PLOS GENETICS

TRPS1 modulates chromatin accessibility to regulate estrogen receptor alpha (ER) binding and ER target gene expression in luminal breast cancer cells

Thomas G. Scott[1], Kizhakke Mattada Sathyan[2,3], Daniel Gioeli[4,5], Michael J. Guertin[2,3]*

1 Department of Biochemistry and Molecular Genetics, University of Virginia, Charlottesville, Virginia, United States of America, 2 Center for Cell Analysis and Modeling, University of Connecticut, Farmington, Connecticut, United States of America, 3 Department of Genetics and Genome Sciences, University of Connecticut, Farmington, Connecticut, United States of America, 4 Department of Microbiology, Immunology, and Cancer, University of Virginia, Charlottesville, Virginia, United States of America, 5 Cancer Center Member, University of Virginia, Charlottesville, Virginia, United States of America

* guertin@uchc.edu

Abstract

Common genetic variants in the repressive GATA-family transcription factor (TF) *TRPS1* locus are associated with breast cancer risk, and luminal breast cancer cell lines are particularly sensitive to *TRPS1* knockout. We introduced an inducible degron tag into the native *TRPS1* locus within a luminal breast cancer cell line to identify the direct targets of TRPS1 and determine how TRPS1 mechanistically regulates gene expression. We acutely deplete over 80 percent of TRPS1 from chromatin within 30 minutes of inducing degradation. We find that TRPS1 regulates transcription of hundreds of genes, including those related to estrogen signaling. TRPS1 directly regulates chromatin structure, which causes estrogen receptor alpha (ER) to redistribute in the genome. ER redistribution leads to both repression and activation of dozens of ER target genes. Downstream from these primary effects, TRPS1 depletion represses cell cycle-related gene sets and reduces cell doubling rate. Finally, we show that high TRPS1 activity, calculated using a gene expression signature defined by primary TRPS1-regulated genes, is associated with worse breast cancer patient prognosis. Taken together, these data suggest a model in which TRPS1 modulates the genomic distribution of ER, both activating and repressing transcription of genes related to cancer cell fitness.

Author summary

Breast cancer is the most common cancer among women. The majority of cases are luminal, which tend to be estrogen receptor alpha (ER)-positive. ER is well-studied among transcription factors (TFs) because it is ligand-activated. This allows for the rapid induction of ER activity and the identification of primary estrogen-responsive genes. Most TFs have not been so extensively characterized, because their activity is not so rapidly

perturbable. TRPS1 is an atypical GATA family TF that is associated with corepressor complexes and transcriptional repression. Here, we use an inducible degron tag system to rapidly deplete endogenously-tagged TRPS1 in luminal breast cancer cells within 30 minutes. We find that TRPS1 directly decreases local chromatin accessibility. This decreases ER binding intensity at TRPS1-proximal ER binding sites. As an indirect effect, ER binding intensity distal to TRPS1 increases in intensity. These effects on ER binding are associated with changes in ER target gene transcription, repressing or activating these genes in concordance with the effect on ER binding intensity. Our work is consistent with a model in which TFs either exclusively activate or exclusively repress transcription of their direct target genes, with indirect primary response genes changing due to the redistribution of limiting activating TFs or coactivators.

Introduction

Breast cancer is the most frequently diagnosed cancer in women, with an estimated lifetime risk of about 1 in 8 for women in the United States [1]. Far from a monolithic disease, breast tumors can be classified into subtypes based on gene expression, histology, and immunohistochemistry [2, 3]. The most common subtype is luminal breast cancer, which is typically estrogen receptor alpha (ER)-positive [4].

High lifetime exposure to endogenous estrogen is a strong risk factor for breast cancer incidence [5]. Estrogen is a potent hormone that binds to ER, a ligand-activated transcription factor (TF), which then homodimerizes, binds to reverse palindromic pairs of AGGTCA motifs on DNA, and recruits cofactors to activate hundreds of genes that promote cell growth and proliferation [6–8]. In additional to surgery, radiation, and traditional chemotherapy, endocrine therapies that inhibit endogenous estrogen production or binding to ER provide a significant survival benefit to luminal breast cancer patients [4, 9, 10]. However, patients with advanced disease frequently develop resistance to these therapies, though many endocrine therapy-resistant luminal tumors still remain dependent on ER activity [11]. Thus, there is a need to identify additional factors that regulate ER activity or genomic binding and contribute to breast tumor progression.

TRPS1 is a member of the GATA-family of TFs that bind to (A/T)GATA(A/G) motifs on DNA [12]. In contrast to the other six members of the GATA family that activate transcription, TRPS1 directly represses transcription of target genes via its unique IKZF1-like zinc fingers [13]. TRPS1 has been shown to interact with multiple corepressors and lysine deacetylases, including members of the NuRD and coREST complexes, to regulate transcription [14–18].

TRPS1 was first described as the gene mutated in cases of tricho-rhino-phalangeal syndrome, an autosomal dominant disorder characterized by developmental abnormalities of the hair, nose, and fingers [19]. *TRPS1* is crucial for the proper development of several tissues, including hair, bone, and kidney [20, 21]. As with many developmentally important genes co-opted during the process of cancer initiation and progression, *TRPS1* is commonly over-expressed in breast tumors, both relative to normal tissue and relative to other tumor types [22, 23].

The transcriptional program that TRPS1 regulates in breast cancer is not fully understood. Knockdown of *TRPS1* in various breast cancer cell lines has been shown to increase markers of epithelial to mesenchymal transition and genome instability [24–27]. Additionally, TRPS1 binding sites on chromatin overlap with those of YAP1 and ER, though this is coupled with a genome-wide activation of YAP1 target genes but repression of ER target genes [16, 17].

A key feature of these previous studies is the use of extended knockdown kinetics with traditional RNA interference methods. Days after knockdown, the resultant effects represent not only the primary TRPS1-responsive genes but also secondary and compensatory effects. As such, we do not know which genes TRPS1 directly regulates and whether these genes are important for breast cancer cell growth and proliferation.

In this study, we set out to directly assay the primary effects of TRPS1 on chromatin accessibility, ER binding, and transcription in luminal breast cancer cells. To do so, we acutely depleted TRPS1 protein levels using an inducible degron tag inserted into the endogenous *TRPS1* locus. By performing sensitive genome-wide assays minutes to hours after TRPS1 depletion, we demonstrated that TRPS1 changes chromatin structure, which allows ER to redistribute in the genome. Along with this redistribution, we propose that TRPS1 both directly represses and indirectly activates dozens of ER target genes at baseline. Furthermore, we defined a signature of primary TRPS1-regulated genes that predicts breast cancer patient prognosis.

Results

TRPS1 is associated with breast cancer incidence and promotes breast cancer cell number accumulation

A recent genome-wide association study (GWAS) identified 32 novel single nucleotide polymorphisms (SNPs) associated with breast cancer susceptibility [28]. When we queried the NHGRI EBI GWAS Catalog to find published associations with genetic variants within the *TRPS1* genomic locus, we found one of the lead SNPs from this study [29]. Furthermore, in the authors' subtype-specific analysis of the results, the association with this variant was strongest for luminal breast cancers relative to other subtypes [28].

We used LocusZoom to plot the data within this locus (Fig 1A). A plot of these data indicates that two sets of SNPs have low linkage disequilibrium with one another, indicating that they are inherited independently and each confer risk. One set of variants is within an intronic region of the *TRPS1* gene, and one is about 400 kilobases upstream from the transcription start site (TSS) of the *TRPS1* gene. There are often many genes in close proximity to the lead SNP in a GWAS, making it difficult to predict which gene mediates the effect on the associated phenotype. However, in this case the nearest gene is almost a megabase upstream from *TRPS1*, suggesting that *TRPS1* contributes to the breast cancer susceptibility associated with one or both of these sets of genetic variants.

Based on this result, we hypothesized that perturbation of *TRPS1* in luminal breast cancer cell lines would affect cell fitness. Using data from the Cancer Dependency Map project, we found that sensitivity to *TRPS1* knockout correlated with sensitivity to knockout of *ESR1*, the gene encoding ER (Fig 1B) [31]. Furthermore, while *TRPS1* knockout led to an increase in cell number for most cancer cell lines, luminal breast cancer cell lines were significantly enriched for *TRPS1* dependency (Fig 1C).

Taken together, these data indicate that *TRPS1* influences breast cancer incidence and is required for maximal breast cancer cell fitness. Next we sought to determine how TRPS1 regulates its target genes to mediate these breast cancer patient and cellular outcomes.

Endogenously degron-tagged TRPS1 is rapidly degraded in T47D cells

To rapidly deplete TRPS1 and isolate primary TRPS1-regulated genes, we employed the dTAG system for targeted protein degradation [32, 33]. We inserted an inducible degron tag into the endogenous *TRPS1* locus in the luminal breast cancer cell line T47D. We generated three
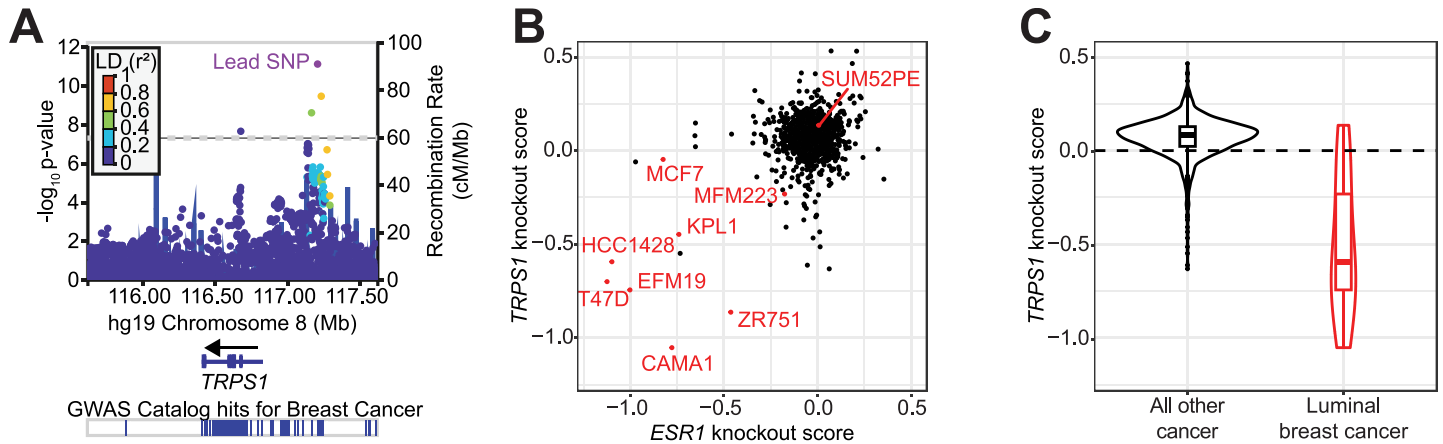
Fig 1. _TRPS1_ is associated with breast cancer incidence and promotes breast cancer cell fitness. A: LocusZoom plot of the _TRPS1_ genomic locus depicting the location and significance of SNPs associated with breast cancer susceptibility. _TRPS1_ is the closest gene to two sets of genetic variants in low linkage disequilibrium with one another. Data from [28], generated with summary statistics downloaded from the NHGRI-EBI GWAS Catalog, using LocusZoom [29, 30]. B: Scatter plot of _TRPS1_ and ESR1 knockout scores for each gene tested. Scores are normalized such that knockout of a gene with a score of 0 has no effect on cell number, and knockout of a gene with a score of -1 has an effect equal to that of knocking out one of a set of universally essential genes. Luminal breast cancer cell lines are colored in red. The data are from the Cancer Dependency Map project [31]. C) Violin and box and whisker plots of _TRPS1_ knockout scores from (B) for luminal breast cancer cell lines versus all other cell lines. Wilcoxon rank sum test p-value of $3.2*10^{-5}$.

independent clones that express the tagged TRPS1 protein that can be degraded by the addition of the small molecules dTAG-13 and dTAG^V-1 at 50nM each (dTAG) (Fig 2A). Of note, we depleted around 50% of TRPS1 from whole cell lysates in 10 minutes of treatment with dTAG, as determined by quantitative western blot, with less than 10% detected as soon as 20 minutes and as late as 48 hours after treatment (Fig 2B).

