transition regulated by ZFP281[36]) and from EL *versus* PT spheres (**Figure 1A**). This analysis revealed that DEGs in EL vs PT and in primed *versus* naïve stem cells were commonly positively correlated in the categories of EMT and Wnt signaling (**Figure 4A**), suggesting that in EL cells ZFP281 may drive similar EMT and Wnt signaling as those found in primed pluripotent stem cells.

To gather deeper insight into the ZFP281 regulated programs in EL *versus* PT cells we performed chromatin immunoprecipitation sequencing (ChIPseq) in EL and PT cells, identifying 4018 ZFP281 targets in EL cells. Strikingly, comparison of the EL ChIPseq data with that derived from primed mEpiSCs showed a significant overlap in the categories of genes and actual genes that are regulated by ZFP281 in these two contexts (**Figure 4B**). ZFP281 seems to regulate cell cycle arrest, EMT, Wnt and FGFR signaling both in EL and primed mEpiSCs, suggesting that HER2-driven early lesion cells activates distinct programs found very early in embryo development, even earlier than EMT in neural crest cells during gastrulation[46].

When comparing our RNAseq and ChIPseq data from EL *versus* PT cells, we found 504 genes with high ZFP281 binding and high expression in EL cells **(Figure S4A-B)**. Some of these genes overlap with the putative ZFP281 target genes from the network analysis in **Figure 3A (Figure S4B)**, but we also identified new ZFP281 target genes that were not computationally predicted. Among

them are Snai1, Vim, Zeb1 (EMT inducers[28]), Cdk2 and Cdkn1a (cell cycle related[47]) and Tgfbr1, Nr2f1 and Bmp7 (dormancy associated genes[48]) (**Figure 4C and S4C**). These genes were exclusively bound by ZFP281 and upregulated in EL cells or bound by ZFP281 in EL and PT cells but only upregulated in EL cells. In contrast to the genes that were bound by ZFP281 in EL cells, we also identified 118 genes with high ZFP281 binding and high expression in PT cells (**Figure S4B**). This suggests that while ZFP281 expression decreases in PT cells it still binds and regulates a different ZFP281-dependent program of yet unknown function in PT cells.

To address the importance of ZFP281 and its target genes in lung DCCs, we examined their expression in our lung DCC scRNAseq data. Strikingly, we found that ZFP281 targets (from EL ChIPseq) score summarizing the averaged expression of ZFP281 targets has a bimodal distribution in lung DCCs (**Figure S4D**): M-like and hybrid DCC clusters display the highest levels of ZFP281-regulated signatures and these scores drop significantly in Ep-like DCCs (**Figure 4D-E and S4E**). However, some clusters like cluster 8 showed a drop in ZFP281 signature score, arguing that some hybrid cluster cells are moving from an M-like to Ep-like state. We conclude that ZFP281-regulated genes in EL cells are still active in M-like dormant DCCs and that they likely make M-like cells permissive to explore these mesenchymal states while repressing an epithelial state. These data also further support that the M-like program driven by ZFP281 is activated in the EL cells, carried over and sustained in DCCs in the target organs.

ZFP281 prevents the acquisition of an Ep-like state and maintains a M-like dormant phenotype in DCCs

Our data support a model whereby ZFP281 regulates programs of dissemination and primed pluripotency that lead to early DCC dormancy. Thus, we set out to functionally test whether indeed ZFP281 holds DCCs in a dormant state in lungs. EL cells are engaged in a M-like invasive program and, when cultured in suspension (in mammosphere medium) or in 3D (on top of Matrigel), EL cells form more invasive (M-like) spheres than PT cells, which grow into large and less invasive (Ep-like) spheres ([19] and **Figure S5**). Using an inducible short hairpin for ZFP281 (shZFP281) we show that ZFP281 downregulation in EL spheres leads to a transition from an M-like to Ep-like phenotype, resembling PT spheres (**Figure 5A**). FACS analysis shows that a population of M-like EL cells with the DOX-induced shZFP281 gain medium and high EpCAM expression, however they do not downregulate Eng expression, leading to an increase in a hybrid phenotype ($EpCAM^+/Eng^+$, dark blue) (**Figure 5D-E**). Additionally, although the frequency of spheres does not change significantly (**Figure 5B**), the size (evaluated by the number of cells per sphere) is increased upon ZFP281 downregulation (**Figure 5C**), supporting enhanced proliferation once the spheres are formed. Further,

when these cells are plated on top of Matrigel, the number of invasive acini is significantly lower in the EL shZFP281+DOX condition (**Figure 5F-G**), supporting a switch from an invasive and motile program to a growth program upon ZFP281 downregulation. Corroborating this partial mesenchymal-to-epithelial transition (MET), we observed a decrease, measured by qPCR, in Twist1, Eng and CDH11 (mesenchymal markers) upon ZFP281 downregulation (**Figure S3A, 3rd column**). This effect is specific to shZFP281 since all the experiments were run in parallel in control EL shCt cells -/+ DOX without significant differences observed (**Figure S5 and S3A, 2rd column**). Conversely, overexpression of ZFP281 in PT spheres (PT ZFP281-OE, **Figure S3A, 4th column**) induced an invasive M-like phenotype (**Figure 5A**), confirmed by FACS (**Figure 5D-E**) and qPCR (**Figure S3A, 4th column**), increased sphere formation potential consistent with a stronger stem program (**Figure 5B**), reduced sphere size consistent with an inhibition of proliferation (**Figure 5C**) and increased organoid invasive phenotype (**Figure 5F-G**).

While the *in vitro* assays employed above are not optimal surrogates to read out dormancy mechanisms, they provide clues as to the phenotypic direction of certain genes in cancer cells. Thus, we next tested the gain and loss of function effects of ZFP281 on tumorigenesis, dissemination and metastasis. EL spheres transduced with the DOX-inducible shZFP281 system were injected in the mammary fat pad (MFP) of mice as reported[19]. Then mice were given control drinking water (-DOX), water with doxycycline from day 0 (+DOX) or starting one month after sphere injection (-DOX +DOX) until the end of the experiment, five months after spheres injection. As previously reported[19] few mice developed palpable slow-growing tumors with static kinetics, with no difference between conditions (data not shown); however, when the injection sites were analyzed after five months, HER2+ EL cells were still found in the MFP of all mice and ZFP281 expression was downregulated both in EL shZFP281 '+DOX' and '-DOX +DOX' groups (**Figure 6A**). Even in the absence of primary tumors, after five months single DCCs and micro-metastasis were found in all lungs, supporting a 100% efficiency of dissemination by EL cells (**Figure 6B**). Comparison of the control and two DOX treatment groups showed that EL shZFP281 -DOX +DOX mice showed less single DCCs per mouse. However, both groups of animals where ZFP281 was downregulated from the beginning ('+DOX') or after one month ('-DOX +DOX') displayed a significant increase in lung metastasis (**Figure 6C**). While solitary $HER2^+$ DCCs in all groups were Ki67 negative, the frequency of proliferative $Ki67^+$ cells in metastasis increased upon ZFP281 downregulation, regardless of the treatment schedule (**Figure S6A**). Given that the M-like clusters mostly enriched in early DCCs were characterized by a ZFP281 enriched signature that also showed expression of dormancy and cell cycle arrest genes (**Figure 2D**), these data strongly support that the M-like dormant phenotype is induced and maintained by ZFP281.

Consistently, loss of ZFP281 signaled early DCCs reactivation from dormancy and switch to a proliferative phenotype.

Next, we studied the effect of PT spheres overexpressing ZFP281 (ZFP281-OE). The control or ZFP281-OE spheres were injected in the MFP of mice as described above. Although ZFP281-OE cells were not impaired in their ability to form the initial tumor, they were significantly slower in their growth kinetics, supporting the growth suppressive function of ZFP281 (**Figure 6E**). Tumor sections showed a heterogeneous increase of ZFP281 expression in the PT ZFP281-OE condition (**Figure 6D**). Nevertheless, even with slower growing tumors (**Figure 6E**), after two months mice carrying ZFP281-OE tumors showed a five-fold increase in the number of lung single-cell DCCs compared to control tumors (**Figure 6F**). Importantly, the increase in single-cell DCCs in the ZFP281-OE group did not result in an increase in micro-metastasis at two months. Thus, ZFP281 suppresses growth of the primary tumor, but enhances dissemination without a subsequent increase in metastatic growth, which is consistent with its ability to induce dormancy. In a longer experiment that allowed reading out better overt metastasis, less PT Control or ZFP281-OE cells were injected and tumors were allowed to grow for 70 days, removed by surgery and then mice were followed and euthanized five months after injection. While no difference in number of lung single DCCs was found, a significant reduction in number and size of metastasis was observed in PT ZFP281-OE mice over PT control (**Figure 6G-I**), as well as reduction of $Ki67^{+}$ cells (**Fig S6B**). These results clearly corroborate the key role of ZFP281 in inducing a growth arrested dormant phenotype in DCCs that is intrinsically active in early DCCs but also enforceable in late DCCs.

DISCUSSION

Early dissemination was documented in various human and mouse studies[5–22]; however, limited information is available as to what is the fate of the early DCCs once lodged in target organs and before metastasis grow out. Further, incomplete modeling of early dissemination biology has prevented determining whether early DCCs turn on active programs of dormancy that delay re-growth and/or if they simply lack sufficient “driver” mutations and require more time and slow proliferation to produce successful clones in target organs.

Our previous work revealed that oncogene and microenvironmental signals in early lesions conspire to activate an EMT program, which appeared to persist in non-proliferative DCCs as marked by a $TWIST1^{high}E\text{-}Cadherin^{low}p\text{-}Rb^{low}p\text{-}histone\text{-}H3^{low}$ profile that regained E-cadherin to resume proliferation[7,19]. Similarly, early pancreatic DCCs also undergo EMT that persists in circulating pancreatic cells that seed the liver[20]. However, all these studies did not functionally link the EMT program to dormancy or reactivation. Nevertheless, together these data hinted at a modulation of M-

like and Ep-like identities in DCCs as a driving mechanism of early DCC fate. Corroborating this, our current study reveals that early lesions driven by the HER2 oncogene activate an M-like program of motility and invasion linked to primed pluripotency that not only allows early lesion cells to spread but also enables them with a program of dormancy where stem cell-like plasticity is operational. Our analysis revealed that ZFP281 serves as a barrier for DCCs to adopt an Ep-like phenotype but also enables M-like DCCs to explore at least 4 major phenotypes described by transcriptional modules A-D. These M-like programs seem to associate with the expression of dormancy markers such as the Cdkn1c/p57^{Kip2}, NR2F1 and TGFβ2, some of which are directly bound by ZFP281 and exclusively induced in EL cells. Hybrid clusters of DCCs appeared to downregulate ZFP281 activity and gain back Ep-like genes, supporting our hypothesis that ZFP281 is restraining this switch. Ep-like clusters are also more homogeneous, arguing that once the DCCs commit to a proliferative phenotype they are funneled into a more phenotypically uniform state, except for cluster 15 that seemed to veer further into a "differentiation state" where lactation genes were upregulated. Interestingly, we had described that early DCC-founded metastasis had a mixed histology with undifferentiated and glandular-like structures reminiscent of lactogenic acini[19]. We interpret that early DCCs enter the lung in a M-like state and can persist dormant until signals, yet to be determined (intrinsic or microenvironmental), cause a final switch. Early DCCs M-like states may allow DCCs to explore different programs (developmental/pluripotency and mixed-lineage differentiation) that best fit them to adapt and survive in secondary organs. In fact, mesenchymal M-like DCCs in these same mouse models have been linked to drug resistance[49]. Our analysis of HER2^{+} early DCCs in tissues showed that they are all vastly negative for Ki67 protein expression. Thus, Hybrid and Ep-like cells may also undergo a dormancy phase, but their transcriptional states may enable them to be more prone to reactivate, a measure we could not capture with Ki67 staining. Additionally, the ZFP281 knockdown suggests that simply reducing this TF can move early DCCs to a more Hybrid or Ep-like phenotype and this correlated with increased metastatic growth. Furthermore, Snail- and Zeb1-driven EMT was previously described to suppress cell-cycle progression through repression of cyclin D1 and D2[50,51]; while mesenchymal-to-epithelial transition (MET) was associated with rapid relapse and reduced survival in metastatic castrate resistant prostate cancer patients[52].

