

RESEARCH

Open Access



Genomic dissection and mutation-specific target discovery for breast cancer *PIK3CA* hotspot mutations

Adam X. Miranda¹, Justin Kemp¹, Brad A. Davidson¹, Sara Erika Bellomo³, Verda E. Miranda², Alexandra Manoni¹, Caterina Marchiò^{3,4}, Sarah Croessmann¹, Ben H. Park^{1,5*} and Emily Hodges^{2,5,6*}

Abstract

Background Recent advancements in high-throughput genomics and targeted therapies have provided tremendous potential to identify and therapeutically target distinct mutations associated with cancers. However, to date the majority of targeted therapies are used to treat all functional mutations within the same gene, regardless of affected codon or phenotype.

Results In this study, we developed a functional genomic analysis workflow with a unique isogenic cell line panel bearing two distinct hotspot *PIK3CA* mutations, E545K and H1047R, to accurately identify targetable differences between mutations within the same gene. We performed RNA-seq and ATAC-seq and identified distinct transcriptional and epigenomic differences associated with each *PIK3CA* hotspot mutation. We used this data to curate a select CRISPR knock out screen to identify mutation-specific gene pathway vulnerabilities. These data revealed AREG as a E545K-preferential target that was further validated through in vitro analysis and publicly available patient databases.

Conclusions Using our multi-modal genomics framework, we discover distinct differences in genomic regulation between *PIK3CA* hotspot mutations, suggesting the *PIK3CA* mutations have different regulatory effects on the function and downstream signaling of the PI3K complex. Our results demonstrate the potential to rapidly uncover mutation specific molecular targets, specifically AREG and a proximal gene regulatory region, that may provide clinically relevant therapeutic targets. The methods outlined provide investigators with an integrative strategy to identify mutation-specific targets for the treatment of other oncogenic mutations in an isogenic system.

Keywords Integrative analysis, *PIK3CA*, Isogenic cell lines, Breast cancer, Epigenomics, Gene expression, Chromatin accessibility

*Correspondence:

Ben H. Park
ben.h.park@vumc.org
Emily Hodges
emily.hodges@vanderbilt.edu

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

In the past few decades, significant strides in precision medicine and the advancement of targeted therapies have led to personalized treatment options and improved outcomes for patients with cancer, while limiting off-target toxicities. However, response to treatment still varies widely and the ability to better identify patients that would benefit from targeted therapies remains complex [1]. For the most part, current clinical practices regard mutations within the same gene as clinically equivalent despite distinct molecular differences, creating a significant obstacle in the implementation of targeted therapies [2, 3]. *PIK3CA*, which encodes the p110 α subunit of phosphoinositide 3-kinase (PI3K), is the most commonly mutated gene in breast cancer and is responsible for regulating a diverse range of cellular functions including cell proliferation and survival [4–6]. *PIK3CA* has two distinct and highly prevalent hotspot mutations, E545K and H1047R, which occur in the helical and kinase domains, respectively. Mutations in *PIK3CA* are more common in the luminal A subtype of breast cancer and occur at a lower frequency in the triple negative subtype. Yet, the hotspot mutations of *PIK3CA* consistently occur at roughly a 2:1 (H1047R:E545K) ratio regardless of breast cancer subtype (Fig. S1 and Table S1)[7]. These two hotspot mutations have been shown to have distinct molecular changes and sensitivity to targeted therapeutics [5, 6, 8–10]. Despite these differences, current clinical application of PI3Kinase inhibitors in breast cancer do not distinguish between different mutations or between normal and mutated PI3K. This results in significant issues of toxicity, often times leading to dose reduction or discontinuation of the drug [2, 11–14]. To date, there is a distinct unmet need in the treatment of cancer, to accurately identify and understand molecular differences between mutations to effectively target cancer cells, improve selectivity, and decrease off-target effects.

Recent advancements in genomics technology and the affordability of generating high throughput genomics data have allowed researchers to begin to better understand the nuanced differences between mutations within the same gene. Furthermore, bioinformatic efforts have begun integrating transcriptomic and epigenomic data to better understand distinct molecular differences among unique mutational profiles from cancer patients. However, due to the significant mutational variability among individual cancers, as well as tumor heterogeneity and clonality, attributing observed differences to a single mutation has proven difficult. To better understand the molecular differences of *PIK3CA* hotspot mutations, our group has developed an integrative discovery platform to better identify key differences induced by different *PIK3CA* hotspot mutations in an isogenic human

breast epithelial cell line panel [15]. The utilization of an isogenic mutation panel allows comparisons of the individual *PIK3CA* mutations under the expression of the endogenous promoter in near isolation, allowing for the identification of potential mutation-specific and mutation-preferential therapeutic targets.

The discovery platform presented here integrates RNA-seq, an assay for transposase-accessible chromatin with sequencing (ATAC-seq), and a select CRISPR knockout (KO) screen to uniquely identify distinct molecular targets attributed to either the *PIK3CA* E545K or H1047R mutations within a well-controlled model. RNA-seq allows for the identification and quantification of genes and pathways with altered expression due to the presence of either mutation [16]. ATAC-seq measures chromatin accessibility and can identify putative gene regulatory elements to provide additional insight into how regulation of genes and binding activities of transcription factors differ between two mutations within the same gene [17–19]. In our framework, data from these two assays are used to tailor a CRISPR screen that can accurately confirm genes with high essentiality in either mutant cell line; in doing so, we identify potential mutation-specific targets for treatment [20–22]. Combined application of these assays provides improved understanding of differences in cell function induced by distinct hotspot mutants as well as providing potential means of mutation-preferential inhibition.

Herein we describe a systematic approach (Fig. 1) to identify potential mutation-preferential therapeutic targets. The utilization of an isogenic mammary epithelial cell model allows for the direct attribution of differences to specific mutations. This in return should improve selectivity of targeted therapies and decrease off-target effects. Our goal is to create a framework with “plug and play” accessibility for the evaluation of other hotspot mutants across cancer types using isogenic cell line models and to provide a foundation for future studies to identify a candidate list to maximize the potential for therapeutic benefit.

Results

RNA sequencing uncovers distinct transcriptional profiles and differential regulation of key cancer pathways in E545K and H1047R *PIK3CA* mutant cells

To evaluate differences in the transcriptomes of cells harboring the *PIK3CA* hotspot mutations E545K and H1047R, we performed RNA-seq on a panel of isogenically modified nontumorigenic breast epithelial MCF-10A cell lines harboring the respective mutations. RNA-seq identified 1271 genes with differential expression between the two mutant cell lines (Fig. 2A and B) [23]. A complete summary of the differentially expressed

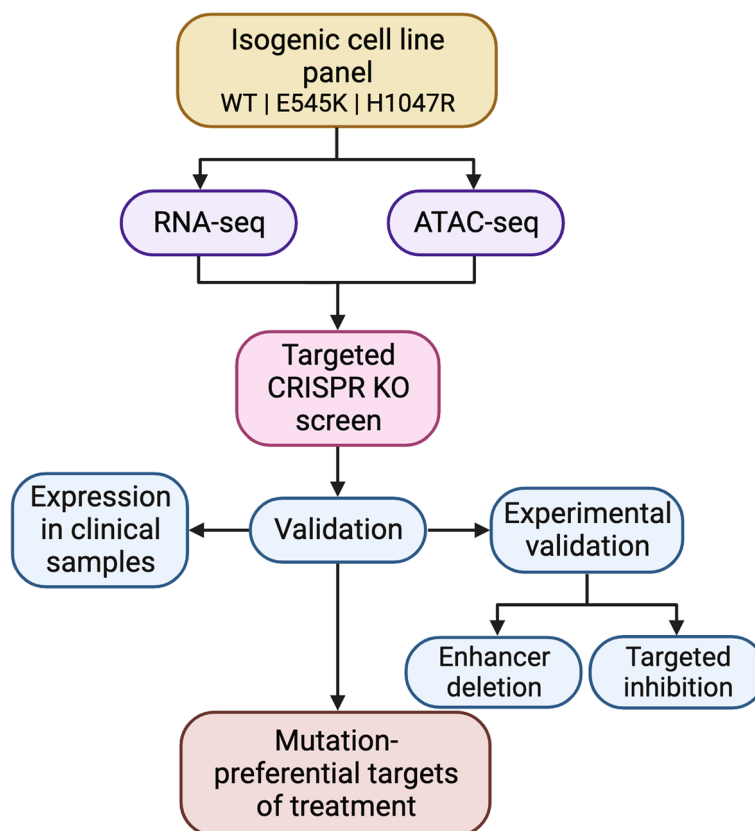


Fig. 1 Discovery platform identifies mutation-preferential gene targets from isogenic cell line models. Flowchart breaking down the process of identifying selective gene targets from an isogenic cell line model

genes (DEGs) can be found in Table S2 and are displayed in Fig. 2B. Interestingly, hierarchical clustering revealed the gene expression patterns of the E545K cell line shared greater similarity with the WT parental cell line than the H1047R mutant cell line (Fig. 2A). It is important to note that there is no differential expression of *PIK3CA* (Fig. S2). Thus, differential gene expression can be attributed to the effects of the mutations and not altered total expression of the mutant transcripts.

