Supplemental Information

EXTENDED EXPERIMENTAL PROCEDURES

Cell culture. The immortalized human mammary epithelial cells (HMLE) and the transformed cells (HMLE-Ras) were kindly provided by Dr. Robert A. Weinberg, and maintained as described (Elenbaas et al., 2001). HeLa and HEK293 cells were maintained in DMEM with10% FBS. Freshly isolated primary normal human MEC or breast cancer cells were cultured in MEGM with supplements (Keller et al., 2012).

Generation of stable cell lines. For overexpression, Pin1 and RAB2A CDS were subcloned into the pBabe retroviral vector or pBybe lentiviral vector. To overexpress Rab2A and the Q58H mutant using lentivirus-mediated gene expression at levels similar to or 3 times over the endogenous level, less optimal Kozak sequences were introduced into the vector, namely GCCTTT and GCCGCC, respectively. Specific point mutations were introduced using the Quickchange kit (Stratagene) and sequences were verified. All lentiviral shRNA constructs were provided by Dr. William C. Hahn. The target sequence of Pin1 shRNA is CCACCGTCACACAGTATTTAT. The target sequences of Rab2A shRNAs are GCTCGAATGATAACTATTGAT and CCAGTGCATGACCTTACTATT. The production of retroviruses or lentiviruses as well as the infection of target cells was described previously(Stewart et al., 2003). Following infection, the cells were selected using puromycin, hygromycin or blasticidin. Cells were used immediately following selection and for up to one month after selection.

Flow Cytometry. For human cell lines, CD24-PE and CD44-APC antibodies (eBioscience) were

used to fractionate the $CD24^{-}CD44^{+}$ population, as described (Al-Hajj et al., 2003). Sorting for $CD24^{-}CD44^{+}$ cells freshly isolated from human breast cancers was performed with Epics Altra flow cytometer (Beckman Coulter). To deplete non-tumor cells from primary cancer samples, a cocktail of lineage marker antibodies including CD2, CD3, CD10, CD16, CD18, CD31, CD64 and CD140b (PharMingen) were used, while to deplete non-tumor cells from mouse specimens, anti-H2 Kd antibody was used.

Microarray Data Processing. Affymetrix CEL files were analyzed with BRB-ArrayTools (Simon et al., 2007). Genes that expressed lower in KO cells than in WT cells with fold change < 0.8 ($P < 0.05$) were selected as "downregulated" ones. Two datasets obtained from NCBI's Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) with GEO Series accession numbers GSE3711 (Stingl et al., 2006) and GSE8863 (Zhang et al., 2008) were reanalyzed together with our raw data. In GSE3711, mammary stem cells (MaSC, defined as lineage- $CD49f^{++}CD24^{+}$) were compared to myoepithelial cells (MYO, defined as lineage- $CD49f^{+}CD24^{+}$) and colony-forming progenitor cells (CFC, defined as lineage- $CD49f^{+}CD24^{++}$). Genes that expressed higher in MaSC than in both MYO and CFC with fold change > 1.5 ($P < 0.05$) were selected as "upregulated" ones. In GSE8863, The Lin-$CD29^{High}CD24^{High}$ subpopulation of CSCs was compared to the Lin-$CD29^{Low}CD24^{Low}$ subpopulation of non-CSCs. Genes that expressed higher in CSC than in non-CSC with fold change > 1.5 ($P < 0.05$) were selected as "upregulated" ones. When comparing the upregulated gene list in these two datasets (SC/non-SC > 1.5, $P < 0.05$) with the downregulated gene list in Pin1 KO cells (KO/WT < 0.8, $P < 0.05$), 14 genes were repeatedly found in the two gene lists and were identified as candidate genes.

Western blot. Primary monoclonal Pin1 antibody (1:5000), polyclonal RAB2A antibody (1:1000) (Proteintech Group), polyclonal Erk1/2 (1:4000) and pErk antibody (1:2000) (Cell Signaling Technology), monoclonal unphosphorylated β-catenin antibody (1:2000) (Millipore), monoclonal M2 antibody for Flag tag (1:2000) (Sigma) and monoclonal Actin antibody (1:5000) (Sigma) are used.

Quantitative RT-PCR. RNA from cells was extracted with the Total RNA isolation mini kit (Agilent). cDNA was prepared with transcriptor first strand cDNA synthesis kit (Roche) and PCR was carried out with iQ SYBR Green Supermix (Bio-Rad). Samples were run on the QIAGEN Rotor-Gene Q real-time cycler. GAPDH was used as an internal control. Analysis was performed with the delta-delta ct method. The following primers were used:

