Published in final edited form as:

Cancer Res. 2020 June 01; 80(11): 2125–2137. doi:10.1158/0008-5472.CAN-19-3018.

Peptidylarginine deiminase IV (PADI4) regulates breast cancer stem cells via a novel tumor cell-autonomous suppressor role

Nellie Moshkovich[1], **Humberto J. Ochoa**[1], **Binwu Tang**[1], **Howard H. Yang**[1], **Yuan Yang**[1], **Jing Huang**[1], **Maxwell P. Lee**[1], **Lalage M. Wakefield**[1]

[1]Laboratory of Cancer Biology and Genetics, National Cancer Institute, Bethesda MD 20892, USA

Abstract

Peptidylarginine deiminases (PADIs) catalyze post-translational modification of many target proteins and have been suggested to play a role in carcinogenesis. Citrullination of histones by PADI4 was recently implicated in regulating embryonic stem and hematopoietic progenitor cells. Here we investigated a possible role for PADI4 in regulating breast cancer stem cells. PADI4 activity limited the number of cancer stem cells (CSC) in multiple breast cancer models in vitro and in vivo. Mechanistically, PADI4 inhibition resulted in a widespread redistribution of histone H3 with increased accumulation around transcriptional start sites. Interestingly, epigenetic effects of PADI4 on the bulk tumor cell population did not explain the CSC phenotype. However, in sorted tumor cell populations, PADI4 downregulated expression of master transcription factors of stemness, NANOG and OCT4, specifically in the cancer stem cell compartment, by reducing the transcriptionally activating H3R17me2a histone mark at those loci; this effect was not seen in the non-stem cells. A gene signature reflecting tumor cell-autonomous PADI4 inhibition was associated with poor outcome in human breast cancer datasets, consistent with a tumor suppressive role for PADI4 in estrogen receptor-positive tumors. These results contrast with known tumor-promoting effects of PADI4 on the tumor stroma and suggest that the balance between opposing tumor cell-autonomous and stromal effects may determine net outcome. Our findings reveal a novel role for PADI4 as a tumor suppressor in regulating breast cancer stem cells and provide insight into context-specific effects of PADI4 in epigenetic modulation.

Keywords

PADI4; Cancer stem cells; Breast cancer; Citrullination; Epigenetics

INTRODUCTION

Breast cancer is a global problem accounting for almost a quarter of all cancers in women and despite therapeutic advances, 20% to 30% of treated patients still suffer from cancer relapse and metastasis (1). The cancer stem cell hypothesis has generated new insights into

Corresponding author: Lalage M. Wakefield D Phil, 37 Convent Drive, MSC 4255, National Cancer Institute, Building 37, Room 4032A, Bethesda MD 20892-4255, Tel: 240-760-6808, lw34g@nih.gov.

Conflict of interest statement: The authors declare no conflicts of interest.

this problem. Many tumors consist of a phenotypic hierarchy of tumor cells, with a small subpopulation of cancer stem cells (CSCs) at the apex, giving rise to proliferative but progressively differentiating non-tumorigenic cells (2,3). These rare CSCs are uniquely capable of the self-renewal necessary for tumor initiation and sustainability. CSCs have been identified in many tumors including breast cancer (4), and are thought to be the key to cancer recurrence, due to their intrinsically elevated resistance to a wide range of therapeutic approaches (3). Thus a better understanding of the mechanisms regulating CSC dynamics may lead to the design of more effective therapeutics.

The hierarchical organization of tissues and tumors is maintained by differences in the epigenome (5,6). Central to this process is the dynamic regulation of histones and other proteins by post-translational modification, of which methylation and acetylation are the most studied (7,8). However, other post-translation modifications are also emerging as important. Peptidylarginine deiminases (PADIs) are a family of Ca^{2+}-dependent enzymes that catalyze citrullination of arginine residues to modulate the activity of target proteins, with the citrullinome now comprising several hundred proteins (reviewed in (9)). Citrullination results in the loss of a positive charge, leading to altered biological activity and/or protein interactions. Among the five member PADI family, only PADI4 has a classic nuclear localization signal (10), suggesting a particularly important role in modulating activity of nuclear targets. Within the nucleus PADI4 targets histones (H1, H2A, H3 and H4) and modulates activity of some transcription factors (eg. p53) through effects on nuclear proteins (eg. ING4), as well as affecting targets outside the nucleus, including extracellular matrix (eg. Col1), and intracellular proteins (eg. vimentin) (reviewed in (9)). In doing so, it regulates many biological processes, including apoptosis, neutrophil extracellular trap (NET) formation and pluripotency (11–13). NET formation, a host defense mechanism, is the most dramatic manifestation of the effect of PADI4 citrullination. In this process, PADI4 hypercitrullinates histones in neutrophils, facilitating global chromatin decondensation and extrusion of a DNA web (the NET) to trap extracellular pathogens (14).

Post-translational modification of histones is key to maintenance of the pluripotent stem cell state in embryonic stem cells, and of the stem phenotype in normal tissue and disease (8). Several recent studies have implicated citrullination of chromatin components by PADI4 in the regulation of stemness and differentiation. Indeed, PADI4 was first identified in human myeloid leukemia cells, where it was induced by differentiating agents such as retinoic acid (15). PADI4 suppressed proliferation of multipotent hematopoietic stem cells, and promoted their differentiation through effects on histone H3 and altered transcriptional regulation (16). Conversely in the early mouse embryo, citrullination of histone H1 by PADI4 displaced it from chromatin resulting in global chromatin decondensation, thereby creating open chromatin necessary for reprogramming of pluripotency/stemness (13).

PADIs have been most extensively studied in the context of inflammation and autoimmunity (17). However, important roles for PADIs in tumorigenesis are emerging. Both PADI2 and PADI4 have been shown to have context-dependent roles in many different types of cancers (9). PADI2 promotes cell proliferation and invasiveness in gastric cancer, while suppressing cell growth and metastasis in liver cancer (18) and tumor cell invasion in breast cancer (19). In MCF7 breast cancer cells stimulated with estrogen, PADI4 transcriptionally represses

estrogen receptor (ER) target genes via H4 citrullination (20), while activating the ELK1 oncogene under EGF simulation (21). Additionally, dysregulation of PADI4 in breast cancer cells activated TGF-β signaling and induced an epithelial-to-mesenchymal transition (EMT), increasing the invasive phenotype of xenografted tumors (22).

The recent connections between PADI4 and stemness in other systems led us to hypothesize that PADI4 may play a role in breast cancer progression by regulating the CSC phenotype. Here, we use pharmacologic and genetic approaches in breast cancer models to generate new mechanistic insights into the role of PADI4 in breast cancer progression.

MATERIALS AND METHODS

Breast cancer cell lines and biological activity assays

The MCF10CA1h and MCF10Ca1a cell lines were obtained from Dr. Fred Miller, Karmanos Cancer Institute, Detroit, MI. MCF7E (early passage MCF7 cells) were a gift from Dr. Michael Brattain, Eppley Institute, Nebraska. MDA-MB-231 LM2 cells were a gift from Dr. Joan Massague, Memorial Sloan Kettering Cancer Center, New York, NY. Additional breast cancer cell lines were all obtained from the ATCC, Manassas, VA. Cell lines not obtained directly from ATCC were authenticated by short terminal repeat (STR) analysis using 15 polymorphic markers and one gender marker (Laragen Inc, Culver City CA). All cell lines were routinely tested using Mycoplasma PCR Detection Kit (Sigma, MP0035) and shown to be mycoplasma-free. Cells were used for experiments within 2–12 passages after recovery from frozen stocks. Growth conditions are given in Supplementary Table S1. Proliferation and apoptosis were assessed using Click-iT ® imaging assays (Invitrogen). Invasion and migration were assessed by Transwell assays with or without Matrigel. Tumorsphere-forming activity was assessed in semisolid MethoCult medium (Stem Cell Technologies).

Modulation of PADI4 activity

Genetic.—Cells were stably transduced with lentiviruses expressing PADI4 shRNAs for knockdown (Supplementary Table S2), or various PADI4 isoforms for overexpression, and were used after recovery from selection.

Pharmacologic.—Cells were treated with 1μM PADI4 inhibitor GSK484 (Cayman Chemical) or its inactive analog GSK106 for 3–10 days prior to assay, as indicated.

In vivo extreme limiting dilution assay (ELDA)

All animal experiments were conducted under protocol LC-070 approved by the Animal Care and Use Committee of the National Cancer Institute. To quantify the frequency of CSCs *in vivo*, serial dilutions of tumor cells (1,000 – 250,000 cells) were orthotopically implanted into immunodeficient mice. After 60 days, tumor incidence data was analyzed using ELDA software (http://bioinf.wehi.edu.au/software/elda)

Enrichment of cancer stem cells

Cells were transduced with a modified version of the SORE6 fluorescent cancer stem cell reporter (23), followed by fluorescence-activated cell sorting (FACS) for SORE6+ cells (CSC enriched) and SORE6− cells (CSC depleted).

Whole genome and transcriptome analysis

To assess possible effects of PADI4 inhibition on the epigenome/transcriptome, breast cancer cells were treated with the GSK484 PADI4 inhibitor or inactive analog GSK106 for 3 days and then harvested for ChIP-Seq for multiple Histone H3 marks (see below), and for RNA-Seq. Since poised promoters play a particularly important role in stem cell biology (24), H3K4me3 and H3K27me3 marks were used to identify loci containing bivalently marked peaks under the different treatment conditions, and the data were integrated with transcriptomic data to address whether PADI4 inhibition led to changes in expression of relevant stem/differentiation factors.

