Targeting the Id1-Kif11-Aurka axis in triple negative breast cancer using combination therapy

Reshma Murali[1$,] Binitha Anu Varghese[1$], Nitheesh Karthikeyan[1], Archana PT[1,2], Wee Siang Teo[3], Andrea McFarland [3],Daniel L Roden[3,4],Holly Holliday [3,4],Christina Konrad[3], Aurelie Cazet[3,4], Eoin Dodson[3,4] , Jason T George[5,6] , Herbert Levine[5,7], Mohit Kumar Jolly[8], Alexander Swarbrick[3,4*], Radhika Nair[1*]

[1]Cancer Research Program, Rajiv Gandhi Center for Biotechnology, Kerala 695014, India.

[2]Manipal Academy of Higher Education, Manipal, Karnataka, 576104, India

[3]Cancer Research Division & Kinghorn Cancer Centre, Garvan Institute of Medical Research, Darlinghurst, NSW 2010, Australia

[4]St Vincent's Clinical School, Faculty of Medicine, University of New South Wales, NSW 2052, Australia

[5]Center for Theoretical Biological Physics, Rice University, Houston, TX 77005, USA

[6]Medical Science Training Program, Baylor College of Medicine, Houston, TX, USA

[7]Departments of Bioengineering and Physics, Northeastern University, Boston, MA 02115, USA

[8]Centre for BioSystems Science and Engineering, Indian Institute of Science, Bangalore 560012, India.

[$] Equal first authors

*Corresponding author

Dr Radhika Nair

[$] Present address

Rajiv Gandhi Center for Biotechnology

Kerala, India

radhikanair@rgcb.res.in

Phone: +91-471-2781251

Fax: +91-471-2346333

Running title: Id drives cancer stem cell phenotypes in Triple Negative Breast Cancer by impacting Kif11 and Aurka.

Key words: Cancer stem cells, Chemoresistance, Self renewal, Combination therapy

Abstract

Evidence points to breast cancer following a hierarchical model, with Cancer Stem Cells (CSCs) driving critical phenotypes of the bulk tumor. Chemoresistant CSCs are not an abstract concept but have clinical consequences as they drive relapse and ultimately lead to mortality in patients, making it imperative to understand how these subpopulations of cells survive. Our previous work (1-2) has demonstrated that the bHLH transcription factor, Inhibitor of Differentiation 1 (Id1) and it's closely related family member Id3, have an important role in maintaining the CSC phenotype in the Triple Negative breast cancer (TNBC) subtype. A genetic screen conducted to further elucidate the molecular mechanism underlying the Id (Id1/3) mediated CSC phenotypes in TNBC revealed critical cell cycle genes such as Kif11 and Aurka as putative Id targets. We take this work forward by investigating how alteration in Kif11 and Aurka via Id proteins promotes the CSC phenotype in TNBC. Cells lacking Id are poised in a state of G0/G1 arrest from which they can re-enter the cell cycle. Intriguingly, depletion of Kif11 and Aurka independently did not phenocopy the G0/G1 arrest observed in Id knockdown (Id KD) cells. We have further explored the hypothesis that we can deplete the chemo resistant Id expressing CSC population by combining chemotherapy with targeted therapy using existing small molecule inhibitors (against Id target Kif11) to more effectively debulk the entire tumor. This work opens up exciting new possibilities of targeting Id targets like Kif11, in the TNBC subtype which is currently refractory to chemotherapy.

Introduction

Breast cancer is a heterogeneous disease with different molecular subtypes displaying distinct pathological-clinical outcomes that have been successfully exploited in the management of the disease (3). The TNBC subtype does not express molecular markers such as ER and Her2 that are the basis of targeted therapies in other molecular subtypes of breast cancer (4). Consequently patients presenting with TNBC are left with few therapeutic choices, resulting in lower five-year survival rates when compared to the other subtypes (4-5). There is hence an urgent need to understand the molecular basis of TNBC in order to identify new drug targets.

The critical role of a subpopulation of cells termed Cancer Stem Cells (CSCs) in self- renewal, chemoresistance and metastasis has assumed great clinical importance in breast cancer (6-7). The Inhibitor of differentiation (Id) proteins are negative regulators of the basic helix-loop-helix (bHLH) transcription factors. The Id proteins are important for maintaining the CSC population and therefore tumour progression in TNBC(8) We have previously shown that Id1/3 (collectively known as Id) are critical for the CSC associated phenotypes in the TNBC molecular subtype(1) . A detailed genetic screen analysis of Id knock down (Id KD) and Id1 expression models led to the identification of Kif11 and Aurka as putative Id targets.

The detailed mechanism by which Id controls the cell cycle is not clear, although Id is known to impact the pathway via decreased expression of Cyclins D1 and E, reduced phosphorylation of Rb as well as reduced Cyclin E-Cdk2 activity (9). In this work, we show how Id acts as a central focal point to coordinate the cell cycle genes Kif11 and Aurka and demonstrate that Id KD leads to cell cycle arrest in the G0/G1 phase of the cell cycle. Interestingly, we found that the depletion of Kif11 and Aurka independently did not phenocopy the G0/G1 arrest we observed in Id KD cells. We demonstrate that Id KD puts the brakes on the cell cycle resulting in a state of arrest at

the G0/G1 phase via impacting cell cycle molecules and Id is a critical driver of self-renewal acting via Kif11 and Aurka. We found that the Id expressing tumor cells are resistant to chemotherapy, which forms the first line of treatment in TNBC. Interestingly, treatment with the Ispinesib, a small molecule inhibitor against Kif11 resulted in the reduced expression of Id in these cells. We finally exploit this finding to treat tumor cells with the chemotherapy Paclitaxel combined with Ispinesib to ablate the Id expressing chemo resistant tumor cells along with bulk tumor cells leading to more effective therapeutic targeting in the TNBC subtype.

Results

Id depletion leads to a G0/G1 cell cycle arrest which is reversible

It has been previously demonstrated that Id KD significantly affects pathways associated with cell cycle progression ((10-11). We first sought to validate this observation in our model system using the pSLIK (single lentivector for inducible knockdown) construct(12-13) . We used the metastatic 4T1 cell line as it is representative of the TNBC subtype and Id proteins have been shown to play an important role in tumorigenesis in TNBC subtype. As reported before, we observed a significant decrease in the proliferative capacity of cells upon Doxycycline (Dox) induced Id KD in comparison to control conditions.

As proliferation is inextricably linked to the cell cycle, we next characterized the effect of Id KD on cell cycle progression. We found that Id depletion prevented cells from entering the S phase with accumulation in the G0/G1 phase, as seen in a significant increase in the G0/G1 fraction when compared to the controls (Figure 1A, B, C). To further elucidate the molecular mechanism through which Id controls the cell cycle, we analysed the effect of Id KD on the expression of key cell cycle genes which are vital at different phases of the cell cycle. The down regulation of Id significantly decreased the expression of Ccna2, Ccnb1, Ccnb2, Cdk1 and c-Myc as shown in Figure 1D. Interestingly, it shows an inverse correlation with Rb and p21, which are the negative regulators of these cell cycle genes (Figure1E).

We have already demonstrated that Id KD significantly compromises other key CSC phenotypes like self renewal and migration (10). As Id marks a CSC in the TNBC subtype, we next looked at the expression of the CSC markers CD24 and CD29 in the Id KD system (10, 14-17). Contrary

to our hypothesis, we found that the percentage of cells marked by CD29+/CD24+ in the Id KD population is significantly higher in comparison to the controls (Supplementary Figure1A). This suggests that CD29+/CD24+ does not mark the CSCs in 4T1 model system unlike other model systems (18). Therefore Id depletion clearly affects the key CSC phenotypes such as proliferation, self-renewal and migration which are closely linked to the cell cycle in the TNBC subtype.

Identification of putative Id regulated genes impacting on the cell cycle

To characterize the network of genes regulated by Id proteins, functional annotation analysis was performed on gene array and RNA sequencing data from two different TNBC models of tumour cells marked by either Id depletion or Id1 expression, as described previously (1) (Figure 2A). The Id1 expression model analysed genes whose expression was associated with Id1, whereas the Id depletion model attempted to identify downstream targets of Id proteins.

To study the phenotypes associated with depletion of Id as well as to assess its downstream targets, the gene expression profile of three independent replicates of control and Id KD cells was compared by microarray analysis to generate a list of differentially expressed genes between Id depleted and control cells (Supplementary Table 1). In addition, we used the Id1C3Tag model system to prospectively isolate Id1+ cells as described earlier(1). The gene expression profiles of the Id1+ and Id1- cells from three independent Id1C3Tag tumours were compared by RNA sequencing. This resulted in a list of differentially expressed genes between the Id+ and Id- mouse TNBC cells (Supplementary Table 2). To characterize the network of genes regulated by Id, enrichment analysis was performed on the candidate genes and visualized by process networks and pathway maps. Interestingly, when we looked at the pathway analysis generated

from the differentially expressed gene lists of both models, cell cycle pathways were among the top hits (Supplementary Figure 1B, C)

Aiming to discover high confidence genes involved in the Id gene regulatory network, lists of differentially expressed genes in the Id depleted TNBC model and their controls were compared using MetaCoreTM software. The data was uploaded in MetaCore and filtered using an adjusted p-value threshold of 0.05, resulted in 4301 network objects that were differentially expressed between the Id KD and the control cells (Figure 2A). Similarly the Id1C3Tag RNA Sequencing data revealed 126 network objects differentially expressed between Id1+ and Id1- cells (Figure 2A). By comparing these two datasets, lists of MetaCore network objects common to both experiments as well as those unique to each of the two data sets were generated (Supplementary Figure 1D).When comparing these lists of network objects from the two TNBC models, 34 high confidence MetaCore network objects were identified as common to both the data sets of differently expressed genes. Finally, the genes were mapped to network objects in MetaCore, resulting in a list of 26 genes that were significantly regulated in both models of TNBC (Table 1).

We first analysed the pathways controlled by the 26 putative Id targets. Interestingly, the main pathways regulated by Id involved the cell cycle, specifically the metaphase checkpoint, spindle assembly and chromosome segregation (Figure 2B, C). Strict regulation of cell cycle progression and proliferation is essential for normal cellular function. Disruption of checkpoint control and aberrant regulation of the cell cycle are thus observed in tumourigenesis resulting in uncontrolled cell proliferation. A key function of Id is the stimulation of cell cycle progression and proliferation by controlling the activity of cell cycle regulators. Studies already reported that over expression of Id has been associated with up regulated cell cycle progression in tumourigenesis

(9, 11, 19-20). Also enriched were pathways involving cytoskeleton remodelling, integrin-mediated cell adhesion and migration and chemotaxis, which are all key steps in epithelial to mesenchymal transition (EMT) and metastasis (Figure 2C). Analysis of each individual experiment, along with the genes common to both data sets, showed a similar result with Id depletion affecting mainly the cell cycle pathway, DNA damage, checkpoints, and cytoskeleton remodelling. Id expression model alone showed enrichment for pathways involving hypoxia-induced epithelial-mesenchymal transition, WNT pathway in development, cytoskeleton remodelling and cell cycle (Figure 2C).

We next investigated the process networks that were significantly enriched by the genes common to both the Id1 expressing and Id depleted models. This analysis identified process networks involved in cell cycle, cell adhesion and the cytoskeleton but also regulation of angiogenesis and inflammation. Casc5, Aspm, Aurka, Cenpf, Dynamin, Kif11, Kif4a, and Ube2c were the main drivers of the cell cycle and cytoskeleton networks, whereas the process networks involved in regulation of angiogenesis were enriched for Ctgf, Il6, Angptl4, and Oxtr. Process networks involved in cell adhesion were enriched for Il6, Pdgfc, and Ctgf. In addition, DNA damage checkpoint and mismatch repair (MMR) as well as regulation of EMT were also enriched in the Id depleted model whereas Id1 expression uniquely affected cell-matrix adhesion and interactions (Figure 2C).

We looked at the EMT program which is an important driver of the CSC state and found a significant change in the E cadherin and Vimentin protein levels (Supplementary Figure 2A, B). We next looked at the EMT state using a bioinformatics approach and interestingly found that the EMT scores of these samples using an inferential scoring metric(21) did not show any significant change in the EMT scores, indicating that the EMT status of the cells was not altered

upon Id KD (Supplementary Figure 2C). This clearly suggests as elaborated in the discussion section that the CSC phenotype was being driven by Id using a unique slew of genes.

Identification of Kif11 and Aurka as potential Id targets in disrupting the cell cycle pathway

Among the 26 differentially expressed genes common to the two TNBC models marked by Id depletion or Id1 expression, 16 genes were prioritized for validation. These were chosen based on a significant p value (< 0.05) and at least 1.5 fold (0.58 log2 FC) down or up regulation compared to the controls. Most of these genes had opposite regulation in the two TNBC models which was consistent with the fact that one model was marked by Id depletion whereas the other was an Id1 expression model. In addition, 8 potential cancer stem cell markers and genes previously implicated in Id biology were chosen based on cell surface localization, significant enrichment in Id+ cells and availability of antibodies. Altogether 60 candidate genes were identified for further validation as putative Id candidate target genes (Supplementary Table 3).