To ensure this treatment depleted TRPS1 from chromatin, we performed chromatin immunoprecipitation with sequencing (ChIP-seq) using an anti-HA antibody to recognize the 2xHA tag within the degron tag. Our ChIP-seq libraries were of high quality, as measured using the quality control metrics of fraction of reads in peaks (S1 Fig) and the peak calling-
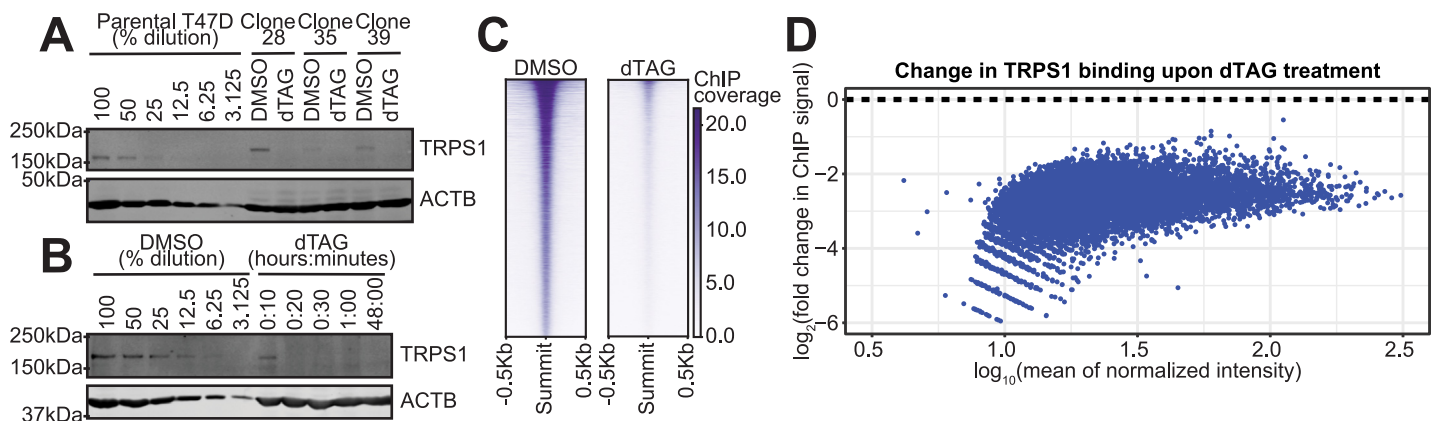
Fig 2. Endogenously degron-tagged TRPS1 is rapidly degraded in T47D cells. A: Quantitative Western blot with a serial dilution of the parental T47D cells followed by three independent dTAG-TRPS1 clones treated with DMSO or dTAG-13 and dTAG^V-1 at 50nM each (dTAG) for 2 hours. Membranes were probed with anti-TRPS1 or anti-ACTB antibodies. B: Quantitative Western blot with a serial dilution of dTAG-TRPS1 Clone 28 followed by a time course of treatment with dTAG. Membranes were probed as in (A). C: Heatmap of TRPS1 ChIP-seq peaks, in rows ranked by intensity, in cells treated with DMSO or dTAG for 30 minutes. D: MA plot of TRPS1 ChIP-seq peaks, with fold change values representing binding intensity in the dTAG condition relative to the DMSO condition. All points are colored blue to indicate they are significantly decreased at an FDR of 0.1.

independent strand cross-correlation (S2 Fig), generated using the ChIPQC R package [34]. We called peaks using MACS2 [35], using all DMSO samples together and all dTAG samples as the control. We observed a genome-wide decrease in TRPS1 binding intensity, with over 80 percent of TRPS1 depleted from chromatin after 30 minutes of dTAG treatment (Fig 2C and 2D).

TRPS1 directly represses regulatory element activity

With this system in hand, we set out to test the effects of TRPS1 depletion on regulatory element (RE) (i.e., enhancer or promoter) activity. To capture the dynamics of chromatin accessibility after TRPS1 depletion, we conducted a time course analysis using the Assay for Transposase-Accessible Chromatin with sequencing (ATAC-seq). The time points included dTAG treatments at 30 minutes, 1 hour, 2 hours, 4 hours, and 24 hours, while a DMSO treatment served as the vehicle control, assigned as the zero minute time point. Our ATAC-seq libraries were of high quality, as determined by plotting the distribution of fragment sizes (S3 Fig) and the enrichment of signal around TSS's (S4 Fig), generated using the ATACseqQC R package [36]. We generated a consensus peak set using MACS2 [35] for all samples together. At the earliest time point after degradation, our best estimate of the primary effects of TRPS1 depletion, we identify 472 peaks that increased in intensity and 36 that decreased in intensity, at a false discovery rate (FDR) of 0.1. (Fig 3A and S1 and S2 Tables). We hypothesized that the increased peaks result from loss of a direct TRPS1 reduction of chromatin accessibility, with the decreased peaks an indirect effect of the redistribution of limiting cofactors.

To test this hypothesis, we performed *de novo* motif identification in the increased and decreased peaks. We identified a GATA motif in the increased peaks but not the decreased peaks (Fig 3B). To explicitly calculate the motif prevalence in each class of peaks, we found individual motif occurrences genome-wide and intersected the peaks with these motif instances. We found a significant enrichment of a representative GATA motif in the increased peaks over the unchanged and decreased peaks (Fig 3C).

To identify additional TFs that might be contributing to changes in chromatin accessibility, we analyzed motif enrichment in both increased peaks relative to unchanged peaks as well as decreased peaks relative to unchanged peaks (S3 and S4 Tables). To our surprise, we identified nuclear receptor motifs in both classes of peaks. To follow up this analysis, we again calculated motif prevalence in each class of peaks for a representative ER half-site. For this motif, we found a significant enrichment in both increased and decreased peaks relative to unchanged peaks (Fig 3D).

We predicted that the enrichment of the GATA motif within increased peaks would wane over time. To test this prediction, we turned to our time course data. We identified ATAC-seq peaks that were significantly changed over the time course at an FDR of 0.1, using a likelihood ratio test within DESeq2 [37] to identify peaks for which including a variable for the time point increased the predictive power of the model over one without this information. We performed hierarchical clustering of these peaks and found that the replicates for each time point clustered together and that the majority of dynamic peaks changed gradually over the time course, with additional clusters displaying different kinetics (Fig 3E). We called increased, unchanged, and decreased peaks as in Fig 3A for each time point relative to the control condition. We then calculated GATA motif prevalence in each set of peaks as in Fig 3C. As we expected, the GATA motif prevalence in the increased peaks decreased over time, but the majority of increased peaks at 24 hours still contained a GATA motif (Fig 3F). This is consistent with the primary effects of TRPS1 depletion driving a large proportion of the changes in chromatin accessibility even as late as 24 hours after TRPS1 depletion.
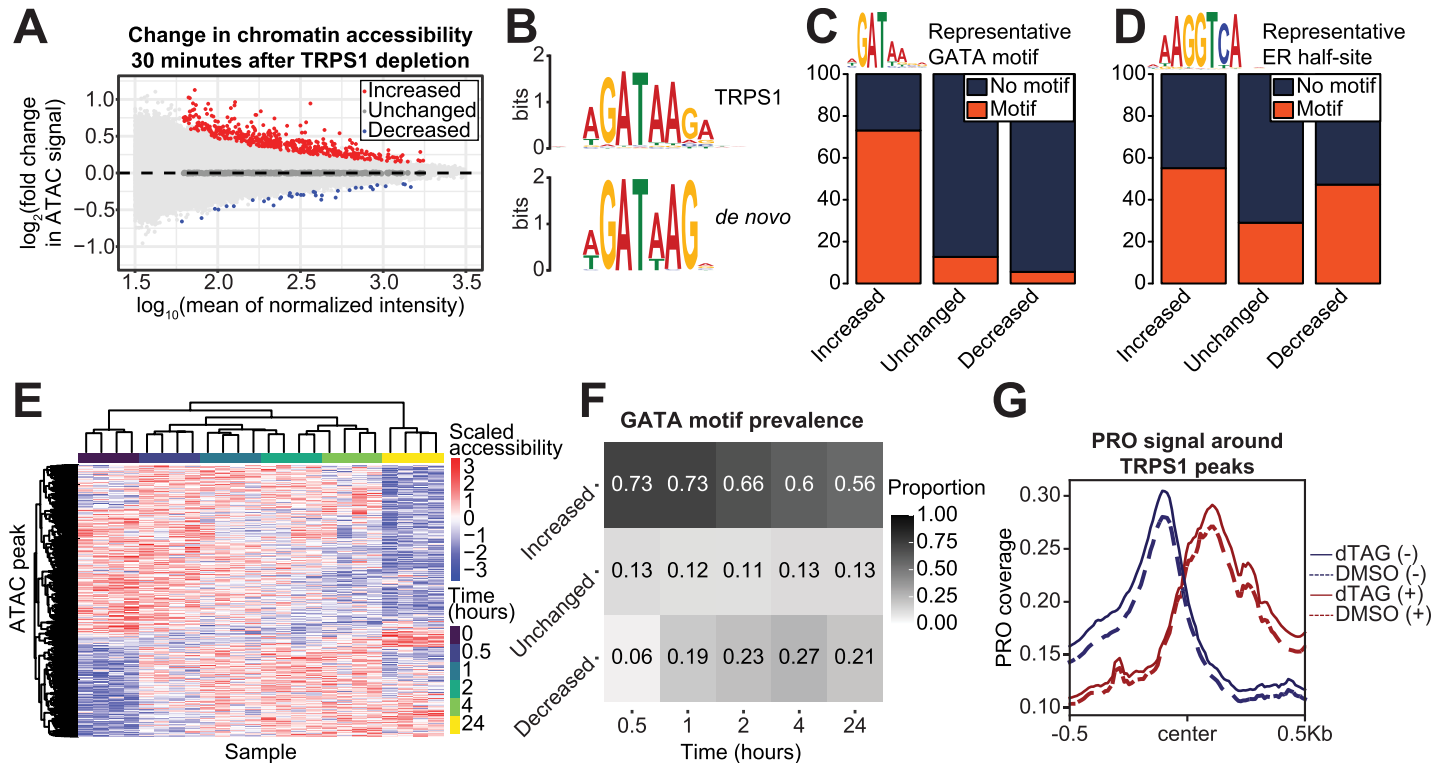
Fig 3. TRPS1 directly represses regulatory element activity. A: MA plot of ATAC-seq peaks, with fold change values representing accessibility in the 30 minute dTAG-13 and dTAG^V-1 at 50nM each (dTAG) treatment condition relative to the DMSO condition. B: *De novo* motif identified in increased peaks from (A) (below), matched to the TRPS1 motif (above). C: Bar charts of prevalence of a representative GATA motif in increased, unchanged, and decreased peaks from (A). Chi-square test p-value $< 2.2*10^{-16}$. D: Bar charts of prevalence of a representative ER half-site in increased, unchanged, and decreased peaks from (A). Chi-square test p-value $4.5*10^{-15}$. E: Heat map with hierarchical clustering of chromatin accessibility in ATAC-seq peaks that are significantly changed over the time course at an FDR of 0.1. F: Heat map of prevalence of the representative GATA motif in increased, unchanged, and decreased peaks, as in (C), for each time point relative to the DMSO condition. G: Density plot for composite PRO-seq signal across all TRPS1 ChIP-seq peaks, separated by strand and treatment condition.

https://doi.org/10.1371/journal.pgen.1011159.g003

As an orthogonal readout of RE activity, we measured bidirectional transcription around TRPS1 binding sites using precision genomic run-on with sequencing (PRO-seq) [38]. Our libraries were of high quality using several quality control metrics (S5 Fig) [39]. Using a window centered on each summit of TRPS1 ChIP-seq intensity, we observed an increase in bidirectional transcription 30 minutes after TRPS1 depletion (Fig 3G).