Previous studies also theorized that M-like states might produce dormant DCCs[28]. Further, Lawson *et al.* found that low-burden (dormant-like) breast cancer metastatic cells (i.e., upregulated CDKN1B, CHEK1, TGFBR3 and TGFB2) were mostly basal and pluripotent stem-like (i.e., upregulated POU5F1 and SOX2), while higher-burden DCCs were more luminal-like and proliferative (i.e., upregulated MYC, CDK2, MMP1 and CD24)[53]. Previously, we also reported that the lineage commitment regulators DEC2/BHLHE41 and NR2F1/COUP-TF1 coordinate stem-like and

quiescence programs[45,54,55]. However, all the latter studies are in late evolution cancer models. Our data is the first to functionally map these basal/stem-like and developmental/pluripotency programs to such early stages of cancer evolution and associate it with an M-like dormant DCC phenotype. Recently, Laughney *et al.* also reported that metastatic cells (from late evolution cancer models) recapitulate a primitive transcriptional program spanning stem-like to regenerative pulmonary epithelial progenitor states, such as the key endoderm and lung-specifying transcription factors, SOX2 and SOX9[56]. Similarly, we observed that early DCCs in our model have a cellular plasticity that may allow them to explore different programs (developmental/pluripotency and mixed-lineage differentiation) that best fit them to adapt and survive in secondary organs. Together, these data suggests that pluripotency and dedifferentiation programs may be common programs present in different cancer types and in early stages of cancer progression. Our findings support that early DCCs display a high degree of cellular plasticity through mesenchymal-like, primed pluripotency and dormancy programs that likely endow them with the necessary fitness to survive and undergo genetic maturation upon reactivation.

As mentioned above, close to 100% of early DCCs are Ki67 negative and when ZFP281 is downregulated while metastases emerge, the Ki67 frequency is very low. Thus, as proposed earlier, it is likely that once early DCCs break out from dormancy, they then initiate slow proliferation and genetic maturation. These gradual kinetics may contribute to the invisible phase of the metastatic disease[1]. Nevertheless, our work shows that ZFP281-regulated (or other) dormancy of early DCCs is a prior barrier to overcome before maturation ensues and together both steps may be very protracted.

A remarkable finding was that the early DCC dormancy program seems to be pre-encoded in the primary site early lesions *via* ZFP281 upregulation and thus, this TF may serve as a new marker of dormant early DCCs. ZFP281 suppresses a fully epithelial phenotype, inducing a growth arrested dormant phenotype in DCCs that is spontaneously operational in early DCCs. Furthermore, loss of ZFP281 leads to reactivation of DCCs and switch to a proliferative phenotype. Thus, an opportunity opens to identify lesions that may carry or not this dormancy program and determine if it informs on dissemination and relapse measures. Indeed, we showed that ZFP281 detection is prevalent in human DCIS samples and significantly decreased in advanced invasive tumors, further supporting the validity of our findings. Since ZFP281 seems to be quite specific for early lesions and early DCCs, and knowing that other TFs, such as NR2F1, when detected in prostate and breast cancer DCCs inform on patient prognosis[45,57], similar studies could be performed for ZFP281. It would be interesting to test if ZFP281 may help measure the abundance of early-like DCCs in patients with early or advanced disease and if it may serve as a marker of relapse.

Several questions that remain unanswered will need additional studies. For example, we have no clear indication of what cues specifically induce ZFP281 in early lesions and early DCCs. The link to M-like phenotypes suggest that signaling like TGFβs, Wnts and BMPs may induce this TF. Further, the role of ZFP281 in PT cells remains unknown. While ZFP281 expression decreases in PT cells it still binds and regulates a different ZFP281-dependent program in PT cells, suggesting that different ZFP281-dependent regulatory networks may operate in EL and PT cells, due likely to the presence/absence of different ZFP281-interacting co-activators and/or repressors[36]. Last, our approach of single cell analysis could not specifically distinguish early DCCs from those exclusively arriving from late lesions. Nevertheless, our data provide unprecedented insight into early DCC fate, demonstrating that ZFP281 regulates an active program of dormancy that must be overridden and precedes a slow proliferation phase towards metastasis. Future studies would also need to pair the analysis of these mechanisms to determine how they influence genetic maturation of early DCCs. Overall, we reveal a unique biology that expands our understanding of metastatic progression that may lead to new markers and strategies to prevent metastasis.

ACKNOWLEDGMENTS

We thank the Aguirre-Ghiso lab for helpful discussions and thank the expertise and assistance of the Dean's Flow Cytometry CoRE and Microscope CoRE, Icahn School of Medicine at Mount Sinai. This work was supported by The National Institute of Health /National Cancer Institute (CA109182, CA216248, CA218024, CA196521) and the Samuel Waxman Cancer Research Foundation Tumor Dormancy Program. A.R.N. was funded by Portuguese Foundation for Science and Technology (SFRH/BD/100380/2014). E.D. was funded by the The National Institute of Health /National Cancer Institute (T32 CA078207). Research in the laboratory of J.W. was funded by grants from NYSTEM (C32569GG; C32583GG) and NIH (R01GM129157; R01HD095938; R01HD097268; R01HL146664).

AUTHOR CONTRIBUTIONS

ARN designed, planned and conducted experiments, analysed data and wrote the manuscript; ED, JY and XH conducted experiments; JY and EK did the bioinformatics analysis; JW provided necessary reagents and developmental biology expertise input; JAAG designed experiments, analysed data and wrote the manuscript.

COMPETING INTERESTS

JAAG is a scientific co-founder of, scientific advisory board member and equity owner in HiberCell and receives financial compensation as a consultant for HiberCell, a Mount Sinai spin-off company

focused on the research and development of therapeutics that prevent or delay the recurrence of cancer.

FIGURE LEGENDS

Figure 1. Early DCCs maintain a global Mesenchymal-like phenotype.

(A) Heatmap of differentially expressed genes (DEGs, STable 1) detected by RNAseq from MMTV-HER2 early lesion (EL) and primary tumor (PT) spheres cultured for 7 days. 2873 upregulated genes (red); 1417 downregulated genes (blue); p-value <0.05 and FC>2 or <0.5.

(B) Gene Set Enrichment analysis[31,32] of EMT (top hallmark hit) and mammary gland luminal, myoepithelial/basal and stem cell signatures[33] in MMTV-HER2 EL *vs*. PT 7-day spheres bulk RNAseq (GSEA STable 3). ES, enrichment score; NES, normalized enrichment score; NOM p-value, nominal p-value <0.05; FDR value, false discovery rate <0.25.

(C) Distribution of the signatures 'Up and Down in EL/PT spheres' (Figure 1A and STable 1) in MMTV-HER2 EL (teal), PT (red), eL (early lungs, blue) and LL (late lungs, orange) DCCs single-cell RNAseq.

(D) MMTV-HER2 EL, PT, eL (early lungs) and LL (late lungs) DCCs single-cell RNAseq sample distribution per cluster. Unsupervised clustering on the DEGs was performed using a previously described batch-aware algorithm[34].

(E) Heatmap of UMI counts of selected epithelial (Ep) and mesenchymal (M) genes (STable 4) in scRNAseq after unsupervised clustering on the DEGs using a previously described batch-aware algorithm[34]. Cell clusters were sub-grouped as EL (1-4), PT (5-6) and DCCs (7-11), according with the predominant cell type in each cluster.

(F) $CD45^{-}HER2^{+}$ cells used for scRNAseq of MMTV-HER2 EL, PT and eL (early lungs) and LL (late lungs) DCCs after tissue dissociation were quantified for expression of EpCAM (epithelial marker) and Eng (mesenchymal marker).

(G) Imaging of sorted $CD45^{-}HER2^{+}$ cells EL, PT, eL and LL DCCs upon 7-day culture in 3D, on top of matrigel. Left: brightfield images, scale 50 um. Right: HER2 (red) expression, scale 25 um.

See also Figure S1.

Figure 2. Early DCCs turn on Mesenchymal- and Pluripotency/Progenitor-like programs that allow them to undergo Dormancy.

(A) Distribution of Epithelial (Ep) and Mesenchymal (M) scores (gene lists in STable 4) in MMTV-HER2 lung DCC clusters after unsupervised clustering on the DEGs using a previously described batch-aware algorithm[34]. Cell clusters were sub-grouped as M-like (1-4, higher M-like score), Hybrid

(5-8) and Ep-like (9-15). Dots color-coded by sample origin: early lung DCCs, eL DCCs, blue; or late lung DCCs, LL DCCs, orange.

(B) Heatmap of UMI counts of selected genes (gene lists in STable 4) in MMTV-HER2 eL (early lungs,) and LL (late lungs) DCCs single-cell RNAseq after unsupervised clustering on the DEGs using a previously described batch-aware algorithm[34]. Clusters 1-15 differentially express gene modules identified in boxed letters. Transcription factors (TF, bottom genes in blue) enriched in each gene module were predicted using Enrichr[29,30]. Lung DCCs show mesenchymal, pluripotency/progenitor-like and dormancy programs that are downregulated once these cells undergo mesenchymal to epithelial transition (MET), fully differentiate and proliferate, potentially giving rise to metastasis (bottom diagram).

(C) Distribution of gene modules B and D in clusters 2, 3, 4, 8 and 9. Dots represent single cells color-coded by cluster (top) and sample origin (eL or LL, bottom).

(D) Distribution of gene modules B and D in clusters 2, 3, 4, 8 and 9. Dots represent single cells color-coded by expression of Ccnd1, Cdkn1c, Nr2f1 and Tgfb2.

See also Figure S2.

Figure 3. Identification of ZFP281 in early lesions (EL) and early DCCs.

(A) Transcription factor (TF) network analysis derived from the RNAseq DEGs from MMTV-HER2 EL and PT spheres. Red, upregulated genes; Green, downregulated genes; Blue, TFs not differentially expressed (DE). Lines connect the TF at the center of the node to target genes indicating that the connected genes have predicted TF binding elements (validated or predicted) in the promoter regions (-500 to +2500 bp from the TSS). Full gene list in STable 6.

(B) Distribution of ZFP281 predicted target scores, summarizing the UMI fraction of ZFP281 predicted targets (ZFP281 node in Figure 3A and STable 6), in all DCC clusters analyzed by scRNAseq (Figure 2).

(C) ZFP281 expression in FvB mammary gland and MMTV- HER2 EL and PT tissues. Representative pictures in Figure S3B. Graph shows n=4, mean, SEM and 2-tailed Mann-Whitney test.

(D and E) ZFP281 (green) and Ki67 (gray) protein expression in MMTV- HER2 (HER2, red) lung DCCs. Scales, 20 um. Graph shows n=4, mean, SEM and 2-tailed Mann-Whitney test.

(F and G) ZFP281 (green) protein detection in human ductal carcinoma in situ (DCIS) and invasive breast cancer (IBC) samples. Scales, 25 um. Graph shows n=14, mean, SEM and 2-tailed Mann-Whitney test.

See also Figure S3.

Figure 4. ZFP281 regulates an EMT-like program both in primed mEpiSCs and early DCCs.

(A) Correlation of EMT and Wnt signaling DEGs in primed *versus* naïve mouse pluripotent stem cells (RNASeq data, GSE81044 from [36]) and EL vs. PT cells (bulk RNAseq, Figure 1A). Gene lists in STable 7 and 8.

(B) Venn diagrams for ZFP281 targets identified by ChIPseq from EL cells and primed mEpiSCs ChIPSeq data, GSE93042 from [58]) and cell cycle arrest, EMT, Wnt and FGF signaling genes. Comparisons and statistics were done in pairs (ZFP281 targets in EL vs. Cell cycle arrest genes; ZFP281 targets in EL vs. EMT genes; ZFP281 targets in EL vs. Wnt genes; ZFP281 targets in EL vs. FGFR genes; and same for ZFP281 targets primed mEpiSCs). Overlapped genes between these paired comparisons were rare so for graphical simplification, the 4 comparisons are displayed together. Common genes (central blue box) correspond to genes from each category that are targeted by ZFP281 both in EL and mEpiSCs. Gene lists in STable 9.

(C) ChIPseq and RNAseq data in EL over PT cells. ZFP281 binds EMT (Snai1, Vim, Zeb1), cell-cycle (Cdk2 and Cdkn1a), and dormancy (Tgfbr1, Nr2f1 and Bmp7) associated genes. All genes were upregulated in EL cell; Snai1, Vim, Zeb1, Cdk2, Tgfbr1, and Nr2f1 were ZFP281 bound genes exclusively in EL cells; Cdkn1a, and Bmp7 were bound by ZFP281 in both EL and PT cells but only upregulated in EL cells.

(D) Distribution of ZFP281 target (ChIP data) scores, summarizing the averaged expression of ZFP281 targets, in all DCC clusters analyzed by scRNAseq (Figure 2).

(E) Distribution of gene modules B and D in all DCC clusters. Dots represent single cells color-coded by ZFP281 target (ChIP data) scores (low, red, to high, green).

See also Figure S4.

Figure 5. ZFP281 induces an M-like slow-cycling phenotype *in vitro*.

(A) Column of representative images of the mammosphere phenotype of EL shZFP281±DOX, PT Control and PT ZFP281-overexpressed (OE) cells. Scale 50 um.

(B) Quantification of mammosphere (MS) frequency of EL shZFP281±DOX, PT Control and PT ZFP281-OE cells. Graph shows n=3 mean, SEM and 2-tailed Mann-Whitney test.

(C) Quantification of mammosphere (MS) size, as number of cells per sphere after dissociation of EL shZFP281±DOX, PT Control and PT ZFP281-OE spheres. Graph shows n=3, mean, SEM and 2-tailed Mann-Whitney test.

