Supplemental Information

Epigenetic Regulation of Epithelial-mesenchymal Transition in Trophoblast Stem Cells

Amy N. Abell, Nicole Vincent Jordan, Weichun Huang, Aleix Prat, Alicia A. Midland, Nancy L. Johnson, Deborah A. Granger, Piotr A. Mieczkowski, Charles M. Perou, Shawn M. Gomez, Leping Li, and Gary L. Johnson

Inventory of Supplemental Information

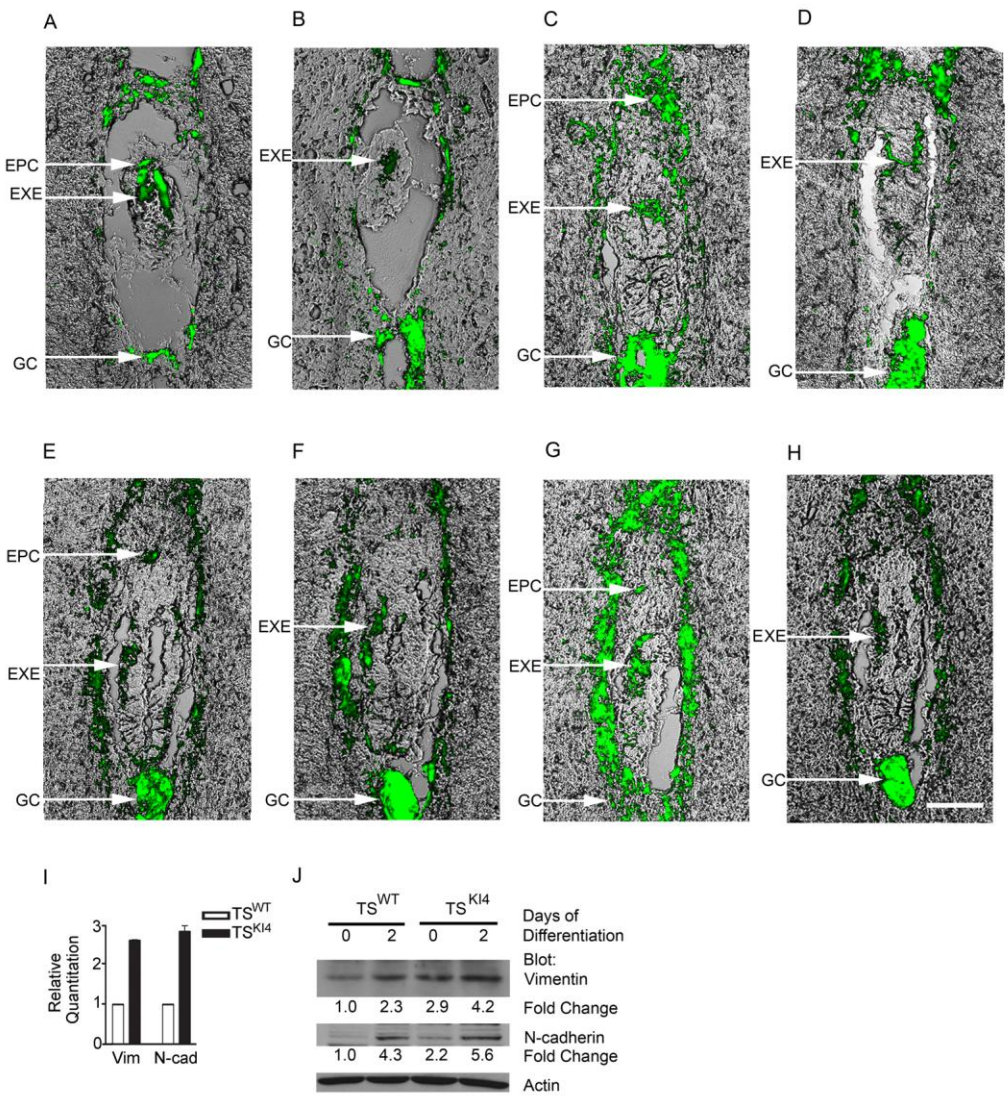


Figure S1, related to Figure 1. Developmental potency and mesenchymal properties of TS^{KI4} cells.

(A-H) Developmental potency of TS^{WT} and TS^{KI4} cells demonstrated by injection of GFP positive TS^{WT} and TS^{KI4} cells into wild-type blastocysts. Representative 18 μm section from TS^{WT} cell (A,B) or TS^{KI4} cell (C-H) E6.0 to E6.5 chimeras. (E-H) Serial sections through an E6.5 chimera generated by injection of GFP positive TS^{KI4} cells. EPC, ectoplacental cone; EXE, extraembryonic ectoderm; GC, giant cells. White bar equals 100 μm. (I) Increased expression of mesenchymal markers in TS^{KI4} cells relative to TS^{WT} cells as measured by qRT-PCR. Data shows the mean ± range of two independent experiments performed in triplicate. (J) Increased expression of mesenchymal markers in TS^{KI4} relative to TS^{WT} cells and with trophoblast differentiation as measured by Western blotting with the indicated antibodies. Fold changes are relative to undifferentiated TS^{WT} cells. Data shows representative blots from two independent experiments.