Since composite profiles have limitations, we chose to look more closely at each TRPS1 peak and determine if increasing bidirectional transcription is a reproducible trend. We counted PRO reads in the window around the TRPS1 peak summits shown in Fig 3G and used DESeq2 to identify differentially transcribed regions. We normalized our counts based on the size factors generated from the DESeq2 analysis of the reads in genes from Fig 4. As read counts in small putative regulatory elements tend to be fewer than in large genes, it is not surprising that only 9 of these regions have individually statistically significantly higher bidirectional transcription and none with significantly lower transcription, at a false discovery rate of 0.1 (S6 Fig). However, we did find a TRPS1 cistrome-wide increase in PRO signal, with over 62% of these regions increasing in bidirectional transcription upon TRPS1 depletion. We performed ANOVA on a linear model predicting the logarithm of the normalized PRO reads for each region based on the DMSO or 30 minute dTAG treatment condition across the four replicates, and the p-value for the F-test was $< 2.2*10^{-16}$.
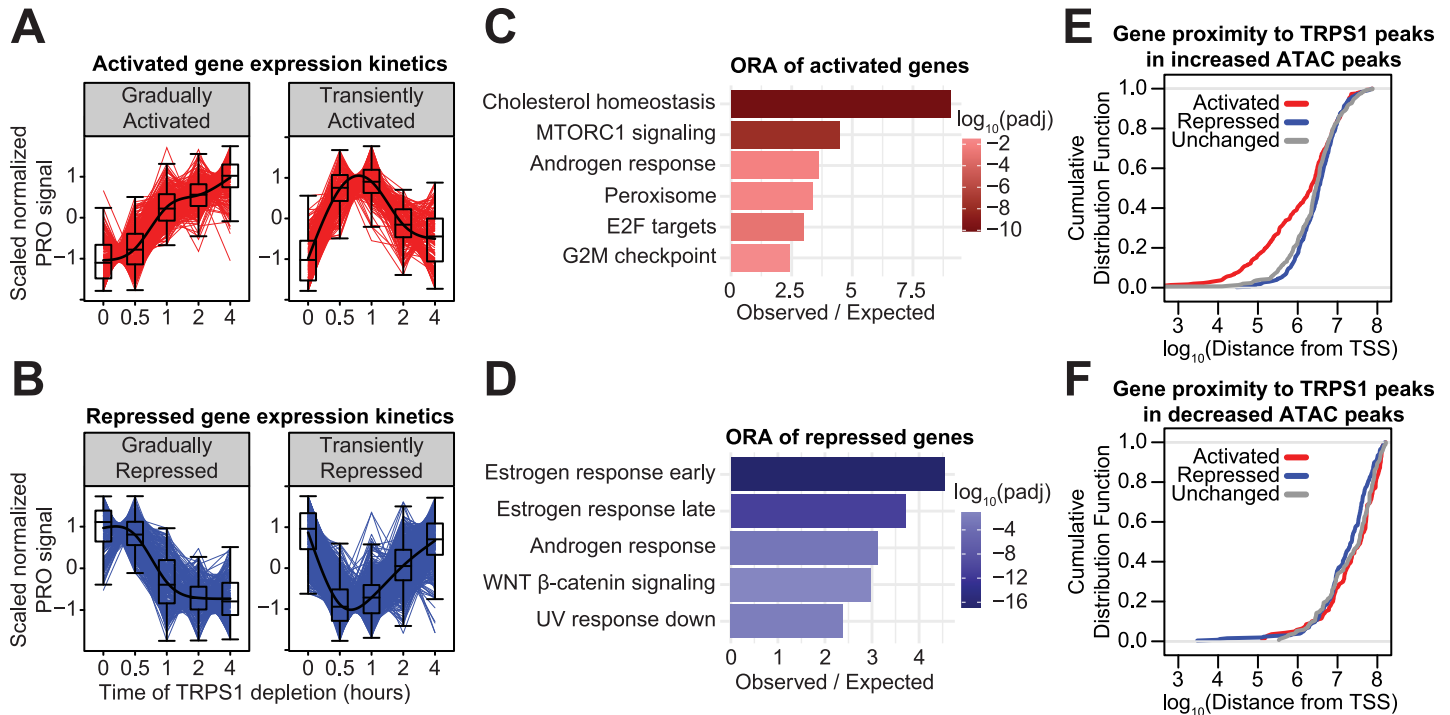
Fig 4. TRPS1 directly represses transcription of target genes. A: Kinetic traces of the two major clusters of activated genes over the time course. B: Kinetic traces of the two major clusters of repressed genes over the time course. C: Over-representation analysis of the activated genes from (A). D: Over-representation analysis of the repressed genes from (B). E: Cumulative distribution function plot of the distance from the TSS of each gene to the nearest TRPS1 ChIP-seq peak overlapping an increased ATAC-seq peak, by gene class. Kolmogorov–Smirnov test between activated and unchanged genes: p-value = 0.011. F: Cumulative distribution function plot of the distance from the TSS of each gene to the nearest TRPS1 ChIP-seq peak overlapping an decreased ATAC-seq peak, by gene class. KS test between repressed and unchanged genes: p-value > 0.1.

https://doi.org/10.1371/journal.pgen.1011159.g004

The biological interpretation is consistent with our other results that indicate a role of TRPS1 in chromatin compaction. We would not necessarily expect an increase in bidirectional transcription at each TRPS1 peak upon TRPS1 degradation, but the increase in chromatin accessibility could facilitate downstream factors that promote transcription initiation. As chromatin accessibility and bidirectional transcription can affect one another, we next isolated ATAC-seq peaks without bidirectional transcription, as identified using dREG [40]. As we saw in Fig 3A, chromatin accessibility predominantly increased upon TRPS1 depletion (S7 Fig). Along with our accessibility data and motif analysis, these data indicate that TRPS1 directly represses RE activity primarily via its effect on chromatin accessibility.

TRPS1 directly represses transcription of target genes

Downstream from the changes in RE activity, we measured changes in nascent transcription within genes with PRO-seq over the same time course of TRPS1 depletion as in the ATAC-seq experiment. As with our ATAC-seq time course analysis, we identified genes that were significantly changed over the time course at an FDR of 0.1, using a likelihood ratio test within DESeq2 [37] to identify genes for which including a variable for the time point increased the predictive power of the model over one without this information. We identified 1,425 dynamic genes over the time course and performed hierarchical clustering to classify the genes based on their expression kinetics (Fig 4A and 4B and S5, S6, S7 and S8 Tables). Over-representation analysis (ORA) of the activated genes identified several enriched Hallmark gene sets [41], most

prominently cholesterol homeostasis genes (Fig 4C). ORA on the repressed genes revealed that the two estrogen response gene sets were the most significantly enriched of the Hallmark gene sets.

We hypothesized that TRPS1 regulates these gene sets by distinct mechanisms. Specifically, we predicted that TRPS1 directly represses the dTAG-activated genes by repressing the activity of REs proximal to these genes. To test this prediction, we measured the distance from the transcription start site (TSS) of each gene to the nearest TRPS1 ChIP-seq peak overlapping an increased ATAC-seq peak. By constructing a cumulative distribution function, we found that the activated genes are significantly closer to these activated REs (Fig 4E).

In contrast to the genes activated by TRPS1 depletion, we predicted that the effects on the genes repressed by TRPS1 depletion are indirect and distal to TRPS1 binding. To test whether TRPS1 directly activates a subset of REs to activate transcription of proximal genes, we measured the distance from the TSS of each gene to the nearest TRPS1 ChIP-seq peak overlapping a decreased ATAC-seq peak. There are few examples of this class of RE, so the distances were much farther, and there was no significant enrichment of repressed genes proximal to these peaks (Fig 4F). These data suggest that, while TRPS1 positively and negatively regulates hundreds of primary response genes, TRPS1 only represses transcription of its direct target genes.

TRPS1 redistributes ER binding to modulate ER target gene transcription

Based on the correlation between cancer cell line sensitivity to *TRPS1* knockout and *ESR1* knockout (Fig 1B), the enrichment of an ER binding motif in both increased and decreased ATAC-seq peaks (Fig 3D), and the over-representation of estrogen response gene sets in the genes repressed by TRPS1 depletion (Fig 4D), we focused on ER target genes to explore a possible mechanism by which TRPS1 indirectly activates transcription. We first defined direct ER target genes using our previously-generated PRO-seq data from parental T47D cells that were hormone starved for three days and then acutely stimulated with estrogen or a DMSO vehicle control for 90 minutes (S8 Fig) [39]. We exclusively focused on estrogen-activated genes because ER directly activates these genes [8, 42, 43]. 65 ER target genes were activated, and 58 were repressed by acute TRPS1 depletion (Fig 5A). To test the robustness of this change in ER target gene transcription, we additionally performed PRO-seq in each of the three independent TRPS1-dTAG clones generated from the parental T47D cells. Indeed, the genes we identified as TRPS1-regulated ER target genes in the one clone used for the time course experiment tend to be regulated in the same direction upon TRPS1 degradation across these three clones, suggesting that this effect is robust across the cell lines in which we can acutely deplete TRPS1 (S9 Fig).

We hypothesized that these changes in ER target gene transcription are mediated by changes in the genomic distribution of ER binding. We performed ER ChIP-seq to test the prediction that ER binding intensity proximal to dynamic ER target genes would change in concordance with the change in gene transcription. As before, our ChIP-seq libraries were of high quality (S1 and S2 Figs). We called peaks using MACS2 [35], using all ER samples together and all IgG samples as the control. Consistent with our hypothesis, we found that ER ChIP-seq peaks within a 100 kilobase (kb) window around the TSS of activated genes tended to increase in intensity, and ER ChIP-seq peaks within a 100kb window around the TSS of repressed genes tended to decrease in intensity (Fig 5B).

We further hypothesized that only the increased ER binding sites represent a direct effect of TRPS1 activity. Consistent with this hypothesis, we found that ER binding proximal to TRPS1 tends to increase in intensity, and ER binding distal to TRPS1 tends to decrease in intensity (Fig 5C). Together, these data suggest a model in which TRPS1 depletion redistributes ER
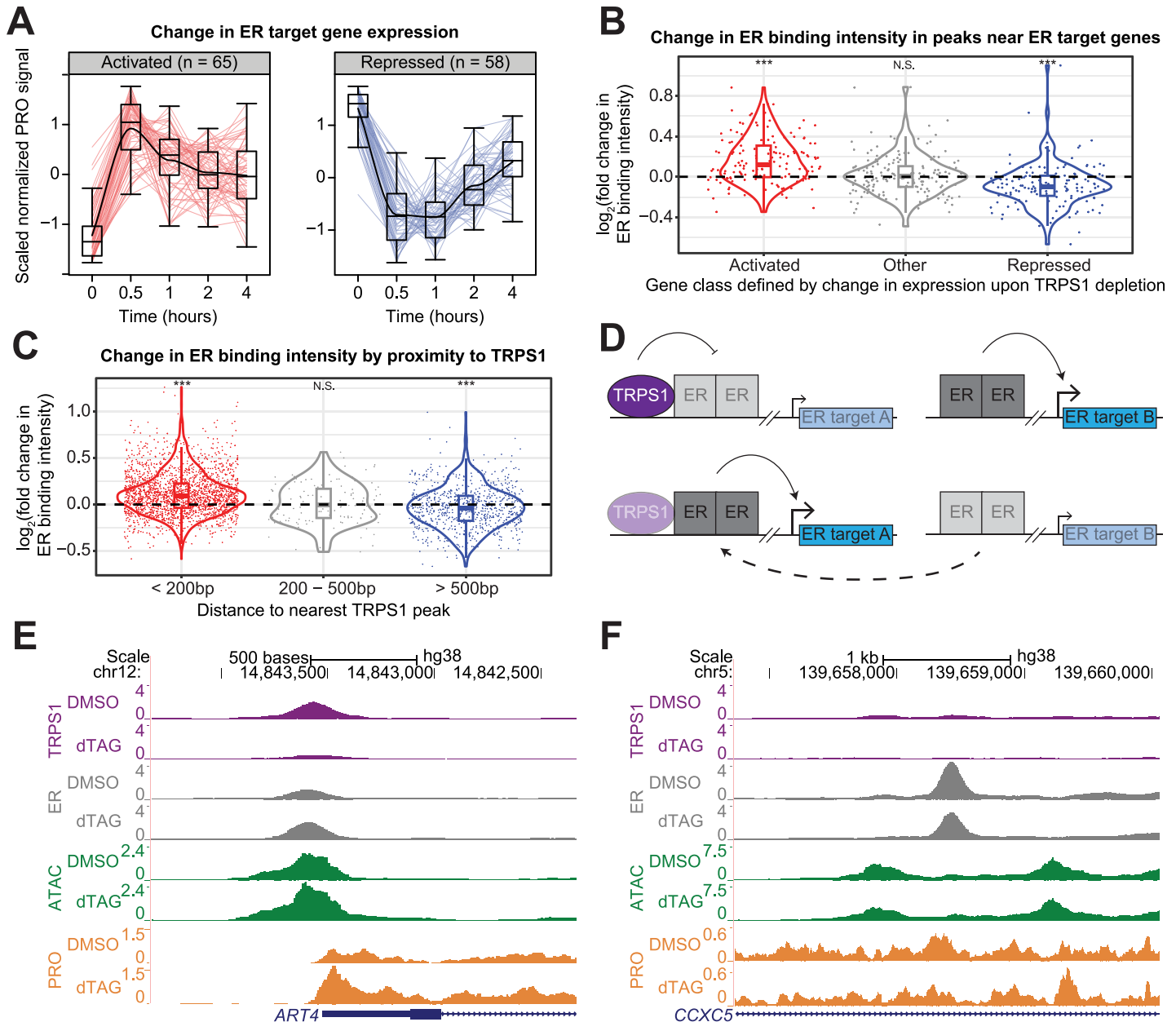
Fig 5. TRPS1 redistributes ER binding to modulate ER target gene transcription. A: Kinetic traces of activated and repressed ER target genes over the time course. B: Violin and box and whisker plots for ER binding intensity fold change upon TRPS1 depletion at ER ChIP-seq peaks within 100kb of the TSS of each gene, grouped by gene class defined by change in expression upon TRPS1 depletion. (In (B) and (C), *** represents a significant one-sample t-test p-value $< 10^{-3}$. N.S. represents a non-significant p-value > 0.1.) C: Violin and box and whisker plots for ER binding intensity fold change upon TRPS1 depletion at ER ChIP-seq peaks, grouped by summit-to-summit distance to the nearest TRPS1 ChIP-seq peak. D: Model of TRPS1-mediated ER redistribution and modulation of ER target gene transcription. Transparent boxes indicate reduced binding intensity or attenuated transcription. Above, at baseline, TRPS1 directly decreases ER binding intensity proximal to TRPS1, attenuating ER activation of proximal ER target genes. Distal to TRPS1, ER binding intensity is not directly affected by TRPS1, and ER fully activates proximal ER target genes. Below, after TRPS1 depletion, ER binding proximal to TRPS1 increases in intensity, augmenting ER activation of proximal ER target genes. Distal to TRPS1, ER binding intensity is indirectly decreased, as limiting ER molecules are redistributed to TRPS1-proximal regulatory elements, attenuating ER activation of proximal ER target genes. E: ChIP, ATAC, and PRO density around an example increased ER binding site near an activated ER target gene. At this TRPS1-proximal ER binding site, upon dTAG treatment, TRPS1 binding intensity decreases, ER binding intensity increases, chromatin accessibility increases, and gene expression increases. F: ChIP, ATAC, and PRO density around an example decreased ER binding site near an repressed ER target gene. At this TRPS1-distal ER binding site, upon dTAG treatment, ER binding intensity decreases, chromatin accessibility decreases, and gene expression decreases. In (F) and (G), dTAG refers to dTAG-13 and dTAG^V-1 at 50nM each.