GAPDH forward CATGAGAAGTATGACAACAGCCT

GAPDH reverse AGTCCTTCCACGATACCAAAGT

Pin1 forward GCCTCACAGTTCAGCGACT

Pin1 reverse ACTCAGTGCGGAGGATGATGT

Ecad forward TGCCCAGAAAATGAAAAAGG

Ecad reverse GTGTATGTGGCAATGCGTTC

Ncad forward ACAGTGGCCACCTACAAAGG

Ncad reverse CCGAGATGGGGTTGATAATG

FN1 forward CAGTGGGAGACCTCGAGAAG

FN1 reverse TCCCTCGGAACATCAGAAAC

Vim forward GAGAACTTTGCCGTTGAAGC

Vim reverse GCTTCCTGTAGGTGGCAATC

Cmpk1 forward TGGGAAGGCAGATGTATCTTTCG

Cmpk 1 reverse TGTTGACTGAAGGTAGGTCTGA

ELAVL1 forward AACCATTAAGGTGTCGTATGCTC

ELAVL1 reverse CGCCCAAACCGAGAGAACA

EMP2 forward CATCCAGCTAATGTCATGTCTGT

EMP2 reverse CTCTGGTCACGGGATAGAATTTC

GLE1 forward ACGCAAGCTCTGCCTTTTC

GLE1 reverse CGTGAGGACTGAAGTACCATAGA

HMGN1 forward GCGAAGCCGAAAAAGGCAG

HMGN1 reverse TCCGCAGGTAAGTCTTCTTTAGT

HTATSF1 forward ATGGTGACACCCAGACCGAT

HTATSF1 reverse GAGAAGCCATAATTGGCCTGAT

LAMP2 forward TCCCAAAGATCTGCCTTCAC

LAMP2 reverse TTCTGCATTGTGCTGAGAGG

Magi3 forward TCTTCTTTTGAGGCCAGGAA

Magi3 reverse GGAAAGACCAAGAAAAGCCC

RAB2A forward AGTTCGGTGCTCGAATGATAAC

RAB2A reverse AATACGACCTTGTGATGGAACG

SEH1L forward TGAATCTCAGCCAGTGGTCTT

SEH1L reverse TCATCACTTCCTACGGCGAT

TM7SF3 forward TTCCTTTTCTCCGACTCTCCTT

TM7SF3 reverse CCCCAAGTACCAAGTGCATGT

TM9SF3 forward TGCCAGCCACTTACTGTGAAA

TM9SF3 reverse GCCTCACCAACAATACCCCATA

ZDHHC3 forward AGATTGGACAACCTATGGACTGA

Zdhhc3 reverse GCACTCTGTCGAACTGAAGTTA

ZYG11B forward GAGGAGGCGTCTCCCTATTC

ZYG11B reverse GCATCTGGTTGCCCCTAAAAA

Luciferase reporter assays. For the reporter assay of RAB2A promoter, two deletion luciferase reporter constructs of RAB2A were generated. The promoter sequences from -1310 and -904, which contain the -1293 and -890 AP-1 binding sites, respectively, were subcloned into pGL3 vector. HEK293 cells were plated in 12-well plates for 24 hr and transfected with luciferase reporter constructs, pRL-tk renilla luciferase and Flag-Pin1 or control vector. Increasing dose of Flag-Pin1 or control vector plasmid were add as 0.15, 0.5 1.5 μg. Cells were harvested and luciferase activity was measured 48 hr later using the Dual-Luciferase Reporter Assay System (Promega).

Chromatin immunoprecipitation (ChIP). ChIP assay was performed according to the manufacturer's instruction (Upstate Biotechnology). Monoclonal Pin1 antibody (generated by our lab) or polyclonal c-Jun antibody (Abcam) were used to precipitate the chromatin-protein complexes. Re-ChIP assay was performed as described (Petruk et al., 2012). Real-time PCR primers for the -1293 locus were CCTGTGGTCTTTTTGAACAGAG and

CAACTGGAGGCCCTGTATGT, and for the -890 locus were ACACACACATAAACAGATCATCTCGG and AGTCTCTGAACCTGTCCTGGTTCTG.

In vitro assays. Mammosphere culture was performed as described(Dontu et al., 2003). Single-cell suspension was plated on ultra-low attachment plates (Corning, Costar) in DMEM/F-12 HAM medium containing bFGF, EGF, heparin and B-27 supplement. The mammospheres were cultured for two weeks. Then the mammospheres with diameter >75 μm were counted.

Soft agar assays were done by seeding cells at a density of 10^3 in 60 mm culture dishes containing 0.3% top low-melt agarose and 0.5% bottom low-melt agarose, as described(Ryo et al., 2002). Cells were fed every 4 days, and colonies were stained with 0.2% p-iodonitrotetrazolium violet and counted after 3 weeks.

For wound healing assays, cells were grown to confluence and then wounded using a yellow pipette tip, and migration was visualized by time-lapse imaging. The rate of wound closure was calculated by a ratio of the average distance between the two wound edges and the total duration of migration.

Transwell migration assay were performed as previously described (Luo et al., 2006). Assay media with EGF (5 ng/ml) was added to the bottom chamber. Cells (5×10^4/100 μl) were added to the top chamber of cell culture inserts (8 mm pore size) (Corning, Costar). After 12 h of incubation, cells that migrated to the bottom surface of the insert were fixed with methanol and stained with 0.4% crystal violet. The number of cells that had migrated was quantified by counting ten random distinct fields using a microscope.

GTP hydrolysis assay. Rab2A GTPase hydrolysis assay were performed as described(Davis et al., 2013) with small modifications. GST-Rab2A or GST-Rab2A Q58H (100 nM) was incubated in 20 mM Tris-HCl (pH 7.5), 50 mM NaCl, 10 mM MgCl2, 0.5uM GTP and 3 pmol of [α–^{32}P] GTP at room temperature for the indicated time. The Rab2a-bound nucleotides were eluted with elution buffer (2 mM EDTA, 0.2% sodium dodecyl sulfate, 1 mM GDP, 1 mM GTP). 1uL of the reaction mixture was spotted onto polyethyleneimine-cellulose sheets. Chromatograms were developed in 0.75M KH2PO4 (pH 3.4). GTP and GDP resolved by thin-layer chromatography were visualized by autoradiography film exposure.

Tumor Implantation

Aliquots of indicated numbers of cells were injected into 5-week-old BALB/c nude mice (Jackson Laboratories), as described (Mani et al., 2008). The tumor incidence was monitored by palpation and determined at two months after injection, with the same tumor incidence at 6 months postinjection. After tumors were detected, tumor size was measured every three days.

Preparation of single-cell suspensions. Human mammary reduction plasty tissues and breast cancer tissues were mechanically disaggregated and then digested with 200 U/ml collagenase (Sigma) and 100 U/ml hyaluronidase (Sigma), as described(Al-Hajj et al., 2003). The resultant organoids were further digested in 0.25% trypsin-EDTA and Dispase/DNaseI, and then filtered through a 40 μm mesh.

Immunohistochemistry analyses on tissue microarrays. Formalin-fixed and paraffin-embedded tissue microarrays of human breast tissue were purchased from Imgenex (IMH-364 and 371). Rab2A (Proteintech Group) and ALDH1 (BD biosciences) staining was

performed following the manufacturer's protocol. Immunolabeling was visualized with a mixture of DAB solution (Vector Laboratories), followed by counterstaining with hematoxylin. Microscopic analysis was assessed in a blinded manner. Immunostaining results were scored using percentage (P) × intensity (I), as described(Ginestier et al., 2002). In brief, percentage of positive cells ranged from 0 to 100, and intensity was categorized into three groups as 1 (negative or weak), 2 (moderate) and 3 (strong). Expression levels are scored as low ($0 < P \times I \leqslant 100$), medium ($100 < P \times I \leqslant 200$) and high ($200 < P \times I \leqslant 300$). For ALDH1, only the intensity was estimated, because the percentages of positive cells were low. Intensity in foci with maximum staining was scored as low, medium and high, as described(Kunju et al., 2011).