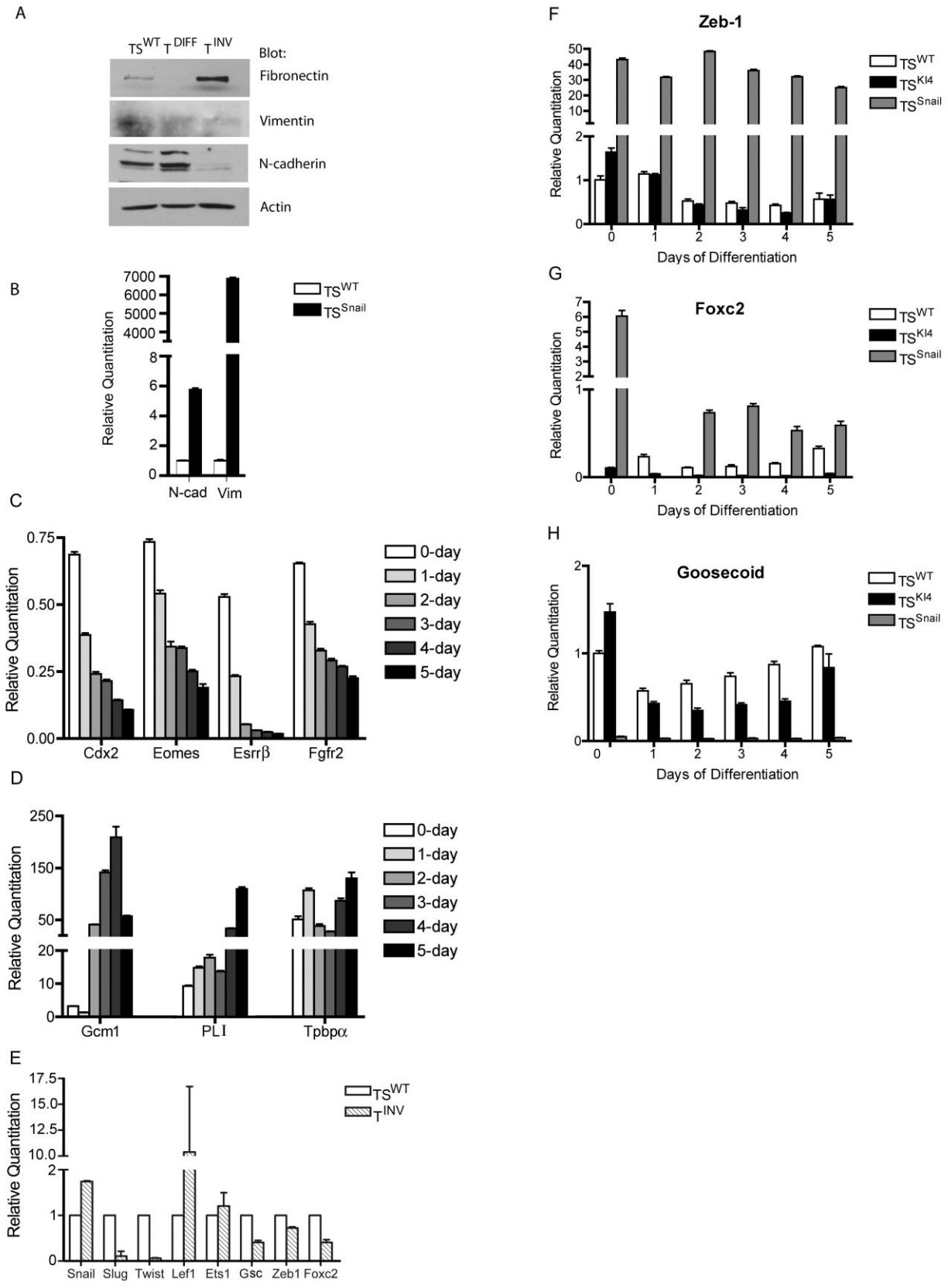


Figure S2, related to Figure 2. Expression of stem and mesenchymal cell markers, and EMT-inducing transcription factors in T^{INV}, TS^{KI4}, and TS^{Snail} cells.

(A) Increased expression of mesenchymal markers in T^{INV} cells relative to TS^{WT} cells as measured by Western blotting with the indicated antibodies. Data shows representative blots from two independent experiments. (B) Increased expression of mesenchymal markers in TS^{Snail} cells relative to TS^{WT} cells as measured by qRT-PCR. Data shows the mean ± range of two independent experiments performed in triplicate. (C) Expression of TS cell stemness markers in TS^{Snail} cells relative to TS^{WT} cells measured by qRT-PCR. Data shows the mean ± range of two independent experiments performed in triplicate. (D) Maintenance of multipotency in TS^{Snail} cells demonstrated by normal induction of trophoblast differentiation markers 0 to 5 days after removal of FGF4. Changes in message relative to TS^{WT} cells were measured by qRT-PCR. Data shows the mean ± range of two independent experiments performed in triplicate. (E) Expression of EMT-inducing transcription factors in TS^{WT} and T^{INV} cells measured by qRT-PCR. Data shows the mean ± range of two independent experiments. (F-H) Expression of indicated EMT-inducing transcription factors 0 to 5 days after differentiation of TS^{WT}, TS^{KI4}, and TS^{Snail} cells measured by qRT-PCR. Data shown are the mean ± range of two independent experiments.

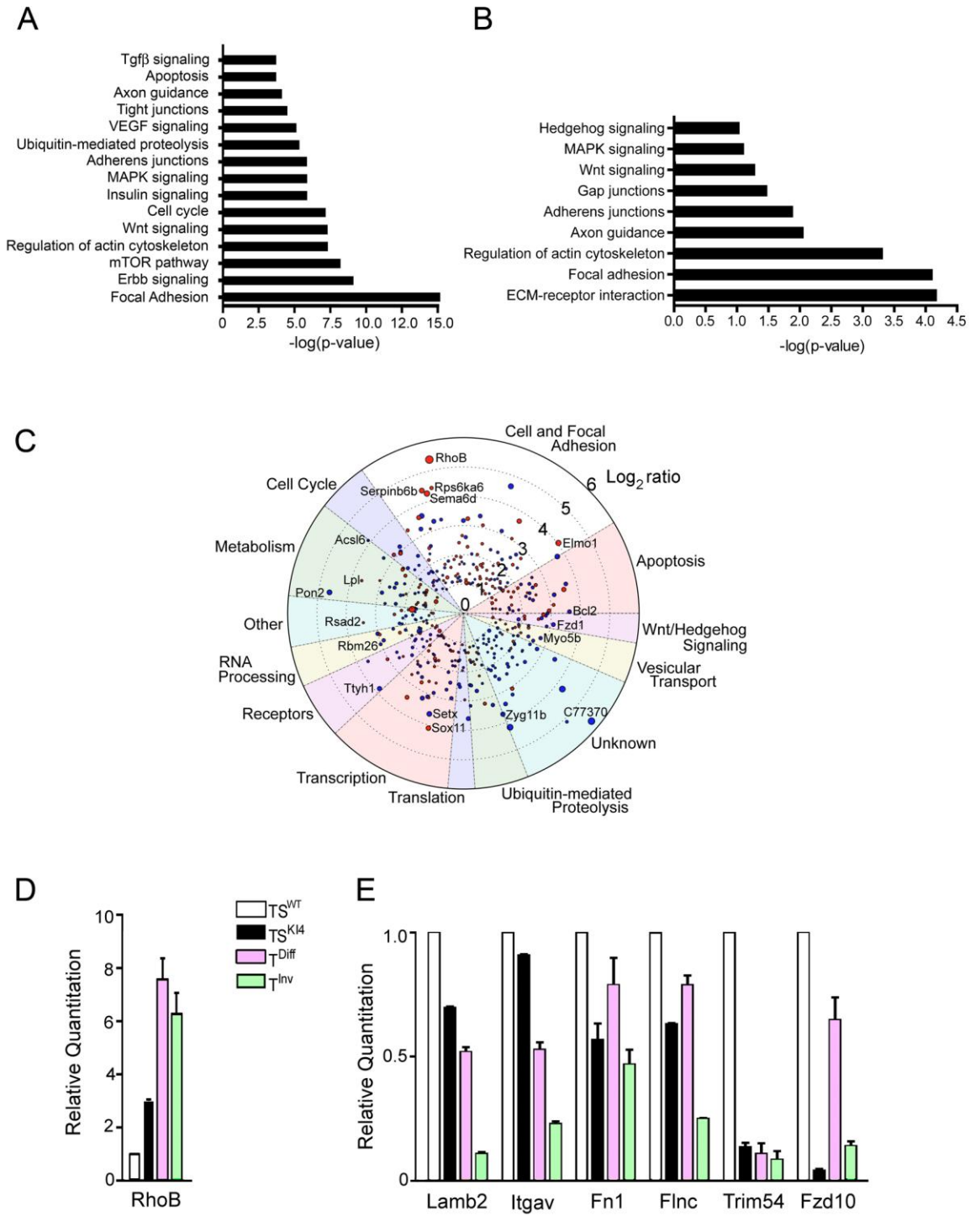


Figure S3, related to Figure 3. Intersecting gene expression changes in focal adhesion and actin cytoskeleton pathways in T^{INV} and TS^{KI4} cells.
(A) Enrichment for focal adhesion GO category in T^{INV} cells. Canonical KEGG pathways are plotted according to p-values of pathway enrichment. (B) Enrichment for focal adhesion and actin cytoskeleton GO categories in TS^{KI4} cells. The p-values for canonical KEGG pathways are shown. (C) Diagram shows T^{INV} cell $\log_2$ ratios for the intersecting 416 genes with TS^{KI4} cells categorized according to biological function. Red

and blue indicate upregulated and downregulated genes, respectively. Circle diameter shows the relative difference in log_2 ratios in T^{INV} cells versus TS^{KI4} cells; abs(T^{INV}-TS^{KI4}). (D,E) Validation of gene expression changes by qRT-PCR in TS^{WT}, TS^{KI4}, T^{DIFF}, and T^{INV} cells. Data represent gene expression changes relative to TS^{WT} cells expressed as the mean ± range for two independent experiments.