(D) EpCAM (epithelial marker) and Eng/CD105 (mesenchymal marker) expression in EL shZFP281±DOX, PT Control and PT ZFP281-OE cells. Representative experiment of n=3 biological replicates.

(E) Fold-change of Ep-like (EpCAM+Eng-), hybrid (EpCAM+Eng+) and M-like (EpCAM-Eng+) populations in EL shZFP281+DOX over –DOX and PT ZFP281-OE over PT Control spheres. Graph shows n=3, mean, SEM and 2-tailed Mann-Whitney test.
(F) Column of representative images of 3D-matrigel organoids and invasive phenotype of EL shZFP281±DOX, PT Control and PT ZFP281-OE cells. Scale 50 um.
(G)Quantification of percentage of 3D-matrigel spheroids with invasive protusions per condition. Graph shows n=4, mean, SEM and 2-tailed Mann-Whitney test.
See also Figure S5.

Figure 6. ZFP281 controls the M-like program that leads to a switch from dormant early DCCs to proliferation *in vivo*.
(A) Representative images of HER2 (red) and ZFP281 (green) protein expression in mammary fat pads of mice 5 month after EL shZFP281 sphere injections. Mice were given water 1) without doxycycline (DOX) for 5 months: '-DOX'; 2) with DOX for 5 months: '+DOX' or 3) 1 month without and 4 months with DOX: '-DOX +DOX'. Arrows point to ZFP281 expression in EL shZFP281-DOX, which is downregulated in groups '+DOX' and '-DOX +DOX'. Scales 25um (top row) and 50um (bottom row, inserts).
(B) Representative images of lung DCCs, single cells and metastasis quantified in C. HER2, red; ZFP281, green; DAPI, blue. Scales 25um.
(C) Frequency of lung single cell (SC) and metastasis per lung section per mouse for all conditions, 5 months after EL shZFP281 sphere injections. 2 lung slides with all lobules represented were scanned and quantified per mouse. Graph shows n=5-10 mice/condition, median and 2-tailed Mann-Whitney test.
(D) Representative images of HER2 (red) and ZFP281 (green) protein expression in primary tumors 71 days post PT Control and PT ZFP281-OE cell mammary fat pad injection. Scales 25um.
(E) Tumor volume over time of PT Control and PT ZFP281-OE. Tumors were removed at day 71 and mice sacrificed 5 months after cancer cells injections (corresponding lungs in G). Graph shows n=10 per condition, median, interquartile range and 2-tailed Mann-Whitney test.
(F and G) Frequency of lung single cell (SC) and metastasis per mice per condition, 2 (F) and 5 (G) months after mammary fat pad injection of PT Control or PT ZFP281-OE spheres. 2 lung slides with all lobules represented were scanned and quantified per mouse. Graph shows n=5 mice/condition, median and 2-tailed Mann-Whitney test.
(H) H&E images of mice lungs 5 months after mammary fat pad injection of PT Control or PT ZFP281-OE spheres. Scales 2mm.

(I) Quantification of lung metastasis burden, normalized for total lung area, of images in B (1 slide per mouse). Graph shows n=5 mice/condition, median and 2-tailed Mann-Whitney test.
See also Figure S6.

Figure 7. Model of ZFP281 regulated dissemination and dormancy states during early cancer evolution. Early upon cancer initiation via HER2 signaling, EL cells activate the primed pluripotency transcription factor ZFP281 (upper section). This transcription factor regulates at least four distinct mesenchymal- and pluripotency-like programs, leading to EL cells to disseminate and to enter a prolonged dormancy as early DCCs in lungs (lower left, Early Stage). The M-like dormancy and primed pluripotency program is associated with dormancy programs, such as those controlled by NR2F1 and TGFβ2, and block the acquisition of an epithelial state. Over time, intrinsic and microenvironmental changes allow the dormant DCCs to disrupt ZFP281 function and adopt an Ep-like phenotype (lower right, Late Stage), which enables a proliferative state. Importantly, M-like, hybrid and Ep-like DCCs co-exist in both early and late stage lungs, with predominance of M-like dormant DCCs in early stage lungs.

Figure S1.

(A) Experimental design of MMTV-Neu bulk and single cell RNA sequencing.
(B) Enrichr analysis[29,30] of differentially expressed genes (DEG) in MMTV-HER2 early lesion (EL) and primary tumor (PT) 7-day spheres bulk RNAseq. Full table in STable 2. Orange, terms mentioned in the text.
(C) Biological negative controls used for FACS gating strategy. FvB mammary gland (MG) was used to set the EL and PT gate and FvB lungs for eL and LL DCCs (see Figure 1F).
(D) Percentage of epithelial ($EpCAM^+Eng^-$), hybrid ($EpCAM^+Eng^+$) and mesenchymal ($EpCAM^-Eng^+$) populations in $CD45^-HER2^+$ MMTV-HER2 EL, PT and eL (early lungs) and LL (late lungs) DCCs after tissue dissociation (representative FACS plots in Figure 1F).

Figure S2.

(A) Distribution of Epithelial (Ep) and Mesenchymal (M) scores (gene lists in STable 4, showed in Figure 2A) in MMTV-HER2 lung DCC clusters. Cell clusters were sub-grouped as M-like (1-4, higher M-like score), Hybrid (5-8) and Ep-like (9-15).
(B) Normal lung cells (grey), eL (early lungs, blue) and LL (late lungs, orange) DCCs single-cell RNAseq sample distribution per cluster. Unsupervised clustering on the DEGs was performed using a previously described batch-aware algorithm[34].

(C) Heatmap of UMI counts of selected genes (gene lists in STable 4) in MMTV-HER2 normal lung cells and eL and LL DCCs single-cell RNAseq. N1-10 are clusters enriched in non-cancer ($HER2^-$) lung cells and excluded in the analysis. Clusters 1-15have with less than 16% of non-cancer ($HER2^-$) lung cells, so non-cancer lung cells were excluded further analysis but these clusters were considered cancer cell clusters.

(D) Distribution of gene modules B and D (M-like) in all DCC clusters. Dots represent single cells color-coded by cluster (left), sample origin (eL or LL, middle) and sub-gourp (Ep-like, hybrid, M-like, right). Gene module lists in STable 4.

(E) Distribution of gene modules I (Ep-like) and D (M-like) in all DCC clusters. Dots represent single cells color-coded by cluster (left), sample origin (eL or LL, middle) and sub-gourp (Ep-like, hybrid, M-like, right). Gene module lists in STable 4.

(F) Heatmap of UMI counts of selected genes (gene lists in STable 4) in MMTV-HER2 eL (early lungs,) and LL (late lungs) DCCs single-cell RNAseq after unsupervised clustering on the DEGs and down-sampling to 500 UMI per cell. 'Per cell' representation of Figure 2B heatmap, which shows UMI averages.

Figure S3.

(A) mRNA expression of ZFP281, its predicted targets (Figure 3A) and EMT genes in EL *vs*. PT cells, EL shCt, EL shZFP281 and PT ZFP281-OE. Red, upregulated genes; Blue, downregulated genes; *p-value <0.05.

(B) Representative images of ZFP281 (1^{st} column, green), E-cadherin (2^{nd} column, green) and Twist1 (3^{rd} column, green) protein expression in consecutive sections of FvB mammary gland (FvB MG, biological negative control) and MMTV- HER2 EL and PT tissues. HER2 expression in red. Arrows point to $ZFP281^+Ecad^{low}Twist1^+$ cells in EL. Dashed arrow points to $ZFP281^+$ adipocytes (internal control). Scales, 20 um.

Figure S4.

(A) Heatmap of combined RNAseq and ChIP seq data from EL/PT cells. 504 genes show higher ZFP281 binding and higher expression in EL *vs* PT cells; 118 genes show higher ZFP281 binding and higher expression in PT *vs* EL cells; 41 genes show higher expression while lower binding in PT *vs* EL cells; 63 genes show higher expression while lower binding in EL *vs* PT cells. Gene lists in STable 10.

(B) Venn diagram of EL/PT RNAseq (Figure 1A), ZFP281 node (Figure 3A) and ChIPseq (Figure 4B) data. Targets of ZFP281 in EL cells and EpiSCs were identified from ChIP-seq data and further used to compare with EMT, Wnt, FGFR, and cell cycle arrest genes.
(C) Representative tracks of EL/PT ChIPseq (Figure 4B-C). Example genes, Snai1, Tgfbr1, Vim, Zeb1, Cdk2, and Cdkn1a are used to show the difference binding between EL and PT cells.
(D) Frequency of ZFP281 target (ChIP) score, summarizing the averaged expression of ZFP281 targets, in all cells analyzed by scRNAseq (Figure 2).
(E) Distribution of gene modules I (Ep-like) and B (M-like) in all DCC clusters. Dots represent single cells color-coded by ZFP281 target scores (low, red to high,green).

Figure S5.

(A) Column of representative images of the mammosphere phenotype of EL, PT and EL shControl±DOX cells. Scale 50 um.
(B) Quantification of mammosphere (MS) frequency of EL, PT and EL shControl±DOX cells. Graph shows n=3, mean, SEM and 2-tailed Mann-Whitney test.
(C) Quantification of mammosphere (MS) size, as number of cells per sphere after dissociation of EL, PT and EL shControl±DOX spheres. Graph shows n=3, mean, SEM and 2-tailed Mann-Whitney test.
(D) EpCAM (epithelial marker) and Eng/CD105 (mesenchymal marker) expression in EL, PT and EL shControl±DOX cells. Representative experiment of n=3 biological replicates.
(E) Fold-change of Ep-like (EpCAM+Eng-), hybrid (EpCAM+Eng+) and M-like (EpCAM-Eng+) populations in EL over PT and EL shControl±DOX spheres. Graph shows n=3, mean, SEM and 2-tailed Mann-Whitney test.
(F) Column of representative images of 3D-matrigel organoids and invasive phenotype of EL, PT and EL shControl±DOX cells. Scale 50 um.
(G) Quantification of percentage of 3D-matrigel spheroids with invasive protusions per condition. Graph shows n=4, mean, SEM and 2-tailed Mann-Whitney test.

Figure S6.

(A) Quantification of Ki67+ cells in lung metastasis 5 months after EL shZFP281 sphere injections. Graph shows n=5 mice/condition, median and 2-tailed Mann-Whitney test.
(B) Quantification of Ki67+ cells in lung metastasis 5 months after PT Control or PT ZFP281-OE sphere injections. Graph shows n=5 mice/condition, median and 2-tailed Mann-Whitney test.

Supplementary Table 1. Differentially expressed genes (DEGs) in MMTV-HER2 early lesion (EL) and primary tumor (PT) 7-day spheres RNAseq. Heatmap on Figure 1A.

Supplementary Table 2. Enrichr analysis[29,30] of differentially expressed genes (DEG) in MMTV-HER2 early lesion (EL) and primary tumor (PT) 7-day spheres bulk RNAseq. Partially showed on Figure S1B.

Supplementary Table 3. GSEA analysis[31,32] of hallmark pathways up and downregulated in EL over PT cells, as well as mammary gland luminal *vs.* myoepithelial and stem cell signatures [33]. Partially showed on Figure 1B.

Supplementary Table 4. Gene lists used to generate each Figure panel.

Supplementary Table 5. Enrichr analysis[29,30] of Gene modules from Figure 2B.

Supplementary Table 6. Transcription factor (TF) network analysis from Figure 3A.

Supplementary Table 7. Comparison of EMT-related DEG in RNAseq and ChIP seq data from Primed/Naïve mouse pluripotent stem cells (RNASeq data, GSE81044 from [36]) and EL/PT cells (bulk RNAseq, Figure 1A). Diagram on Figure 4A.

Supplementary Table 8. Comparison of Wnt signaling DEG in RNAseq and ChIP seq data from Primed/Naïve mouse pluripotent stem cells (RNASeq data, GSE81044 from [36]) and EL/PT cells (bulk RNAseq, Figure 1A). Diagram on Figure 4A.

Supplementary Table 9. Comparison of ZFP281 targets in EL (ChIPseq data) and primed mEpiSCs (ChIPSeq data, GSE93042 from [58]) per category: cell cycle arrest, EMT, Wnt and FGF signaling. Venn diagrams on Figure 4B.

Supplementary Table 10. Gene lists of DEG in RNAseq and ChIP seq data from EL/PT cells. Heatmap on Figure S4A.

Supplementary Table 11. Gene lists of ZFP281 targets identified by ChIP seq.

Supplementary Table 12. Table of primers.

Supplementary Table 13. Table of antibodies and conditions.

EXPERIMENTAL PROCEDURES

Animal experiments. MMTV-HER2/Neu mice were maintained on FvB background and bred and crossed in our facilities. 14 to 18-week-old female mice were used as early ('pre-malignant') stage mice and 20-week-old or older females with palpable tumor(s) were used as late stage of cancer progression. No randomization or blinding was used to allocate experimental groups. Tumous were not allowed to grow beyond the IACUC allowed limit of 1 cm^3 per animal. All experimental procedures were approved by the Institutional Animal Care and Use Committee (IACUC) of Icahn School of

Medicine at Mount Sinai.