ChIP-seq and ChIP-qPCR

All chromatin immunoprecipitations (ChIPs) were done with SimpleChIP® Enzymatic Chromatin IP Kit (Cell Signaling Technology). Briefly, cells (4×10^6 cells/IP) treated with GSK484 or control GSK106 for 72 hours were fixed with formaldehyde and lysed, and chromatin was fragmented by partial digestion with Micrococcal Nuclease. For immunoprecipitation (IP), 5–10 μg of digested, cross-linked chromatin per IP was incubated with antibodies for total Histone H3, H3K4me3, H3K27me3, H3R2R8R17citr, H3R17me2, H3R2me2 or IgG. ChIPed DNA was purified and 5 ng was processed for NG-sequencing DNA library construction for all marks except H3R17me2 and H3R2me2, which were just analyzed by ChIP-qPCR. Sequencing libraries were run on Illumina NextSeq platform SR 75bp cycle runs of 50M reads/sample. The enrichment of specific DNA sequences during IP was validated by ChIP-qPCR. In some experiments, cells transduced with the SORE6 reporter were FACS-sorted into SORE6+ (CSC) and SORE6− (non-CSC) compartments prior to ChIP-qPCR at targeted loci. Primers and antibodies used for ChIP are listed in Supplementary Tables S3 and S4 respectively. The fastq sequence reads were mapped to the human hg19 reference genome and peaks were identified using MACS2. Heatmaps were generated using deepTools2.

RNA-seq and Microarray analysis

Total RNA was extracted using Trizol Reagent (Ambion) and all RINs were 9 or above. For RNA-seq, sequencing libraries were generated according to Illumina's TruSeq® Stranded mRNA Sample Preparation protocol and run on Illumina HiSeq platform PE 125bp cycle runs of 200M reads/sample. For microarray analysis, RNA samples were labeled and hybridized to the Clariom S Human Array (Affymetrix).

Statistical analysis

Experiments were analyzed for statistical significance using GraphPad Prism, version 8 (GraphPad Software Inc). Where indicated, the t-test (2 samples) or Dunnett's multiple

comparison test (>2 samples) were used. A p-value of 0.05 was considered statistically significant.

Supplementary Methods

More detailed methods are available in the Supplementary Data and Methods.

Data availability

ChIPSeq, RNASeq and microarray datasets relating to this study are available from GEO as SuperSeries GSE138506.

RESULTS

PADI4 is expressed in human breast cancer cell lines

To address the prevalence of PADI4 in breast cancer, expression was assessed in a panel of breast cancer cell lines. RT-qPCR showed that PADI4 mRNA trended to slightly higher expression in luminal subtype cell lines when compared to basal subtype cell lines, with the notable exceptions of MDA-MB-231 LM2, a lung-tropic subclone of MDA-MB-231 (25), and MDA-MB-436 (Fig. 1A). A similar non-significant trend was seen in a larger publically-available dataset of 50 breast cancer cell lines (Supplementary Fig. S1A), while in human breast tumors in the TCGA cohort there was no significant difference in PADI4 expression between basal and luminal subtypes (Supplementary Fig. S1B). In three lines examined in more detail, PADI4 mRNA expression was equal to or higher than the other three PADI family members (Fig. 1B). PADI4 protein was not readily detectable by Western blot analysis of total cell lysates, but analysis of nuclear extracts revealed multiple PADI4 protein bands in the range of 65–43 kDa (Fig. 1C–D). PADI4 full-length transcript is predicted to translate to a protein of 73 kDa. However, several bands have previously been detected, depending on the tissue of origin (https://www.proteinatlas.org/ENSG00000159339-PADI4/antibody#western_blot), and RefSeq predicts eight alternatively spliced PADI4 transcripts encoding seven protein isoforms (Supplementary Fig. S2A,B) of unknown functional significance. The protein bands marked by arrows in Fig. 1C–D were confirmed to be PADI4 by shRNA knockdown in three representative cell lines (see next section). Consistent with the RNA data, luminal breast cancer cell lines generally expressed slightly higher levels of PADI4 protein than basal lines, though only MDA-MB-231 LM2 cells expressed high levels of full-length PADI4 (Fig. 1C–D). We selected luminal MCF10Ca1h, MCF7E and basal MDA-MB-231 LM2 models for further study, as these breast cancer cell lines had the highest PADI4 protein levels and captured the range of different protein isoforms expressed.

PADI4 knockdown increases pro-progression responses and stemness *in vitro*

PADI4 function in the breast cancer models was addressed by shRNA knockdown. Three shRNAs reduced PADI4 protein expression by 60–70% (shown for sh79 and sh80 in Fig. 1E–F), and significantly decreased PADI4 mRNA expression (Fig. 1G). Citrullination of histone H3 at arginine R2, R8, R17 and R26 was reduced by 2-5-fold, confirming a reduction in PADI4 catalytic activity (Fig. 1H). Addressing biological responses, PADI4 knockdown had no effect on proliferation *in vitro* (Fig. 2A; Supplementary Fig. S3A), but

enhanced cell survival in response to staurosporine, a potent inducer of apoptosis (Fig. 2A, Supplementary Fig. S3B), consistent with known inhibitory effects of PADI4 in cell survival (26,27). No effect on apoptosis under basal or low serum conditions was seen (Supplementary Fig. S3C). PADI4 knockdown led to significantly increased cell invasion and migration (Fig. 2A; Supplementary Fig. S3D), as seen by others (22). Clonogenicity was also increased in all lines except MDA-MB-231 LM2 (Fig. 2A; Supplementary Fig. S3E). Thus, PADI4 suppresses multiple biological processes relevant to tumor progression.

The ability to form tumorspheres under anchorage-independent conditions is an *in vitro* surrogate for cancer stem cell activity (28). Tumorsphere-forming efficiency increased upon PADI4 knockdown in all cell lines tested (Fig. 2B–C), suggesting that PADI4 inhibition may increase the size of the CSC subpopulation. To demonstrate a dependency on PADI4 enzymatic activity, we inhibited PADI4 pharmacologically using GSK484, a citrullination inhibitor with >35-fold selectivity for PADI4 over other PADs, or its inactive analog GSK106 (12), and observed a 3-fold reduction in Histone H3 citrullination after 3 days (Fig. 2D). Pretreatment with GSK484 for 3 days significantly increased tumorsphere formation (Fig. 2E) and invasion and migration (Fig 2F). Prolonged treatment (8–10 days) with GSK484 further increased tumorsphere formation to levels comparable to the genetic knockdown, while having no additional effect on invasion/migration (Supplementary Fig. S3F,G). Overall, these data suggest that endogenous PADI4 activity may reduce the size of the CSC population and that this effect is dependent on its citrullination activity.

Overexpression of any of the five catalytically active isoforms of PADI4 can decrease the CSC population *in vitro*

To further explore the effect of PADI4 on CSCs, we overexpressed PADI4 in MCF10Ca1h cells. In our Western blots, we observed several endogenous PADI4 bands (Fig. 1C–D). Given that the PADI4 effect on the CSC population was dependent on its citrullination activity (Fig 2D–E), we overexpressed the five RefSeq-predicted protein isoforms of PADI4 that retained most or all of the C-terminal catalytic domain; full-length PADI4 cDNA (FL), and four shorter protein isoforms: X1, X2, X3 and X7 (see Supplementary Fig S2). We detected all five protein isoforms in the transduced MCF10Ca1h cells by Western blot (Fig. 2G,H). All isoforms ran 2–8 kDa smaller than predicted (Supplementary Fig. S4A), suggesting some structural feature or post-translational modification that reduces the apparent molecular weights. This observation is consistent with our inability to detect a band at 75kDa corresponding to the predicted size of the full-length transcript endogenously in any breast cancer cell line. X3 and X7, which lack the predicted nuclear localization signal, were still found in the nucleus at levels higher than the endogenous protein (Supplementary Fig. S4B). All isoforms, including X2 which lacks two of the four predicted catalytic residues, gave up to a 4-fold increase in PADI4 catalytic activity over control, that was further enhanced following estradiol and calcium ionophore treatment (Fig. 2I; Supplementary Fig. S4C). Despite the wide range of protein expression, all isoforms showed similar maximal catalytic activity (varying <2-fold), suggesting that PADI4 protein levels may not be the major limiting factor for PADI4 activity. Indeed, all five PADI4 isoforms similarly suppressed tumorsphere formation in MCF10Ca1h cells (Fig. 2J). These results

support the hypothesis that PADI4 may inhibit the CSC subpopulation, and show that this occurs with all predicted isoforms that retain the majority of the catalytic domain.

PADI4 knockdown increases tumor initiating ability of breast cancer cells *in vivo*

The gold standard assay for cancer stem cell activity is the ability to initiate tumorigenesis *in vivo* (29). To confirm the increase in CSC frequency following PADI4 knockdown, we performed extreme limiting dilution assay (ELDA) *in vivo* in immunocompromised mice using four different breast cancer models. Groups of female virgin athymic nu/nu mice were transplanted with progressively lower tumor cell inocula and tumor incidence was assessed at day 60 (Table 1). The relative stem cell frequency was 2-7-fold higher following PADI4 inhibition in all four breast cancer models tested. These results confirm that endogenous PADI4 limits the size of the CSC population in breast tumors *in vivo*.

PADI4 inhibition modulates the breast cancer epigenome

Earlier work had suggested that PADI4 could affect transcriptional programs of self-renewal and differentiation through histone citrullination and modification of the epigenome (15,16). To address effects of PADI4 inhibition on the cancer cell epigenome, MCF10Ca1h cells were treated with the PADI4 inhibitor GSK484, or its inactive analog (GSK106, hereafter “CON”) for 3 days, and then subjected to ChIP-seq analysis for histone H3, H3R2/8/17citr, H3K4me3 and H3K27me3, and RNA-Seq. The short treatment timeframe was selected to highlight direct effects of PADI4 inhibition. Concurrent enrichment of the repressive H3K27me3 mark with an activating H3K4me3 mark identifies “bivalent domains” that are characteristic of poised promoters of many genes involved in lineage specification and differentiation in embryonic stem cells (24).