https://doi.org/10.1371/journal.pgen.1011159.g005

binding from TRPS1-distal sites to TRPS1-proximal sites and modulates ER target gene transcription proximal to the dynamic ER binding sites (Fig 5D). To illustrate this phenomenon, we provide an example of a TRPS1-proximal, increased ER peak near an activated ER target gene in (Fig 5E), and a TRPS1-distal, decreased ER peak near a repressed ER target gene in (Fig 5F).

TRPS1 activity is associated with breast cancer patient outcomes

We next sought to connect the primary TRPS1-responsive genes with downstream cellular and patient-related outcomes. We defined a new steady state of transcription with the 24 hour time point after TRPS1 depletion. We ranked genes based on their shrunken fold change in PRO signal (Fig 6A). Using this ranking, we performed gene set enrichment analysis with the Hallmark gene sets and found multiple cell-cycle-related gene sets to be negatively enriched, including E2F Targets (Fig 6B) [44, 45]. Consistent with this, we observed a significant decrease in cell number doubling rate of T47D dTAG-TRPS1 cells upon TRPS1 depletion (Fig 6C). Importantly, the isogenic parental T47D cells do not display a cell number defect with dTAG treatment, so we attribute this effect to TRPS1 depletion and not a non-specific effect of the compounds (S10 Fig).

Finally, we calculated a TRPS1 activity score by adapting methods developed by [48, 49]. We used our PRO-seq data to determine a primary TRPS1 regulon based on the differentially expressed genes 30 minutes after TRPS1 depletion. We classified breast cancer patients from the METABRIC cohort as having high TRPS1 activity if both a) TRPS1-repressed genes are negatively enriched and b) TRPS1-activated genes are positively enriched, relative to all other patients in the cohort (example patient in (Fig 6D)) [46, 47]. Similarly, we classified patients as having low TRPS1 activity if both a) TRPS1-repressed genes are positively enriched and b) TRPS1-activated genes are negatively enriched. We classified the remaining patients as having intermediate TRPS1 activity. We ranked patients based on their TRPS1 activity and found no association with other clinical covariates (Fig 6E). When we stratified patients by TRPS1 activity, we found TRPS1 activity to be significantly associated with shorter survival time (Logrank p-value $4.99*10^{-4}$) (Fig 6F). When we first separated tumors by ER-positivity or by intrinsic subtype, we found that this association was specific for ER-positive tumors and Luminal A tumors (S11 Fig). We also performed this analysis using genes differentially expressed after 24 hours of TRPS1 depletion (S12 Fig). However, by this time cell cycle genes dominate, and we speculate that the association with breast cancer patient survival is due to these genes. We believe using the primary response genes offers a unique measure of TRPS1 activity not achievable using previous surrogates.

Discussion

In this study, we used rapidly inducible targeted protein degradation to systematically determine the primary effects of acute TRPS1 depletion on chromatin accessibility, ER binding, and nascent transcription in a luminal breast cancer cell line. We focused on TRPS1 based on two orthogonal, genome-wide, unbiased assays that implicated *TRPS1* in the processes of breast tumor incidence and breast cancer cell number accumulation.

First, we used the summary statistics from a recent GWAS to plot two sets of common genetic variants in the *TRPS1* locus associated with breast cancer incidence [28]. These genetic variants were independently identified as significantly associated with breast cancer incidence in a previous GWAS [50], but Zhang *et al.* determined that the association was strongest among luminal breast tumors. Second, we analyzed data from the Cancer Dependency Map project and found that sensitivity to *TRPS1* knockout was correlated with sensitivity to *ESR1*
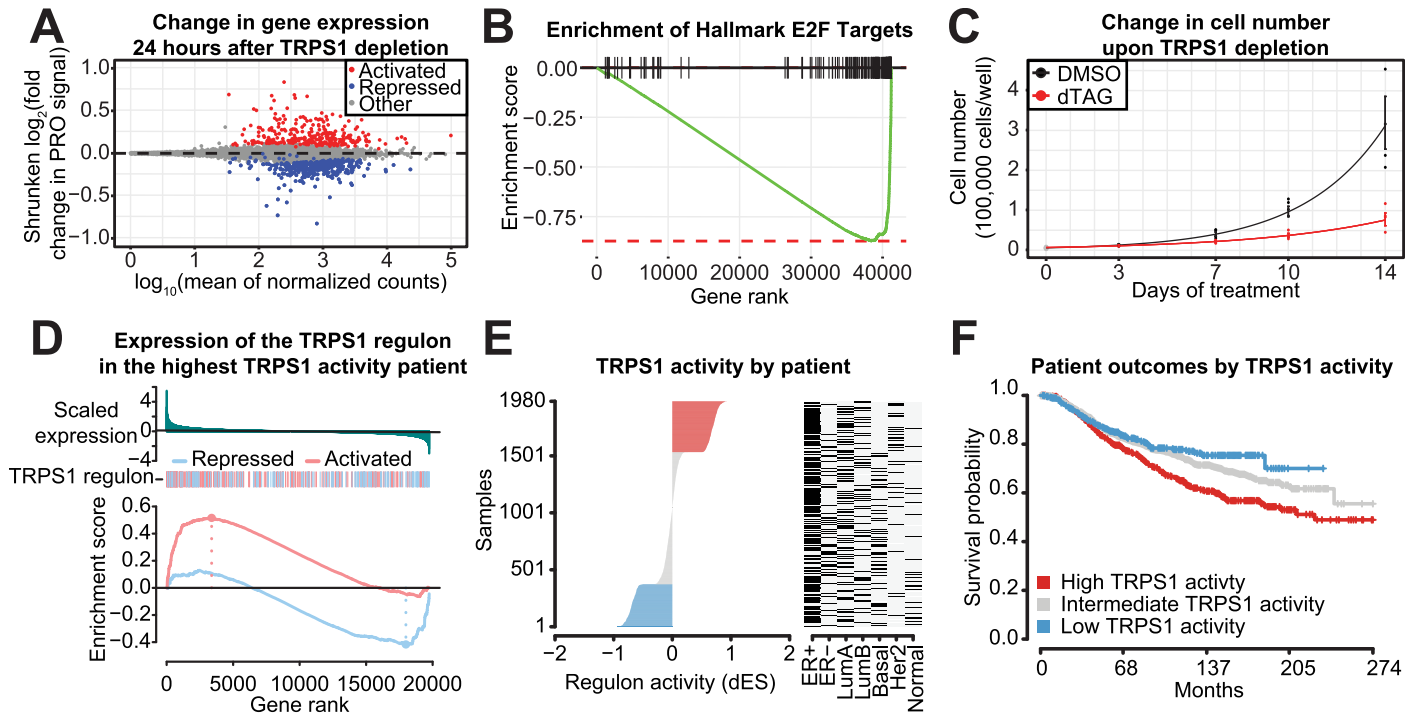
Fig 6. TRPS1 activity is associated with breast cancer patient outcomes. A: MA plot of PRO signal in each gene, with shrunken log fold change values representing transcription in the 30 minute dTAG-13 and dTAG^V-1 at 50nM each (dTAG) treatment condition relative to the DMSO condition. B: Mountain plot of the Hallmark E2F Targets gene set, using genes ranked by shrunken fold change from (A). A negative enrichment score indicates an enrichment of the gene set among repressed genes. Adjusted p-value $2.5*10^{-19}$. C: Cell number over time of dTAG-TRPS1 cells treated with dTAG or DMSO. Analysis of variance for the coefficient corresponding to the difference in doubling rates between the conditions in a linear model of the logarithm of cell number versus time: p-value $1.1*10^{-5}$. D: Differential enrichment score (dES) calculation for an example patient with the highest TRPS1 activity. Above, genes ranked by scaled expression in this patient relative to all other patients in the METABRIC cohort [46, 47]. Below, gene set enrichment analysis of TRPS1-repressed and TRPS1-activated genes defined by response after 30 minutes of TRPS1 depletion. E: Patients from the METABRIC cohort, ranked by dES as calculated in (D), with classifications of the tumors on the right. F: Kaplan-Meier curves for patients in the METABRIC cohort, stratified by TRPS1 activity as in (E). Logrank p-value $4.99*10^{-4}$.

https://doi.org/10.1371/journal.pgen.1011159.g006

knockout and significantly enriched among luminal breast cancer cell lines [31]. Both of these unbiased screens indicate that *TRPS1* contributes to luminal breast cancer cell fitness and led us to the hypothesis that TRPS1 influences ER activity or genomic binding.

As TFs regulate the transcription of many other chromatin-associated factors that themselves regulate RE activity, TF binding, and transcription, we sought to isolate the primary effects of TRPS1 depletion. To do so, we used the dTAG inducible degron tag system to acutely deplete endogenous TRPS1 protein abundance within minutes of induction [32]. This is in contrast to traditional RNA interference or gene knockout methods, which can take days to deplete the target of interest.

Minutes to hours after TRPS1 depletion, we performed several sensitive, genome-wide assays. ATAC-seq and ChIP-seq can be performed at any time point after a perturbation, as they measure chromatin accessibility and chromatin-associated factor binding, which can change with rapid kinetics. In contrast, changes in messenger RNA abundance accumulate more slowly, with kinetics that depend not only on the rate of nascent transcription but also on the ratio of abundance to synthesis and degradation rates. In contrast, nascent transcriptional profiling measures the immediate change in RNA synthesis rates after a perturbation. Here we use PRO-seq coupled with acute TRPS1 depletion to identify primary TRPS1-responsive genes.

With our cell lines and assays in hand, we first measured changes in chromatin accessibility upon TRPS1 depletion. Consistent with previous studies linking TRPS1 to corepressor complexes, we found that the predominant effect of TRPS1 depletion is an increase in chromatin accessibility and bidirectional transcription at REs [14–18]. We observed decreasing enrichment of the GATA motif prevalence in increased peaks over time, indicating that our shortest time point was the most specific for isolating the primary effects of TRPS1 depletion. Intriguingly, we identified a significant enrichment of ER half-site motifs in increased as well as decreased ATAC peaks, suggesting that ER binding intensity was changing in a site-specific manner.

We next measured changes in nascent transcription minutes to hours after TRPS1 depletion and clustered the gene responses. Activated genes were enriched for cholesterol homeostasis genes. Of note, several recent GWAS have identified SNPs in the *TRPS1* locus associated with blood cholesterol levels [51–53]. However, as of yet, no mechanistic follow-up studies into the role of TRPS1 in cholesterol biology have been performed. After TRPS1 depletion, repressed genes were enriched for estrogen response gene sets. Consistent with our previous data, activated genes were closer to increased TRPS1-bound ATAC peaks, suggesting TRPS1 directly represses these target genes at steady state. On the other hand, repressed genes were not closer to decreased TRPS1-bound ATAC peaks, suggesting a distinct mechanism of transcriptional regulation of this gene class.