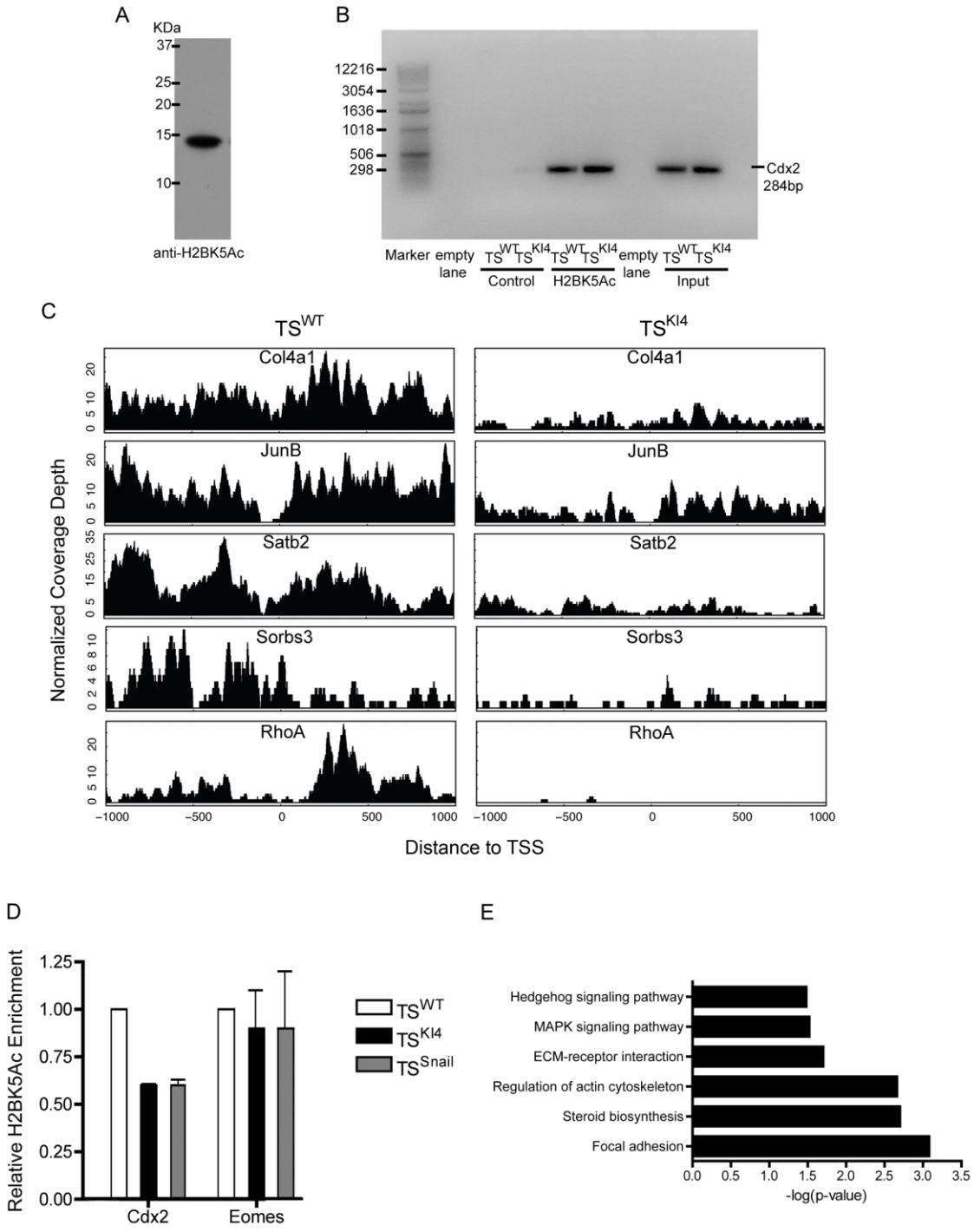


Figure S4, related to Figure 5. Validation of H2BK5Ac ChIP-seq analysis.
(A) Validation of the H2BK5Ac antibody used for ChIP-seq**.** Specificity of anti-acetylated H2BK5 antibody was measured by Western blotting of acid-extracted histone lysates from TS^{WT} cells using anti-H2BK5Ac antibody. (B) Specificity of anti-acetylated H2BK5 antibody by ChIP. Semi-quantitative RT-PCR for the promoter of the TS cell marker Cdx2 from ChIP samples isolated from TS^{WT} or TS^{KI4} cells is shown. Results are representative of three independent experiments. (C) Promoter acetylation density plots

of additional genes with significant loss of H2BK5 acetylation. Comparison of normalized coverage depth of acetylation density within 1 kb of the transcription start site (TSS) between TS^{WT} and TS^{KI4} cells for indicated genes with p-values $\leq 1.4e^{-14}$. The x-axis for each subplot is the distance to the TSS, and the y-axis is the normalized coverage depth. (D) Validation of H2BK5Ac enrichment on promoters of TS cell stemness markers, Cdx2 and Eomes using H2BK5Ac ChIP-qRT-PCR. Data represent the mean $\pm$ range of two independent experiments performed in triplicate. (E) GO categories enriched for genes with altered H2BK5Ac in TS^{KI4} cells. The p-values of canonical KEGG pathways are shown.