Gene set enrichment analysis using the MSigDB Hallmark pathway collection was performed to identify patterns of shared function across DEGs [24]. Multiple uniquely enriched pathways were associated with each mutant (Fig. 2C). Genes within pathways related to cell cycle and proliferation, as well as epithelial-mesenchymal transition genes, exhibited greater increased expression in the E545K cells, while genes in estrogen response pathways and K-ras associated genes had greater increased expression in the H1047R cells. It is important to note that while the MCF-10A lineage is considered an ER- cell line, these cells do still express *ESR1* mRNA (Fig. S2) and may therefore still exhibit changes in estrogen regulated genes. A more detailed look at the

altered expression of genes within the estrogen response early and estrogen response late pathways can be found in Fig. S3. These differences in gene expression patterns suggest distinct modes of tumorigenic activity between the different mutants, despite being treated as clinically equivalent. Using patient data from The Cancer Genome Atlas (TCGA) Breast Cancer (BRCA) data set, RNA-seq samples from tumors bearing each *PIK3CA* mutation confirmed observations made within our isogenic MCF-10A panel (Fig. 2D). Of the 14 pathways found to be significantly enriched in our panel, all were confirmed to be significantly enriched in differentially expressed gene sets from corresponding TCGA mutant samples. These findings demonstrate single amino acid substitutions in the same gene can have wide-ranging and distinct disruption of gene expression, which translates directly to expression differences observed in clinical samples.

PIK3CA mutants demonstrate unique differences in chromatin accessibility and gene regulation

Considering the gene expression changes observed with RNA-seq, we performed ATAC-seq to identify genomic regions with altered regulatory landscapes, which may

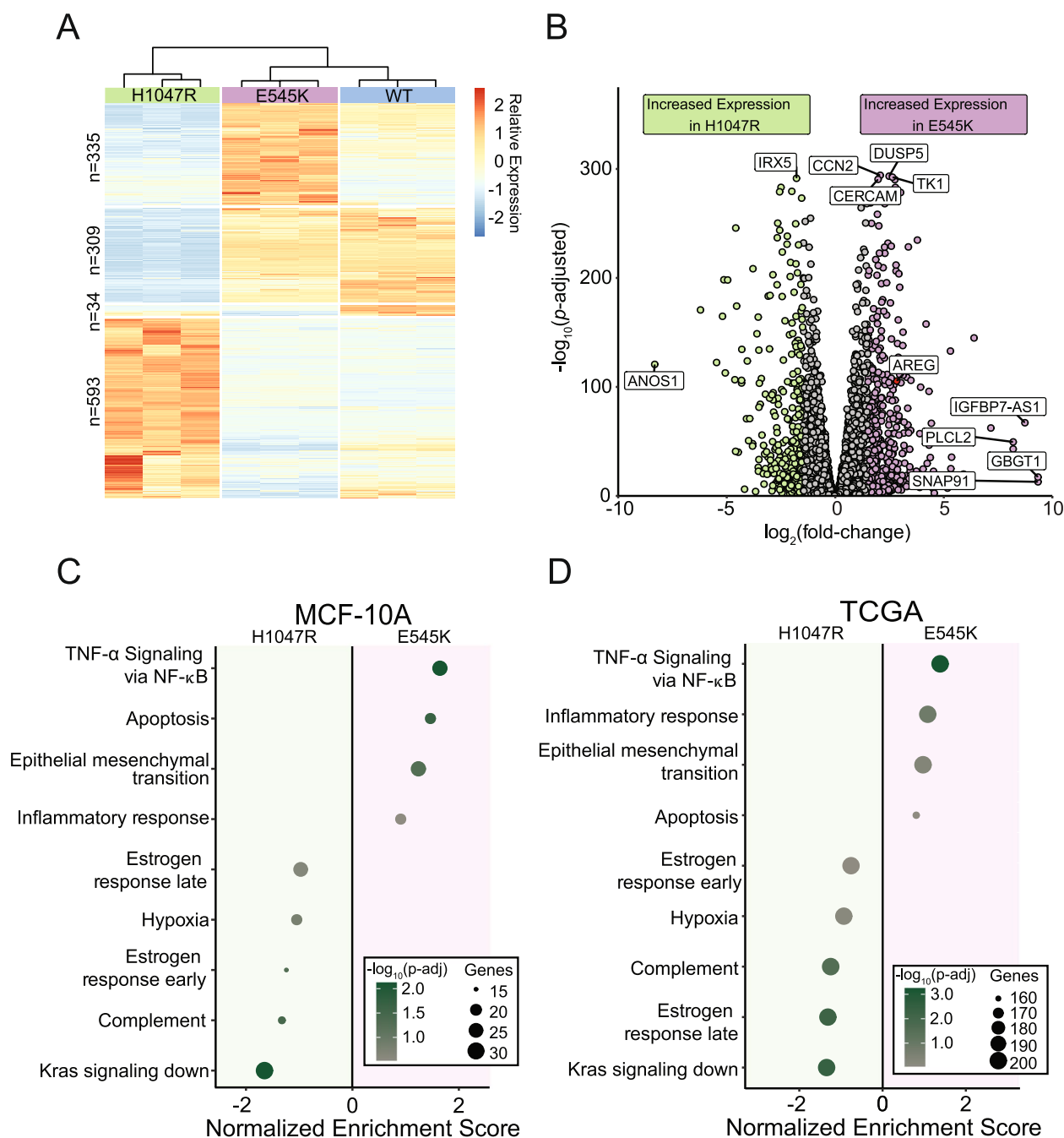


Fig. 2 RNA-seq captures distinct gene expression differences induced by *PIK3CA* hotspot mutations in isogenic cell line models which are reflected in TCGA patient samples. **(A)** Heatmap of normalized counts for 1271 differentially expressed genes. Hierarchical clustering of these genes reveals that E545K cells bear more similarity to WT than H1047R. **(B)** Volcano plot of differential expression between mutant isogenic cell lines. Differentially expressed genes (DEGs) were defined by the criteria: fold change $>|1.5|$, $P_{\text{adj}} < 0.05$. **(C, D)** Dot plot showing results from GSEA pathway enrichment analyses using the hallmark gene sets for **(C)** MCF-10A DEGs and **(D)** expression data from TCGA-BRCA samples. Pathways shown in panel D are those that are significantly enriched, shared and concordant with those identified as significant in the MCF-10A cell line

contribute to changes in gene expression. In addition to the identification of dynamic regions of chromatin accessibility (ChrAcc), differences in transcription factor (TF) binding activities were also estimated from Tn5 cut-site

profiles. Comparing accessibility profiles between E545K and H1047R mutants identified 8672 differentially accessible regions. We performed unsupervised clustering to define 4 distinct groups of accessibility patterns (Fig. 3A).

