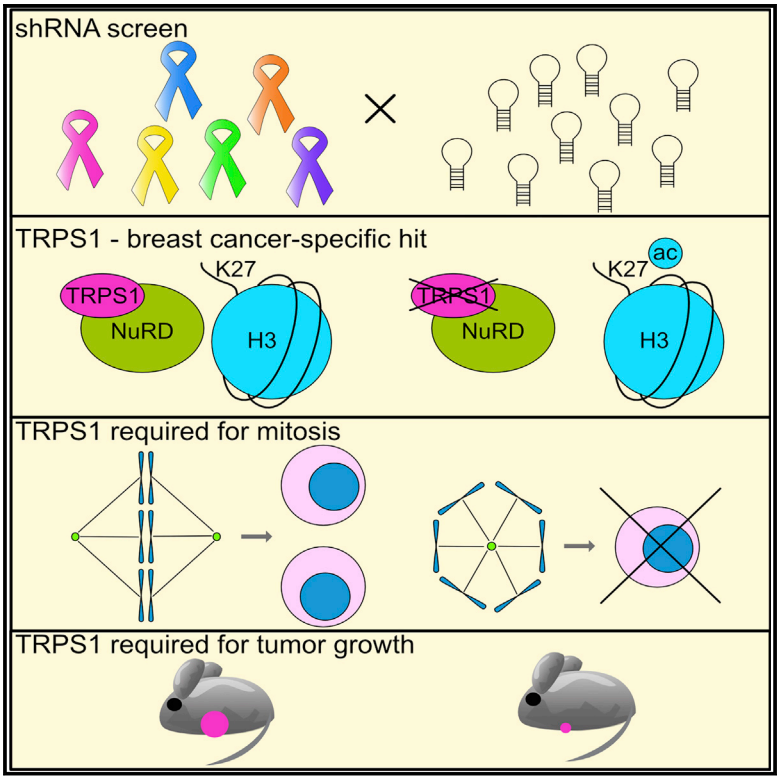
Article

Cell Reports

TRPS1 Is a Lineage-Specific Transcriptional Dependency in Breast Cancer

Graphical Abstract



Authors

Robert M. Witwicki, Muhammad B. Ekram, Xintao Qiu, ..., Ramesh A. Shivdasani, David Pellman, Kornelia Polyak

Correspondence

kornelia_polyak@dfci.harvard.edu

In Brief

Witwicki et al. use a targeted shRNA screening strategy to identify transcriptional and epigenomic dependencies in poorly differentiated human cancers. TRPS1 is a lineage-specific transcription factor that is required for mitosis in breast cancer cells. TRPS1 is associated with the NuRD complex, and it regulates cell adhesion, cytoskeleton, and G2-M phase-related genes.

Highlights

- Poorly differentiated human cancers have limited shared epigenomic dependencies
- TRPS1 is a breast lineage-specific transcriptional dependency
- TRPS1 is required for normal mitosis by directly regulating genes in G2-M
- TRPS1 is associated on the chromatin with the NuRD repressor complex

Witwicki et al., 2018, Cell Reports *25*, 1255–1267
October 30, 2018
https://doi.org/10.1016/j.celrep.2018.10.023

CellPress

Cell Reports
Article

TRPS1 Is a Lineage-Specific Transcriptional Dependency in Breast Cancer

Robert M. Witwicki,[1,2,8] Muhammad B. Ekram,[1,2,8,9] Xintao Qiu,[3] Michalina Janiszewska,[1,2] Shaokun Shu,[1,2] Mijung Kwon,[4,5] Anne Trinh,[1,2] Elizabeth Frias,[6] Nadire Ramadan,[6] Greg Hoffman,[6] Kristine Yu,[6] Yingtian Xie,[3] Gregory McAllister,[6] Rob McDonald,[6] Javad Golji,[6] Michael Schlabach,[6] Antoine deWeck,[6] Nicholas Keen,[6] Ho Man Chan,[6] David Ruddy,[6] Tomas Rejtar,[6] Sosathya Sovath,[6] Serena Silver,[6] William R. Sellers,[6] Zainab Jagani,[6] Michael D. Hogarty,[7] Charles Roberts,[4,5,10] Myles Brown,[1,2,3] Kimberly Stegmaier,[4,5] Henry Long,[3] Ramesh A. Shivdasani,[1,2,3] David Pellman,[4,5] and Kornelia Polyak[1,2,3,11,*]

[1]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA 02215, USA
[2]Department of Medicine, Harvard Medical School, Boston, MA 02115, USA
[3]Center for Functional Cancer Epigenetics, Dana-Farber Cancer Institute, Boston, MA 02215, USA
[4]Department of Pediatric Oncology, Dana-Farber Cancer Institute, Boston, MA 02215, USA
[5]Department of Pediatrics, Harvard Medical School, Boston, MA 02115, USA
[6]Novartis Institutes for Biomedical Research, Cambridge, MA 02139, USA
[7]Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA
[8]These authors contributed equally
[9]Present address: WuXi NextCODE, Cambridge, MA 02142, USA
[10]Present address: St. Jude Children's Research Hospital, Memphis, TN 38105, USA
[11]Lead Contact
*Correspondence: kornelia_polyak@dfci.harvard.edu
https://doi.org/10.1016/j.celrep.2018.10.023

SUMMARY

Perturbed epigenomic programs play key roles in tumorigenesis, and chromatin modulators are candidate therapeutic targets in various human cancer types. To define singular and shared dependencies on DNA and histone modifiers and transcription factors in poorly differentiated adult and pediatric cancers, we conducted a targeted shRNA screen across 59 cell lines of 6 cancer types. Here, we describe the TRPS1 transcription factor as a strong breast cancer-specific hit, owing largely to lineage-restricted expression. Knockdown of *TRPS1* resulted in perturbed mitosis, apoptosis, and reduced tumor growth. Integrated analysis of TRPS1 transcriptional targets, chromatin binding, and protein interactions revealed that TRPS1 is associated with the NuRD repressor complex. These findings uncover a transcriptional network that is essential for breast cancer cell survival and propagation.

INTRODUCTION

Modulation of chromatin structure plays an important role in regulating cell-type- and differentiation stage-specific gene expression programs in normal tissues and cancers (Brien et al., 2016; Jones et al., 2016). Genome sequencing studies have demonstrated that epigenetic regulator genes, including transcription factor (TF), are among the most commonly mutated in human cancers (Lawrence et al., 2014; Plass et al., 2013). These mutations represent good cancer-specific therapeutic targets, with likely low toxicity to heterologous cells. However, with the exception of oncogenic kinases and metabolic enzymes that produce gain-of-function phenotypes, the exploitation of most cancer-specific mutations for therapy remains a major challenge. Identifying the mechanisms of TF dependencies can lead to new targetable therapeutic approaches.

Perturbed epigenetic programs are particularly relevant in pediatric tumors, which generally carry few somatic mutations and are thought to initiate from stem/progenitor cells that failed to follow the normal differentiation path (Gröbner et al., 2018). Rhabdoid tumors are exemplary because they contain only a single recurrent mutation in the *SMARCB1* gene (Kim and Roberts, 2014), while infantile ependymomas lack any significant somatic genetic initiating event and may entirely be a result of epigenetic dysregulation (Mack et al., 2014). A significant fraction of neuroblastomas, the most common extracranial pediatric solid tumor, is driven largely by the amplified *NMYC* TF (Pugh et al., 2013; Sausen et al., 2013). Adult prostate and breast carcinomas, particularly triple-negative breast cancer (TNBC), also have relatively few recurrent genetic alterations, suggesting the importance of epigenetic drivers in these tumor types (Lawrence et al., 2014). While most prostate tumors are strongly dependent on the androgen receptor AR, a ligand-dependent TF, the epigenetic drivers and transcriptional dependencies of TNBC remain poorly characterized.

High-throughput knockdown and knockout screens have been useful for the discovery of cellular dependencies and for identifying synthetic lethal and resistant mechanisms. Although various RNAi screens have generated functional dependency maps in large panels of cancer cell lines (Marcotte et al., 2016; McDonald et al., 2017; Tsherniak et al., 2017; Wang et al., 2015), pediatric tumors and poorly differentiated adult cancers, including androgen-independent prostate cancer and TNBC, are generally underrepresented in these screens.

Here, we report the results of a lentiviral small hairpin RNA (shRNA) screen coupled with comprehensive molecular profiling to assess epigenetic and transcriptional dependencies in 59 cell lines representing 6 poorly differentiated cancer types.

RESULTS

Functional Genomic Screen Reveals Lineage Specificity

To determine whether poorly differentiated adult and pediatric cancers of different tissues of origin share common epigenetic and transcriptional dependencies, we performed a targeted pooled shRNA screen. We screened 59 cell lines (Table S1) using 2 shRNA libraries with 650 genes in each, designed to target genes encoding epigenetic regulators and TFs (Table S2). To enhance specificity and sensitivity, we used libraries with a median of 20 shRNAs targeting each gene. We chose cell lines encompassing multiple cancer types, including pediatric leukemia (5), neuroblastoma (15), triple-negative (8) and human epidermal growth factor receptor 2-positive ($HER2^+$) (5) breast cancer, androgen-independent prostate cancer (7), and colorectal (13) and rhabdoid (6) tumors. We prioritized cell lines not included in the Cancer Cell Lines Encyclopedia (CCLE) and not available through public repositories. Thus, almost none were included in prior RNAi screens. All breast cancer cell lines were hormone receptor-negative non-luminal cancers, 5 *HER2*-amplified and 8 TNBCs, with 3 of the lines derived from inflammatory breast cancer (IBC), a rare but lethal subset. Each cell line was individually optimized for efficient lentiviral infection and selection, then screened at an average depth of 1,000 cells per shRNA. We also obtained comprehensive molecular profiles, analyzing all cell lines for genetic alterations by exome sequencing (exome-seq), gene expression by RNA sequencing (RNA-seq), and epigenetic patterns by histone mass spectrometry and Illumina 450K DNA methylation arrays. Contrary to most prior high-throughput screens, we grew each cell line in optimal conditions, including optimized growth medium, seeding density, and time between passages, hence minimizing stresses that could introduce bottlenecks and bias the results (Figure 1A). Normalized sequencing counts of shRNAs after the screen compared to the input libraries were analyzed by ATARiS (Shao et al., 2013) and RSA (König et al., 2007) algorithms. We detected 26–87 significant dependencies in each cell line, for a total of 684 confident hits (Figure 1B; Table S2). Confirming the quality of the screen, top hits included several known cancer dependencies: *AR*, *EED*, and *FOXA1* in prostate cancer lines; *KRAS*, *BRAF*, and *CTNNB1* in colon cancer; *CEBPB* and *NR3C1* in breast cancer; *MYCN*, *PHOX2B*, and *ASCL2* in neuroblastoma; and *SMARCC2* in rhabdoid tumors (Figures 1C and S1). After excluding genes whose deficiencies were broadly lethal or previously described in the literature, we found 111 dependencies; slightly more than half of these were epigenetic regulators and the rest were TFs. For example, *KDM6A* encoding the H3K27 demethylase UTX was a significant hit in colon cancer and in some neuroblastomas, whereas *ARID1A* was a top prostate cancer-specific hit (Figure S1). As expected, the number of hits was higher when more cell lines of a given tumor type were analyzed (Figure 1D).

Most hits occurred in >1 cancer type, and tumors with higher cell line representation tended to have more lineage-specific hits (Figure 2A). Unsupervised clustering of cell lines using ATARiS scores of significant hits revealed overall alignment according to lineage, with neuroblastoma, prostate cancer, and leukemias forming the tightest clusters (Figures 2B and S2A). To determine whether this clustering was due to lineage-specific expression of the significant hits, we analyzed associations between ATARiS scores and mRNA levels. mRNA levels of significant hits were generally higher than those of non-hits, implying that cellular dependencies reflect more abundant expression (Figure 2C), but within the same lineage, the mRNA levels of specific hits were not always higher than the background (Figure 2D). The mRNA levels of genes included in the screen also did not differ significantly among tumor types (Figure 2E). Analyzing similarities among cell lines based on the expression of all genes or genes included in the screen again yielded lineage-specific clusters (Figures 2F, S2B, and S2C). Notably, breast cancer cell lines split into 2 groups based on both expression and ATARiS scores, one showing higher similarity to colon cancer cell lines, while another subset resembled rhabdoid tumors and neuroblastoma.

