bioRxiv preprint doi: https://doi.org/10.1101/497313; this version posted December 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Id proteins promote a cancer stem cell phenotype in triple negative breast cancer via Robo1-dependent c-Myc activation

Wee S Teo[1,2], Aurélie S Cazet[1,2], Daniel L Roden[1,2], Kate Harvey[1],Nitheesh K[3],Holly Holliday[1,2],Christina Konrad[1], Reshma Murali[3], Binitha Anu Varghese[3], Archana P T[3,4], Eoin Dodson[1,2], Andrea McFarland[1], Simon Junankar[1,2], Sunny Ye[1], Jessica Yang[1], Iva Nikolic[1,2], Jaynish Shah[5], Laura A Baker[1,2], Ewan K.A.Millar[1,6,7], Mathew J Naylor[1,2,8], Christopher J Ormandy[1,2], Sunil R Lakhani[9], Warren Kaplan[1,10],Albert Mellick[11,12], Sandra A O'Toole[1,6,7], Radhika Nair[1,2,3*], Alexander Swarbrick[1,2*].

Affiliations

[1]Garvan Institute of Medical Research, Darlinghurst, New South Wales, 2010, Australia
[2]St Vincent's Clinical School, Faculty of Medicine, UNSW Sydney, New South Wales, 2010, Australia
[3] Cancer Research Program, Rajiv Gandhi Centre for Biotechnology, Kerala 695014, India
[4]Manipal Academy of Higher Education, Manipal, India
[5] School of Medical Sciences, UNSW, New South Wales, 2052, Australia
[6]Department of Tissue Pathology and Diagnostic Oncology, Royal Prince Alfred Hospital, Camperdown
[7]Sydney Medical School, University of Sydney, Camperdown, New South Wales, 2050, Australia
[8]Discipline of Physiology & Bosch Institute, School of Medical Sciences, University of Sydney, NSW 2006
[9]The University of Queensland, UQ Centre for Clinical Research, School of Medicine and Pathology Queensland, The Royal Brisbane & Women's Hospital, Herston, Brisbane, Queensland 4029, Australia
[10]Peter Wills Bioinformatics Centre, Garvan Institute of Medical Research, Darlinghurst, NSW 2010, Australia
[11]School of Medicine, Western Sydney University, Locked Bag 1797, Penrith, NSW 2751, Australia.
[12]Ingham Institute for Applied Medical Research, South Western Sydney Clinical School UNSW & CONCERT Translational Cancer Research Centre, 1 Campbell Street, Liverpool, NSW 2170, Australia.

* Co-corresponding authors

*Dr Radhika Nair
Present address
Rajiv Gandhi Centre for Biotechnology
Kerala, India
radhikanair@rgcb.res.in
Phone: +91-471-2781251

Fax: +91-471-2346333

* A/Prof Alexander Swarbrick
Cancer Research Division, Garvan Institute of Medical Research

bioRxiv preprint doi: https://doi.org/10.1101/497313; this version posted December 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

370 Victoria St, Darlinghurst, NSW, 2010
Australia
a.swarbrick@garvan.org.au
Phone: +61-2-9355-5780
Fax: +61-2-355-5869

Running title: Id1/3 controls CSC phenotypes via negative transcriptional regulation of Robo1

bioRxiv preprint doi: https://doi.org/10.1101/497313; this version posted December 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Abstract

Breast cancers display phenotypic and functional heterogeneity and several lines of evidence support the existence of cancer stem cells (CSCs) in certain breast cancers, a minor population of cells capable of tumour initiation and metastatic dissemination. Identifying factors that regulate the CSC phenotype is therefore important for developing strategies to treat metastatic disease. The Inhibitor of Differentiation Protein 1 (Id1) and its closely related family member Inhibitor of Differentiation 3 (Id3) are expressed by a diversity of stem and progenitor cells and are required for metastatic dissemination in experimental models of breast cancer. Here, we show that Id1 is expressed in rare neoplastic cells within ER-negative breast cancers and enriched in brain metastases compared to patient matched primary tissues. To address the function of Id1 expressing cells within tumours, we developed two independent murine models of Triple Negative Breast Cancer (TNBC) in which a genetic reporter permitted the prospective isolation of Id1^{+} cells. Id1^{+} cells are enriched for self-renewal in tumoursphere assays *in vitro* and for tumour initiation *in vivo*. Conversely, depletion of Id1 and Id3 in the 4T1 murine model of TNBC demonstrates that Id1/3 are required for cell proliferation and self-renewal *in vitro*, as well as primary tumour growth and metastatic colonization of the lung *in vivo*. We defined a novel mechanism of Id function via negative regulation of the Roundabout Axon Guidance Receptor Homolog 1 (Robo1) leading to activation of a Myc transcriptional programme.

bioRxiv preprint doi: https://doi.org/10.1101/497313; this version posted December 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Statement of significance

Using three independent models of Id1 expression and Id1/3 depletion, we demonstrate that Id1/3 is a functional marker of CSCs in the TNBC subtype, acting by negative transcriptional regulation of Robo1 via a Myc signature. Targeting Id proteins or Robo1 may represent a novel therapeutic approach to the prevention or treatment of metastatic TNBC.

bioRxiv preprint doi: https://doi.org/10.1101/497313; this version posted December 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Introduction

Several lines of evidence suggest that rare sub-populations of tumour cells, commonly termed cancer stem cells (CSCs), drive key tumour phenotypes such as self-renewal, drug resistance and metastasis and contribute to disease relapse and associated patient mortality (1-4). Recent evidence points to the hypothesis that CSCs are not static, but they rather exist in dynamic states, driven by critical transcription factors and are highly dependent on the microenvironmental cues (5-7). Understanding the molecular networks that are critical to the survival and plasticity of CSCs are fundamental to resolving clinical problems associated with chemo-resistance and metastatic residual disease.

The Inhibitor of DNA binding (ID) proteins have previously been recognized as regulators of CSCs and tumour progression (8). These proteins constitute a family of four highly conserved transcriptional regulators (ID1-4) that act as dominant-negative inhibitors of basic helix–loop–helix (bHLH) transcription factors. ID proteins are expressed in a tissue-specific and stage-dependent manner and are required for the maintenance of self-renewal and multipotency of embryonic and many tissue stem cells (9-10). For example, high levels of Id1 identified type B adult neuronal stem and progenitor cells and Id is required to sustain self-renewal capacity of this population (11). Id proteins are sensitive to a number of external stimuli such as TGF-β signalling and integrate the signals from the extracellular milieu leading to the regulation of key genes involved in cell proliferation, self-renewal and other programs critical for normal development and cancer progression (12-13).

Previous studies have reported a functional redundancy among the four members of the mammalian Id family, in particular Id1 and Id3 (Referred to collectively here as Id), and their overlapping expression patterns during normal development and cancer (14-18). The dependency of CSCs towards the expression of ID proteins for tumour initiation has been well

bioRxiv preprint doi: https://doi.org/10.1101/497313; this version posted December 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

studied in colon cancer and malignant glioma. Experimental work has shown that the combined expression of Id1 and Id3 governed the maintenance of colon CSCs by regulating the cell-cycle inhibitor p21 (18). In contrast, suppression of Id lead to a loss of tumour initiating potential and sensitised CSCs to the chemotherapeutic agent Oxaliplatin, demonstrating that Id controls the CSC phenotype in colon cancer. This requirement was also seen in high-grade human glioblastoma, where genetic ablation or pharmacological inhibition of the TGF-β- ID axis reduced sphere formation and invasiveness *in vitro* and tumour re-initiation *in vivo* (14).

ID1 and ID3 have also been found to play a key role in breast cancer progression and metastasis. Significant co-expression of ID1 and ID3 was reported specifically in TNBC and both the transcription factors were required for primary tumour formation and lung metastatic colonization (15). We have previously demonstrated that Id1 cooperates with activated Ras signalling and promotes mammary tumour initiation and metastasis *in vivo* by supporting long-term self-renewal and proliferative capacity (19). Id1 acts in the genesis of high grade metastatic TNBC by averting cellular senescence and acting downstream of p21 Waf1/Cip1 (19). Id1 also stimulates cell proliferation by promoting the expression of Cyclin D1 and Cyclin E and their associated cyclin-dependant kinases, CDK4 and CDK2 in human breast epithelial cells (20). Additional work has clearly implicated Id1 as direct target of important factors like the matrix metalloproteinase MT1-MMP (21), KLF17 (22), Cyclin D1 (23), Bcl-2 (24), and BMI1 (25) among others.

Even though several Id1-dependent targets have been identified, we still lack a comprehensive picture of the downstream molecular mechanisms controlled by Id and their associated pathways mediating breast cancer progression and metastasis. Understanding the importance of Id in driving a CSC phenotype is of particular interest in the poor prognostic TNBC subtype for which targeted therapies are still urgently needed. In this study, we unambiguously demonstrate using three independent mouse model of TNBC that Id is important for the

bioRxiv preprint doi: https://doi.org/10.1101/497313; this version posted December 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

maintenance of a cancer stem cell phenotype. Importantly, we also describe a novel mechanism by which Id controls the CSC state by negatively regulating Robo1 to control proliferation and self-renewal via activation of a Myc transcriptional programme.

bioRxiv preprint doi: https://doi.org/10.1101/497313; this version posted December 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Results

Id1 marks a subset of cells with stem-like properties

We investigated the role of Id in the context of CSC biology in the TNBC molecular subtype. This was based on observations by our group and others that ID1 is expressed in a minority of neoplastic cells within ER-negative disease, which includes the TNBC and Her2+ subtypes (15) (Supplementary Figure 1A, B). No significant difference in the distribution of ID3 expression was observed across different subtypes (Data not shown). Immunohistochemistry (IHC) analysis revealed that ID1 is expressed by a small minority of cells (range 0.5-6% of total cancer cells) in ~50 % of TNBC and of Her2+ tumours. Interestingly, when we looked at the expression of ID1 and ID3, we found only three cases where both were highly expressed in the TNBC cases and none correlated in the ER+ and Her2+ cases (disregarding cases with H score<10) (Supplementary Figure 1C). This suggests that though ID1 and ID3 may have functional redundancy, their function in TNBC is not necessarily correlated. To test the hypothesis that $Id1^+$ cells have a unique malignant phenotype, we developed two murine models of TNBC that permit the prospective isolation of $Id1^+$ cells for functional assays. In the first, we used the p53-/- TNBC model, in which TNBC tumours arise spontaneously following transplantation of Tp53-null mammary epithelium into the mammary fat pads of naïve FVB/n mice. The tumours were then transplanted into naïve recipients; this method has been previously used to study murine TNBC CSCs (26-27). IHC analysis of p53-/- TNBC tumours revealed that ~ 5% of neoplastic cells expressed Id1, consistent with the observation in the clinical samples, while Id3 marked a majority of the tumour cells in this model (Figure 1A).

To create a genetic reporter cell line, p53-/- mammary tumour cells were transduced with a lentiviral GFP reporter construct under the control of the Id1 promoter (Id1/GFP), as described previously (28) (Supplementary Figure 1D). FACS sorting for GFP expression followed by

bioRxiv preprint doi: https://doi.org/10.1101/497313; this version posted December 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

immunoblotting confirmed the ability of the Id1/GFP construct to prospectively enrich for Id1+ cells from this model (Supplementary Figure 1E). We next sought to understand if Id1 marked cells with high self-renewal capacity in this model using tumoursphere assays, a well-established surrogate for cells with high self-renewal capacity (29-30). We observed an increase in the self-renewal capacity of Id1/GFP$^+$ cells when compared to the unsorted cell population in the p53-/- model (Figure 1B).

Our next step was to establish the *in vivo* relevance of the increased self-renewal capacity of the Id1/GFP$^+$ tumour cells observed *in vitro*. We determined the tumour initiating capacity (TIC) of the Id1/GFP$^+$ cells using the limiting dilution assay (31). Id1/GFP$^+$ cells (1/42) showed more than a 7-fold increase in tumour initiating cell frequency over Id1/GFP$^-$ cells (1/314) after serial passage (Figure 1C).

We used a second murine model to assess the phenotypes associated with Id1$^+$ cells, in which the expression of SV40-large T antigen in the mammary epithelium under the control of the C3 promoter leads to the development of TNBC in mice (32-33). These tumours (C3-Tag) closely model the TNBC subtype as assessed by gene expression profiling (Supplementary Figure 1F). To generate a genetic reporter of Id1 promoter activity in TNBC, the C3-Tag model was crossed to a genetic reporter mouse model in which GFP is knocked into the intron 1 of the Id1 gene. The resulting Id1GFPC3-Tag mice (called Id1C3-Tag model) developed mammary tumours with similar kinetics as the parental C3-Tag mice and have a classical CK14$^+$/CK8$^-$ phenotype (Supplementary Figure 1G). 5% and 60% of cells in the Id1C3-Tag tumour were stained positive for Id1 and Id3 expression, respectively, as observed by IHC (Figure 1D). We were able to isolate Id1$^+$ tumour cells with a high degree of purity by FACS based on GFP expression followed by q-RT PCR (Figure 1E). Similar to the p53-/- Id1/GFP model, Id1$^+$/GFP$^+$ cells from the Id1C3-Tag model were enriched for sphere-forming capacity (Figure 1F).

bioRxiv preprint doi: https://doi.org/10.1101/497313; this version posted December 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Using the Id1C3-Tag model, we also looked at the association of Id1/GFP expression with the expression of established CSC markers CD29, CD24 and CD61. $CD29^+/CD24^+$ status was previously reported to mark the tumourigenic subpopulation of cells in murine mammary tumours (34-35). The $Id1^+/GFP^+$ cells are predominantly of the $CD29^+/CD24^+$ phenotype (Figure 1G), with a 1.6-fold higher proportion of cells expressing both CD29 and CD24 compared to the $Id1^-/GFP^-$ cells which comprise the bulk of the tumour. Interestingly, $Id1^+/GFP^+$ cells are also highly enriched for $CD24^+/CD61^+$ expression (more than 6-fold increase in $Id1^+/GFP^+$ cells), which was also reported to mark a murine breast CSC population (36) (Figure 1G).