The heatmaps in Fig. 3A–D show relative ChIP-seq intensity for the GSK484-treated vs CON condition in +/− 10kb regions flanking the transcriptional start site (TSS) for genes encoding all RefSeq NM transcripts. The most prominent effect of PADI4 inhibition was a widespread increase in H3 binding to chromatin, particularly between the TSS and +400bp (Fig. 3A). Total cellular H3 protein levels were unaffected by PADI4 inhibition (Supplementary Fig. S5A), and PADI4 inhibition had little effect on genomic distribution of the other histone marks (Fig. 3B–D). The increased H3 loading was seen at the TSS (Fig. 3F), and at proximal enhancers within +/− 4kb of the TSS (Fig. 3G), but was much weaker at distal enhancers (Fig. 3H). The increased H3 binding on PADI4 inhibition is consistent with the expectation that blockade of arginine citrullination increases the net positive charge on H3, resulting in tighter binding to DNA. We noted that the H3 binding increase on PADI4 inhibition was negatively correlated with H3 binding in the control (Pearson r=−0.178, p=2.35E-281; Supplementary Fig. S5B,C), suggesting the possibility that TSS sites associated with the more open (nucleosome-depleted) chromatin may be more accessible to binding of new H3. Interestingly, the heatmap appeared almost identical between H3.CON and H3R2/8/17citr.CON (Supplementary Fig. S5B, D), indicating that all H3 histones carried a similar level of citrullination across the region surrounding the TSS. Matched heatmaps for the other two histone marks and RNA-Seq are given in Supplementary Fig. S5E–K. Surprisingly, there was essentially no change in levels of citrullinated H3 around the TSS following 3 days of PADI4 inhibition (Fig. 3B), suggesting that the citrullination

modification is very stable, and that PADI4 inhibition leads to loading of unmarked H3 that dilutes but does not displace the pre-existing citrullinated H3. The functional consequence of the increased H3 loading at the TSS is unclear, asKit occurred to a similar extent at loci where transcription was increased or decreased by PADI4 inhibition (Supplementary Fig. S6A).

RNA expression across the transcriptome was negatively correlated with H3 binding near TSS (Pearson r=−0.127; p=2.58E-143), and was positively correlated with the activation mark H3K4me3 (r=0.17; p=4.15E-262) but negatively correlated with the repressive H3K27me3 mark (r=−0.05; p=4.52E-262) in the control state (Supplementary Fig. S5B–K, Fig. S6B–K). Thus, the major histone marks were behaving as expected with respect to transcriptional activity. However, there were no widespread changes in global H3K3me3, H3K27me3 occupancy or global RNA expression following PADI4 inhibition (Fig. 3C–E). Indeed, even at an individual level, surprisingly few genes were differentially expressed following the 3 days of PADI4 inhibition. Only 94 genes were affected by GSK484 and not control compound ($p < 0.05$) and of these, only 24 showed an absolute fold-change >=1.5 (Supplementary Table S5 Tab1). The minimal effect of GSK484 treatment on gene expression may reflect a difference in kinetics between PADI4 control of chromatin and its readout in gene expression. However, since we saw clear biological effects of PADI4 inhibition in this 3-day timeframe, it is also possible that the biological output is driven by the integrated effect of multiple small changes in gene expression, most of which are below our detection limit.

In relation to the original hypothesis that PADI4 might preferentially target poised promoters, we identified genetic loci that had both H3K4me3 and H3K27me3 marks within +/− 4kb of the TSS (Supplementary Fig. S7). Further analysis at the peak level identified 448 promoters that were bivalent in either GSK484 or control samples or both (Supplementary Fig. S8A and Supplementary Table S5, Tab2,3). However, there was almost no overlap of these poised promoters with the differentially expressed gene list generated by RNA-Seq on GSK484-treated cells, or with a larger genelist generated by microarray analysis following prolonged PADI4 inhibition by genetic knockdown (Supplementary Fig. S8B,C). Thus the observed biological effects of PADI4 inhibition could not be attributed to changes in state of poised promoters. Overall, PADI4 inhibition had a profound effect on histone H3 loading at transcriptional start site, but this altered epigenetic state had minimal effects on the transcriptome of the bulk tumor cell population, at least in the short-term.

PADI4 has different effects on stemness-related transcription factors in the CSC and non-CSC compartments

Our epigenetic and transcriptomic analyses of the bulk tumor cell population following PADI4 inhibition gave no clear clues as to the molecular mechanism underlying the effect of PADI4 on the CSCs. We therefore reasoned that PADI4 might be having effects specifically in the CSC compartment, that our bulk analyses did not have the sensitivity to detect. To test this hypothesis, we transduced MCF10Ca1h and MDA-MB-231 LM2 cells with the SORE6 cancer stem cell reporter (Fig. 4A) (23) to facilitate separation of CSCs from non-CSCs. This reporter has been rigorously validated in these cell lines as enriching for cells with the

expected CSC properties, including the ability to initiate and sustain tumorigenesis through multiple transplant generations *in vivo* (23). The SORE6 reporter identified a minority population of cells as CSCs in both cell lines (Fig. 4B), and treatment of the unsorted cultures with the GSK484 PADI4 inhibitor led to the expected increase in CSCs (Fig. 4C). Cells were treated with either the PADI4 inhibitor (GSK484) or inactive analog (GSK106) for 3 days, and then FACS sorted into CSCs (SORE6+) and non-CSCs (SORE6−) for transcriptomic and ChIP-qPCR analysis (Fig. 4D). RT-qPCR showed no significant differences in *PADI4* mRNA expression between CSCs and non-CSCs in multiple cell lines (Supplementary Fig S9A). As expected, basal expression of stem cell transcription factors was higher in the CSC compartment of untreated cells, with *NANOG*, *POU5F1 (OCT4)*, *SOX9*, and *ZEB1* being higher in the CSC compartment in both cell lines, and MCF10Ca1h additionally showing higher expression of *SNAI* (Supplementary Fig. S9B), confirming enrichment of the stem phenotype. Pathway analysis of RNA-Seq datasets for compartment-specific effects of PADI4 inhibition indicated an increase in invasion, migration and metastasis and a reduction in cell death specifically in the non-CSCs and unsorted tumor cells (Fig. 4E), suggesting that the non-CSC compartment may be the main target of these previously described tumor suppressive effects of PADI4. In contrast, PADI4 inhibition was predicted to suppress cell cycle progression and transcription specifically in the CSC compartment (Fig. 4E). These data suggest that PADI4 has distinct biological activities in the CSC and non-CSC compartments, despite similar levels of expression, with the compartment-specific effects possibly due to predicted differences in epigenetic context (5,30).

While we were unable to detect core stem cell transcription factors in our RNA-Seq libraries, they were detectable in the same RNA samples by RT-qPCR. Importantly, we found that PADI4 inhibition upregulated *NANOG* and *POU5F1/OCT4* mRNA specifically in the CSC population in both breast cancer models (Fig. 4F,G), while having no effect *(POU5F1)* or the opposite effect *(NANOG)* in the non-CSC compartment. This upregulation of *NANOG* and *POU5F1* was not evident in bulk (unsorted) cells following short-term GSK484 PADI4 inhibitor treatment (Supplementary Fig. S9C) or long-term genetic PADI4 knockdown (Supplementary Fig. S9D). Indeed, PADI4 inhibition actually decreased *NANOG* expression in the bulk (unsorted) cell population in the MCF10Ca1h model, consistent with the observed effect on the non-CSCs. We speculate that PADI4 inhibition may enforce the stem phenotype in CSCs while enforcing a more differentiated phenotype in the non-CSCs. Other stem cell transcription factors were relatively unaffected by PADI4 inhibition in the CSC compartment (Supplementary Fig. S9E). Although we did not directly test the functional role of *NANOG* and *POU5F1/OCT4* by the gold standard *in vivo* limiting dilution assay in the current study, *NANOG* and *POU5F1/OCT4* have previously been shown to induce/maintain a stem phenotype in many breast cancer models, see eg. (31,32). Importantly, knockdown of *NANOG* in the MCF7 and MDA-MB-231 models that we used in this study decreased the CD44+CD24− fraction that is enriched for stem phenotype, as well as decreasing mammosphere formation (31). Thus PADI4 may decrease the size of the CSC population by inhibiting transcription of these master stemness transcription factors specifically in the CSC compartment. Due to the relatively small size of the CSC

compartment, this effect cannot readily be seen when analyzing the bulk tumor cell population.

PADI4 affects the epigenetic status of *NANOG* and *POU5F1* specifically in the CSC compartment

Citrullination of histone arginines by PADI4 can counteract arginine methylation by protein arginine methyltransferases (PRMTs), with variable effects depending on whether the specific arginine methylation is normally activating or repressive (20,33,34). A recent study found that PADI4 repressed *CTCF* transcription by keeping the activating H3R17me2a histone mark low, while stimulating *IL6ST* transcription by lowering the repressive H3R2me2a mark (35). We confirmed these effects of PADI4 at the CTCF and IL6ST loci in the MCF10Ca1h cells, and observed them to occur primarily in the non-CSC compartments (Supplementary Fig. S10A,B). No effect of PADI4 inhibition was seen on either mark at the *GAPDH* or *MYOD1* promoters used as negative controls (Supplementary Fig. S10C,D. To address whether a similar mechanism might underlie the effect of PADI4 on *POU5F1* and *NANOG* expression, we assessed H3K4me3, H3R17me2a, and H3R2me2a histone marks at the *NANOG* and *POU5F1* promoters by ChIP-qPCR in sorted CSCs and non-CSCs from MCF10Ca1h or MDA-MB-231 LM2 cells treated with either the PADI4 inhibitor GSK484 or the inactive analog GSK106. Normalizing to total H3, we found that PADI4 inhibition increased the activating H3R17me2a mark at the *NANOG* and *POU5F1* promoters in CSCs in both cell lines (Fig. 4H,I). We also observed an enrichment of the activating H3K4me3 mark at these loci in the CSCs following PADI4 inhibition (Fig. 4J,K). No effect was seen on the repressive mark H3R2me2a at *NANOG* and *POU5F1* promoters in the CSCs of either cell lines (Supplementary Figure S10E,F). These data show that endogenous PADI4 down-regulates *NANOG* and *POU5F1* expression specifically in CSCs by blocking establishment of the activating H3R17me2a histone mark at their promoters, and thus suppressing the transcription of these master transcription factors of stemness.