Epigenetic profiles were more variable among cell lines, with DNA methylation profiles-based correlation maps showing higher similarity to expression-based clusters than histone modification profiles (Figures 2G and S2D–S2F). Neuroblastoma, breast cancer, and rhabdoid tumor lines showed especially tight clustering according to lineage and also displayed higher similarity to one another, largely driven by levels of H3K27me0 (unmethylated H3K27) and H3K36me1/2 peptides (Figure S2F), suggesting potentially common chromatin features. Despite the similarity in histone modification patterns, almost no epigenetic regulator was a shared hit in breast, rhabdoid, and neuroblastoma lines (Table S2). In contrast, a number of TFs, including *GATA1*, *NEUROD1*, *YY1*, and *GLI2*, were common dependencies in several lines of these 3 lineages. The limited number of common but not pan-essential hits among cell lines from different tumor types was generally true for the full screen, with very few genes giving hits in $\geq$3 lineages (Figure 2H). Finally, most hits did not correspond to mutated genes, as determined by exome-seq (data not shown). Thus, our data suggest that most identified cellular dependencies are driven by lineage-restricted gene expression, which is in agreement with findings from recent large-scale genome-wide shRNA screens (McDonald et al., 2017; Tsherniak et al., 2017).

TRPS1 Is Required for Breast Cancer Cell Survival

The most significant hits in breast cancer cell lines included *ADAR* encoding for an RNA-adenosine deaminase and regulator of the double-stranded RNA-sensing pathway (Pasquinelli, 2018); *DCAF1*, an E3 ubiquitin ligase (Romani and Cohen, 2012); and the TF gene *TRPS1*, which is implicated in the rare inherited tricho-rhino-phalangeal syndrome (TRPS) (Momeni et al., 2000). *TRPS1* showed the most breast-specific dependency and a requirement for viability in nearly every breast cancer cell line we tested (Figure 3A). Compared to other tissues, *TRPS1* transcript levels are significantly higher in the normal breast (Figures 3B and S3A) and in breast cancer (Figure S3B), and *TRPS1* expression correlated with ATARiS scores in individual breast

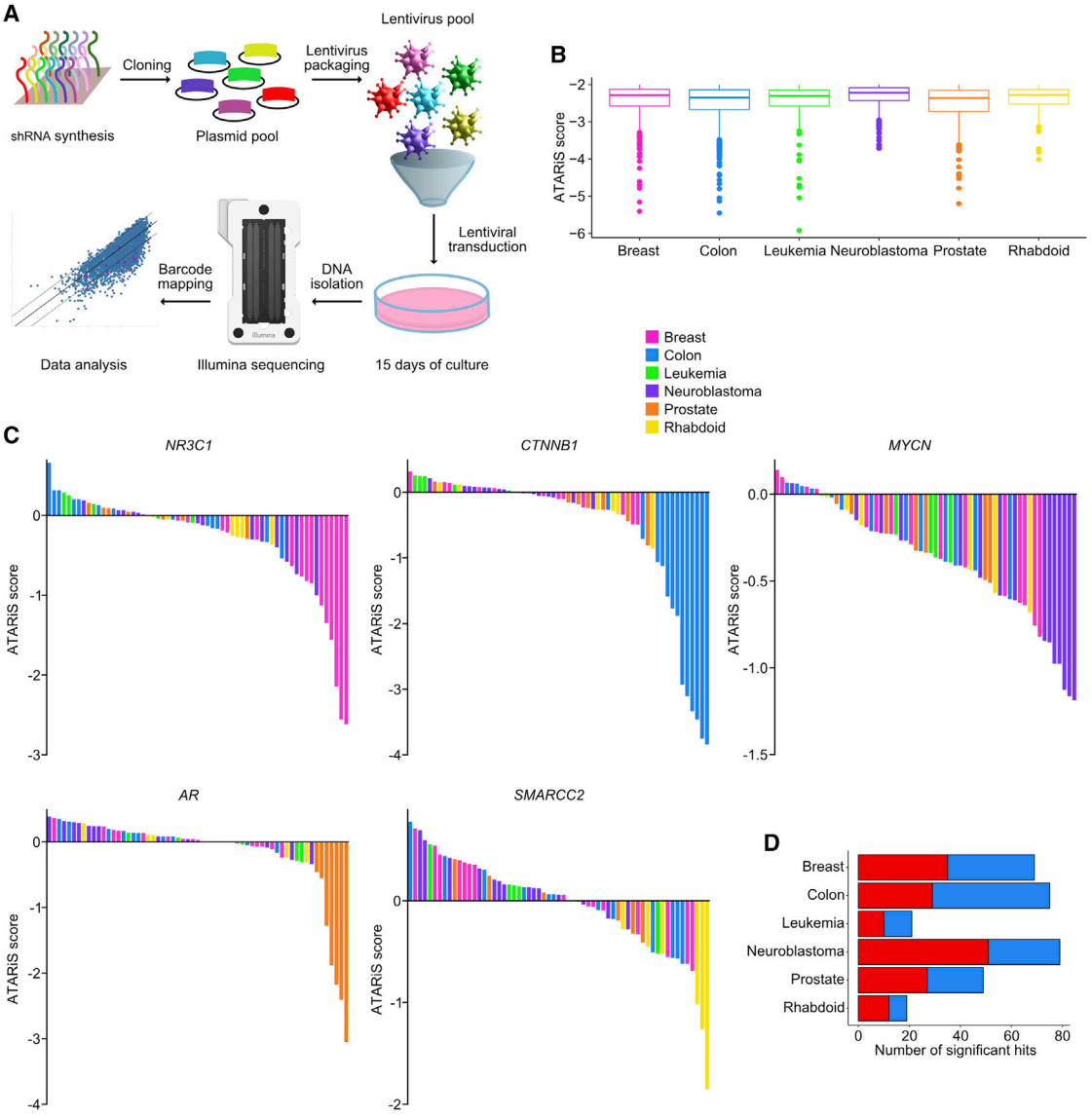


Figure 1. Results of Pooled shRNA Screen in 6 Different Tumor Types

(A) Schematic outline of the shRNA screen.

(B) Boxplots depicting the numbers of statistically significant hits and their degree of dependency, based on ATARiS score among different lineages. Box plots shown as median, 25th and 75th percentiles; points are outliers below 1.5 × the interquartile range.

(C) Examples for top tumor type-specific screen hits.

(D) Number of genes scored as significant hit in at least 1 cell line in the transcription factor (red) and epigenetic regulators (blue) shRNA library.

See also Figure S1 and Tables S1 and S2.

cancer ($R^2 = 0.44$) but not in other cell lines ($R^2 = 0.07$). These observations imply that the specific *TRPS1* dependency of breast cancer reflects lineage-restricted expression.

To investigate the consequences of *TRPS1* depletion in breast cancer cells, we generated derivatives of the HCC3153 and SUM159 TNBC cell lines, which were highly (HCC3153) and mildly (SUM159) affected by *TRPS1* knockdown in the shRNA screen. Using the tetracycline-inducible lentiviral knock out plasmid (pLKO-Tet-ON) system, we expressed doxycycline-inducible *TRPS1*-targeting shRNAs, and the resulting downregulation of *TRPS1* led to rapid loss of viability in HCC3153 cells (Figures 3C and S3C), whereas SUM159 cells showed significant growth suppression with limited death (Figure 3C). *In vivo* xenograft assays using the same tetracycline (Tet)-inducible *TRPS1* shRNAs demonstrated diminished tumor growth after *TRPS1* knockdown in both cell lines, with nearly complete regression in HCC3153-derived tumors (Figures 3D, 3E, S3D, and S3E). Histology of the residual tissue revealed reduced cellularity in both cell lines, with very few tumor cells left at HCC3153 injection sites (Figure S3F).

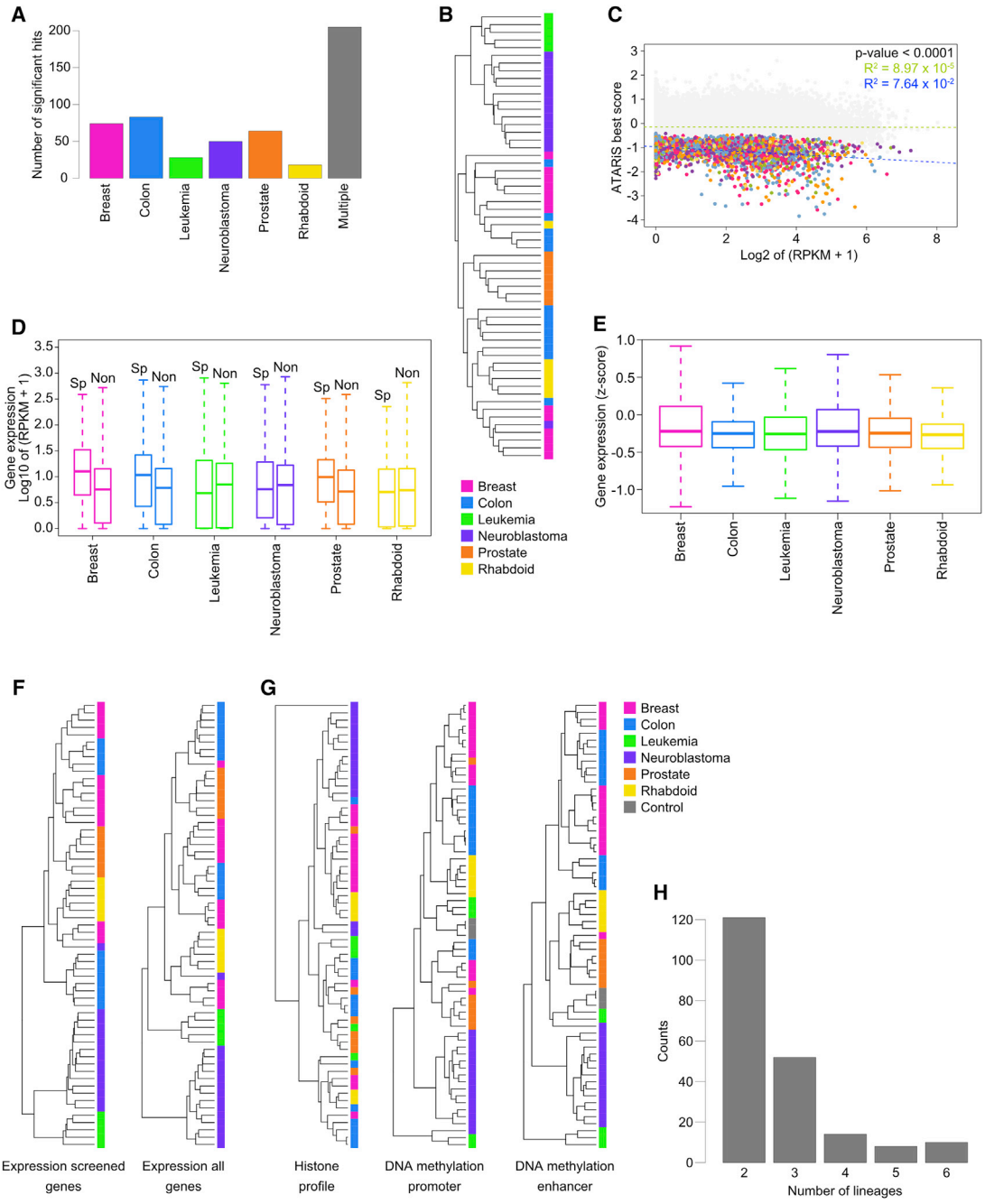


Figure 2. Associations between Genomic Profiles and Screen Hits

(A) Graph depicting numbers of significant hits specific to a lineage or hit in multiple (>3) lineages.

(B) Dendogram depicting similarity of cell lines based on unsupervised clustering of ATARiS scores.

(C) Correlation between ATARiS scores and mRNA levels (blue line, hits; green line, non-hits). Each dot is a screen hit; colors correspond to tumor types. RPKM, reads per kilobase million.

(D) Expression levels of lineage-specific (Sp) and non-specific (Non) hits in the different lineages. Box plots shown as median, 25th and 75th percentiles; whiskers represent 1.5 × the interquartile range.