As Id functions as a key regulator of cell cycle progression and unlimited proliferation is a key phenotype of cancer cells, the effect of target gene knockdown on the proliferative phenotype of the Id KD cells was assessed for the 61 candidate genes identified using a targeted siRNA screen. Interestingly, the target genes that showed the greatest effect on the viability and thus the proliferative phenotype of the Id KD cells with more than 50% were Kif11, Casc5, Ccnd1 and Aurka (Figure 3A, Supplementary Table 3).

To investigate the effect of putative Id targets such as Kif11, Aurka, Casc5 and Ccnd1 on cell cycle, relative mRNA levels were analysed. We observed significant reduction in the mRNA

levels of Kif11, Aurka, Ccnd1 and Casc5 in Id KD compared to the controls (Figure 3B). We also detected a decrease in the expression of Kif11, Aurka, Ccnd1 and Casc5 at the protein level by western blot (Figure 3C, D). Kif11 and Aurka were also down regulated at the transcript level in spheres generated in the Id KD cells when compared to control (Figure 3E, Supplementary Figure 2D).

To confirm the effect of putative Id targets on proliferation, we used an independent pMission siRNA system in 4T1 parental cells. Loss of Kif11 and Aurka lead to significant decrease in the proliferative capacity of the 4T1 cells (Figure 3F, Supplementary Figure 3A, B). We continued our studies with Kif11 and Aurka as we did not observe any significant loss of proliferative phenotype with Casc5 and Ccnd1 knock down (Supplementary Figure 3C, D).

The effect of Id KD on key molecular cell cycle genes is different from Kif11 and Aurka

As proliferation is inextricably linked to the cell cycle, we next characterized the effect of Kif11 and Aurka knock down using the pMission system on the cell cycle. We found that Kif11 and Aurka depletion lead to cell cycle arrest in the G2/M phase as evidenced by cell cycle analysis (Figure 4 A,B,C). Intriguingly, this observation was fundamentally different from the G0/ G1 arrest that cells undergo when Id is depleted (Figure 1A, Supplementary 3E).

To determine the molecular effect of Kif11 and Aurka on cell cycle, relative mRNA levels of key cell cycle genes were analysed. Kif 11 KD significantly reduced the expression of Aurka even though the expression of Id1 was not altered (Figure 4D). Decreased Kif11 expression had a positive effect on the mRNA levels of Id3, Casc5, Rb and p21. It reduced the expression of Cdk1, but had no significant effect on Ccnd1, c-Myc and p53. Knock down of Aurka did not show any effect on Id1, Id3 and Kif11, but decreased the mRNA level of Ccnd1 and increased

that of Casc5 and p21 (Figure 4E). Altogether these results suggests that Id is having an upstream effect on Kif11 and Aurka is possibly a downstream target of Kif11.It also gives a hint that Id is acting through these target molecules via a unique molecular pathway which is independent of the cell cycle.

Kif11 or Aurka depletion does not phenocopy loss of Id

We next compared our microscopic observations on the phenotype of the Id system with that of depleting the cells of the Id putative targets, Kif11 and Aurka. We noticed monoastral bodies with misformed mitotic spindle, indicating that the majority of the Kif11 KD and Aurka KD cells are arrested in M phase of the cell cycle (Figure 4F, Supplementary Figure 3E)). The formation of monoastral bodies is indicative of duplicated chromatin (4N) not being able to separate due perturbations in the spindle and centrosome, thus indicates G2/M arrest and matches the cell cycle analysis (22) (Figure 4A). Quantification of cells in M-phase (Figure 4G) and cells exhibiting the monoastral body phenotype (Figure 4H) in Control, Id KD, Kif11 KD, and Aurka KD clearly demonstrates that Id depletion results in a phenotype that is distinct from Kif11 and Aurka which points to a unique mechanism controlled by Id.

Therapeutic targeting of CSCs through Id1-Kif11/Aurka axis

There is currently no effective targeted therapy for TNBC and chemotherapy is usually the first line of treatment with a relapse rate of 25% (23). Our previous work has demonstrated that Id is critical for CSC associated phenotypes in TNBC such as proliferation, self-renewal, migration and metastasis(1) .We have now identified that these phenotypes are controlled by the Id-Kif11/Aurka axis. We hypothesized that targeting Kif11 or Aurka to block the Id1-Kif11/Aurka

axis may make the Id expressing CSC more vulnerable to chemotherapy, thus more effectively debulking the entire tumor mass.

To test this hypothesis, we first determined the IC50 values for two commonly used chemotherapy drugs in breast cancer treatment, Paclitaxel and Doxorubicin, in 4T1 cells (Supplementary Figure4A). Interestingly, we find that there is a significant enrichment for Id1+ tumor cells after treatment with Doxorubicin and Paclitaxel, suggesting that the Id1+ CSCs are chemo resistant (Figure 5A). We next determined the IC50 values for the small molecule inhibitors of Kif11 and Aurka, Ispinesib and Alisertib, respectively (Supplementary Figure 4B). Cells treated with Ispinesib showed a significant reduction in the percentage of Id1+ cells but no significant change was observed in those treated with Alisertib as compared to the Control (Figure 5A). We decided to continue with Paclitaxel and Ispinesib based on these results.

We next asked whether the reduction in Id1 levels by Ispinesib can increase the sensitivity of these cells to Paclitaxel in TNBC cells. As proof of principle, we found that the survival fraction and Id1, Kif11 expression in cells treated with a combination of Paclitaxel and Ispinesib was significantly less when compared to cells treated with either Paclitaxel or Ispinesib alone (Figure 5B,C, D). This suggests that the disruption of the Id1-Kif11 axis by Ispinesib sensitizes CSCs to chemotherapy.

We next checked the effect of combination therapy on *in vitro* self-renewal of TNBC cells. A significant reduction in the self-renewal capacity was observed in cells treated with the combination of Paclitaxel and Ispinesib when compared to cells treated with either of the drugs alone (Figure 5 E,F). The expression levels of Id1and Aurka decreased upon combination treatment when compared to the chemotherapy Paclitaxel (Supplementary Figure 4C). We have

also checked the expression of known CSC markers CD24/CD29 in these cells and interestingly found that the CSC fraction was significantly reduced with combination therapy when compared to using Paclitaxel alone (Supplementary Figure 4D).

In conclusion we show that targeting the Id1/ Kif11 molecular pathway in the Id1+ CSCs in combination with chemotherapy results in more effective debulking in TNBC.

Discussion

The current body of work sheds light on the role Id proteins (specifically Id1 and Id3) play in affecting the CSC phenotype of proliferation due to the striking G0/G1 cell cycle arrest we observed when cancer cells were depleted of Id proteins. The Gap phase (G1) is not simply a time delay between the M and S phase. It is a time period within which the cell can monitor the internal and external environment to ensure that the conditions are optimal for the S and M phases. The G1 phase is especially important as the cell can stop in a special G0 resting phase if it finds the conditions are unfavourable for the cell to undergo further cell division. The most distinctive feature of Id depleted cells is the lack of DNA division as also reflected in the G1 block. Combining the phenotype with the change in expression levels of critical cell cycle genes, Id pauses or checks the cells in the G1 state in a manner that they can re-enter the normal cell cycle once the stress is removed. This supports the theory of Id as a master regulator that on sensing unfavourable conditions, can “brake” the cells in the G1 phase through multiple means (molecular regulation of cell cycle genes, DNA division inhibition, protein complex perturbation at the centrosome and spindle fibres). This strategy would allow cells to survive in a state of stasis till conditions favourable to growth of the tumor cell arise. Interestingly, this state could be reversed suggesting that the cells are poised in a state of cell arrest under unfavourable conditions (like chemotherapy treatment) and re enter the cell cycle subsequently.

The idea that CSCs are more plastic and can exist in more than one state may be supported by looking at the EMT program. This makes sense from the point of view of the EMT scores, i.e. the Id KD is not pushing the cells clearly towards a more E or a more M state, as the levels of both canonical markers decrease. Also, our bioinformatic model uses a ratio of CDH1/VIM as a predictor to calculate the scores; so the relatively proportional change that we see at RNA/protein

303 levels for both CDH1 and VIM is consistent. The data adds to the evidence that EMT and MET
304 are not binary (24)for different stages of EMT and their varying degrees of causal contribution to
305 metastasis.
306 Using two independent models of Id1 gene expression and gene depletion, we were intrigued to
307 identify the critical cell cycle genes, Kif11and Aurka as Id putative gene targets. Interestingly
308 previous work in nasopharyngeal cells have linked Id1 and Aurka mechanistically in the
309 induction of tetraploidy. Id1 was found to affect Aurka degradation which normally occurs
310 during exit from mitosis by the APC/C Cdh1 mediated proteolysis pathway. Id1 stabilized Aurka
311 by actively competition with Cdh1, thus preventing Aurka degradation(25). Interestingly, while
312 individual knock down of Kif11 and Aurka also led to a proliferative arrest, it did not phenocopy
313 the G0/G1 cell cycle arrest with the Id knock down or the formation of monoastral bodies. This
314 suggested that the impediment of the cell cycle by Id protein is through different mechanisms
315 and not the canonical mitotic pathways involving the microtubules by Kif11/ Aurka which forms
316 a part of our future investigation.
317 There is no targeted therapy for TNBC and chemotherapy is the first line of treatment. Thus, we
318 checked the effect of commonly used chemotherapy drugs paclitaxel and doxorubicin which are
319 used in the clinic for TNBCs. Studies by our group and others have already reported that Id1
320 marks a chemoresistant breast cancer cell in cancers like Hepatocarcinomas. Interestingly,
321 previous work has targeted the Kif11 pathway in Docetaxel resistant TNBC cells(26). But the
322 most compelling reason to target the Id1/ Kif11 pathway came from work by Chattopadhya et
323 al(27) who identified the drug BRD9876 as a Kinesin-5 inhibitor in Multiple Myeloma which
324 led to significant down regulation of ID1. Based on our work, we used Id as a marker for the
325 chemoresistant CSC population in TNBC. We tested our hypothesis that we can achieve a better

response by combining traditional chemotherapy along with ablation of the Id expressing chemoresistant cells using small molecule inhibitors against the Id targets Kif11. We achieved a significant decrease in proliferative and self renewal capacity when the cells were treated with Paclitaxel and Ispiniseb by successfully targeting sub populations of cells including the Id+ CSCs within a tumor.

Thus a combination of targeted drugs with chemotherapy would be an effective strategy for the complete treatment of TNBC and give women currently living with this disease a better long term prognosis.

Figure legends

Figure 1. Effect of Id knockdown on cell cycle genes.

(A) Flow cytometric analysis of cell cycle on Control and Id KD cells after labelling with Hoechst 33342 stain. (B) Comparing the percentage cell count in each phase of the cell cycle showing a significant increase in the G0/G1 phase in Id KD conditions. (C) Percentage cell count in G0/G1 phase of Control and Id KD cells. (D)The relative mRNA expression level of cell cycle genes Ccna2, Ccnb1,Ccnb2, Ccne1, Cdk1, Cdk2, Cdk4, c-Myc and p53 in Control and Id KD cells quantified using qRT-PCR. (E) Relative mRNA expression of Rb and p21 in Id KD cells with respect to Control cells were quantified using qRT-PCR. Data were normalized to beta-actin and analyzed by the $2^{-\Delta\Delta Ct}$ method. .

Figure 2.Identification of putative ID regulated genes.

(A)To characterize the network of genes regulated by Id proteins, functional annotation analysis was performed on gene array and RNA sequencing data from two different TNBC models of tumour cells marked by either Id depletion or expression. Aiming to discover high confidence genes involved in the Id gene regulatory network, lists of differentially expressed genes between the Id expressing or depleted TNBC models and their controls were compared using MetaCoreTM software. By comparing these two datasets, lists of MetaCore network objects common to both experiments as well as those unique to each of the two data sets was generated The data was uploaded in MetaCore and filtered using an adjusted p-value threshold of 0.05, resulted in 4301 network objects that were differentially expressed between the Id KD and the 4T1 control cells. Similarly the Id1C3Ttg RNA Sequencing data revealed 126 network objects

specific to mouse differentially expressed between Id+ and Id- cells .When comparing these lists of network object for the two TNBC models, 34 high confidence MetaCore network objects were identified as common to both data sets of differently expressed genes. Finally, the genes were mapped to network objects in MetaCore, resulting in a list of 34 genes that were significantly regulated in both models of TNBC. (B) To characterize the network of genes regulated by Id, enrichment analysis was performed on the 34 candidate genes and visualized by process networks and pathway maps. The enriched pathways were involving cytoskeleton remodelling, integrin-mediated cell adhesion and migration, and chemotaxis, which are all key steps in EMT and metastasis. (C) Id expression showed enrichment for pathways involving hypoxia-induced epithelial-mesenchymal transition, WNT pathway in development, cytoskeleton remodelling and cell cycle.