A gene signature reflecting PADI4 activity is associated with good outcome in human breast cancer

Finally we wished to address whether the tumor cell-autonomous suppressive effects of PADI4 might have clinical relevance. In this study we found a poor correlation between PADI4 mRNA, protein and enzymatic activity, likely reflecting multiple layers of control over PADI4 activity. This observation complicates the interpretation of studies that seek to correlate PADI4 mRNA or protein levels directly with outcome in clinical cancer datasets. Furthermore, stromal PADI4 expression can have the opposite effect on cancer progression (36,37). To overcome this problem, we sought to generate a gene expression signature that specifically reflects tumor cell-autonomous responses to PADI4, and is enriched for PADI4-driven tumor suppressor effects described by us and others.

In our RNA-Seq analyses of GSK484-treated cells, we saw relatively little impact on gene expression following short-term (3 days) pharmacologic PADI4 inhibition, despite this treatment clearly reducing citrullination activity (Fig. 2D), and affecting migration, invasion and tumorsphere formation (Fig. 2E,F). Given the stability of the citrullination mark, we reasoned that there might be a delay before readily detectable changes in the transcriptome

are seen. We therefore performed microarray analysis of MCF10Ca1h cells with and without stable shPADI4 knockdown to look for longer-term effects (Fig. 5A). shRNA knockdown reduced PADI4 mRNA levels by 5-fold (Fig. 1G) and citrullination activity by >2-fold (Fig. 1H). Consistent with the RNA-Seq data, principal component analysis showed that PADI4 knockdown had a relatively small effect on the transcriptome as the samples did not clearly segregate by genotype (Supplementary Fig. S11). However, differential gene expression analysis using DESeq2 identified 181 genes that were affected specifically by PADI4 knockdown ($p<0.01$, $FC\geq1.5$; Figure 5A; Supplementary Table S5, Tab4). Also consistent with results from our smaller RNA-Seq genesets (Fig. 4E), Ingenuity Pathway Analysis (IPA) showed the expected enrichment for functional annotations relating to cell migration and survival (Fig.5B). IPA analysis further identified a connection between PADI4 and estrogen signaling as seen in the literature (20,33), and suggested provocative links between PADI4 and inflammatory mediators such as TGF-β, IL1 and TNF (Fig. 5C). Hallmark geneset enrichment analysis identified only the EMT program as showing a weak enrichment on PADI4 knockdown (Fig. 5D), consistent with our observations of increased cancer stem activity and a previous study implicating PADI4 in suppressing EMT (22). Thus despite the relatively small number of affected genes and the low magnitude of the expression changes on PADI4 knockdown, the genelist nevertheless captures components of tumor suppressive effects of PADI4 on multiple biological processes, with processes affecting the bulk tumor cell population being most prominent.

Next we generated a 20-gene PADI4 inhibition signature from the differentially expressed genes identified above (Supplementary Table S5, Tab5 for signature genes, and Supplementary Fig. S12 for functional annotation). Expression of the signature was lowest in luminal A subtype tumors, suggesting that these have the highest PADI4 tumor suppressor activity (Supplementary Fig. S13). In the METABRIC dataset of 2000 breast cancer patients (38), patients whose tumors showed high expression of the PADI4 inhibition signature had significantly worse overall survival compared to those with low expression (Fig. 5E), an effect that was more prominent in estrogen receptor-positive (ER+) breast cancers (Fig. 5E). We saw similar results with the same signature using the GOBO tool for a metanalysis of independent breast cancer datasets (Supplementary Fig. S14A–C). These results suggest that PADI4 has tumor cell-autonomous suppressive effects that can affect outcome in human breast cancer, particularly within the ER+ subtype.

DISCUSSION

Epigenetic regulation plays an essential role in specifying cell phenotype in the differentiation hierarchies of normal and diseased tissue (39). Histones and their associated code of post-translational modifications are major epigenetic modifiers that regulate gene expression (7). Citrullination catalyzed by PADIs is an under-studied post-translational modification that has recently been implicated in the regulation of stemness and differentiation in embryonic and hematopoietic stem cells (13,15,16). Here we demonstrate for the first time that endogenous PADI4 suppresses breast cancer progression by limiting the size of the cancer stem cell population in multiple breast cancer models *in vitro* and *in vivo*. This effect is mediated through epigenetic regulation of *NANOG* and *POU5F1* expression specifically in the CSC compartment.

Our data suggest an important role for PADI4 as a tumor suppressor in breast cancer. The existence of multiple PADI4 isoforms, and the poor correlation between expression levels and citrullination activity, pose challenges for interpreting studies that correlate PADI4 expression with outcome in human cancer cohorts. In one large study, PADI4 was shown to be overexpressed in many different human tumors, including breast cancer, suggesting it might play a tumor promoting role (40). In contrast, our preclinical findings clearly demonstrate that endogenous PADI4 functionally restricts the size of the CSC population in multiple breast cancer models and suppresses tumor formation. We also confirmed the findings of others that PADI4 can suppress additional pro-tumorigenic activities such as cell survival, invasion and migration (22,27). The ability of PADI4 to promote differentiation of promyelocytic leukemia cells, and suppress EMT and the development of more invasive tumors in the MCF7 breast cancer model is also consistent with a tumor suppressor role (22,41). Supporting the potential clinical relevance of PADI4-driven tumor suppression, our gene signature reflecting tumor cell-autonomous PADI4 activity was associated with good outcome in clinical breast cancer cohorts. Additionally, the PADI4 locus is genetically altered in 8/816 sequenced cases of breast cancer in the TCGA cohort, with 1 amplification, 5 deep deletions and 2 missense mutations of unknown significance. The deep deletions are also consistent with a potential role for PADI4 as a tumor suppressor.

Reconciling these disparate observations, we propose that PADI4 may have opposing effects on the tumor cell and on the tumor stroma in regulating tumor progression. Under physiological conditions, PADI4 is mainly expressed in peripheral blood neutrophils to provide antibacterial innate immunity via the formation of neutrophil extracellular traps (NETs) (14). Interestingly, metastatic breast cancer cells were recently shown to induce neutrophils to form metastasis-supporting NETs in the absence of infection, in a PADI4-dependent manner (42). Furthermore NETs could awaken dormant metastatic breast cancer cells in mice (36), and facilitate premetastatic niche formation in ovarian cancer models (43). In a mechanistically distinct process, citrullination of the extracellular matrix by PADI4 secreted in tumor exosomes facilitated human liver colorectal cancer metastasis through effects on tumor cell motility and plasticity (44). These data strongly suggest that PADI4 in stromal cells, or secreted into the extracellular microenvironment by tumor cells, can promote tumor progression. Thus, it seems likely that the net effect of PADI4 in cancer progression depends on a context-dependent balance between tumor cell-autonomous suppressive effects and stromally-mediated tumor promoting effects.

Given the importance of histones as PADI4 targets (10,13,33), we addressed whether PADI4 might regulate stemness through global effects on the epigenome via changes in histone citrullination. Our ChIP-seq analysis revealed a strikingly widespread effect of PADI4 activity in maintaining a reduced level of H3 around transcriptional start sites. Since nucleosome occupancy is typically decreased upstream of transcriptionally active genes and increased in regulatory regions of silenced genes (45,46), we speculate that PADI4 may maintain a state of elevated genomic responsiveness to transcriptional regulation by creating a more permissive and accessible genome (47). This finding is consistent with a previous study showing PADI4 enrichment near transcriptional start sites of actively transcribed genes in MCF7 breast cancer cells (21). However, we observed relatively little impact of PADI4 inhibition on the tumor cell transcriptome in the present study. This suggests either that there

is a kinetic uncoupling of the effects of PADI4 inhibition on genomic organization and downstream transcriptional responses, potentially due to a requirement for additional regulatory signals, or that biological responses to PADI4 may be elicited by the integration of many small magnitude effects on the transcriptome that are difficult to detect over the noise. These findings merit further exploration, but did not explain our observed effects on the CSC. Instead, using a reporter-based strategy to enrich for CSCs, we showed that PADI4 regulates breast CSCs via epigenetic down-regulation of the master transcription factors of stemness, *NANOG* and *POU5F1,* specifically in the CSC compartment. Since CSCs are a relatively small subpopulation, this mechanism is not evident in analyses of the bulk tumor cell population. PADI4 represses the promoters of these key stemness genes by inhibiting establishment of the activating H3R17me2a histone mark, thus suppressing transcription of *NANOG* and *POU5F1.* A similar mechanism has been described for PADI4 suppression of CTCF transcription in erythroleukemia cells (35).

In conclusion, our findings reveal important insights into PADI4 function in breast cancer progression, suggesting a novel tumor suppressive role involving epigenetic modulation of the tumor-initiating breast CSCs. PADI4 inhibitors are in pre-clinical development for treatment of autoimmune disease (48). Also, given compelling data showing a tumor-promoting role for PADI4 in the tumor stroma, PADI4 inhibition has been proposed as a cancer therapy (37,49). The relative balance between the pathologic and homeostatic effects of PADI4 will determine the potential benefit of systemic PADI4 inhibition in different disease states. Further exploration of this complex biology will help guide clinical intervention with PADI4 as a target.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We acknowledge the expert technical assistance of Dr. Xiaolin Wu (Laboratory of Molecular Technology, Frederick National Laboratory for Cancer Research) with the microarray analysis, and Dr. Bao Tran (Center for Cancer Research Illumina Sequencing Facility) with the NextGen Sequencing, and Dr. Huaitian Liu with data deposition in the GEO repository. This work was funded by the Intramural Research Program of the National Cancer Institute, Center for Cancer Research, NIH project ZIA BC 005785 (to LMW).