We found no correlation between Id1 expression (as indicated by GFP^+) and the $CD29^+/CD24^+$ phenotype in the first transplantation round using the p53-/- model, as the percentage of $CD29^+/CD24^+$ cells was similar across each gating group (Supplementary Figure 1H). Interestingly, the $Id1^+$ cells, , which are the putative cells that give rise to the increased TIC as shown in Figure 1C, showed 10 times less $CD24^+/CD29^+$cells in the second transplantation round (36). The ability of the markers like CD24, CD29 and CD61 to identify the CSC population is clearly model-dependent. In addition to CD29 and CD24, the percentage of GFP^+ cells were also analysed and a higher percentage of GFP^+ cells was found in the second transplantation round compared to the first round tumour result (Supplementary Figure 1I), consistent with the increase in TICs reported in Figure 1B.

Id1 and Id3 are required for self-renewal *in vitro* and metastatic competency *in vivo*

We next assessed the requirement for Id1 and Id3 in maintaining the CSC phenotypes. Numerous studies have shown that there exists a functional redundancy between Id1 and Id3, so functional studies typically require depletion of both the factors to reveal a phenotype (37). We used the transplantable syngeneic 4T1 TNBC model, which has a high propensity to

bioRxiv preprint doi: https://doi.org/10.1101/497313; this version posted December 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

spontaneously metastasize to distant sites (including bone, lung, brain and liver), mimicking the aggressiveness of human breast cancers (38-43). We first checked the expression of Id1 and Id3 in 4T1 primary tumours by IHC and western blotting. IHC analysis showed that 15% of the tumour cells express high levels of Id1, and 35% have intermediate levels of Id1 expression, whereas the expression of Id3 was found in most of the cells (Figure 2A).

We used an inducible lentiviral shRNA system (44) that permits reversible knock down of Id1 and Id3 in response to doxycycline (Dox) treatment. Two clonal 4T1 cell lines (pSLIK (single lentivector for inducible knockdown) shId clonal cell lines 8 and 12, hereafter referred to as C8 and C12) were chosen, along with a control line (C1), based on the efficiency of Id knock down (Figure 2B). Morphological and proliferative characterization of the Id knock down cells when compared to the control showed that Id depletion resulted in a significant decrease in cell proliferation and migration *in vitro* (Supplementary figure 2 A, B, C, D)

We next interrogated the effect of Id depletion on the self-renewal capacity of the C1, C8 and C12 cell lines. Dox-dependent shRNA induction significantly reduced the ability of the C8 and C12 cells to form primary tumourspheres in the suspension culture (Figure 2C). This effect was not observed in the control cell line (C1; Figure 2C). A significant further decrease in self-renewal capacity of C8 and C12 lines was observed when primary tumourspheres were passaged to the secondary stage (Figure 2D). The Id depleted tumourspheres were also significantly smaller in size compared to tumourspheres that expressed Id (Figure 2E, Supplementary Figure 2E).

To assess if the self-renewal phenotype controlled by Id is reversible, we firstly passaged primary tumourspheres (previously treated with Dox) to secondary tumourspheres. The secondary tumourspheres were then cultured in the presence or absence of Dox, to maintain the Id knockdown status or to allow the re-expression of Id, respectively (Supplementary Figure 2F, G). The secondary tumourspheres cultured without Dox re-established their self-

bioRxiv preprint doi: https://doi.org/10.1101/497313; this version posted December 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

renewal capacity as evidenced by the ability to form new tumourspheres (Figure 2F-H), suggesting that Id depletion does not lead to a permanent loss of self-renewal capacity.

To verify the specificity and to rule out any potential off-target effects of the Id1 shRNA, we rescued Id1 expression in the shRNA expressing C8 cells. As shown in Supplementary Figure 3A, the rescue clone showed restored ID1 protein expression level. Cell viability assays performed on the C8 cells showed that overexpression of ID1 significantly restored cell proliferation in Id knockdown cells (Supplementary Figure 3B).

To determine whether Id1 and Id3 are required for primary tumour and metastatic growth *in vivo*, C8 cells were orthotopically transplanted into the mammary fat pad of BALB/c mice. Dox-mediated knockdown of Id resulted in modest inhibition of primary tumour growth, with control tumours significantly growing faster and reaching the ethical end point earlier when compared to the Id knockdown group (Figure 2I). More significantly, mice transplanted with Id depleted C8 cells presented far fewer lung metastatic lesions compared to the control despite growing in the host for a longer time ($p<0.0001$; Figure 2J). We have also checked whether there is an enrichment of Id expression in lung metastasis compared to primary tumours in mice injected with C8 cells +/- Dox. An increase in the expression of Id1 was observed in the lung metastasis in all the samples, while no significant enrichment of Id3 expression was observed (Supplementary Figure 3C).

These results were correlated with human clinical data. To determine whether altered expression patterns of ID1 are associated with metastasic progression, ID1 IHC was performed on a cohort of 49 cases with matching primary tumour and brain metastatic lesions surgically removed from breast cancer patients. Amongst the 13 cases in which ID1 was detected by IHC, an enrichment of ID1 expression was observed in brain metastases over the patient-matched primary tumour in 11 cases (Supplementary Figure 3D, Supplementary Table 2). Together with

bioRxiv preprint doi: https://doi.org/10.1101/497313; this version posted December 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

data from the animal model, this result suggests that ID1 promotes metastatic dissemination in a subset of human breast cancers.

Identification of genes and pathways regulated by Id

The canonical role for Id proteins is to regulate gene expression through association with transcription factors. A comprehensive analysis of Id transcriptional targets in cancer has not been reported yet. We performed microarray expression profiling of Id depleted C8 cells. The gene expression profiles of three independent replicates (R1, R2 and R3± doxycycline treatment) were compared by microarray analysis (Supplementary Figure 4A). 6081 differentially expressed genes were identified (Q<0.05), with 3310 up-regulated and 2771 down-regulated genes in Id KD cells (Table 1). Network and pathway enrichment analysis was conducted using the MetaCore^{TM} software. 4301 significant network objects were identified for the Id knockdown microarray data (adjusted p-value of 0.05). The top pathways affected by Id knockdown were mostly associated with the cell cycle (Figure 3A, B) consistent with the loss of proliferative phenotype described previously (Supplementary Figure 2B, Supplementary Figure 3B). Similar results were obtained using Gene Set Enrichment Analysis (GSEA) with significant down regulation of proliferative signatures (CELL_CYCLE_PROCESS) and mitosis (M_PHASE) as seen inTable2. Enrichment for genes involved in several oncogenic pathways such as Mek, Vegf, Myc and Bmi1 signalling have also been highlighted (Table 3). Interestingly, Myc targets were enriched in both the 'up' and 'down' sets. Specifically, genes down regulated by Myc overexpression in breast epithelial cells (MYC_UP.V1_DN) were upregulated by Id knockdown, whilst genes upregulated by Myc (MYC_UP.V1_UP) were down regulated by Id knockdown, consistent with a role of Id in promoting Myc transcriptional activity (45).

bioRxiv preprint doi: https://doi.org/10.1101/497313; this version posted December 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

In addition, in order to identify whether Id specifically regulate genes controlling breast cancer metastasis as phenotypically demonstrated in Figure 2J, GSEA analysis was performed with a collection of custom "metastasis gene sets". This collection (Table 4) consists of several metastatic signatures from the C2 collection (MSigDB database; Table 5), combined with a group of custom gene sets described in major studies (46-55) as shown in Figure 3C. Genes differentially expressed in this set included *Robo1 (56-57),, Tnc (52), Postn (58), Il6 (59) Fermt1 (58)* and *Foxc2 (59)*. Three putative Id1 targets, *Robo1, Tnc* and *Postn* were then validated using q-RT PCR (Figure 3D, Table 6) and found to be up regulated in the C8 cell line upon Id KD.

Id mediated inhibition of *Robo1* controls the proliferative phenotype via activation of Myc transcription

We next sought to determine which targets of Id have an epistatic interaction with Id loss of function using a targeted siRNA screen followed by proliferation assay. 63 predefined target candidates of Id1 along with 10 controls were knocked down using siRNA libraries in C8 cells with and without dox-induced Id knockdown. These genes included negative effectors of canonical Wnt signalling pathway (*Sfrp2*, *Axin2*) and *Robo1*. Figure 4A shows the relative proliferation of C8 cells following transfection with candidate siRNAs in the presence or absence of dox-induced Id knockdown. The X-axis shows the relative proliferation of siRNA KD cells normalized to that of the control siRNA. The Y-axis shows the relative proliferation of siRNA KD cells normalised to the impact of Id KD alone. Genes above or below the diagonal present a possible genetic interaction with Id. Of interest, four genes had minimal impact alone (X axis) but rescued the proliferative defect caused by Id knockdown (Y axis). These included *Cdkn1a*, encoding the p21 CDK inhibitor, *Chad*, a putative tumour suppressor in HCC (60)

bioRxiv preprint doi: https://doi.org/10.1101/497313; this version posted December 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

and *Robo1*. Knockdown of *Robo1* using multiple siRNAs removed the requirement for Id and most effectively restored the proliferation in Id KD cells (Figure 4A, B).

To understand the mechanisms by which *Robo1* rescues cell proliferation in Id depleted cells *in vitro*, we performed RNA-Sequencing (RNA-Seq) experiments on C8 cells with dox-inducible Id KD and/or Robo1 depletion using siRNA. Four replicates per condition were generated; PCA plots presented in Supplementary figure 4B showed that the replicates cluster together. Id KD alone in the C8 cells down regulated 4409 genes and up regulated 5236 genes (FDR<0.05), respectively. Approximately 30% of the differentially expressed genes determined by RNA-Seq were commonly found by microarray analysis (Supplementary Figure 4C). Id depletion led to an increase in *Robo1* expression, as observed in the previous microarray experiment (Figure 3C, D; Figure 4 C, D). Importantly, under Id depletion conditions, *Robo1* KD restored expression of a large subset (~45%) of Id target genes to basal levels (Figure 4E). In comparison, knockdown of Id or Robo1 regulated few targets in the same direction (e.g. both up or both down). This implies that that large proportion of Id targets may be regulated via suppression of Robo1.

Genes whose expression was repressed by Id1 KD and rescued by concomitant Robo1 KD were termed 'Intersect 1' (Fig 4E, Table 7), and genes inhibited by Id and activated by Robo1 (in the absence of Id) were annotated 'Intersect 2' (Fig 4E, Table 8). Interestingly, we found that Robo1 targets represented a subset of Id targets, as Id also regulated over half of Robo1 targets. To investigate the function of these intersect group of genes, we performed GSEA analysis using the MSigDB hallmark gene set. The top signatures in Intersect 1 were all involved in cell proliferation, with enrichment for G2M checkpoint, E2F and Myc targets as well as mTOR signalling (Table 7). Rank-based analysis revealed strong enrichment for the hallmark Myc targets signature upon Id knockdown alone or in comparison to Robo1 knockdown (Figure 4F). This suggests that following Id KD, Robo1 is induced and exerts anti-

bioRxiv preprint doi: https://doi.org/10.1101/497313; this version posted December 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

proliferative effects *via* suppression of Myc and its target genes. Transcription factor motif analysis using EnrichR revealed that Myc and its binding partner Max, have a high combined score in the Intersect 1 gene list further implicating Myc as downstream effector of Robo1 and Id (Supplementary Figure 4D, F).

In order to determine the interaction between ID and ROBO1 in human TNBCs, 82 publically available TNBC datasets from The Cancer Genome Atlas (TCGA, (61)) were queried for the mRNA expressions of ID1, ID3, and ROBO1 and the expression heats maps were generated using cBioportal (62-63). Consistent with our results, a negative interaction between ID and ROBO1 was observed in the majority of the samples (Figure 4G).

In summary, we have demonstrated that Id depletion leads to a loss in the proliferative and self-renewal cancer stem cell phenotypes associated with TNBC. Id1 acts by negatively regulating Robo1 which in turn leads to the activation of a Myc transcriptional program.

bioRxiv preprint doi: https://doi.org/10.1101/497313; this version posted December 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Discussion

There is increasing evidence that all cells within a tumour are not equal with some cells having the plasticity to adapt and subvert cellular and molecular mechanisms to be more tumorigenic than others. Transcription factors like the Id family of proteins can affect a number of key molecular pathways, allowing switching of phenotypes and are known to be master regulators in a CSC. In this study, we conclusively demonstrate that Id1 and its closely related family member Id3 are critical for the CSC phenotype in the TNBC subtype. Clinical data shows that Id expression is elevated in Her2+ and TNBC cases. The lack of correlation between Id1 and Id3 expression in TNBC suggests that though Id1 and Id3 may be redundant, their function in TNBC is not necessarily correlated.

We present a novel role for Id in breast CSC biology. The CSC phenotype as marked by Id is plastic, fitting with the latest evidence that CSC are not necessarily be hierarchically organised, but rather represent a transient inducible state dependent on the local microenvironment. Id proteins are highly sensitive to local cues such as transforming growth factor-β (TGF-β) (13, 64), receptor tyrosine kinase signalling (65)and steroid hormones (66) and therefore are able to transduce a multitude of cues into competency for proliferation and self-renewal. Interestingly, in a neurological context, a continuum of Id1 expression is associated with different states of differentiation of the neuronal cell lineage. This seems to be conserved in a corrupted manner even in glioblastomas. Unpublished data from our group shows that high Id1 expression in cells isolated from the normal mammary gland have high self-renewal capacity (in the mammosphere assay). It would be interesting to look into whether the same pattern of corruption of Id1 expressing cells is preserved in the transition of mammary cells to breast tumours.

bioRxiv preprint doi: https://doi.org/10.1101/497313; this version posted December 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

In order to understand the transcriptional network driving the ability of Id1 as a transcription factor to control the plasticity in TNBC cells, we report the first comprehensive analysis of Id transcriptional targets. We found known Id1 targets like p21 (cdkn1a) in our screen. Our data is in concordance with the ID1- mediated repression of p21 affecting self-renewal in endothelial cells supporting the hypothesis that ID1KD resulted in increased p21 levels and a subsequent decrease in self-renewal capacity. In the context of mammary tumors too, ID1 over expression is associated with highly proliferative tumors that can overcome high p21 expression (19). This is in contrast to findings in colorectal cancer where ID kd resulted in loss of p21 and self renewal (ID1 and ID3 regulate the self-renewal capacity of human colon cancer (18)). These data strongly suggest that the function of ID through p21 is dependent on the context of the cell type.