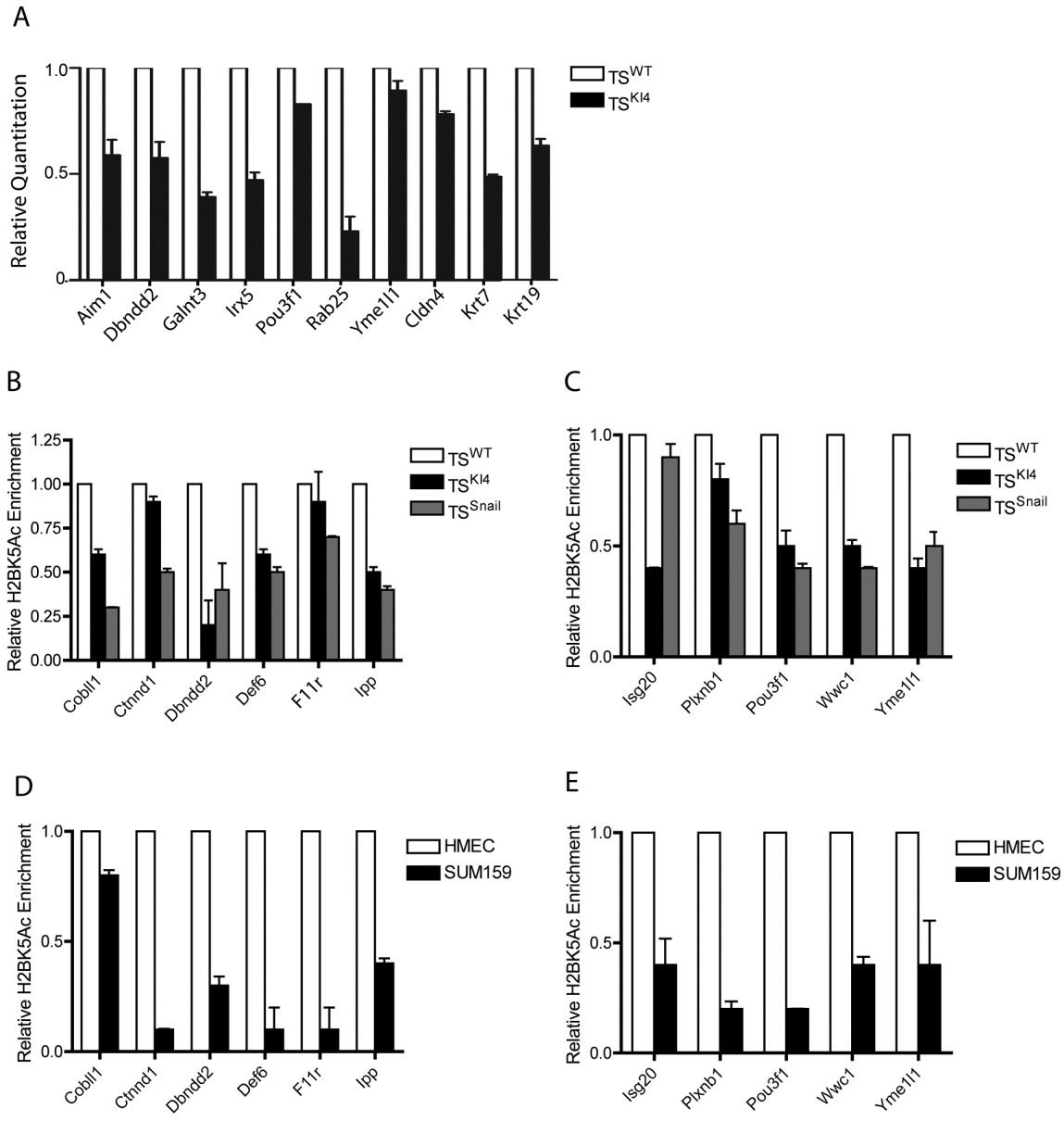


Figure S5, related to Figure 6. Reduced gene expression and H2BK5 acetylation in intersecting gene sets between TS^{KI4}, TS^{Snail}, and CL cells.
(A) Validation of reduced expression of TS^{KI4}/CL intersecting genes in TS^{KI4} cells relative to TS^{WT} cells measured by qRT-PCR. (B-E) Validation of loss of H2BK5Ac on promoters of TS^{KI4}/CL intersecting genes measured by H2BK5Ac ChIP coupled to qRT-PCR. Data shows loss of H2BK5Ac from shared set of genes between TS^{KI4}, TS^{Snail}, and CL cells. (A-E) Data represent the mean $\pm$ range of two independent experiments performed in triplicate.

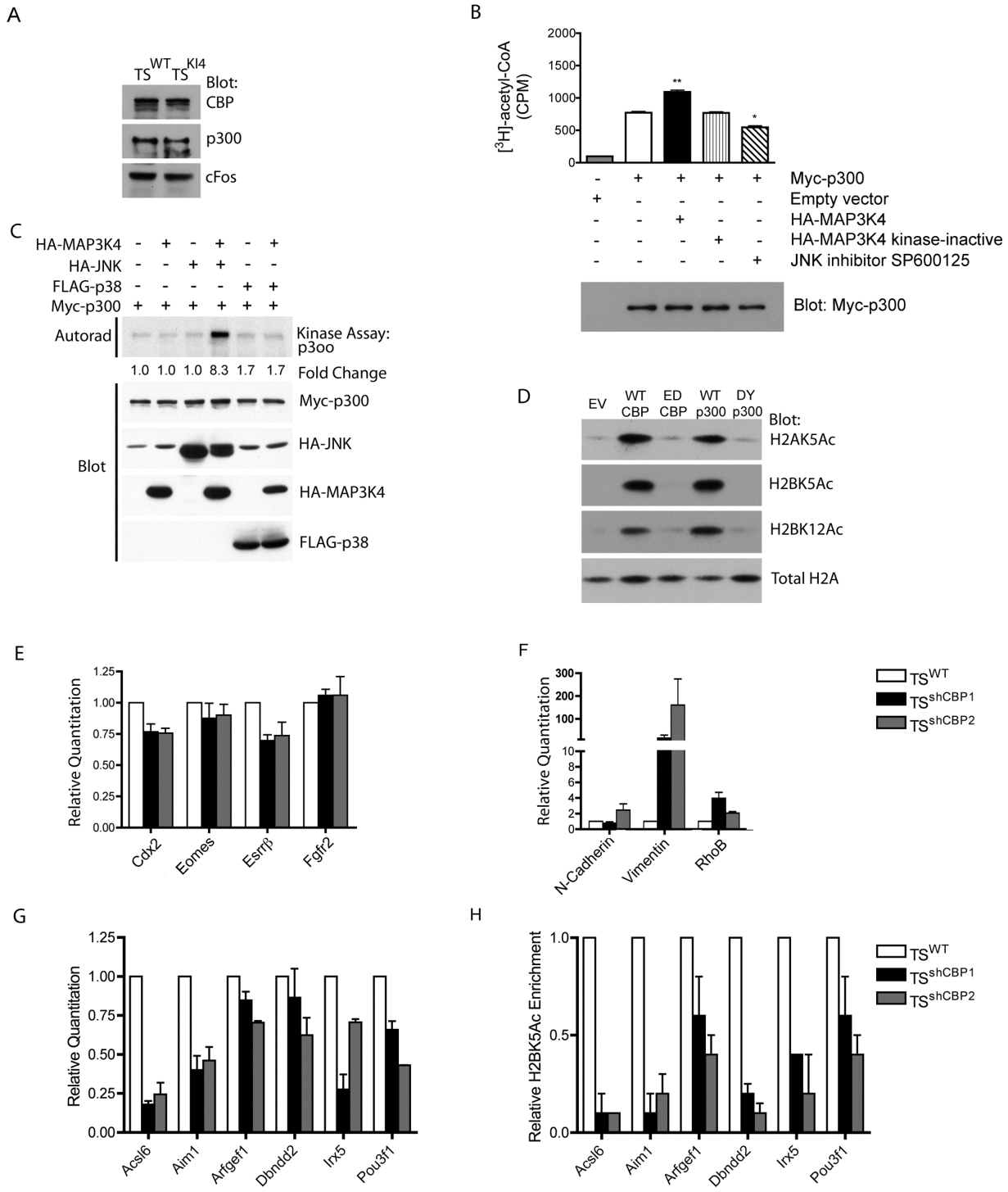


Figure S6, related to Figure 7. TS^{shCBP} cells display properties of EMT and stemness with reduced expression and H2BK5 acetylation of TS^{KI4}/CL intersecting genes.