REFERENCES

1. Global Burden of Disease Cancer C, Fitzmaurice C, Abate D, Abbasi N, Abbastabar H, Abd-Allah F, et al. Global, Regional, and National Cancer Incidence, Mortality, Years of Life Lost, Years Lived With Disability, and Disability-Adjusted Life-Years for 29 Cancer Groups, 1990 to 2017: A Systematic Analysis for the Global Burden of Disease Study. JAMA Oncol 2019
2. Batlle E, Clevers H. Cancer stem cells revisited. Nat Med 2017;23:1124–34 [PubMed: 28985214]
3. Alison MR, Lin WR, Lim SM, Nicholson LJ. Cancer stem cells: in the line of fire. Cancer Treat Rev 2012;38:589–98 [PubMed: 22469558]
4. Al-Hajj M, Wicha MS, Benito-Hernandez A, Morrison SJ, Clarke MF. Prospective identification of tumorigenic breast cancer cells. Proc Natl Acad Sci U S A 2003;100:3983–8 [PubMed: 12629218]
5. Poli V, Fagnocchi L, Zippo A. Tumorigenic Cell Reprogramming and Cancer Plasticity: Interplay between Signaling, Microenvironment, and Epigenetics. Stem Cells Int 2018;2018:4598195 [PubMed: 29853913]

6. Vincent A, Van Seuningen I. On the epigenetic origin of cancer stem cells. Biochim Biophys Acta 2012;1826:83–8 [PubMed: 22495062]
7. Strahl BD, Allis CD. The language of covalent histone modifications. Nature 2000;403:41–5 [PubMed: 10638745]
8. Wang YC, Peterson SE, Loring JF. Protein post-translational modifications and regulation of pluripotency in human stem cells. Cell Res 2014;24:143–60 [PubMed: 24217768]
9. Yuzhalin AE. Citrullination in Cancer. Cancer Res 2019;79:1274–84 [PubMed: 30894374]
10. Nakashima K, Hagiwara T, Yamada M. Nuclear localization of peptidylarginine deiminase V and histone deimination in granulocytes. J Biol Chem 2002;277:49562–8 [PubMed: 12393868]
11. Asaga H, Yamada M, Senshu T. Selective deimination of vimentin in calcium ionophore-induced apoptosis of mouse peritoneal macrophages. Biochem Biophys Res Commun 1998;243:641–6 [PubMed: 9500980]
12. Lewis HD, Liddle J, Coote JE, Atkinson SJ, Barker MD, Bax BD, et al. Inhibition of PAD4 activity is sufficient to disrupt mouse and human NET formation. Nat Chem Biol 2015;11:189–91 [PubMed: 25622091]
13. Christophorou MA, Castelo-Branco G, Halley-Stott RP, Oliveira CS, Loos R, Radzisheuskaya A, et al. Citrullination regulates pluripotency and histone H1 binding to chromatin. Nature 2014;507:104–8 [PubMed: 24463520]
14. Li P, Li M, Lindberg MR, Kennett MJ, Xiong N, Wang Y. PAD4 is essential for antibacterial innate immunity mediated by neutrophil extracellular traps. J Exp Med 2010;207:1853–62 [PubMed: 20733033]
15. Nakashima K, Hagiwara T, Ishigami A, Nagata S, Asaga H, Kuramoto M, et al. Molecular characterization of peptidylarginine deiminase in HL-60 cells induced by retinoic acid and 1alpha,25-dihydroxyvitamin D(3). J Biol Chem 1999;274:27786–92 [PubMed: 10488123]
16. Nakashima K, Arai S, Suzuki A, Nariai Y, Urano T, Nakayama M, et al. PAD4 regulates proliferation of multipotent haematopoietic cells by controlling c-myc expression. Nat Commun 2013;4:1836 [PubMed: 23673621]
17. Baka Z, Gyorgy B, Geher P, Buzas EI, Falus A, Nagy G. Citrullination under physiological and pathological conditions. Joint Bone Spine 2012;79:431–6 [PubMed: 22366145]
18. Guo W, Zheng Y, Xu B, Ma F, Li C, Zhang X, et al. Investigating the expression, effect and tumorigenic pathway of PADI2 in tumors. Onco Targets Ther 2017;10:1475–85 [PubMed: 28331341]
19. Horibata S, Rogers KE, Sadegh D, Anguish LJ, McElwee JL, Shah P, et al. Role of peptidylarginine deiminase 2 (PAD2) in mammary carcinoma cell migration. BMC Cancer 2017;17:378 [PubMed: 28549415]
20. Wang Y, Wysocka J, Sayegh J, Lee YH, Perlin JR, Leonelli L, et al. Human PAD4 regulates histone arginine methylation levels via demethylimination. Science 2004;306:279–83 [PubMed: 15345777]
21. Zhang XS, Gamble MJ, Stadler S, Cherrington BD, Causey CP, Thompson PR, et al. Genome-Wide Analysis Reveals PADI4 Cooperates with Elk-1 to Activate c-Fos Expression in Breast Cancer Cells. Plos Genetics 2011;7
22. Stadler SC, Vincent CT, Fedorov VD, Patsialou A, Cherrington BD, Wakshlag JJ, et al. Dysregulation of PAD4-mediated citrullination of nuclear GSK3beta activates TGF-beta signaling and induces epithelial-to-mesenchymal transition in breast cancer cells. Proc Natl Acad Sci U S A 2013;110:11851–6 [PubMed: 23818587]
23. Tang B, Raviv A, Esposito D, Flanders KC, Daniel C, Nghiem BT, et al. A flexible reporter system for direct observation and isolation of cancer stem cells. Stem Cell Reports 2015;4:155–69 [PubMed: 25497455]
24. Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell 2006;125:315–26 [PubMed: 16630819]
25. Minn AJ, Gupta GP, Siegel PM, Bos PD, Shu W, Giri DD, et al. Genes that mediate breast cancer metastasis to lung. Nature 2005;436:518–24 [PubMed: 16049480]

26. Tanikawa C, Espinosa M, Suzuki A, Masuda K, Yamamoto K, Tsuchiya E, et al. Regulation of histone modification and chromatin structure by the p53-PADI4 pathway. Nat Commun 2012;3:676 [PubMed: 22334079]
27. Liu GY, Liao YF, Chang WH, Liu CC, Hsieh MC, Hsu PC, et al. Overexpression of peptidylarginine deiminase IV features in apoptosis of haematopoietic cells. Apoptosis 2006;11:183–96 [PubMed: 16502257]
28. Dontu G, Abdallah WM, Foley JM, Jackson KW, Clarke MF, Kawamura MJ, et al. In vitro propagation and transcriptional profiling of human mammary stem/progenitor cells. Genes Dev 2003;17:1253–70 [PubMed: 12756227]
29. Rycaj K, Tang DG. Cell-of-Origin of Cancer versus Cancer Stem Cells: Assays and Interpretations. Cancer Res 2015;75:4003–11 [PubMed: 26292361]
30. Avgustinova A, Benitah SA. The epigenetics of tumour initiation: cancer stem cells and their chromatin. Curr Opin Genet Dev 2016;36:8–15 [PubMed: 26874045]
31. Hu C, Xu L, Liang S, Zhang Z, Zhang Y, Zhang F. Lentivirus-mediated shRNA targeting Nanog inhibits cell proliferation and attenuates cancer stem cell activities in breast cancer. J Drug Target 2016;24:422–32 [PubMed: 26339994]
32. Beltran AS, Rivenbark AG, Richardson BT, Yuan X, Quian H, Hunt JP, et al. Generation of tumor-initiating cells by exogenous delivery of OCT4 transcription factor. Breast Cancer Res 2011;13:R94 [PubMed: 21952072]
33. Cuthbert GL, Daujat S, Snowden AW, Erdjument-Bromage H, Hagiwara T, Yamada M, et al. Histone deimination antagonizes arginine methylation. Cell 2004;118:545–53 [PubMed: 15339660]
34. Denis H, Deplus R, Putmans P, Yamada M, Metivier R, Fuks F. Functional connection between deimination and deacetylation of histones. Mol Cell Biol 2009;29:4982–93 [PubMed: 19581286]
35. Kolodziej S, Kuvardina ON, Oellerich T, Herglotz J, Backert I, Kohrs N, et al. PADI4 acts as a coactivator of Tal1 by counteracting repressive histone arginine methylation. Nat Commun 2014;5:3995 [PubMed: 24874575]
36. Albrengues J, Shields MA, Ng D, Park CG, Ambrico A, Poindexter ME, et al. Neutrophil extracellular traps produced during inflammation awaken dormant cancer cells in mice. Science 2018;361
37. Demers M, Wong SL, Martinod K, Gallant M, Cabral JE, Wang Y, et al. Priming of neutrophils toward NETosis promotes tumor growth. Oncoimmunology 2016;5:e1134073 [PubMed: 27467952]
38. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature 2012;486:346–52 [PubMed: 22522925]
39. Lunyak VV, Rosenfeld MG. Epigenetic regulation of stem cell fate. Hum Mol Genet 2008;17:R28–36 [PubMed: 18632693]
40. Chang X, Han J, Pang L, Zhao Y, Yang Y, Shen Z. Increased PADI4 expression in blood and tissues of patients with malignant tumors. BMC Cancer 2009;9:40 [PubMed: 19183436]
41. Song G, Shi L, Guo Y, Yu L, Wang L, Zhang X, et al. A novel PAD4/SOX4/PU.1 signaling pathway is involved in the committed differentiation of acute promyelocytic leukemia cells into granulocytic cells. Oncotarget 2016;7:3144–57 [PubMed: 26673819]
42. Park J, Wysocki RW, Amoozgar Z, Maiorino L, Fein MR, Jorns J, et al. Cancer cells induce metastasis-supporting neutrophil extracellular DNA traps. Sci Transl Med 2016;8:361ra138
43. Lee W, Ko SY, Mohamed MS, Kenny HA, Lengyel E, Naora H. Neutrophils facilitate ovarian cancer premetastatic niche formation in the omentum. J Exp Med 2019;216:176–94 [PubMed: 30567719]
44. Yuzhalin AE, Gordon-Weeks AN, Tognoli ML, Jones K, Markelc B, Konietzny R, et al. Colorectal cancer liver metastatic growth depends on PAD4-driven citrullination of the extracellular matrix. Nat Commun 2018;9:4783 [PubMed: 30429478]
45. Lee CK, Shibata Y, Rao B, Strahl BD, Lieb JD. Evidence for nucleosome depletion at active regulatory regions genome-wide. Nat Genet 2004;36:900–5 [PubMed: 15247917]