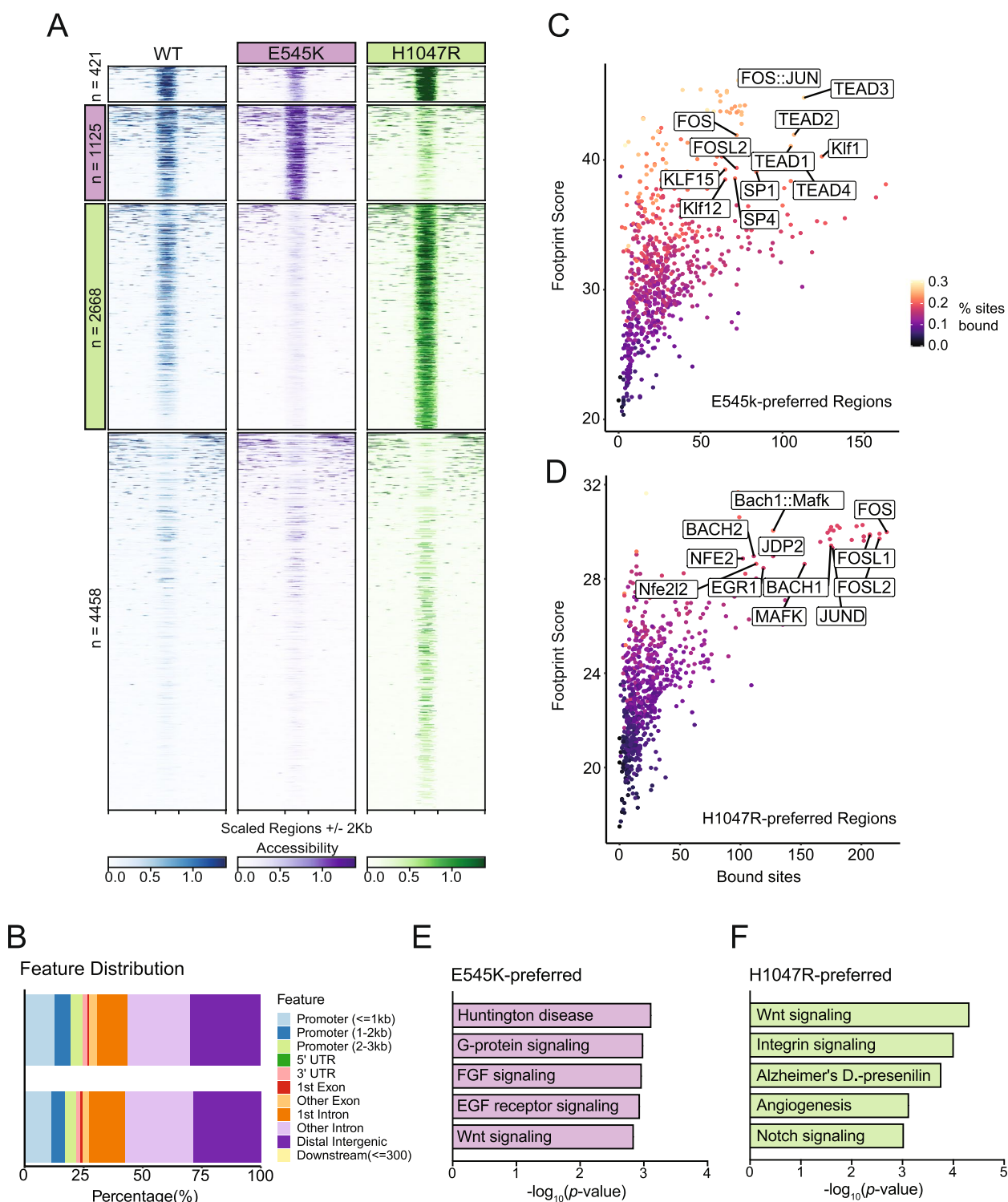


Fig. 3 ATAC-Seq identifies mutation-specific gene regulatory mechanisms near genes of key pathways. **(A)** Heatmap of accessibility at 8672 peaks exhibiting differential accessibility between the mutant isogenic cell lines. Peaks are divided based on k-means clustering. The second cluster highlighted in pink has been designated as E545K-preferred. The third cluster highlighted in green has been designated as H1047R-preferred **(B)** Distribution of genomic feature annotations of regions within the E545K-preferred and H1047R-preferred clusters. **(C)** Scatter plots of TOBIAS transcription factor footprinting of accessibility in the E545K-preferred regions. **(D)** Scatter plots of TOBIAS transcription factor footprinting of accessibility in the H1047R-preferred regions. **(E)** Bar plot of pathway enrichment from the PANTHER pathway database analysis performed on genes uniquely annotated to the E545K-preferred cluster. **(F)** Bar plot of pathway enrichment from the PANTHER pathway database analysis performed on genes uniquely annotated to the H1047R-preferred cluster regions

Two distinct groups, designated as E545K-preferred and H1047R-preferred, represented putative regulatory loci with increased accessibility in either E545K or H1047R mutant cells, respectively. In addition to providing insight into gene regulatory mechanisms, ChrAcc provides insight into *cis*-regulatory elements including enhancers. Within both the E545K-preferred and H1047R-preferred region clusters, over 50% of the regions fall within intronic and distal intergenic sequences (Fig. 3B). These results suggest that a significant amount of chromatin remodeling that is driven by the different *PIK3CA* mutations occurs at noncoding enhancer elements that can bind TFs and influence gene expression [25].

Indeed, TF motif analysis revealed that each accessibility cluster is enriched for distinct families of TF binding sites identified from the JASPAR database [26, 27] (Fig. S4). In the E545K-preferred cluster, strong enrichment for the hormone receptor transcription factors ARE and PGR were observed as well as the TEAD transcription factor family. The TEAD TF family has been shown to have a strong association with canonical PI3K/AKT signaling and can promote epithelial to mesenchymal transition [28, 29]. In the H1047R-preferred cluster, there was increased enrichment of AP-1 family TFs. AP-1 family TFs have been shown to interact with chromatin remodelers and promote a proliferative gene expression program [30, 31], and have also been associated with signaling through the MAPK cascade [32].

While TF motif analysis informs which sequences are enriched within ChrAcc regions, it does not predict TF occupancy. To better understand differential TF binding activities, we performed TF footprinting using TOBIAS, which uses Tn5 cut-site profiles to identify differences in proteins bound at TF binding motifs [18]. Our results show high levels of TEAD TF binding in E545K-preferred regions (Fig. 3C) and high levels of AP-1 binding (FOS, FOSL1, FOSL2, JUND) in H1047R-preferred regions (Fig. 3D). These results are consistent with the motif enrichment results and point to activity of the TEAD and AP-1 TF families as key regulators of differential gene expression between the *PIK3CA* hotspot mutants. The differential binding activity of TFs from these TF families are influenced by the *PIK3CA* mutation status of the cells and cofactors of these TFs likely alter the ChrAcc at these mutation-preferred regions. See discussion for more detailed description.

Nearest neighbor gene annotation using GREAT was used for gene ontology analysis to identify genes uniquely associated with either mutation-preferred accessibility cluster and analyzed for pathway enrichment using Enrichr [33–36]. Using the PANTHER database, we identified enrichment of distinct pathways promoted by either *PIK3CA* mutants [37, 38]. Within the E545K-preferred

cluster regions, unique enrichment of multiple growth factor receptor signaling pathways was observed and is likely due to changes in PI3K α signaling induced by the E545K mutant cells (Fig. 3E) [4]. Interestingly, enrichment from the H1047R-preferred cluster regions showed enrichment for both the Notch and Wnt signaling pathways. Both of these pathways are associated with the promotion of tumor growth in breast cancers, but neither are canonically associated with increased activity of PI3K α (Fig. 3F) [39, 40]. This suggests H1047R mutant cells may drive alternative proliferative cell signaling outside of canonical PI3K signaling. These gene ontology results are consistent with the observed TF enrichment and footprinting between clusters, and provide additional context to the differential gene expression observed from RNA-seq. The differences in ChrAcc demonstrate distinct differences in genomic regulation between *PIK3CA* mutations and suggest the *PIK3CA* mutations have different effects on the function and downstream signaling of the PI3K complex.

Select CRISPR-Cas9 knockout screen identifies genes with mutation-specific essentiality

A key advantage of our isogenic cell line model is the ability to compare both mutants to the unmodified parental cell line. Therefore, a CRISPR KO screen could accurately identify essential genes specific to *PIK3CA* mutations, but not *PIK3CA* WT cells, and may provide a list of promising therapeutic targets with limited off-target effects in normal cells. Performing a whole genome CRISPR screen can be both time and resource intensive. To circumvent these limitations, we used the data generated from both the previously performed RNA-seq and ATAC-seq assays to curate a select list of genes to investigate within our CRISPR KO screen. Analysis of RNA-seq data identified 616 unique DEGs, 160 for E545K mutants and 456 for H1047R mutants, with significantly upregulated expression in a mutant cell line relative to the parental (Fig. 4A and B). An additional 410 genes were identified to have increased expression in both mutant cell lines compared to WT; however, gene selection was limited to those with uniquely increased expression in either the E545K or H1047R cells (Fig. 4B). Among the 9677 combined mutant-specific genes that annotate to regions of differential chromatin accessibility, 312 were identified as DEGs (Fig. 4C).