MMTV-HER2/Neu mice were euthanized using isoflurane and cervical dislocation. All 5 pairs of mammary glands were checked for the presence of any visible small lesions or palpable tumors. Mice were perfused with PBS and organs were collected. For histopathology, organs were fixed in 4% paraformaldehyde (PFA, Thermoscientific) for 24 hours, processed, embedded in paraffin and sections were cut. For FACS and cell culture preparations, whole mammary glands, primary tumors and/or lungs were digested in 0.15% Collagenase 1A (SIGMA, C-9891) 2.5% bovine serum albumin (BSA) at 37°C with agitation for 30 min. Red blood-cell lysis buffer (Lonza) was used for 2-5 minutes, cells were filtered through a 40-μm filter, passed through a 25-gauge needle and counted. CD45 depletion (MACS, mouse CD45 MicroBeads) was performed for some experiments following manufactures' instructions.

RNA sequencing. RNA from EL and PT spheres (after 7 days in cultures) was extracted using RNeasy protocol (Qiagen) and sequenced using Illumina MiSeq. The RNA-Seq data was analyzed using Basepair software (https://urldefense.proofpoint.com/v2/url?u=https-3A__www.basepairtech.com_&d=DwIFaQ&c=shNJtf5dKgNcPZ6Yh64b-A&r=2eTrMIlet1N-adSlZc06oeRObUwVMmPYgdoyfpjVSin5Kgoy-anjqY2o5qXmg6s_&m=JP2DaO0EeJHYY-pErtgneKGYD4_scJPda-QP6fmXaok&s=Pk2IgzqD_rcZILd5PYdfYTol9LmHQNL3dNIMjv5M6bE&e=) with a pipeline that included the following steps. Reads were aligned to the transcriptome derived from UCSC genome assembly (((hg19))) using STAR[59] with default parameters. Read counts for each transcript was measured using featureCounts[60]. Differentially expressed genes were determined using DESeq2[61] and a cut-off of 0.05 on adjusted p-value (corrected for multiple hypotheses testing) was used for creating lists and heatmaps. GSEA was performed on normalized gene expression counts, using gene permutations for calculating p-value. The RNA-Seq data was further analyzed using Enrichr[29,30] and GSEA[31,32].

Network analysis. Bioinf2bio did this analysis. The genomic sequence corresponding to the promoter (ranging from 2500 bp upstream from the TSS until 500 bp after the TSS) of each DEG was extracted using the database UCSC (mm10). Next, we retrieved from JASPAR all available position weight matrices (PWM) corresponding to all known mouse TFs and for the identification of the TF binding site (TFBS) we used the TFBSTools package[62] (http://bioconductor.org/packages/release/bioc/html/TFBSTools.html) designed to be a computational framework for TF binding analysis. We screened each DE-gene promoter sequence for all putative TFBS predicted in each of the retrieved PWMs. TFBS were scanned in both strands with a 'min.score.percentage' parameter set to 95%. For the network construction we have used a matrix built from all the TFBS predicted within the promoters of all DE-genes selected. Of notice, we

detected 581121 TFBS with p-value below $1E^{-04}$, so we selected only the TFBS with high scores (above 21).

Single-cell RNA sequencing. Mammary glands of 'early stage' mice, primary tumors from 'late stage' mice and lungs from 'early' and 'late stage' mice were dissected and digested (see 'Animal experiments' section). For the 1st scRNAseq experiment (Figure 1), early lesion (EL), primary tumor (PT) and early and late DCCs were sorted ($CD45^-$ $HER2^+$), while for the 2nd experiment (Figure 2 to 4) non-cancer lung cells ($CD45^-HER2^-$) and early and late DCCs ($CD45^-$ $HER2^+$) were sorted. After sorting, cells were encapsulated using the 10X Chromium 3' v2 (1st experiment) or v3 (2nd experiment) and chemistry kit according to manufacturer instructions. Sequencing, libraries were prepared according to manufacturer instructions. QC of cDNA and final libraries was performed by CyberGreen qPCR library quantification assay (KAPA). Samples were sequenced on an Illumina Nextseq 550 using the 75-cycle kit to a depth of 100 million reads per library. Single-cell clustering: Single-cell datasets (1st and 2nd experiments) were clustered separately using an unsupervised clustering algorithm previously described[34]. In this EM-like algorithm, parameters for multinomial cell-type specific gene-expression models are learned together with a parameter for the fraction of background noise that is associated with cells in each sample. Mitochondrial genes, Malat1 were excluded from the clustering process. The clustering parameters were chosen to accommodate the different UMI and cell counts between the experiments. In the 1st experiment, the minimum number of UMIs per cell was 300, the number of clusters k was 12, and $(P_1,P_2) = (30^{th},60^{th})$ percentiles. In the 2nd experiment, the minimum number of UMI per cell was 800, k was set to 20, and $(P_1,P_2) = (0^{th},20^{th})$ percentiles. Gene modules and gene scores: Gene modules were based on a gene-covariance analysis as was applied as in [34]. Briefly, cells were down-sampled and variable genes were selected based on the variance to mean ratio. Gene-to-gene correlations were estimated per sample and were averaged following z-transformation. The averaged correlation matrices were hierarchically clustered into gene-"modules". Given a gene-list, we calculated its score per cell by summing up the UMIs of the genes in this cell and divided the sum by the total sum of UMI in the cell. The score therefore equals to the fraction of UMIs associated with the genes in the list.

ChIP sequencing. ChIP was performed using EZ ChIP protocol (Millipore, Massachusetts, USA) with Zfp281 antibody (ab101318, Abcam) for EL and PT samples (after 7 days iin cultures). High-throughput sequencing was then used to get the ChIP-seq data. ChIP-seq reads were aligned to the mm10 genome using bowtie2 (v2.3.4.3), followed by removing PCR duplicates using Picard with the parameter REMOVE_DUPLICATES=true. ChIP-seq peaks were determined by the MACS program (v.2.1.2) using input ChIP-seq as the control data, and all other parameters followed the default setting. Binding difference around the transcription start sites [-5kb, +5kb] between the EL and PT

samples are analyzed using the DiffBind (v2.1.6). ChIP-seq data was compared with RNA-seq data analysis after RNA-seq reads were aligned to the mouse mm10 genome using Bowtie2 (v2.3.4.3). The aligned bam files were sorted by name using the parameter -n. We used the HTSeq software (v0.11.2) and mm10 annotation file from GENCODE (version M19) to count reads for each gene using parameters -r name -f bam, and BioMart[63] to retrieve corresponding genes names. Finally, read counts were normalized with the trimmed mean of M-values (TMM) method[64] for differential expression analysis using edgeR (v3.26.8)[65]. Public RNA-seq data were downloaded (refer to Key Resource Table) and aligned to mm10, and followed with the same processing setting. Cell cycle arrest, EMT, FGF signaling, and Wnt signaling gene sets were downloaded from MsigDB (https://www.gsea-msigdb.org/gsea/msigdb/index.jsp) with systematic names M1134, M5930, M1090, and M7847, respectively. All processed and index sorted bam files of high-throughput sequencing data were converted to TDF files using count command of igvtools, followed by visualization using IGV software[66,67].

Flow cytometry and cell sorting. Mammary glands of 'early stage' mice, primary tumors from 'late stage' mice and lungs from 'early' and 'late stage' mice were dissected and digested (see 'Animal experiments' section). In case of cells in culture, single cell suspensions were obtained by incubating the cells in accutase (Sigma) for 20 minutes, at 37C. Cells were stained using antibodies and conditions in Supplementary Table 13. All experiments were performed using BD FACSAria II sorter equipped with FACS Diva software (BD Biosciences) or analyzed using Aurora analyzer (Cytek Biosciences) equipped with SpectroFlo software. Dead cells and debris were excluded by FCS, SSC and DAPI (4',6-diamino-2-phenylindole) (Fisher Scientific) staining profiles. Data were analyzed with FACS Diva (BD Biosciences) or FCS Express Cytometry 7 (De Novo) softwares.

Cell culture. Mammary glands of 'early stage' mice, primary tumors from 'late stage' mice and lungs from both stages were dissected and digested (see 'Animal experiments' section). For sphere cultures, $5x10^5$ cells were seeded in 6-well ultra-low adhesion plates in 1 ml mammosphere media (DMEM/F12 (Gibco 11320-082), 1:50 B27 (Invitrogen 17504-044), 500ng/ml Hydrocortisone (Lonza CC-403), 40 µg/ml Insulin (Gibco 12585-014), 20 ng/ml EGF (Peprotech AF-100-15-A), 100 units/ml penicillin and 100ug/ml streptomycin (Corning)) supplemented with 0.5% methylcellulose (R&D systems HSC001). Sphere-forming capacity was measured by quantification of number of spheres per well after 7 days in culture. Spheres were then dissociated with accutase (Sigma) and number of cells per sphere was calculated as a measurement of sphere size.

EL cells were transduced with lentivirus pTRIPZ (shControl) or shZFP281 (V2THS_42594, Open Biosystems) as previously described[68] at day 0 of sphere formation. Cultures were treated every 24 hours, starting at day 1, with 2ug/ml DOX. PT cells were transfected with ZFP281-OE plasmid (pB-

3XFL-ZFP281) using Lipofectamine 3000 transfection reagent (Invitrogen) according to the manufacturer's instructions.

For organoid cultures, cells were seeded in 8-well chamber slides coated with 50ul of Matrigel (Corning, growth factor reduced) per slide in 400ul of assay medium (DMEM/F12, 5% horse serum, 500ng/ml Hydrocortisone (Lonza CC-403), 40 µg/ml Insulin (Gibco 12585-014), 100 units/ml penicillin and 100ug/ml streptomycin (Corning), and 2% Matrigel (Corning, growth factor reduced). 4 pictures of random fields per well were analyzed to quantify the percentage of invasive structures. All in vitro experiments were performed and analyzed using 4 wells per condition (technical replicates) and at least 3 independent experiments (biological replicates).

Immunofluorescence. Tissues slides (see 'Animal experiments' section) were dehydrated, followed by antigen retrieval 10 mM citrate buffer pH 6.0 ($Na_3H_6H_5O_7$). Blocking was done using 0.5% BSA in PBS with 5% normal goat serum (Thermofisher PCN5000) for 1 hour. Antibodies and incubation conditions used are summarized in Supplementary Table 13. For ZFP281 detection, Alexa Fluor™ 488 Tyramide SuperBoost™ Kit goat anti-mouse IgG (Invitrogen) was used for amplification of the signal. All slides were mounted with ProLong Gold Antifade reagent with DAPI (Invitrogen P36931).

3D cultures were fixed with 4% PFA for 20 min at room temperature, permeabilized with 0.5% Triton X-100 in PBS for 20 min and blocking was done using 1Å~ immunofluorescence PBS wash buffer (130 mM NaCl; 7 mM Na2HPO4; 3.5 mM NaH2PO4; 7.7 mM NaN3; 0.1 %BSA; 0.2% Triton X-100; 0.05% Tween-20) containing 5% normal goat serum (Thermofisher PCN5000) for 1h. Antibodies and conditions used are summarized in Supplementary Table 13. Chambers were removed from slides and wells were fixed and mounted with ProLong Gold Antifade reagent with DAPI (Invitrogen P36931). Images were obtained using a Leica SPE high-resolution spectral confocal microscope and Leica software.

Quantitative PCR. Spheres were processed using Cell-to-CT 1-Step Power SYBR Green kit (Invitrogen, A25600) and primers from Supplementary Table 12. GAPDH was used as housekeeping control for all experiments.

***In vivo* experiments.** 300 EL or 150-300 PT spheres were injected per site into nude mice (BALB/cnu/nu, Charles River) in 100ul of a 1:1 PBS-Matrigel solution (Corning, growth factor reduced). Spheres were injected in the two fourth inguinal gland fat pad using a 27-gauge needle. Mice injected with sh-TRIPZ-shZFP281 were given control drinking water (-DOX), water with doxycycline from day 0 (+DOX) or water with doxycycline starting 1 month after sphere injection (-DOX +DOX) until the end of the experiment, 5 months after spheres injection. In the case of mice injected with tumour-derived spheres, tumours were removed before reaching $1cm^3$, according to IAUCU regulations. Mice were euthanized and organs were collected and processed 2 or 5 month

after cancer cell injections. Immunofluorescences were performed 2 sections per mice were used to quantify and characterize single DCCs and metastasis. H&E slides were scanned using NanoZoomer S60 Digital slide scanner and NDP.view2 software (Hamamatsu) and metastasis area was calculated and normalized for the total area of the lungs.

Patient samples. Paraffin-embedded sections from ductal carcinoma *in situ* (DCIS) and invasive breast cancer (IBC) lesions were collected from the Cancer Biorepository at Icahn School of Medicine at Mount Sinai, New York, New York. Samples were de-identified and obtained with Institutional Review Board approval, which indicated that this work does not meet the definition of human subject research according to the 45 CFR 46 and the Office of Human Subject Research. 28 samples were analyzed, 14 DCIS and 14 IBC.