46. Shivaswamy S, Bhinge A, Zhao Y, Jones S, Hirst M, Iyer VR. Dynamic remodeling of individual nucleosomes across a eukaryotic genome in response to transcriptional perturbation. PLoS Biol 2008;6:e65 [PubMed: 18351804]

47. Gaspar-Maia A, Alajem A, Meshorer E, Ramalho-Santos M. Open chromatin in pluripotency and reprogramming. Nat Rev Mol Cell Biol 2011;12:36–47 [PubMed: 21179060]

48. Cully M Deal watch: Bristol-Myers Squibb acquires potential keys to treating rheumatoid arthritis. Nat Rev Drug Discov 2016;15:301 [PubMed: 27139990]

49. Cedervall J, Zhang Y, Olsson AK. Tumor-Induced NETosis as a Risk Factor for Metastasis and Organ Failure. Cancer Res 2016;76:4311–5 [PubMed: 27402078]

STATEMENT OF SIGNIFICANCE

Findings demonstrate a novel activity of the citrullinating enzyme PADI4 in suppressing breast cancer stem cells through epigenetic repression of stemness master transcription factors NANOG and OCT4.

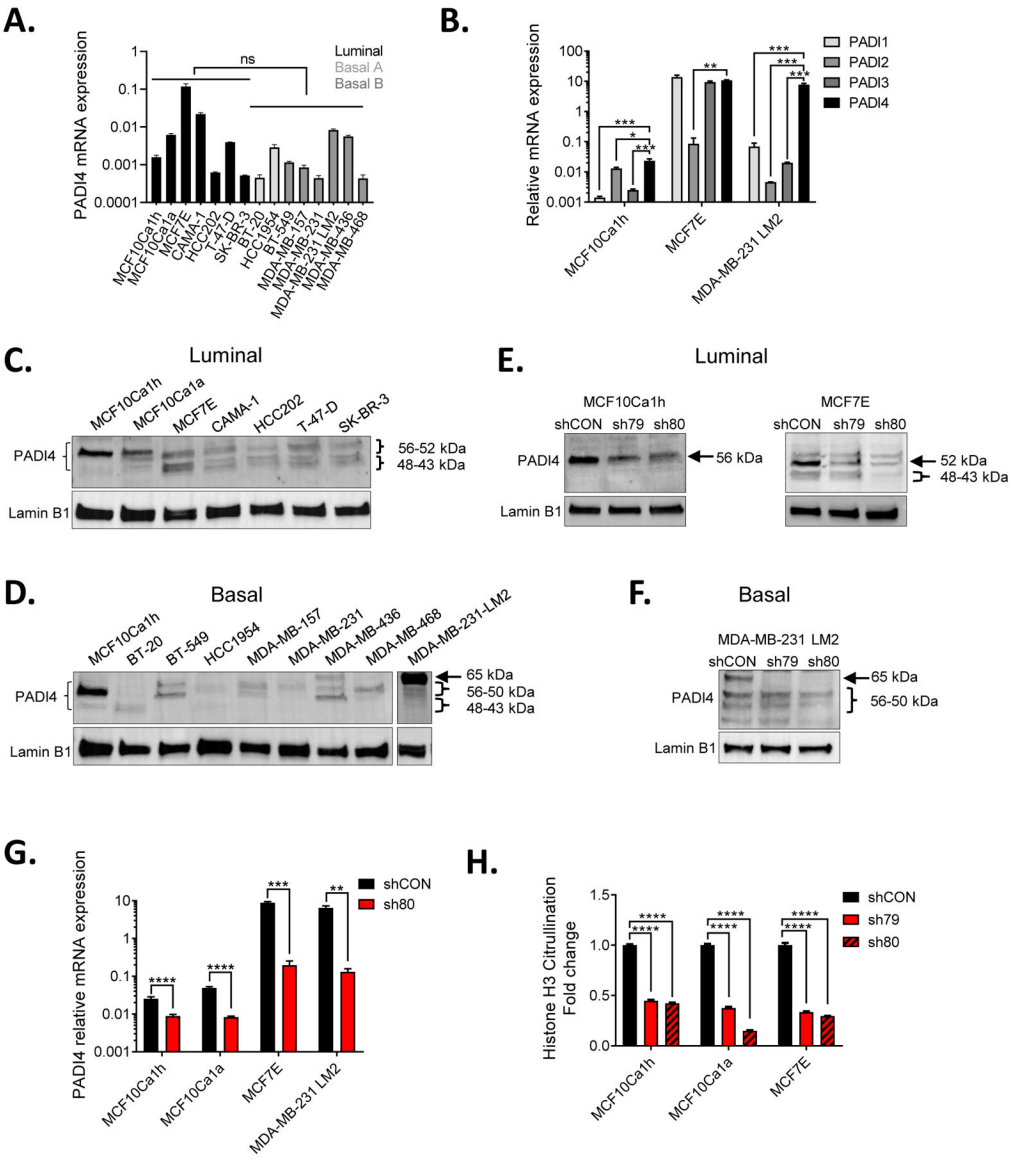

Figure 1. PADI4 is expressed in human breast cancer cell lines.

(A) PADI4 mRNA expression normalized to GAPDH in human breast cancer cell lines as assessed by RT-qPCR. t-test for difference between basal and luminal subclasses. **(B)** PADI1-4 mRNA expression in representative breast cancer cell lines as assessed by RT-qPCR. Dunnett's multiple comparison test with PADI4 as comparator. **(C-D)** Western blots of nuclear PADI4 protein expression in luminal (panel C) and basal A and basal B (panel D) breast cancer cell lines. Lamin B1, loading control. **(E-F)** Western blots of nuclear PADI4 following shRNA knockdown in MCF10Ca1h, MCF7E cells (E) and MDA-MB-231 LM2 cells (F) cells. **(G)** RT-qPCR of PADI4 following knockdown (mean +/−SD, n=3; t-test). **(H)** ELISA for total citrullinated histone H3. Results are normalized to shCON (mean +/− SD, n=3; Dunnett's multiple comparison test with shCON as the comparator. *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001.

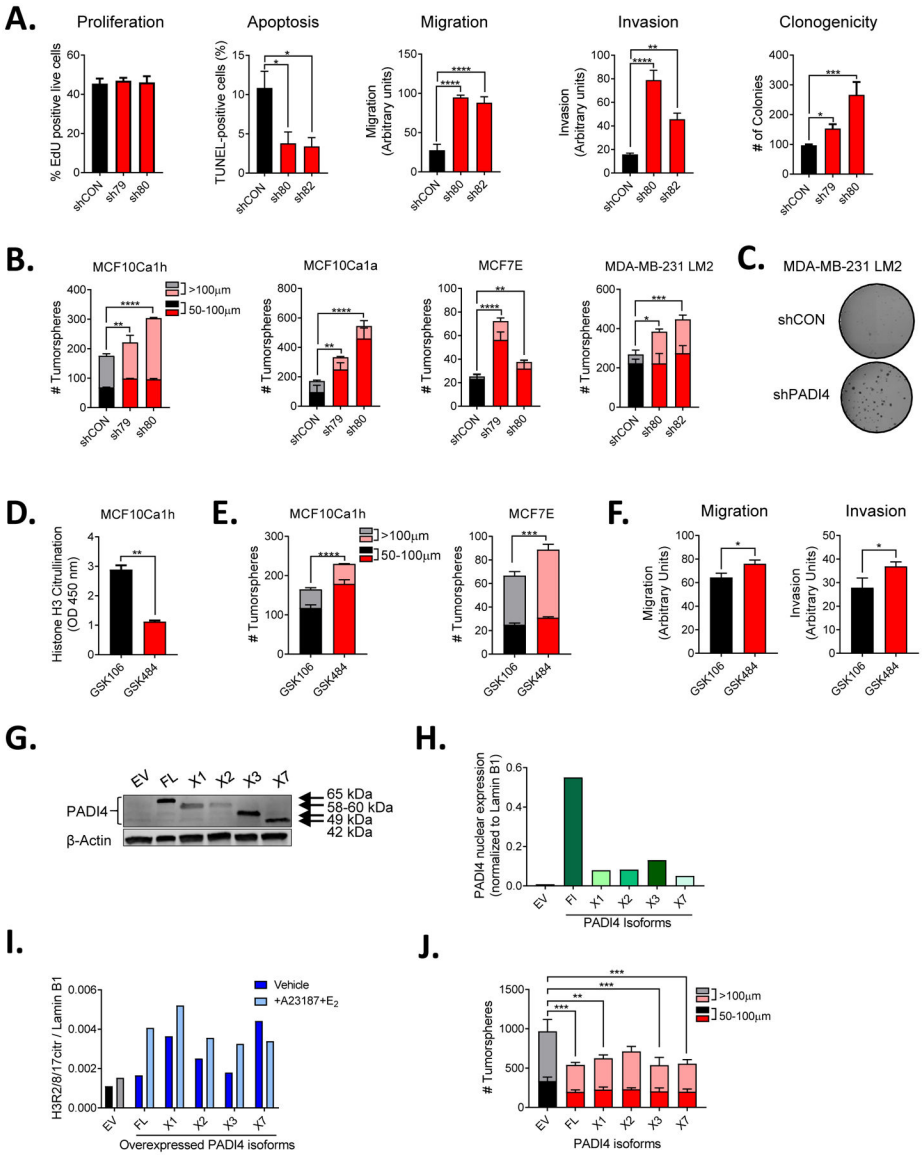

Figure 2. PADI4 suppresses pro-tumorigenic biological responses in the tumor cell.