(E) Graph depicting the relative expression levels of the screened genes across the various lineages. Box plots shown as median, 25th and 75th percentiles; whiskers represent 1.5 × the interquartile range.

(F) Dendrogram depicting similarity of cell lines on the basis of unsupervised clustering of mRNA levels of all genes or genes included in the screen.

(G) Dendrogram depicting similarity of cell lines on the basis of unsupervised clustering of histone modification and DNA methylation profiles.

(H) Graph depicting numbers of significant hits in the indicated number of lineages.

See also Figure S2.

To investigate the mechanisms underlying this growth suppression, we analyzed cell-cycle profiles and apoptosis by flow cytometry for DNA content and annexin V, respectively, in HCC3153 cells at different times after inducing sh*TRPS1*. The fraction of cells with 4N DNA content increased from 27% to 48% ($p = 0.0005$, chi-square test) by 4 days after shRNA induction, which is consistent with G2-M arrest (Figure 3F). Increased apoptosis was significant ($p = 0.00005$, chi-square test) after 6 days, with a mean 4-fold increase in annexin V^+PI^- cells (Figure 3G). Immunofluorescence analysis of mitotic cells labeled for phospho-histone H3 and tubulin revealed hyperaccumulation of pre-anaphase cells with severely disorganized mitotic spindles after *TRPS1* downregulation (Figures 3H and 3I), a 10-fold decrease in the percentage of post-anaphase cells after TRPS1 depletion (0.8% relative to 9.4% in control). To further characterize this perturbed mitosis phenotype in more detail, we expressed red fluorescent protein (RFP)-tagged histone H2B in HCC3153 cells with *TRPS1* shRNA and performed live-cell imaging of the cells 4 days after doxycycline treatment. We found a strong delay in M phase progression after TRPS1 depletion, with the average time from initiation of mitosis to mitotic exit increased by 8-fold (Figures 3J and 3K; Video S1). The majority (80%) of these cells exited mitosis without chromosome separation after prolonged mitotic arrest. These results suggest that *TRPS1* is essential for normal mitotic progression and is required for the proper organization of mitotic spindle in breast cancer cells.

TRPS1 as Both an Activator and a Repressor of Transcription in Breast Cancer

Because TRPS1 is a TF that may directly regulate genes required for G2-M progression and survival in breast cancer cells, we performed RNA-seq of HCC3153 and SUM159 cell lines 3, 4, and 5 days after inducing *TRPS1* knockdown. We found 2,039 and 892 genes with a consistent change in expression at all 3 time points in HCC3153 and SUM159 cells, respectively (Figure 4A; Table S3). Although 349 affected genes were common to the 2 cell lines, implying a likely common regulatory basis, the lines also showed some differences in directionality and timing of expression changes. The number of genes increased in SUM159 cells was higher than the genes reduced at all 3 time points, and the numbers of differentially expressed genes were similar on successive days. In contrast, more genes were downregulated than upregulated in HCC3153 cells at days 3 and 4, and the numbers of differential genes increased over time. Pathway analysis of the commonly or uniquely differentially expressed genes in the 2 cell lines by Metacore (Nikolsky et al., 2009) correlated very well with the observed cellular phenotypes. Many of the top 10 pathways enriched in genes dependent on TRPS1 in both cell lines were associated with mitosis, microtubules, and cell-cycle G2-M regulation, while upregulated genes were enriched in transcription, G1-S regulation, and various stress and cellular survival pathways (Figures 4B, S4A, and S4B). In contrast, the top enriched pathways for genes uniquely (excluding common ones) differentially expressed in a cell-type-specific manner included apoptosis, hedgehog signaling, and blood vessel morphogenesis (Figures S4C and S4D).

To identify direct transcriptional targets of TRPS1, we performed chromatin immunoprecipitation sequencing (ChIP-seq) in HCC3153, SUM159, and MDA-IBC3 cells, which were selected based on endogenous TRPS1 levels. A subset of TRPS1 binding sites was shared among all 3 cell lines, but most peaks were unique, with the highest number observed in HCC3153 cells (Figure 4C; Table S4). The distribution of binding sites also varied among cell lines, with a higher fraction of TRPS1 peaks located in promoters and exons in MDA-IBC3 cells compared to HCC3153 and SUM159 cells, where distal sites and introns (putative enhancers) dominated the profile (Figure S4E).

TRPS1 was initially described as a transcriptional repressor (Malik et al., 2001), but more recently it was also shown to activate certain genes via direct promoter binding (Fantauzzo and Christiano, 2012). To determine whether TRPS1 localizes at transcriptionally active or repressed chromatin regions, we analyzed overlaps between TRPS1 peaks and H3K27ac and H3K27me3 histone marks, which characterize transcriptionally active and silenced regions, respectively (Justin et al., 2010), in HCC3153 cells. TRPS1 binding did not overlap significantly with H3K27me3-enriched regions (promoters), while approximately one-third of TRPS1-bound chromatin was associated with H3K27ac marks (Figures 4D and S4F). To assess whether H3K27ac-overlapping and non-overlapping binding of TRPS1 was associated with changes in gene expression, we analyzed these sites with respect to gene expression changes after TRPS1 downregulation in HCC3153 cells. Binding and expression target analysis (BETA) (Wang et al., 2013) of TRPS1-unique and H3K27ac-overlapping peaks revealed that TRPS1 loss led to the upregulation of genes associated with TRPS1-unique peaks, whereas H3K27ac-overlapping sites were associated with both upregulated and downregulated genes (Figure 4E). Examples of TRPS1 peaks include *WEE1*, *BCL2L1*, and *PTK2* (Figure 4F). Downregulation of TRPS1 also led to significantly increased H3K27ac signal, with more pronounced increases at TRPS1-unique than at TRPS1-H3K27ac-overlapping peaks (Figures 4G and S4G). Metacore analysis of pathways significantly enriched among genes that are differentially expressed and TRPS1 bound gave results that are similar to the enrichments based on expression changes alone. Cytoskeleton, cell adhesion, and G2-M regulation were the top enriched pathways in downregulated genes, while apoptosis, epithelial-to-mesenchymal transition (EMT), and cell adhesion pathways dominated among upregulated genes (Figure 4H). Several key proteins involved in osteogenesis, including BMPs, osteoprotegerin, osteonectin, and parathyroid hormone-related protein, are known TRPS1 targets in developing bone (Nishioka et al., 2008) and were upregulated after TRPS1 knockdown in HCC3153 breast cancer cells.

TRPS1, an atypical member of the GATA family (Malik et al., 2001), also contains C-terminal zinc fingers related to the IKAROS-family dimerization domain (Momeni et al., 2000). To explore the sequence specificity of TRPS1 binding, we performed *de novo* motif analysis of TRPS1-unique and H3K27ac-overlapping peaks. Although the motifs most enriched in both groups were for general TFs such as AP-1 proteins, GATA3, TEAD3/4, NF1, RUNX3, CEBPA/E, and FOXA1 were also among

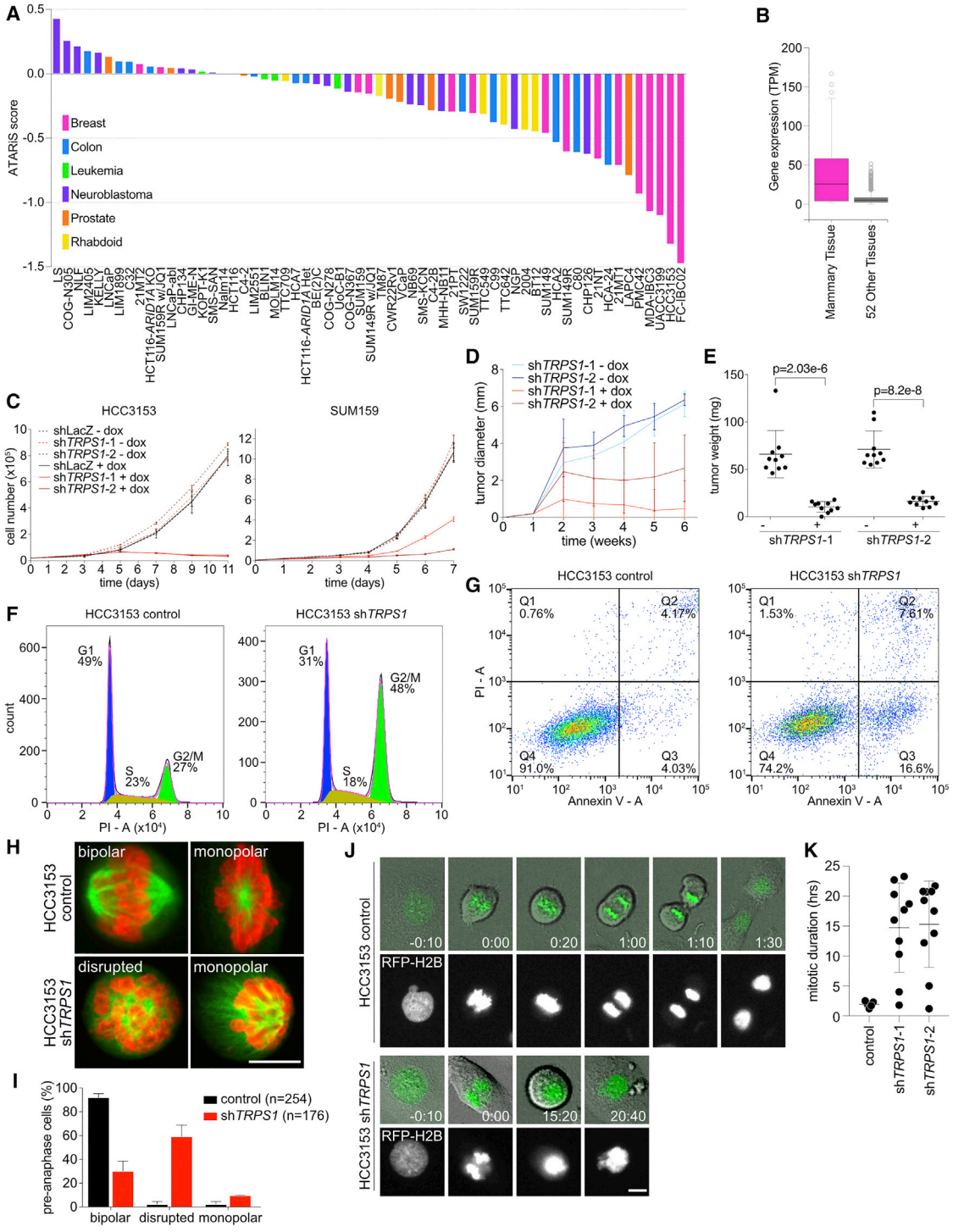


Figure 3. *TRPS1* as a Breast Lineage-Specific Dependency

(A) ATARiS score of *TRPS1* across the panel of cell lines.

(B) mRNA levels of *TRPS1* in normal human tissues. TPM, transcripts per million. Box plots shown as median, 25th and 75th percentiles; points are outliers above 1.5 × the interquartile range.

(C) Cellular viability after *TRPS1* knockdown in HCC3153 and SUM159 cell lines. Error bars represent SD; n = 3.

(D) Volume of HCC3153 tumors with and without doxycycline in diet. Error bars represent SD.

(E) Final weight of HCC3153 tumors with and without doxycycline in diet; unpaired t test. Error bars represent SD, n = 10.

(F) Cell-cycle profile of HCC3153 cells expressing control and TRPS1-targeting shRNA.

(legend continued on next page)

the top enriched motifs (Figure S4H). While GATA3 and FOXA1 are expressed at very low levels in TNBC, several other members of these TF families (e.g., GATAD2B, FOXA2) have high levels in TNBCs. Because the DNA binding domains of different members of these TF families are nearly identical, the DNA sequences they recognize are also very similar, and the most well characterized ones show up as the most significant hits in motif analyses. These data suggest that TRPS1 may participate in different chromatin complexes and, depending on the complex composition, may affect transcription positively or negatively.