Figure 3. Effect of target gene knockdown on the proliferative phenotype

A) The effect of the candidate genes on proliferation of4T1 cells was asseses by reverse transfection with 40nM siGENOMESMART pool siRNA against each of the 57 candidate genes. Cell viability was quantified at 72h post-transfection using the CellTiter-Glo® luminescent assay. IncuCyteZOOM® live cell imaging every 2 hours was also performed, which allowed us to quantify cell growth (confluence) over time throughout the experiment. Interestingly, the target genes that showed the greatest effect on the viability and thus the proliferative phenotype of the 4T1 cells with more than 50% were Kif11,Ccnd1, Casc5, and Aurka. (B) The relative mRNA expression level of Kif11, Aurka, Ccnd1 and Casc5 in Id KD cells with respect to control cells were quantified with qRT-PCR. Data were normalized to beta-actin and analyzed by the $2^{-\Delta\Delta Ct}$ method. (C) The expression of main target genes Casc5, Kif11, Aurka,Ccnd1 at protein level were analysed in both control and Id KD cells by western

blot.(D) Quantification of relative protein expression of Kif11, Aurka, Ccnd1 and Casc5 in control and Id KD, normalized with b-actin showed a decrease in knock down conditions. (E) Relative mRNA expression level of Kif11, Aurka in primary tumour spheres was assesed. (F) Phase object confluency on Kif11 KD and Aurka KD showing significant decrease in cell proliferation compared to the Controls.

Figure 4.Kif11 and Aurka depletion does not phenocopy loss of Id1

(A) Flow cytometric analysis of cells after knocking down Kif11 (Kif11 KD), Aurka (Aurka KD) along with scrambled control (Control). (B) Comparing the percentage cell count in each phase of the cell cycle after knocking down Kif11 or Aurka. (C) Percentage of cells in G2/M phase in Control, Kif11 KD, Aurka KD. Statistical significance was analyzed by an unpaired t-test. ***p< 0.001, **p< 0.01, *p< 0.05. (D)The relative mRNA expression level of important cell cycle genes were analysed in Kif11 KD and Aurka KD (E) with respect to scrambled control using qRT-PCR. Data were normalized to beta-actin and analyzed by the $2^{-\Delta\Delta Ct}$ method .Statistical significance was analyzed by an unpaired t-test. ***p< 0.001, **p< 0.01, *p< 0.05. (F) shows the Immunofluoresence staining for Id1, Kif11 and Aurka on 4T1 control, Id KD, Kif11 KD andAurka KD cell lines. Representative images are taken using Nicon A1R+ confocal system at 100x magnification with scale bar corresponds to 100um. Dapi shows the nuclear staining, aster(*) shows the monoastral spindle formation in siRNA KD system, Δ shows normal cell division , inset shows the 100x zoomed images of the same. (G) shows the percentage of cells in M-phase in Control, Id KD, Kif11 KD and Aurka KD cell lines. (H) Quantification of cells exhibiting monoastral bodies in Control, Id KD, Kif11KD and Aurka KD conditions.

Figure 5. Therapeutic targeting of CSC through Id1-Kif11/Aurka axis

(A) Cells expressing Id1 after treatment with Paclitaxel, Doxorubicin, Ispinesib and Alisertib compared to the Control. (B) Cell viability was determined using the MTT assay after treating the cells with Paclitaxel, Ispinesib and the combination therapy of Paclitaxel + Ispinesib. (C) Percentage cells expressing Id1 in Control, Paclitaxel, Ispinesib and combination treatment showing a significant decrease in Id1 cells when treated with Ispinesib and combination of chemotherapy and small molecule inhibitor. (D) Percentage cells expressing Kif11 under control, Paclitaxel, Ispinesib and combination treatment. (E) Phase contrast image showing the tumour sphere size under Control, Paclitaxel, Ispinesib and combination treatment (F) There is a significant reduction in sphere forming ability with combination treatment when compared to chemotherapy Paclitaxel alone.

Supplementary Figure 1. Identification of putative ID regulated genes

(A) Cancer stem cell markers CD29/CD24 in Id KD conditions is not significantly changed from controls. (B) The gene expression profile of three independent replicates (with and without doxycycline treatment), was compared by microarray analysis to generate a list of differentially expressed genes between Id depleted and control 4T1 cells. (C) The gene expression profiles of the Id+ and Id- cells from three independent Id1C3Ttg tumours were compared by RNA sequencing. This resulted in a list of differentially expressed genes between the Id+ and Id- mouse TNBC tumour cells. (D) Aiming to discover high confidence genes involved in the Id gene regulatory network, lists of differentially expressed genes between the Id expressing or depleted TNBC models and their controls were compared using MetaCoreTM software. By comparing these two datasets, lists of MetaCore network objects common to both experiments as well as those unique to each of the two data sets was generated. Results are visualized using the enrichment map plug-in for Cytoscape. Each circular node is a gene set with diameter proportional to the number of genes. The outer node color represents the magnitude and direction of enrichment (see scale) in Id1C3Tag cells, inner node color enrichment in Id KD cells. Thickness of the edges (green lines) is proportional to the similarity of gene sets between linked nodes. The most related clusters are placed nearest to each other. The functions of prominent clusters are shown.

Supplementary Figure 2. Effect of Id knockdown on Id targeted genes.

(A) and (B) Relative protein expression level and quantification of E-cadherin, Vimentin in Id KD cells with respect to Control cells were quantified with western blot. β-actin was used as the

loading control. (C) EMT score calculated for all samples, on a scale of 0 (fully epithelial) to 2 (fully mesenchymal). (D) Phase contrast images and quantification of primary spheres under control and Id KD conditions.

Supplementary Figure 3. Id KD leads to cell cycle arrest

(A) Phase contrast images of 4T1 cells under control, Kif11 KD and Aurka KD. (B) A significant decrease in cell viability was determined in Kif11 KD and Aurka KD cells compared to Controls. (C) Phase contrast images and cell viability of 4T1 cells under Control, Ccnd1KD and Casc5 KD conditions.(D) Relative mRNA expression level of Ccnd1 and Casc5 under Ccnd1 and Casc5 KD condition compared to the control. (E) Immunofluorescence images showing Dapi staining of Control, Id KD, Kif11 KD and Aurka KD cells under fluorescence microscope. Representative images are taken using Nikon A1R+ confocal system at 100x magnification with scale bar corresponds to 100um, inset shows the 100x zoomed images of the same.

Supplementary Figure 4. Combination therapy is more effective.

(A) IC50 values in 4T1 cells for two commonly used chemotherapy drugs in breast cancer treatment, Paclitaxel and Doxorubicin. (B) IC50 values for the small molecule inhibitors of Kif11 and Aurka, Ispinesib and Alisertib. (C) Relative protein level expression of Id1, Kif11 and Aurka on Paclitaxel, Ispinesib and Paclitaxel +Ispinesib treated 4T1 cells. (D) Expression of known CSC markers CD24/CD29 on single and combination treated 4T1 cells.

460 ## Materials and Methods

461 ### Mammalian cell culture

462 The cell lines 4T1pSLIK cell lines and Parental cell lines used in this study were obtained from
463 the American Type Culture Collection (ATCC). All cell lines were cultured at 37°C in 5% CO2
464 and 95% air to no more than 80% confluence. Cells were passaged by washing with PBS twice
465 and trypsinised with trypsin-EDTA (0.05%), followed by incubation at 37°C until the cells were
466 detached from the tissue culture flask. An equal or greater volume of culture medium was added
467 to neutralise the trypsin-EDTA. Appropriate volume of cell suspension was then added into a
468 new tissue culture flask for passaging. All cell lines were preserved by cryopreservation. Each
469 cryovial contained 1x106 cells and were frozen in a solution consisting of 50% (v/v) foetal
470 bovine serum (FBS), 40% (v/v) growth media, and 10% (v/v) DMSO to -80°C at a rate of
471 1°C/min for a minimum of 4hr before transferring to liquid nitrogen. Cells were revived by
472 warming individual cryovials to 37°C and seeding into 10mL of culture medium in a T75 tissue
473 culture flask.

474 ### Immunofluoresence

475 Cells were seeded on coverslips in tissue culture dishes and cultured for 2 days. Cells were
476 washed in PBS and fixed with 4% paraformaldehyde. Fixed cells were washed with PBS and
477 resuspended in 0.2% Triton-X (Sigma) in PBS solution for 20 minutes. Cells were blocked using
478 1%BSA in PBS for 1 hr at room temperature and primary antibody was added and incubated at 4
479 degrees overnight. Next day the cells were washed with PBS thriceand secondary antibody was
480 added with Dapi(1mg/ml)and incubated for 1 hr at room temperature. Cover slips were mounted
481 in Prolong gold antifade reagent and visualized under Fluorescence microscope.

Microscopic imaging

Cells on tissue culture plates were magnified with Fluorescence microscope (Olympus, Germany), under both high and low magnification. Confocal images were captured by the Leica DFC280 digital camera system (Leica Microsystems, Wetzlar, Germany).

Cell Cycle Analysis

Cells were harvested and spin down at 1200rpm for 5 minutes. Cells were counted and Hoechst (Sigma) (4ug/mL) was added to the cell suspension(1 million cells) and incubated at 37^0C for 30 minutes. The cell cycle distribution was determined with a flow cytometer (BD Aria III).The data were analysed using the BD FACS analyser software.

RNA extraction and Real-time PCR (RT-PCR)

Total RNA was isolated using TRIzol® Reagent (Invitrogen) and cDNA was synthesized using High-Capacity cDNA Reverse Transcription Kits (Applied Biosystems; Thermo fisher Scientific, Inc.). Real-time PCR was performed on the QuantStudio 7 Flex Real-Time PCR System using Power SYBR Green PCR Master Mix (Applied Biosystems). The PCR conditions were 950C for 10 min, followed by 40 cycles of 950C for 30s, and 600C for 1 min. All reactions were done in triplicates and the transcript levels were normalized to those of b-act. The relative fold change was determined by $2-\Delta\Delta CT$ method as described (PMID: 11846609). The gene specific primers used for RT-PCR are listed in (Table)

CSC markers staining

The cells were collected at 1200rpm for 5 minutes at 4°C(There should be atleast 1 million cells). Pellets were resuspended in FcBlock (Miltenyibiotec, 1:10) and incubated on ice for

10minutes. Cells were pelleted (1200rpm, 5mins, 4°C) and washed by resuspending with cold PBS+salts, then pelleted again. Cells were resuspended in lineage marker cocktail CD29 (Miltenyibiotec, 1:10), CD61 (Miltenyibiotec, 1:10) and incubated on ice for 20 minutes. Cells were pelleted, washed with PBS and pelleted. Cells were resuspended in FACS buffer1:400) and incubated on ice for 20 minutes. Cells were pelleted, washed with PBS+salts, pelleted then resuspended in 200ul FACS buffer(PBS containing salt + 2% FBS + 2% HEPES). Appropriate single stain and unstained controls were performed alongside CSC marker staining.

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

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 titerglo 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.

4T1 cells were reverse transfected with a 40nM siGENOMESMART pool siRNA against each of the 57 candidate genes. Cell viability was quantified at 72h post-transfection using the CellTiter-Glo® luminescent assay (Table 1). IncuCyte ZOOM® live cell imaging every 2 hours was also performed, which allowed us to quantify cell growth (confluence) over time throughout the experiment.

Extracting protein lysates

Protein lysates were obtained by direct lysis from the tissue culture plates. Cells were washed once with chilled PBS (Invitrogen, India) before adding ice cold RIPA lysis buffer supplemented with complete protease inhibitor cocktail solution (Sigma Aldrich, India) to inhibit protein degradation. The cell lysate was then transferred to 1.5mL tubes. All steps above were performed on ice. The cell lysates were then centrifuged at 14000rpm for 10min, 4°C. The supernatants were transferred to new 1.5mL tubes and were stored at -80°C for later use or on ice for immediate use.

Quantifying protein concentration

The protein concentration of each sample was measured by using a BCA assay using the micro bicinchoninic acid (BCA) kit (Thermofisher scientific, India) following the manufacturer's instructions. The assay was performed in a clear- bottomed flat surface 96-well plate. Briefly, the BSA (2mg/ml) was serially diluted in distilled water to generate a dilution range of: 0.0µg/µL to 2µg/µL. Protein lysate was diluted 1 in 10 with distilled water. The BCA reagents were then mixed (50:1 Part A:Part B) and 200ul was added to each well. The plate was then incubated at 37°C for 30min, followed by a measurement of absorbance at 562nm using the TECAN plate reader. The protein concentration of each sample was calculated by using GraphPad prism software.

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

SDS was performed by using the Biorad system (Biorad, India). 20µg of protein from each sample was made up in 1x Lammelli sample buffer and denatured at 95°C for 10min. The denatured protein samples were loaded onto polyacrylamide gels. Gels consisted of a 5% acrylamide stacking gel and a 12% acrylamide gradient separating gel. Electrophoresis was performed for 45min at 140V in 1x SDS running buffer.

Protein transfer and immunoblotting

Following electrophoresis, proteins were transferred onto PVDF membranes (Biorad, India) at 120V for 50min in 1x transfer buffer (Biorad, India). The PVDF membranes were blocked in a solution containing 5% (w/v) skim milk powder and 0.1% TBS- tween for 1hr at room temperature. After blocking, the membranes were washed 3 times in TBS-tween (5min each time). The primary antibodies were diluted in an antibody diluting solution containing 5% (w/v) BSA, 0.025% (w/v) sodium azide in 1% (v/v) TBST. The washed membranes were incubated

with primary antibody solutions at concentration as per Table for 1hr at room temperature or overnight at 4°C. Following primary antibody incubation, the membranes were washed 3 times (10min each time). The secondary antibodies used were anti-rabbit IgG or anti-mouse IgG conjugated with horseradish peroxidise (HRP). Membranes were incubated in secondary antibodies at a concentration of 1:10000 in 5% (w/v) skim milk/TBS- tween buffer for 1hr at room temperature. Excess secondary antibody was washed in TBS-tween four times (15min each time). Specific protein bands were detected by ECL chemiluminesence (Biorad, India).