SUPPLEMENTAL REFERENCES

Al-Hajj, M., Wicha, M.S., Benito-Hernandez, A., Morrison, S.J., and Clarke, M.F. (2003). Prospective identification of tumorigenic breast cancer cells. Proc Natl Acad Sci U S A *100*, 3983-3988.

Davis, M.J., Ha, B.H., Holman, E.C., Halaban, R., Schlessinger, J., and Boggon, T.J. (2013). RAC1P29S is a spontaneously activating cancer-associated GTPase. Proc Natl Acad Sci U S A *110*, 912-917.

Dontu, G., Abdallah, W.M., Foley, J.M., Jackson, K.W., Clarke, M.F., Kawamura, M.J., and Wicha, M.S. (2003). In vitro propagation and transcriptional profiling of human mammary stem/progenitor cells. Genes Dev *17*, 1253-1270.

Elenbaas, B., Spirio, L., Koerner, F., Fleming, M.D., Zimonjic, D.B., Donaher, J.L., Popescu, N.C., Hahn, W.C., and Weinberg, R.A. (2001). Human breast cancer cells generated by oncogenic transformation of primary mammary epithelial cells. Genes Dev *15*, 50-65.

Ginestier, C., Charafe-Jauffret, E., Bertucci, F., Eisinger, F., Geneix, J., Bechlian, D., Conte, N., Adelaide, J., Toiron, Y., Nguyen, C., *et al.* (2002). Distinct and complementary information provided by use of tissue and DNA microarrays in the study of breast tumor markers. Am J Pathol *161*, 1223-1233.

Keller, P.J., Arendt, L.M., Skibinski, A., Logvinenko, T., Klebba, I., Dong, S., Smith, A.E., Prat, A., Perou, C.M., Gilmore, H., *et al.* (2012). Defining the cellular precursors to human breast cancer. Proc Natl Acad Sci U S A *109*, 2772-2777.

Kunju, L.P., Cookingham, C., Toy, K.A., Chen, W., Sabel, M.S., and Kleer, C.G. (2011). EZH2 and ALDH-1 mark breast epithelium at risk for breast cancer development. Mod Pathol *24*, 786-793.

Luo, M.L., Shen, X.M., Zhang, Y., Wei, F., Xu, X., Cai, Y., Zhang, X., Sun, Y.T., Zhan, Q.M., Wu, M., *et al.* (2006). Amplification and overexpression of CTTN (EMS1) contribute to the metastasis of esophageal squamous cell carcinoma by promoting cell migration and anoikis resistance. Cancer Res *66*, 11690-11699.

Mani, S.A., Guo, W., Liao, M.J., Eaton, E.N., Ayyanan, A., Zhou, A.Y., Brooks, M., Reinhard, F., Zhang, C.C., Shipitsin, M., *et al.* (2008). The epithelial-mesenchymal transition generates cells with properties of stem cells. Cell *133*, 704-715.

Petruk, S., Sedkov, Y., Johnston, D.M., Hodgson, J.W., Black, K.L., Kovermann, S.K., Beck, S., Canaani, E., Brock, H.W., and Mazo, A. (2012). TrxG and PcG proteins but not methylated histones remain associated with DNA through replication. Cell *150*, 922-933.

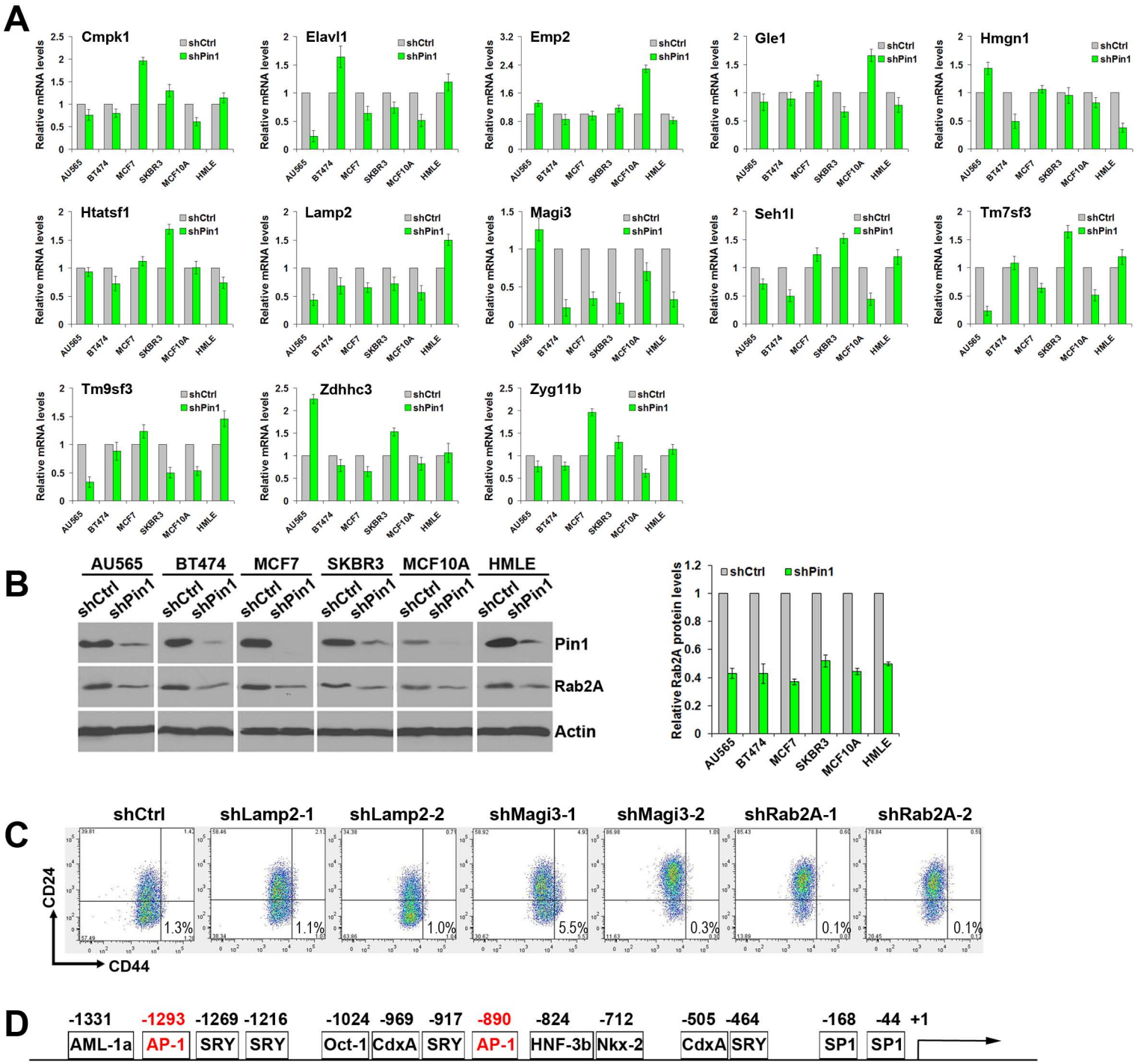
Ryo, A., Liou, Y.C., Wulf, G., Nakamura, M., Lee, S.W., and Lu, K.P. (2002). PIN1 is an E2F target gene essential for Neu/Ras-induced transformation of mammary epithelial cells. Mol Cell Biol *22*, 5281-5295.