We identify a novel epistatic relationship with Robo1, with Robo1 loss sufficient to remove the necessity for Id in proliferation, suggesting that suppression of Robo1 is a central function for Id in this setting. Robo1 is receptor for SLIT1 and SLIT2 that mediates cellular responses to molecular guidance cues in cellular migration. Previous work with mammary stem cells showed that the extracellular cue SLIT2 signals ROBO1 to regulate the asymmetric self-renewal of basal stem cells during mammary gland development by controlling the expression through the transcription factor Snail . This finding may have implications for tumour biology because SLIT/ROBO signaling is altered in 40.7% of basal breast tumours.

Interestingly, Minn and colleagues had earlier reported a lung metastatic signature which included ID1 (52). In the screen, Robo1 was also initially identified as a significantly up regulated gene in cancer cells with a propensity to migrate to the lung from breast. Depleting cells of Id1 alone was sufficient to ablate the ability of lung metastatic potential while Robo1 knock down did not have the same effect. Combined with our data, this supports the role of Id1 being able to regulate Robo1 through feedback loop mechanisms.

bioRxiv preprint doi: https://doi.org/10.1101/497313; this version posted December 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Activation of the PI3K/Akt signaling pathway results in tumorigenesis and metastasis. In most of the human cancers, Akt-mediated β-catenin nuclear translocation increased transcriptional activity of oncogenes such as c-Myc, Cyclin D that promotes tumor development. Existing data shows that elevated levels of Id1 promotes tumorigenesis, mostly through Akt –related pathway. Our observation showed that Id1 depletion leads to activation of Robo1. Studies also reported Slit/ ROBO1 can suppresses Akt phosphorylation and indirectly blocks β-catenin translocation from cytoplasm to nucleus. The role of the microenvironment in halting tumor cell proliferation using the Robo1pathway has also been studied. Slit2 from the stromal fibroblasts leads to repression of the PI3K/ Akt pathway via activation of Robo1 in the tumor cells. Based on this work, we hypothesised that Id1/Robo1 promotes tumorigenesis by activating the Akt-βcatenin pathway. But we did not find evidence of the repression of the Akt pathway via Robo1 in our model system. There could be two explanations for this contrary finding- one the activity of the Id1/Robo1 pathway is different in the TNBC scenario. Alternatively, the activation of this pathway requires cognate Slit2 signalling from the microenvironment, which we are currently investigating. Studies also showed that SLIT2 restricts stem cell renewal by signaling through ROBO2 in a subset of basal cells to negatively regulate WNT signalling . The absence of SLIT/ROBO2 signalling leads to increased levels of nuclear b-catenin which results enhanced self-renewal. We did not see any evidence of this pathway in our data strongly suggesting that Id1/ Robo1 acts on the CSC phenotypes via a different pathway.

The significant decrease in the Myc levels on Id1 knock down strongly implicate an Id1/ Robo1/ Myc axis and suggest a novel role for Robo1 in regulating Myc directly as opposed to through the PI3K/Akt or Wnt pathways. Id1 and Robo1 clearly regulate Myc as shown by previous studies and our work. The ability to regulate expression of 15% of genes via Myc

bioRxiv preprint doi: https://doi.org/10.1101/497313; this version posted December 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

would allow the Id1/Robo1 axis to exert enormous control over genes essential for tumor proliferation, self renewal and metastatic dissemination.

Our data provides further evidence that Robo1 is an important tumor suppressor gene in TNBC. Prior data on association of high Robo1 expression with good outcome in breast cancer, whereas low Slit2 expression correlated with increased metastasis, is consistent with our finding. There has been huge interest in targeting Myc and Id1, but until now has been very challenging. We show that Id1 is able to reprogram Myc activity via Robo1 thus providing an alternative strategy to target these transcriptional pathways. An interesting approach has involved the reversal of promoter hypermethylation by HDAC inhibitors that would eventually restore Robo1 expression. These strategies will have to be extensively tested in model systems first but still hold promise as a new therapeutic approach of targeting the ID1/Robo1/ Myc pathway in TNBC.

bioRxiv preprint doi: https://doi.org/10.1101/497313; this version posted December 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Material and Methods

Plasmids

pEN_TmiRc3 parental entry plasmid, pSLIK-Venus and pSLIK-Neo destination vectors were obtained from the ATCC (Manassas, VA,USA).

Cell cultur 4T1 and HEK293T cells were cultured as per obtained from the American Type Culture Collection (ATCC guidelines.). 4T1 cells were maintained in RPMI 1640 (Gibco, Grand Island, NY, USA) supplemented with 10% (v/v) FBS (Thermo Fisher Scientific, Scoresby, Vic, Australia), 20mM HEPES (Gibco, Grand Island, NY, USA), 1mM sodium pyruvate (Gibco, Grand Island, NY, USA), and 0.25% (v/v) glucose. HEK293T cells were grown in DMEM (Gibco, Grand Island, NY, USA) supplemented with 10% (v/v) FBS (Thermo Fisher Scientific, Scoresby, Vic, Australia), 6mM L-glutamine (Gibco, Grand Island, NY, USA), 1mM sodium pyruvate (Gibco, Grand Island, NY, USA) and 1% (v/v) MEM Non-essential Amino Acids (Gibco, Grand Island, NY, USA). All cell lines were cultured at 37°C in a humidified incubatorwith 5% CO2..

Animals

All experiments involving animal work were performed in accordance with the rules and regulations stated by the Garvan Institute Animal Ethics Committee. The BALB/c mice were sourced from the Australian BioResources Ltd. (Moss Vale, NSW, Australia). FVBN mice, p53 null mice, C3-Tag mice were a generous gift from Tyler Jacks, Cambridge, MA. Doxycycline (Dox) food, which contains 700mg Dox/kg, was manufactured by Gordon's Specialty Stock Feed (Yanderra, NSW, Australia) and fed to the mice during studies involving Dox-induced knockdown of Id1/3.

bioRxiv preprint doi: https://doi.org/10.1101/497313; this version posted December 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

p53-/- model

We developed and validated an Id1/GFP molecular reporter construct in which 1.2kb of the Id1 proximal promoter is placed upstream of the GFP cDNA. Cells with active Id1 promoter can be visualized and isolated based on GFP expression by FACS from primary mouse tumours and cell lines. A similar approach has been successfully used to isolate CSCs with activeβ-catenin signalling (32). Using the reporter construct, we typically see between 2-15% of cancers cells are GFP+ by FACS, depending on the clone analysed. We experimentally validated the Id1/GFP system to ensure that GFP expression accurately marks the Id1^{+} cells within the bulk tumour cell population. After transfection of the Id1/GFP reporter into cultured p53-/- tumour cells, both the sorted GFP^{+} and unsorted cells were able to generate new tumours when transplanted into wild-type recipient mice. Tumors were harvested, dissociated into single cells, expanded briefly *in vitro*, and then FACS sorted once more to collect GFP^{+} and GFP+ cell fractions.

Generation of shRNA lentiviral vectors

Single stranded cDNA sequences of mouse Id1 and Id3 shRNAs were purchased from Sigma-Aldrich (Lismore, NSW, Australia). The Id1 shRNA sequence which targets 5'-GGGACCTGCAGCTGGAGCTGAA-3' has been validated earlier (69).The Id3 sequence was adopted from Gupta et al 2007 (15) and targets the sequence 5'-ATGGATGAGCTTCGATCTTAA-3'.shRNA directed against EGFP was used as the control. The shRNA linkers were designed as described earlier (44). The sense and antisense oligonucleotides with BfuAI restriction overhangs were annealed and cloned into the BfuAI restriction siteofpEN_TmiRc3 entry plasmid. pSLIK lentiviral vectors expressing shRNA

bioRxiv preprint doi: https://doi.org/10.1101/497313; this version posted December 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

against Id1 and Id3namely pSLIK-Venus-TmiR-shId1 and pSLIK-Neo-TmiR-shId3,were generated by Gateway recombination between the pEN_TmiR_Id1 or the pEN_TmiR_Id3 entry vector and the pSLIK-Venus or pSLIK-Neo destination vector respectively. Control pSLIK vector expressing shRNA against EGFP (pSLIK-Neo-TmiR-shEGFP) was generated by recombination between the pEN_TmiR_EGFP vector and the pSLIK-Neo vector. The Gateway recombination was performed using the LR reaction according to the manufacturer's protocol (Invitrogen, Mulgrave, Vic, Australia).

Lentivirus production

Lentiviral supernatant was produced by transfecting each lentiviral expression vector along with third-generation lentiviral packaging and pseudotyping plasmids (70) into the packaging cell line HEK293T. Briefly, 1.4 x 10^6 cells were seeded in a 60mm tissue culture dish and grown to 80% confluence. 3µg of expression plasmid was co-transfected with lentiviral packaging and pseudotyping plasmids (2.25µg each of pMDLg/pRRE and pRSV-REV and 1.5µg of pMD2.G), using Lipofectamine 2000 (Invitrogen, Mulgrave, Vic, Australia) according to the manufacturer's protocol. Cell culture medium was replaced after 24hr. The viral supernatant was collected 48hr post transfection and filtered using a 0.45µm filter. The filtered lentiviral supernatant was concentrated 20-fold by using Amicon Ultra-4 filter units (100 kDa NMWL) (Millipore, North Ryde, NSW, Australia).

Lentiviral infection

4T1 cells were plated at a density of 1.0 x 10^5 cells per well in 6-well tissue culture plates and culture medium was replaced after 24hr withmedium containing 8µg/mL of polybrene (Sigma-Aldrich, Lismore, NSW, Australia). Thecells were infected overnight with the concentrated virus at 1:5 dilution. Culture medium was changed 24hr post infection and cells were grown until reaching confluence. Cellstransduced with both pSLIK-Venus-TmiR-Id1 and pSLIK-

bioRxiv preprint doi: https://doi.org/10.1101/497313; this version posted December 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Neo-TmiR-Id3 were sorted on FACS using Venus as a marker followed by selection with neomycin at 400µg/mL for 5 days.Cellstransduced with pSLIK-Neo-TmiR-EGFP were also selected with neomycin.

Analytical flow cytometry for mammary CSC markers on Id1/GFP or Id1C3-Tag tumours

A single cell suspensions of the harvested tumours from p53-/- or Id1C3Tag tumours were obtained by first chopping the tumours into small fragments, followed by enzymatic digestion using the GentleMACS Tissue Dissociator according to the manufacturer's instructions. Cells were blocked with anti-CD16/CD-32 (1:200, BD Biosciences) antibodies in FACS buffer (PBS with Calcium and Magnesium, 2% FBS, 2%, HEPES) for 10 minutes. Cells were then stained with a lineage positive cocktail containing anti-CD31-biotin (1:40, BD Biosciences, Clone 390), anti-CD45-biotin (1:100, BD Biosciences Clone 30-F11) and CD-TER11-biotin (1:80, BD Biosciences, Clone TER119) antibodies for 20 minutes. All antibodies were diluted in FACS buffer. Cells were then incubated with Streptavidin-APC-Cy7 (1:400, BD Biosciences) and the following mammary epithelial cell markers; anti-CD24-PE-Cy7 (1:400, BD Biosciences, Clone M1/69), anti-CD29-Pacific Blue (1:100, Biolegend, Clone HMβ1-1) and anti-CD61-APC (1:100, Invitrogen, Clone HMβ1-1) for a further 20 minutes. The cells were stained with DAPI immediately prior to flow cytometric analysis. BD CompBeads (BD Biosciences) were used as single colour compensation controls for each fluorophore used. Flow cytometry was performed using the LSRII SORP using BD FACS DIVA software. The results were analysed using FlowJo software (Treestar).

bioRxiv preprint doi: https://doi.org/10.1101/497313; this version posted December 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Tumoursphere assay

Cells dissociated from modified 4T1 cells and p53-/- null Id1/GFP, Id1C3-Tag tumors were put into tumorsphere assay as described previously (31).

mRNA and Protein expression analysis

Total RNA from the cells were isolated using Qiagen RNeasy minikit (Qiagen, Doncaster, VIC, Australia) and cDNA was generated using from 500 ng of RNA using the Superscript III firststrand synthesis system (Invitrogen, Mulgrave, VIC, Australia) according to the manufacturer's protocol. Protein lysates were prepared and run as demonstrated before (31).

Tumourigenesis and spontaneous metastasis assays

For orthotopic transplantation, 7.0×10^3cells/10µL (4T1, C8) cells were injected into the fat pad of the 4th mammary gland of six-week old female BALB/c mice. Primary tumour and organs including the lungs, liver, lymph node, spleen, pancreas and brain were harvested and visually examined for metastatic lesions and foci. The lung and brain were also examined under the LEICA MZ16 FA fluorescence microscope (Leica Microsystems, Wetzlar, Germany) to detect and quantify the presence of any metastatic lesions.

***In vivo* and *ex vivo* imaging**

The 4T1 cells were lentivirally modified with the pLV4311-IRES-Thy1.1 vector, a luciferase expressing vector (a kind gift from Dr Brian Rabinovich, The University of Texas M.D. Anderson Cancer Center, Houston, TX, USA). Animals were imaged twice weekly. Briefly, mice were first injected intraperitoneally with 200µL of 30% D-luciferin (Xenogen, Hopkinton, MA, USA) in PBS with calcium and magnesium (Life Technologies, Mulgrave, Vic, Australia) and imaged under anesthesia using the IVIS Imaging System 200 Biophotonic Imager (Xenogen, Alameda, CA, USA). Bioluminescent intensity was analysed and quantified using

bioRxiv preprint doi: https://doi.org/10.1101/497313; this version posted December 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

the Image Math feature in Living Image 3.1 software (Xenogen, Alameda, CA, USA). For *ex vivo* imaging, 200µL of 30% D-luciferin was injected into the mice just before autopsy. Tissues of interest were collected, placed into 6-well tissue culture plates in PBS, and imaged for 1–2min.At ethical endpoint, lungs were harvested and visually examined to detect the presence of metastases and later quantified based on 4T1 GFP fluorescence under a dissecting microscope.