TRPS1 Interacts with the NuRD Repressor Complex

The mechanism by which TRPS1 represses transcription has not been characterized. TRPS1 was previously shown to bind CtBP via a PXDLS region in its C terminus (Perdomo and Crossley, 2002). To characterize TRPS1-containing chromatin complexes in HCC3153 and MDA-IBC3 cells, we performed rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME) (Mohammed et al., 2013) and identified 214 and 147 TRPS1-associated proteins, respectively, compared to immunoglobulin G (IgG) controls. Limiting the analysis to proteins with known roles in transcription and epigenetic regulation, we detected TRPS1 interactions with multiple components of the nucleosome remodeling deacetylase (NuRD) complex, including CHD4, MTA1–3, HDAC1/2, GATAD2B/A (Figure 5A). Pathway enrichment analysis of all TRPS1-associated proteins detected by RIME also identified Sin3 and NuRD as the top enriched pathway, followed by the HP1 heterochromatin and RE1-silencing TF corepressor (CoREST) complex, suggesting that TRPS1 is mainly associated with transcriptional repression (Figure 5B). Immunoblot analysis of TRPS1 immunoprecipitates for GATAD2A, GATAD2B, and HDAC2, and MTA2 validated the RIME data (Figure S5A).

To determine whether GATAD2A and GATAD2B overlap with TRPS1-occupied chromatin regions, we performed ChIP-seq experiments for these factors. We found that 72% of GATAD2A peaks and 38% of GATAD2B peaks overlap with TRPS1 binding sites (Figures 5C and S5B; Table S4). Motif analysis of overlapping peaks identified similarities to TEAD, RUNX, and CEBP binding sites (Figure S5C). We then integrated the TRPS1-unique and GATAD2A/B-overlapping peaks with gene expression changes after *TRPS1* knockdown. TRPS1 unique peaks again were associated with genes upregulated after TRPS1 loss, whereas TRPS1/GATAD2B-overlapping peaks were linked with both upregulated and downregulated genes (Figure 5D). Functional analysis of TRPS1-GATAD2B overlapping peaks associated with genes differentially expressed after *TRPS1* knockdown revealed cell junction and cytoskeleton regulation among the top pathways enriched in both upregulated and downregulated genes, while bone remodeling and EMT-related genes were enriched only in the upregulated pathways and WNT and Hh signaling only in the downregulated pathways, respectively (Figure 5E). Examples of genes with TRPS1-GATAD2B-overlapping peaks include MYC and VIM (Figure 5F). To confirm that TRPS1 associates on the chromatin with the NuRD complex and not only with GATAD2A and GATAD2B, we also performed ChIP-seq for HDAC2 and MTA2 and found that essentially all 5 factors co-localize to the same genomic regions (Figures S5D and S5E). However, this nearly complete overlap of chromatin binding does not prove direct interaction of TRPS1 with any of these proteins. Delineating the details of TRPS1-NuRD complex interaction would require additional studies.

To further interrogate the functional relevance of the TRPS1 and GATAD2A/B interaction, we performed ChIP-seq experiments after depleting TRPS1. Loss of TRPS1 resulted in the almost complete dissociation of GATAD2B from the chromatin, with a 52-fold reduction (or 61-fold with fold change [FC] >10) in peak numbers, whereas GATAD2A peaks did not change significantly (Figure 5G). This dramatic loss in chromatin-bound GATAD2B was not due to the downregulation of GATAD2B itself, since its mRNA and protein levels were not significantly affected by sh*TRPS1* expression (data not shown; Figure S5F). However, since TRPS1 knockdown perturbs mitosis and the majority of the cells are arrested in M phase, when most TFs dissociate from the chromatin, the observed loss of chromatin-bound GATAD2B could be the consequence of this.

Clinical Relevance of TRPS1 in TNBC

To assess the clinical relevance of *TRPS1* in human cancer, we analyzed its mutation status and expression in all cancers and in breast cancer using cBioPortal. We found that *TRPS1* is commonly amplified in prostate, breast, and ovarian carcinomas, in which up to 40% of the tumors have a *TRPS1* copy number gain, while a significant fraction of gastric and lung adenocarcinomas have point mutations in the *TRPS1* gene (Figures S5G and S5H). *TRPS1* and *GATAD2B* are also commonly co-amplified (cBioPortal significant co-occurrence, $p = 0.005$) in breast tumors; 24% of cases with *GATAD2B* amplification also have *TRPS1* amplification, while 16% of cases with TRPS1 amplification also have GATAD2B amplification.

We also analyzed the expression of genes associated with TRPS1-GATAD2B-overlapping peaks in human breast cancer cohorts. We computed single-sample gene set enrichment analysis (ssGSEA) scores for each patient based on 2 signatures of interest: (1) a cistrome featuring candidate targets of both

(G) Fluorescence-activated cell sorting (FACS) analysis of annexin V^+ cells in control and sh*TRPS1*-expressing HCC3153 cells.
(H) Immunofluorescence images of HCC3153 cells stained for phospho-histone H3 and tubulin after expression of control or *TRPS1*-targeting shRNA. Scale bar corresponds to 10 μm.
(I) Percentage of pre-anaphase cells with normal bipolar spindles, disrupted spindles, or monopolar spindles in the indicated conditions. Note that, unlike in control cells, the monopolar spindles in *TRPS1* shRNA-expressing cells showed chromosomes localized proximal to the spindle poles, suggestive of defective kinetochore-microtubule attachment. Error bars represent SD.
(J) Time-lapse still images from live-cell imaging of RFP-histone H2B-expressing HCC3153 cells before and after *TRPS1* downregulation. Time is in hr:min (t = 0 mitotic cell rounding). Scale bar corresponds to 10 μm.
(K) Quantification of mitotic duration in indicated conditions. Error bars represent SD.
See also Figure S3.

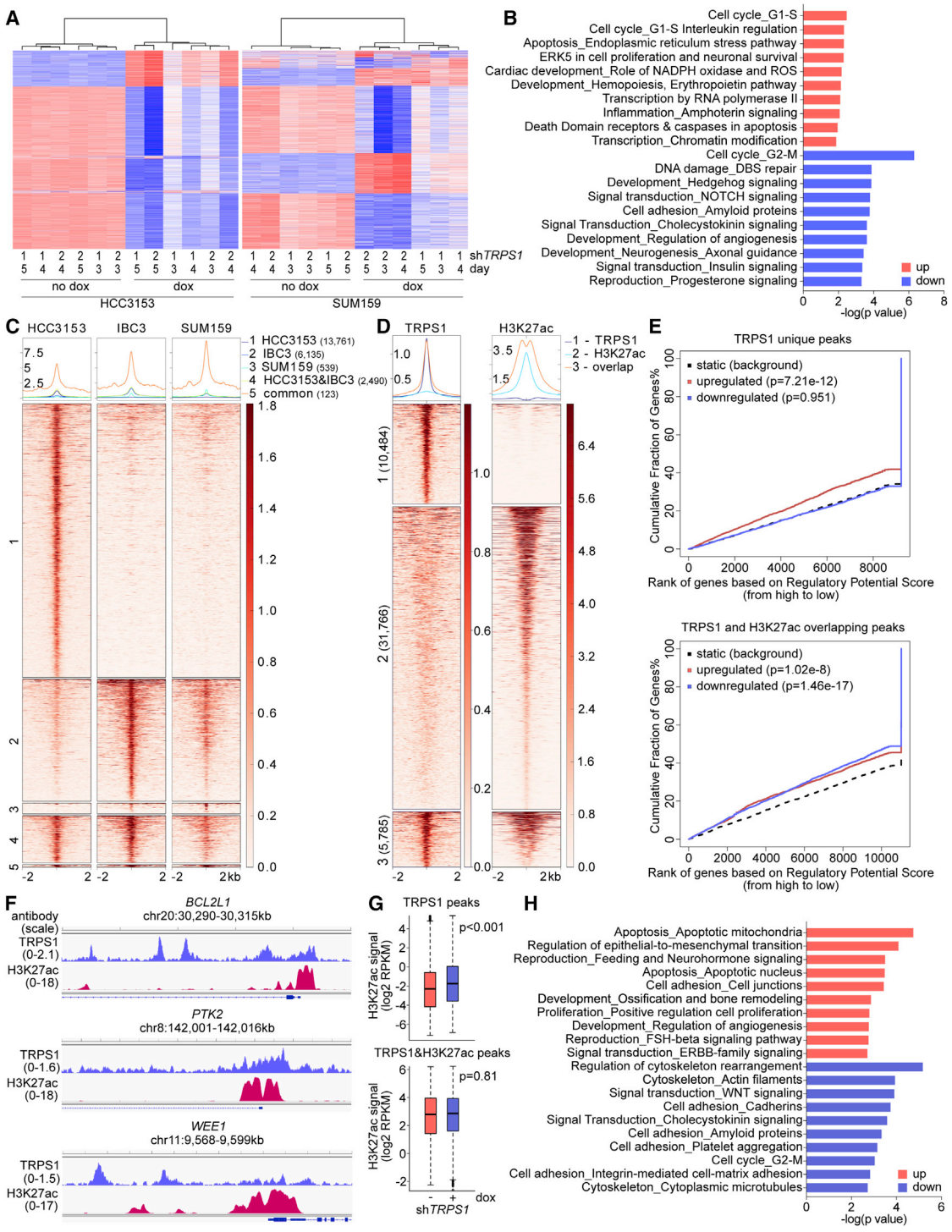


Figure 4. Genomic and Transcriptional Targets of TRPS1 in Breast Cancer

(A) K-means clustering of top 2,000 differentially expressed genes in HCC3153 and SUM159 lines after *TRPS1* knockdown using doxycycline (dox)-inducible sh*TRPS1*.

(B) Top 10 upregulated and downregulated process networks enriched in differentially expressed genes that are common for HCC3153 and SUM159 cell lines.

(C) Heatmap of TRPS1 peaks in HCC3153, MDA-IBC3, and SUM159 cell lines categorized into common and unique peaks. Each group represents peaks: 1, specific to HCC3153; 2, specific to MDA-IBC3; 3, specific to SUM159; 4, common for HCC3153 and MDA-IBC3; and 5, common for all 3 cell lines. Numbers of peaks in each group are indicated.

(legend continued on next page)

TRPS1 and GATAD2B and (2) a gene signature that integrates ChIP-seq peaks with upregulated genes. Both signatures appeared to be more highly enriched in estrogen receptor-negative (ER^-) tumors (including $HER2^+$) compared to ER^+ (Figure S5I; $p << 0.0001$ for both cases, Wilcoxon rank sum test), suggesting a higher level of TRPS1/GATAD2B-mediated expression of these target genes. To determine the prognostic value of these signatures, survival analysis was performed using Cox proportional hazards models adjusted for age, Nottingham prognostic index (NPI), and ER/HER2 status. We found that patients with medium and high ssGSEA scores had worse outcomes (hazard ratio [HR] = 1.18 [0.95–1.46] for medium and HR = 1.3 [1.04–162] for high patients, log rank $p << 0.0001$; Figure S5J), suggesting that tumors with higher expression of TRPS1-GATAD2B targets have unfavorable biological properties. These data suggest a functionally and clinically relevant role for TRPS1 and the TRPS1-GATAD2B complex in breast tumorigenesis.

DISCUSSION

Based on an epigenome-targeted high-density shRNA screen of cell lines from 6 different tumor types, we identified a number of TFs and epigenetic regulators required for proliferation or survival of diverse cancer types. This study culminated in our identification and characterization of *TRPS1* as a lineage-specific transcriptional repressor required for breast cancer cell survival and mitosis.