Among the 312 selected genes, 280 genes with targeting single guide RNAs (sgRNAs) were available from the Brunello full genome library for a select CRISPR KO screen [41] (Table S3). Using the MAGeCK software package, Z-score differences between each of the mutant cell lines were compared to the parental cell line for each gene [42] (Fig. 4D). From this analysis, we

identified 36 genes with a Z-score difference greater than a significance threshold of 1.65, which corresponds to a confidence interval of 95% (Table S4). When knocked out, these genes specifically disrupt the survival of either mutant cell line with minimal disruption to the parental line (Table S4 and Fig. 4D). The top five genes (NAT1, PPM1H, AREG, ACSS1, CXCL1) with the greatest differences in Z-scores were evaluated in the *PIK3CA* mutant breast cancer samples from TCGA (Fig. 4E). Of the top 5 genes, AREG was the only gene with significant differential expression between the *PIK3CA* mutations. Samples with E545K mutations demonstrated a significant increase in expression when compared to the H1047R, recapitulating the differential expression of AREG observed in our isogenic panel (Fig. 4F). This association was independently confirmed using data from two other databases. The first of these databases was the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) database, which shows increased expression of AREG in E545K mutant samples with an increased difference within the luminal B subtype (Fig. S5) [43]. We also show that E545K mutant breast cancer cell lines show unique sensitivity to loss of AREG in data from project Achilles and DepMap (Fig. S6) [44]. This project contains a collection of CRISPR KO screen results from 1100 cancer cell lines, and this result is consistent with the findings from our CRISPR KO screen in the MCF-10A model, which shows a loss of AREG is much more deleterious to cells with the E545K mutation compared to other *PIK3CA* genotypes. Clinical confirmation of a unique molecular target identified from this select CRISPR screen emphasizes the translational potential of hits identified from isogenic mutant cell lines analyzed with our strategy.

Disruption of differentially accessible locus identified by ATAC-seq exhibits regulatory function over AREG expression

Selection criteria for inclusion of the ATAC-seq data in the CRISPR screen required identification of differentially accessible peaks between the two mutants. The

accessible peak annotated to AREG (chr4:74,435,384–74,435,596 locus) was identified as significantly more accessible in the E545K mutant cells and exhibits many qualities of a gene regulatory region (Fig. 5A). A previous study using CTCF ChIA-PET in MCF-10A cells published by ENCODE (ENCSR403ZYJ) showed that this locus interacts with the promoter of the AREG gene and could influence expression [45, 46]. Furthermore, the Genotype-Tissue Expression project (GTEx) identifies 5 different AREG expression quantitative trait loci (eQTLs) single nucleotide polymorphisms (SNPs) within 1 kb of this region (Table S5) and these SNPs have been shown to influence the expression of AREG in multiple tissue types (Fig. 5A). Specifically, the rs28570600 SNP (gold square, Fig. 5A) has previously been shown to be significantly associated with breast cancer susceptibility [47].

To investigate the function of this peak, the region was deleted using CRISPR-Cas9 and a pair of sgRNAs targeting the discussed locus upstream of the AREG TSS (oligo sequences in Table S6). Loss of this enhancer region significantly reduced AREG expression in all cell lines in the isogenic model (Fig. 5B). There was also an observed decline in the proliferation/survivability of the mutant cell lines, however only significant within the E545K cells. (Fig. 5C). This experiment was also performed in a previously developed isogenic model for the E545K mutation in the MCF7 breast cancer cell line background [48]. Results from this experiment also show a specific decrease in proliferation/survivability of E545K mutant cells (Figure S7). These assays demonstrate the capability of our approach to identify regulatory regions that may themselves provide targets for mutation-specific treatment of *PIK3CA* mutant disease.

PIK3CA mutant cells exhibit specific dependency on AREG

To confirm the role of AREG expression on cell survival, short interfering RNA (siRNAs) were used to inhibit AREG in the isogenic panel (Fig. S8). With an siRNA knockdown, our goal was to assess the effect of a reduction of AREG expression without complete loss of expression in the system. Consistent with observations

(See figure on next page.)

Fig. 4 Genes with altered expression and nearby chromatin accessibility were selected for a CRISPR KO screen to identify gene targets with mutation-specific essentiality. **(A)** Scatter plots of gene expression in mutant cells relative to parental cells. Unique genes meeting the fold change and significance threshold (fold change >1.5, $P_{adj} < 0.05$) shown in color. **(B)** Euler plot showing the overlap of genes with increased expression in cells with either mutation compared to the parental cell line. **(C)** Euler plot showing the overlap between genes annotated to regions of differential accessibility with those exhibiting increased expression in a mutant cell line compared to the parental cell line. **(D)** Scatter plot showing the results of the CRISPR KO screen. Significant hits with a Z-score difference >1.65] are shown in red. **(E)** Box plots showing expression of the top 5 hits from the CRISPR screen in TCGA-BRCA samples. AREG shows significant difference in expression between samples bearing either of the hotspot mutations. Significance was calculated using an ANOVA with a post-hoc Fisher's LSD test. Significance threshold is 0.05. **(F)** Bar plots showing the expression of the top 5 hits from the CRISPR screen in the MCF-10A RNA-seq samples. Significance was calculated using an ANOVA with a post-hoc Fisher's LSD test; * = p-value 0.05–0.0332, ** = p-value .0332–0.0021, *** = p-value 0.0021–0.0002, **** = p-value < 0.0002