MTT assay

4 T1 cells were seeded at a density of 500 cells/well in 96 well plate.When the cells become 80% confluency, freshly prepared MTT reagent (5ug/mL) was added to the culture media. Incubate the plate for 3 hrs in dark at 37^0C. Remove the media containing reagent from each well and add 200ul of DMSO to each well. Take the absorbance readings using a microplate reader at 570nm.

IC50 values for Chemotherapeutic drugs

4T1 cells were harvested and seeded 1000 cells/well in 96 well plate. When the cells become 20% confluency, culture media was removed and replenished with media containing chemotherapeutic drugs. MTT reading was taken 48hrs post drug treatment.

Tumoursphere assay

4T1 cells were trypsinised and washed twice with PBS (Invitrogen, India). Cells were then resuspended in RPMI1640 medium without FBS and sieved through a 40μM cell strainer (BD Falcon, India) twice to ensure at least 95-99% of cells were in single cell suspension before being counted on the haemocytometer. Single cells were plated in ultralow attachment 6-well

plates (Corning, India) at a density of 2.0 x 104 viable cells/well in triplicate. Cells were cultured in serum-free RPMI1640 medium, supplemented with B27 (Invitrogen, India) and 20ng/mL bFGF (Millipore, India) and 4µg/mL heparin (Sigma-Aldrich, India). Serum-free media supplemented with the additives mentioned above was added to the cells every 3 days. The plate was tapped very gently to ensure even distribution of the cells. Primary tumourspheres were counted at day 8.

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. 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.

603

604 **Table 1.** List of differentially expressed genes common to both the models.

605

606

		Id1_KD_4T1_metacore		Id1C3Tag_RNASeq_metacore	
#	Input IDs	Signal	p-value	Signal	p-value
1	**Adamtsl3**	-0.3927	0.020463	1.3402333	0.03359618
2	**Casc5**	-1.0215	0.001402	0.9961643	0.002269021
3	**Aspm**	-1.3871	0.000268	0.9586754	0.000409133
4	**Aurka**	-1.1447	0.0005806	0.9813335	0.01574964
5	**Casz1**	0.2437	0.0479447	0.9948092	0.002230733
6	**Cenpf**	-0.9544	0.001877	0.7718833	0.02742762
7	**Ctla2a**	0.4903	0.0023071	1.3066035	0.04998167
8	**Cxcl15**	2.0181	0.0004083	-6.916629	0.01322961
9	**Angptl7**	-0.631	0.0023779	-1.489744	0.002084568
10	**Cldn6**	-0.5278	0.0131683	-7.315199	0.004137659
11	**Gpr133**	0.284	0.0448178	-4.065336	0.000459154
12	**Hmga1;Hmga1-rs1**	-0.4779	0.0332492	-1.080467	0.001750527
13	**Il6**	1.2691	0.0003497	1.2769696	0.004137659
14	**Kif4**	-1.0266	0.0007331	0.7764488	0.01698252
15	**Kif11**	-1.3438	0.0002452	0.7487605	0.02222188
16	**Lphn1**	0.4649	0.0027409	0.8679558	0.04907909

17	**Ltbp2;Ltbp3**	0.2667	0.0244267	0.9104315	0.002134988
18	**Mylk**	0.5396	0.0154085	1.1309435	0.00340821
19	**Lnp;Nusap1**	-1.011	0.0001961	1.0915483	0.008424267
20	**Pdgfc**	1.7513	0.00001065	0.7934828	0.03592082
21	**Angptl2;Angptl4**	3.5985	5.221E-07	-1.456264	1.46715E-05
22	**Prc1**	-0.9945	0.0009206	0.7781182	0.02082709
23	**Stc2**	-1.2088	0.0005806	-2.135721	8.46387E-05
24	**Mylk**	0.5396	0.0154085	1.1309435	0.00340821
25	**Ube2c**	-0.6684	0.0060464	1.2384745	0.004860396
26	**Upp1**	-0.5451	0.0022003	-1.767675	0.000253826

607

608

609 **Supplementary Table 1.** The top 50 DE genes generated from thegene expression profile of
610 three independent replicates of control and Id KD cells was compared by microarray analysis to
611 generate a list of differentially expressed genes between Id depleted and control cells.

612

S. No.	***Gene.Symbol.x***	***Fold change***	***Direction***	***P Value***	***Q Value***
1	Mx2 :: myxovirus (influenza virus) resistance 2	26.6022	up	2.889E-12	7.591E-08
2	Oas1g :: 2'-5' oligoadenylate synthetase 1G	14.9575	up	7.929E-12	1.041E-07
3	Oas3 :: 2'-5' oligoadenylate synthetase 3	15.1238	up	2.227E-11	0.00000 0195
4	Cmpk2 :: cytidine monophosphate (UMP-CMP) kinase 2, mitochondrial	24.3302	up	3.53E-11	2.318E-07
5	Stat1 :: signal transducer and activator of transcription 1	6.8185	up	6.255E-11	3.286E-07
6	Xaf1 :: XIAP associated factor 1	9.0698	up	9.273E-11	0.00000 0406
7	Usp18 :: ubiquitin specific peptidase 18	30.4897	up	1.111E-10	4.104E-07
8	Oas2 :: 2'-5' oligoadenylate synthetase 2	36.13	up	1.25E-10	4.104E-07

9	Ifit1 :: interferon-induced protein with tetratricopeptide repeats 1	18.532	up	1.567 E-10	4.573E-07
10	Gpr56 :: G protein-coupled receptor 56	21.8354	up	1.817 E-10	0.000000472
11	Zbp1 :: Z-DNA binding protein 1	13.5822	up	2.098 E-10	0.000000472
12	Olfr65 :: olfactory receptor 65	8.3324	up	2.156 E-10	0.000000472
13	Parp14 :: poly (ADP-ribose) polymerase family, member 14	5.9386	up	2.691 E-10	5.218E-07
14	Angptl4 :: angiopoietin-like 4	12.1131	up	2.867 E-10	5.218E-07
15	Irf7 :: interferon regulatory factor 7	13.4439	up	2.979 E-10	5.218E-07
16	Gbp3 :: guanylate binding protein 3	10.1348	up	3.456 E-10	5.674E-07
17	Stat2 :: signal transducer and activator of transcription 2	5.0425	up	4.398 E-10	6.796E-07
18	Oasl2 :: 2'-5' oligoadenylate synthetase-like 2	5.7616	up	5.125 E-10	7.454E-07
19	Bst2 :: bone marrow stromal cell antigen 2	8.43	up	5.396 E-10	7.454E-07
20	Lypd3 :: Ly6/Plaur domain containing 3	4.2202	up	5.675	7.454E-

				E-10	07
21	Iigp1 :: interferon inducible GTPase 1	14.3363	up	6.404 E-10	7.932E-07
22	Pvrl1 :: poliovirus receptor-related 1	4.7013	up	6.643 E-10	7.932E-07
23	Oas1b :: 2'-5' oligoadenylate synthetase 1B	9.7559	up	7.787 E-10	8.894E-07
24	Megf10 :: multiple EGF-like-domains 10	11.0028	up	9.708 E-10	0.00000 1044
25	Rtp4 :: receptor transporter protein 4	9.0025	up	9.932 E-10	0.00000 1044
26	Dhx58 :: DEXH (Asp-Glu-X-His) box polypeptide 58	9.4499	up	1.133 E-09	0.00000 1145
27	Irgm1 :: immunity-related GTPase family M member 1	3.8284	up	1.624 E-09	0.00000 158
28	Fst :: follistatin	6.2073	up	1.906 E-09	0.00000 1788
29	Twf2 :: twinfilin, actin-binding protein, homolog 2 (D rosophila)	3.1538	up	2.147 E-09	0.00000 1931
30	Ifih1 :: interferon induced with helicase C domain 1	4.3871	up	2.215 E-09	0.00000 1931
31	Gbp7 :: guanylate binding protein 7	7.9921	up	2.279	0.00000

				E-09	1931
32	17549062	3.6692	up	2.363 E-09	0.00000 1939
33	BC006779 :: cDNA sequence BC006779	4.0514	up	2.523 E-09	0.00000 2009
34	Eif2ak2 :: eukaryotic translation initiation factor 2-alpha kinase 2	3.4611	up	2.674 E-09	0.00000 2066
35	Oas1a :: 2'-5' oligoadenylate synthetase 1A	8.1417	up	2.816 E-09	0.00000 2114
36	Shf :: Src homology 2 domain containing F	2.7862	up	2.911 E-09	0.00000 2124
37	Ecscr :: endothelial cell surface expressed chemotaxis and apoptosis regulator	4.0889	up	2.994 E-09	0.00000 2126
38	Ifi44 :: interferon-induced protein 44	36.7437	up	3.392 E-09	0.00000 2345
39	Gbp9 :: guanylate-binding protein 9	6.043	up	4.69E-09	0.00000 3159
40	Lcp1 :: lymphocyte cytosolic protein 1	3.8831	up	4.994 E-09	0.00000 328
41	Gstm5 :: glutathione S-transferase, mu 5	5.9088	up	5.454 E-09	0.00000 3495
42	17549150	4.0644	up	5.665 E-09	0.00000 3543

43	Sp100 :: nuclear antigen Sp100	6.721	up	5.864 E-09	0.00000 3582
44	Igtp :: interferon gamma induced GTPase	14.6896	up	6.216 E-09	0.00000 3639
45	Ubash3b :: ubiquitin associated and SH3 domain containing, B	3.4232	up	6.234 E-09	0.00000 3639
46	Scube3 :: signal peptide, CUB domain, EGF-like 3	3.1924	up	6.474 E-09	0.00000 3678
47	Cercam :: cerebral endothelial cell adhesion molecule	9.0449	up	6.753 E-09	0.00000 3678
48	Cxcl11 :: chemokine (C-X-C motif) ligand 11	14.2065	up	6.962 E-09	0.00000 3678
49	Ddx60 :: DEAD (Asp-Glu-Ala-Asp) box polypeptide 60	8.6685	up	6.964 E-09	0.00000 3678
50	H19 :: H19 fetal liver mRNA	2.8527	down	7.001 E-09	0.00000 3678

613

614

615

616 **Supplementary Table 2.** List of differentially expressed genes between the Id+ and Id- mouse

617 TNBC cells generated from the Id1C3Tag model.

S. No.	Gene name	logFC	P-value	Q-value
1	Ibsp	34.70987	5.90E-32	1.49E-27
2	Car3	22.99498	2.63E-31	3.34E-27
3	Chad	29.2403	2.44E-22	2.06E-18
4	Alpl	13.80174	7.81E-18	4.95E-14
5	Fgg	0.016064	1.27E-13	6.43E-10
6	Wif1	11.50008	1.56E-13	6.60E-10
7	Upk1b	5.2646	1.24E-12	4.50E-09
8	Cyp1b1	3.053822	1.55E-10	4.91E-07
9	Comp	7.095195	1.70E-09	4.78E-06
10	Col8a2	5.33003	2.50E-09	6.33E-06
11	Thy1	0.072133	3.77E-09	7.34E-06
12	Daam2	5.753571	3.72E-09	7.34E-06
13	Oxtr	5.647227	3.68E-09	7.34E-06
14	Angptl4	0.364436	8.10E-09	1.47E-05
15	Bmp8a	35.35223	1.48E-08	2.50E-05
16	Cilp2	6.256408	2.79E-08	4.16E-05
17	Spink5	19.48619	2.73E-08	4.16E-05
18	Panx3	8.509771	4.80E-08	6.75E-05

19	Stc2	0.227554	6.35E-08	8.46E-05
20	Upp1	0.293682	2.00E-07	0.000254
21	C1qtnf3	0.17218	3.11E-07	0.000375
22	Aspm	1.943525	3.55E-07	0.000409
23	Gpr133	0.059733	4.17E-07	0.000459
24	Olfml2a	4.777964	7.71E-07	0.000814
25	Smad9	4.26899	8.80E-07	0.000892
26	BC106179	0.007113	1.01E-06	0.000986
27	Smoc2	3.83852	1.06E-06	0.00099
28	Zfhx4	0.027723	1.34E-06	0.00121
29	Id1	3.148029	1.48E-06	0.00129
30	Hmga1-rs1	0.472876	2.07E-06	0.001751
31	Cadm3	0.007485	2.45E-06	0.002
32	Angptl7	0.356076	2.63E-06	0.002085
33	Ltbp2	1.879608	2.78E-06	0.002135
34	Gas2l3	2.020165	2.93E-06	0.002182
35	Casz1	1.992817	3.08E-06	0.002231
36	Casc5	1.99469	3.22E-06	0.002269
37	Trim71	6.6267	3.50E-06	0.002394
38	3110079O15Rik	2.516533	3.65E-06	0.002435
39	Mylk	2.190019	5.24E-06	0.003408
40	Itga10	3.333803	6.08E-06	0.003853
41	Sulf1	0.155514	6.27E-06	0.003877