We hypothesized that at steady state TRPS1 directly represses and indirectly activates its primary response genes. Using ER target genes and ER genomic binding as a case study, we found evidence supporting a model of acute ER redistribution. ER binding sites proximal to TRPS1 tended to increase in intensity upon TRPS1 knockdown, with distal ER binding sites tending to decrease in intensity. Furthermore, genes activated upon TRPS1 depletion were surrounded by ER binding sites that increased in intensity, and repressed genes were near decreased ER binding sites. Taken together, we propose a model in which TRPS1 directly decreases chromatin accessibility at steady state. Upon acute TRPS1 depletion, TRPS1-proximal REs increase in accessibility, an effect which we propose allows ER to redistribute from TRPS1-distal REs. In this proposed model, subsets of ER target genes are activated or repressed by TRPS1 depletion via distinct mechanisms.

First described in the 1980s, the concept of coactivator "squelching" has been debated as a mechanism of indirect activity distal from a TF's genomic binding sites [54, 55]. Squelching has been proposed as a mechanism by which nuclear receptors like ER acutely repress transcription of a subset of primary response genes by competing for limiting coactivators [43, 56–58]. Here we propose not the redistribution of coactivators by an activating TF, but a redistribution of activating TFs themselves via a rapid increase in local chromatin accessibility after the acute depletion of a repressive TF.

Our findings of both increased and decreased ER genomic binding and target gene transcription are distinct from previous studies of the effects of TRPS1 on TF binding and activity. Elster *et al.* used an unbiased screen to identify TRPS1 as a repressor of YAP1 activity in another luminal breast cancer cell line, MCF7 [17]. After TRPS1 knockdown, the authors observed a genome-wide activation of YAP target genes. We did not find a YAP gene signature among dynamic genes in our PRO-seq data, though we did observe an enrichment of TEAD motifs in increased ATAC peaks, suggesting differences between the cells used in each study and perhaps their baseline YAP-TEAD activity. While we would predict that our acute redistribution model is generalizable to other TFs and sets of TRPS1-regulated genes beyond ER and its target genes, it remains possible that TRPS1 modulates the genomic binding intensity of additional TFs in a unidirectional manner via a distinct mechanism.

Serandour *et al.* knocked down *TRPS1* in MCF7 cells and reported both a genome-wide repression of ER target genes as well as a genome-wide increase in ER binding [16]. We did not perform ChIP-seq at a comparable time point to their days-long knockdown, so we cannot directly compare our ER binding data. Our latest PRO-seq time point was 24 hours after TRPS1 depletion, at which time we do not observe a genome-wide repression of ER target genes. This could once again be attributable to a difference in cell lines. However, we would also speculate that the unidirectional and nonconcordant changes in ER binding and target gene expression at later time points could be due to non-primary effects of extended *TRPS1* knockdown.

Finally, we used PRO-seq data from both late and early time points to identify genes that represent cells at a new steady state after TRPS1 depletion, as well as primary TRPS1-responsive genes. After 24 hours of TRPS1 depletion, repressed genes were enriched for cell cycle related genes, consistent with a decrease in cell number doubling rate. Unique to this study, we used primary TRPS1-responsive genes to define a TRPS1 activity score, adapting a method based on predicted TF target genes [48, 49]. Using this method, we were able to stratify breast cancer patients into groups with differing survival probabilities.

Using TRPS1 activity score to classify patients may provide additional insight into the transcriptional program within a patient's tumor that might not be immediately apparent based on previous surrogates for TRPS1 activity. For example, *TRPS1* is frequently amplified in breast tumors, and this amplification is associated with worse prognosis [16, 59]. However, *TRPS1* is often co-amplified along with the rest of the chromosomal segment 8q23–q24, where the proto-oncogene *MYC* resides, making it difficult to discern whether *TRPS1* amplification is a driver of breast cancer progression [60].

In contrast, higher *TRPS1* expression has been associated with better breast cancer patient outcomes, though its expression is highly correlated with ER and GATA3, both favorable prognostic indicators [22, 61]. As relative TF expression and activity across patients are not always identical, our data uses primary TRPS1-responsive genes as a measure of TRPS1 activity. Our TRPS1 activity score is not correlated with ER-positivity and effectively stratifies patients. Though our patient outcome analysis, as with all similar analyses, describes an association and does not necessarily imply a causative relationship, the direction is consistent with the effect on cell number observed in this study as well as the Cancer Dependency Map, suggesting that TRPS1 drives breast cancer cell number accumulation.

Altogether, we provide a systematic study of the primary effects of rapid TRPS1 depletion in luminal breast cancer cells. We propose a model in which TRPS1 depletion leads to decondensation of local chromatin structure, allowing for the acute redistribution of ER, both activating and repressing subsets of ER target genes. This TRPS1-regulated transcription appears to be relevant for cancer cell fitness, as TRPS1 depletion decreases cell number doubling rate, and high TRPS1 activity is associated with worse breast cancer patient outcomes. These methods of inducible targeted protein degradation coupled with genomic chromatin assays and nascent RNA transcriptional profiling should in principle be applicable to the study of any TF, allowing us to better understand the mechanisms behind the phenotypes associated with additional GWAS hits.

Materials and methods

GWAS and DepMap data visualization

Summary statistics from [28] were downloaded from the NHGRI-EBI GWAS Catalog [29]. SNPs in the *TRPS1* locus were plotted using LocusZoom [30]. Knockout scores and luminal

breast cancer identifiers were downloaded from the Cancer Dependency Map project [31] and plotted using the statistical programming language R [62].

Cell culture

T47D cells (RRID:CVCL_0553) (ATCC) were cultured in RPMI 1640 medium (Gibco) supplemented with 10% fetal bovine serum (Gemini) and 10μg/ml insulin from bovine pancreas (Sigma, made as a 1000x solution in 1% aqueous glacial acetic acid).

Plasmid generation for gene editing

DNA for transfection was prepared as previously described [63, 64]. A CRISPR sgRNA (TTATCTTTGCAGATATGGTC) targeting the 5' end of the *TRPS1* coding sequences was designed using Benchling. The sgRNA was cloned into hSpCas9 plasmid PX458 (Addgene #48138) as previously described [65], using the following primers:

5'-CACCGTTATCTTTGCAGATATGGTC-3' and
5'-AAACGACCATATCTGCAAAGATAAC-3'.

A plasmid harboring a synthetic HygR-P2A-2xHA-FKBP_F36V insert was generated with Cold Fusion (System Biosciences), starting with the HygR-P2A-AID cassette in pMGS58 (Addgene #135311) [64] and the Puro-P2A-2xHA-FKBP_F36V casette in (Addgene #91793) [32]. The linear donor was generated by PCR using primers (IDT) that contain 50-nucleotide homology tails and gel-purified. The primers contained 5' phosphorothioate modifications to increase PCR product stability in the cell [66]. The primers used for making PCR donor fragments were:

5'-G*T*AACTTTCAGATAACACTGTATCTGCCTTTT
CCCTTTATCTTTGCAGATATGAAAAAGCCTGAACT
CACCG-3'
and
5'-T*T*CACTTGCAACGTTTCTCAGAGGGGGGTTCT
TTTTCCGGACACCTGAACCTGAACCTCCAGATCCA
CCAGATCTTTCCAGTTTTAGAAGCTCCACATCG-3'
with asterisks representing the phosphorothioate modifications.

dTAG-TRPS1 clone generation

Clones were generated as previously described [63, 64], with modifications. An initial round of cloning was performed using puromycin selection, but upon genomic DNA sequencing this clone did not appear to have a dTAG insertion event within *TRPS1*. Nevertheless, this clone was used for a second round of cloning using hygromycin selection. $3*10^6$ cells were plated in 10cm plates. The next day, cells were cotransfected with 15μg of CRISPR/Cas9-sgRNA plasmid and 1.85μg of linear donor PCR product using Lipofectamine 3000 (Thermo Fisher Scientific) in Optimem (Gibco). One day after transfection, the media was replaced. Starting four days after transfection, cells were selected for two weeks with 200μg/mL of Hygromycin B (Invitrogen) with 20% conditioned media, replaced twice per week. Colonies were then grown in 20% conditioned media, replaced twice per week, until they were large enough to be picked and passaged to a 24-well plate. Clones were expanded and frozen at 8 passages after transfection. Integration was tested with Western blotting, PCR, and Sanger sequencing. In each of the three clones, two to three of the four genomic copies of *TRPS1* are knocked out, and only

tagged TRPS1 protein is expressed. Details about each of the determined alleles are available at
https://guertinlab.github.io/TRPS1_ER_analysis/Vignette.html#allele-sequencing.

Western blotting

$8*10^5$ cells per sample were plated in each well of a 6-well plate. Cells were treated with DMSO
or 50nM dTAG-13 and 50nM dTAG^V-1 in DMSO at various time points and collected simul-
taneously. At the time of harvest, cells were scraped and lysed in RIPA buffer (1% Nonidet P-
40, 1% sodium deoxycholate, 0.1% sodium dodecyl sulfate, 2mM EDTA, 150mM NaCl, 10mM
sodium phosphate, 50mM NaF, 50mM Tris pH 7.5), with 100µM benzamidine, 5µg/mL apro-
tinin, 5µg/mL leupeptin, 1µg/mL pepstatin, 1mM phenylmethylsulfonyl fluoride, and 2mM
sodium orthovanadate added fresh. Lysates were sonicated in a Biorupter UCD-200 (Diage-
node) on high for 30 seconds on and 30 seconds off for 5 cycles, and clarified by centrifugation
at 14,000rpm for 15 min in 4˚C. Protein concentration was measured by BCA assay and
diluted to the same concentration. 10× Laemmli buffer was added to a final concentration of
1x, and 2-mercaptoethanol was added to a final concentration of 1%. Samples were boiled at
95˚C for 10 minutes, and 30µg of each was loaded into a 10% polyacrylamide gel. Samples
were separated by gel electrophoresis and transferred to nitrocellulose membranes. Mem-
branes were incubated in blocking buffer (3% bovine serum albumin, 1X Tris buffered saline)
for 1 hour at room temperature with rocking. Primary antibodies (anti-TRPS1, Cell Signaling
#17936S, and anti-ACTB, Cell Signaling #3700S) were diluted 1:1,000 in primary buffer (3%
bovine serum albumin, 0.1% sodium azide, 0.1% Tween-20, 1X Tris buffered saline) at 4˚C
with rocking overnight. Fluorescent secondary antibodies were diluted 1:10,000 in secondary
buffer (5% bovine serum albumin, 0.1% sodium azide, 0.1% Tween-20, 1X Tris buffered saline)
and incubated for 1 hour at room temperature with rocking, and fluorescence was measured
(Odyssey, Licor).

ChIP-seq library preparation

$2.4*10^7$ cells per sample were plated across 3 15cm dishes 2 days before harvest. Cells were
treated with DMSO or 50nM dTAG-13 and 50nM dTAG^V-1 in DMSO for 30 minutes and col-
lected simultaneously. At the time of harvest, cells were fixed with 1% formaldehyde (Sigma)
for 10 minutes at 37˚C and quenched with 125mM Glycine (Fisher) for 10 minutes at 37˚C.
Plates were moved to ice, and cells were washed and scraped into ice cold PBS containing
Complete EDTA-free Protease Inhibitor Cocktail (Roche). Cells were pelleted in aliquots of
$3.6*10^7$ cells, snap frozen in liquid nitrogen, and stored at -80˚C. Pellets were thawed, and cells
were lysed in 1mL Cell Lysis Buffer (85mM KCl, 0.5%NP40, 5mM PIPES pH 8.0), with prote-
ase inhibitor cocktail added fresh, for 10 minutes with rotation at 4˚C. Nuclei were pelleted at
3300g at 4˚C for 5 minutes and resuspended in 500µL ChIP lysis buffer (0.5% SDS, 10mM
EDTA, 50mM Tris-HCl pH 8.1), with protease inhibitor cocktail added fresh, for 10 minutes
with rotation at 4˚C. Lysates were moved to 15ml polystyrene conical tubes (Falcon) and soni-
cated in a Biorupter UCD-200 (Diagenode) on high for 30 seconds on and 30 seconds off for 4
sets of 5 cycles. Before each set, ice in the water bath was replaced, and samples were gently
vortexed to mix. Sonicated lysates were then move to 1.5ml tubes and clarified by centrifuga-
tion at 14,000rpm for 15 min in 4˚C. 500µL of the supernatant was diluted into 6.5mL Dilution
Buffer (0.01% SDS, 1.1% Triton X-100, 1.2mM EDTA, 167mM NaCl, 16.6mM Tris-HCl pH
8.0), with protease inhibitor cocktail added fresh ($1*10^6$ cells in 200µL). 1ml ($5*10^6$ cells) was
aliquoted into each of 3 tubes with antibody (1.25 µg anti-HA, Cell Signaling #3724S, 2.5µg
anti-ER, Millipore #06–935, or 2.5µg IgG control, Cell Signaling #2729S), and incubated with
end-over-end rotation at 4˚C overnight.