(A) Effects of PADI4 knockdown on biological responses of MCF10Ca1h cells *in vitro*. **(B, C)** Effect of PADI4 knockdown on tumorsphere formation in multiple breast cancer cells lines (B), with representative images shown for MDA-MB-231 LM2 (C). **(D-F)** Effect of 3-day pharmacological inhibition of PADI4 by GSK484 on H3 citrullination in MCF10Ca1h cells **(**D**),** tumorsphere formation assay in MCF10Ca1h cells and MCF7E cells (E), and invasion and migration in MCF10Ca1h cells (F). GSK106 is the inactive control compound. **(G)** Western blots for overexpression of PADI4 isoforms in total cell lysates from MCF10Ca1h cells. FL, full-length; X1, X2, X3, X7, variants. For details of isoforms, see Supplementary Fig. S2. EV, Empty vector. β-Actin, loading control. **(H)** Quantification of nuclear PADI4 protein by Western blot, normalized to Lamin B1. **(I)** Quantification of PADI4 citrullination activity following PADI4 isoform overexpression in MCF10Ca1h cells. Cells treated with vehicle or 17-β-estradiol (E_2) and calcium ionophore (A23187) were

analyzed by Western blotting with H3R2/8/17citr antibodies and normalized to Lamin B1. **(J)** Effect of PADI4 overexpression on tumorsphere formation in MCF10Ca1h cells. Results are mean +/− SD for n=3. Statistics are Dunnett's multiple comparison test (A-B, J) with shCON or EV as the comparators; and unpaired t-test in D-E. *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001.

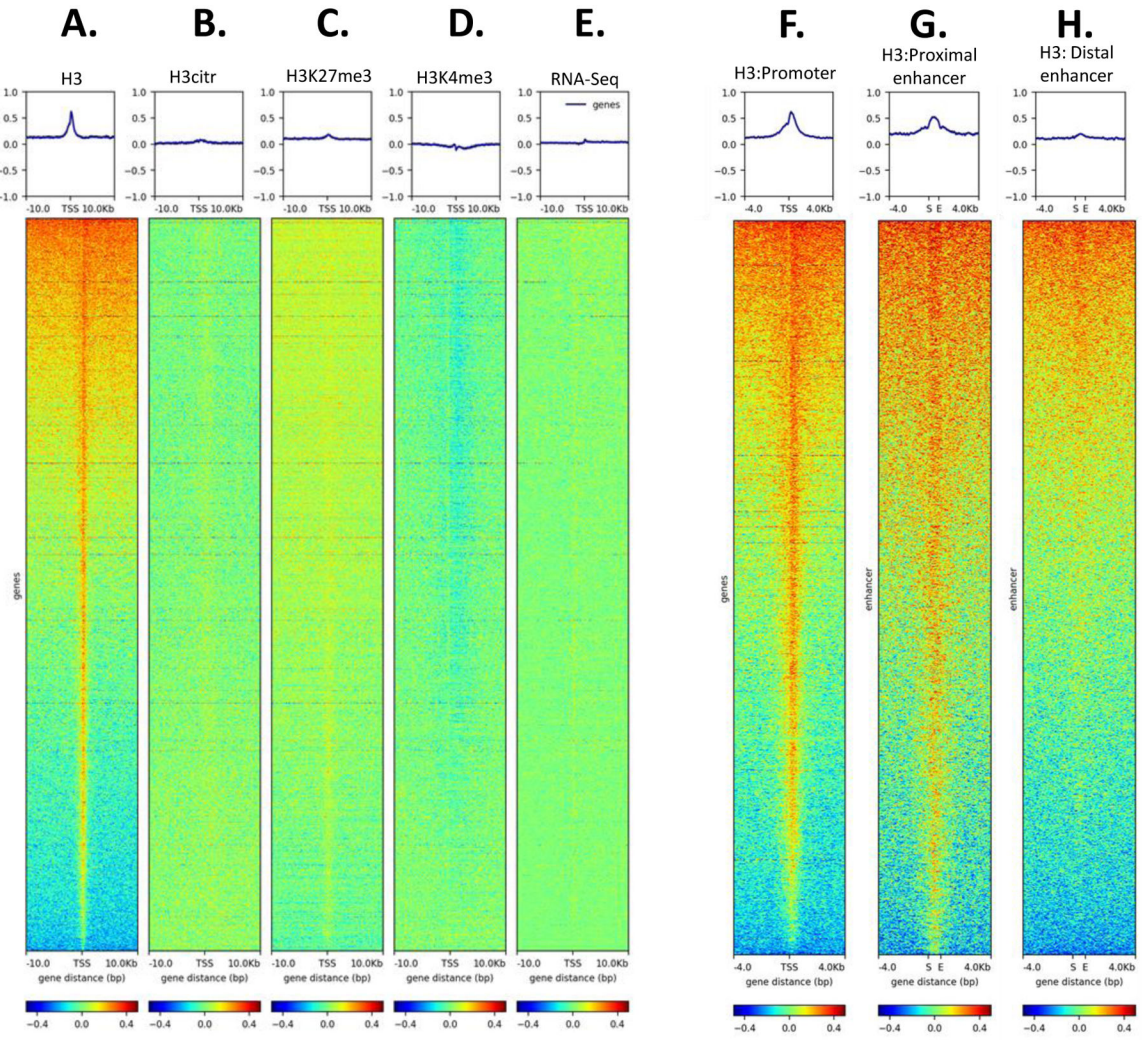

Figure 3. PADI4 inhibition has widespread effects on Histone H3 distribution but little effect on the transcriptome.

(A-D) Heatmaps for the difference in ChIP-seq signal in GSK484 (PADI4 inhibitor) vs GSK106 (control) treated MCF10Ca1h cells for H3 (A), H3R2/8/17citr (B), H3K27me3 (C) and H3K4me3 (D) at the transcriptional start site (TSS) +/− 10kb regions for all Ref-Seq NM transcripts. The heatmaps are sorted by H3. GSK484 and show ChIP-seq intensity (log2(ChIP/IgG)). Treatment duration was 3 days. **(E)** Corresponding heatmap for the difference in RNA-seq data (GSK484 vs CON). **(F-H)** Heatmaps for the difference in H3 ChIP-seq signal at promoters (F: +/− 4kb of TSS); proximal enhancers (G: enhancer elements located within 4kb of the TSS); and distal enhancers (H: enhancers located >4 kb distal to a TSS). S and E show start and end of enhancer element. The graphs above each heatmap show the average signal intensity across the region for all loci.

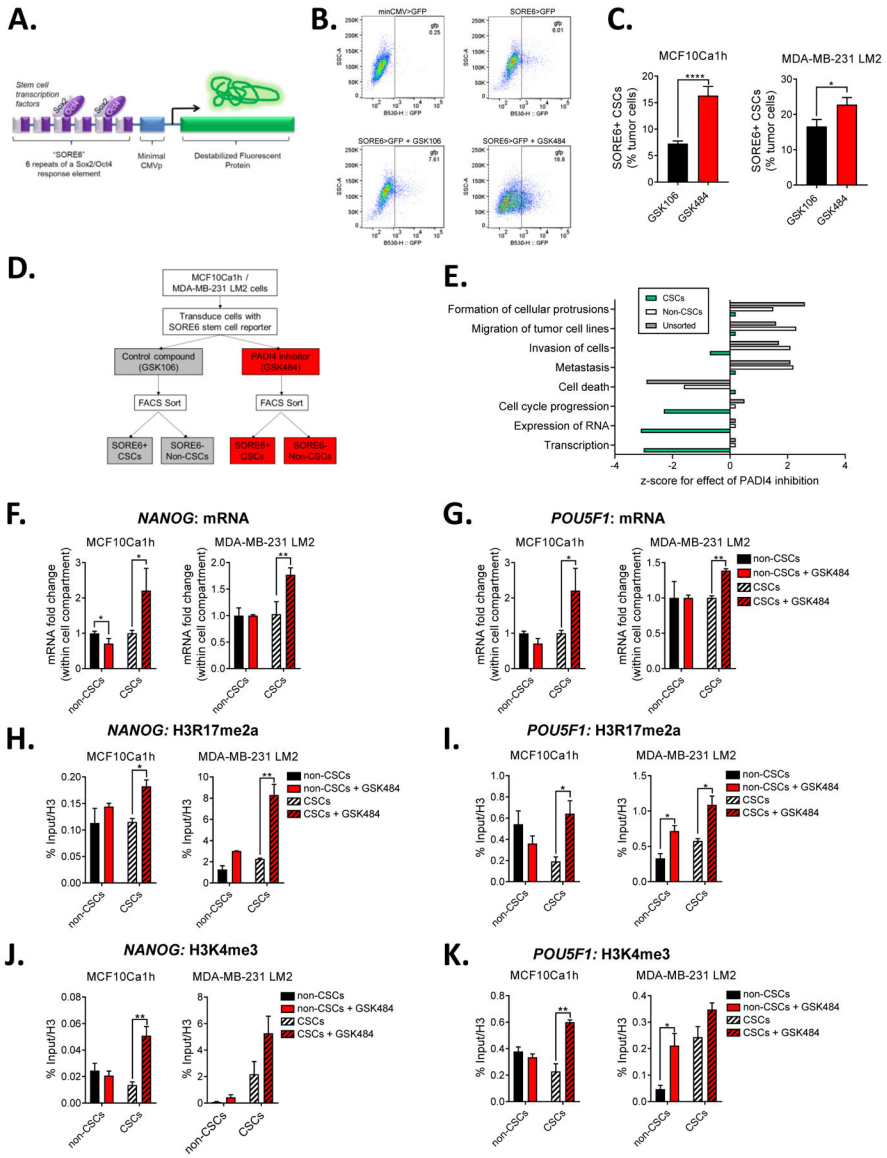

Figure 4. Compartment-specific regulation of stem cell transcription factors by PADI4.