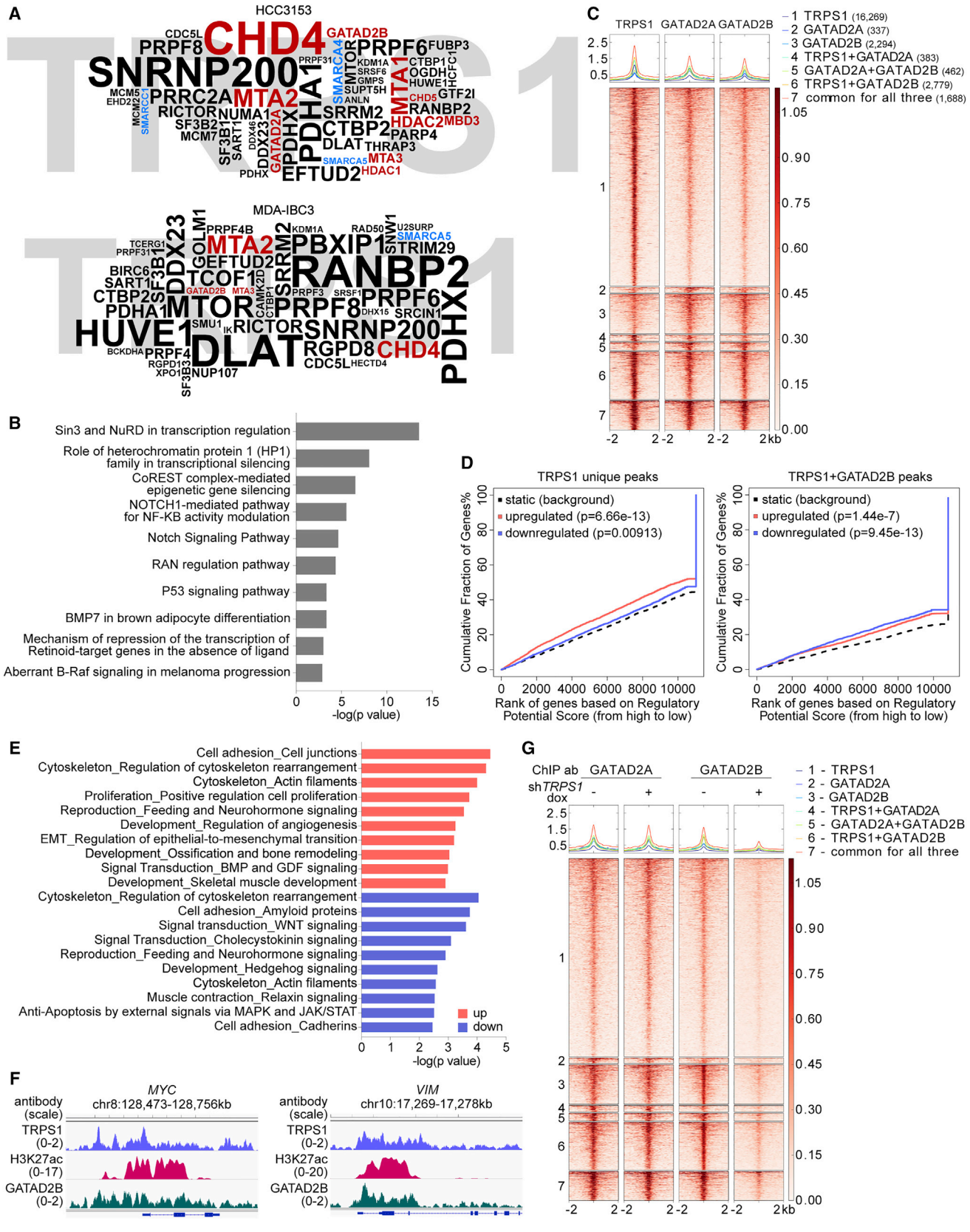
TRPS1 is a TF with classic GATA-type zinc fingers (Malik et al., 2001), but contrary to most GATA family members, TRPS1 is known to repress transcription, a function that requires its C-terminal IKAROS domain-containing CtBP binding sites (Malik et al., 2001; Perdomo and Crossley, 2002). Heterozygous *TRPS1* germline mutations are responsible for a rare hereditary autosomal dominant disorder, the TRPS, which is characterized by skeletal and facial abnormalities and sparse scalp hair (Momeni et al., 2000). Phenotypic severity depends on the location of the mutation within *TRPS1* (Maas et al., 2015), with those in the GATA domain causing more severe disease, as confirmed in mice with genetic deletion of this domain (Malik et al., 2002). These results suggest a critical role for the canonical GATA domain in TRPS1 function, potentially as a result of direct DNA binding.

Prior studies have hinted at roles for *TRPS1* in tumorigenesis, including breast cancer, but the underlying mechanism is poorly defined. *TRPS1* is overexpressed in breast tumors, as confirmed by immunohistochemistry (Radvanyi et al., 2005). In line with these findings, our analyses of The Cancer Genome Atlas (TCGA) and Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) breast cancer cohorts revealed frequent amplifications and overexpression. *Trps1* was also identified by a transposon mutagenesis screen in *Pten* mutant mice as a tumor suppressor gene that suppresses EMT (Rangel et al., 2016). Our demonstration that TRPS1 genomic targets are upregulated after *TRPS1* depletion and enriched for EMT-related genes are consistent with these findings. Characterization of $Trps1^{-/-}$ mice has also shown a role in chondrocyte proliferation and differentiation, mediated in part by Runx2 (Napierala et al., 2008; Wuelling et al., 2013). However, direct genomic targets of TRPS1 have remained poorly characterized and the mammary epithelium not examined in $Trps1^{-/-}$ mice.

Our shRNA screen and follow-up validation provide firm evidence of *TRPS1* transcriptional dependency in breast cancer owing to its lineage-specific expression. It is required for breast cancer cell growth *in vitro* and *in vivo* and a critical regulator of M phase progression, an unusual function for a TF because most TFs vacate chromatin during M phase. Our integrated RNA-seq and ChIP-seq profiling of multiple breast cancer cell lines reveal that TRPS1 suppresses EMT-related and osteogenic genes in breast cancer cells, while enhancing expression of genes involved in cell adhesion (e.g., P-, M-, R-cadherins), cytoskeleton organization (e.g., *PTK2*, *RDX*, *TMOD2*, *FMN1)*, and G2-M progression (e.g., *BUBR1*); these observations provide a mechanistic explanation for the defects we observed reproducibly after *TRPS1* knockdown. TRPS1-containing chromatin complexes, identified by RIME, and ChIP-seq for H3K27ac indicated that TRPS1 represses genes by recruiting the NuRD complex to chromatin, hence triggering histone deacetylation. Thus, TRPS1 may exert its function by the proper regulation of histone deacetylation that is required for normal mitotic progression and chromosome segregation (Cimini et al., 2003). However, our data on enrichment of various TF motifs in TRPS1 peaks, coupled with RNA defects in knockdown cells, suggest that TRPS1 is also associated with transcriptional activation when complexed with other TFs such as TEAD, RUNX, and GATA family members. In line with our findings, a recent study identified TRPS1 as a repressor of YAP activity via interacting with the YAP-TEAD1 complex (Elster et al., 2018), while another reported TRPS1 interaction with ER and repression of ER-dependent transcription in ER^+ breast cancer (Serandour et al., 2018).

Based on our observations, we speculate that TRPS1 has a critical role in maintaining epithelial barrier function, tissue architecture, and mechano-transduction. It may in this regard control mammary epithelial duct branching, similar to its role in ureteric bud branching via regulating transforming growth factor-β

(D) Heatmap of TRPS1 peaks overlapping or not overlapping with H3K27ac peaks in HCC3153 cells. Group 1 shows peaks specific to TRPS1 (10,484 peaks), group 2 shows peaks specific to H3K27ac (31,766 peaks), and group 3 shows peaks common for both proteins (5,785 peaks).
(E) Binding and expression target analysis (BETA) of associations between gene expression changes after *TRPS1* knockdown and TRPS1 chromatin occupancy at unique and TRPS1-H3K27ac-overlapping sites.
(F) Examples of TRPS1 peaks.
(G) Boxplot depicting changes in H3K27ac signal at TRPS1-unique and TRPS1-H3K27ac-overlapping peaks after shTRPS1 expression. Box plots shown as median, 25th and 75th percentiles; points are outliers above or below 1.5 × the interquartile range.
(H) Top 10 upregulated and downregulated process networks for genes differentially expressed after TRPS1 knockdown and associated with TRPS1 peaks in HCC3153 cells.
See also Figure S4 and Tables S3 and S4.



(legend on next page)

(TGF-β) signaling (Gui et al., 2013). Coordinated regulation of cell adhesion and mitosis are critical during epithelial growth, and abnormalities in these processes could lead to chromosome missegregation and tumorigenesis (Ragkousi and Gibson, 2014; Ramkumar and Baum, 2016). Although further studies are required to test these hypotheses, our data highlight TRPS1 as a crucial TF in normal mammary epithelial cells and in breast cancer. Because toxicities resulting from TRPS1 interference in adults may be restricted to the mammary gland, targeting TRPS1 may be a feasible strategy for treating a sizable fraction of breast tumors.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Cell line characterization and shRNA screen
  - shRNA libraries
  - Breast cancer cohort data and gene expression data
- METHOD DETAILS
  - Cell line derivatives and *in vitro* and *in vivo* growth assays
  - FACS analysis, immunofluorescent staining, and live cell imaging
  - ChIP-seq profiling
  - Immunoblotting, co-immunoprecipitation (CoIP), and rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME)
  - Genomic profiling
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Genomic data analyses
  - Gene signature analysis
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes five figures, four tables, and one video and can be found with this article online at https://doi.org/10.1016/j.celrep.2018.10.023.

ACKNOWLEDGMENTS

We thank members of our laboratories for their critical reading of this manuscript and useful discussions. We thank Patrick Reynolds and the Children's Oncology Group/Alex's Lemonade Stand Foundation (COG/ALSF) Childhood Cancer Repository (www.CCcells.org) for the neuroblastoma cell lines. This research was supported by the National Cancer Institute (NCI) R35 CA197623 (K.P.), K99 CA201606 (M.J.), and P01 CA080111 (K.P., M.B.), National Institute of Neurological Disorders and Stroke (NINDS) R01 NS088355 (K.S.), Susan G. Komen Foundation (S.S.), the Ludwig Center at Harvard (K.P. and M.B.), and the Novartis Institute for Biomedical Research (K.P., M.B., K.S., C.R., R.S.). The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the NIH, and by the NCI, the National Human Genome Research Institute (NHGRI), the National Heart, Lung, and Blood Institute (NHLBI), the National Institute on Drug Abuse (NIDA), the National Institute of Mental Health (NIMH), and NINDS. The data used for the analyses described in this manuscript were obtained from the GTEx Portal on December 12, 2017.

AUTHOR CONTRIBUTIONS

R.M.W. performed the ChIP-seq, RNA-seq, and cell culture experiments, M.B.E., X.Q., A.T., and Y.X. completed the data analyses and software development. M.J. and S. Shu. Performed the *in vivo* experiments. M.K. performed the live-cell imaging. E.F., N.R., G.H., K.Y., G.M., R.M., J.G., M.S., A.d.W., N.K., H.M.C., D.R., T.R., S. Sovath, S. Silver, W.R.S., and Z.J. helped with the shRNA screen and data analyses. M.D.H., H.L., M.B., K.S., C.R., D.P., and R.A.S. provided the reagents and resources. K.P. supervised the study. All of the authors helped to design the study and write the manuscript.

DECLARATION OF INTERESTS

K.P., M.B., K.S., C.R., and R.A.S. received research support and were consultants to Novartis Institutes for BioMedical Research during the execution of this study. K.P. and M.B. serve on the scientific advisory board (SAB) of Mitra Biotech and Kronos Bio, respectively. E.F., N.R., G.H., K.Y., G.M., R.M., J.G., M.S., A.d.W., N.K., H.M.C., D.R., T.R., S. Sovath, S. Silver, W.R.S., and Z.J. are current and former employees of Novartis Institutes for BioMedical Research. H.M.C. is a current employee of Foghorn Therapeutics. W.R.S. is a shareholder in Novartis and Peloton Therapeutics; serves on the SAB of Ideaya Biosciences, Epidarex Capital, and Ibsen Pharmaceuticals; and is a consultant to Astex, Array, Servier, and Sanofi Pharmaceuticals.

Received: May 29, 2018
Revised: September 9, 2018
Accepted: October 3, 2018
Published: October 30, 2018

Figure 5. TRPS1 Chromatin Binding Partners and Their Changes after TRPS1 Knockdown

(A) Word cloud depicting the top 50 TRPS1-associated proteins identified by RIME in HCC3153 and MDA-IBC3 cell lines. Font size reflects the ratio of peptide count in anti-TRPS1 versus control IgG immunoprecipitation. Colors indicate NuRD (red) and switch/sucrose non-fermentable (SWI/SNF) (blue) complex proteins. TRPS1 signal (gray) is represented in the background.

(B) Top 10 pathways enriched for TRPS1-interacting proteins in both HCC3153 and MDA-IBC3 cell lines.

(C) Heatmap of TRPS1 (group 1), GATAD2A (group 2), and GATAD2B (group 3)-unique, TRPS1 and GATAD2A (group 4), GATAD2A and GATAD2B (group 5), TRPS1 and GATAD2B (group 6), and common for all 3 marks (group 7) peaks in HCC3153 cells. Numbers of peaks in each group are indicated.