50μL Protein A/G Magnetic Beads (Pierce) per sample were washed with bead washing buffer (PBS with 0.1% BSA and 2mM EDTA) and then incubated with samples for 2 hours with rotation at 4˚C. The samples were washed once each with low salt immune complex buffer (0.1% SDS, 1% Triton x-100, 2mM EDTA, 150mM NaCl, 20mM Tris HCl pH 8.0), high salt immune complex buffer (0.1% SDS, 1% Triton x-100, 2mM EDTA, 500mM NaCl, 20mM Tris Hcl pH8.0), LiCl immune complex buffer (0.25M LiCl, 1% NP-40, 1% deoxycholate, 1mM EDTA, 10mM Tris-HCl pH8.0), and 1xTE (10mM Tris-HCl, 1mM EDTA pH8.0). Immune complexes were eluted in elution solution, (1% SDS, 0.1M sodium bicarbonate) in a thermomixer for 30 min at 65˚C at 1,200rpm. Crosslinks were reversed and proteins were digested with the addition of 200mM NaCl and 2ul Proteinase K in a thermocycler at 65˚C for 16 hours. DNA was purified with a Qiaquick PCR cleanup (Qiagen), and libraries were prepared with a NEBNext Ultra II Library Prep Kit (New England Biolabs).

ChIP-seq analysis

Adapters were removed using `cutadapt` [67]. Reads were aligned to the *hg38* genome assembly with `bowtie2` [68]. Duplicate reads were removed, and the remaining reads were sorted into *BAM* files and converted to *bed* format for counting with `samtools` [69]. Reads were also converted to *bigWig* format with `deeptools` [70]. Peaks were called with `MACS2` [35]. Reads were counted in peaks using `bedtools`, and differentially bound peaks were identified with `DESeq2` [37, 71]. Heatmaps were generated with `deeptools`. Peak proximity to and overlap with other features were calculated with `bedtools`.

ATAC-seq library preparation

ATAC-seq libraries were prepared as previously described [72], with modifications. 4 replicates were performed from cells treated and collected at different times in the same day. $4*10^5$ cells per sample were plated in each well of a 6-well plate 2 days before harvest. Cells were treated with DMSO or 50nM dTAG-13 and 50nM $dTAG^V$-1 in DMSO at various time points and collected simultaneously. At the time of harvest, cells were moved to ice and scraped in 1mL ice cold PBS, and 100μL ($\sim 5 \times 10^4$ cells) were transferred to 1.5 mL tubes. Cells were centrifuged at 500 x *g* for 5 minutes at 4˚C, and the pellets were resuspended in 50 μL cold lysis buffer (10mM Tris-HCl, 10mM NaCl, 3mM MgCl$_2$, 0.1% NP-40, 0.1% Tween-20, 0.01% Digitonin, adjusted to pH 7.4) and incubated on ice for 3 minutes. Samples were washed with 1 mL cold wash buffer (10mM Tris-HCl, 10mM NaCl, 3mM MgCl$_2$, 0.1% Tween-20). Cells were centrifuged at 500 x *g* for 10 minutes at 4˚C, and pellets were resuspended in the transposition reaction mix (25 μL 2X TD buffer (Illumina), 2.5 μL TDE1 Tn5 transposase (Illumina), 16.5 μL PBS, 0.5 μL 1% Digitonin, 0.5 μL 10% Tween-20, 5 μL nuclease-free water) and incubated in a thermomixer at 37˚C and 100rpm for 30 minutes. DNA was extracted with the DNA Clean and Concentrator-5 Kit (Zymo Research). Sequencing adapters were attached to the transposed DNA fragments using NEBNext Ultra II Q5 PCR mix (New England Biolabs), and libraries were amplified with 8 cycles of PCR. PEG-mediated size fractionation [73] was performed on the libraries by mixing SPRIselect beads (Beckman) with each sample at a 0.5:1 ratio, then placing the reaction vessels on a magnetic stand. The right side selected sample was transferred to a new reaction vessel, and more beads were added for a final ratio of 1.8:1. The final size-selected sample was eluted into nuclease-free water. This size selection protocol was repeated to further remove large fragments.

ATAC-seq analysis

Adapters were removed using `cutadapt` [67]. Reads aligning to the mitochondrial genome with `bowtie2` [68] were removed. The remaining reads were aligned to the *hg38* genome assembly with `bowtie2`. Duplicate reads were removed, and the remaining reads were sorted into *BAM* files with `samtools` [69]. Reads were converted to *bed* format with `seqOutBias` and *bigWig* format with `deeptools` [70, 74]. Accessibility peaks were called with `MACS2` [35]. Reads were counted in peaks using `bedtools`, and differentially accessible peaks were identified with `DESeq2` [37, 71]. *de novo* motif identification was performed on dynamic peaks with `MEME`, and `TOMTOM` was used to match motifs to the HOMER, Jaspar, and Uniprobe TF binding motif databases [75–78]. `AME` was used to identify motifs enriched in increased or decreased peaks relative to unchanged peaks [79]. `FIMO` and `bedtools` were used to assess motif enrichment around peak summits [80]. Dynamic peaks were clustered into response groups using `DEGreport` [81].

PRO-seq library preparation

Cell permeabilization was performed as previously described [82], with modifications. 4 replicates were performed from cells treated and collected at different times in the same day. For the time course experiment, $8*10^6$ dTAG-TRPS1 Clone 28 cells per sample were plated in 15cm dishes 2 days before harvest. Cells were treated with DMSO or 50nM dTAG-13 and 50nM dTAG^V-1 in DMSO at various time points and collected simultaneously. For the three clone experiment, $4*10^6$ cells per sample were plated in 10cm dishes 1 day before harvest. Cells were treated with DMSO or 100nM dTAG-13 in DMSO for 90 minutes and collected simultaneously.

At the time of harvest, cells were scraped in 10mL ice cold PBS and washed in 5mL buffer W (10mM Tris-HCl pH 7.5, 10mM KCl, 150mM sucrose, 5mM $MgCl_2$, 0.5mM $CaCl_2$, 0.5mM DTT, 0.004U/mL SUPERaseIN RNase inhibitor (Invitrogen), Complete protease inhibitors (Roche)). Cells were permeabilized by incubating with buffer P (10 mM Tris-HCl pH 7.5, KCl 10 mM, 250 mM sucrose, 5 mM $MgCl_2$, 1 mM EGTA, 0.05% Tween-20, 0.1% NP40, 0.5 mM DTT, 0.004 units/mL SUPERaseIN RNase inhibitor (Invitrogen), Complete protease inhibitors (Roche)) for 3 minutes on ice. Cells were washed with 10 mL buffer W before being transferred into 1.5mL tubes using wide bore pipette tips. Finally, cells were resuspended in 50μL buffer F (50mM Tris-HCl pH 8, 5mM $MgCl_2$, 0.1mM EDTA, 50% Glycerol, 0.5 mM DTT). Cells were snap frozen in liquid nitrogen and stored at -80˚C.

PRO-seq libraries were prepared as previously described [83], with modifications. RNA extraction after the run-on reaction was performed with 500μL Trizol LS (Thermo Fisher) followed by 130μL chloroform (Sigma). The equivalent of 1μL of 50μM for each adapter was used. A random eight base unique molecular identifier (UMI) was included at the 5' end of the adapter ligated to the 3' end of the nascent RNA. 37˚C incubations were performed with rotation with 1.5mL tubes placed in 50mL conical tubes in a hybridization oven. For the reverse transcription reaction, RP1 was used at 100μM and dNTP mix was used at 10mM each. Libraries were amplified by PCR for a total of 8 cycles in 100μL reactions with Phusion polymerase (New England Biolabs). No PAGE purification was performed to ensure that our libraries were not biased against short nascent RNA insertions.

PRO-seq analysis

Adapters were removed using `cutadapt` [67]. Libraries were deduplicated using `fqdedup` and the 3' UMIs [84]. UMIs were removed, and reads were reverse complemented with the-`seqtk`. Reads aligning to the rDNA genome with `bowtie2` [68] were removed. The

remaining reads were aligned, sorted, and convert to *bed* and *bigWig* files with `bowtie2`, `samtools`, `seqOutBias`, and `deeptools`, respectively [69, 70, 74]. Composite profiles around TRPS1 peaks were generated with `deeptools`. Reads were counted in genes using `bedtools`, and differentially expressed genes were identified with `DESeq2` [37, 71]. Dynamic genes were clustered into response groups using `DEGreport` [81]. Over-representation analysis was performed with `enrichr` [85], and gene set enrichment analysis was performed with `fgsea` [86], both using the Hallmark gene sets [41].

Genome browser visualization

Genome browser [87] images were taken from the following session: https://genome.ucsc.edu/s/tgscott/dTAG_TRPS1_ChIP_PRO_ATAC.

Cell number enumeration

$1.25*10^4$ cells per sample were plated in a 24-well plate. The next day (day 0), cells were treated with DMSO or 50nM dTAG-13 and 50nM dTAG^V-1 in DMSO. Media was replaced, maintaining the treatment condition, every 2 days. Cells were enumerated using a hemocytometer. 2 technical replicates were used on day 0, 2 for each treatment on day 3, and 3 for each treatment on days 7, 10, 14. The technical replicates were merged, and the experiment was performed in 4 biological replicates from different cell passages. The data were imported into R [62] for visualization and statistical analysis. A linear model was fit for the log-transformed cell number and the time. A second linear model was fit that included an interaction term between the time and the treatment condition, representing the effect of treatment on the doubling rate. Analysis of variance was performed on the two models to test for the significance of the interaction term.

TRPS1 activity score and patient outcome stratification

Primary TRPS1-regulated genes were defined based on the 30 minute time point using `DESeq2` [37]. This TRPS1 regulon was then used in `RTN` [48, 49] to define a TRPS1 activity score for each patient within the METABRIC cohort [46, 47].

Supporting information

S1 Fig. Fraction of reads in peaks (FRiP) for ChIP-seq libraries. FRiP scores for each library, calculated using the ChIPQC R package [34].
(PDF)

S2 Fig. Strand cross-correlation (CC) plots for ChIP-seq libraries. CC values for each library, calculated using the ChIPQC R package [34].
(PDF)

S3 Fig. Fragment size distribution plots for ATAC-seq libraries. A plot for each library was generated using the ATACseqQC R package [36].
(PDF)

S4 Fig. Plots of signal enrichment around TSS's for ATAC-seq libraries. A plot for each library was generated using the ATACseqQC R package [36].
(PDF)

S5 Fig. Quality control metrics for PRO-seq libraries. Quality control metrics are defined as in [39]. Each metric is a row, and each sample is a column. The green region for each metric is

the goal for a high quality library.
(PDF)

S6 Fig. Bidirectional transcription at TRPS1 peaks increases upon TRPS1 depletion. MA plot of TRPS1 ChIP-seq peaks from Fig 3G, with fold change values representing bidirectional transcription in the 30 minute dTAG-13 and dTAGV -1 at 50nM each (dTAG) treatment condition relative to the DMSO condition. Testing for a TRPS1 cistrome-wide increase in bidirectional transcription, the ANOVA F-test p-value was $< 2.2*10^{-16}$.
(PDF)

S7 Fig. Chromatin accessibility at ATAC-seq peaks without bidirectional transcription increases upon TRPS1 depletion. MA plot of ATAC-seq peaks, with fold change values representing accessibility in the 30 minute dTAG-13 and dTAGV -1 at 50nM each (dTAG) treatment condition relative to the DMSO condition, as in Fig 3.
(PDF)

S8 Fig. Acute estrogen treatment identifies direct ER target genes in T47D cells. MA plot of PRO signal, with fold change values representing transcription in the 90 minute estrogen treatment condition relative to the DMSO condition. Each point represents a gene, and black points represent the estrogen-activated genes that we use in Fig 5.
(PDF)

S9 Fig. The effect of TRPS1 depletion on ER target gene transcription is consistent across three independent clones. Fold changes in normalized PRO signal across three independent clones for ER target genes that are (A) activated or (B) repressed upon TRPS1 depletion, as defined in Fig 5A.
(PDF)

S10 Fig. T47D cells do not display a cell number defect with dTAG treatment. Cell number over time of parental T47D cells treated with dTAG or DMSO, as in Fig 6C.
(PDF)