(A) Schematic of the SORE6 cancer stem cell reporter (for more details, see text). **(B)** Flow cytometry of MCF10Ca1h cells transduced with control vector (minCMV>GFP) or CSC reporter (SORE6>GFP) and treated with PADI4 inhibitor (GSK484) or control (GSK106) for 3 days. **(C)** Quantification of SORE6+ CSCs in MCF10Ca1h and MDA-MB-231 LM2 cells. Results are mean +/− SD for n=3; unpaired t-test. **(D)** Experimental schematic for analysis of cell-compartment specific effects of PADI4. **(E)** Ingenuity Pathway Analysis of Biofunction enrichment in RNA-Seq generated genelists following PADI4 inhibition in unsorted cells, sorted CSCs and non-CSCs. Enriched biofunctions with an absolute z-score>2 in any cell compartment are shown. A positive z-score reflects activation on PADI4 inhibition. **(F,G)** Fold change in *NANOG* (F) and *POU5F1* (G) mRNA in CSCs and non-CSCs following PADI4 inhibition in MCF10Ca1h and MDA-MB-231 LM2 cells as assessed by RT-qPCR. **(H,I)** H3R17me2a ChIP of *NANOG* (H) and *POU5F1* (I) promoters in CSCs

and non-CSCs in MCF10Ca1h and MDA-MB-231 LM2 cells. **(J,K)** H3K4me3 ChIP of *NANOG* (J) and *POU5F1* (K) promoters. Results are percent enrichment compared to input and normalized to H3; mean +/− SD, n=3; unpaired t-test. *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001.

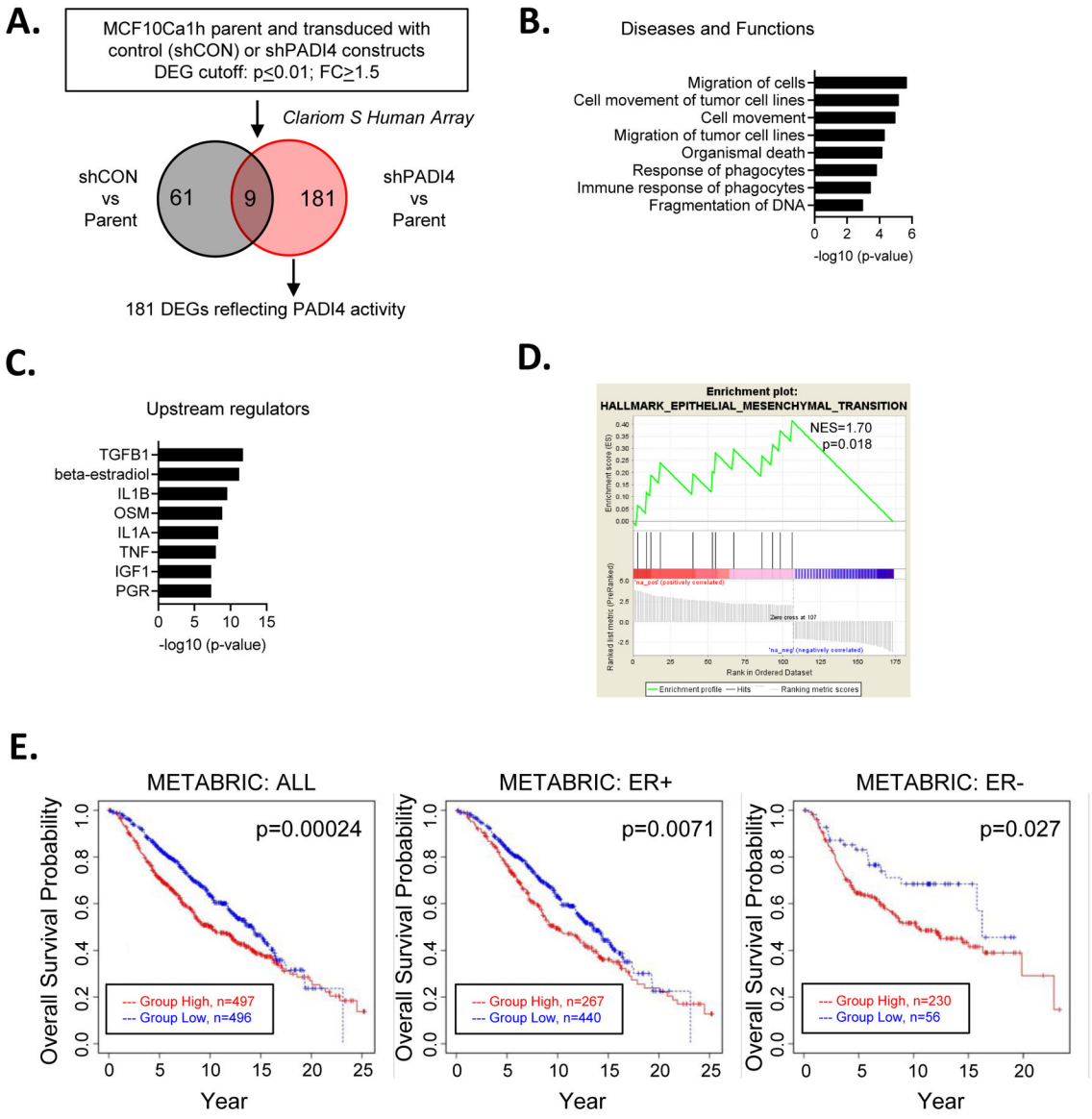

Figure 5. The PADI4-regulated transcriptome is associated with outcome in clinical breast cancer datasets.

(A) Schematic for identification of PADI4-regulated transcripts in MCF10Ca1h cells following long-term inhibition of PADI4 by genetic knockdown. **(B-C)** Ingenuity Pathway analysis) of top upregulated diseases and functions (B), and upstream regulators (C) derived from core analysis of the 181 differentially expressed genes (DEGs) reflecting PADI4 inhibition. **(D)** Hallmark GSEA analysis identifies “Epithelial Mesenchymal Transition” gene set as significantly enriched. NES, normalized enrichment score. **(E)** Kaplan Meier overall survival curves for breast cancer patients in METABRIC dataset, stratified by higher than median (red) or lower than median (blue) expression of a PADI4 inhibition signature. n=number of patients/group. ER, estrogen receptor. For details of signature generation, see Supplementary Methods.

Table 1.

Effect of PADI4 knockdown on tumor-initiating ability of breast cancer cell lines *in vivo*.

Cells were orthotopically implanted in nude mice at the specified cell inocula, and tumor formation was assessed at day 60. Cancer stem cell frequency was calculated from tumor incidence rates using ELDA software.

		Tumor incidence at different cell inocula							
Model	***Cell inoculum***	***2.5×10^5***	***5×10^4***	***1×10^4***	***1×10^3***	**1/stem cell frequency**	**Confidence interval**	**Relative stem cell frequency**	**Chi2 p-value**
MCF10Ca1h	Parent	10/10	5/10	0/10	0/10	75410	139349-40809	1	
	shCON	10/10	8/10	0/10	0/10	45474	85992-24047	1.65	0.29
	shPADI4 79	10/10	10/10	3/10	0/10	18760	18760-10053	4.02	0.00368
	shPADI4 80	10/10	10/10	3/10	0/10	17575	17575-9217	4.29	0.00263
	Cell inoculum	***2.5×10^5***	***1×10^5***	***1×10^4***	***1×10^3***	**1/stem cell frequency**	**Confidence interval**	**Relative stem cell frequency**	**Chi2 p-value**
MCF10Ca1a	Parent	ND	8/10	3/10	0/10	52177	103175-26387	1	
	shCON	ND	8/10	5/10	0/10	41253	82167-20712	1.26	0.61
	shPADI4 79	ND	8/10	5/10	0/10	8469	17493-4101	6.16	0.000511
	shPADI4 80	ND	10/10	7/10	1/10	7366	14958-3627	7.08	0.000128
	Cell inoculum	***2.5×10^5***	***1×10^5***	***1×10^4***	***1×10^3***	**1/stem cell frequency**	**Confidence interval**	**Relative stem cell frequency**	**Chi2 p-value**
MCF7E	Parent	ND	8/10	2/10	0/10	59215	117500-29842	1	
	shCON	ND	8/10	1/10	0/10	67767	135682-33846	0.87	0.787
	shPADI4 79	ND	10/10	3/10	0/10	25058	53620-11710	2.36	0.101
	shPADI4 80	ND	10/10	4/10	2/10	20275	44941-9148	2.92	0.0435
	Cell inoculum	***2.5×10^5***	***2×10^5***	***2×10^4***	***2×10^3***	**1/stem cell frequency**	**Confidence interval**	**Relative stem cell frequency**	**Chi2 p-value**
MDA-MB-231 LM2	shCON	ND	10/10	7/10	1/10	16938	34985-8201	1	
	shPADI4 80	ND	10/10	10/10	6/10	2180	4957-959	7.77	0.000265