(D) Integrated analysis of gene expression changes after *TRPS1* knockdown and TRPS1 chromatin occupancy for chromatin regions bound by TRPS1 only or TRPS1 and GATAD2B.

(E) Top 10 upregulated and downregulated process networks for genes differentially expressed after TRPS1 knockdown and associated with TRPS1-GATAD2B-overlapping peaks in HCC3153 cells.

(F) Examples of TRPS1 and GATAD2B-overlapping peaks.

(G) Heat-map of GATAD2A and GATAD2B peaks in HCC3153 cells before and after *TRPS1* knockdown. Each group represents peaks for TRPS1 (group 1), GATAD2A (group 2), GATAD2B (group 3), TRPS1 and GATAD2A (group 4), GATAD2A and GATAD2B (group 5), TRPS1 and GATAD2B (group 6), common for all 3 marks (group 7).

See also Figure S5.

REFERENCES

Barbie, D.A., Tamayo, P., Boehm, J.S., Kim, S.Y., Moody, S.E., Dunn, I.F., Schinzel, A.C., Sandy, P., Meylan, E., Scholl, C., et al. (2009). Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. Nature *462*, 108–112.

Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A.A., Kim, S., Wilson, C.J., Lehár, J., Kryukov, G.V., Sonkin, D., et al. (2012). The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature *483*, 603–607.

Brien, G.L., Valerio, D.G., and Armstrong, S.A. (2016). Exploiting the epigenome to control cancer-promoting gene-expression programs. Cancer Cell *29*, 464–476.

Cimini, D., Mattiuzzo, M., Torosantucci, L., and Degrassi, F. (2003). Histone hyperacetylation in mitosis prevents sister chromatid separation and produces chromosome segregation defects. Mol. Biol. Cell *14*, 3821–3833.

Cornwell, M., Vangala, M., Taing, L., Herbert, Z., Köster, J., Li, B., Sun, H., Li, T., Zhang, J., Qiu, X., et al. (2018). VIPER: Visualization Pipeline for RNA-seq, a Snakemake workflow for efficient and complete RNA-seq analysis. BMC Bioinformatics *19*, 135.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Du, P., Zhang, X., Huang, C.C., Jafari, N., Kibbe, W.A., Hou, L., and Lin, S.M. (2010). Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. BMC Bioinformatics *11*, 587.

Elster, D., Tollot, M., Schlegelmilch, K., Ori, A., Rosenwald, A., Sahai, E., and von Eyss, B. (2018). TRPS1 shapes YAP/TEAD-dependent transcription in breast cancer cells. Nat. Commun. *9*, 3115.

Fantauzzo, K.A., and Christiano, A.M. (2012). Trps1 activates a network of secreted Wnt inhibitors and transcription factors crucial to vibrissa follicle morphogenesis. Development *139*, 203–214.

Gröbner, S.N., Worst, B.C., Weischenfeldt, J., Buchhalter, I., Kleinheinz, K., Rudneva, V.A., Johann, P.D., Balasubramanian, G.P., Segura-Wang, M., Brabetz, S., et al.; ICGC PedBrain-Seq Project; ICGC MMML-Seq Project (2018). The landscape of genomic alterations across childhood cancers. Nature *555*, 321–327.

Gui, T., Sun, Y., Gai, Z., Shimokado, A., Muragaki, Y., and Zhou, G. (2013). The loss of Trps1 suppresses ureteric bud branching because of the activation of TGF-β signaling. Dev. Biol. *377*, 415–427.

Jones, P.A., Issa, J.P., and Baylin, S. (2016). Targeting the cancer epigenome for therapy. Nat. Rev. Genet. *17*, 630–641.

Justin, N., De Marco, V., Aasland, R., and Gamblin, S.J. (2010). Reading, writing and editing methylated lysines on histone tails: new insights from recent structural studies. Curr. Opin. Struct. Biol. *20*, 730–738.

Kim, K.H., and Roberts, C.W. (2014). Mechanisms by which SMARCB1 loss drives rhabdoid tumor growth. Cancer Genet. *207*, 365–372.

König, R., Chiang, C.Y., Tu, B.P., Yan, S.F., DeJesus, P.D., Romero, A., Bergauer, T., Orth, A., Krueger, U., Zhou, Y., and Chanda, S.K. (2007). A probability-based approach for the analysis of large-scale RNAi screens. Nat. Methods *4*, 847–849.

Lawrence, M.S., Stojanov, P., Mermel, C.H., Robinson, J.T., Garraway, L.A., Golub, T.R., Meyerson, M., Gabriel, S.B., Lander, E.S., and Getz, G. (2014). Discovery and saturation analysis of cancer genes across 21 tumour types. Nature *505*, 495–501.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

Maas, S.M., Shaw, A.C., Bikker, H., Lüdecke, H.J., van der Tuin, K., Badura-Stronka, M., Belligni, E., Biamino, E., Bonati, M.T., Carvalho, D.R., et al. (2015). Phenotype and genotype in 103 patients with tricho-rhino-phalangeal syndrome. Eur. J. Med. Genet. *58*, 279–292.

Mack, S.C., Witt, H., Piro, R.M., Gu, L., Zuyderduyn, S., Stütz, A.M., Wang, X., Gallo, M., Garzia, L., Zayne, K., et al. (2014). Epigenomic alterations define lethal CIMP-positive ependymomas of infancy. Nature *506*, 445–450.

Malik, T.H., Shoichet, S.A., Latham, P., Kroll, T.G., Peters, L.L., and Shivdasani, R.A. (2001). Transcriptional repression and developmental functions of the atypical vertebrate GATA protein TRPS1. EMBO J. *20*, 1715–1725.

Malik, T.H., Von Stechow, D., Bronson, R.T., and Shivdasani, R.A. (2002). Deletion of the GATA domain of TRPS1 causes an absence of facial hair and provides new insights into the bone disorder in inherited tricho-rhino-phalangeal syndromes. Mol. Cell. Biol. *22*, 8592–8600.

Marcotte, R., Sayad, A., Brown, K.R., Sanchez-Garcia, F., Reimand, J., Haider, M., Virtanen, C., Bradner, J.E., Bader, G.D., Mills, G.B., et al. (2016). Functional genomic landscape of human breast cancer drivers, vulnerabilities, and resistance. Cell *164*, 293–309.

McDonald, E.R., 3rd, de Weck, A., Schlabach, M.R., Billy, E., Mavrakis, K.J., Hoffman, G.R., Belur, D., Castelletti, D., Frias, E., Gampa, K., et al. (2017). Project DRIVE: a compendium of cancer dependencies and synthetic lethal relationships uncovered by large-scale, deep RNAi screening. Cell *170*, 577–592.e10.

Mohammed, H., D'Santos, C., Serandour, A.A., Ali, H.R., Brown, G.D., Atkins, A., Rueda, O.M., Holmes, K.A., Theodorou, V., Robinson, J.L., et al. (2013). Endogenous purification reveals GREB1 as a key estrogen receptor regulatory factor. Cell Rep. *3*, 342–349.

Momeni, P., Glöckner, G., Schmidt, O., von Holtum, D., Albrecht, B., Gillessen-Kaesbach, G., Hennekam, R., Meinecke, P., Zabel, B., Rosenthal, A., et al. (2000). Mutations in a new gene, encoding a zinc-finger protein, cause tricho-rhino-phalangeal syndrome type I. Nat. Genet. *24*, 71–74.

Napierala, D., Sam, K., Morello, R., Zheng, Q., Munivez, E., Shivdasani, R.A., and Lee, B. (2008). Uncoupling of chondrocyte differentiation and perichondrial mineralization underlies the skeletal dysplasia in tricho-rhino-phalangeal syndrome. Hum. Mol. Genet. *17*, 2244–2254.

Nikolsky, Y., Kirillov, E., Zuev, R., Rakhmatulin, E., and Nikolskaya, T. (2009). Functional analysis of OMICs data and small molecule compounds in an integrated "knowledge-based" platform. Methods Mol. Biol. *563*, 177–196.

Nishioka, K., Itoh, S., Suemoto, H., Kanno, S., Gai, Z., Kawakatsu, M., Tanishima, H., Morimoto, Y., Hatamura, I., Yoshida, M., and Muragaki, Y. (2008). Trps1 deficiency enlarges the proliferative zone of growth plate cartilage by upregulation of Pthrp. Bone *43*, 64–71.

Pasquinelli, A.E. (2018). A rADAR defense against RNAi. Genes Dev. *32*, 199–201.

Perdomo, J., and Crossley, M. (2002). The Ikaros family protein Eos associates with C-terminal-binding protein corepressors. Eur. J. Biochem. *269*, 5885–5892.

Plass, C., Pfister, S.M., Lindroth, A.M., Bogatyrova, O., Claus, R., and Lichter, P. (2013). Mutations in regulators of the epigenome and their connections to global chromatin patterns in cancer. Nat. Rev. Genet. *14*, 765–780.

Pugh, T.J., Morozova, O., Attiyeh, E.F., Asgharzadeh, S., Wei, J.S., Auclair, D., Carter, S.L., Cibulskis, K., Hanna, M., Kiezun, A., et al. (2013). The genetic landscape of high-risk neuroblastoma. Nat. Genet. *45*, 279–284.

Qin, Q., Mei, S., Wu, Q., Sun, H., Li, L., Taing, L., Chen, S., Li, F., Liu, T., Zang, C., et al. (2016). ChiLin: a comprehensive ChIP-seq and DNase-seq quality control and analysis pipeline. BMC Bioinformatics *17*, 404.

Radvanyi, L., Singh-Sandhu, D., Gallichan, S., Lovitt, C., Pedyczak, A., Mallo, G., Gish, K., Kwok, K., Hanna, W., Zubovits, J., et al. (2005). The gene associated with trichorhinophalangeal syndrome in humans is overexpressed in breast cancer. Proc. Natl. Acad. Sci. USA *102*, 11005–11010.

Ragkousi, K., and Gibson, M.C. (2014). Cell division and the maintenance of epithelial order. J. Cell Biol. *207*, 181–188.

Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. *44*, W160–W165.

Ramkumar, N., and Baum, B. (2016). Coupling changes in cell shape to chromosome segregation. Nat. Rev. Mol. Cell Biol. *17*, 511–521.

Rangel, R., Lee, S.C., Hon-Kim Ban, K., Guzman-Rojas, L., Mann, M.B., Newberg, J.Y., Kodama, T., McNoe, L.A., Selvanesan, L., Ward, J.M., et al. (2016). Transposon mutagenesis identifies genes that cooperate with mutant Pten in breast cancer progression. Proc. Natl. Acad. Sci. USA *113*, E7749–E7758.

Romani, B., and Cohen, E.A. (2012). Lentivirus Vpr and Vpx accessory proteins usurp the cullin4-DDB1 (DCAF1) E3 ubiquitin ligase. Curr. Opin. Virol. *2*, 755–763.

Sausen, M., Leary, R.J., Jones, S., Wu, J., Reynolds, C.P., Liu, X., Blackford, A., Parmigiani, G., Diaz, L.A., Jr., Papadopoulos, N., et al. (2013). Integrated genomic analyses identify ARID1A and ARID1B alterations in the childhood cancer neuroblastoma. Nat. Genet. *45*, 12–17.

Serandour, A.A., Mohammed, H., Miremadi, A., Mulder, K.W., and Carroll, J.S. (2018). TRPS1 regulates oestrogen receptor binding and histone acetylation at enhancers. Oncogene, Published online June 12, 2018. https://doi.org/10.1038/s41388-018-0312-2.

Shao, D.D., Tsherniak, A., Gopal, S., Weir, B.A., Tamayo, P., Stransky, N., Schumacher, S.E., Zack, T.I., Beroukhim, R., Garraway, L.A., et al. (2013). ATARiS: computational quantification of gene suppression phenotypes from multisample RNAi screens. Genome Res. *23*, 665–678.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. *28*, 511–515.

Tsherniak, A., Vazquez, F., Montgomery, P.G., Weir, B.A., Kryukov, G., Cowley, G.S., Gill, S., Harrington, W.F., Pantel, S., Krill-Burger, J.M., et al. (2017). Defining a cancer dependency map. Cell *170*, 564–576.e16.