(A) CBP and p300 protein expression levels are unchanged in TS^{KI4} cells relative to TS^{WT} cells. Western blots were performed on nuclear lysates from either TS^{WT} or TS^{KI4} cells using anti-CBP or anti-p300 antibodies. Results are representative of two independent experiments. (B) MAP3K4/JNK increases HAT activity of p300 measured

as incorporation of [^{3}H]-acetyl-CoA in counts per minute (CPM) in 293 cells co-expressing p300 with the constitutively active kinase domain of MAP3K4 or kinase-inactive. Significance of change in HAT activity was evaluated by an unpaired Student's T-test. **p-value < 0.01; *p-value < 0.05. Data represents the mean ± SEM of three independent experiments performed in triplicate. (C) MAP3K4 and JNK promote phosphorylation of p300 as measured by kinase assay. γ^{32}P-ATP labeled proteins were visualized by autoradiography and quantified by phosphorimaging. Protein expression levels were monitored by Western blotting with the indicated antibodies. Data are representative of two independent experiments. (D) CBP and p300 HAT activity is required to selectively acetylate histones H2A and H2B as HAT deficient CBP and p300 mutants are unable to acetylate H2A and H2B. Western blots were performed using acid-extracted histone cell lysates from 293 cells expressing wild-type (WT) or HAT-inactive CBP (ED) and p300 (DY) and probed with the indicated anti-acetylated H2A and H2B antibodies. Results are representative of two independent experiments. (E) Expression of TS cell stemness genes in TS^{shCBP} cells compared to control virus infected cells measured by qRT-PCR. (F) Increased expression of mesenchymal markers in TS^{shCBP} cells relative to control virus infected cells as measured by qRT-PCR. (G) Reduced expression in TS^{shCBP} cells of genes common to TS^{KI4}, TS^{Snail} and CL cells as measured by qRT-PCR. (H) Reduced H2BK5Ac on the promoters of the same genes in G shown by H2BK5Ac ChIP-qRT-PCR. (E-H) Data shows the mean ± range of two independent experiments performed in triplicate.

Table S5 Phenotypic similarity between MAP3K4 (KI4) and CBP/p300 -/- mice

Gene	Skeletal	Neural Tube	Cranio-facial	Embryonic Lethality	Growth Retardation	Cardiac	Ref.
MAP3K4	+	+	+	+	+	-	1,2,4
CBP	+	+	+	+	+	-	11,12
P300	+	+	-	+	+	+	15
TRAF4	+	+	-	+	-	-	9
GCN5	+	+	-	+	+	-	3,7,13,14
Dvl1/2	+	+	-	+	-	+	5
JNK1/2	-	+	-	+	+	-	6,10
TWIST	-	+	+	+	+	-	8
P38IP	-	+	-	+	+	-	17
AXIN	+	+	-	+	+	+	16
PCAF	-	-	-	-	-	-	14

Yellow indicates phenotypic overlap of the MAP3K4 (KI4) mouse with the knockout mice for the genes listed. + indicates presence of the phenotype specified. - indicates absence of the phenotype specified. Table is based upon comprehensive literature review, as indicated by the references below.

1. Abell, A. N., Granger, D. A., Johnson, N. L., Vincent-Jordan, N., Dibble, C. F., and Johnson, G. L. (2009). Trophoblast stem cell maintenance by fibroblast growth factor 4 requires MEKK4 activation of Jun N-terminal kinase. Mol Cell Biol *29*, 2748-2761.
2. Abell, A. N., Rivera-Perez, J. A., Cuevas, B. D., Uhlik, M. T., Sather, S., Johnson, N. L., Minton, S. K., Lauder, J. M., Winter-Vann, A. M., Nakamura, K.*, et al.* (2005). Ablation of MEKK4 kinase activity causes neurulation and skeletal patterning defects in the mouse embryo. Mol Cell Biol *25*, 8948-8959.
3. Bu, P., Evrard, Y. A., Lozano, G., and Dent, S. Y. (2007). Loss of Gcn5 acetyltransferase activity leads to neural tube closure defects and exencephaly in mouse embryos. Mol Cell Biol *27*, 3405-3416.
4. Chi, H., Sarkisian, M. R., Rakic, P., and Flavell, R. A. (2005). Loss of mitogen-activated protein kinase kinase kinase 4 (MEKK4) results in enhanced apoptosis and defective neural tube development. Proc Natl Acad Sci U S A *102*, 3846-3851.
5. Hamblet, N. S., Lijam, N., Ruiz-Lozano, P., Wang, J., Yang, Y., Luo, Z., Mei, L., Chien, K. R., Sussman, D. J., and Wynshaw-Boris, A. (2002). Dishevelled 2 is essential for cardiac outflow tract development, somite segmentation and neural tube closure. Development *129*, 5827-5838.
6. Kuan, C. Y., Yang, D. D., Samanta Roy, D. R., Davis, R. J., Rakic, P., and Flavell, R. A. (1999). The Jnk1 and Jnk2 protein kinases are required for regional specific apoptosis during early brain development. Neuron *22*, 667-676.
7. Lin, W., Zhang, Z., Srajer, G., Chen, Y. C., Huang, M., Phan, H. M., and Dent, S. Y. (2008). Proper expression of the Gcn5 histone acetyltransferase is required for neural tube closure in mouse embryos. Dev Dyn *237*, 928-940.
8. Ota, M. S., Loebel, D. A., O'Rourke, M. P., Wong, N., Tsoi, B., and Tam, P. P. (2004). Twist is required for patterning the cranial nerves and maintaining the viability of mesodermal cells. Dev Dyn *230*, 216-228.

9. Regnier, C. H., Masson, R., Kedinger, V., Textoris, J., Stoll, I., Chenard, M. P., Dierich, A., Tomasetto, C., and Rio, M. C. (2002). Impaired neural tube closure, axial skeleton malformations, and tracheal ring disruption in TRAF4-deficient mice. Proc Natl Acad Sci U S A *99*, 5585-5590.
10. Sabapathy, K., Jochum, W., Hochedlinger, K., Chang, L., Karin, M., and Wagner, E. F. (1999). Defective neural tube morphogenesis and altered apoptosis in the absence of both JNK1 and JNK2. Mech Dev *89*, 115-124.
11. Tanaka, Y., Naruse, I., Hongo, T., Xu, M., Nakahata, T., Maekawa, T., and Ishii, S. (2000). Extensive brain hemorrhage and embryonic lethality in a mouse null mutant of CREB-binding protein. Mech Dev *95*, 133-145.
12. Tanaka, Y., Naruse, I., Maekawa, T., Masuya, H., Shiroishi, T., and Ishii, S. (1997). Abnormal skeletal patterning in embryos lacking a single Cbp allele: a partial similarity with Rubinstein-Taybi syndrome. Proc Natl Acad Sci U S A *94*, 10215-10220.
13. Xu, W., Edmondson, D. G., Evrard, Y. A., Wakamiya, M., Behringer, R. R., and Roth, S. Y. (2000). Loss of Gcn5l2 leads to increased apoptosis and mesodermal defects during mouse development. Nat Genet *26*, 229-232.
14. Yamauchi, T., Yamauchi, J., Kuwata, T., Tamura, T., Yamashita, T., Bae, N., Westphal, H., Ozato, K., and Nakatani, Y. (2000). Distinct but overlapping roles of histone acetylase PCAF and of the closely related PCAF-B/GCN5 in mouse embryogenesis. Proc Natl Acad Sci U S A *97*, 11303-11306.
15. Yao, T. P., Oh, S. P., Fuchs, M., Zhou, N. D., Ch'ng, L. E., Newsome, D., Bronson, R. T., Li, E., Livingston, D. M., and Eckner, R. (1998). Gene dosage-dependent embryonic development and proliferation defects in mice lacking the transcriptional integrator p300. Cell *93*, 361-372.
16. Zeng, L., Fagotto, F., Zhang, T., Hsu, W., Vasicek, T. J., Perry, W. L., 3rd, Lee, J. J., Tilghman, S. M., Gumbiner, B. M., and Costantini, F. (1997). The mouse Fused locus encodes Axin, an inhibitor of the Wnt signaling pathway that regulates embryonic axis formation. Cell *90*, 181-192.
17. Zohn, I. E., Li, Y., Skolnik, E. Y., Anderson, K. V., Han, J., and Niswander, L. (2006). p38 and a p38-interacting protein are critical for downregulation of E-cadherin during mouse gastrulation. Cell *125*, 957-969.