Limiting dilution assay

Single-cell suspensions of FACS sorted Id1/GFP+ or unsorted viable tumour cells were prepared as described previously. Tumour cells were transplanted in appropriate numbers into the fourth mammary fat pad of 8- to 12-week-old FVB/N mice and aged till ethical end point. Extreme limiting dilution analysis71 software was used to calculate the TPF.

siRNA screen to assess proliferation

Reverse transfection of 4T1 cells in 384 well plates was performed with 400 cells and 0.08uL Dharmafect1 per well using a Caliper Zephyr and Biotek EL406 liquid handling robots. Media was change at 24hr post-transfection. Cell titer glo assay was performed using a BMG Clariostar plate reader (luminescence assay). Final data presented is generated from three biological replicates each consisting of two technical replicates. Viability measurements were normalized to the treatment-matched scrambled control after subtracting the blank empty wells.

Id1 Overexpression

Briefly, Id1/3 shRNA-expressing 4T1 clonal cell line (C8) and the Control cells (C1) were transduced with a lentiviral vector (pLenti6.3) encoding a full-length human cDNA of the Id1 gene and 3xHA tag with expression driven by a CMV promoter. Cells were also transduced

bioRxiv preprint doi: https://doi.org/10.1101/497313; this version posted December 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

with a vector expressing LacZ gene as a control. The over expression was confirmed using western blotting.

Immunohistochemistry

Immunohistochemistry analysis was performed as described earlier (31). Briefly, 4µm-thick sections of formalin-fixed, paraffin-embedded (FFPE) tissue blocks were antigen retrieved by heat-induced antigen retrieval and were incubated with primary and secondary antibodies.

Microarray and bioinformatics analysis

Total RNA from the samples were isolated using Qiagen RNeasy minikit (Qiagen, Doncaster, VIC, Australia. cDNA synthesis, probe labelling, hybridization, scanning and data processing were all conducted by the Ramaciotti Centre for Gene Function Analysis (The University of New South Wales). Gene expression profiling was performed using the AffymetrixGeneChip® Gene 1.0 ST Array, a whole-transcript array which covers >28000 coding transcripts and >7000 non-coding long intergenic non-coding transcripts. Data analysis was performed using the Genepattern software package from the Broad Institute. Three different modules, Hierarchical Clustering Viewer, Comparative Marker Selection Viewer and Heatmap Viewer were used to visualize the data. In addition to identifying candidate molecules and pathways of interest, Gene Set Enrichment Analysis (GSEA) (http://www.broadinstitute.org/gsea) was performed using the GSEA Pre-ranked module. Briefly, GSEA compares differentially regulated genes in an expression profiling dataset with curated and experimentally determined sets of genes in the MSigDB database to determine if certain sets of genes are statistically over-represented in the expression profiling data.

bioRxiv preprint doi: https://doi.org/10.1101/497313; this version posted December 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

MTS proliferation assay

Cell viability assay (MTS assay) was carried out using the CellTiter 96 AQueous Cell Proliferation Assay (G5421; Promega, Alexandria, NSW, Australia) according to the manufacturer's recommendation.

Statistical analysis

Statistical analyses of the data were performed using GraphPad Prism 6. All *in vitro* experiments were done in 3 biological replicates each with 2 or more technical replicates. 5-10 mice were used per condition for the *in vivo* experiments. Data represented are means ± standard deviation. Statistical tests used are Unpaired student t-test and two-way-ANOVA. p-values <0.05 were considered statistically significant with *$P< 0.05$, **$P< 0.01$, *** $P< 0.001$, **** $P< 0.0001$.

bioRxiv preprint doi: https://doi.org/10.1101/497313; this version posted December 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

References

1. Chen J, Li Y, Yu TS, McKay RM, Burns DK, Kernie SG, et al. A restricted cell population propagates glioblastoma growth after chemotherapy. Nature. 2012 Aug 23;488(7412):522-6.
2. Lawson DA, Bhakta NR, Kessenbrock K, Prummel KD, Yu Y, Takai K, et al. Single-cell analysis reveals a stem-cell program in human metastatic breast cancer cells. Nature. 2015 Oct 1;526(7571):131-5.
3. Li X, Lewis MT, Huang J, Gutierrez C, Osborne CK, Wu MF, et al. Intrinsic resistance of tumorigenic breast cancer cells to chemotherapy. J Natl Cancer Inst. 2008 May 7;100(9):672-9.
4. Malanchi I, Santamaria-Martinez A, Susanto E, Peng H, Lehr HA, Delaloye JF, et al. Interactions between cancer stem cells and their niche govern metastatic colonization. Nature. 2011 Dec 7;481(7379):85-9.
5. da Silva-Diz V, Lorenzo-Sanz L, Bernat-Peguera A, Lopez-Cerda M, Munoz P. Cancer cell plasticity: Impact on tumor progression and therapy response. Seminars in cancer biology. 2018 Aug 18.
6. Lee G, Hall RR, 3rd, Ahmed AU. Cancer Stem Cells: Cellular Plasticity, Niche, and its Clinical Relevance. J Stem Cell Res Ther. 2016 Oct;6(10).
7. Wahl GM, Spike BT. Cell state plasticity, stem cells, EMT, and the generation of intra-tumoral heterogeneity. NPJ breast cancer. 2017;3:14.
8. Lasorella A, Benezra R, Iavarone A. The ID proteins: master regulators of cancer stem cells and tumour aggressiveness. Nat Rev Cancer. 2014 Feb;14(2):77-91.
9. Aloia L, Gutierrez A, Caballero JM, Di Croce L. Direct interaction between Id1 and Zrf1 controls neural differentiation of embryonic stem cells. EMBO Rep. 2015 Jan;16(1):63-70.
10. Hong SH, Lee JH, Lee JB, Ji J, Bhatia M. ID1 and ID3 represent conserved negative regulators of human embryonic and induced pluripotent stem cell hematopoiesis. J Cell Sci. 2011 May 1;124(Pt 9):1445-52.
11. Nam HS, Benezra R. High levels of Id1 expression define B1 type adult neural stem cells. Cell Stem Cell. 2009 Nov 6;5(5):515-26.
12. Liang YY, Brunicardi FC, Lin X. Smad3 mediates immediate early induction of Id1 by TGF-beta. Cell Res. 2009 Jan;19(1):140-8.
13. Stankic M, Pavlovic S, Chin Y, Brogi E, Padua D, Norton L, et al. TGF-beta-Id1 Signaling Opposes Twist1 and Promotes Metastatic Colonization via a Mesenchymal-to-Epithelial Transition. Cell reports. 2013 Dec 12;5(5):1228-42.
14. Anido J, Saez-Borderias A, Gonzalez-Junca A, Rodon L, Folch G, Carmona MA, et al. TGF-beta Receptor Inhibitors Target the CD44(high)/Id1(high) Glioma-Initiating Cell Population in Human Glioblastoma. Cancer Cell. 2010 Dec 14;18(6):655-68.
15. Gupta GP, Perk J, Acharyya S, de Candia P, Mittal V, Todorova-Manova K, et al. ID genes mediate tumor reinitiation during breast cancer lung metastasis. Proc Natl Acad Sci U S A. 2007 Dec 4;104(49):19506-11.
16. Lyden D, Young AZ, Zagzag D, Yan W, Gerald W, O'Reilly R, et al. Id1 and Id3 are required for neurogenesis, angiogenesis and vascularization of tumour xenografts. Nature. 1999 Oct 14;401(6754):670-7.
17. Niola F, Zhao X, Singh D, Sullivan R, Castano A, Verrico A, et al. Mesenchymal high-grade glioma is maintained by the ID-RAP1 axis. J Clin Invest. 2013 Jan 2;123(1):405-17.

bioRxiv preprint doi: https://doi.org/10.1101/497313; this version posted December 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

18. O'Brien CA, Kreso A, Ryan P, Hermans KG, Gibson L, Wang Y, et al. ID1 and ID3 regulate the self-renewal capacity of human colon cancer-initiating cells through p21. Cancer cell. 2012 Jun 12;21(6):777-92.
19. Swarbrick A, Roy E, Allen T, Bishop JM. Id1 cooperates with oncogenic Ras to induce metastatic mammary carcinoma by subversion of the cellular senescence response. Proceedings of the National Academy of Sciences of the United States of America. 2008 Apr 8;105(14):5402-7.
20. Swarbrick A, Akerfeldt MC, Lee CS, Sergio CM, Caldon CE, Hunter LJ, et al. Regulation of cyclin expression and cell cycle progression in breast epithelial cells by the helix-loop-helix protein Id1. Oncogene. 2005 Jan 13;24(3):381-9.
21. Fong S, Itahana Y, Sumida T, Singh J, Coppe JP, Liu Y, et al. Id-1 as a molecular target in therapy for breast cancer cell invasion and metastasis. Proc Natl Acad Sci U S A. 2003 Nov 11;100(23):13543-8.
22. Gumireddy K, Li A, Gimotty PA, Klein-Szanto AJ, Showe LC, Katsaros D, et al. KLF17 is a negative regulator of epithelial-mesenchymal transition and metastasis in breast cancer. Nat Cell Biol. 2009 Nov;11(11):1297-304.
23. Tobin NP, Sims AH, Lundgren KL, Lehn S, Landberg G. Cyclin D1, Id1 and EMT in breast cancer. BMC Cancer. 2011;11:417.
24. Kim H, Chung H, Kim HJ, Lee JY, Oh MY, Kim Y, et al. Id-1 regulates Bcl-2 and Bax expression through p53 and NF-kappaB in MCF-7 breast cancer cells. Breast Cancer Res Treat. 2008 Nov;112(2):287-96.
25. Qian T, Lee JY, Park JH, Kim HJ, Kong G. Id1 enhances RING1b E3 ubiquitin ligase activity through the Mel-18/Bmi-1 polycomb group complex. Oncogene. 2010 Oct 28;29(43):5818-27.
26. Herschkowitz JI, Simin K, Weigman VJ, Mikaelian I, Usary J, Hu Z, et al. Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. Genome Biol. [Comparative Study
Research Support, N.I.H., Extramural
Research Support, Non-U.S. Gov't]. 2007;8(5):R76.
27. Hochgrafe F, Zhang L, O'Toole SA, Browne BC, Pinese M, Porta Cubas A, et al. Tyrosine phosphorylation profiling reveals the signaling network characteristics of Basal breast cancer cells. Cancer research. 2010 Nov 15;70(22):9391-401.
28. Mellick AS, Plummer PN, Nolan DJ, Gao D, Bambino K, Hahn M, et al. Using the transcription factor inhibitor of DNA binding 1 to selectively target endothelial progenitor cells offers novel strategies to inhibit tumor angiogenesis and growth. Cancer research. 2010 Sep 15;70(18):7273-82.
29. Lee CH, Yu CC, Wang BY, Chang WW. Tumorsphere as an effective in vitro platform for screening anti-cancer stem cell drugs. Oncotarget. 2016 Jan 12;7(2):1215-26.
30. Pastrana E, Silva-Vargas V, Doetsch F. Eyes wide open: a critical review of sphere-formation as an assay for stem cells. Cell Stem Cell. 2011 May 6;8(5):486-98.
31. Nair R, Roden DL, Teo WS, McFarland A, Junankar S, Ye S, et al. c-Myc and Her2 cooperate to drive a stem-like phenotype with poor prognosis in breast cancer. Oncogene. 2014 Jul 24;33(30):3992-4002.
32. Green JE, Shibata MA, Yoshidome K, Liu ML, Jorcyk C, Anver MR, et al. The C3(1)/SV40 T-antigen transgenic mouse model of mammary cancer: ductal epithelial cell targeting with multistage progression to carcinoma. Oncogene. 2000 Feb 21;19(8):1020-7.
33. Pfefferle AD, Herschkowitz JI, Usary J, Harrell JC, Spike BT, Adams JR, et al. Transcriptomic classification of genetically engineered mouse models of breast cancer identifies human subtype counterparts. Genome Biol. 2013 Nov 12;14(11):R125.

bioRxiv preprint doi: https://doi.org/10.1101/497313; this version posted December 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

34. Herschkowitz JI, Zhao W, Zhang M, Usary J, Murrow G, Edwards D, et al. Comparative oncogenomics identifies breast tumors enriched in functional tumor-initiating cells. Proc Natl Acad Sci U S A. 2012 Feb 21;109(8):2778-83.
35. Zhang M, Behbod F, Atkinson RL, Landis MD, Kittrell F, Edwards D, et al. Identification of tumor-initiating cells in a p53-null mouse model of breast cancer. Cancer Res. [Research Support, N.I.H., Extramural
Research Support, Non-U.S. Gov't]. 2008 Jun 15;68(12):4674-82.
36. Vaillant F, Asselin-Labat ML, Shackleton M, Forrest NC, Lindeman GJ, Visvader JE. The mammary progenitor marker CD61/beta3 integrin identifies cancer stem cells in mouse models of mammary tumorigenesis. Cancer Res. 2008 Oct 1;68(19):7711-7.
37. Konrad CV, Murali R, Varghese BA, Nair R. The role of cancer stem cells in tumor heterogeneity and resistance to therapy. Canadian journal of physiology and pharmacology. [Review]. 2017 Jan;95(1):1-15.
38. Yoneda T, Michigami T, Yi B, Williams PJ, Niewolna M, Hiraga T. Actions of bisphosphonate on bone metastasis in animal models of breast carcinoma. Cancer. [Research Support, U.S. Gov't, P.H.S.
Review]. 2000 Jun 15;88(12 Suppl):2979-88.
39. Lelekakis M, Moseley JM, Martin TJ, Hards D, Williams E, Ho P, et al. A novel orthotopic model of breast cancer metastasis to bone. Clin Exp Metastas. 1999 Mar;17(2):163-70.
40. Tao K, Fang M, Alroy J, Sahagian GG. Imagable 4T1 model for the study of late stage breast cancer. BMC cancer. 2008;8:228.
41. Aslakson CJ, Miller FR. Selective events in the metastatic process defined by analysis of the sequential dissemination of subpopulations of a mouse mammary tumor. Cancer research. 1992 Mar 15;52(6):1399-405.
42. Pulaski BA, Ostrand-Rosenberg S. Reduction of established spontaneous mammary carcinoma metastases following immunotherapy with major histocompatibility complex class II and B7.1 cell-based tumor vaccines. Cancer research. 1998 Apr 1;58(7):1486-93.
43. Eckhardt BL, Parker BS, van Laar RK, Restall CM, Natoli AL, Tavaria MD, et al. Genomic analysis of a spontaneous model of breast cancer metastasis to bone reveals a role for the extracellular matrix. Mol Cancer Res. 2005 Jan;3(1):1-13.
44. Shin KJ, Wall EA, Zavzavadjian JR, Santat LA, Liu J, Hwang JI, et al. A single lentiviral vector platform for microRNA-based conditional RNA interference and coordinated transgene expression. Proc Natl Acad Sci U S A. [Research Support, N.I.H., Extramural]. 2006 Sep 12;103(37):13759-64.
45. Nair R, Teo WS, Mittal V, Swarbrick A. ID proteins regulate diverse aspects of cancer progression and provide novel therapeutic opportunities. Molecular therapy : the journal of the American Society of Gene Therapy. 2014 Aug;22(8):1407-15.
46. Aceto N, Bardia A, Miyamoto DT, Donaldson MC, Wittner BS, Spencer JA, et al. Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis. Cell. 2014 Aug 28;158(5):1110-22.
47. Bos PD, Zhang XH, Nadal C, Shu W, Gomis RR, Nguyen DX, et al. Genes that mediate breast cancer metastasis to the brain. Nature. 2009 Jun 18;459(7249):1005-9.
48. Charafe-Jauffret E, Ginestier C, Iovino F, Wicinski J, Cervera N, Finetti P, et al. Breast cancer cell lines contain functional cancer stem cells with metastatic capacity and a distinct molecular signature. Cancer Res. 2009 Feb 15;69(4):1302-13.
49. Dontu G, Al-Hajj M, Abdallah WM, Clarke MF, Wicha MS. Stem cells in normal breast development and breast cancer. Cell proliferation. 2003 Oct;36 Suppl 1:59-72.