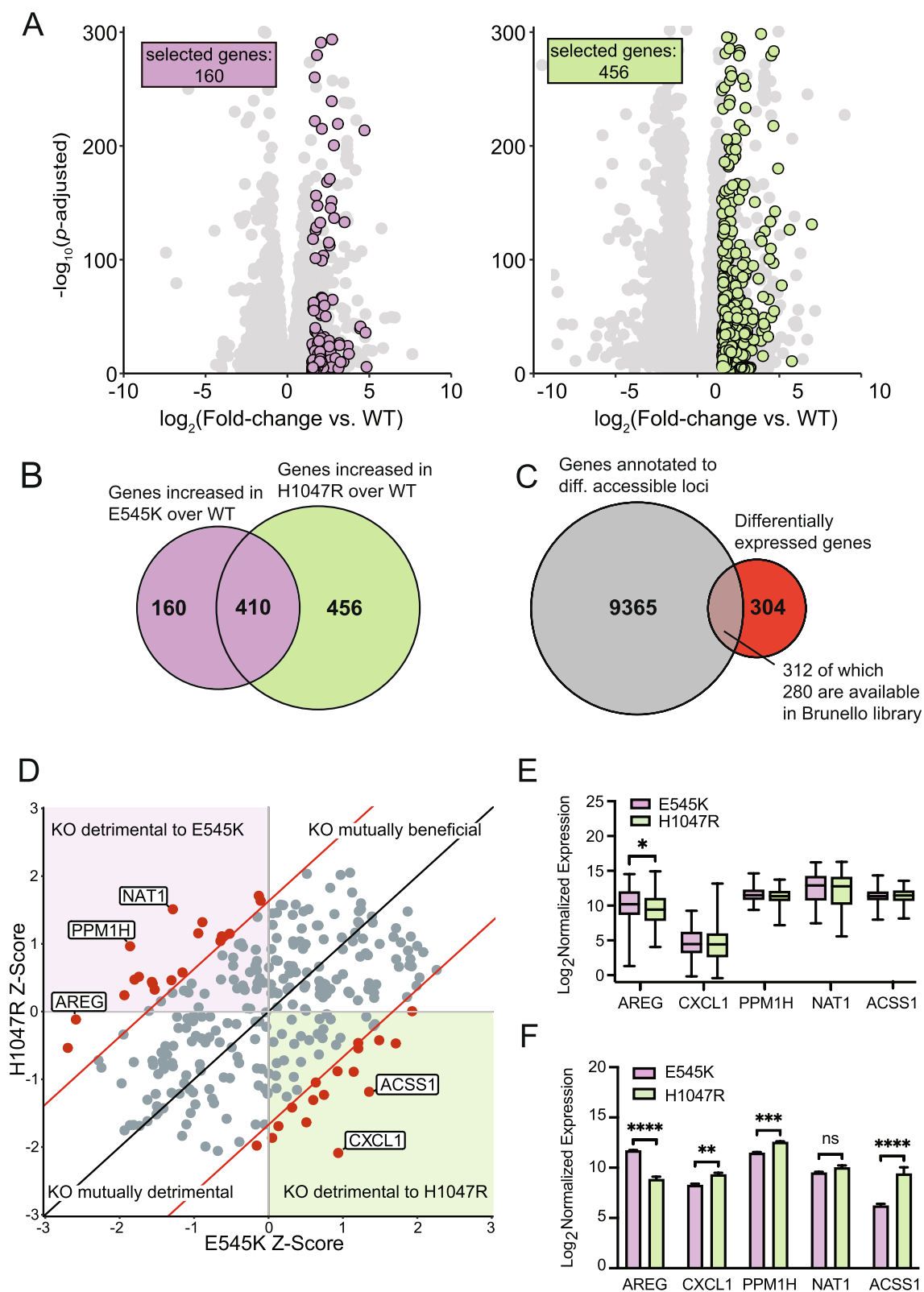


Fig. 4 (See legend on previous page.)

from the CRISPR KO screen, reduction of AREG expression significantly disrupted the survival and proliferation of E545K mutant cells, while exhibiting no significant changes in WT or H1047R mutant cells (Fig. 5D). To validate this observation, cells were treated with a neutralizing antibody that disrupts the extracellular signaling of AREG. Both mutant cell lines exhibited sensitivity to AREG perturbation, while WT cells showed no significant change (Fig. 5E). This effect is not exclusive to the E545K mutant cells but is consistent with the effect of AREG loss shown in Fig. 4D. Knockout of the AREG gene had a small deleterious effect on H1047R cells compared to a larger effect observed in E545K cells. The extracellular nature of AREG makes it a particularly attractive target as inhibitors of AREG would not necessarily need to penetrate the cell membrane to be effective. This could simplify drug design and reduce potential off target toxicity [49, 50]. Taken together, these results demonstrate the utility of our research strategy to identify potential molecular targets as an option for mutation-preferential therapeutic strategy. The in vitro and translational confirmation of our findings demonstrate the power of our model to accurately identify actionable gene targets and gene regulatory regions with high selectivity for mutant cells and minimal impact to WT cells.

Discussion

Increased availability and advancements in multi-omics technology have begun to revolutionize translational research to better understand the interplay of molecular changes and provide new opportunities for targeted therapies. However, integration and implementation of multi-omics data for identifying new molecular targets for therapeutic development remains underutilized in the cancer setting. This study presents an analytical framework for employing an isogenic mutant panel to better understand and uniquely identify the molecular differences between mutations within the same gene. Traditionally, most cancer-associated mutations have been clinically evaluated and treated as a monolithic group with variable success. More contemporary targeted therapies such as inhibitors specific to mutant KRAS G12C,

for example, highlight the success and feasibility of developing mutation specific inhibitors [51]. To improve upon this current paradigm, our workflow takes advantage of a model of isogenically incorporated mutations in a genetically stable background, integrating both RNA-seq and ATAC-seq data to design a uniquely tailored CRISPR KO screen enabling the detection of mutation-selective targets. The utilization of a mutant model that incorporates an isogenic background provides a system to identify a candidate list that demonstrates mutation-preferential gene regulatory dependencies. Previous studies have made use of this model to demonstrate how the mutant cells differ from the parental line, but our approach differs in focus in that the unique differences within the mutant cell lines are the priority [9, 15, 52]. Furthermore, the accessibility of CRISPR-Cas9 gene editing systems makes the development of isogenic models for cancer-associated mutations a relatively fast and straightforward process and can be scaled for a variety of mutations across tumor types. In addition to identifying mutation-preferential molecular targets, our comprehensive process paired with the isogenic panel can identify and characterize potential enhancer regions with mutation-specific activity that may offer alternative targets for treatment. These putative enhancers have affinity for distinct TF families that result in unique expression profiles and may be exploited as therapeutic vulnerabilities. The true utility of our process is in the identification of potential targets. Hits from our analyses still require additional validation to determine their effects beyond the CRISPR KO screen and ultimately their translation for potential clinical impact.

Evidence for the applicability of our workflow in breast cancer was used to analyze the two most common *PIK3CA* mutations in breast cancer to identify distinct molecular differences that impact downstream signaling, chromatin accessibility, and gene expression. RNA-seq and ATAC-seq analysis identified the disruption of epithelial-mesenchymal transition associated genes in E545K mutant cells and the MAPK cascade in H1047R mutant cells. These results suggest a model in which the hotspot mutations promote the activation of different

(See figure on next page.)

Fig. 5 E545K mutant cells exhibit specific dependence on *AREG*, which is regulated by nearby accessibility peak/putative enhancer. **(A)** Genomic tracks showing the ATAC-seq data across the isogenic cell lines alongside key SNPs at the *AREG* gene locus. The red box highlights the differentially accessible region annotated to the *AREG* gene. The green bars designate GTEx *AREG* eQTLs. The gold bar designates a GWAS SNP associated with breast cancer. **(B)** Bar plot of *AREG* expression following CRISPR-mediated deletion of the putative *AREG* enhancer. **(C)** Bar plot of cell counts following CRISPR-mediated deletion of the putative *AREG* enhancer. Significance of B and C was calculated using an ANOVA with a post-hoc Šidák's test. **(D)** Bar plot showing differences in survival/proliferation following inhibition of *AREG* expression using siRNA (Oligos on Table S6, expression of *AREG* shown in Fig. S6). Significance was calculated using an ANOVA with a post-hoc Tukey's test. **(E)** Bar plots showing differences in survival/proliferation following inhibition of *AREG* using a neutralizing antibody (R&D Systems, MAB262-SP). Significance was calculated using an ANOVA with a post-hoc Šidák's test; * = p-value 0.05–0.0332, ** = p-value .0332–0.0021, *** = p-value 0.0021–0.0002, **** = p-value < 0.0002