42	Cldn6	0.006279	7.02E-06	0.004138
43	Il6	2.423294	6.99E-06	0.004138
44	Ube2c	2.359489	8.44E-06	0.00486
45	Lox	0.252412	1.03E-05	0.005689
46	Fat2	3.213024	1.03E-05	0.005689
47	Grem2	0.008294	1.16E-05	0.006248
48	Scn7a	0.06646	1.32E-05	0.006944
49	Nusap1	2.131026	1.63E-05	0.008424
50	Ccnd1	0.48054	1.86E-05	0.00945
51	Chst13	6.796579	2.05E-05	0.010165
52	Cytl1	2.994287	2.62E-05	0.012746
53	Cxcl15	0.008277	2.77E-05	0.01323
54	Gfra2	0.328484	3.09E-05	0.014486
55	Ptx3	2.57852	3.23E-05	0.014626
56	Sfrp2	3.039032	3.23E-05	0.014626
57	Aurka	1.974289	3.54E-05	0.01575
58	Kif4	1.712909	3.89E-05	0.016983
59	Slpi	1.960889	3.98E-05	0.017109
60	Arhgap42	1.708361	4.67E-05	0.019716
61	Fgb	0.022514	4.81E-05	0.019924
62	Gm12324	6.61234	4.87E-05	0.019924
63	Aire	0.36178	5.06E-05	0.02035
64	Serpinb8	2.309948	5.28E-05	0.020453

65	Fam78b	0.007857	5.22E-05	0.020453
66	Cyp2b19	3.242025	5.33E-05	0.020453
67	Ctgf	2.026126	5.59E-05	0.020827
68	Prc1	1.714893	5.62E-05	0.020827
69	Mfap4	0.103597	5.67E-05	0.020827
70	Ptges	0.522245	5.77E-05	0.020886
71	Kif11	1.680348	6.37E-05	0.022222
72	Sdk2	2.449539	6.26E-05	0.022222
73	Sytl5	3.537568	6.40E-05	0.022222
74	Chil1	2.190271	6.58E-05	0.022522
75	Lef1	2.453506	7.00E-05	0.022795
76	Mmp3	0.458453	7.02E-05	0.022795
77	Fcgr3	0.013883	6.84E-05	0.022795
78	Wfdc3	3.703518	6.86E-05	0.022795
79	Tmeff2	3.820828	7.85E-05	0.025178
80	Chrdl2	4.840907	8.28E-05	0.026135
81	Dnm3	2.903093	8.42E-05	0.026135
82	Gramd2	2.748813	8.46E-05	0.026135
83	Cenpf	1.707497	8.98E-05	0.027428
84	Stmn2	3.600841	9.43E-05	0.028459
85	Frem1	4.723742	9.59E-05	0.028608
86	Lect1	4.066957	9.95E-05	0.029319
87	Sctr	16.04509	0.000102	0.029784

88	Tmem252	15.01066	0.000103	0.029793
89	Adam19	0.122191	0.000111	0.031557
90	Shc2	2.932148	0.000121	0.033596
91	Adamtsl3	2.531923	0.00012	0.033596
92	Pdgfc	1.733254	0.00013	0.035921
93	Peg3	2.447937	0.000141	0.037694
94	Syn1	2.457699	0.000141	0.037694
95	Sema3e	4.518764	0.000139	0.037694
96	Tmem56	0.28973	0.000149	0.039273
97	Zakit	5.98488	0.000156	0.040878
98	Plxdc2	2.138441	0.000162	0.041887
99	Scrg1	7.269112	0.00017	0.04347
100	Dsg1a	3.900087	0.00018	0.045529
101	Tshz3	0.016964	0.000183	0.045653
102	Loxl2	0.520959	0.000189	0.045653
103	Myo1d	1.681598	0.000186	0.045653
104	Ttc12	11.81049	0.000185	0.045653
105	Nhsl2	1.927395	0.000189	0.045653
106	Arsi	3.65595	0.000199	0.04762
107	Hus1b	0.016583	0.000207	0.048961
108	Lphn1	1.825075	0.000209	0.049079
109	Tmem47	3.039314	0.000211	0.049131
110	Lgr6	1.96859	0.000221	0.049982

111	Ctla2a	2.473585	0.000219	0.049982
112	Omp	2.821668	0.000218	0.049982

618

619

620 **Supplementary Table 3.** 61 candidate genes were identified for further validation as putative Id

621 candidate target genes by siRNA screen.

S. No.	miRNA.mimicsiRNA	Viability1	Viability2	Mean_Viability	Replicate_Disparity
1	Actb	119.2385	69.44172	94.34012	0.199673
2	Adamtsl3	59.5834	72.47348	66.02844	0.051686
3	Alpl	90.50029	96.10138	93.30084	0.022459
4	ANGPTL4	47.11264	43.44723	45.27994	0.014697
5	Aspm	73.0994	49.45445	61.27693	0.094811
6	AURKA	47.13871	62.01435	54.57653	0.059648
7	Axin2	120.1843	168.8596	144.522	0.195177
8	BMI1	76.35547	93.35345	84.85446	0.068158
9	BMP8A	54.38554	78.88393	66.63473	0.098233
10	Casc5	12.91458	20.8636	16.88909	0.031874
11	CCND1	2.362772	2.487312	2.425042	0.000499
12	Ccne1	107.1427	101.5924	104.3675	0.022255
13	Cdkn1a	151.815	187.5555	169.6852	0.143311
14	Cdkn2a	9.258506	12.23492	10.74671	0.011935
15	CENPF	104.4639	126.2672	115.3655	0.087426
16	CHAD	161.2562	192.6356	176.9459	0.125824
17	CLDN6	5.437018	16.92079	11.1789	0.046047
18	Col8a2	92.66478	91.96999	92.31739	0.002786
19	Comp	112.4995	166.9346	139.717	0.218272

20	CXCL15	11.52211	22.00924	16.76568	0.042051
21	Dpysl2	83.76739	73.22459	78.49599	0.042274
22	FOXC2	32.42075	51.92788	42.17431	0.078219
23	GAPDH	74.66308	129.214	101.9385	0.218736
24	Gp9	67.24151	72.35828	69.79989	0.020517
25	GPR133	66.20689	72.52594	69.36641	0.025338
26	GREM2	14.9732	27.7484	21.3608	0.051226
27	Gypa	55.29883	41.61613	48.45748	0.054864
28	Ibsp	56.76146	76.95146	66.85646	0.080957
29	ID1	119.4619	135.7036	127.5827	0.065126
30	ID2	25.04487	48.90737	36.97612	0.095683
31	ID3	26.8322	26.7894	26.8108	0.000172
32	Id4	60.58005	61.8837	61.23187	0.005227
33	Il6	54.41804	38.86627	46.64216	0.062359
34	Kif11	5.929065	10.06528	7.997174	0.016585
35	Kif4	137.4777	145.0792	141.2785	0.03048
36	Lef1	69.74422	75.35658	72.5504	0.022504
37	Lgr6	13.77349	27.60068	20.68709	0.055444
38	Lox	30.49407	53.24569	41.86988	0.091229
39	Loxl2	33.93647	41.46456	37.70052	0.030186
40	Myom1	78.83341	66.06561	72.44951	0.051196
41	Nusap1	26.12422	18.19966	22.16194	0.031776
42	Oxtr	14.45927	29.72044	22.08986	0.061194

43	Panx3	70.37883	95.81948	83.09915	0.102011
44	Pdgfc	48.00148	46.44553	47.2235	0.006239
45	POSTN	57.7375	78.37966	68.05858	0.08277
46	Prc1	44.28233	36.23709	40.25971	0.03226
47	ROBO1	249.3912	225.3732	237.3822	0.096306
48	Scrn1	219.7686	218.585	219.1768	0.004746
49	Sctr	102.5398	85.73719	94.13851	0.067375
50	Sfrp2	113.2776	131.8863	122.582	0.074616
51	Smad9	37.48249	43.25865	40.37057	0.023161
52	Spaca3	46.28752	72.89763	59.59258	0.1067
53	Stom	174.221	165.451	169.836	0.035165
54	TGFBR3	123.9611	98.65301	111.307	0.101479
55	Tmem252	9.216349	8.13346	8.674905	0.004342
56	Tmem47	33.99214	59.51196	46.75205	0.102328
57	Ube2c	66.20978	93.01105	79.61042	0.107467
58	VCAM1	20.82949	33.1387	26.9841	0.049357
59	Vps51	46.86967	34.84365	40.85666	0.048222
60	WIF1	31.32273	52.26417	41.79345	0.08397
61	ylk	103.3257	62.85685	83.09127	0.162271

622

623

References

1. Wee S. Teo HH NK, Aurélie S. Cazet , Daniel L.Roden, Kate Harvey,Christina Valbirk Konrad, Reshma Murali,Binitha Anu Varghese, Archana P. T. Chia-LingChan, Andrea McFarland,Simon Junankar, Sunny Ye,Jessica Yang,Iva Nikolic, Jaynish S. Shah,Laura A. Baker, Ewan K.A. Millar, Mathew J. Naylor, Christopher J. Ormandy, Sunil R. Lakhani, Warren Kaplan, Albert S. Mellick, Sandra A. O'Toole, Radhika Nair$, Alexander Swarbrick$. Id proteins promote a cancer stem cell phenotype in triple negative breast cancer via Robo1-dependent c-Myc activation. Biorxiv. 2019.

2. Cubillo E D-LA CE, Moreno-Bueno G, Peinado H. E47 and Id1 Interplay in Epithelial-Mesenchymal Transition Plos one. 2013.

3. Ashley G. Rivenbark SMOC, and William B. Coleman. Molecular and Cellular Heterogeneity in Breast Cancer Challenges for Personalized Medicine. Am J Pathol. 2013.

4. Clinton Yam aSAM, b and Stacy L. Moulder. Targeting the Molecular Subtypes of Triple Negative Breast Cancer: Understanding the Diversity to Progress the Field. Oncologist. 2017.

5. Sharma P. Biology and Management of Patients With Triple-Negative Breast Cancer. Oncologist. 2016.

6. Lan Thi Hanh Phi INS, Ying-Gui Yang, Sang-Hyun Lee, Nayoung Jun, Kwang Seock Kim, Yun Kyung Lee,corresponding author and Hyog Young Kwon. Cancer Stem Cells (CSCs) in Drug Resistance and their Therapeutic Implications in Cancer Treatment. Stem Cells Int. 2018.

7. Marta Prieto-Vila R-uT, Wataru Usuba, Isaku Kohama, and Takahiro Ochiya*. Drug Resistance Driven by Cancer Stem Cells and Their Niche. Int J Mol Sci. 2017.

8. Nair R1 TW, Mittal V2, Swarbrick A3. ID proteins regulate diverse aspects of cancer progression and provide novel therapeutic opportunities. Mol Ther 2014.

9. Swarbrick A1 AM, Lee CS, Sergio CM, Caldon CE, Hunter LJ, Sutherland RL, Musgrove EA. Regulation of cyclin expression and cell cycle progression in breast epithelial cells by the helix-loop-helix protein Id1. Oncogene. 2005 Jan 13.

10. Wee S. Teo1. Id proteins promote a cancer stem cell phenotype in triple negative breast cancer via Robo1-dependent c-Myc activation. Biorxiv. 2019.

11. O'Brien CA1 KA, Ryan P, Hermans KG, Gibson L, Wang Y, Tsatsanis A, Gallinger S, Dick JE. ID1 and ID3 regulate the self-renewal capacity of human colon cancer-initiating cells through p21. Cancer Cell. 2012 Jun 12.

12. Swarbrick A RE, Allen T, Bishop JM. Id1 cooperates with oncogenic Ras to induce metastatic mammary carcinoma by subversion of the cellular senescence response. Proc Natl Acad Sci. 2008 Apr 8.

13. Kum-Joo Shin EAW, Joelle R. Zavzavadjian, Leah A. Santat, Jamie Liu, Jong-Ik Hwang, Robert Rebres, Tamara Roach, William Seaman, Melvin I. Simon, and Iain D. C. Fraser. A single lentiviral vector platform for microRNA-based conditional RNA interference and coordinated transgene expression. PNAS. September 12, 2006

14. Christopher J O'Conor‡ TC, Iván González, Dengfeng Cao & Yan Peng. Cancer stem cells in triple-negative breast cancer: a potential target and prognostic marker. BIOMARKERS IN MEDICINE. 2018.