Supplemental Experimental Procedures

293T culture conditions and transfections

293T cells were cultured in Dulbecco's modified Eagle's high glucose medium with 10% FBS and 1% PS. Transfection of HEK293 cells was performed in 10 cm dishes for 24 hours with LipofectaminePlus (Invitrogen) according to manufacturer's specifications. The Snail expression plasmid was a kind gift from Dr. Paul Wade. Snail was subcloned between EcoRI and XhoI restriction sites of pcDNA3.1(+). G418 was used to stably select for TS cells overexpressing Snail. Published plasmids are specified below under **Plasmids and sources.**

Lentivirus production, infections of TS cells and shRNA knockdown of CBP and p300

To produce replication incompetent lentivirus, 293T cells were cotransfected with either GFP or shRNA constructs in pLKO.1 (Open Biosystems) in combination with pMD2.G and psPAX2 (Addgene) using Lipofectomine Plus reagent (Invitrogen). 24 hours later, cell media was changed. Viral supernatants were harvested at 48 and 72 hours post-transfection by ultracentrifugation. Viral pellets were resuspended in 100 μl 30% TS media (RPMI 1640, 20% fetal bovine serum, 1% penicillin and streptomycin, 1% L-glutamine, 1% sodium pyruvate and 100 μM β-mercaptoethanol) and 70% MEF conditioned TS cell media, supplemented with FGF4 (37.5 ng/ml) and Heparin (1 mg/ml) and stored at -80° C. Infection of TS cells was performed essentially as previously described (Odiatis and Georgiades, 2010). Briefly, TS^{WT} cells were plated overnight at 1 X 10^4 cells per 24 mm well in a Nunc 4-well dish, and then infected with 100 μl of lentivirus and polybrene (5 μg/ml). The next day, cells were washed five times in PBS and TS media was replaced. Three days after infection, cells were examined for

GFP positive cells as a indicator of transduction efficiency, showing a 40-50% infection efficiency. Puromycin (2 μg/ml) was used to select for transduced cells. Selection was complete after seven days of puromycin treatment. Knockdown of CBP or p300 was measured by qRT-PCR as described in the main methods under **Real-time Quantitative RT-PCR**.

Generation and analysis of chimeras

GFP expressing TS^{WT} and TS^{KI4} cells generated as described above for **Lentivirus production.** These cells were used to generate chimeric embryos. The Animal Models Core at the University of North Carolina at Chapel Hill performed blastocyst injections using standard procedures. Conceptuses were dissected at E6.5 and E9.5, where date of blastocyst injection was considered E3.5. Samples were fixed for one hour in PBS containing 4% paraformaldehyde and 1% sucrose, washed, and put through a sucrose gradient. After embedding and freezing in OCT, samples were cryosectioned at 18 μm sections. All animal and embryo work was performed according to both university and federal guidelines for the use of animals.

Invasion assays and isolation of invasive trophoblasts (T^{INV})

Invasion assays were performed essentially as previously described (Abell et al., 2009). Briefly, cells were plated on GF-reduced Matrigel (BD Biosciences) coated 8 μM pore transwell chambers. After 48 hours, invasion assays were terminated. Non-invading cells were removed from the top of the transwells by washing and swabbing. Invaded cells (T^{INV}) were isolated from the bottom of the transwell chambers by either direct lysis in RNAeasy buffer for RNA isolation or by trypsinization for immunofluorescence, ChIP, and Western blotting. For quantitation of invasion, chambers were fixed in 3%

paraformaldehyde for ten minutes and stained with DAPI. For each transwell, five 10X fields were imaged and counted.

Immunofluorescence

Immunostaining was performed as previously described (Abell et al., 2009). Briefly, TS cells cultured on glass coverslips for two days were fixed in 3% paraformaldehyde, 1% Sucrose, 1% PBS for 10 min. Coverslips were washed with PBS, permeabilized for 5 min with 0.2% Triton and blocked for one hour with 10% goat serum. Coverslips were incubated for one hour at RT with anti-E-cadherin antibody (1:1000), washed for 30 min and then incubated at RT for one hour with Alexa 555 anti-mouse secondary antibody (Invitrogen) and DAPI.

Agilent gene expression arrays

For gene arrays, cDNA was competitively hybridized to a 4 x 24,000 Agilent gene array. Raw gene expression data was analyzed using Agilent GeneSpringGX software to produce a ranked gene list based upon the $\log_2$ ratio of Agilent probe intensity appearing on at least three gene arrays with a Benjamini-Hochberg p-value < 0.05. Heat maps were prepared by hierarchical clustering in Cluster 3.0 and JavaTreeView. Published breast cancer cell line data from Neve et al. was clustered according to the methods described above (Neve et al., 2006).

***In vitro* histone acetyltransferase (HAT), immunoprecipitation, and kinase assays**

HAT assays were performed using Millipore (cat. # 17-289) HAT assay kit. Briefly, HA-CBP and/or Myc-p300 were immunoprecipitated in PBS with 5 μg anti-HA or anti-myc antibodies specified below under **Antibodies and sources**. Following immunoprecipitation, CBP/p300-bound protein G sepharose beads were incubated for 30 min at 30° C in a reaction mixture containing 100 μM acetyl-CoA, 1.4 μCi [^{3}H]-acetyl-

CoA (specific activity 200 mCi/mmol), 10 μg purified core histones and HAT buffer provided by the manufacturer (Millipore). The reaction mixture was spotted onto P18 filter paper, washed with sodium phosphate buffer (pH 9) and quantitated with a Beckman liquid scintillation counter. Total HAT activity was measured from nuclear lysates according to the same method. Kinase assays were performed as previously described (Abell et al., 2007). Briefly, HA-CBP or Myc-p300 immunoprecipitates were incubated for 20 min at 30° C with HA-MAP3K4 and HA-JNK or FLAG-p38 in kinase buffer containing 0.5 mM ATP and 10 μCi γ- ^{32}P[ATP]. Proteins were separated by SDS-PAGE and quantitated by autoradiography using a Phosphorimager.

Cdx2 PCR conditions for ChIP-seq validation

PCR for TS cell marker Cdx2 was carried out at 61^{o}C for 35 cycles to generate a 284 bp product. The following primers were used for semi-quantitative RT-PCR: sense 5'-CATCCCCGCCTCTACAGCTTA-3' and 5'-ATACATGCTCACGTCCTTGTCC-3'.