Simon, R., Lam, A., Li, M.C., Ngan, M., Menenzes, S., and Zhao, Y. (2007). Analysis of gene expression data using BRB-ArrayTools. Cancer Inform *3*, 11-17.

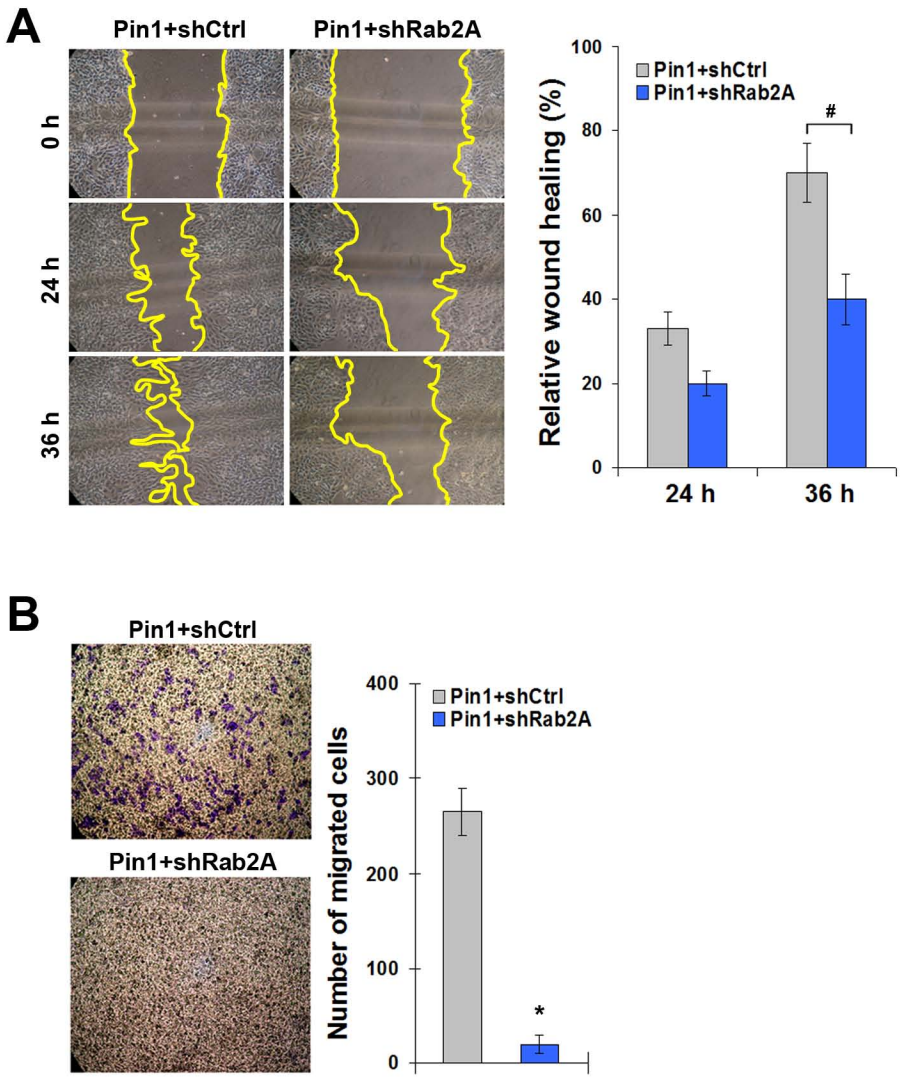
Stewart, S.A., Dykxhoorn, D.M., Palliser, D., Mizuno, H., Yu, E.Y., An, D.S., Sabatini, D.M., Chen, I.S., Hahn, W.C., Sharp, P.A., *et al.* (2003). Lentivirus-delivered stable gene silencing by RNAi in primary cells. RNA *9*, 493-501.

Stingl, J., Eirew, P., Ricketson, I., Shackleton, M., Vaillant, F., Choi, D., Li, H.I., and Eaves, C.J. (2006). Purification and unique properties of mammary epithelial stem cells. Nature *439*, 993-997.