S11 Fig. TRPS1 activity is associated with breast cancer patient outcomes specifically for ER-positive and Luminal A tumors. A) Kaplan-Meier curves for patients in the METABRIC cohort, stratified by TRPS1 activity as in Fig 6F, separated by ER-posivity. Logrank p-value $3.83*10^{-6}$ for ER-positive tumors and not significant for ER-negative tumors. B) Kaplan-Meier curves for patients in the METABRIC cohort, stratified by TRPS1 activity as in Fig 6F, separated by intrinsic subtype. Logrank p-value $4.23*10^{-4}$ for Luminal A tumors and not significant for the other subtypes.
(PDF)

S12 Fig. Genes differentially expressed after 24 hours of TRPS1 depletion are associated with breast cancer patient outcomes. Kaplan-Meier curves for patients in the METABRIC cohort, stratified by TRPS1 activity as in Fig 6F, but using genes differentially expressed after 24 hours of TRPS1 depletion. Logrank p-value $2.09*10^{-13}$.
(PDF)

S1 Table. ATAC-seq peaks that increase in accessibility upon 30 minutes of TRPS1 depletion (red dots in Fig 3A).
(TXT)

S2 Table. ATAC-seq peaks that decrease in accessibility upon 30 minutes of TRPS1 depletion (blue dots in Fig 3A).
(TXT)

S3 Table. Top 50 motifs significantly enriched in increased ATAC peaks relative to unchanged ATAC peaks at 30 minutes. Results generated using AME [79].
(CSV)

S4 Table. Motifs significantly enriched in decreased ATAC peaks relative to unchanged ATAC peaks at 30 minutes. Results generated using AME [79].
(CSV)

S5 Table. Genes that are gradually activated upon TRPS1 depletion (from Fig 4A).
(TXT)

S6 Table. Genes that are transiently activated upon TRPS1 depletion (from Fig 4A).
(TXT)

S7 Table. Genes that are gradually repressed upon TRPS1 depletion (from Fig 4B).
(TXT)

S8 Table. Genes that are transiently repressed upon TRPS1 depletion (from Fig 4B).
(TXT)

S9 Table. ER target genes that are activated upon TRPS1 depletion (from Fig 5A).
(TXT)

S10 Table. ER target genes that are repressed upon TRPS1 depletion (from Fig 5A).
(TXT)

Author Contributions

Conceptualization: Thomas G. Scott, Kizhakke Mattada Sathyan, Daniel Gioeli, Michael J. Guertin.

Data curation: Thomas G. Scott.

Formal analysis: Thomas G. Scott, Daniel Gioeli, Michael J. Guertin.

Funding acquisition: Michael J. Guertin.

Investigation: Thomas G. Scott.

Methodology: Thomas G. Scott, Kizhakke Mattada Sathyan, Daniel Gioeli, Michael J. Guertin.

Project administration: Michael J. Guertin.

Resources: Michael J. Guertin.

Software: Thomas G. Scott, Michael J. Guertin.

Supervision: Daniel Gioeli, Michael J. Guertin.

Validation: Thomas G. Scott.

Visualization: Thomas G. Scott, Daniel Gioeli, Michael J. Guertin.

Writing – original draft: Thomas G. Scott.

Writing – review & editing: Thomas G. Scott, Kizhakke Mattada Sathyan, Daniel Gioeli, Michael J. Guertin.

References

1. Siegel RL, Miller KD, Wagle NS, Jemal A. Cancer statistics, 2023. CA Cancer J Clin. 2023; 73(1):17–48. https://doi.org/10.3322/caac.21763 PMID: 36633525

2. Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. Nature. 2000; 406(6797):747–752. https://doi.org/10.1038/35021093 PMID: 10963602

3. Cheang MCU, Martin M, Nielsen TO, Prat A, Voduc D, Rodriguez-Lescure A, et al. Defining breast cancer intrinsic subtypes by quantitative receptor expression. Oncologist. 2015; 20(5):474–482. https://doi.org/10.1634/theoncologist.2014-0372 PMID: 25908555

4. Harbeck N, Penault-Llorca F, Cortes J, Gnant M, Houssami N, Poortmans P, et al. Breast cancer (Primer). Nature Reviews: Disease Primers; London. 2019; 5(1):s41572–019.

5. Henderson BE, Ross RK, Pike MC, Casagrande JT. Endogenous hormones as a major factor in human cancer. Cancer Res. 1982; 42(8):3232–3239. PMID: 7046921

6. Hall JM, Couse JF, Korach KS. The multifaceted mechanisms of estradiol and estrogen receptor signaling. J Biol Chem. 2001; 276(40):36869–36872. https://doi.org/10.1074/jbc.R100029200 PMID: 11459850

7. Kumar V, Chambon P. The estrogen receptor binds tightly to its responsive element as a ligand-induced homodimer. Cell. 1988; 55(1):145–156. https://doi.org/10.1016/0092-8674(88)90017-7 PMID: 3167974

8. Hah N, Danko CG, Core L, Waterfall JJ, Siepel A, Lis JT, et al. A rapid, extensive, and transient transcriptional response to estrogen signaling in breast cancer cells. Cell. 2011; 145(4):622–634. https://doi.org/10.1016/j.cell.2011.03.042 PMID: 21549415

9. Francis PA, Pagani O, Fleming GF, Walley BA, Colleoni M, Láng I, et al. Tailoring Adjuvant Endocrine Therapy for Premenopausal Breast Cancer. N Engl J Med. 2018; 379(2):122–137. https://doi.org/10.1056/NEJMoa1803164 PMID: 29863451

10. Gnant M, Mlineritsch B, Stoeger H, Luschin-Ebengreuth G, Knauer M, Moik M, et al. Zoledronic acid combined with adjuvant endocrine therapy of tamoxifen versus anastrozol plus ovarian function suppression in premenopausal early breast cancer: final analysis of the Austrian Breast and Colorectal Cancer Study Group Trial 12. Ann Oncol. 2015; 26(2):313–320. https://doi.org/10.1093/annonc/mdu544 PMID: 25403582

11. Dodwell D, Wardley A, Johnston S. Postmenopausal advanced breast cancer: options for therapy after tamoxifen and aromatase inhibitors. Breast. 2006; 15(5):584–594. https://doi.org/10.1016/j.breast.2006.01.007 PMID: 16504510

12. Ko LJ, Engel JD. DNA-binding specificities of the GATA transcription factor family. Mol Cell Biol. 1993; 13(7):4011–4022. https://doi.org/10.1128/mcb.13.7.4011 PMID: 8321208

13. Malik TH, Shoichet SA, Latham P, Kroll TG, Peters LL, Shivdasani RA. Transcriptional repression and developmental functions of the atypical vertebrate GATA protein TRPS1. EMBO J. 2001; 20(7):1715–1725. https://doi.org/10.1093/emboj/20.7.1715 PMID: 11285235

14. Wang Y, Lin X, Gong X, Wu L, Zhang J, Liu W, et al. Atypical GATA transcription factor TRPS1 represses gene expression by recruiting CHD4/NuRD(MTA2) and suppresses cell migration and invasion by repressing TP63 expression. Oncogenesis. 2018; 7(12):96. https://doi.org/10.1038/s41389-018-0108-9 PMID: 30563971

15. Wang Y, Zhang J, Wu L, Liu W, Wei G, Gong X, et al. Tricho-rhino-phalangeal syndrome 1 protein functions as a scaffold required for ubiquitin-specific protease 4-directed histone deacetylase 2 de-ubiquitination and tumor growth. Breast Cancer Res. 2018; 20(1):83. https://doi.org/10.1186/s13058-018-1018-7 PMID: 30071870

16. Serandour AA, Mohammed H, Miremadi A, Mulder KW, Carroll JS. TRPS1 regulates oestrogen receptor binding and histone acetylation at enhancers. Oncogene. 2018; 37(39):5281–5291. https://doi.org/10.1038/s41388-018-0312-2 PMID: 29895970

17. Elster D, Tollot M, Schlegelmilch K, Ori A, Rosenwald A, Sahai E, et al. TRPS1 shapes YAP/TEAD-dependent transcription in breast cancer cells. Nat Commun. 2018; 9(1):3115. https://doi.org/10.1038/s41467-018-05370-7 PMID: 30082728

18. Cornelissen LM, Drenth AP, van der Burg E, de Bruijn R, Pritchard CEJ, Huijbers IJ, et al. TRPS1 acts as a context-dependent regulator of mammary epithelial cell growth/differentiation and breast cancer development. Genes Dev. 2020; 34(3-4):179–193. https://doi.org/10.1101/gad.331371.119 PMID: 31879358

19. Momeni P, Glöckner G, Schmidt O, von Holtum D, Albrecht B, Gillessen-Kaesbach G, et al. Mutations in a new gene, encoding a zinc-finger protein, cause tricho-rhino-phalangeal syndrome type I. Nat Genet. 2000; 24(1):71–74. https://doi.org/10.1038/71717 PMID: 10615131

20. Malik TH, Von Stechow D, Bronson RT, Shivdasani RA. 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-pha-langeal syndromes. Mol Cell Biol. 2002; 22(24):8592–8600. https://doi.org/10.1128/MCB.22.24.8592-8600.2002 PMID: 12446778

21. Gai Z, Zhou G, Itoh S, Morimoto Y, Tanishima H, Hatamura I, et al. Trps1 functions downstream of Bmp7 in kidney development. J Am Soc Nephrol. 2009; 20(11):2403–2411. https://doi.org/10.1681/ASN.2008091020 PMID: 19820125

22. Lin HY, Zeng D, Liang YK, Wei XL, Chen CF. GATA3 and TRPS1 are distinct biomarkers and prognos-tic factors in breast cancer: database mining for GATA family members in malignancies. Oncotarget. 2017; 8(21):34750–34761. https://doi.org/10.18632/oncotarget.16160 PMID: 28423734

23. Ai D, Yao J, Yang F, Huo L, Chen H, Lu W, et al. TRPS1: a highly sensitive and specific marker for breast carcinoma, especially for triple-negative breast cancer. Mod Pathol. 2021; 34(4):710–719. https://doi.org/10.1038/s41379-020-00692-8 PMID: 33011748

24. Huang JZ, Chen M, Zeng M, Xu SH, Zou FY, Chen D, et al. Down-regulation of TRPS1 stimulates epi-thelial-mesenchymal transition and metastasis through repression of FOXA1. J Pathol. 2016; 239 (2):186–196. https://doi.org/10.1002/path.4716 PMID: 26969828

25. Hu J, Su P, Jiao M, Bai X, Qi M, Liu H, et al. TRPS1 Suppresses Breast Cancer Epithelial-mesenchymal Transition Program as a Negative Regulator of SUZ12. Transl Oncol. 2018; 11(2):416–425. https://doi.org/10.1016/j.tranon.2018.01.009 PMID: 29471243

26. Stinson S, Lackner MR, Adai AT, Yu N, Kim HJ, O'Brien C, et al. TRPS1 targeting by miR-221/222 pro-motes the epithelial-to-mesenchymal transition in breast cancer. Sci Signal. 2011; 4(177):ra41. https://doi.org/10.1126/scisignal.2001538 PMID: 21673316

27. Yang J, Liu X, Huang Y, He L, Zhang W, Ren J, et al. TRPS1 drives heterochromatic origin refiring and cancer genome evolution. Cell Rep. 2021; 34(10):108814. https://doi.org/10.1016/j.celrep.2021.108814 PMID: 33691114

28. Zhang H, Ahearn TU, Lecarpentier J, Barnes D, Beesley J, Qi G, et al. Genome-wide association study identifies 32 novel breast cancer susceptibility loci from overall and subtype-specific analyses. Nat Genet. 2020; 52(6):572–581. https://doi.org/10.1038/s41588-020-0609-2 PMID: 32424353

29. Sollis E, Mosaku A, Abid A, Buniello A, Cerezo M, Gil L, et al. The NHGRI-EBI GWAS Catalog: knowl-edgebase and deposition resource. Nucleic Acids Res. 2023; 51(D1):D977–D985. https://doi.org/10.1093/nar/gkac1010 PMID: 36350656

30. Boughton AP, Welch RP, Flickinger M, VandeHaar P, Taliun D, Abecasis GR, et al. LocusZoom.js: interactive and embeddable visualization of genetic association study results. Bioinformatics. 2021; 37 (18):3017–3018. https://doi.org/10.1093/bioinformatics/btab186 PMID: 33734315

31. Meyers RM, Bryan JG, McFarland JM, Weir BA, Sizemore AE, Xu H, et al. Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. Nat Genet. 2017; 49(12):1779–1784. https://doi.org/10.1038/ng.3984 PMID: 29083409

32. Nabet B, Roberts JM, Buckley DL, Paulk J, Dastjerdi S, Yang A, et al. The dTAG system for immediate and target-specific protein degradation. Nat Chem Biol. 2018; 14(5):431–441. https://doi.org/10.1038/s41589-018-0021-8 PMID: 29581585

33. Nabet B, Ferguson FM, Seong BKA, Kuljanin M, Leggett AL, Mohardt ML, et al. Rapid and direct control of target protein levels with VHL-recruiting dTAG molecules. Nat Commun. 2020; 11(1):4687. https://doi.org/10.1038/s41467-020-18377-w PMID: 32948771

34. Carroll TS, Liang Z, Salama R, Stark R, de Santiago I. Impact of artifact removal on ChIP quality metrics in ChIP-seq and ChIP-exo data. Front Genet. 2014; 5:75. https://doi.org/10.3389/fgene.2014.00075 PMID: 24782889

35. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). Genome biology. 2008; 9(9):1–9. https://doi.org/10.1186/gb-2008-9-9-r137 PMID: 18798982

36. Ou J, Liu H, Yu J, Kelliher MA, Castilla LH, Lawson ND, et al. ATACseqQC: a Bioconductor package for post-alignment quality assessment of ATAC-seq data. BMC Genomics. 2018; 19(1):169. https://doi.org/10.1186/s12864-018-4559-3 PMID: 29490630

37. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology. 2014; 15(12):1–21. https://doi.org/10.1186/s13059-014-0550-8 PMID: 25516281

38. Kwak H, Fuda NJ, Core LJ, Lis JT. Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. Science. 2013; 339(6122):950–953. https://doi.org/10.1126/science.1229386 PMID: 23430654

39. Scott TG, Martins AL, Guertin MJ. Processing and evaluating the quality of genome-wide nascent transcription profiling libraries. bioRxiv. 2022; p. 2022–12.