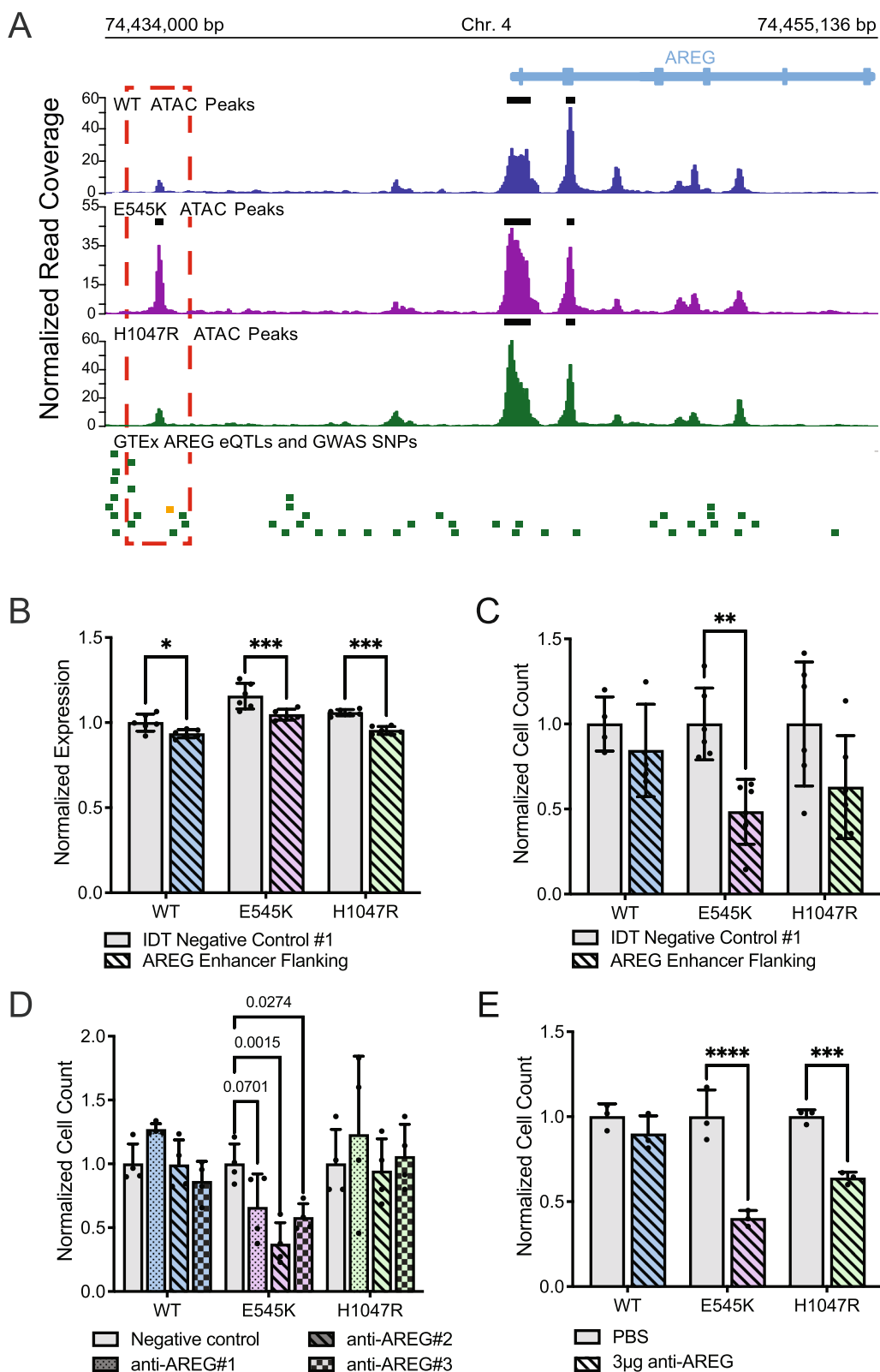


Fig. 5 (See legend on previous page.)

biochemical pathways that in turn signal through different TF families. These TFs activate mutation-specific chromatin remodeling and expression of target genes (Fig. 6). Integration of these assays to create a uniquely tailored, focused CRISPR screen allowed us to identify AREG as an E545K-specific exploitable molecular difference in a highly efficient manner. This mutation-preferential dependence on AREG suggests a positive feedback loop in which increased AREG expression further promotes signaling through the E545K-mutant PI3K complex (Fig. 6). AREG has been previously and independently established as a signaling molecule required for the growth of PIK3CA-mutant breast cancer cells [53]. Independent identification of this molecular target utilizing our approach demonstrates its immediate biological application. Furthermore, we were able to confirm translational applicability through retrospective analyses of publicly available patient data. While our experiments were performed in a single isogenic cell line model, these clinical findings suggest the applicability of our results to actual breast cancer patients. Taken together, this study provides a framework for the independent evaluation of oncogenic hotspot mutations from a functional genomics perspective. This implies that in the era of patient-specific treatment and pharmacogenomics, our process

may allow for the discovery of new targets and improved personalized medicine with the potential for increased specificity and decreased toxicity.

Conclusions

This work highlights the utility of integrating multiomics data collected from an isogenic mutant model to better identify molecular targets for therapy. With the increased accessibility to genome editing technology and services, our strategy can provide investigators with a clear method for studying specific mutants in other cancer cell line models. Our workflow was able to identify AREG as an E545K-preferential molecular target, which was confirmed through in vitro assays and retrospective analyses of patient data, highlighting the potential clinical utility of our work.

Materials and methods

Cell Culture

MCF-10A parental cell lines were purchased from American Type Culture Collection (ATCC). MCF-10A cell line knock-ins were generated as previously described [15]. All cell lines were grown in 5% CO₂ at 37 °C with 1% Penicillin/Streptomycin in respective media conditions. Parental MCF-10A cell lines were cultured in DMEM/

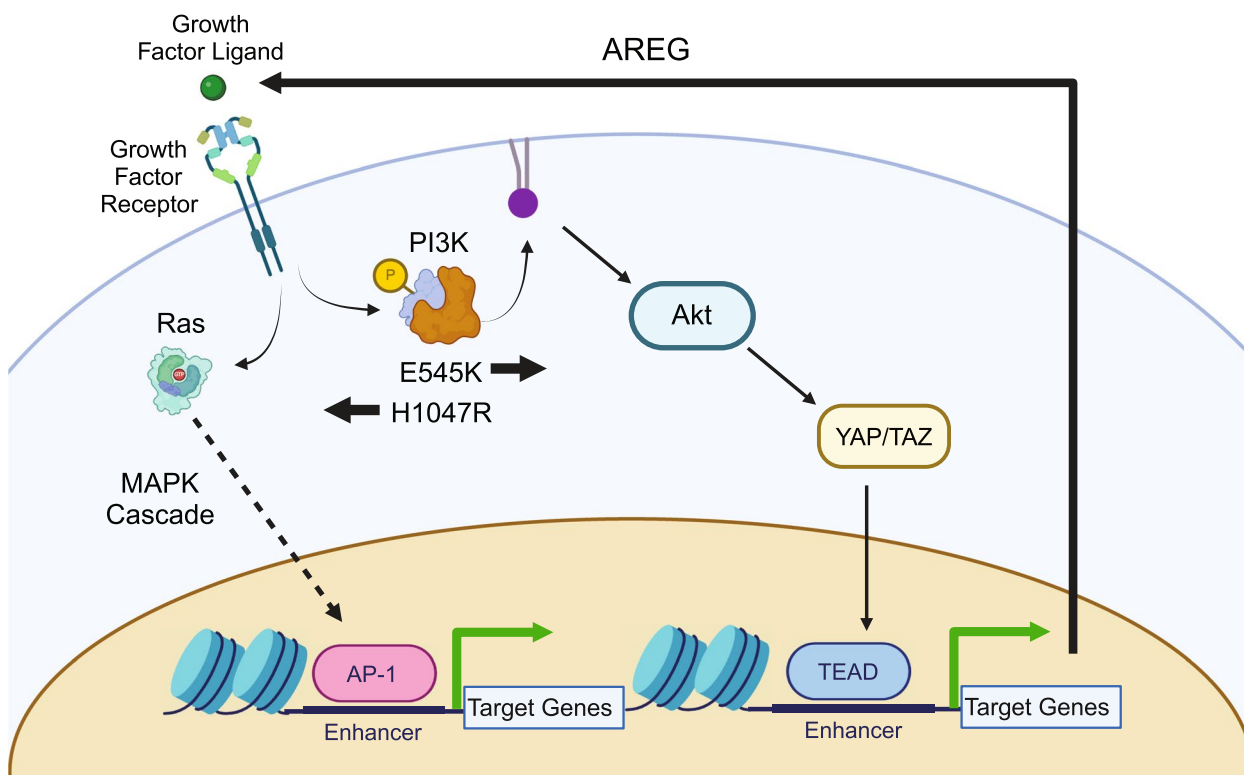


Fig. 6 Model of Mutation-specific cell signaling. Pathway diagram depicting the effects of *PIK3CA* hotspot mutations on the signaling of breast cells in the MCF-10A model

F12 (1:1) supplemented with 5% horse serum, 20 ng/ml epidermal growth factor (EGF), 10 µg/ml insulin (Roche), 0.5 µg/ml hydrocortisone (Sigma), and 100 ng/ml cholera toxin (Sigma). Knock-in cell lines were maintained in MCF-10A media in the absence of EGF. For all sequencing assays, cells were transferred to assay media 24 h prior to sample collection. Assay media contains phenol red-free DMEM/F12 (1:1) supplemented with 1% charcoal–dextran stripped FBS (Fisher), 0.2 ng/ml EGF, 10 µg/ml insulin, 0.5 µg/ml hydrocortisone, and 100 ng/ml cholera toxin.