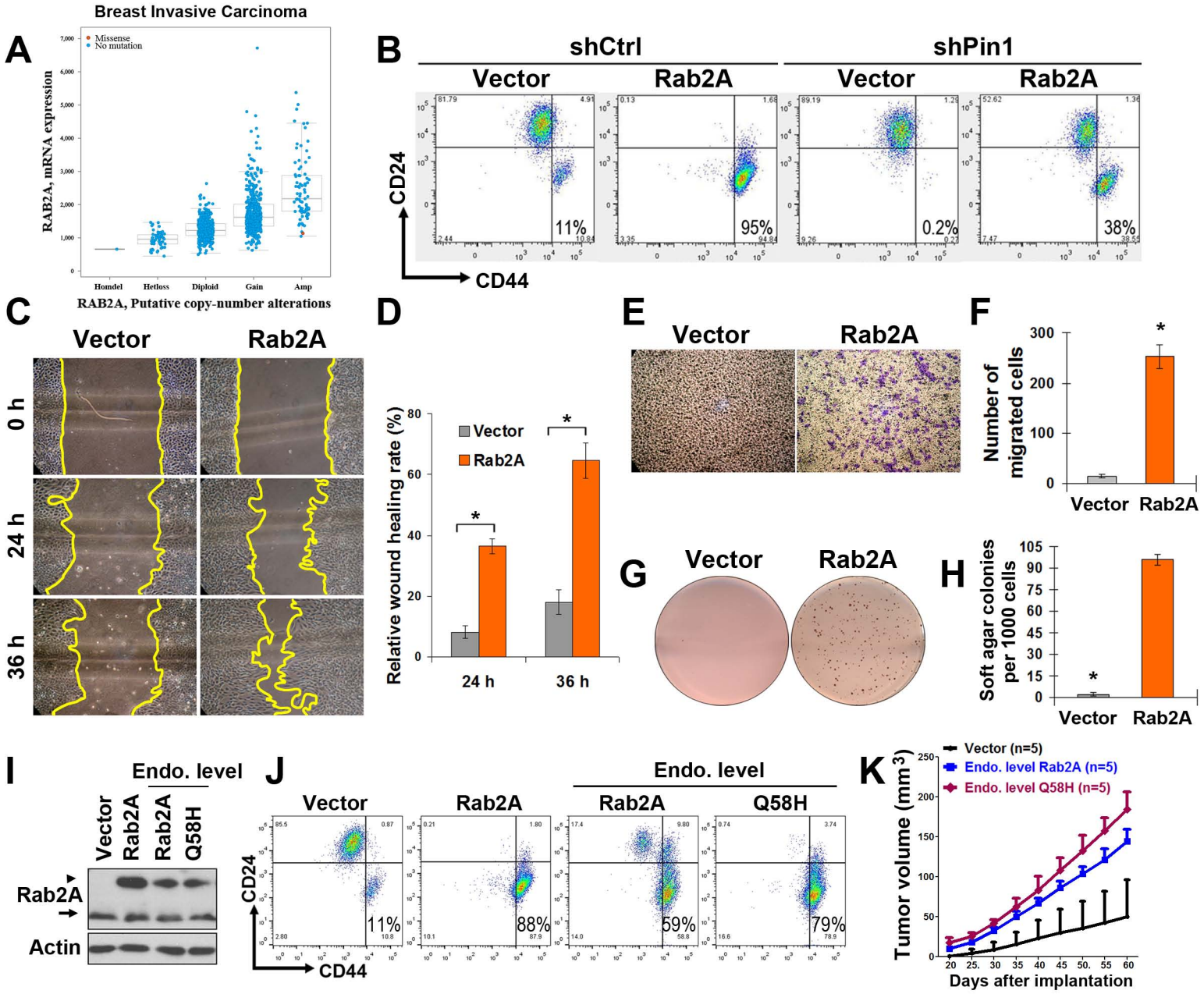
Zhang, M., Behbod, F., Atkinson, R.L., Landis, M.D., Kittrell, F., Edwards, D., Medina, D., Tsimelzon, A., Hilsenbeck, S., Green, J.E., *et al.* (2008). Identification of tumor-initiating cells in a p53-null mouse model of breast cancer. Cancer Res *68*, 4674-4682.



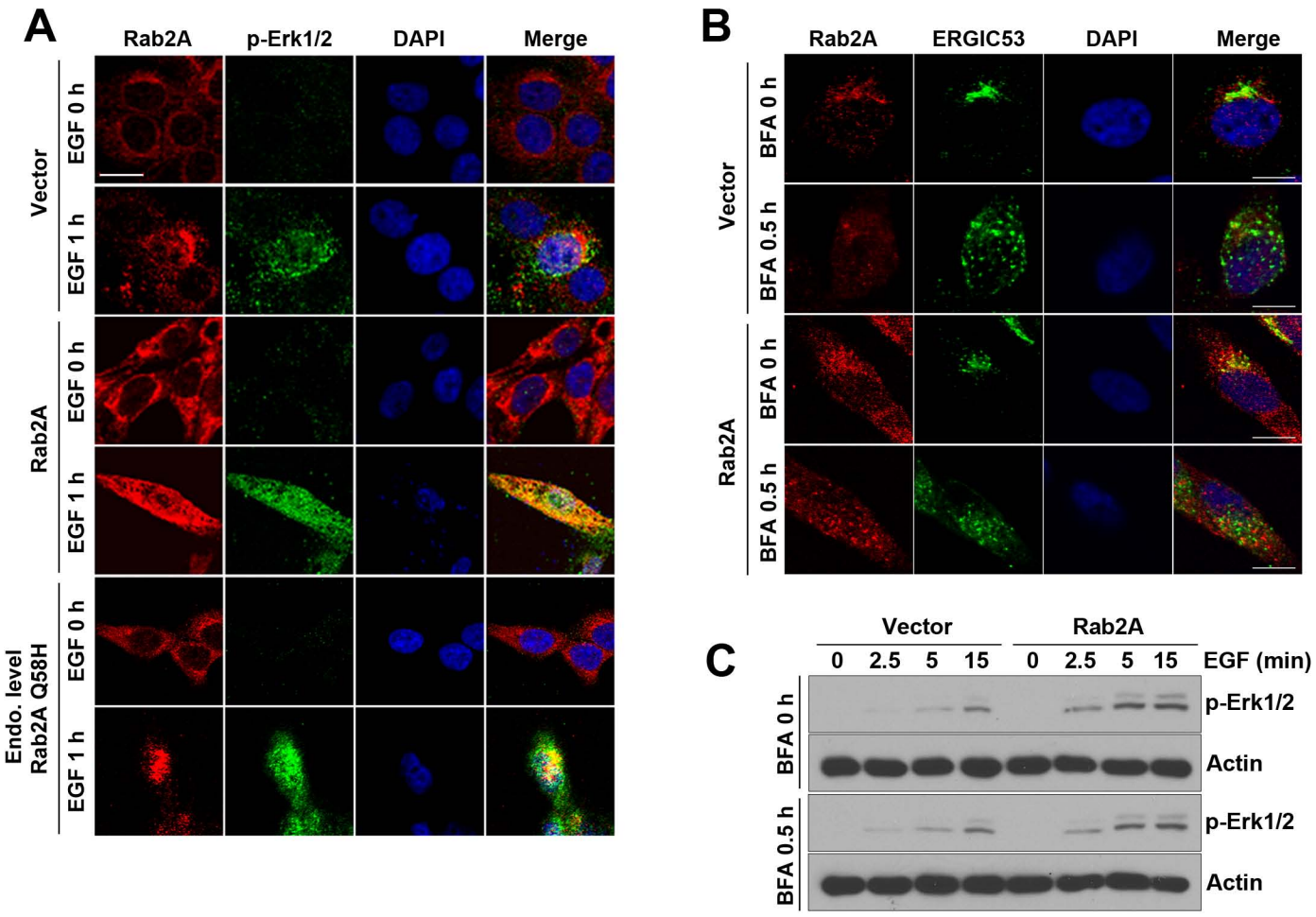
Luo et al., Figure S1. A series of genomic profiling analyses identify Rab2A as a target gene regulated by Pin1, Related to Figure 1. (A) Real-time PCR analysis of mRNA expression of 13 candidate genes in six Pin1 KD breast cell lines. **(B)** Pin1 KD reduced Rab2A expression in six human breast cancer cell cells at the protein level. **(C)** Lamp2, Magi3 and Rab2A expressions were knocked down by two shRNAs in MCF10A cells. Only Rab2A , but not Lamp2 or Magi3 knockdown consistently reduced the CD24-CD44+ population. **(D)** Schematic representation of Rab2A promotor with predicted transcription factor binding sites in TFsearch. In all panals, bar graphs present mean±SD.