Statistical analysis. Sample sizes were chosen empirically and no exclusion criteria were applied. The investigators were not blinded to allocation during experiments but quantifications were done in coded samples to reduce operator bias. Statistical analyses were done using Prism Software and differences were considered significant if $p<0.05$. Unless otherwise specified, 3 or more independent experiments were performed, all values were included and median, interquartile range and 2-tailed Mann–Whitney U-tests were performed.

Data availability. The datasets generated during and/or analyzed during the current study are available within the paper (and its Supplementary Information) and/or from the corresponding author on reasonable request. All sequencing data will be made available in a public data repository.

REFERENCES

1. Klein, C. A. Cancer progression and the invisible phase of metastatic colonization. *Nature Reviews Cancer* (2020) doi:10.1038/s41568-020-00300-6.
2. Pantel, K. & Brakenhoff, R. H. Dissecting the metastatic cascade. *Nat. Rev. Cancer* **4**, 448 (2004).
3. Valastyan, S. & Weinberg, R. A. Tumor metastasis: Molecular insights and evolving paradigms. *Cell* (2011) doi:10.1016/j.cell.2011.09.024.
4. Yates, L. R. *et al.* Genomic Evolution of Breast Cancer Metastasis and Relapse. *Cancer Cell* (2017) doi:10.1016/j.ccell.2017.07.005.
5. Schardt, J. A. *et al.* Genomic analysis of single cytokeratin-positive cells from bone marrow reveals early mutational events in breast cancer. *Cancer Cell* **8**, 227–39 (2005).
6. Sanger, N. *et al.* Disseminated tumor cells in the bone marrow of patients with ductal carcinoma in situ. *Int. J. Cancer* (2011) doi:10.1002/ijc.25895.
7. Hosseini, H. *et al.* Early dissemination seeds metastasis in breast cancer. *Nature* (2016)

doi:10.1038/nature20785.

8. Ullah, I. *et al.* Evolutionary history of metastatic breast cancer reveals minimal seeding from axillary lymph nodes. *J. Clin. Invest.* (2018) doi:10.1172/JCI96149.
9. Casasent, A. K. *et al.* Multiclonal Invasion in Breast Tumors Identified by Topographic Single Cell Sequencing. *Cell* (2018) doi:10.1016/j.cell.2017.12.007.
10. Rhim, A. D. *et al.* Detection of circulating pancreas epithelial cells in patients with pancreatic cystic lesions. *Gastroenterology* (2014) doi:10.1053/j.gastro.2013.12.007.
11. Makohon-Moore, A. P. *et al.* Precancerous neoplastic cells can move through the pancreatic ductal system. *Nature* (2018) doi:10.1038/s41586-018-0481-8.
12. Muzumdar, M. D. *et al.* Clonal dynamics following p53 loss of heterozygosity in Kras-driven cancers. *Nat. Commun.* (2016) doi:10.1038/ncomms12685.
13. Shain, A. H. *et al.* The genetic evolution of metastatic uveal melanoma. *Nat. Genet.* (2019) doi:10.1038/s41588-019-0440-9.
14. Hu, Z. *et al.* Quantitative evidence for early metastatic seeding in colorectal cancer. *Nat. Genet.* (2019) doi:10.1038/s41588-019-0423-x.
15. Turajlic, S. *et al.* Tracking Cancer Evolution Reveals Constrained Routes to Metastases: TRACERx Renal. *Cell* (2018) doi:10.1016/j.cell.2018.03.057.
16. Li, C. *et al.* Mutational landscape of primary, metastatic, and recurrent ovarian cancer reveals c-MYC gains as potential target for BET inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* (2019) doi:10.1073/pnas.1814027116.
17. Dorssers, L. C. J. *et al.* Molecular heterogeneity and early metastatic clone selection in testicular germ cell cancer development. *Br. J. Cancer* (2019) doi:10.1038/s41416-019-0381-1.
18. Wang, D. *et al.* Multiregion sequencing reveals the genetic heterogeneity and evolutionary history of osteosarcoma and matched pulmonary metastases. *Cancer Res.* (2019) doi:10.1158/0008-5472.CAN-18-1086.
19. Harper, K. L. *et al.* Mechanism of early dissemination and metastasis in Her2+mammary cancer. *Nature* (2016) doi:10.1038/nature20609.
20. Rhim, A. D. *et al.* EMT and dissemination precede pancreatic tumor formation. *Cell* (2012) doi:10.1016/j.cell.2011.11.025.
21. Werner-Klein, M. *et al.* Genetic alterations driving metastatic colony formation are acquired outside of the primary tumour in melanoma. *Nat. Commun.* (2018) doi:10.1038/s41467-017-02674-y.
22. Eyles, J. *et al.* Tumor cells disseminate early, but immunosurveillance limits metastatic outgrowth, in a mouse model of melanoma. *J. Clin. Invest.* (2010) doi:10.1172/JCI42002.

23. Brisken, C. *et al.* Essential function of Wnt-4 in mammary gland development downstream of progesterone signaling. *Genes Dev.* (2000) doi:10.1101/gad.14.6.650.
24. Linde, N. *et al.* Macrophages orchestrate breast cancer early dissemination and metastasis. *Nat. Commun.* (2018) doi:10.1038/s41467-017-02481-5.
25. Guy, C. T. *et al.* Expression of the neu protooncogene in the mammary epithelium of transgenic mice induces metastatic disease. *Proc. Natl. Acad. Sci. U. S. A.* (1992) doi:10.1073/pnas.89.22.10578.
26. Herschkowitz, J. I. *et al.* Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. *Genome Biol.* (2007) doi:10.1186/gb-2007-8-5-r76.
27. Lu, P., Takai, K., Weaver, V. M. & Werb, Z. Extracellular Matrix degradation and remodeling in development and disease. *Cold Spring Harb. Perspect. Biol.* (2011) doi:10.1101/cshperspect.a005058.
28. Nieto, M. A., Huang, R. Y. Y. J., Jackson, R. A. A. & Thiery, J. P. P. EMT: 2016. *Cell* (2016) doi:10.1016/j.cell.2016.06.028.
29. Chen, E. Y. *et al.* Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* (2013) doi:10.1186/1471-2105-14-128.
30. Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* (2016) doi:10.1093/nar/gkw377.
31. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* (2005) doi:10.1073/pnas.0506580102.
32. Mootha, V. K. *et al.* PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat. Genet.* (2003) doi:10.1038/ng1180.
33. Lim, E. *et al.* Transcriptome analyses of mouse and human mammary cell subpopulations reveal multiple conserved genes and pathways. *Breast Cancer Res.* (2010) doi:10.1186/bcr2560.
34. Martin, J. C. *et al.* Single-Cell Analysis of Crohn's Disease Lesions Identifies a Pathogenic Cellular Module Associated with Resistance to Anti-TNF Therapy. *Cell* (2019) doi:10.1016/j.cell.2019.08.008.
35. Barbara, N. P., Wrana, J. L. & Letarte, M. Endoglin is an accessory protein that interacts with the signaling receptor complex of multiple members of the transforming growth factor-β superfamily. *J. Biol. Chem.* (1999) doi:10.1074/jbc.274.2.584.
36. Fidalgo, M. *et al.* Zfp281 Coordinates Opposing Functions of Tet1 and Tet2 in Pluripotent

States. *Cell Stem Cell* (2016) doi:10.1016/j.stem.2016.05.025.

37. Seo, K. W. *et al.* ZNF281 knockdown induced osteogenic differentiation of human multipotent stem cells in vivo and in vitro. *Cell Transplant.* (2013) doi:10.3727/096368912X654948.
38. Nicolai, S. *et al.* ZNF281/Zfp281 is a target of miR-1 and counteracts muscle differentiation. *Mol. Oncol.* (2020) doi:10.1002/1878-0261.12605.
39. Hahn, S., Jackstadt, R., Siemens, H., Hünten, S. & Hermeking, H. SNAIL and miR-34a feed-forward regulation of ZNF281/ZBP99 promotes epithelial-mesenchymal transition. *EMBO J.* (2013) doi:10.1038/emboj.2013.236.
40. Micalizzi, D. S., Farabaugh, S. M. & Ford, H. L. Epithelial-mesenchymal transition in cancer: Parallels between normal development and tumor progression. *Journal of Mammary Gland Biology and Neoplasia* (2010) doi:10.1007/s10911-010-9178-9.
41. Kim, D. *et al.* Epithelial Mesenchymal Transition in Embryonic Development, Tissue Repair and Cancer: A Comprehensive Overview. *J. Clin. Med.* (2017) doi:10.3390/jcm7010001.
42. Gu, P. *et al.* Orphan Nuclear Receptor LRH-1 Is Required To Maintain Oct4 Expression at the Epiblast Stage of Embryonic Development. *Mol. Cell. Biol.* (2005) doi:10.1128/MCB.25.9.3492-3505.2005.
43. Liu, L. *et al.* Nr5a2 promotes tumor growth and metastasis of gastric cancer AGS cells by Wnt/beta-catenin signaling . *Onco. Targets. Ther.* (2019) doi:10.2147/ott.s201228.
44. Luo, Z. *et al.* Effect of NR5A2 inhibition on pancreatic cancer stem cell (CSC) properties and epithelial-mesenchymal transition (EMT) markers. *Mol. Carcinog.* (2017) doi:10.1002/mc.22604.
45. Sosa, M. S. *et al.* NR2F1 controls tumour cell dormancy via SOX9- and RARβ-driven quiescence programmes. *Nat. Commun.* **6**, 1–14 (2015).
46. Acloque, H., Adams, M. S., Fishwick, K., Bronner-Fraser, M. & Nieto, M. A. Epithelial-mesenchymal transitions: The importance of changing cell state in development and disease. *Journal of Clinical Investigation* (2009) doi:10.1172/JCI38019.
47. Johnson, D. G. & Walker, C. L. Cyclins and cell cycle checkpoints. *Annual Review of Pharmacology and Toxicology* (1999) doi:10.1146/annurev.pharmtox.39.1.295.
48. Risson, E., Nobre, A. R., Maguer-Satta, V. & Aguirre-Ghiso, J. A. The current paradigm and challenges ahead for the dormancy of disseminated tumor cells. *Nat. Cancer* (2020) doi:10.1038/s43018-020-0088-5.
49. Fischer, K. R. *et al.* Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance. *Nature* (2015) doi:10.1038/nature15748.
50. Vega, S. *et al.* Snail blocks the cell cycle and confers resistance to cell death. *Genes Dev.* (2004) doi:10.1101/gad.294104.

51. Mejlvang, J. *et al.* Direct repression of cyclin D1 by SIP1 attenuates cell cycle progression in cells undergoing an epithelial mesenchymal transition. *Mol. Biol. Cell* (2007) doi:10.1091/mbc.E07-05-0406.
52. Stylianou, N. *et al.* A molecular portrait of epithelial–mesenchymal plasticity in prostate cancer associated with clinical outcome. *Oncogene* (2019) doi:10.1038/s41388-018-0488-5.
53. Lawson, D. A. *et al.* Single-cell analysis reveals a stem-cell program in human metastatic breast cancer cells. *Nature* (2015) doi:10.1038/nature15260.
54. Adam, A. P. *et al.* Computational identification of a p38 SAPK-regulated transcription factor network required for tumor cell quiescence. *Cancer Res.* **69**, 5664–5672 (2009).
55. Bragado, P. *et al.* TGF-β2 dictates disseminated tumour cell fate in target organs through TGF-β-RIII and p38α/β signalling. *Nat. Cell Biol.* **15**, 1351–61 (2013).
56. Laughney, A. M. *et al.* Regenerative lineages and immune-mediated pruning in lung cancer metastasis. *Nat. Med.* (2020) doi:10.1038/s41591-019-0750-6.
57. Borgen, E. *et al.* NR2F1 stratifies dormant disseminated tumor cells in breast cancer patients. *Breast Cancer Res.* (2018) doi:10.1186/s13058-018-1049-0.
58. Huang, X. *et al.* Zfp281 is essential for mouse epiblast maturation through transcriptional and epigenetic control of Nodal signaling. *Elife* (2017) doi:10.7554/eLife.33333.
59. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* (2013) doi:10.1093/bioinformatics/bts635.
60. Liao, Y., Smyth, G. K. & Shi, W. FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* (2014) doi:10.1093/bioinformatics/btt656.
61. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* (2014) doi:10.1186/s13059-014-0550-8.
62. Tan, G. & Lenhard, B. TFBSTools: An R/bioconductor package for transcription factor binding site analysis. *Bioinformatics* (2016) doi:10.1093/bioinformatics/btw024.
63. Durinck, S. *et al.* BioMart and Bioconductor: A powerful link between biological databases and microarray data analysis. *Bioinformatics* (2005) doi:10.1093/bioinformatics/bti525.
64. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* (2010) doi:10.1186/gb-2010-11-3-r25.
65. McCarthy, D. J., Chen, Y. & Smyth, G. K. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res.* (2012) doi:10.1093/nar/gks042.
66. Robinson, J. T., Thorvaldsdóttir, H., Wenger, A. M., Zehir, A. & Mesirov, J. P. Variant review

with the integrative genomics viewer. *Cancer Research* (2017) doi:10.1158/0008-5472.CAN-17-0337.

67. Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Brief. Bioinform.* (2013) doi:10.1093/bib/bbs017.

68. Fidalgo, M. *et al.* Zfp281 mediates Nanog autorepression through recruitment of the NuRD complex and inhibits somatic cell reprogramming. *Proc. Natl. Acad. Sci. U. S. A.* (2012) doi:10.1073/pnas.1208533109.

Figures

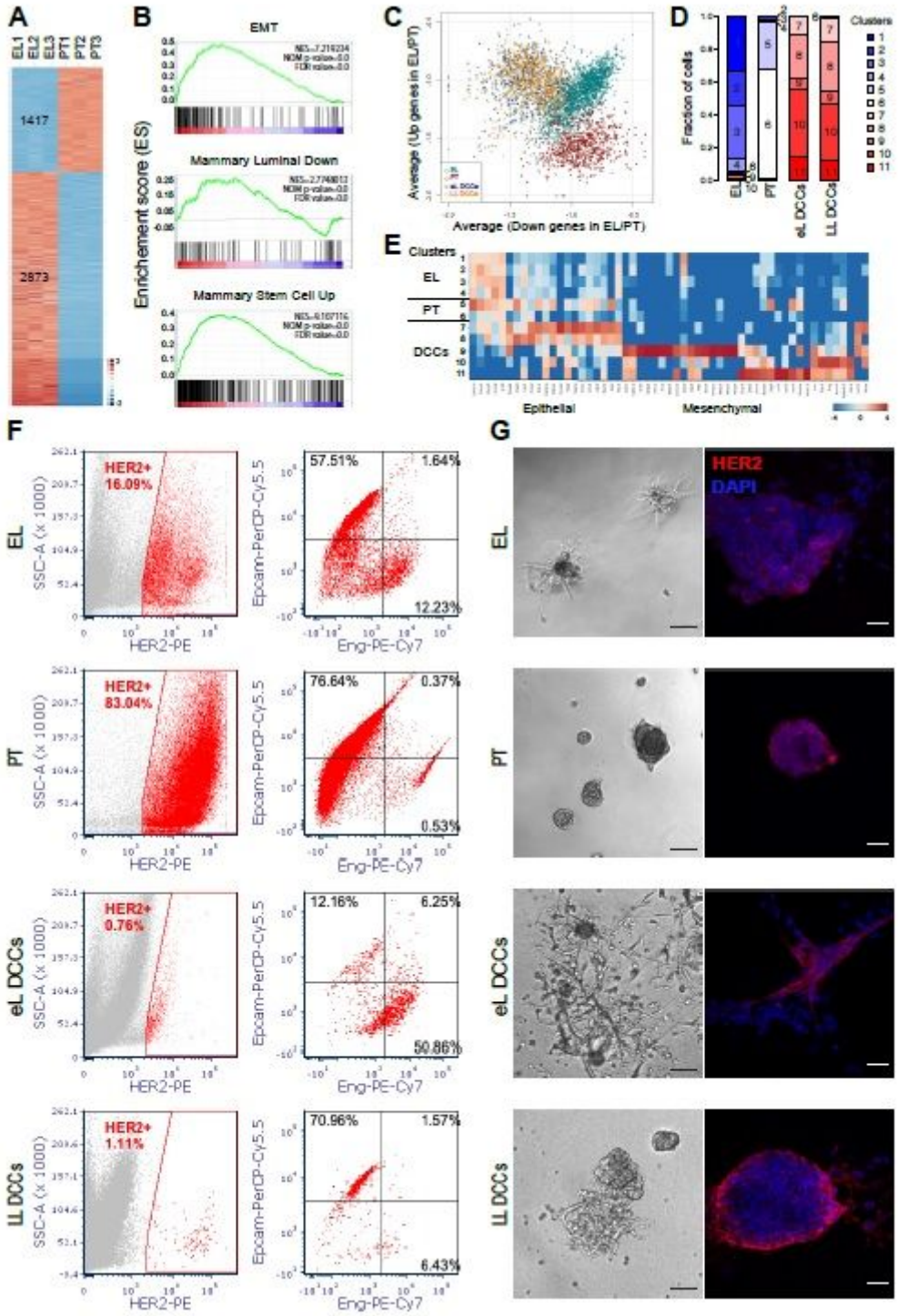

Figure 1

Early DCCs maintain a global Mesenchymal-like phenotype. (A) Heatmap of differentially expressed genes (DEGs, STable 1) detected by RNAseq from MMTVHER2 early lesion (EL) and primary tumor (PT) spheres cultured for 7 days. 2873 upregulated genes (red); 1417 downregulated genes (blue); p-value <0.05 and FC>2 or <0.5. (B) Gene Set Enrichment analysis31,32 of EMT (top hallmark hit) and mammary

gland luminal, myoepithelial/basal and stem cell signatures33 in MMTV-HER2 EL vs. PT 7-day spheres bulk RNAseq (GSEA STable 3). ES, enrichment score; NES, normalized enrichment score; NOM p-value, nominal p-value <0.05; FDR value, false discovery rate <0.25. (C) Distribution of the signatures 'Up and Down in EL/PT spheres' (Figure 1A and STable 1) in MMTV-HER2 EL (teal), PT (red), eL (early lungs, blue) and LL (late lungs, orange) DCCs single-cell RNAseq. (D) MMTV-HER2 EL, PT, eL (early lungs) and LL (late lungs) DCCs single-cell RNAseq sample distribution per cluster. Unsupervised clustering on the DEGs was performed using a previously described batch-aware algorithm34. (E) Heatmap of UMI counts of selected epithelial (Ep) and mesenchymal (M) genes (STable 4) in scRNAseq after unsupervised clustering on the DEGs using a previously described batch-aware algorithm34. Cell clusters were sub-grouped as EL (1-4), PT (5-6) and DCCs (7-11), according with the predominant cell type in each cluster. (F) CD45-HER2+ cells used for scRNAseq of MMTV-HER2 EL, PT and eL (early lungs) and LL (late lungs) DCCs after tissue dissociation were quantified for expression of EpCAM (epithelial marker)and Eng (mesenchymal marker). (G) Imaging of sorted CD45-HER2+ cells EL, PT, eL and LL DCCs upon 7-day culture in 3D, on top of matrigel. Left: brightfield images, scale 50 um. Right: HER2 (red) expression, scale 25 um.See also Figure S1.

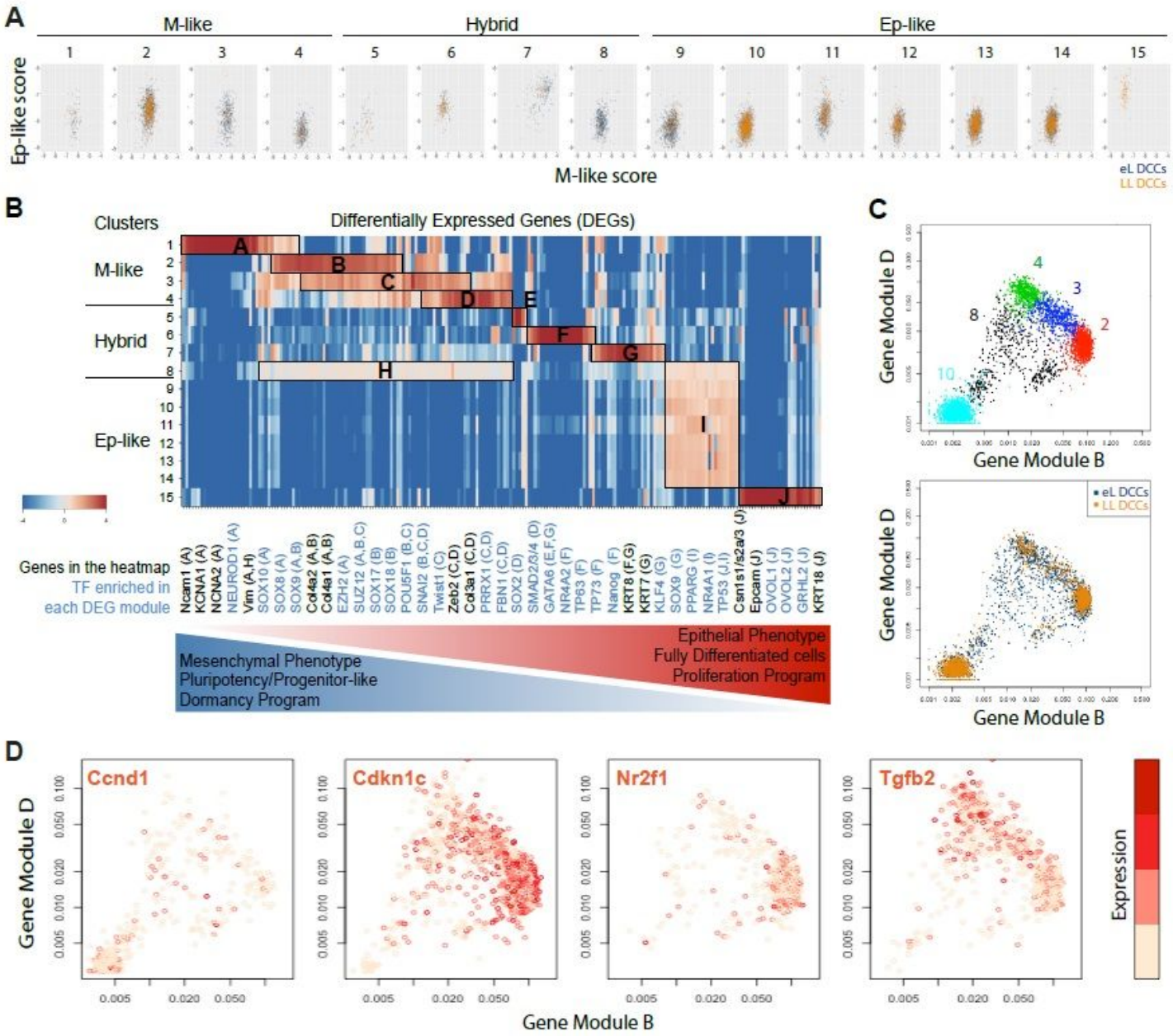

Figure 2

Early DCCs turn on Mesenchymal- and Pluripotency/Progenitor-like programs that allow them to undergo Dormancy. (A) Distribution of Epithelial (Ep) and Mesenchymal (M) scores (gene lists in STable 4) in MMTVHER2 lung DCC clusters after unsupervised clustering on the DEGs using a previously described batch-aware algorithm34. Cell clusters were sub-grouped as M-like (1-4, higher M-like score), Hybrid 17 (5-8) and Ep-like (9-15). Dots color-coded by sample origin: early lung DCCs, eL DCCs, blue; or late lung DCCs, LL DCCs, orange. (B) Heatmap of UMI counts of selected genes (gene lists in STable 4) in MMTV-HER2 eL (early lungs,) and LL (late lungs) DCCs single-cell RNAseq after unsupervised clustering on the DEGs using a previously described batch-aware algorithm34. Clusters 1-15 differentially express gene modules identified in boxed letters. Transcription factors (TF, bottom genes in blue) enriched in each gene module were predicted using Enrichr29,30. Lung DCCs show mesenchymal, pluripotency/progenitorlike and dormancy programs that are downregulated once these cells undergo

mesenchymal to epithelial transition (MET), fully differentiate and proliferate, potentially giving rise to metastasis (bottom diagram). (C) Distribution of gene modules B and D in clusters 2, 3, 4, 8 and 9. Dots represent single cells colorcoded by cluster (top) and sample origin (eL or LL, bottom). (D) Distribution of gene modules B and D in clusters 2, 3, 4, 8 and 9. Dots represent single cells colorcoded by expression of Ccnd1, Cdkn1c, Nr2f1 and Tgfb2. See also Figure S2.