40. Wang Z, Chu T, Choate LA, Danko CG. Identification of regulatory elements from nascent transcription using dREG. Genome Res. 2019; 29(2):293–303. https://doi.org/10.1101/gr.238279.118 PMID: 30573452

41. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. 2015; 1(6):417–425. https://doi.org/10.1016/j.cels.2015.12.004 PMID: 26771021

42. Carroll JS, Liu XS, Brodsky AS, Li W, Meyer CA, Szary AJ, et al. Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. Cell. 2005; 122(1):33–43. https://doi.org/10.1016/j.cell.2005.05.008 PMID: 16009131

43. Guertin MJ, Zhang X, Coonrod SA, Hager GL. Transient estrogen receptor binding and p300 redistribution support a squelching mechanism for estradiol-repressed genes. Mol Endocrinol. 2014; 28(9):1522–1533. https://doi.org/10.1210/me.2014-1130 PMID: 25051172

44. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005; 102(43):15545–15550. https://doi.org/10.1073/pnas.0506580102 PMID: 16199517

45. Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, et al. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat Genet. 2003; 34(3):267–273. https://doi.org/10.1038/ng1180 PMID: 12808457

46. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature. 2012; 486(7403):346–352. https://doi.org/10.1038/nature10983 PMID: 22522925

47. Pereira B, Chin SF, Rueda OM, Vollan HKM, Provenzano E, Bardwell HA, et al. The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. Nat Commun. 2016; 7:11479. https://doi.org/10.1038/ncomms11479 PMID: 27161491

48. Castro MAA, de Santiago I, Campbell TM, Vaughn C, Hickey TE, Ross E, et al. Regulators of genetic risk of breast cancer identified by integrative network analysis. Nat Genet. 2016; 48(1):12–21. https://doi.org/10.1038/ng.3458 PMID: 26618344

49. Fletcher MNC, Castro MAA, Wang X, de Santiago I, O'Reilly M, Chin SF, et al. Master regulators of FGFR2 signalling and breast cancer risk. Nat Commun. 2013; 4:2464. https://doi.org/10.1038/ncomms3464 PMID: 24043118

50. Michailidou K, Lindström S, Dennis J, Beesley J, Hui S, Kar S, et al. Association analysis identifies 65 new breast cancer risk loci. Nature. 2017; 551(7678):92–94. https://doi.org/10.1038/nature24284 PMID: 29059683

51. Ripatti P, Rámó JT, Mars NJ, Fu Y, Lin J, Sóderlund S, et al. Polygenic Hyperlipidemias and Coronary Artery Disease Risk. Circ Genom Precis Med. 2020; 13(2):e002725. https://doi.org/10.1161/CIRCGEN.119.002725 PMID: 32154731

52. Richardson TG, Sanderson E, Palmer TM, Ala-Korpela M, Ference BA, Davey Smith G, et al. Evaluating the relationship between circulating lipoprotein lipids and apolipoproteins with risk of coronary heart disease: A multivariable Mendelian randomisation analysis. PLoS Med. 2020; 17(3):e1003062. https://doi.org/10.1371/journal.pmed.1003062 PMID: 32203549

53. Sakaue S, Kanai M, Tanigawa Y, Karjalainen J, Kurki M, Koshiba S, et al. A cross-population atlas of genetic associations for 220 human phenotypes. Nat Genet. 2021; 53(10):1415–1424. https://doi.org/10.1038/s41588-021-00931-x PMID: 34594039

54. Gill G, Ptashne M. Negative effect of the transcriptional activator GAL4. Nature. 1988; 334(6184):721–724. https://doi.org/10.1038/334721a0 PMID: 3412449

55. Tasset D, Tora L, Fromental C, Scheer E, Chambon P. Distinct classes of transcriptional activating domains function by different mechanisms. Cell. 1990; 62(6):1177–1187. https://doi.org/10.1016/0092-8674(90)90394-T PMID: 2205398

56. Meyer ME, Gronemeyer H, Turcotte B, Bocquel MT, Tasset D, Chambon P. Steroid hormone receptors compete for factors that mediate their enhancer function. Cell. 1989; 57(3):433–442. https://doi.org/10.1016/0092-8674(89)90918-5 PMID: 2720778

57. Bocquel MT, Kumar V, Stricker C, Chambon P, Gronemeyer H. The contribution of the N- and C-terminal regions of steroid receptors to activation of transcription is both receptor and cell-specific. Nucleic Acids Res. 1989; 17(7):2581–2595. https://doi.org/10.1093/nar/17.7.2581 PMID: 2717402

58. Schmidt SF, Larsen BD, Loft A, Mandrup S. Cofactor squelching: Artifact or fact? Bioessays. 2016; 38 (7):618–626. PMID: 27273739

59. Radvanyi L, Singh-Sandhu D, Gallichan S, Lovitt C, Pedyczak A, Mallo G, et al. The gene associated with trichorhinophalangeal syndrome in humans is overexpressed in breast cancer. Proc Natl Acad Sci U S A. 2005; 102(31):11005–11010. https://doi.org/10.1073/pnas.0500904102 PMID: 16043716

60. Savinainen KJ, Linja MJ, Saramäki OR, Tammela TLJ, Chang GTG, Brinkmann AO, et al. Expression and copy number analysis of TRPS1, EIF3S3 and MYC genes in breast and prostate cancer. Br J Cancer. 2004; 90(5):1041–1046. https://doi.org/10.1038/sj.bjc.6601648 PMID: 14997205

61. Chen JQ, Litton J, Xiao L, Zhang HZ, Warneke CL, Wu Y, et al. Quantitative immunohistochemical analysis and prognostic significance of TRPS-1, a new GATA transcription factor family member, in breast cancer. Horm Cancer. 2010; 1(1):21–33. https://doi.org/10.1007/s12672-010-0008-8 PMID: 21761348

62. R Core Team. R: A Language and Environment for Statistical Computing; 2021. Available from: https://www.R-project.org/.

63. Sathyan KM, Scott TG, Guertin MJ. ARF-AID: A Rapidly Inducible Protein Degradation System That Preserves Basal Endogenous Protein Levels. Curr Protoc Mol Biol. 2020; 132(1):e124. https://doi.org/10.1002/cpmb.124 PMID: 32757370

64. Sathyan KM, McKenna BD, Anderson WD, Duarte FM, Core L, Guertin MJ. An improved auxin-inducible degron system preserves native protein levels and enables rapid and specific protein depletion. Genes & development. 2019; 33(19-20):1441–1455. https://doi.org/10.1101/gad.328237.119 PMID: 31467088

65. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. Nat Protoc. 2013; 8(11):2281–2308. https://doi.org/10.1038/nprot.2013.143 PMID: 24157548

66. Zheng Q, Cai X, Tan MH, Schaffert S, Arnold CP, Gong X, et al. Precise gene deletion and replacement using the CRISPR/Cas9 system in human cells. Biotechniques. 2014; 57(3):115–124. https://doi.org/10.2144/000114196 PMID: 25209046

67. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet journal. 2011; 17(1):10–12. https://doi.org/10.14806/ej.17.1.200

68. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nature methods. 2012; 9 (4):357–359. https://doi.org/10.1038/nmeth.1923 PMID: 22388286

69. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. Bioinformatics. 2009; 25(16):2078–2079. https://doi.org/10.1093/bioinformatics/btp352 PMID: 19505943

70. Ramírez F, Dündar F, Diehl S, Grüning BA, Manke T. deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res. 2014; 42(Web Server issue):W187–91. https://doi.org/10.1093/nar/gku365 PMID: 24799436

71. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010; 26(6):841–842. https://doi.org/10.1093/bioinformatics/btq033 PMID: 20110278

72. Grandi FC, Modi H, Kampman L, Corces MR. Chromatin accessibility profiling by ATAC-seq. Nat Protoc. 2022; 17(6):1518–1552. https://doi.org/10.1038/s41596-022-00692-9 PMID: 35478247

73. Lis JT. [42] Fractionation of DNA fragments by polyethylene glycol induced precipitation. In: Methods in enzymology. vol. 65. Elsevier; 1980. p. 347–353.

74. Martins AL, Walavalkar NM, Anderson WD, Zang C, Guertin MJ. Universal correction of enzymatic sequence bias reveals molecular signatures of protein/DNA interactions. Nucleic acids research. 2018; 46(2):e9–e9. https://doi.org/10.1093/nar/gkx1053 PMID: 29126307

75. Bailey TL, Johnson J, Grant CE, Noble WS. The MEME suite. Nucleic acids research. 2015; 43(W1): W39–W49. https://doi.org/10.1093/nar/gkv416 PMID: 25953851

76. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Molecular cell. 2010; 38(4):576–589. https://doi.org/10.1016/j.molcel.2010.05.004 PMID: 20513432

77. Khan A, Fornes O, Stigliani A, Gheorghe M, Castro-Mondragon JA, Van Der Lee R, et al. JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. Nucleic acids research. 2018; 46(D1):D260–D266. https://doi.org/10.1093/nar/gkx1126 PMID: 29140473

78. Hume MA, Barrera LA, Gisselbrecht SS, Bulyk ML. UniPROBE, update 2015: new tools and content for the online database of protein-binding microarray data on protein–DNA interactions. Nucleic acids research. 2015; 43(D1):D117–D122. https://doi.org/10.1093/nar/gku1045 PMID: 25378322

79. McLeay RC, Bailey TL. Motif Enrichment Analysis: a unified framework and an evaluation on ChIP data. BMC Bioinformatics. 2010; 11:165. https://doi.org/10.1186/1471-2105-11-165 PMID: 20356413

80. Grant CE, Bailey TL, Noble WS. FIMO: scanning for occurrences of a given motif. Bioinformatics. 2011; 27(7):1017–1018. https://doi.org/10.1093/bioinformatics/btr064 PMID: 21330290

81. Pantano L. DEGreport: Report of DEG analysis. New Jersey, NJ: R package version. 2019;1(0).

82. Mahat DB, Kwak H, Booth GT, Jonkers IH, Danko CG, Patel RK, et al. Base-pair-resolution genome-wide mapping of active RNA polymerases using precision nuclear run-on (PRO-seq). Nature protocols. 2016; 11(8):1455–1476. https://doi.org/10.1038/nprot.2016.086 PMID: 27442863

83. Judd J, Wojenski LA, Wainman LM, Tippens ND, Rice EJ, Dziubek A, et al. A rapid, sensitive, scalable method for Precision Run-On sequencing (PRO-seq); 2020.

84. Martins AL, Guertin MJ. Remove PCR duplicates from FASTQ files. https://github.com/guertinlab/fqdedup. 2018;.

85. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics. 2013; 14:128. https://doi.org/10.1186/1471-2105-14-128 PMID: 23586463

86. Korotkevich G, Sukhov V, Budin N, Shpak B, Artyomov MN, Sergushichev A. Fast gene set enrichment analysis; 2021.

87. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. Genome research. 2002; 12(6):996–1006. https://doi.org/10.1101/gr.229102 PMID: 12045153