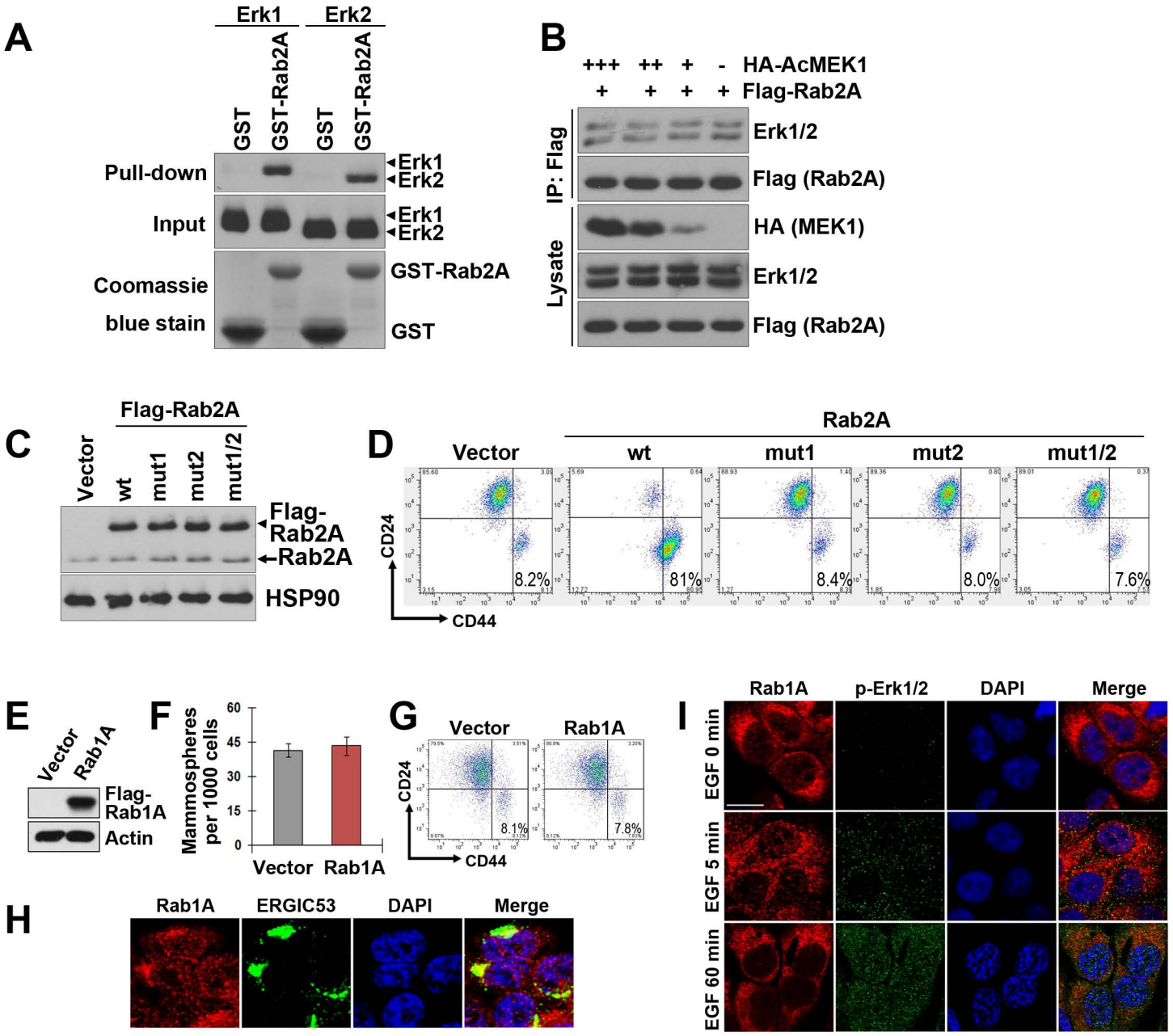
Luo et al., Figure S2. Rab2A knockdown abrogates the EMT phenotypes of Pin1 overexpression, Related to Figure 1. (A) Rab2A knockdown in Pin1-overexpressing HMLE cells impaired the wound healing capability. **(B)** Rab2A knockdown impaired the ability of Pin1 overexpression to increase cell migration, as measured by the transwell assay. In all panels, bar graphs present mean±SD of three independent experiments. # P<0.01, *P<0.001.