bioRxiv preprint doi: https://doi.org/10.1101/497313; this version posted December 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

50. Kang Y, Siegel PM, Shu W, Drobnjak M, Kakonen SM, Cordon-Cardo C, et al. A multigenic program mediating breast cancer metastasis to bone. Cancer cell. 2003 Jun;3(6):537-49.
51. Liu H, Patel MR, Prescher JA, Patsialou A, Qian D, Lin J, et al. Cancer stem cells from human breast tumors are involved in spontaneous metastases in orthotopic mouse models. Proceedings of the National Academy of Sciences of the United States of America. 2010 Oct 19;107(42):18115-20.
52. Minn AJ, Gupta GP, Siegel PM, Bos PD, Shu W, Giri DD, et al. Genes that mediate breast cancer metastasis to lung. Nature. 2005 Jul 28;436(7050):518-24.
53. Minn AJ, Kang Y, Serganova I, Gupta GP, Giri DD, Doubrovin M, et al. Distinct organ-specific metastatic potential of individual breast cancer cells and primary tumors. The Journal of clinical investigation. 2005 Jan;115(1):44-55.
54. Padua D, Zhang XH, Wang Q, Nadal C, Gerald WL, Gomis RR, et al. TGFbeta primes breast tumors for lung metastasis seeding through angiopoietin-like 4. Cell. 2008 Apr 4;133(1):66-77.
55. Tang B, Yoo N, Vu M, Mamura M, Nam JS, Ooshima A, et al. Transforming growth factor-beta can suppress tumorigenesis through effects on the putative cancer stem or early progenitor cell and committed progeny in a breast cancer xenograft model. Cancer research. 2007 Sep 15;67(18):8643-52.
56. Chang PH, Hwang-Verslues WW, Chang YC, Chen CC, Hsiao M, Jeng YM, et al. Activation of Robo1 signaling of breast cancer cells by Slit2 from stromal fibroblast restrains tumorigenesis via blocking PI3K/Akt/beta-catenin pathway. Cancer research. 2012 Sep 15;72(18):4652-61.
57. Qin F, Zhang H, Ma L, Liu X, Dai K, Li W, et al. Low Expression of Slit2 and Robo1 is Associated with Poor Prognosis and Brain-specific Metastasis of Breast Cancer Patients. Scientific reports. 2015 Sep 24;5:14430.
58. Landemaine T, Jackson A, Bellahcene A, Rucci N, Sin S, Abad BM, et al. A six-gene signature predicting breast cancer lung metastasis. Cancer research. 2008 Aug 1;68(15):6092-9.
59. Mani SA, Yang J, Brooks M, Schwaninger G, Zhou A, Miura N, et al. Mesenchyme Forkhead 1 (FOXC2) plays a key role in metastasis and is associated with aggressive basal-like breast cancers. Proceedings of the National Academy of Sciences of the United States of America. 2007 Jun 12;104(24):10069-74.
60. Deng X, Wei W, Huang N, Shi Y, Huang M, Yan Y, et al. Tumor repressor gene chondroadherin oppose migration and proliferation in hepatocellular carcinoma and predicts a good survival. Oncotarget. 2017 Sep 1;8(36):60270-9.
61. Cancer Genome Atlas Research N, Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, et al. The Cancer Genome Atlas Pan-Cancer analysis project. Nature genetics. 2013 Oct;45(10):1113-20.
62. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Science signaling. 2013 Apr 2;6(269):pl1.
63. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer discovery. 2012 May;2(5):401-4.
64. Kang Y, Chen CR, Massague J. A self-enabling TGFbeta response coupled to stress signaling: Smad engages stress response factor ATF3 for Id1 repression in epithelial cells. Molecular cell. 2003 Apr;11(4):915-26.

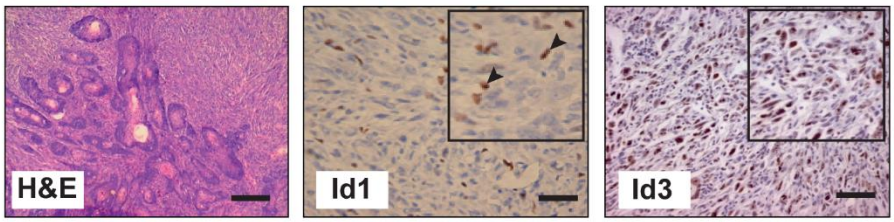
bioRxiv preprint doi: https://doi.org/10.1101/497313; this version posted December 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

65. Tam WF, Gu TL, Chen J, Lee BH, Bullinger L, Frohling S, et al. Id1 is a common downstream target of oncogenic tyrosine kinases in leukemic cells. Blood. 2008 Sep 1;112(5):1981-92.
66. Lin CQ, Singh J, Murata K, Itahana Y, Parrinello S, Liang SH, et al. A role for Id-1 in the aggressive phenotype and steroid hormone response of human breast cancer cells. Cancer research. 2000 Mar 1;60(5):1332-40.
67. Gao D, Nolan DJ, Mellick AS, Bambino K, McDonnell K, Mittal V. Endothelial progenitor cells control the angiogenic switch in mouse lung metastasis. Science. 2008 Jan 11;319(5860):195-8.
68. Dull T, Zufferey R, Kelly M, Mandel RJ, Nguyen M, Trono D, et al. A third-generation lentivirus vector with a conditional packaging system. Journal of virology. 1998 Nov;72(11):8463-71.

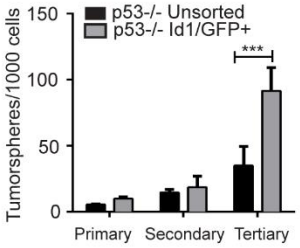
bioRxiv preprint doi: https://doi.org/10.1101/497313; this version posted December 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Figure1

A.



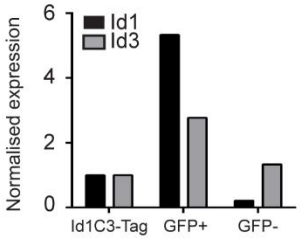
B.



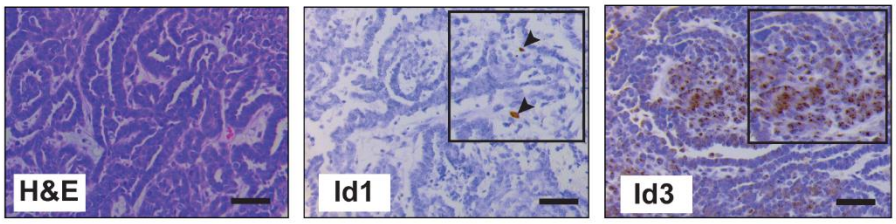
C.

Transplant round	Cell population	No. of cells			Tumour-initiating cell frequenzy
		1000	100	10	(95% confidence interval)
First round (from *in vivo* culture)	Id1/GFP+	9/9	4/9	1/9	1/151(1/64-1/352)
	Unsorted	10/10	2/9	0/8	1/285(1/133-1/607)
Second round (from harvested tumour)	Lin- Id1/GFP+		6/7	1/15	1/42(1/20-1/87)
	Lin- Id1/GFP-		3/10	0/10	1/314(1/102-1/970)
	Lin- Unsorted		4/10	0/10	1/221(1-83-1/588)

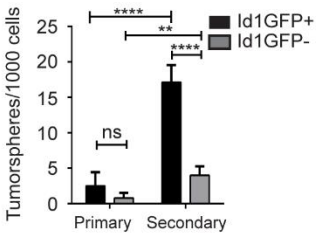
E.



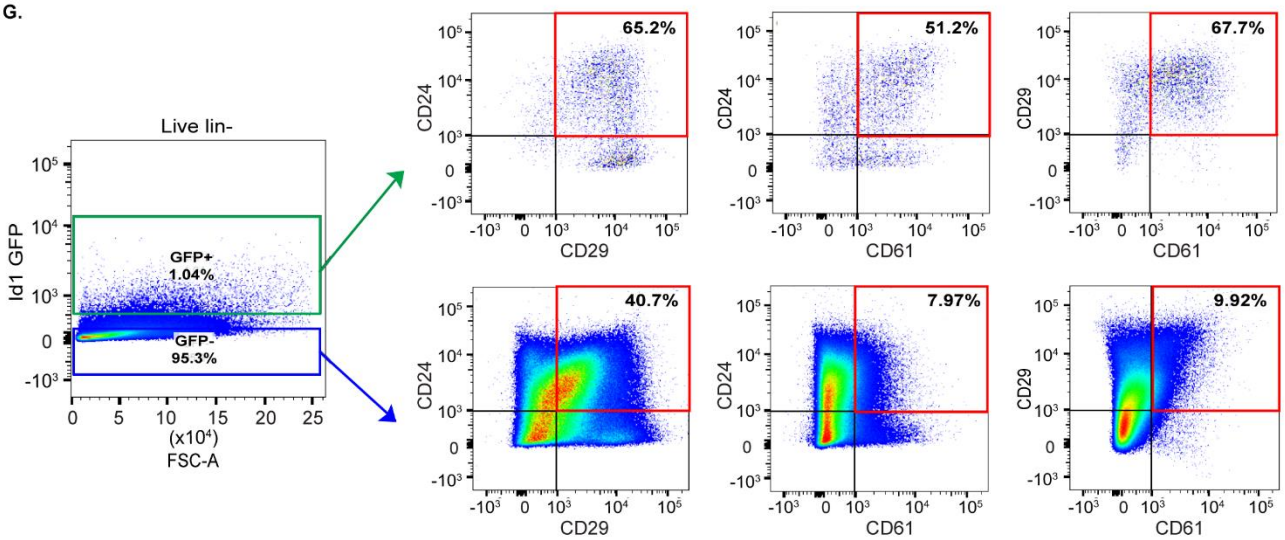
D.



F.



G.



bioRxiv preprint doi: https://doi.org/10.1101/497313; this version posted December 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Figure 1. Id1 marks tumour cells with high self-renewal in murine models of TNBC.

A. Representative IHC images of Id1 and Id3 expression in p53-/- tumour model. Black arrows in the inset indicate rare Id1+ cells. Scale bars = 50 μ m **B.** p53-/- tumour cells were transfected with the Id1/GFP reporter and subsequently sorted for GFP expression. The self-renewal capacity of Id1/GFP$^+$p53-/- cells was significantly higher than Id1/GFP$^-$ p53-/- cells upon passage to tertiary tumourspheres. Data are means± SD (n=3). (*** P< 0.001; Two-way ANOVA).**C.** Id1 expressing cells were sorted from the p53-/- Id1-GFP tumour model and transplanted into recipient mice by limiting dilution assay. Based on limiting dilution calculations (ELDA), the Id1$^+$ cells demonstrated a 7-fold enrichment in tumour initiating capacity (TIC) when compared to the Id1$^-$ cells in serial passage. **D.** Representative IHC images of the Id1C3-Tag model, confirming its suitability as a model system. Black arrows in the inset indicate rare Id1+ cells. Expression of Id1 was less than 5% as determined by IHC. Bars = 50 μ m. **E.** Tumour cells from the Id1C3-Tag tumour model were FACS sorted based on their GFP expression. qRT-PCR analyses on the sorted GFP$^+$ and GFP$^-$ cell populations showed a significant increase for Id1 expression in the GFP$^+$ cells compared to cells lacking GFP expression. **F.** *In vitro* self-renewal capacity was measured using the tumoursphere assay. The secondary sphere forming capacity of Id1$^+$ tumour cells from the Id1C3-Tag model was significantly enriched in comparison to the Id1$^-$tumour cells. Data are means ± SD (n=3). (**P< 0.01, **** P< 0.0001; Two-way ANOVA). **G.** RepresentativeFACS scatterplots and histograms from Id1C3-Tag tumors showing the expression of the CSC markers CD24, CD29, and CD61 in the Id1$^-$/GFP$^-$ and Id1$^+$/GFP$^+$ cancer cells. Putative CSC populations are highlighted within the red box.

bioRxiv preprint doi: https://doi.org/10.1101/497313; this version posted December 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Figure 2