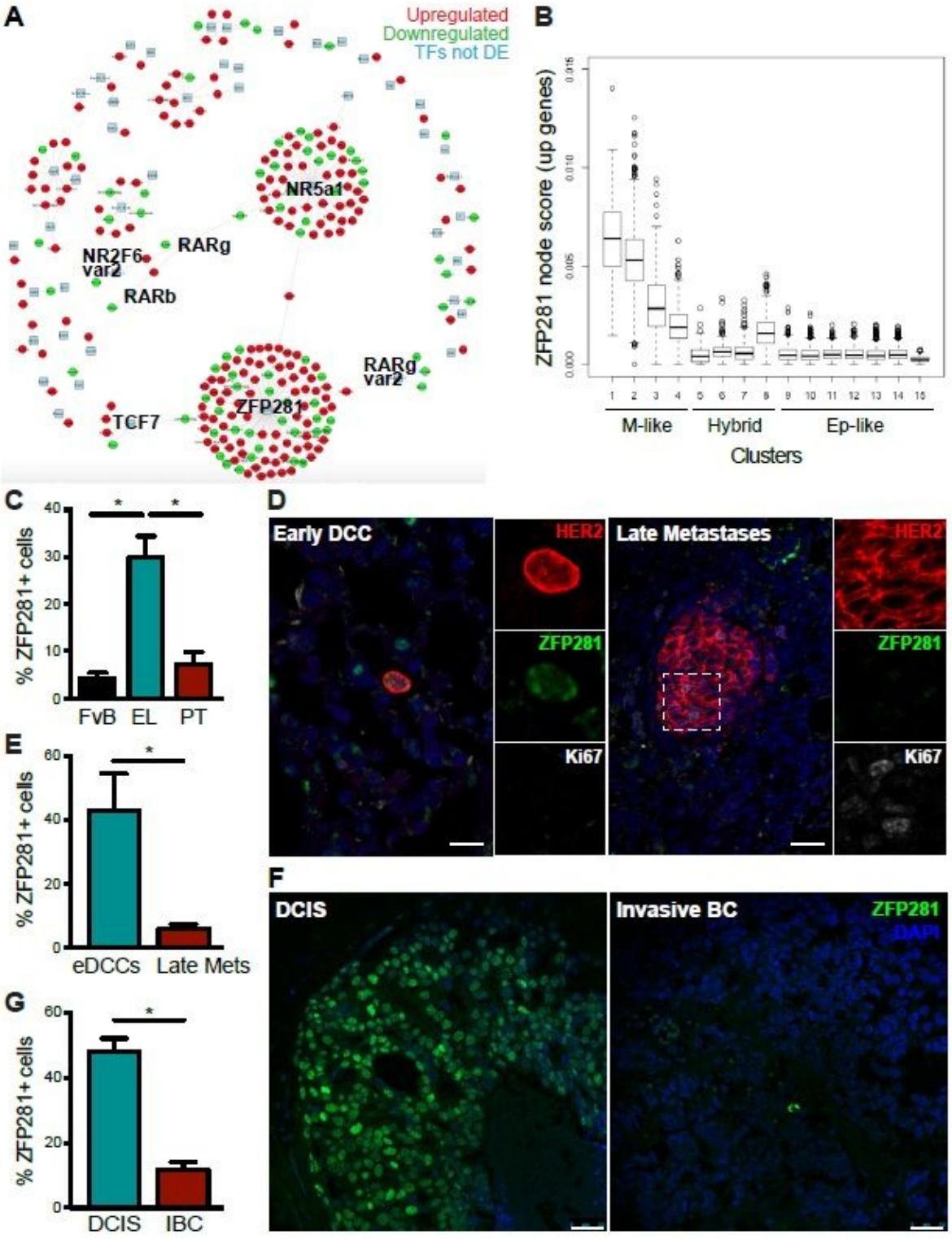

Figure 3

Identification of ZFP281 in early lesions (EL) and early DCCs. (A) Transcription factor (TF) network analysis derived from the RNAseq DEGs from MMTV-HER2 EL and PT spheres. Red, upregulated genes; Green, downregulated genes; Blue, TFs not differentially expressed (DE). Lines connect the TF at the center of the node to target genes indicating that the connected genes have predicted TF binding elements (validated or predicted) in the promoter regions (-500 to +2500 bp from the TSS). Full gene list in STable 6. (B) Distribution of ZFP281 predicted target scores, summarizing the UMI fraction of ZFP281 predicted targets (ZFP281 node in Figure 3A and STable 6), in all DCC clusters analyzed by scRNAseq (Figure 2). (C) ZFP281 expression in FvB mammary gland and MMTV- HER2 EL and PT tissues. Representative pictures in Figure S3B. Graph shows n=4, mean, SEM and 2-tailed Mann-Whitney test. (D and E) ZFP281 (green) and Ki67 (gray) protein expression in MMTV- HER2 (HER2, red) lung DCCs. Scales, 20 um. Graph shows n=4, mean, SEM and 2-tailed Mann-Whitney test. (F and G) ZFP281 (green) protein detection in human ductal carcinoma in situ (DCIS) and invasive breast cancer (IBC) samples. Scales, 25 um. Graph shows n=14, mean, SEM and 2-tailed Mann-Whitney test.See also Figure S3.

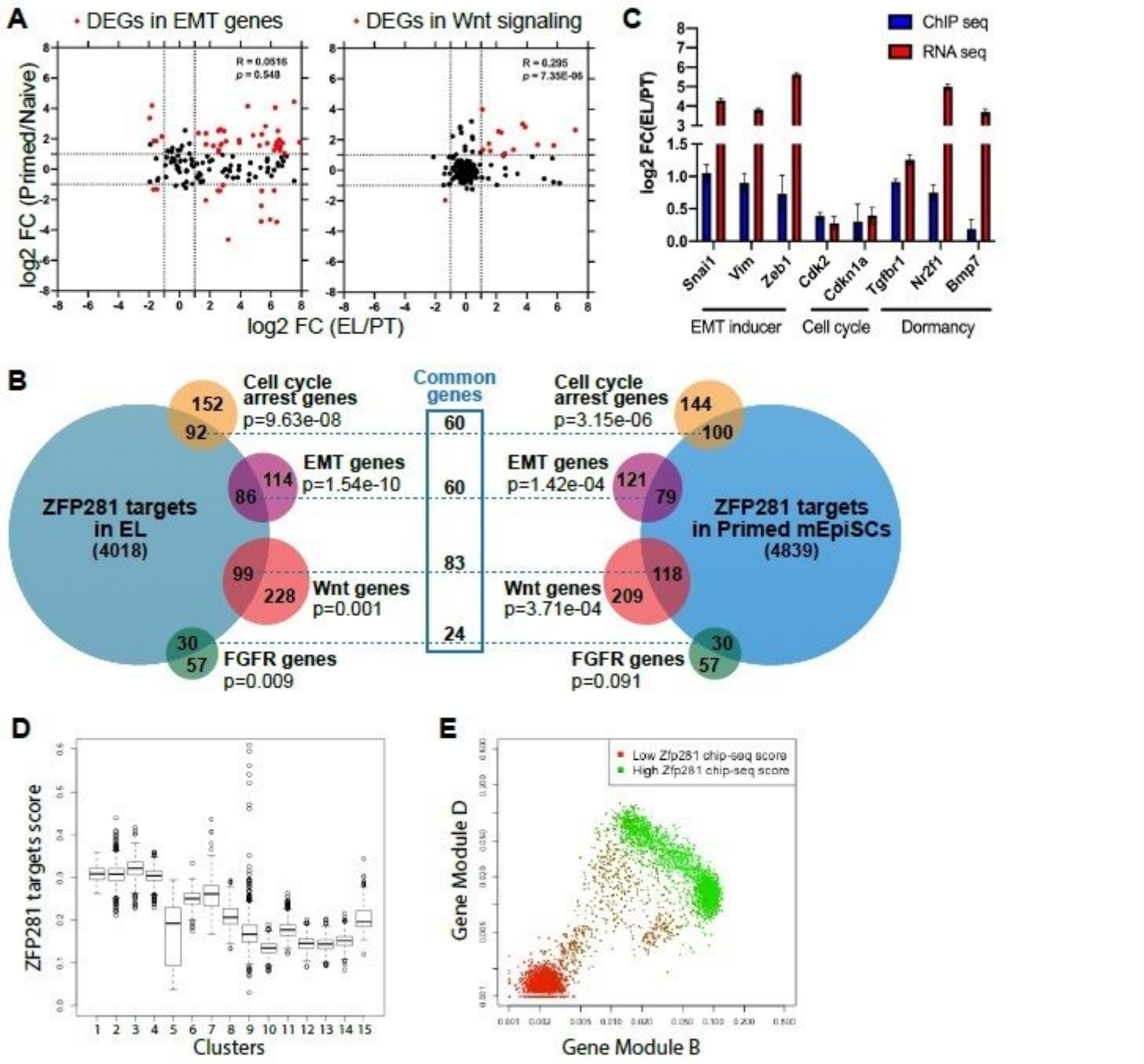

Figure 4

ZFP281 regulates an EMT-like program both in primed mEpiSCs and early DCCs. (A) Correlation of EMT and Wnt signaling DEGs in primed versus naïve mouse pluripotent stem cells (RNASeq data, GSE81044 from 36) and EL vs. PT cells (bulk RNAseq, Figure 1A). Gene lists in STable 7 and 8. (B) Venn diagrams for ZFP281 targets identified by ChIPseq from EL cells and primed mEpiSCs ChIPSeq data, GSE93042 from 58) and cell cycle arrest, EMT, Wnt and FGF signaling genes. Comparisons and statistics were done in pairs (ZFP281 targets in EL vs. Cell cycle arrest genes;ZFP281 targets in EL vs. EMT genes; ZFP281 targets in EL vs. Wnt genes; ZFP281 targets in EL vs.FGFR genes; and same for ZFP281 targets primed mEpiSCs). Overlapped genes between these paired comparisons were rare so for graphical simplification, the 4 comparisons are displayed together. Common genes (central blue box) correspond to genes from

each category that are targeted by ZFP281 both in EL and mEpiSCs. Gene lists in STable 9. (C) ChIPseq and RNAseq data in EL over PT cells. ZFP281 binds EMT (Snai1, Vim, Zeb1), cell-cycle (Cdk2 and Cdkn1a), and dormancy (Tgfbr1, Nr2f1 and Bmp7) associated genes. All genes were upregulated in EL cell; Snai1, Vim, Zeb1, Cdk2, Tgfbr1, and Nr2f1 were ZFP281 bound genes exclusively in EL cells; Cdkn1a, and Bmp7 were bound by ZFP281 in both EL and PT cells but only upregulated in EL cells. (D) Distribution of ZFP281 target (ChIP data) scores, summarizing the averaged expression of ZFP281 targets, in all DCC clusters analyzed by scRNAseq (Figure 2). (E) Distribution of gene modules B and D in all DCC clusters. Dots represent single cells color-coded by ZFP281 target (ChIP data) scores (low, red, to high, green).See also Figure S4.

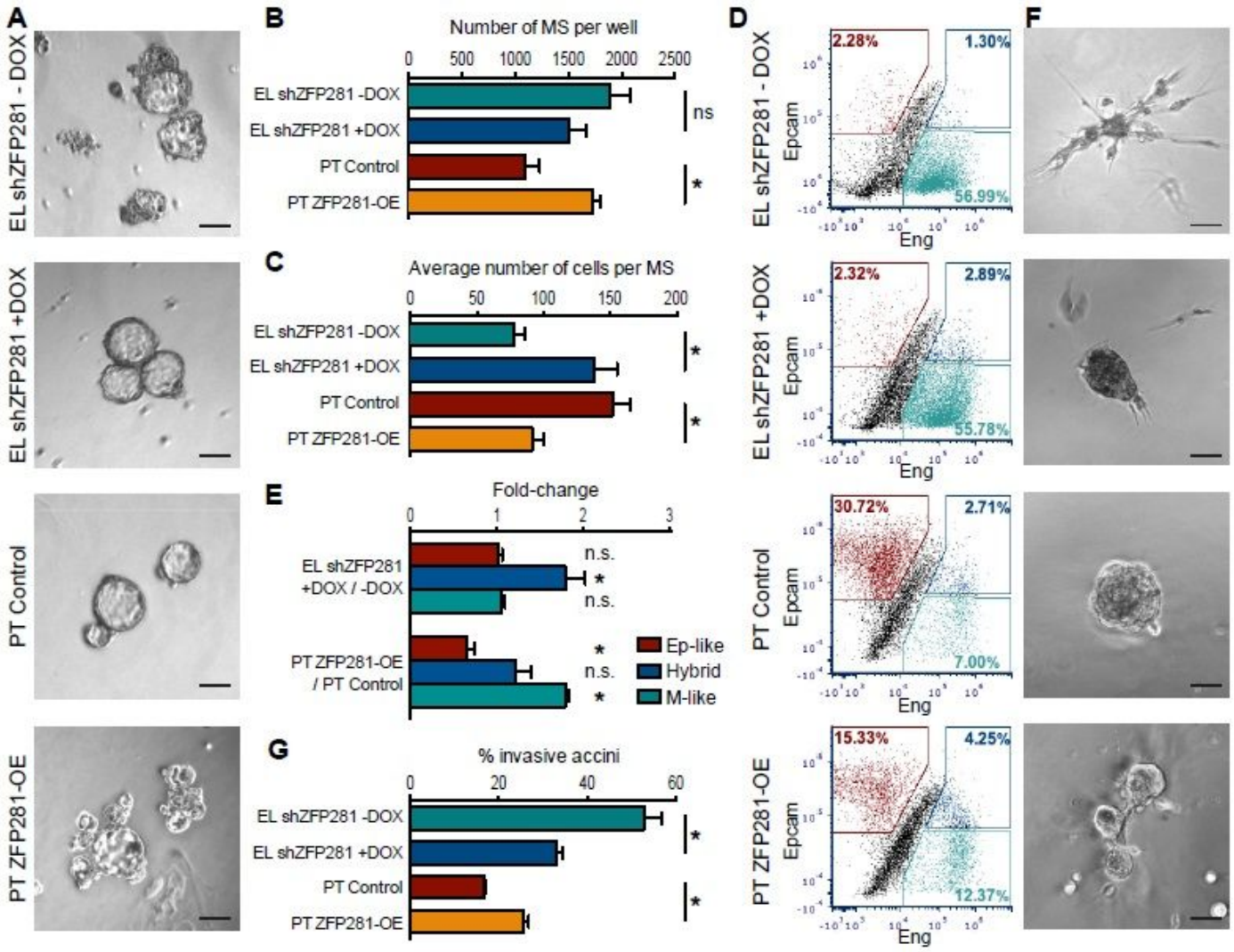

Figure 5

ZFP281 induces an M-like slow-cycling phenotype in vitro. (A) Column of representative images of the mammosphere phenotype of EL shZFP281±DOX, PT Control and PT ZFP281-overexpressed (OE) cells. Scale 50 um. (B) Quantification of mammosphere (MS) frequency of EL shZFP281±DOX, PT Control and PT ZFP281-OE cells. Graph shows n=3 mean, SEM and 2-tailed Mann-Whitney test. (C) Quantification of mammosphere (MS) size, as number of cells per sphere after dissociation of EL shZFP281±DOX, PT