668 15. Zhang M1 BF, Atkinson RL, Landis MD, Kittrell F, Edwards D, Medina D, Tsimelzon
669 A, Hilsenbeck S, Green JE, Michalowska AM, Rosen JM. Identification of tumor-initiating cells
670 in a p53-null mouse model of breast cancer. Cancer Res. 2008 Jun 15.
671 16. Jason I. Herschkowitz aWZ, b,c Mei Zhang,a Jerry Usary,c,d George Murrow,c David
672 Edwards,a Jana Knezevic,a Stephanie B. Greene,a David Darr,c,d Melissa A. Troester,c Susan
673 G. Hilsenbeck,e Daniel Medina,a Charles M. Perou,c,d,f and Jeffrey M. Rosena,1. Comparative
674 oncogenomics identifies breast tumors enriched in functional tumor-initiating cells. Proc Natl
675 Acad Sci U S A. 2012 Feb 21.
676 17. Vaillant F1 A-LM, Shackleton M, Forrest NC, Lindeman GJ, Visvader JE. The
677 mammary progenitor marker CD61/beta3 integrin identifies cancer stem cells in mouse models
678 of mammary tumorigenesis. Cancer Res. 2008 Oct 1.
679 18. Nair R1 RD, Teo WS2, McFarland A1, Junankar S2, Ye S2, Nguyen A1, Yang J1,
680 Nikolic I1, Hui M1, Morey A3, Shah J4, Pfefferle AD5, Usary J6, Selinger C7, Baker LA8,
681 Armstrong N9, Cowley MJ10, Naylor MJ11, Ormandy CJ2, Lakhani SR12, Herschkowitz JI13,
682 Perou CM14, Kaplan W15, O'Toole SA16, Swarbrick A2. c-Myc and Her2 cooperate to drive a
683 stem-like phenotype with poor prognosis in breast cancer. Oncogene. 2014 Jul 24.
684 19. Nair R1 TW, Mittal V2, Swarbrick A3. ID proteins regulate diverse aspects of cancer
685 progression and provide novel therapeutic opportunities. Mol Ther. 2014 Aug 22.
686 20. C. Elizabeth Caldon AS, Christine S.L. Lee, Robert L. Sutherland and Elizabeth A.
687 Musgrove. The Helix-Loop-Helix Protein Id1 Requires Cyclin D1 to Promote the Proliferation
688 of Mammary Epithelial Cell Acini. Cancer Res. April 2008.

689 21. George JT1, 3, Jolly MK1, Xu S4, Somarelli JA4, Levine H5,2,6. Survival Outcomes in
690 Cancer Patients Predicted by a Partial EMT Gene Expression Scoring Metric. Cancer Res. 2017
691 No 15.
692 22. Henderika M.J. Hut WL, * Engbert H. Blaauw,† Gert W.A. van Cappellen,‡ Harm H.
693 Kampinga,* and Ody C.M. Sibon. Centrosomes Split in the Presence of Impaired DNA Integrity
694 during Mitosis. Mol Biol Cell. 2003 May.
695 23. Lauren Steward M, LC, BS1, FG, PhD2, aJAM, MD, FACS1. Predictive Factors and
696 Patterns of Recurrence in Patients
697 with Triple Negative Breast Cancer. Ann Surg Oncol 2014.
698 24. Jolly MK1 WK, Gilja S2, Somarelli JA2, Levine H1. EMT and MET: necessary or
699 permissive for metastasis? Mol Oncol. 2017 Jul;11.
700 25. Cornelia Man JR, † Y. L. Yip,* Annie Lai-Man Cheung,* Y. L. Kwong,‡, Stephen J.
701 Doxsey aSWT. Id1 Overexpression Induces Tetraploidization and
702 Multiple Abnormal Mitotic Phenotypes by Modulating
703 Aurora A. Molecular Biology of the Cell. June 2008.
704 26. Jiang M#1, Zhuang H#3, Xia R1,2, Gan L1,2, Wu Y1,2, Ma J1,2, Sun Y4, Zhuang Z1,2.
705 KIF11 is required for proliferation and self-renewal of docetaxel resistant triple negative breast
706 cancer cells. Oncotarget. 2017.
707 27. Chattopadhyay S1 SA, Mukherjee S3, Huang C4, Hartwell KA5, Miller PG6,
708 Subramanian R7, Carmody LC2, Yusuf RZ8, Sykes DB8, Paulk J9, Vetere A2, Vallet S10,
709 Santo L10, Cirstea DD11, Hideshima T11, Dančík V2, Majireck MM9, Hussain MM12, Singh
710 S9, Quiroz R13, Iaconelli J14, Karmacharya R15, Tolliday NJ2, Clemons PA2, Moore MAS16,
711 Stern AM17, Shamji AF2, Ebert BL18, Golub TR19, Raje NS10, Scadden DT20, Schreiber

712 SL21. Niche-Based Screening in Multiple Myeloma Identifies a Kinesin-5 Inhibitor with
713 Improved Selectivity over Hematopoietic Progenitors. Cell Rep. 2015.

714

715

716 **Declarations**

717 **Consent for publication**

718 All authors have given their consent for publication.

719 **Competing interests**

720 The authors declare that they have no competing interests.

721 **Funding**

722 This work was supported by funding from an Early Career Research (ECR) Award from Science
723 and Engineering Research Board (SERB), Government of India (ECR/2015/000031), the
724 National Health and Medical Research Council (NHMRC) of Australia, The National Breast
725 Cancer Foundation and John and Deborah McMurtrie. A.S. is the recipient of a Senior Research
726 Fellowship from the NHMRC. RN is the recipient of the Ramanujan Fellowship from the
727 Government of India (SERB) (SB/S2/RJN/182/2014). MKJ was supported by Ramanujan
728 Fellowship (SB/S2/RJN-049/2018) provided by SERB, DST, Government of India. WST is
729 funded by International Postgraduate Research Scholarship and the Beth Yarrow Memorial
730 Award in Medical Science. APT and BAV is funded through CSIR- Junior Research Fellowship,
731 and DST- INSPIRE Fellowship, respectively.

732 **Author's contributions**

733 RN contributed to the conceptualizaton. RS, BAV, NK and APT contributed to the methodology.
734 DLR to the genomic analysis. RM, BAV, APT, , NK, AC, ED, AM, CK and HH contributed to
735 the investigations. RN wrote the original draft of the manuscript. JTG, HL and MKJ contributed
736 to the bioinformatic analysis of the EMT program. AS reviewed and edited the manuscript. AS

737 and RN contributed to the funding acquisition. All authors read and approved the final
738 manuscript.

739

740 **Acknowledgements**

741 We would like to thank the following people for their assistance with this manuscript:
742 Mr. Tilak Prasad and Mrs. Surabhi S.V for helping with Flow cytometry. Mr. Anurup K.G for
743 technical support with Confocal imaging. Dr. Vishnu Sunil Jayakumar for Animal Handling. Dr.
744 T. R. Santhosh Kumar's Lab for helping with lab consumables and technical support.

745

746 **Author information**

747 Radhika is senior author.
748 Correspondence to radhikanair@rgcb.res.in

749

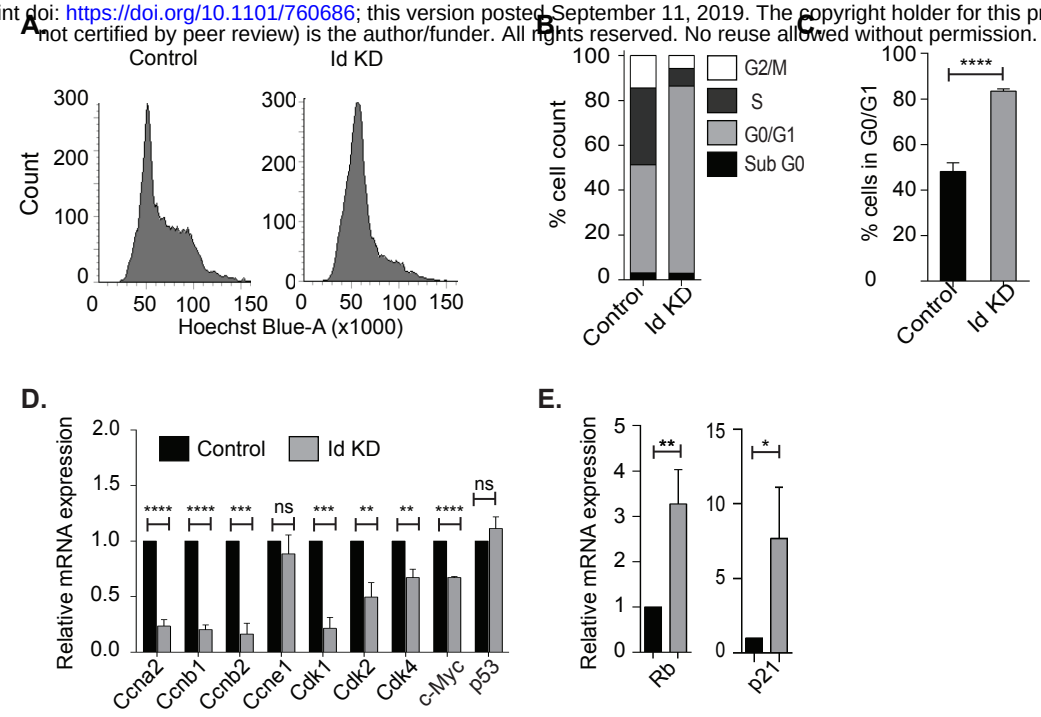
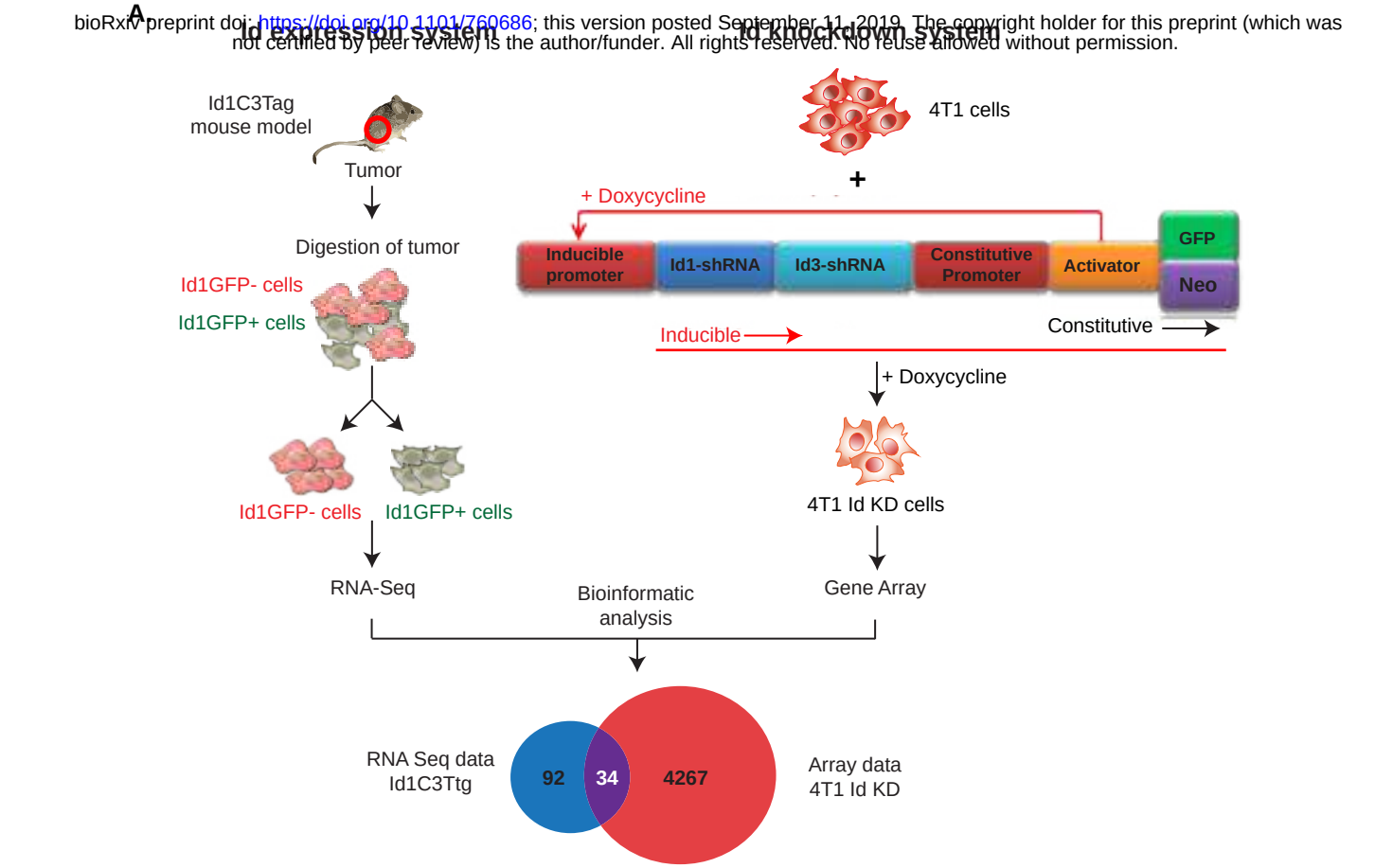
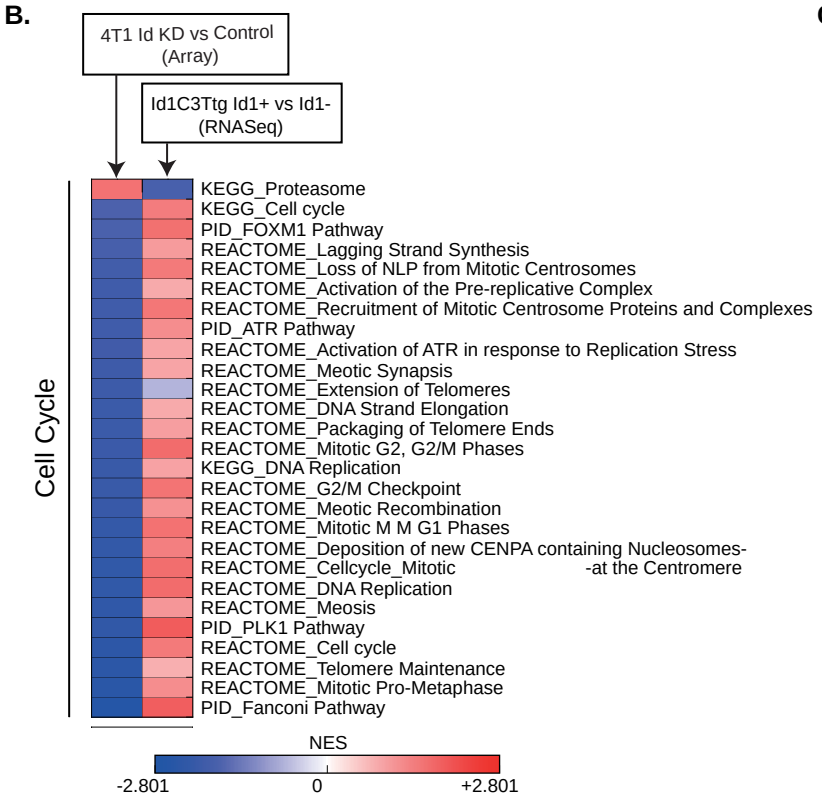
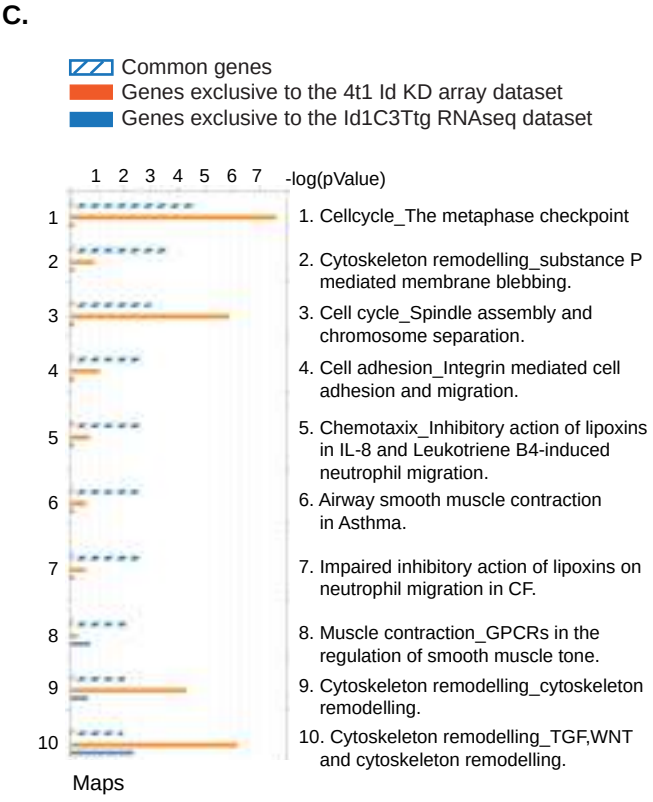