RNA-Seq

RNA was isolated and prepared using the Qiagen RNeasy kit. Libraries were prepared by the Vanderbilt Technologies for Advanced Genomics (VANTAGE) Core using the Illumina Ribo-Zero Plus rRNA Depletion Kit. Each library was sequenced on an Illumina NovaSeq, PE150, at a requested depth of 50 million reads. All code and the specific parameters used in all data analyses can be found at: (<https://github.com/adamxmiranda/PIK3CA>). All sequencing library reads were trimmed of adapters and assessed for quality using the Trim Galore! (version 0.4.0) Wrapper of Cutadapt and FastQC [54, 55]. Trimmed reads were mapped to the human genome assembly hg38 using the Spliced Transcripts Alignment to a Reference (STAR) aligner (version 2.5.4b) [56]. Mapped reads were sorted and filtered for a mapping quality score over 30 using the SAMtools package (version 1.5) [57]. Reads were counted to gene transcripts using featureCounts (version 2.0.0) to version 32 of the GENCODE transcripts [58, 59]. A summary of the sequencing preprocessing can be found in Table S7. The degree to which gene expression could be affected by the genetic modification process was also evaluated by comparing the correlation of gene expression in each of our cell lines to RNA-seq data of the TWT MCF-10A cell line from Dalton et al. 2019 [60]. This cell line underwent the same genetic modification process, but a WT clone was selected. Clustering of these samples showed the targeted WT (TWT) samples cluster between the WT and E545K cells. This suggests that although there are some differences in gene expression introduced by the modification process, that these do not completely explain the observed differences between the lines in our model (Fig. S9). This aligns with previous work from our lab that showed limited phenotypic changes in targeted WT controls using this method of genomic modification [61]. Batch correction for this analysis was performed using the limma package (version 3.50.3). Differential gene expression was identified between conditions using the DESeq2 package [23]. Pathway analysis was performed using the fgsea package on the Hallmark gene set from MsigDB [24, 62].

ATAC-seq

Nuclei were isolated and ATAC-seq libraries were prepared using previously published methods [17, 63]. Libraries were sequenced by the VANTAGE Core on an Illumina NovaSeq PE150, at a requested depth of 50 million reads. Reads from the ATAC-seq libraries were trimmed using the same process described in the RNA-seq section. All code and specific parameters used in all data analyses can be found at: (<https://github.com/adamxmiranda/PIK3CA>). Trimmed reads were mapped to the human genome assembly hg38 using the BBTools (version 38.69) package and Burrows-Wheeler Aligner (version 0.7.17) [20, 64]. Quality filtering was performed on the mapped reads using SAMtools [57]. A summary of the sequencing preprocessing can be found in Table S7. Peaks of accessibility were called using Genrich (version 0.6.1) and differential accessibility was determined using DESeq2 (version 1.34.0) [23]. Accessible regions were clustered using k-means clustering. Gene annotation and pathway enrichment was performed using GREAT (version 4.0.4) [33, 34]. The gene annotation parameters used for GREAT were the default parameters of 5 kb upstream of the transcriptional start site (TSS), 1 kb downstream of the TSS, or up to 1000 kb in either direction for distal regions. Pathway enrichment was performed on uniquely annotated genes using the Enrichr web browser tool and the PANTHER database [35–38]. Motif enrichment and transcription factor (TF) footprinting were performed using HOMER (version 4.10) and TOBIAS (version 0.13.3), respectively, to identify TF potentially binding to identified accessible peak clusters [18, 27].

CRISPR KO Screen

A modified CRISPR screen was performed with a select cohort of gene targets selected based on two criteria: 1. genes exhibited significantly increased RNA expression (log₂ fold change greater than 1.5 and p-adjusted value less than 0.05) specifically in one of the *PIK3CA*-mutant cell lines compared to wild type, and 2. genes were annotated to regions that demonstrated significantly increased accessibility in either mutant cell line. Differential accessibility was assessed using DESeq2 and nearest neighbor annotation for these regions was performed using ChIPseeker (version 1.30.3) with the default annotation conditions of ± 3 kb from the TSS [23, 65]. Based on this selection criteria, 312 genes were selected, for which 280 had guides available in the Brunello whole genome single guide RNA (sgRNA) library. The Brunello whole genome sgRNA library was modified for these 280 genes and prepared by the Vanderbilt Functional Genomics core in the lentiCRISPRv2 plasmid background (Full list of guides Table S3) [20, 66].

MCF-10A cells were cultured in the maintenance media conditions at a density of 500,000 cells/well in a 6-well plate. 24 h after seeding, cells were infected with viral supernatant in maintenance media containing 5 µg/mL polybrene. 24 h post-infection cells were placed in selection media containing 1 µg/mL of puromycin and maintained for two weeks. Following two weeks of selection, libraries were prepared and sequenced following the protocol described in Sanjana et al. [66]. Libraries were sequenced by the VANTAGE core and analysis was performed using the maximum likelihood estimation (MLE) algorithm within the MAGeCK software package (version 0.5.9.5) [42, 67].

Deletion of putative AREG Enhancer

Two sgRNAs were designed targeting the locus, chr4:74,435,384–74,435,596, which is annotated to the AREG gene as well as nearby AREG eQTLs identified from the GTEx Portal [68–70] (Table S5). The guide RNAs were purchased as sgRNAs from IDT with custom targeting sequences (full guide sequences in Table S6). Cells were plated in respective maintenance media conditions at a density of 50,000 cells/well in a 12-well plate. Transfection mixtures were prepared and added to each of the cell lines according to the guidelines described in the Lipofectamine CRISPRMAX Transfection Reagent kit (Invitrogen, CMAX00001) using the provided reagents and the Alt-R S.p. HiFi Cas9 Nuclease V3 (IDT, 1081060) in Opti-MEM Reduced Serum Media (ThermoFisher, 31985062). Separate mixtures were prepared containing either a 1:1 mixture of the designed AREG enhancer flanking guides or a control mixture of Alt-R[®] CRISPR-Cas9 Negative Control crRNA #1 (IDT, 1072544) duplexed to Alt-R[®] CRISPR-Cas9 tracrRNA (IDT, 1072532). DNA and RNA were collected from separate experimental replicates 24 h after transfection (at least 4 technical replicates were prepared for each cell line in each treatment condition). Cell counts and viability were measured 72 h following transfection using a Vi-CELL BLU cell viability analyzer (Beckman Coulter).

DNA was collected using the Wizard SV 96 Genomic DNA Purification System (Promega, A2370). PCR was performed using Phusion High-Fidelity PCR Master Mix (NEB, M0531) and the enhancer deletion validation primer set (Table S6). PCR mix and thermocycler conditions were set according to the Phusion Master Mix protocol provided by NEB. PCR products were measured and visualized using a D5000 ScreenTape System (Agilent, 5067) (Fig. S10).

RNA was isolated from cells using the RNeasy Plus Mini Kit (Qiagen, 74,134). RNA was converted to cDNA using the iScript cDNA Synthesis Kit (Bio-rad, 1708890). qPCR was performed using the AREG and ACTB qPCR primer sets (Table S6) for each sample with the SYBR

Green PCR Master Mix (Applied Biosystems, 4309155). Expression of AREG was calculated relative to the expression of housekeeping gene ACTB.

Anti-AREG antibody assay

Cells were plated in their respective maintenance media conditions at a density of 50,000 cells/well in 12-well plates. 24 h following seeding, cells were treated with 1, 3, or 5 µg of AREG neutralizing antibody (R&D Systems, MAB262-SP) or an equivalent volume of PBS. This experiment was performed with three replicates for each cell line and each treatment condition. Cells were counted and viability was measured 72 h following treatment using a Vi-CELL BLU cell viability analyzer (Beckman Coulter).

siRNA assay

Cells were plated in their respective maintenance media at a density of 50,000 cells/well in a 12-well plate. 24 h following seeding, cells were treated with three different commercially validated AREG targeting siRNA (Ambion, see oligo sequences on Table S6), a negative control siRNA (Invitrogen, 4390843), or a null transfection condition using the Lipofectamine RNAiMAX Transfection Reagent (Invitrogen, 13778100) at a concentration of 10 pmol siRNA per well. 24 h post-transfection, RNA was prepared from cells using the Qiagen RNeasy kit. Four replicates were prepared from each cell line and each treatment condition. RNA was converted to cDNA using the iScript cDNA Synthesis Kit (Bio-rad, 1708890). qPCR was performed using the AREG and ACTB qPCR primer sets (Table S6) for each sample with the SYBR Green PCR Master Mix (Applied Biosystems, 4309155). Expression of AREG was calculated relative to the expression of housekeeping gene ACTB.