Control and PT ZFP281-OE spheres. Graph shows n=3, mean, SEM and 2-tailed Mann-Whitney test. (D) EpCAM (epithelial marker) and Eng/CD105 (mesenchymal marker) expression in EL shZFP281±DOX, PT Control and PT ZFP281-OE cells. Representative experiment of n=3 biological replicates. (E) Fold-change of Ep-like (EpCAM+Eng-), hybrid (EpCAM+Eng+) and M-like (EpCAM-Eng+) populations in EL shZFP281+DOX over –DOX and PT ZFP281-OE over PT Control spheres. Graph shows n=3, mean, SEM and 2-tailed Mann-Whitney test. (F) Column of representative images of 3D-matrigel organoids and invasive phenotype of EL shZFP281±DOX, PT Control and PT ZFP281-OE cells. Scale 50 um. (G) Quantification of percentage of 3D-matrigel spheroids with invasive protusions per condition. Graph shows n=4, mean, SEM and 2-tailed Mann-Whitney test. See also Figure S5.

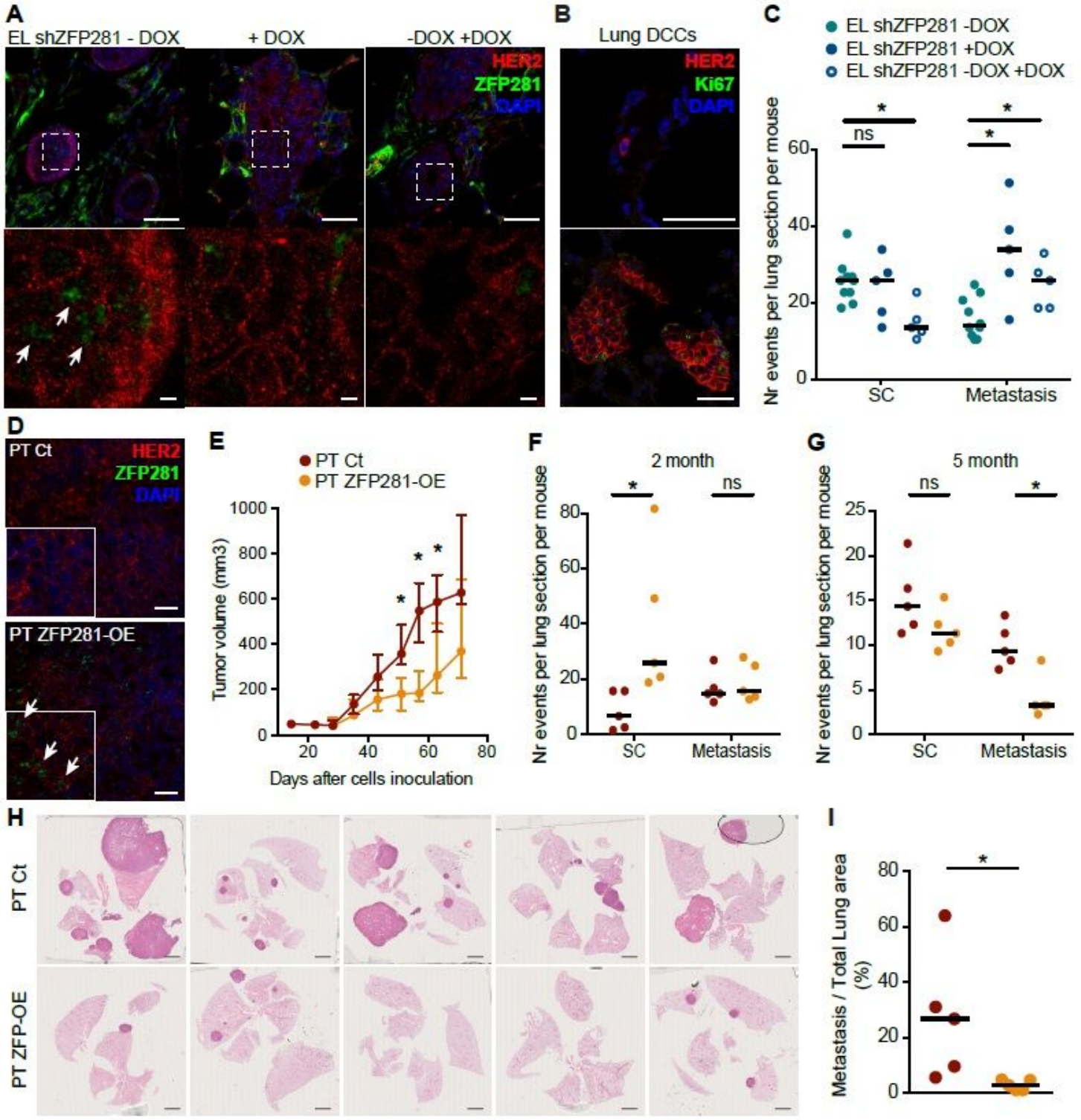

Figure 6

ZFP281 controls the M-like program that leads to a switch from dormant early DCCs to proliferation in vivo. (A) Representative images of HER2 (red) and ZFP281 (green) protein expression in mammary fat pads of mice 5 month after EL shZFP281 sphere injections. Mice were given water 1) without doxycycline (DOX) for 5 months: '-DOX'; 2) with DOX for 5 months: '+DOX' or 3) 1 month without and 4 months with DOX: '-DOX +DOX'. Arrows point to ZFP281 expression in EL shZFP281-DOX, which is

downregulated in groups '+DOX' and '-DOX +DOX'. Scales 25um (top row) and 50um (bottom row, inserts). (B) Representative images of lung DCCs, single cells and metastasis quantified in C. HER2, red; ZFP281, green; DAPI, blue. Scales 25um. (C) Frequency of lung single cell (SC) and metastasis per lung section per mouse for all conditions, 5 months after EL shZFP281 sphere injections. 2 lung slides with all lobules represented were scanned and quantified per mouse. Graph shows n=5-10 mice/condition, median and 2-tailed Mann-Whitney test. (D) Representative images of HER2 (red) and ZFP281 (green) protein expression in primary tumors 71 days post PT Control and PT ZFP281-OE cell mammary fat pad injection. Scales 25um. (E) Tumor volume over time of PT Control and PT ZFP281-OE. Tumors were removed at day 71 and mice sacrificed 5 months after cancer cells injections (corresponding lungs in G). Graph shows n=10 per condition, median, interquartile range and 2-tailed Mann-Whitney test. (F and G) Frequency of lung single cell (SC) and metastasis per mice per condition, 2 (F) and 5 (G) months after mammary fat pad injection of PT Control or PT ZFP281-OE spheres. 2 lung slides with all lobules represented were scanned and quantified per mouse. Graph shows n=5 mice/condition, median and 2-tailed Mann-Whitney test. (H) H&E images of mice lungs 5 months after mammary fat pad injection of PT Control or PT ZFP281-OE spheres. Scales 2mm. (I) Quantification of lung metastasis burden, normalized for total lung area, of images in B (1 slide per mouse). Graph shows n=5 mice/condition, median and 2-tailed Mann-Whitney test. See also Figure S6.

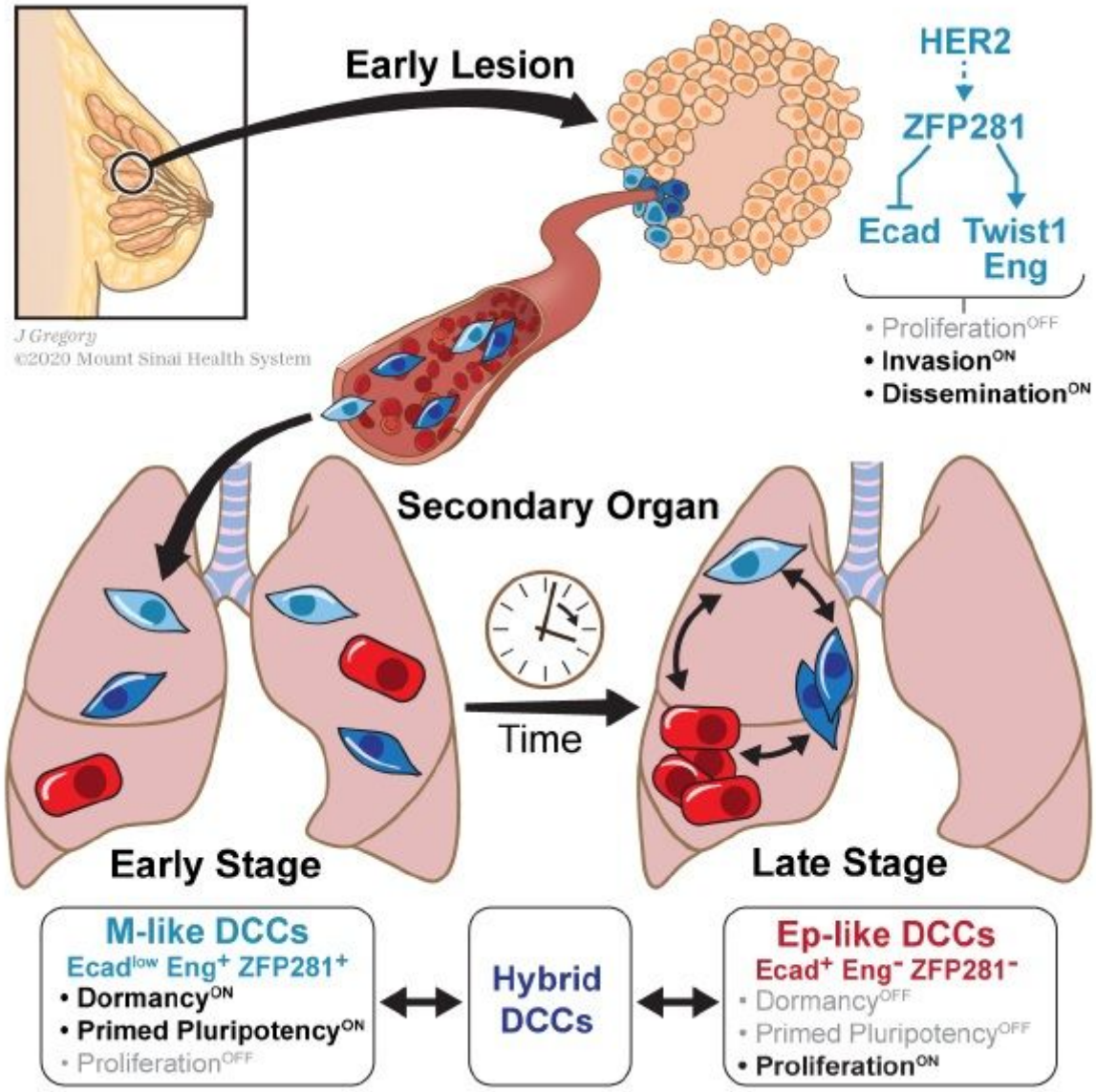

Figure 7

Model of ZFP281 regulated dissemination and dormancy states during early cancer evolution. Early upon cancer initiation via HER2 signaling, EL cells activate the primed pluripotency transcription factor ZFP281 (upper section). This transcription factor regulates at least four distinct mesenchymal- and pluripotency-like programs, leading to EL cells to disseminate and to enter a prolonged dormancy as early DCCs in lungs (lower left, Early Stage). The M-like dormancy and primed pluripotency program is associated with dormancy programs, such as those controlled by NR2F1 and TGFβ2, and block the acquisition of an epithelial state. Over time, intrinsic and microenvironmental changes allow the dormant DCCs to disrupt ZFP281 function and adopt an Eplike phenotype (lower right, Late Stage), which enables a proliferative state. Importantly, M-like, hybrid and Ep-like DCCs co-exist in both early and late stage lungs, with predominance of M-like dormant DCCs in early stage lungs.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Nobreetal2021STables.xlsx
- SupplementaryFigures.pdf