Figure 1







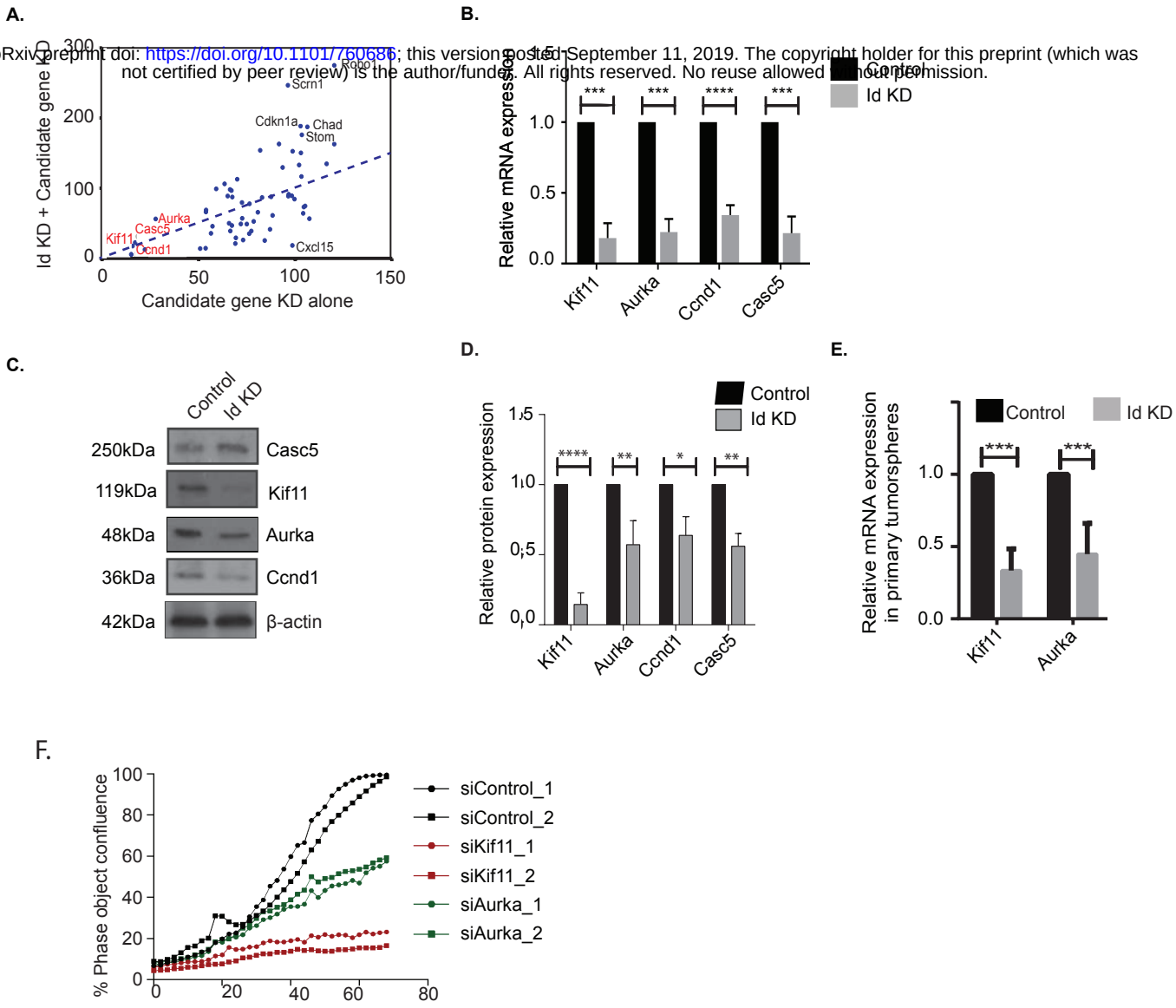


Figure 3

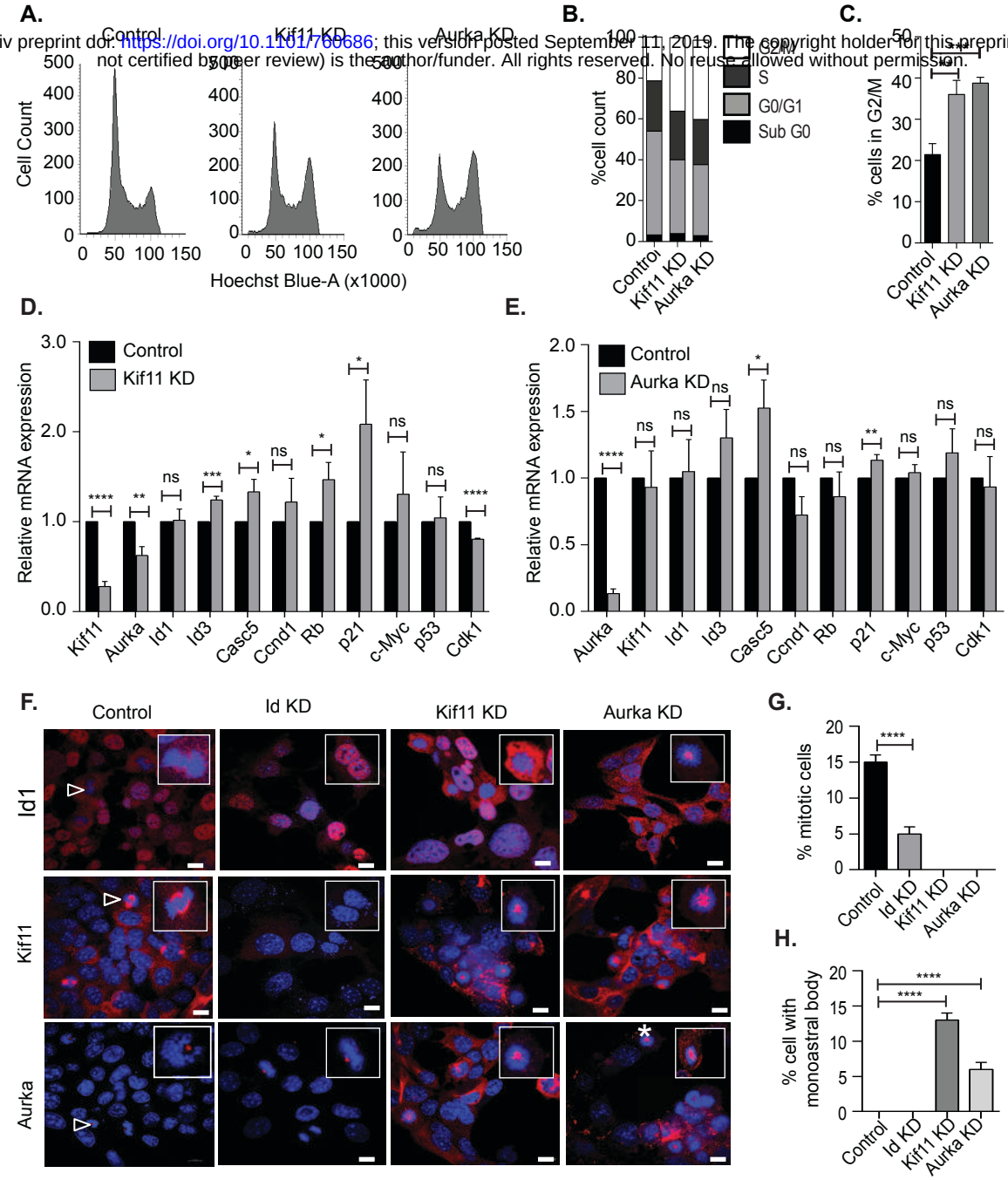
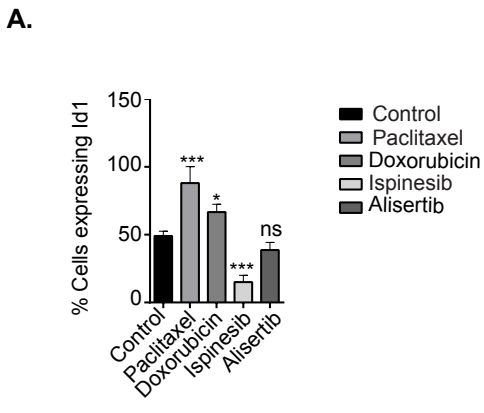
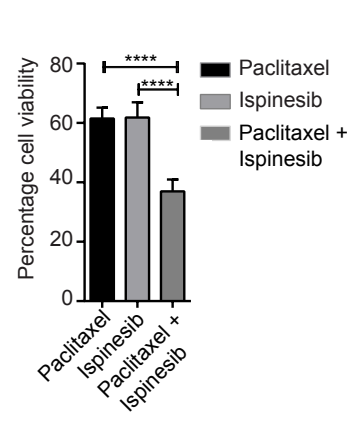
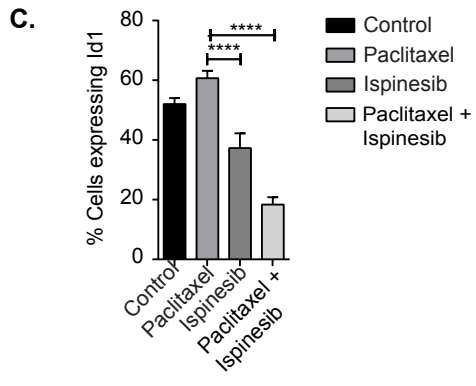
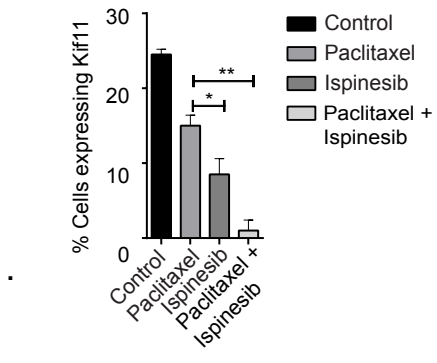