Wang, S., Sun, H., Ma, J., Zang, C., Wang, C., Wang, J., Tang, Q., Meyer, C.A., Zhang, Y., and Liu, X.S. (2013). Target analysis by integration of transcriptome and ChIP-seq data with BETA. Nat. Protoc. *8*, 2502–2515.

Wang, T., Birsoy, K., Hughes, N.W., Krupczak, K.M., Post, Y., Wei, J.J., Lander, E.S., and Sabatini, D.M. (2015). Identification and characterization of essential genes in the human genome. Science *350*, 1096–1101.

Wiederschain, D., Wee, S., Chen, L., Loo, A., Yang, G., Huang, A., Chen, Y., Caponigro, G., Yao, Y.M., Lengauer, C., et al. (2009). Single-vector inducible lentiviral RNAi system for oncology target validation. Cell Cycle *8*, 498–504.

Wuelling, M., Pasdziernik, M., Moll, C.N., Thiesen, A.M., Schneider, S., Johannes, C., and Vortkamp, A. (2013). The multi zinc-finger protein Trps1 acts as a regulator of histone deacetylation during mitosis. Cell Cycle *12*, 2219–2232.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-seq (MACS). Genome Biol. *9*, R137.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit polyclonal anti-TRPS1	Bethyl	Cat # A303-563A; RRID: AB_10949357
Rabbit polyclonal anti-p66 alpha	abcam	Cat # ab87663; RRID: AB_1952305
Rabbit polyclonal anti-p66 beta	Bethyl	Cat # A301-283A; RRID: AB_937930
Rabbit polyclonal anti-SNF2H	abcam	Cat # ab3749; RRID: AB_2191856
Mouse monoclonal anti-alpha tubulin	Sigma-Aldrich	Cat # T9026; RRID: AB_477593
Rabbit polyclonal anti-pHistone 3 (Ser 10)	Millipore	Cat # 06-570; RRID: AB_310177
Rabbit polyclonal anti-MTA2	abcam	Cat # ab8106; RRID: AB_306276
Mouse monoclonal anti-HDAC2	abcam	Cat # ab12169; RRID: AB_2118547
Rabbit polyclonal anti-H3K27ac	Diagenode	Cat # C15410196; RRID: AB_2637079
Mouse monoclonal anti-H3K27me3	abcam	Cat # ab6002; RRID: AB_305237
Chemicals, Peptides, and Recombinant Proteins		
TransIT-293 Transfection Reagent	Mirus	Cat # MIR2700
Polybrene	Millipore	Cat # TR-1003-G
Puromycin	InvivoGen	Cat # ant-pr-1
Critical Commercial Assays		
CellTiter-Glo Luminescent Cell Viability Assay	Promega	Cat # G7573
ThruPLEX DNA-seq 48S Kit	Rubicon	Cat # R400427
QIAamp DNA Blood Maxi kit	QIAGEN	Cat # 51194
QIAamp DNA Mini kit	QIAGEN	Cat # 51304
RNeasy Mini Kit	QIAGEN	Cat # 74106
Alexa Fluor 488 Annexin V/Dead Cell Apoptosis Kit	Life Technologies	Cat # V13241
Deposited Data		
Raw genomic data	NCBI GEO	GEO: GSE114600
Experimental Models: Cell Lines		
HEK293T cells	ATCC	Cat # CRL-11268
59 cancer cell lines used in the screen (Table S1)	various	N/A
HCC3153 sh*TRPS1*-1 cells	This paper	N/A
HCC3153 sh*TRPS1*-1 RFP-H2B cells	This paper	N/A
HCC3153 sh*TRPS1*-2 cells	This paper	N/A
HCC3153 sh*TRPS1*-2 RFP-H2B cells	This paper	N/A
Oligonucleotides		
sh*TRPS1*-1a: CCGGCTGAGGTCCTGACAAGCGATACTCGAGTA TCGCTTGTCAGGACCTCAGTTTTT	This paper	N/A
sh*TRPS1*-1b: AATTAAAAACTGAGGTCCTGACAAGCGATACTC GAGTATCGCTTGTCAGGACCTCAG	This paper	N/A
sh*TRPS1*-2a: CCGGGAGGTCCTGACAAGCGATAACCTCGAGG TTATCGCTTGTCAGGACCTCTTTTT	This paper	N/A
sh*TRPS1*-2b: AATTAAAAAGAGGTCCTGACAAGCGATAA CCTCGAGGTTATCGCTTGTCAGGACCTC	This paper	N/A
Recombinant DNA		
pRSI16cb-U6-sh-13kCB18-HTS6-UbiC-TagRFP-2A-Puro	McDonald et al., 2017	Cellecta Cat # SVSHU616-L
13K-hTF	McDonald et al., 2017	Cellecta Cat # 13K-hTF-GH-NOVA
13K-hEPI2	McDonald et al., 2017	Cellecta Cat # 13K-hEPI2-GHNOVA
Lentiviral Packaging Plasmid Mix	Cellecta	Cat # CPCP-K2A
Tet-pLKO-puro	Wiederschain et al., 2009	Addgene Cat # 21915
pBABE-RFP-H2B	David Pellman (DFCI)	N/A

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Kornelia Polyak, Dana-Farber Cancer Institute, 450 Brookline Ave., Boston, MA 02215, USA. E-mail: kornelia_polyak@dfci.harvard.edu; tel: 617-632-2106; fax: 617-582-8490.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell line characterization and shRNA screen

All cell lines used in this study, together with correspondent culture media, are listed in Table S1. The identity of the cell lines was confirmed based on STR and exome-seq analyses. Cells were regularly tested for mycoplasma. Each cell line was tested for optimal plating density and puromycin sensitivity by plating at three different densities ($1x10^5$ to $4x10^5$ cells/well in 6-well plates), culturing without and with 5 different concentrations of puromycin (0.25 to 4 μg/ml; InvivoGen, cat# ant-pr-1) for 72hrs, and tested by CellTiter-Glo assay (Promega, cat# G7573) as a readout to achieve over 95% selection efficacy. To ensure infection with only one shRNA/cell, multiplicity of infection (MOI) was determined by transducing the cell lines with 12 different doses of reference lentivirus suspension (0 to 400 μl) containing the same vector backbone as a library vector, but expressing RFP (Red Fluorescent Protein), in the presence of 5 μg/ml polybrene (Millipore, cat# TR-1003-G). 72hrs after starting puromycin selection cells were analyzed by FACS and MOI yielding 50% RFP^+ cells was chosen for each cell line. For the shRNA screen all cell lines were transduced with lentiviral libraries at 1,000 cells/shRNA and cultured for 15 days after puromycin selection. For each split, cells were re-plated to ensure the same minimum representation of 1,000 cells/shRNA. After selection cells were harvested and genomic DNA was isolated using QIAamp DNA Blood Maxi kit (QIAGEN, cat# 51194). Library preparation and DNA sequencing was performed by Novartis Institutes for Biomedical Research as described (McDonald et al., 2017).

shRNA libraries

Pooled shRNA libraries were designed by Novartis Institutes for Biomedical Research based on pRSI16 backbone from Cellecta (Cellecta, cat# SVSHU616-L). EpiGenome2.0 library (Cellecta, cat# 13K-hEPI2-GHNOVA) was directed at epigenetic regulators, while hTF1 library (Cellecta, cat# 13K-hTF-GH-NOVA) was targeting TFs highly expressed in human cancers based on the Cancer Cell Line Encyclopedia (Barretina et al., 2012). Each library contained 13,000 shRNAs targeting 650 genes with 20 different shRNAs/gene. Plasmid pools were packaged into lentivirus by transfecting HEK293T cells (ATCC, cat# CRL-11268) using TransIT-293 transfection reagent (Mirus, cat# MIR2700) and Cellecta packaging mix (Cellecta, cat# CPCP-K2A) diluted in Opti-MEM (Life Technologies, cat#31985070). Virus was harvested 72hrs after transfection, aliquoted, and stored at −80°C until use.

Breast cancer cohort data and gene expression data

We obtained the genetic alteration data from the cBioPortal database (http://www.cbioportal.org), gene expression data for various cancer types from The Cancer Genome Atlas (TCGA) database (https://gdac.broadinstitute.org/), and gene expression data in normal tissues from the GTEx Portal (https://gtexportal.org/home/).

METHOD DETAILS

Cell line derivatives and *in vitro* and *in vivo* growth assays

HCC3153 and SUM159 cell lines were transduced with Tet-pLKO-puro lentiviral vectors containing shRNAs against *LacZ* or *TRPS1* transcripts, in the presence of 5 μg/ml polybrene and subsequently selected for 72hrs with puromycin. Viability of the cells expressing sh*TRPS1* was assessed using Celigo system (Nexcelom, Celigo Image Cytometer). Cells expressing doxycycline inducible sh*TRPS1* or sh*LacZ* were plated in triplicates in 6-well plates and cultured at 37°C with 5% CO2. 24hrs after plating cells were treated with 100ng/ml doxycycline to induce shRNA expression. Culture medium was replaced every 48hrs with or without freshly prepared doxycycline. Cell viability was measured every 24hrs (SUM159) or 48hrs (HCC3153) beginning at 48hrs from the start of treatment. All xenograft experiments were performed following Dana-Farber Cancer Institute ACUC approved protocol #11-023. Mammary fat pads of 6-weeks-old female NOG (NOD.Cg-*Prkdc*^{scid} *Il2rg*^{tm1Sug}/JicTac) mice (Taconic) were injected with $2x10^6$ cells in 50% Matrigel (BD Biosciences, cat# 354234), 50% DMEM media (Corning, cat# 10-013-CV) in a total volume of 50 μl. shRNA expression was induced by switching the animals to doxycycline diet after all tumors reached palpable size. Tumor growth was monitored weekly (for HCC3153) or twice a week (for SUM159) with caliper measurements. At the end of the experiment mice were euthanized, tumors were collected, fixed overnight in 4% formalin, stored in 70% ethanol, followed by paraffin embedding, sectioning, and hematoxylin and eosin staining by the Pathology Core of the Brigham and Women's Hospital.

FACS analysis, immunofluorescent staining, and live cell imaging

For the cell cycle analysis cells expressing sh*TRPS1* and control cells were fixed with 70% ethanol for at least 24hrs at 4°C and then washed with ice-cold PBS. Cells were resuspended in 500μl of staining solution (0.1% Triton X-100, 200μg/ml RNase A, 20μg/ml propidium iodide; in PBS) and incubated for 30 min. at room temperature. For the apoptosis analysis cells were stained with annexin

V and propidium iodide using Alexa Fluor 488 Annexin V/Dead Cell Apoptosis kit (Life Technologies) according to the manufacturer's protocol. Signal was acquired on a BD LSRFortessa cytometer (BD Biosciences) and analyzed using FlowJo software. Immunofluorescent staining was performed on cells grown on coverslips and fixed using 4% paraformaldehyde for 20 min. at room temperature. Cells were washed with PBS, permeabilized with 0.3% Triton X-100 in PBS for 3 min. and then blocked with 5% BSA in PBS for 30 min. Samples were incubated with a primary antibody for 45 min., washed with PBS,

incubated with a secondary antibody for 45 min., washed with PBS and distilled water. Coverslips were mounted on glass slides and observed using a Nikon Ti-E inverted microscope (Nikon Instruments, Melville, NY) with a Yokogawa CSU-22 spinning disk confocal head with the Borealis modification. Laser excitation was with 488 nm and 561 nm lasers. Images were acquired using a X60 Plan Apo NA 1.4 oil objective with a CoolSnapHQ2 CCD camera (Photometrics). Acquisition parameters, shutters, filter positions and focus were controlled by Metamorph software (Molecular Devices).

Live-cell imaging was performed on HCC3153 cells expressing RFP-H2B cells that were plated onto glass-bottomed imaging μ-dishes (ibidi). Images were acquired from 5 z sections with a 2 μm step size using a X20 objective lens at 10 min intervals during 24 hr. All images were collected on a Nikon Ti inverted microscope equipped with a Yokogawa CSU-X1 spinning disk confocal head, Spectral Applied Precision LMM-5 with AOTF, a Hamamatsu ORCA ER cooled CCD camera and the Nikon Perfect Focus System. RFP was imaged using a 561nm laser for excitation and a 620/60 emission filter. An Okolab cage incubator was used to maintain samples at 37°C and 5% humidified CO2. Image acquisition was controlled with MetaMorph software (Molecular Devices). For the analysis of mitotic duration, time between mitotic cell rounding and mitotic exit, as judged by cell re-adhesion, was quantified. This is because the majority (~80%) of cells expressing TRPS1 shRNA exited mitosis without chromosome separation after prolonged mitotic arrest, thus making it impossible to define anaphase onset for mitotic duration. Bright field images were used to invariably determine mitotic cell rounding and re-adhesion.