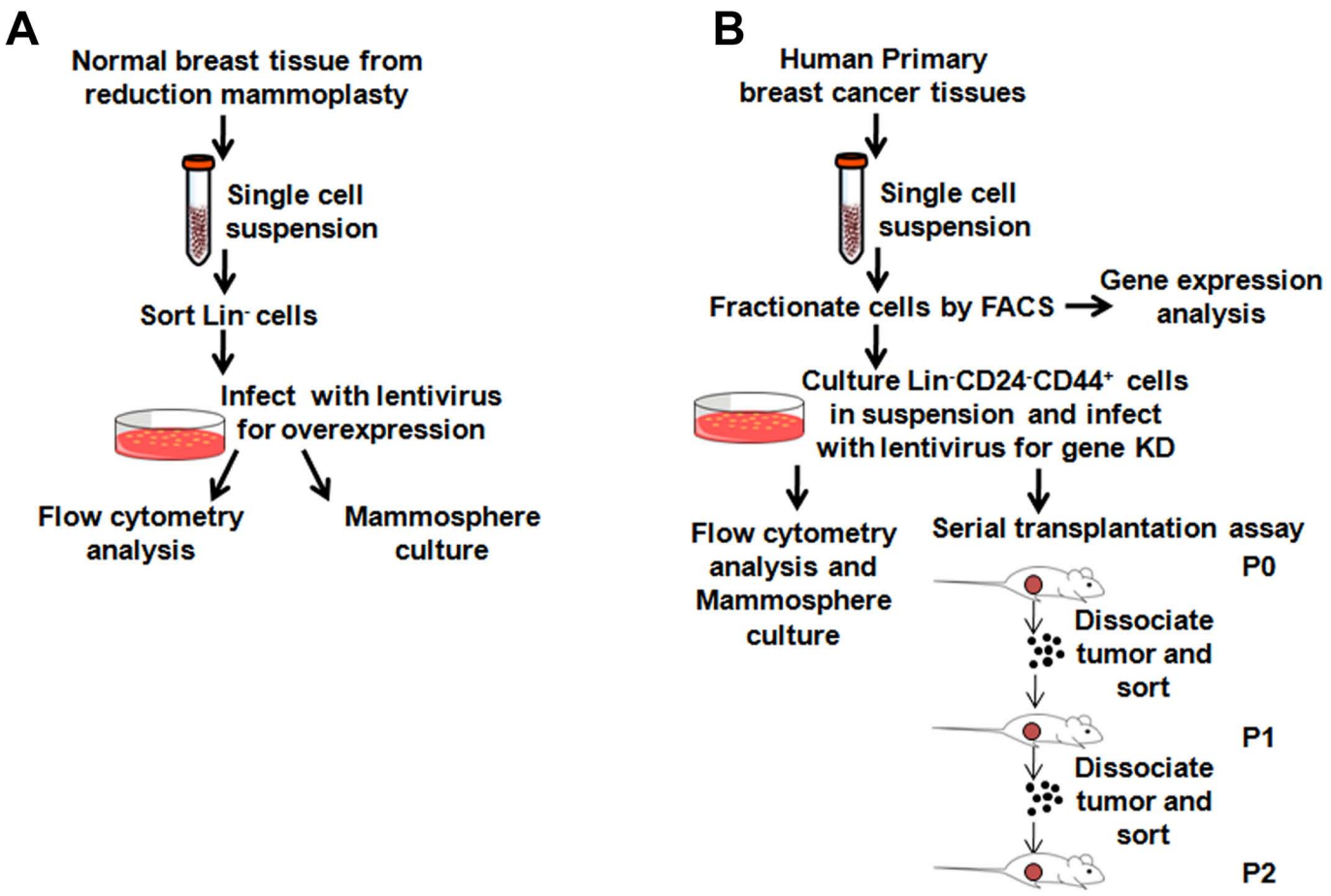
Luo et al., Figure S3. Rab2A overexpression and Q58H mutation drives BCSC expansion, Related to Figure 2. (A) Increased Rab2A copy number is associated with higher mRNA level in the breast cancer (TCGA, Provisional) (P = 1.56E-84). **(B)** Rab2A overexpression in HMLEs increased the CD24-CD44+ population and rescued the phenotypes inhibited by Pin1 KD. **(C-F)** Rab2A overexpression enhances cell migration, as measured by wound healing assay **(C, D)** and transwell migration assay **(E, F)**. **(G and H)** Rab2A overexpression potently increased colony formation in soft agar. * P<0.001. **(I)** Lentivirus-mediated overexpression of Flag-Rab2A and its Q58H mutant at levels similar to or 3 times over the endogenous level in HMLEs. Arrowhead, exogenous Flag-Rab2A; Arrow, endogenous Rab2A. **(J)** Overexpressed Rab2A Q58H mutant in HMLE cells at the endogenous level increased the CD24-CD44+ population as potently as Rab2Aoverexpressed at 3-time over the endogenous level. **(K)** The subcutaneous tumors in nude mice formed by HMLER cells infected with endogenous level of Q58H mutant grew faster than those infected with WT Rab2A.



Luo et al., Figure S4. Rab2A interacts with and activates Erk1/2 at the ERGIC, which may not be relevant to its vesice trafficking function, Related to Figure 4. (A) P-Erk1/2 co-localized with Rab2A overexpressed at three times of the endogenous level and Q58H mutant overexpressed at the endogenous level after EGF stimulation. **(B)** Treatment of 10 μg/ml BFA on vector control or Rab2A-overexpressing HMLEs for 0.5 h destroyed the ERGIC strcture, as measured by ERGIC53 staining. **(C)** BFA treatment, which blocked retrograde transportation, didn't affect Erk1/2 activation in either vector control or Rab2A-overexpressing HMLEs. Scale bars, 10 μm.