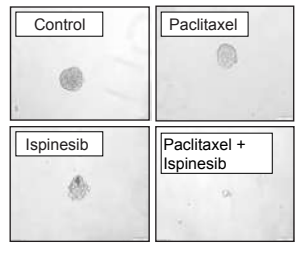
Figure 4

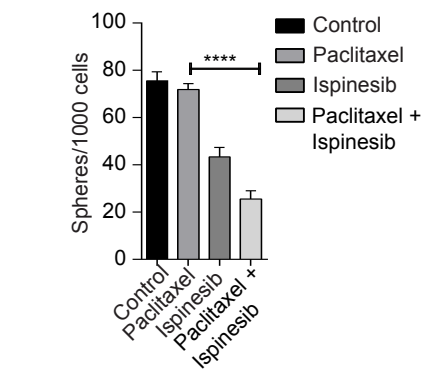




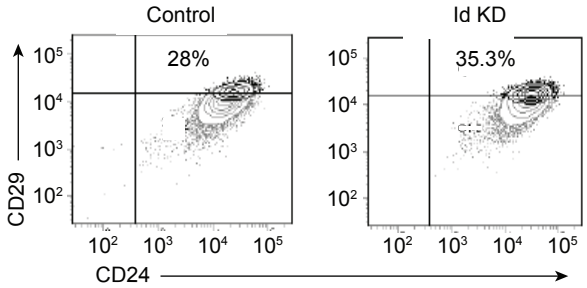




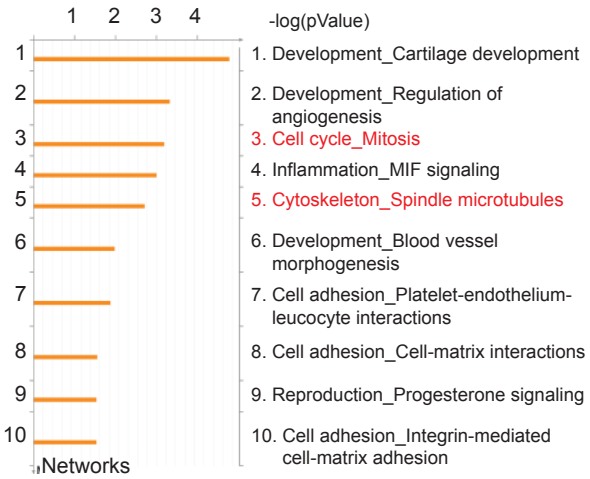




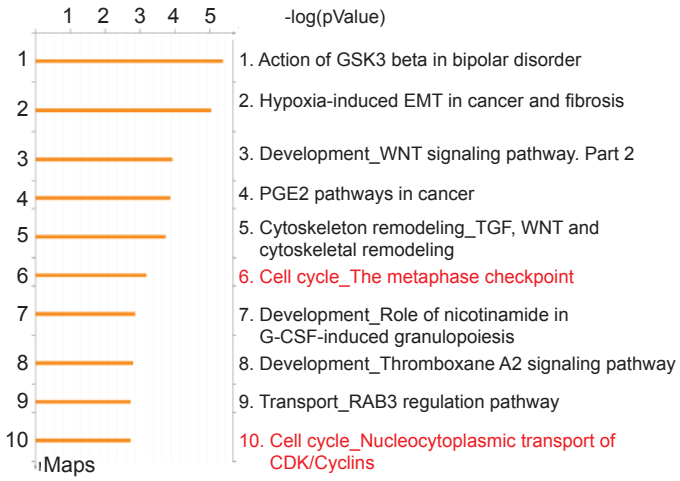
A.



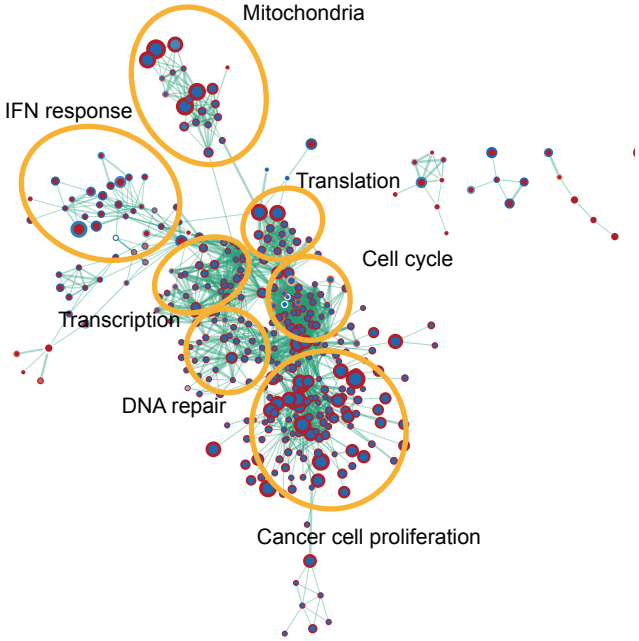
B.



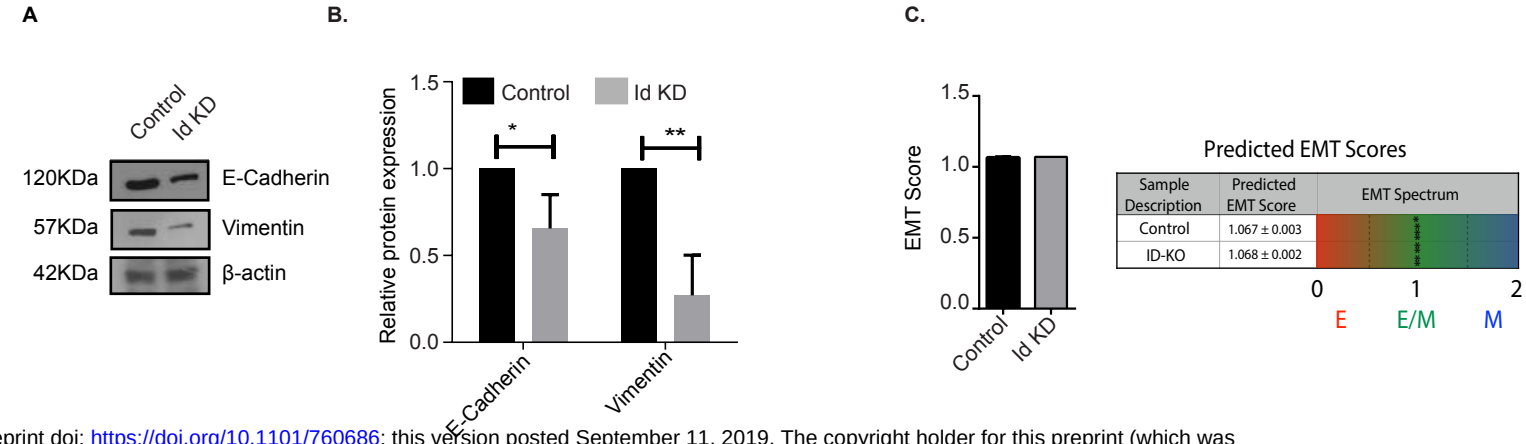
C.

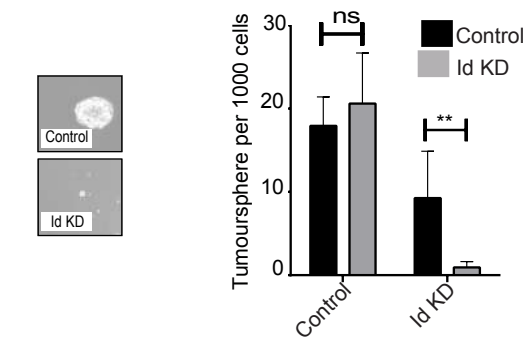


D.

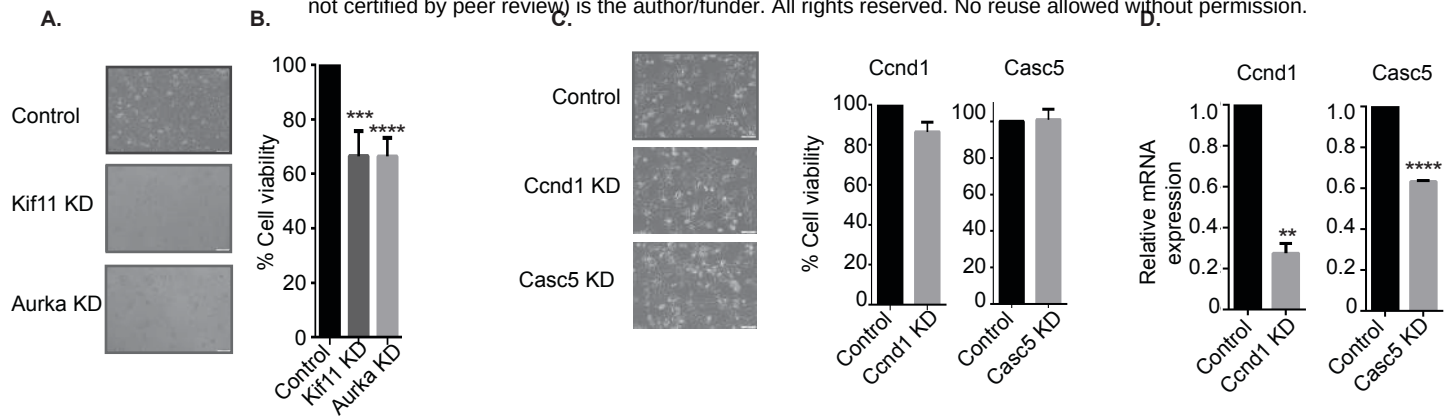


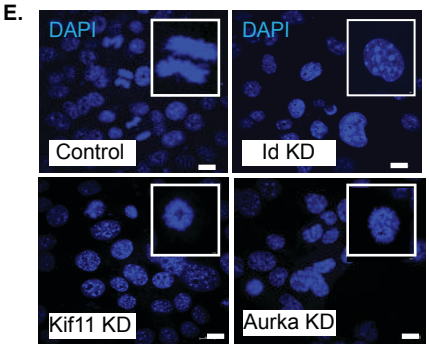
Supplementary Figure 1



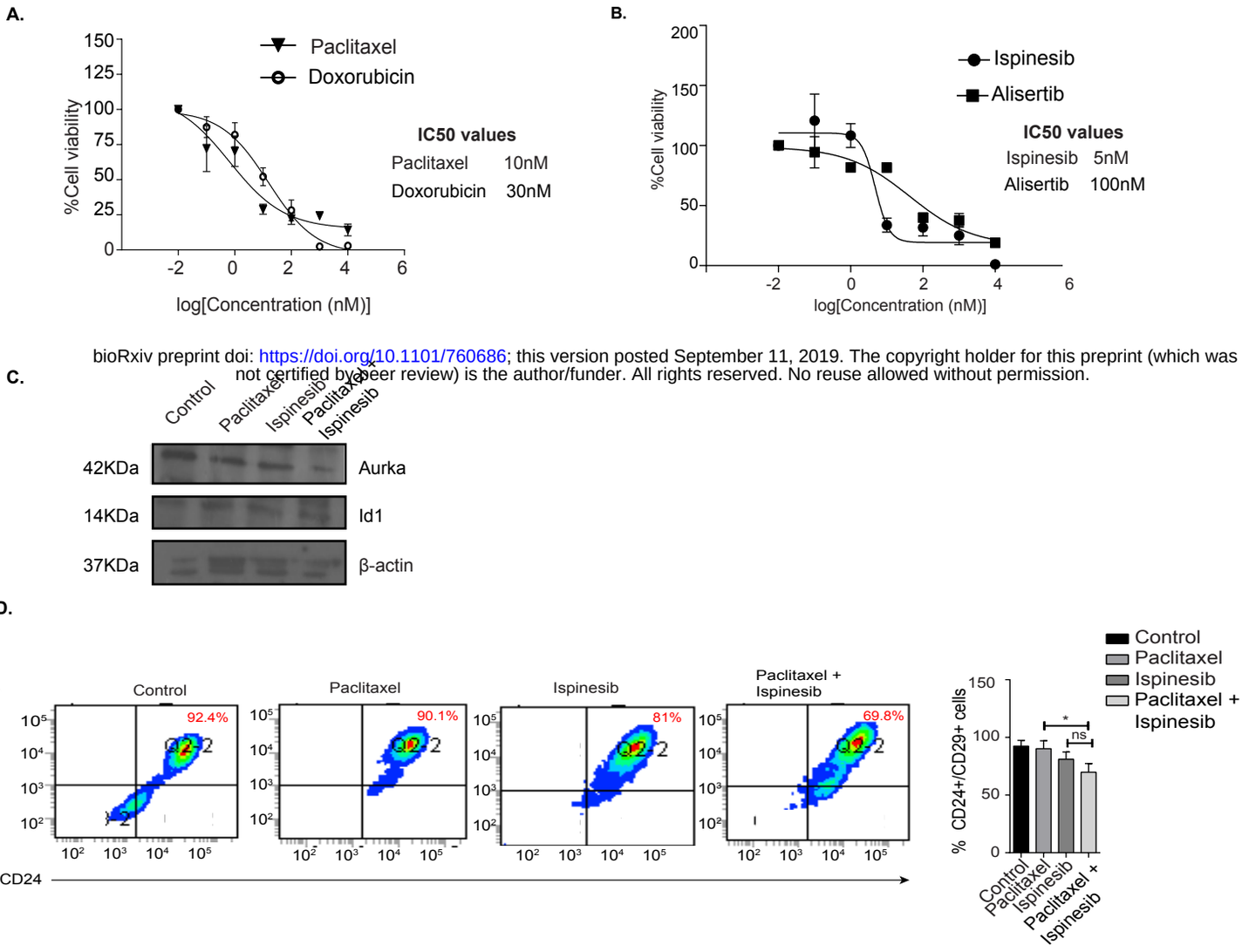


Supplementary Figure 2





Supplementary Figure 3



Supplementary Figure 4