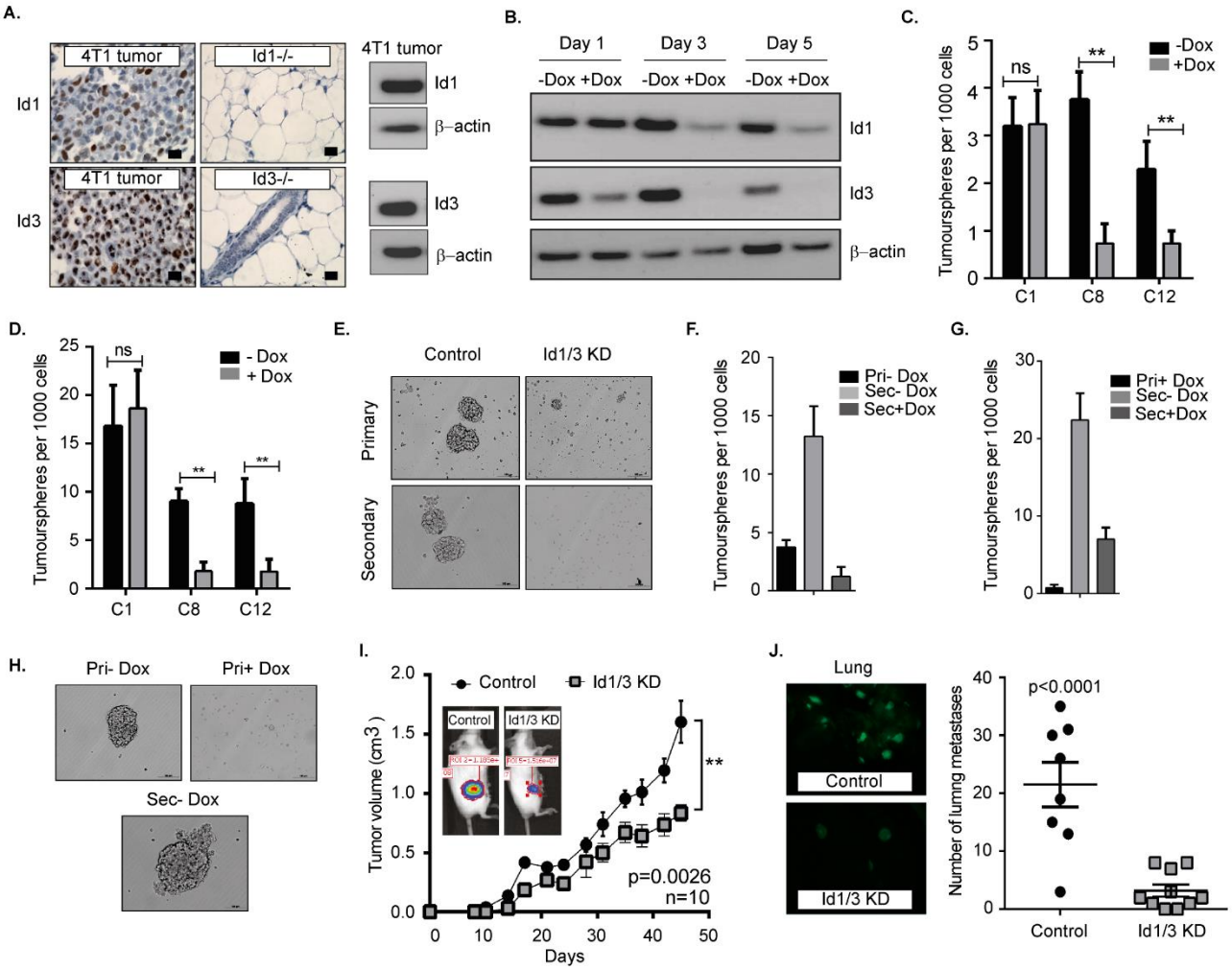


Figure 2. **Depletion of Id1 and Id3 leads to a reduced self-renewal capacity *in vitro* and metastatic potential *in vivo*. A.** Endogenous levels of Id1 and Id3 expression in 4T1primary mammary tumours were determined. 4T1 were cells stained for Id1 and Id3 expression (brown) and counterstained with haematoxylin. Mammary gland tissue from Id1 and Id3 null (Id1-/- and Id3-/-) mice served as negative controls. Scale bars = 50 μm. Western blot analysis of protein lysate from 4T1 tumour cells served as positive controls for Id1and Id3 expression. **B.** Kinetics of conditional Id knockdown in 4T1 cells. Representative Western blot analysis of Id protein levels pSLIK C8 cells over time. Cells were cultured in the presence of 1 μg/ml of

bioRxiv preprint doi: https://doi.org/10.1101/497313; this version posted December 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Doxycycline (Dox) for 1, 3 and 5 days. Protein lysates were subjected to western analysis with an anti-Id1 and anti-Id3 antibody. β-actin was used as loading control. **C.** 4T1 Control, pSLIKC8 and C12 clones were assayed for their tumoursphere forming potential. Dox was added into the culture medium at day 0. Number of primary tumourspheres formed was quantified by visual examination on day 7. Id knockdown leads to a decrease in tumoursphere-forming ability of C8 and C12 cell lines. Data are means ± SD (n=3). (**P< 0.01; Two-way ANOVA).**D.** Primary tumourspheres were then passaged and the number of secondary tumourspheres was quantified on day 14. Knockdown of Id significantly reduces the ability of the C8 and C12 cells to form secondary tumourspheres in the suspension cultureData are means ± SD (n=3). (**P< 0.01; Two-way ANOVA).**E.** Representative images of primary and secondary tumoursphere formation for the clone C8 ±Dox. **F, G.** Primarytumourspheres, previously treated with Dox and depleted of Id expression, were passaged to secondary tumourspheres in the absence of Dox to allow re-expression of Id. Secondary tumourspheres from C8 (C) and C12 (D) cells cultured without Dox re-established their self-renewal potential by resuming the ability to form tumourspheres. Data are means ± SD (n=3). **H.** Representative images of primary tumoursphere formation of the clone C8 control and Id knockdown (±Dox). Dox was removed in the secondary tumoursphere assay. **I.** Knockdown of Id delays tumour growth of the 4T1 syngeneic model. (n = 10 mice; **p value<0.01,Student's t-test). **J.** Id knockdown suppresses spontaneous lung metastasis. Tumours depleted of Id expression generated fewer spontaneous lung macrometastatic lesions compared to the control despite growing in the host for a longer time. Inset shows representative images of lungs bearing the control (C8 - Dox) and Id KD (C8 + Dox) lung metastases at ethical end point. Control; n = 8 mice, Id KD; n=10 mice. Scale bar = 50 um.

bioRxiv preprint doi: https://doi.org/10.1101/497313; this version posted December 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Figure 3

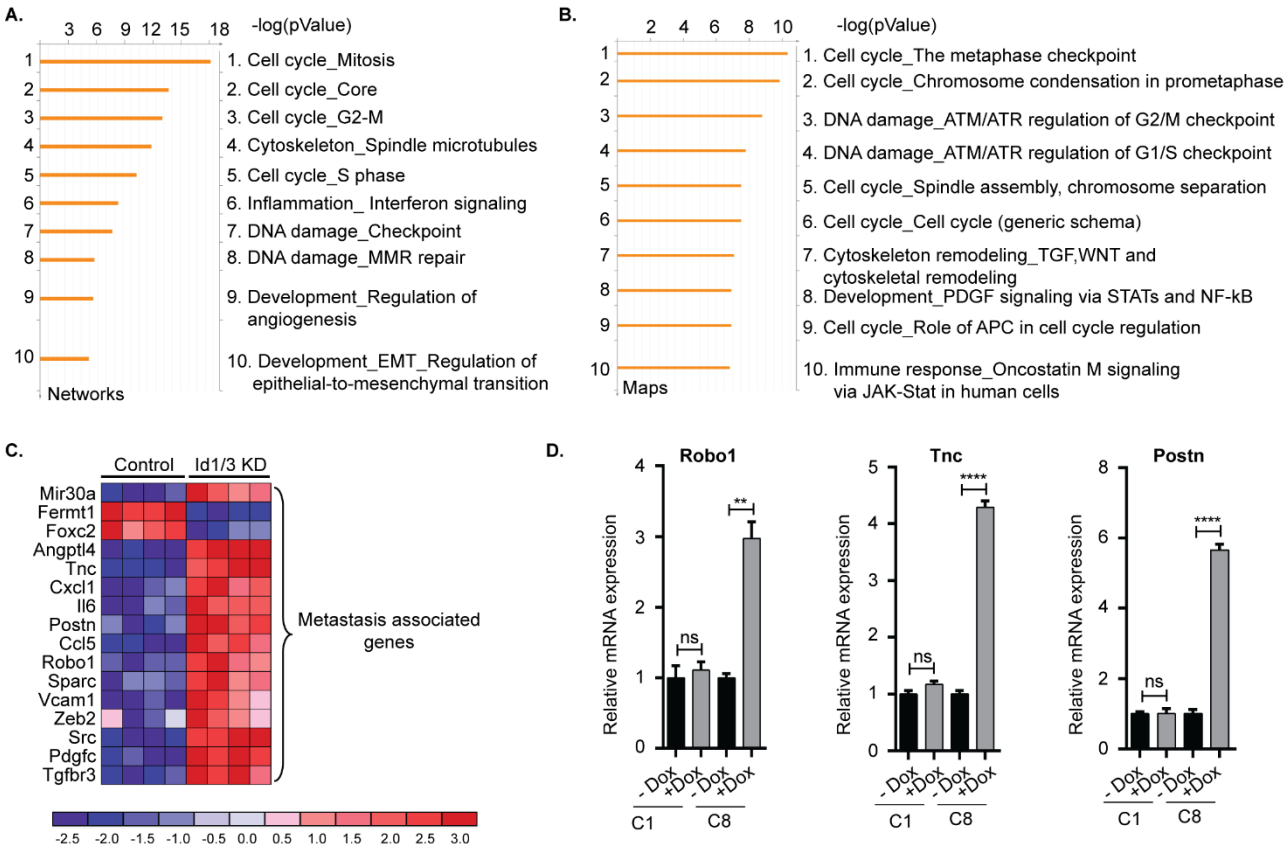


Figure 3. Bioinformatics analysis reveals genes controlling the Id mediated CSC phenotype

A, B. To characterize the network of genes regulated by Id, functional annotation analyses were performed on the gene array data from the 4T1 TNBC model. The Id depletion model attempted to identify downstream targets of Id through a loss of function approach. The gene expression profile of three independent replicates of the C8 shId clone, with and without doxycycline treatment, was compared by microarray analysis. This resulted in a list of differentially expressed genes between control and Id depleted cells, which by further network and map analysis using Metacore demonstrated was largely driven by genes controlling cell cycle pathways.

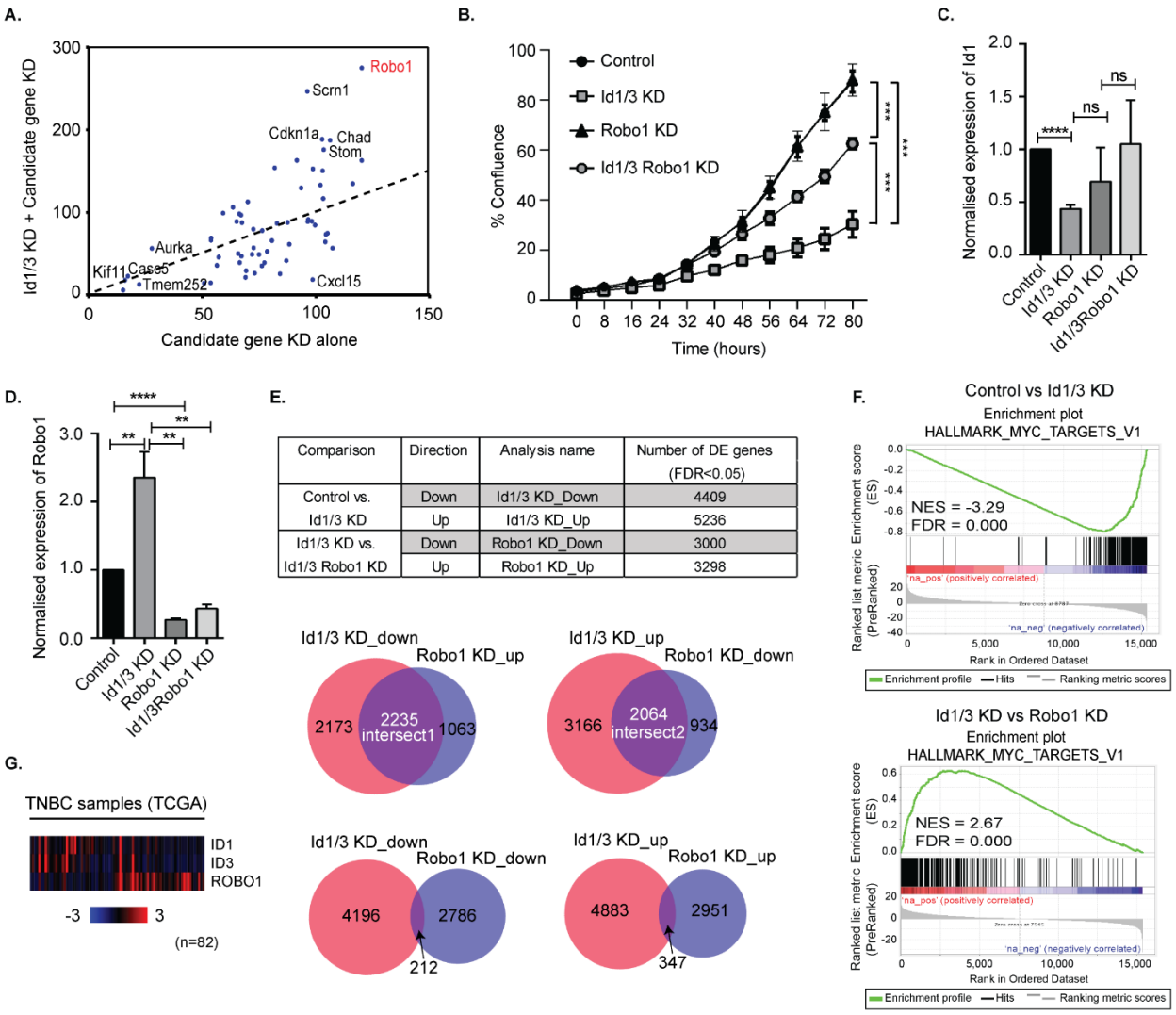
bioRxiv preprint doi: https://doi.org/10.1101/497313; this version posted December 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

C. GSEA analysis identified metastasis-related genes that were differentially expressed in response to Id knockdown. To determine if genes that mediate metastasis were enriched in the Id signature, GSEA analysis was performed using a manually curated set of metastasis gene sets. Genes differentially expressed in response to Id knockdown as well as associated with pathways regulating metastasis were identified based on reports from the literature which included Robo1.