Real-time PCR for ChIP

ChIP assays were quantified by real-time PCR using Absolute Blue SYBR green PCR mix (Thermoscientific) and the Applied Biosystems Fast 7500 Real-Time PCR System. Fold enrichment was determined by the $2^{-\Delta CT}$ method. PCR primers were designed to amplify approximately 75-100 bp fragments from genomic DNA using Primer Express 3.0 (Applied Biosystems). Primers used for ChIP-qRT-PCR include the following:

Gene Name	Forward Primer	Reverse Primer
Acsl6 (mouse)	ACCCCCGCTGCCATTT	GACCGGCCTCCAAGGTTT
Itgav (mouse)	AAAGTCCCGCCGAGTATGC	AGAAGTCCACGGCGAATCC
Lamb2 (mouse)	TTCCCTTGCTTTCCAACTTTACTC	GCAGTCACTCCTCCACAATCTG
Trim54 (mouse)	TAGCCGGGATCCCAAGAAT	TGCCTGTGCACCCTTCCT

Cdx2 (mouse)	TGGCTCCGCAGAACTTTGTC	CCGCCACGTGGTAACCAA
Eomes (mouse)	GCCTTCCACCTTTGATGTATCC	AAAGCTTTGGCGCCTTCTCT
Aim1 (mouse)	CCAGCTTTGCACTTGTTGCA	GCACGCTTGGCTCTCAGTTT
Arfgef1 (mouse)	GGCAAGAAGACGAAGAAGAACATGTT	ACTTCCTTGTCGGCCAGAATC
Galnt3 (mouse)	CCTTAAGCGACTAGTAAAATTGCACAT	TGACTGCACCCAGCTTCCA
Irx5 (mouse)	AACCAGACGGTGTTGAATCGT	GGCTCCGCAACATTTTCG
Krt7 (mouse)	CCGCATGGAGGAATAAAAGG	GAGGAGCAGCAGCGCGAACTC
Krt19 (mouse)	AGTCCGCGGTGGAAGTTTTA	CCAGACAGCAGCCCATCAG
Rab25 (mouse)	TCTCCACCCGCACTGTAATG	GTCCCAGATCTGTGCCTTGAC
Scyl3 (mouse)	AAGGCTTCGGTGCCAGAGT	TGGAAACTGGCCTCAGTCAAG
Cobll1 (mouse)	TCAGACACCTGACACTGAACTTTTT	AAAAGGTTCCTATGTGTCAGGATGA
Ctnnd1 (mouse)	CGCCGCTCCTCTTCCA	CTACGACGAGCTGGGCAATT
Dbndd2 (mouse)	CCCTGGAGCGCCAACA	GGCTGTAAAATGTCCTCAAAGAACT
Def6 (mouse)	GCACTTCCGGGATGACGAT	TGAGGTAGGGCATGTAACCTTGA
F11r (mouse)	AGGACCGATTGGTTGGAAAAG	GCACCCTGAGGAAAACTGTCA
Ipp (mouse)	GACAAACATGCCCAGCTCATC	TCACAGAAATGCTGTCCACTTCTC
Isg20 (mouse)	CGCTGCAGCATTGTGAACA	CGGGTCGGATGTACTTGTCA
Plxnb1 (mouse)	TGCAGAGGAAGGCCGTGTAC	ATACTCGCCCATCCCATGAG
Pou3f1 (mouse)	GCCCACCGCAGGATCTC	TTCTGGAGGTCCCTGCTGTAG
Wwc1 (mouse)	GACGGCAAGGTCTACTACATAGATCA	CCGCGGGTCGATCCA
Yme1l1 (mouse)	GTGTTACCCTGCGTCAGAAATG	CTATAACAGCATGTTCATATTCAC AAAATC
Aim1 (human)	AACGCGGCCAGCGATA	CCGCGGAGATTTCACTTTCT
Arfgef1 (human)	GGCCGACAAGGAAGTGAAGA	CCACCTCGCAAGCTTTGC
Galnt3	GCTTGGTGCAGTAATTTTTTTCTTT	TTTCCATCCTTGATTCCTCTTTG

(human)		
Irx5 (human)	CCATTGCCTTGGTCAAAAATG	GAGGCGGACGGCTGGTA
Krt7 (human)	CAGATCAAGACCCTCAACAACAAG	GTCCAGTCCCGCTCACCTT
Krt19 (mouse)	AAGCTAACCATGCAGAACCTCAA	CAGGGCGCGCACCTT
Rab25 (human)	TCTCCACCCGCACTGTGAT	TGTCCCAGATCTGAGCCTTGA
Scyl3 (human)	ACACTGAGAGAACCACCATTTACCT	TCTTGCAGTACAGCGGGATAAAC
Cobll1 (human)	CCAAAGCTCTGACCCAGAACA	TCTCCCGAACGGATTGCA
Ctnnd1 (human)	CGACCGCCAGCATCTTG	CCCGGGTCAGCTTCTCAAA
Dbndd2 (human)	GGCAAAAATTCTTCGAGGACAT	GATGCAGATGGGACAGAGGAA
Def6 (human)	GCCCTGCGCAAGGAACT	ACGTCCAGCGCGGTAAAG
F11r (human)	CCTGGCATTGGGCAGTGT	TTCTCAGGAATTCTGACTTCAGGTT
Ipp (human)	AGGCTGCTGATAGTCCTTTTTCA	CATCTTATTGATTTGGGCCAAGA
Isg20 (human)	TCGTTGCAGCCTCGTGAA	GGCCGGATGAACTTGTCGTA
Plxnb1 (human)	CCCCTTCCACCAACTGCAT	TTGCCAGGTGCTGCAGATAC
Pou3f1 (human)	CCGAATTTCTGACCCATCTCTATT	ACAAACAGAAGAAGAAACGGAGAAG
Wwc1 (human)	TCGACGGCAAGGTCTACTACATAG	CCGCGGGTCGATCCA
Yme1l1 (human)	CGGTGCAACCCCAGGTTA	TGGTGTATGGAAGGCATTGATG

Comparative expression and GO pathway analysis

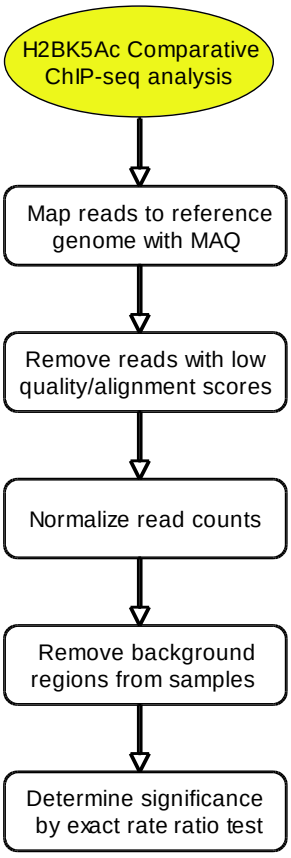
Comparative gene expression analysis was used to determine the T^{INV} gene signature by filtering for genes enriched in the T^{INV} versus T^{DIFF} statistically analyzed gene sets (Benjamini-Hochberg p-value < 0.05 and a $\log_2$ ratio $\geq$ +/- 1). Enrichment was quantified by the absolute value of the $\log_2$ ratios between T^{INV} and T^{DIFF} gene sets using a secondary cutoff $\log_2$ ratio $\geq$ 0.5. The T^{INV} gene signature was assessed for

overlap with the TS^{KI4} statistically analyzed gene set to produce the minimal trophoblast EMT signature. The NIH DAVID gene ontology bioinformatics database was used to cluster genes into canonical KEGG signaling pathways from both the ChIP-seq and Agilent array datasets of statistically significant genes with a p-value < 0.05 and a $\log_2$ ratio $\leq$ -1.

Microarray analyses of human breast tumors and cell lines. Two publicly available gene expression microarray datasets (UNC337 [GSE18229] for tumors and Neve et al. for cell lines) were analyzed as previously described (Prat et al., 2010). For human tumors, we obtained the CL human tumor expression signature of up- and down-regulated genes from Supplemental Data.xls in Prat et al. (http://breast-cancer-research.com/content/supplementary/bcr2635-s2.xls). For human cell lines, we performed a Significance Analysis of Microarrays (SAM) by comparing the previously described CL cell lines versus the rest (false discovery rate <5%). The resulting up- and down-regulated gene sets defined the CL cell line signature analyzed in this manuscript. Finally, overlap between two signatures was estimated using the exact hypergeometric probability implemented in R package (http://cran.r-project.org). In each overlap analysis, only genes present in both platforms were considered.