ChIP-seq profiling

For each ChIP-seq, 5-8x10^6 cells were fixed by adding 1:10 volume of fixing buffer (11% paraformaldehyde (Electron Microscopy Sciences, cat# 15714), 0.1M NaCl, 1mM EDTA pH 8.0, 50mM HEPES pH 8.0) directly to the tissue culture medium. All ChIP-seq experiments were performed in duplicates. For histone modifications, fixation was performed for 10 min. at room temperature, while for TFs for 10 min. at 37°C. Fixing agent was quenched with 0.125M glycine for 5 min. and cells were washed twice with ice cold PBS, collected by centrifugation, and stored at −80°C. Cells were lysed in 1ml of cell lysis buffer (50mM HEPES pH 8.0, 140mM NaCl, 1mM EDTA pH 8.0, 10% glycerol, 0.5% NP-40, 0.25% Triton X-100) for 10 min. at 4°C, pelleted, and washed in 1ml of wash buffer (10mM Tris-HCl pH 8.0, 200mM NaCl, 1mM EDTA pH 8.0, 0.5mM EGTA) for 10 min. at 4°C. Pellets were resuspended in 1ml of sonication buffer (10mM Tris-HCl pH 7.4, 1mM EDTA pH 8.0, 0.1% SDS, 1% Triton X-100, 0.1% sodium deoxycholate, 0.25% sarkosyl, 1mM DTT, protease and phosphatase inhibitors (Thermo Fisher Scientific, cat# 1861279 and 78427)) and sonicated using a Covaris E220 sonicator. Debris was removed by centrifugation (5 min. at 10,000 g) and 970μl of the lysate was mixed with 30μl of 5M NaCl. Chromatin was then pre-cleared with 40μl of Dynabeads Protein G (Life Technologies, cat# 10003D) washed in 0.5% BSA in PBS for 1hr at 4°C. Magnetic beads were removed and cleared lysates were incubated with the antibody overnight at 4°C. Complexes were collected by incubation with 40μl of washed Dynabeads Protein G for 2hrs at 4°C, washed with low salt wash buffer (20mM Tris-HCl pH 8.0, 150mM NaCl, 2mM EDTA pH 8.0, 1% Triton X-100, 0.1% SDS), high salt wash buffer (20mM Tris-HCl pH 8.0, 500mM NaCl, 2mM EDTA pH 8.0, 1% Triton X-100, 0.1% SDS), LiCl wash buffer (10mM Tris-HCl pH 8.0, 250mM LiCl, 1mM EDTA pH 8.0, 1% NP40, 1% sodium deoxycholate), and twice with TE pH 8.0. DNA was eluted in 100μl of elution buffer (100mM NaHCO3, 1% SDS) and cross-link was reversed by incubating overnight at 65°C, followed by 30 min. treatment with RNase A at 37°C and 2hrs treatment with proteinase K at 55°C. DNA was purified with phenol-chloroform extraction and precipitated with isopropanol. Libraries were prepared using Rubicon ThruPLEX DNA-seq kit (Rubicon, cat# R400427) from 1ng DNA or 10ng of input DNA following manufacturer's protocol and sequenced on the Illumina NextSeq500 platform with 75bp single end reads at the Molecular Biology Core Facilities at the Dana-Farber Cancer Institute.

Immunoblotting, co-immunoprecipitation (CoIP), and rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME)

For immunoblotting, cell pellets were lysed in RIPA buffer for 30 min. on ice and sonicated in a cup sonicator (QSonica, cat# Q500). For CoIP, 5-8x10^6 cells were lysed in 1ml of cell lysis buffer (10mM HEPES pH 8.0, 1.5mM MgCl2, 10mM KCl, 0.5% NP-40, 0.5mM DTT, protease and phosphatase inhibitors) for 10 min. at 4°C with rotation, pelleted, resuspended in 350μl of nuclear lysis buffer (10mM HEPES pH 8.0, 1.5mM MgCl2, 420mM NaCl, 0.2mM EDTA pH 8.0, 0.5% NP-40, 25% glycerol, 0.5mM DTT, protease and phosphatase inhibitors), and incubated for 20 min. at 4°C with rotation. DNA was sheared by passing the solution through a 28G needle 10 times and debris was removed by centrifugation (10 min. at 13,000rpm). 1/10 of the supernatant volume was kept as an input and the rest was diluted 1:1 with dilution buffer (20mM Tris-HCl pH 8.0, 1.5mM MgCl2, 0.2mM EDTA pH 8.0, 0.5% NP-40), supplemented with MgCl2 and DNase I (final concentrations of 3mM and 20U/ml, respectively), and incubated 30 min. at 37°C. 50μl of Dynabeads Protein G washed twice in 1:1 solution of nuclear lysis and dilution buffers were added to the sample and incubated for 60 min at 4°C with rotation to pre-clear the lysates. Beads were removed with a magnet and the sample was incubated with the antibody of interest or IgG control overnight at 4°C. 50μl of washed beads were added and incubated for 2hrs at 4°C to precipitate the complexes and subsequently washed twice with 1ml of CoIP wash buffer (20mM Tris-HCl pH 8.0, 1.5mM MgCl2, 0.2mM EDTA

pH 8.0, 0.15mM NaCl, 0.5% NP-40) and once with 1ml of LoTE. Proteins for immunoblotting and CoIP were resolved on 3%–8% polyacrylamide gels (Life Technologies, cat# EA03785BOX) and transferred to PVDF membranes (EMD Millipore, cat# IPVH00010). Membranes were blocked with 5% milk in 0.1% Tween20 in TBS (TBS-T) for 1hr at room temperature. Incubation with primary antibodies was performed overnight at 4°C in 2.5% milk TBS-T and with secondary antibodies for 1hr at room temperature in 2.5% milk TBS-T. Membranes were washed and then developed with Immobilon (EMD Millipore, cat# WBKLS0500) or Pierce ECL substrate (Thermo Fisher Scientific, cat# 32106). For RIME, $2x10^8$ cells were fixed using the same fixing buffer as used for ChIP-seq experiments, but with time extended to 15 min. at room temperature, and then quenched with 0.125M glycine for 5 min. After two washes with ice cold PBS cells were collected, flash-frozen in liquid nitrogen, and submitted to Active Motif for RIME analysis.

Genomic profiling

For exome sequencing genomic DNA was isolated using QIAamp DNA Mini kit (QIAGEN, cat# 51304). Exome sequencing was performed by Personal Genome Diagnostics using their CancerXOME sequencing service. DNA methylation profiling was carried out on Infinium HumanMethylation450K BeadChip arrays (Illumina, discontinued) at the Harvard Medical School-Partners HealthCare Center for Genetics and Genomics. For RNA-seq, mRNA from exponentially growing cells was isolated using RNeasy Mini kit (QIAGEN, cat# 74106) according to the manufacturer's protocol. Libraries were prepared at the Center for Functional Cancer Epigenetics and sequenced on the Illumina NextSeq500 platform with 75bp single end reads at the Molecular Biology Core Facilities at the Dana-Farber Cancer Institute. For histone mass spectrometry cells were collected by trypsinization and three aliquots of $2x10^6$ cells were analyzed by Novartis Institutes for Biomedical Research as previously described (McDonald et al., 2017).

QUANTIFICATION AND STATISTICAL ANALYSIS

Genomic data analyses

The genomic data for the pooled shRNA screen has been analyzed both the ATARiS (Shao et al., 2013) and the RSA (König et al., 2007) algorithms. The ATARiS scores of each gene are Z-score normalized across all the cell lines and any gene that is at least 2 standard deviations below the mean is considered to be significant for that cell line. The RSA Down scores of each cell line are Z-score normalized across all the genes and any gene that is at least 2 standard deviations below the mean is considered to be significant in that cell line. Lastly all the list of significant genes from both methods are combined (present as significant in at least one method) to have a complete list of significant screened genes in all cell lines.

The methylation status of the cell lines was obtained using the *Infinium HumanMethylation450 BeadChip Kit* from Illumina. The unnormalized probe-level beta-values (Du et al., 2010) of each sample were obtained from the raw data using the *Illumina GenomeStudio* software (GSGX v1.1.0). The normalization of the beta-values was performed using the R Bioconductor package *lumi*. The normalized M-value thus obtained were converted back to normalized beta-values using the following formula:

$$\text{Beta - value} = \left(2^{M}\text{-value}\right) / \left\{\left(2^{M}\text{-value}\right) + 1\right\} \text{(Du et al., 2010)}$$

The normalized beta-values of each sample were then collapsed to promoters and enhancers of genes. Briefly, each probe is attempted to be assigned to either a promoter, enhancer, or gene-body of a gene using the annotation provided in the output of the GenomeStudio software mentioned above. The methylation status of the enhancer of a gene is determined by taking the average normalized beta-values of all the probes assigned to the enhancer of that. Depending on the position of the probes relative to the TSS of a gene, the probes are also assigned to either promoters (TSS200, TSS1500, 1stExon, and 5′UTR) or gene-body (Body, 3′UTR) of genes. Following that, single gene-level methylation values of each gene's promoter and gene-body are obtained using the average normalized beta-values as explained for enhancers. The gene-level enhancer, promoter, and gene-body methylation values of each sample have been used to determine the correlation between the samples and their tissue of origin (lineage).

ChIP-seq QC and preprocessing of the data was performed using ChiLin pipeline 2.0.0 (Qin et al., 2016). Reads were mapped with Burrows-Wheeler Aligner tool (BWA) (Li and Durbin, 2009) and peaks called with Model-based Analysis of ChIP-Seq peak caller (MACS2) (Zhang et al., 2008). Based on a dynamic Poisson distribution MACS2 can effectively capture local biases in the genome sequence, allowing for more sensitive and robust prediction of binding sites. Unique read for a position for peak calling is used to reduce false positive peaks, statically significant peaks are finally selected by calculated false discovery rate of reported peaks. Only the peaks that were present in all replicates were used for further analysis. Motif analysis was performed using the HOMER package. Deeptools (Ramírez et al., 2016) were used to create heatmap plots, using average signal of the replicates. BETA (Wang et al., 2013) was used for integration of the ChIP-seq of TFs or chromatin regulators with differential gene expression data to infer direct target genes. RNA-seq read alignment, quality control, and data analysis were performed using VIPER (Cornwell et al., 2018). RNA-seq reads were mapped using STAR (Dobin et al., 2013). Read counts for each gene were generated by Cufflinks (Trapnell et al., 2010). Differential expression was called by Deseq2 (Love et al., 2014). For the Sample-Sample Correlation heatmap we used the expression matrix from Cufflinks (Trapnell et al., 2010). All snoRNA and miRNA genes were first filtered out as were all genes that were only significantly expressed in 3 or fewer samples based on a FPKM threshold of 5. A custom R script was used to calculate the Pearson correlation between all the samples on a pairwise basis and to generate the heatmap.

Gene signature analysis

A gene signature corresponding to the 1451 genes co-regulated by both TRPS1 and GATAD2B was determined by taking the intersection of all TRPS1 and GATAD2B peaks which lie within 10kb of a transcription start site (“all targets”). This signature was further refined to a set of 466 genes by intersecting it with the list of upregulated genes with BH adjusted p value less than 0.1 in HCC3153 sh*TRPS1* cells (“upregulated targets”). Each patient in the METABRIC set (n = 1904) was assigned a “TSS” or a “upregulated” signature score using ssGSEA, single sample gene set enrichment analysis (Barbie et al., 2009). These scores were rescaled to a [0,1] range. Survival analysis was performed using Cox proportional hazards models taking into account age, Nottingham Prognostic Index and ER/HER2 status.

DATA AND SOFTWARE AVAILABILITY

The accession number for the RNA-seq, ChIP-seq, and DNA methylation data reported in this paper is GEO: GSE114600.