D. Validation of expression profiling results by quantitative real-time-PCR using the Taqman® probe based system. Relative mRNA expression of Robo1, Tnc, and Postn in the 4T1 pSLIK shId Clonal cell line 8 (C8) and pSLIKcontrol (C1), as indicated.Data are means ± SD (n=3).(**P< 0.01, **** P< 0.0001; unpaired t-test).

bioRxiv preprint doi: https://doi.org/10.1101/497313; this version posted December 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Figure 4



Comparison	Direction	Analysis name	Number of DE genes (FDR<0.05)
Control vs. Id1/3 KD	Down	Id1/3 KD_Down	4409
	Up	Id1/3 KD_Up	5236
Id1/3 KD vs. Id1/3 Robo1 KD	Down	Robo1 KD_Down	3000
	Up	Robo1 KD_Up	3298

bioRxiv preprint doi: https://doi.org/10.1101/497313; this version posted December 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Figure 4. Identification of Myc signature activation by Id1 via negative regulation of Robo1

A. Id knockdown (Dox treatment) caused a decrease in cell viability that was rescued by siRNAs against genes like Robo1 and Chad in two independent biological replicates (each containing two technical replicates). A siRNA knockdown screen was performed with putative Idtargetgenes identified. Cell viability at 72h-endpoint following reverse-transfection is displayed on the horizontal X-axes. Cell viability following doxycycline-mediated Id-knockdown is displayed on the vertical Y-axis. All data is normalized to treatment-matched scrambled transfection controls. **B.** Proliferation of C8 cells treated with non-targeting (NT) control siRNA or Robo1 siRNA in the absence or presence of Doxycycline to induce Id knockdown was measured by the live-cell imaging system IncuCyte^{TM} (Essen Instruments). Data shown as mean ±SD (n=3). (*** $p< 0.001$; Unpaired two-tailed t-test). **C, D**. Id1(**C**) and Robo1 (**D**) expression in Control, Id1 KD, Robo1 KD and Id1Robo1 KD cells was measured by quantitative PCR. Ct values were normalised to β actin and GAPDH housekeeping genes. Data shown as mean ± SEM (n=4). (** $p< 0.01$, **** $p < 0.0001$; Unpaired two-tailed t-test). **E**. Transcriptional profiling was performed on Id KD cells or Robo1 KD cells. The Table shows the number of significantly differentially expressed genes (voom/limma FDR<0.05) for the comparisons shown. Proportional Venn diagrams (BioVenn) were generated to visualise the overlapping genes between the different comparisons. **F**. GSEA Enrichment plots of the hallmark Myc targets version 1 signature from MSigDB. NES = normalised enrichment score. **G**. The 82 TNBC samples from The Cancer Genome Atlas (TCGA) were queried for the mRNA expressions of ID1, ID3, and ROBO1 and the heats maps were generated using cBioportal.

bioRxiv preprint doi: https://doi.org/10.1101/497313; this version posted December 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Table 1. Top 25 up and down regulated genes identified in Id-depleted C8 cells ranked based on Q-value. Column 1 shows the gene symbols and annotation. Fold change in column 2 represents the ratio of gene expression change in Id1/3 knockdown versus control. Column 3 represents the Q-value for the difference between knockdown and control group. Column 4 shows the direction of change in gene expression.

Gene	Fold Change	Q-value	Direction
Mx2 :: myxovirus (influenza virus) resistance 2	26.6015	7.59E-08	UP
Oas1g :: 2'-5' oligoadenylatesynthetase 1G	14.9574	1.04E-07	UP
Oas3 :: 2'-5' oligoadenylatesynthetase 3	15.1240	1.95E-07	UP
Cmpk2 :: cytidine monophosphate (UMP-CMP) kinase 2, mitochondrial	24.3300	2.32E-07	UP
Stat1 :: signal transducer and activator of transcription 1	6.8186	3.29E-07	UP
Xaf1 :: XIAP associated factor 1	9.0697	4.06E-07	UP
Usp18 :: ubiquitin specific peptidase 18	30.4890	4.10E-07	UP
Oas2 :: 2'-5' oligoadenylatesynthetase 2	36.1304	4.10E-07	UP
Ifit1 :: interferon-induced protein with tetratricopeptide repeats 1	18.5318	4.58E-07	UP
Gpr56 :: G protein-coupled receptor 56	21.8355	4.72E-07	UP
Zbp1 :: Z-DNA binding protein 1	13.5820	4.72E-07	UP
Olfr65 :: olfactory receptor 65	8.3323	4.72E-07	UP
Parp14 :: poly (ADP-ribose) polymerase family, member 14	5.9386	5.22E-07	UP
Angptl4 :: angiopoietin-like 4	12.1130	5.22E-07	UP
Irf7 :: interferon regulatory factor 7	13.4442	5.22E-07	UP
Gbp3 :: guanylate binding protein 3	10.1348	5.68E-07	UP
Stat2 :: signal transducer and activator of transcription 2	5.0425	6.80E-07	UP
Oasl2 :: 2'-5' oligoadenylatesynthetase-like 2	5.7617	7.46E-07	UP
Bst2 :: bone marrow stromal cell antigen 2	8.4297	7.46E-07	UP
Lypd3 :: Ly6/Plaur domain containing 3	4.2202	7.46E-07	UP
Iigp1 :: interferon inducible GTPase 1	14.3364	7.93E-07	UP
Pvrl1 :: poliovirus receptor-related 1	4.7014	7.93E-07	UP
Oas1b :: 2'-5' oligoadenylatesynthetase 1B	9.7557	8.90E-07	UP
Megf10 :: multiple EGF-like-domains 10	11.0025	1.04E-06	UP
Rtp4 :: receptor transporter protein 4	9.0028	1.04E-06	UP
H19 :: H19 fetal liver mRNA	2.8526	3.68E-06	DOWN
Mpzl2 :: myelin protein zero-like 2	3.1982	3.86E-06	DOWN
Cpox :: coproporphyrinogen oxidase	2.5431	4.32E-06	DOWN
Gpr116 :: G protein-coupled receptor 116	2.9257	5.34E-06	DOWN
Cep78 :: centrosomal protein 78	2.5606	8.09E-06	DOWN
Gpt2 :: glutamic pyruvate transaminase (alanine aminotransferase) 2	3.2592	8.29E-06	DOWN

bioRxiv preprint doi: https://doi.org/10.1101/497313; this version posted December 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Cth :: cystathionase (cystathionine gamma-lyase)	4.9146	1.01E-05	DOWN
Zfp420 :: zinc finger protein 420	2.6201	1.07E-05	DOWN
Gja1 :: gap junction protein, alpha 1	2.2993	1.73E-05	DOWN
Cldn9 :: claudin 9	4.2349	1.82E-05	DOWN
Rtkn2 :: rhotekin 2	3.3638	1.93E-05	DOWN
Aqp1 :: aquaporin 1	3.4829	2.80E-05	DOWN
Fermt1 :: fermitin family homolog 1 (Drosophila)	2.1462	3.05E-05	DOWN
Taf5 :: TAF5 RNA polymerase II, TATA box binding protein (TBP)-associated factor	1.9337	3.62E-05	DOWN
Slc7a11 :: solute carrier family 7 (cationic amino acid transporter, y+ system), member 11	7.3157	3.72E-05	DOWN
B4galnt4 :: beta-1,4-N-acetyl-galactosaminyl transferase 4	2.2432	3.81E-05	DOWN
Gls2 :: glutaminase 2 (liver, mitochondrial)	2.3353	3.93E-05	DOWN
Fbln2 :: fibulin 2	2.0405	4.05E-05	DOWN
2610021K21Rik :: RIKEN cDNA 2610021K21 gene	2.9111	4.51E-05	DOWN
Gstcd :: glutathione S-transferase, C-terminal domain containing	1.8478	0.0000465	DOWN
Rps6ka6 :: ribosomal protein S6 kinase polypeptide 6	2.1709	5.28E-05	DOWN
Hspa1l :: heat shock protein 1-like	1.9999	5.89E-05	DOWN
Dyrk3 :: dual-specificity tyrosine-(Y)-phosphorylation regulated kinase 3	2.8598	6.15E-05	DOWN
Hmgn5 :: high-mobility group nucleosome binding domain 5	3.2189	6.36E-05	DOWN
Deptor :: DEP domain containing MTOR-interacting protein	4.1093	6.37E-05	DOWN

bioRxiv preprint doi: https://doi.org/10.1101/497313; this version posted December 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Table 2. Top 20 gene sets identified from the C5 GO gene sets that are enriched in the Id-knockdown C8 cells.

C5 GO Gene Set	**Normalised Enrichment Score**	**P value**	**Direction**
IMMUNE_RESPONSE	2.136929	<0.0001	UP
G_PROTEIN_COUPLED_RECEPTOR_BINDING	2.1243143	<0.0001	UP
CHEMOKINE_RECEPTOR_BINDING	2.111146	<0.0001	UP
CHEMOKINE_ACTIVITY	2.0914385	<0.0001	UP
I_KAPPAB_KINASE_NF_KAPPAB_CASCADE	2.0558612	<0.0001	UP
DEFENSE_RESPONSE	2.0529644	<0.0001	UP
JAK_STAT_CASCADE	2.0520313	<0.0001	UP
HEMATOPOIETIN_INTERFERON_CLASSD200_DOMAIN_CYTOKINE_RECEPTOR_ACTIVITY	2.0426128	<0.0001	UP
LOCOMOTORY_BEHAVIOR	2.0126882	<0.0001	UP
REGULATION_OF_I_KAPPAB_KINASE_NF_KAPPAB_CASCADE	2.0115595	<0.0001	UP
M_PHASE	-2.5211284	<0.0001	DOWN
CELL_CYCLE_PROCESS	-2.5124967	<0.0001	DOWN
RNA_PROCESSING	-2.4857798	<0.0001	DOWN
CHROMOSOME	-2.3979888	<0.0001	DOWN
M_PHASE_OF_MITOTIC_CELL_CYCLE	-2.3873682	<0.0001	DOWN
MITOSIS	-2.386327	<0.0001	DOWN
CHROMOSOMEPERICENTRIC_REGION	-2.366616	<0.0001	DOWN
SPINDLE	-2.3466449	<0.0001	DOWN
CHROMOSOME_ORGANIZATION_AND_BIOGENESIS	-2.3239574	<0.0001	DOWN
MITOTIC_CELL_CYCLE	-2.3222191	<0.0001	DOWN

bioRxiv preprint doi: https://doi.org/10.1101/497313; this version posted December 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Table 3. Top 20 gene sets identified from the C6 Oncogenic Signatures that are enriched in the Id- knockdown C8 cells

C6 Oncogenic Gene Set	Normalised Enrichment Score	P value	Direction
LTE2_UP.V1_DN	2.389604	<0.0001	UP
MEK_UP.V1_DN	2.244798	<0.0001	UP
VEGF_A_UP.V1_UP	2.124735	<0.0001	UP
WNT_UP.V1_DN	2.070847	<0.0001	UP
PKCA_DN.V1_UP	1.969476	<0.0001	UP
MYC_UP.V1_DN	1.865835	<0.0001	UP
BMI1_DN_MEL18_DN.V1_DN	1.860512	<0.0001	UP
MEL18_DN.V1_DN	1.845022	<0.0001	UP
BMI1_DN.V1_DN	1.83551	<0.0001	UP
STK33_UP	1.819187	<0.0001	UP
RPS14_DN.V1_DN	-2.3749187	<0.0001	DOWN
HOXA9_DN.V1_DN	-2.214685	<0.0001	DOWN
CSR_LATE_UP.V1_UP	-2.1530662	<0.0001	DOWN
PRC2_EZH2_UP.V1_UP	-2.0872	<0.0001	DOWN
VEGF_A_UP.V1_DN	-2.0221975	<0.0001	DOWN
RB_P107_DN.V1_UP	-2.0133445	<0.0001	DOWN
E2F1_UP.V1_UP	-2.0126765	<0.0001	DOWN
MYC_UP.V1_UP	-1.8403969	<0.0001	DOWN
STK33_DN	-1.8023152	<0.0001	DOWN
NFE2L2.V2	-1.7664328	<0.0001	DOWN

bioRxiv preprint doi: https://doi.org/10.1101/497313; this version posted December 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Table 4. Gene expression signatures of breast cancer metastasis and breast cancer stem cells. This table showed a collection of gene sets which comprised several metastatic signatures that were picked from the C2 collection on the MSigDB database and several other signatures that were manually curated. GSEA analysis was carried out to identify whether any of the Id1/3 targets from the profiling experiment are enriched in these signatures.