ChIP-seq bioinformatics analysis

Flow chart of bioinformatics analyses



Analysis tools

For both ChIP-seq and mRNA-seq data analysis, we used MAQ (Li et al., 2008) to map read tags to the mouse reference genome (NCBI build 37.1/mm9), and used EpiCenter (manuscript in preparation) to identify significant changes between TS^{WT} and TS^{KI4} cells. EpiCenter is a versatile tool for comparative analysis of both ChIP-seq and mRNA-seq data to identify differential changes. EpiCenter is freely available at http://www.niehs.nih.gov/research/resources/software/epicenter.

Read mapping and quality filtering

In the analysis of H2BK5Ac ChIP-seq data, read tags were mapped directly to the mouse reference genome. Reads with MAQ alignment quality scores <10 (i.e. number of mismatches > 2) were filtered out to remove poorly-aligned reads or non-uniquely mapped reads. After the quality filter, about 38.7 (87.8%) and 24.1 (88.9%) million reads were uniquely mapped to the mouse reference genome in TS^{WT} and TS^{KI4} cells, respectively.

In the analysis of gene expression mRNA-seq data, read tags were directly mapped to cDNA sequences of all mouse RefSeq genes. After filtering out reads with alignment quality scores <10, about 41.6 and 33.6 million reads were uniquely mapped to the RefSeq cDNA sequences. We also mapped the reads to the mouse reference genome, but we had about 20% fewer mapable reads because reads from exon junctions could not be mapped back to the reference genome. The results reported here were based on cDNA sequence mapping data.

Data normalization

Since H2BK5Ac ChIP-seq data showed broad change between TS^{WT} and TS^{KI4} samples, we chose the ratio of the numbers of mapped reads to adjust the difference of read coverage depths. We also used EpiCenter's linear least squares regression method to estimate the read coverage ratio for normalization. The 300 most significant genes were very similar between the two approaches, although fewer genes were identified with the regression-based approach. For mRNA-seq data analysis, we used EpiCenter's "parsimony" normalization procedure to adjust read coverage difference. The "parsimony" automatically adjusted normalization ratio to minimize the number of significant genes by the EpiCenter's exact rate ratio test.

Statistical tests of significance

EpiCenter first tested whether the read tag count in each region or gene was significantly larger than that expected under cell-type specific background noise. Regions/genes whose read-tag counts did not significantly exceed the background noise in both TS^{WT} and TS^{KI4} samples were filtered out. For those regions/genes retained, we then used EpiCenter exact rate ratio test to find whether the difference in read counts was significantly larger than that expected from Poisson variation. In mRNA-seq data analysis, we also used EpiCenter's Z-test to determine if the difference in read counts exceeded normal variation. We used Benjamini-Hochberg correction to adjust p-value for multiple testing in analysis of both ChIP-seq and mRNA-seq data. The technical details of EpiCenter will be published elsewhere.

Antibodies and sources

Antibody specificity	Company	Catalogue number
vimentin (rabbit mAb)	Cell Signaling	5741
vinculin (mouse mAb)	Sigma	V-9131
E-cadherin (mouse mAb)	BD Transduction	610181
actin (mouse mAb)	Sigma	A-3853
fibronectin (rabbit Ab)	EMD Calbiochem	CP70-100UG
N-cadherin (rabbit Ab)	Abcam	AB76057
gamma tubulin	Sigma	T6557
p300 (mouse mAb)	Millipore	05-257
CBP (rabbit pAb)	Abcam	ab2832
H2AK5Ac (rabbit pAb)	Cell Signaling	2576

H2BK5Ac (rabbit pAb)	Cell Signaling	2574
H2BK5Ac (rabbit pAb, ChIP)	Abcam	ab40886
H2BK12Ac (rabbit pAb)	Epitomics	1755-1
H2BK15Ac (rabbit pAb)	Epitomics	2170-1
H2BK20Ac (rabbit pAb)	Cell Signaling	2571S
H3K9Ac (rabbit pAb)	Cell Signaling	9671
H4K8Ac (rabbit pAb)	Cell Signaling	2594
Total H2A (rabbit pAb)	Cell Signaling	2572
Total H3 (rabbit pAb)	Cell Signaling	9715
H3K9me3 (rabbit pAb)	Abcam	ab8898
H3K4me3 (rabbit pAb)	Abcam	ab8580
Anti-HA (rabbit pAb)	Santa Cruz	12CA5
Anti-Myc (mouse mAb)	Santa Cruz	sc-40

Plasmids and sources

Plasmid	Source/Reference
HA-MAP3K4	(Gerwins et al., 1997)
HA-MAP3K4 kinase domain	(Gerwins et al., 1997)
Kinase-dead HA-MAP3K4 kinase domain	(Gerwins et al., 1997)
HA-JNK	(Gerwins et al., 1997)
FLAG-p38	(Abell et al., 2004)
HA-CBP	(Bradney et al., 2003)
HAT-dead HA-CBP	(Bradney et al., 2003)
Myc-p300	(Bradney et al., 2003)
HAT-dead Myc-p300	(Bradney et al., 2003)

Supplemental References

Abell, A. N., DeCathelineau, A. M., Weed, S. A., Ambruso, D. R., Riches, D. W., and Johnson, G. L. (2004). Rac2D57N, a dominant inhibitory Rac2 mutant that inhibits p38 kinase signaling and prevents surface ruffling in bone-marrow-derived macrophages. J Cell Sci *117*, 243-255.

Abell, A. N., Granger, D. A., and Johnson, G. L. (2007). MEKK4 Stimulation of p38 and JNK Activity Is Negatively Regulated by GSK3beta. J Biol Chem *282*, 30476-30484.

Abell, A. N., Granger, D. A., Johnson, N. L., Vincent-Jordan, N., Dibble, C. F., and Johnson, G. L. (2009). Trophoblast Stem Cell Maintenance by Fibroblast Growth Factor 4 Requires MEKK4 Activation of Jun N-Terminal Kinase. Molecular and Cellular Biology *29*, 2748-2761.

Bradney, C., Hjelmeland, M., Komatsu, Y., Yoshida, M., Yao, T. P., and Zhuang, Y. (2003). Regulation of E2A activities by histone acetyltransferases in B lymphocyte development. J Biol Chem *278*, 2370-2376.

Gerwins, P., Blank, J. L., and Johnson, G. L. (1997). Cloning of a novel mitogen-activated protein kinase kinase kinase, MEKK4, that selectively regulates the c-Jun amino terminal kinase pathway. J Biol Chem *272*, 8288-8295.

Li, H., Ruan, J., and Durbin, R. (2008). Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res *18*, 1851-1858.

Neve, R. M., Chin, K., Fridlyand, J., Yeh, J., Baehner, F. L., Fevr, T., Clark, L., Bayani, N., Coppe, J.-P., Tong, F.*, et al.* (2006). A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. Cancer Cell *10*, 515-527.

Odiatis, C., and Georgiades, P. (2010). New insights for Ets2 function in trophoblast using lentivirus-mediated gene knockdown in trophoblast stem cells. Placenta *31*, 630-640.

Prat, A., Parker, J. S., Karginova, O., Fan, C., Livasy, C., Herschkowitz, J. I., He, X., and Perou, C. M. (2010). Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. 1-18.