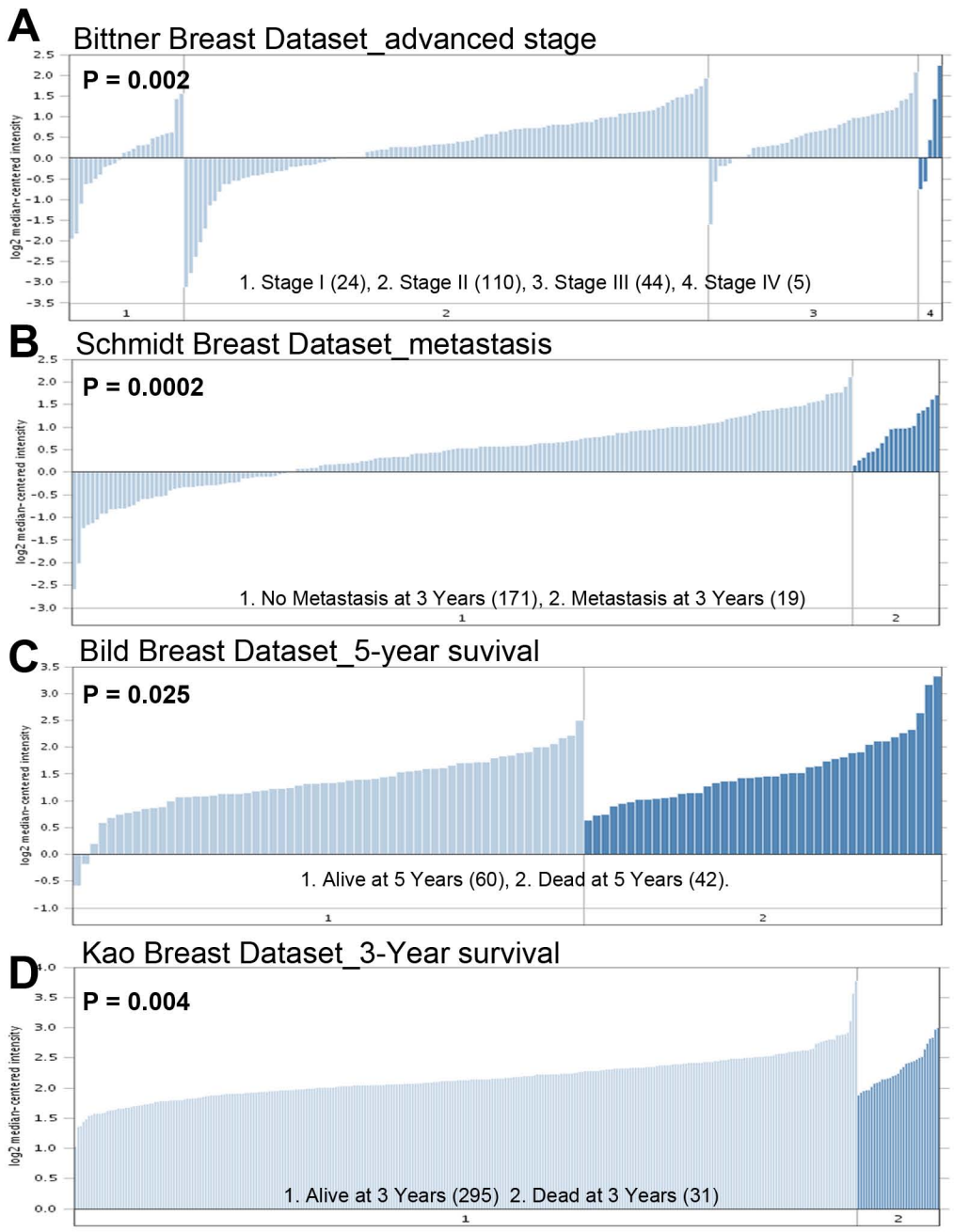
Luo et al., Figure S5. The binding of Rab2A to Erk1/2 is critical for Rab2A to promote Erk1/2 activation, Related to Figure 4. (A) Recombinant Erk1 or Erk2 interacted with GST-Rab2A directly. **(B)** Rab2A didn't compete with MEK1 to bind Erk1/2. Lysates of 293T cells transfected with decreasing doses of HA-AcMKP3 and a constant dose of Flag-Rab2A were immunoprecipitated with M2 (Flag) antibody. **(C and D)** Overexpression of Rab2A mutants with impaired binding to Erk failed to increase the abundance of CD24-CD44+ cells. **(G-I)** Ectopic expression of Flag-Rab1A in HMLE cells, as shown by immunoblot **(G),** didn't affect mammosphere formation **(H)** and the abundance of CD24-CD44+ cells **(I) (J and K)** Overexpressed Flag-Rab1A, which co-localized with ERGIC53 **(J)**, didn't promote Erk1/2 activation or co-localize with p-Erk1/2 **(K)**.



Luo et al., Figure S6. Schematic of the experiments on human tissues, Related to Figure 6.

(A) schematic of the experiments on normal human MECs from reduction mammoplasty tissues.

(B) schematic of the experiments on freshly isolated primary human BCSCs.



Luo et al., Figure S7. Correlation of Rab2A expression with clinical parameters in the oncomine microarray datasets, Related to Figure 7. (A) Rab2A expression correlates with advanced stage in Bittner Breast dataset (ductal breast carcinoma). **(B)** Rab2A expression correlates with metastatic event at 3 years in Schmidt Breast dataset (invasive breast carcinoma). **(C)** Rab2A expression correlates with death at 5 years in Bild Breast dataset (breast carcinoma). **(D)** Rab2A expression correlates with death at 3 year in Kao Breast dataset (breast carcinoma). Each bar in the graph represents the Rab2A level in one patient.

Table S1. Downregulated genes in Pin1 KO mammary epithelial cells.

Table S2. Upregulated genes in Pin1 KO mammary epithelial cells.

Table S3. Gene ontology analysis of dysregulated genes in Pin1 KO mammary and neuron cells comparing to WT cells.

Table S4. Genes that are downregulated in Pin1 KO mammary epithelial cells and are upregulated in normal or tumorigenic mammary stem cells.

Table S5. Patient information for isolation of $Lin^-CD24^-CD44^+$ cells from human breast cancer

Case No.	Age	Tumor size (cm)	ER	PR	HER	Grade	Lymp node	Stage	$Lin^-CD24^-CD44^+$
1	35	3.8×3.0	(+)	(+)	(-)	III	N1	II	1.35%
2	44	3.6 ×2.8	(+)	(-)	(+)	II	N2	III	28.61%
3	40	4.5 ×2.6	(-)	(-)	(-)	III	N2	III	14.00%
4	48	3.3 ×2.5	(-)	(-)	(+)	III	N1	II	5.22%
5	42	3.6 ×2.8	(+)	(+)	(-)	III	N2	III	4.61%
6	45	4.3 ×2.9	(+)	(-)	(-)	III	N2	III	2.07%
7	45	3.0 ×1.8	(+)	(-)	(-)	II	N1	II	1.99%
8	40	3.9 ×2.5	(-)	(-)	(-)	III	N1	II	25.54%

Notes: All eight patients were diagnosed as invasive ductal breast cancer. The tumor size is the original tumor size measured after surgery. ER, Estrogen receptor; PR, progesteron receptor; HER2, Human Epidermal Growth Factor Receptor 2. The hormonal receptor status was determined by the immunohistochemistry, which were done on the paraffin-embedded sections of the primary tumors. The lymph node information is according to the breast cancer TNM staging system. The $Lin^-CD24^-CD44^+$ cells were isolated for primary breast tumors.