Study	Signature	Type	Available on GSEA MSigDB database?
Landemaine et al. A six-gene signature predicting breast cancer lung metastasis. Cancer Res. 2008 Aug 1;68(15):6092-9	Lung metastasis signature of breast cancer	Metastatic tissue tropism	Yes
Bild et al. Oncogenic pathway signatures in human cancers as a guide to targeted therapies. Nature 2006, 439:353-357.	Expression profile of 4 individual genes -- Myc, E2F3, Ras, Src, β-catenin	Signalling pathway	Yes
van 't Veer et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature 2002, 415:530-536.	Poor prognosis signature of breast cancer	Classifier that classifies patients as having good or poor prognosis	Yes
Wang et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. Lancet 2005, 365:671-679.	Poor prognosis signature of breast cancer	Classifier	Yes
Ramaswamy et al. A molecular signature of metastasis in	General metastasis	Classifier	Yes

bioRxiv preprint doi: https://doi.org/10.1101/497313; this version posted December 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

primary solid tumors. Nat Genet 2003, 33:49-54.			
Finak et al. Stromal gene expression predicts clinical outcome in breast cancer. Nat Med 2008, 14:518-527.	Breast tumour stromal gene expression signature	Classifier	Yes
Farmer et al. A stroma-related gene signature predicts resistance to neoadjuvant chemotherapy in breast cancer. Nat Med 2009, 15:68-74.	Stromal gene expression signature of breast tumour treated with chemotherapy	Classifier	Yes
Kang et al. A multigenic program mediating breast cancer metastasis to bone. Cancer Cell 2003, 3:537-549.	Bone metastasis signature of breast cancer	Metastatic tissue tropism	No
Minn et al. Genes that mediate breast cancer metastasis to lung. Nature 2005, 436:518-524	Lung metastasis signature of breast cancer	Metastatic tissue tropism	No
Bos et al. Genes that mediate breast cancer metastasis to the brain. Nature 2009, 459:1005-1009.	Bone metastasis signature of breast cancer	Metastatic tissue tropism	No
Padua et al. TGFbeta primes breast tumors for lung metastasis seeding through angiopoietin-like 4. Cell 2008, 133:66-77.	TGF-b signature in lung metastasis of breast cancer	Signalling pathway	No
Aceto et al. Tyrosine phosphatase SHP2 promotes breast cancer progression and maintains tumor-initiating cells via activation of key transcription factors and a positive feedback signaling loop. Nat Med. 2012 Mar 4;18(4):529-37.	Shp2 signature in breast cancer metastasis	Signalling pathway	No

bioRxiv preprint doi: https://doi.org/10.1101/497313; this version posted December 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Minnet al. Distinct organ-specific metastatic potential of individual breast cancer cells and primary tumors. J Clin Invest. 2005 Jan;115(1):44-55.	Poor prognosis signature of breast cancer ; Breast cancer metastasis signature; Bone metastasis signature of breast cancer	Metastatic tissue tropism	No
Tang et al. Transforming growth factor-beta can suppress tumorigenesis through effects on the putative cancer stem or early progenitor cell and committed progeny in a breast cancer xenograft model. Cancer Res. 2007 Sep 15;67(18):8643-52.	TGF-b signature in lung metastasis of breast cancer	Signalling pathway	No
Liu et al. The prognostic role of a gene signature from tumorigenic breast-cancer cells. The New England journal of medicine. 2007. 356(3), 217-26.	Gene signatures of CD44+CD24-/low tumorigenic breast-cancer cell-lines and normal breast epithelium	Cancer stem cell	No
Charafe-Jauffret et al. Breast cancer cell lines contain functional cancer stem cells with metastatic capacity and a distinct molecular signature. 2009. Cancer research, 69(4), 1302-13.	Breast cancer stem cell signature	Cancer stem cell	No

bioRxiv preprint doi: https://doi.org/10.1101/497313; this version posted December 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Dontu, et al. In vitro propagation and transcriptional profiling of human mammary stem/progenitor cells. 2003. Genes & development, 17(10), 1253-70.	Gene signature of human mammary stem and progenitor cells	Cancer stem cell/ Differentiation	No

bioRxiv preprint doi: https://doi.org/10.1101/497313; this version posted December 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Table 5. Top 20 gene sets identified from the C2 curated gene sets that are enriched in the Id-knockdown C8 cells.

C2 Curated Gene Set	**Normalised Enrichment Score**	**P value**	**Direction**
BROWNE_INTERFERON_RESPONSIVE_GENES	2.893714	<0.0001	UP
TAKEDA_TARGETS_OF_NUP98_HOXA9_FUSION_3D_UP	2.889862	<0.0001	UP
ICHIBA_GRAFT_VERSUS_HOST_DISEASE_D7_UP	2.805356	<0.0001	UP
SANA_TNF_SIGNALING_UP	2.760641	<0.0001	UP
SANA_RESPONSE_TO_IFNG_UP	2.652481	<0.0001	UP
DER_IFN_ALPHA_RESPONSE_UP	2.650679	<0.0001	UP
DAUER_STAT3_TARGETS_DN	2.568591	<0.0001	UP
DER_IFN_BETA_RESPONSE_UP	2.544266	<0.0001	UP
MOSERLE_IFNA_RESPONSE	2.528648	<0.0001	UP
RADAEVA_RESPONSE_TO_IFNA1_UP	2.456697	<0.0001	UP
ROSTY_CERVICAL_CANCER_PROLIFERATION_CLUSTER	-3.2496483	<0.0001	DOWN
SOTIRIOU_BREAST_CANCER_GRADE_1_VS_3_UP	-3.1307108	<0.0001	DOWN
WONG_EMBRYONIC_STEM_CELL_CORE	-3.0442135	<0.0001	DOWN
GRAHAM_CML_DIVIDING_VS_NORMAL_QUIESCENT_UP	-2.9951138	<0.0001	DOWN
KOBAYASHI_EGFR_SIGNALING_24HR_DN	-2.950584	<0.0001	DOWN
LEE_EARLY_T_LYMPHOCYTE_UP	-2.9474769	<0.0001	DOWN
PUJANA_BRCA2_PCC_NETWORK	-2.91798	<0.0001	DOWN
LI_WILMS_TUMOR_VS_FETAL_KIDNEY_1_DN	-2.8696544	<0.0001	DOWN
SHEDDEN_LUNG_CANCER_POOR_SURVIVAL_A6	-2.8691382	<0.0001	DOWN
FURUKAWA_DUSP6_TARGETS_PCI35_DN	-2.841147	<0.0001	DOWN

bioRxiv preprint doi: https://doi.org/10.1101/497313; this version posted December 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Table 6. Candidate genes chosen to be validated by quantitative real-time PCR by using either the Lightcycler or the Taqman® probe based system

Gene	Reference	qPCR validation method
EMT and Invasion		
miR30a	(59-61)	Taqman
TGFBR3	(62, 63)	Taqman
FOXC2	(64, 65)	Taqman
ANGPT4	(38, 39, 53)	Taqman
CXCL1	(42, 43)	Taqman
IL6	(43)	Taqman
ZEB2	(8, 66, 67)	Taqman
PDGFC	(68)	Taqman
FERMT1	(69, 70)	Taqman
SPARC	(71)	Taqman
Survival in Circulation or Distant Organs		
IRF7	(72)	Lightcycler
IRF9	(72)	Lightcycler
STAT1	(72)	Lightcycler
STAT2	(72)	Lightcycler
VCAM1	(17, 18)	Taqman
SRC	(73)	Taqman
CXCL1	(42, 43)	Taqman
ROBO1	(74, 75)	Taqman
Metastasis Niches		
TNC	(76)	Taqman

bioRxiv preprint doi: https://doi.org/10.1101/497313; this version posted December 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

POSTN	(14)	Taqman
CCL5	(13)	Taqman

bioRxiv preprint doi: https://doi.org/10.1101/497313; this version posted December 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Table 7. GSEA on the Intersect 1 genes from Figure 4E against the MSigDB hallmark gene sets.

INTERSECT 1 GSEA						
Gene Set Name	**# Genes in Gene Set (K)**	**Description**	**#GenesinOverlap (k)**	**k/K**	**p-value**	**FDR (q-value)**
HALLMARK_E2F_TARGETS	200	Genes encoding cell cycle related targets of E2F transcription factors.	158	0.79	5.02E-178	2.51E-176
HALLMARK_G2M_CHECKPOINT	200	Genes involved in the G2/M checkpoint, as in progression through the cell division cycle.	116	0.58	2.32E-105	5.80E-104
HALLMARK_MYC_TARGETS_V1	200	A subgroup of genes regulated by MYC - version 1 (v1).	113	0.565	7.79E-101	1.30E-99
HALLMARK_OXIDATIVE_PHOSPHORYLATION	200	Genes encoding proteins involved in oxidative phosphorylation.	96	0.48	1.04E-76	1.30E-75
HALLMARK_MYC_TARGETS_V2	58	A subgroup of genes regulated by MYC - version 2 (v2).	42	0.7241	4.60E-45	4.60E-44
HALLMARK_MTORC1_SIGNALING	200	Genes up-regulated through activation of mTORC1 complex.	66	0.33	1.70E-40	1.42E-39
HALLMARK_MITOTIC_SPINDLE	200	Genes important for mitotic spindle assembly.	60	0.3	2.76E-34	1.97E-33
HALLMARK_DNA_REPAIR	150	Genes involved in DNA repair.	51	0.34	2.54E-32	1.59E-31
HALLMARK_UNFOLDED_P	113	Genes up-regulated during unfolded protein	31	0.2743	3.62E-17	2.01E-16

bioRxiv preprint doi: https://doi.org/10.1101/497313; this version posted December 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

ROTEIN_RESPONSE		response, a cellular stress response related to the endoplasmic reticulum.				
HALLMARK_FATTY_ACID_METABOLISM	158	Genes encoding proteins involved in metabolism of fatty acids.	35	0.2215	5.10E-16	2.55E-15
HALLMARK_ADIPOGENESIS	200	Genes up-regulated during adipocyte differentiation (adipogenesis).	39	0.195	1.08E-15	4.89E-15
HALLMARK_CHOLESTEROL_HOMEOSTASIS	74	Genes involved in cholesterol homeostasis.	24	0.3243	1.95E-15	8.11E-15
HALLMARK_ESTROGEN_RESPONSE_LATE	200	Genes defining late response to estrogen.	35	0.175	8.70E-13	3.11E-12
HALLMARK_GLYCOLYSIS	200	Genes encoding proteins involved in glycolysis and gluconeogenesis.	35	0.175	8.70E-13	3.11E-12
HALLMARK_UV_RESPONSE_UP	158	Genes up-regulated in response to ultraviolet (UV) radiation.	28	0.1772	1.18E-10	3.95E-10
HALLMARK_SPERMATOGENESIS	135	Genes up-regulated during production of male gametes (sperm), as in spermatogenesis.	25	0.1852	4.39E-10	1.37E-09
HALLMARK_ANDROGEN_RESPONSE	101	Genes defining response to androgens.	20	0.198	7.15E-09	2.10E-08
HALLMARK_ESTROGEN_RESPONSE_EARLY	200	Genes defining early response to estrogen.	28	0.14	2.76E-08	7.68E-08
HALLMARK_IL2_STAT5_SIGNALING	200	Genes up-regulated by STAT5 in response to IL2 stimulation.	25	0.125	1.31E-06	3.27E-06
HALLMARK_KRAS_SIGNALING_UP	200	Genes up-regulated by KRAS activation.	25	0.125	1.31E-06	3.27E-06

bioRxiv preprint doi: https://doi.org/10.1101/497313; this version posted December 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Table 8 .GSEA on the Intersect 2 genes from Figure 4E against the MSigDB hallmark gene sets.

INTERSECT 2 GSEA						
Gene Set Name	**# Genes in Gene Set (K)**	**Description**	**# Genes in Overlap (k)**	**k/K**	**p-value**	**FDR (q-value)**
HALLMARK_INTERFERON_GAMMA_RESPONSE	200	Genes up-regulated in response to IFNG [GeneID=3458].	60	0.3	3.53E-38	1.76E-36
HALLMARK_INTERFERON_ALPHA_RESPONSE	97	Genes up-regulated in response to alpha interferon proteins.	43	0.4433	4.26E-36	1.06E-34
HALLMARK_HYPOXIA	200	Genes up-regulated in response to low oxygen levels (hypoxia).	39	0.195	5.11E-18	8.51E-17
HALLMARK_P53_PATHWAY	200	Genes involved in p53 pathways and networks.	33	0.165	2.66E-13	3.33E-12
HALLMARK_APOPTOSIS	161	Genes mediating programmed cell death (apoptosis) by activation of caspases.	28	0.1739	4.49E-12	4.49E-11
HALLMARK_ESTROGEN_RESPONSE_EARLY	200	Genes defining early response to estrogen.	31	0.155	7.51E-12	4.70E-11
HALLMARK_HEME_METABOLISM	200	Genes involved in metabolism of heme (a cofactor consisting of iron and porphyrin) and erythroblast differentiation.	31	0.155	7.51E-12	4.70E-11
HALLMARK_MYOGENESIS	200	Genes involved in development of skeletal muscle (myogenesis).	31	0.155	7.51E-12	4.70E-11

bioRxiv preprint doi: https://doi.org/10.1101/497313; this version posted December 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

HALLMARK_PROTEIN_SECRETION	96	Genes involved in protein secretion pathway.	21	0.2188	2.31E-11	1.28E-10
HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION	200	Genes defining epithelial-mesenchymal transition, as in wound healing, fibrosis and metastasis.	30	0.15	3.77E-11	1.89E-10
HALLMARK_ESTROGEN_RESPONSE_LATE	200	Genes defining late response to estrogen.	29	0.145	1.82E-10	8.29E-10
HALLMARK_APICAL_JUNCTION	200	Genes encoding components of apical junction complex.	28	0.14	8.47E-10	3.53E-09
HALLMARK_IL2_STAT5_SIGNALING	200	Genes up-regulated by STAT5 in response to IL2 stimulation.	26	0.13	1.62E-08	5.77E-08
HALLMARK_KRAS_SIGNALING_DN	200	Genes down-regulated by KRAS activation.	26	0.13	1.62E-08	5.77E-08
HALLMARK_ALLOGRAFT_REJECTION	200	Genes up-regulated during transplant rejection.	25	0.125	6.63E-08	2.21E-07
HALLMARK_UNFOLDED_PROTEIN_RESPONSE	113	Genes up-regulated during unfolded protein response, a cellular stress response related to the endoplasmic reticulum.	18	0.1593	1.16E-07	3.62E-07
HALLMARK_ADIPOGENESIS	200	Genes up-regulated during adipocyte differentiation (adipogenesis).	24	0.12	2.60E-07	6.83E-07
HALLMARK_TNFA_SIGNALING_VIA_NFKB	200	Genes regulated by NF-kB in response to TNF [GeneID=7124].	24	0.12	2.60E-07	6.83E-07
HALLMARK_XENOBIOTIC_METABOLISM	200	Genes encoding proteins involved in processing of drugs and other xenobiotics.	24	0.12	2.60E-07	6.83E-07

bioRxiv preprint doi: https://doi.org/10.1101/497313; this version posted December 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

HALLMARK_IL6_JAK_STAT3_SIGNALING	87	Genes up-regulated by IL6 [GeneID=3569] via STAT3 [GeneID=6774], e.g., during acute phase response.	15	0.1724	4.51E-07	1.13E-06

bioRxiv preprint doi: https://doi.org/10.1101/497313; this version posted December 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.