To assess the impact on survival and proliferation, cells were plated in their respective maintenance media at a density of 30,000 cells/well in a 24-well plate. Cells were treated with one of three different commercially validated AREG targeting siRNA (Ambion, see oligo sequences on Table S6) or a negative control siRNA (Invitrogen, 4390843) using the Lipofectamine RNAiMAX Transfection Reagent (Invitrogen, 13778100) at a concentration of 5 pmol siRNA per well. Cell counts and viability were measured 24 h following treatment using a Vi-CELL BLU cell viability analyzer (Beckman Coulter). This experiment was performed with four replicates in each cell line and each treatment condition.

Visualization and figure creation

Images and figures were generated using ggplot2 (version 3.4.1), plotgardener (1.4.1), deeptools (3.5.1), pheamap (1.0.12), and graphpad Prism (Version 10) [71–74]. Schematic images and flow chart were created using Biorender.com.

Abbreviations

PIK3CA	Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha
PI3K	Phosphatidylinositol-4,5-Bisphosphate 3-Kinase
ATAC-seq	Assay for transposase-accessible chromatin with sequencing
CRISPR	Clustered Regularly Interspaced Palindromic Repeats
KO	Knock out
AREG	Amphiregulin
DEGs	Differentially expressed genes
ChrAcc	Chromatin accessibility
TF	Transcription factor
eQTL	Expression quantitative trait locus

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-024-10368-1>.

Supplementary Material 1.
Supplementary Material 2.
Supplementary Material 3.
Supplementary Material 4.
Supplementary Material 5.
Supplementary Material 6.
Supplementary Material 7.
Supplementary Material 8.

Acknowledgements

The results shown here are in part based upon data generated by the TCGA Research Network: <https://www.cancer.gov/tcga>. Schematic images created with biorender.com.

The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The data used for the analyses described in this manuscript were obtained from: the GTEx Portal on 11/02/2022.

Authors' contributions

AXM, BHP and EH conceived of the project. AXM cultured the cells and performed the data analysis for the RNA-seq, ATAC-seq and CRISPR KO screen experiments. AXM performed the enhancer deletion and siRNA experiments. JK performed the anti-AREG antibody experiments. BD and AM assisted with cell culture. BD and VA contributed to experimental design. SEB and CM provided data and analysis of METABRIC RNA-seq samples. AXM, SC, BHP and EH wrote and edited the manuscript and figures. All authors reviewed and approved the final manuscript.

Funding

We are grateful for support of the project and the time invested in producing this manuscript by NIH awards [1R01GM147078-01 to E.H.], Department of Defense Idea Award [W81XWH-20-1-0522 to E.H.], American Cancer Society (ACS) Institutional Research Grant (#IRG-15-169-56), the Vanderbilt Stanley Cohen Innovation Fund and funds from the Vanderbilt Ingram Cancer Center. We would also like to acknowledge support from the Breast Cancer Research Foundation, the Susan G. Komen Foundation, and the NIH CA214494, CA194024 (B.H.P.). We would also like to thank and acknowledge the support of The Canney Foundation, the Sage Patient Advocates, the Marcie and Ellen Foundation, The Eddie and Sandy Garcia Foundation, the support of Amy and Barry Baker, the support of John and Donna Hall, and the Vanderbilt-Ingram Cancer Center support grant (NIH CA068485) and Breast Cancer SPORE (NIH CA098131).

Availability of data and materials

Raw sequencing datasets are available through the Gene Expression Omnibus (GEO) with the accession number: GSE247822. Detailed workflows and all code used in data analysis can be found at: (<https://github.com/adamxmira/da/PIK3CA>).

Declarations**Ethics approval and consent to participate**

Not applicable.

Consent for publication

Not applicable.

Competing interests

C.M. reports personal consultancy fees from Bayer, Roche, AstraZeneca, Novartis, outside the scope of the present work. B.H.P. is a paid consultant for Guardant Health, AstraZeneca, Caris and is a paid scientific advisory board member for Celcuity Inc. B.H.P. is an unpaid consultant for Tempus Inc. B.H.P. also receives research funding from Eli Lilly and Guardant Health. Under separate licensing agreements between Horizon Discovery, LTD and The Johns Hopkins University, B.H.P. and S.E.C. are entitled to a share of royalties received by the University on sales of products. The terms of this arrangement are being managed by the Johns Hopkins University in accordance with its conflict-of-interest policies. All other authors declare no competing interests.

Author details

¹Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA. ²Department of Biochemistry, Vanderbilt University School of Medicine, Nashville, TN, USA. ³Candiolo Cancer Institute, FPO-IRCCS, Candiolo, TO, Italy. ⁴Department of Medical Sciences, University of Turin, Turin, Italy. ⁵Vanderbilt-Ingram Cancer Center, Nashville, TN, USA. ⁶Vanderbilt Genetics Institute, Vanderbilt University School of Medicine, Nashville, TN, USA.

Received: 21 December 2023 Accepted: 2 May 2024

Published online: 27 May 2024

References

1. Waarts MR, Stonestrom AJ, Park YC, Levine RL. Targeting mutations in cancer. *J Clin Invest*. 2022;132(8):e154943.
2. Alaklabi S, Roy AM, Attwood K, George A, O'Connor T, Early A, Levine EG, Gandhi S. Real world outcomes with alpelisib in metastatic hormone receptor-positive breast cancer patients: A single institution experience. *Front Oncol*. 2022;12:1012391.
3. Lei JT, Gou X, Seker S, Ellis MJ. ESR1 alterations and metastasis in estrogen receptor positive breast cancer. *J Cancer Metastasis Treat*. 2019;5:38.
4. Yang J, Nie J, Ma X, Wei Y, Peng Y, Wei X. Targeting PI3K in cancer: mechanisms and advances in clinical trials. *BioMed Central*. 2019;18:1–28.
5. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov*. 2012;2(5):401–4.
6. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal*. 2013;6(269):p11.
7. Martinez-Saez O, Chic N, Pascual T, Adamo B, Vidal M, Gonzalez-Farre B, Sanfeliu E, Schettini F, Conte B, Braso-Maristany F, et al. Frequency and spectrum of PIK3CA somatic mutations in breast cancer. *Breast Cancer Res*. 2020;22(1):45.
8. Garay JP, Smith R, Devlin K, Hollern DP, Liby T, Liu M, Boddapati S, Watson SS, Esch A, Zheng T, et al. Sensitivity to targeted therapy differs between HER2-amplified breast cancer cells harboring kinase and helical domain mutations in PIK3CA. *Breast Cancer Res*. 2021;23(1):81.
9. Hart JR, Zhang Y, Liao L, Ueno L, Du L, Jonkers M, Yates JR 3rd, Vogt PK. The butterfly effect in cancer: a single base mutation can remodel the cell. *Proc Natl Acad Sci*. 2015;112(4):1131–6.
10. Janku F, Wheler JJ, Naing A, Falchook GS, Hong DS, Stepanek VM, Fu S, Piha-Paul SA, Lee JJ, Luthra R, et al. PIK3CA mutation H1047R is associated with response to PI3K/AKT/mTOR signaling pathway inhibitors in early-phase clinical trials. *Cancer Res*. 2013;73(1):276–84.
11. Jacobson A. Alpelisib Plus Fulvestrant or Letrozole Demonstrates Sustained Benefits Across Subgroups of Patients with PIK3CA-Mutated HR+/HER2- Advanced Breast Cancer. *Oncologist*. 2022;27(Suppl 1):S13–4.

12. Miller MS, Maheshwari S, McRobb FM, Kinzler KW, Amzel LM, Vogelstein B, Gabelli SB. Identification of allosteric binding sites for PI3Kalpha oncogenic mutant specific inhibitor design. *Bioorg Med Chem*. 2017;25(4):1481–6.
13. Andre F, Ciruelos E, Rubovszky G, Campone M, Loibl S, Rugo HS, Iwata H, Conte P, Mayer IA, Kaufman B, et al. Alpelisib for PIK3CA-Mutated, Hormone Receptor-Positive Advanced Breast Cancer. *N Engl J Med*. 2019;380(20):1929–40.
14. Andre F, Ciruelos EM, Juric D, Loibl S, Campone M, Mayer IA, Rubovszky G, Yamashita T, Kaufman B, Lu YS, et al. Alpelisib plus fulvestrant for PIK3CA-mutated, hormone receptor-positive, human epidermal growth factor receptor-2-negative advanced breast cancer: final overall survival results from SOLAR-1. *Ann Oncol*. 2021;32(2):208–17.
15. Gustin JP, Karakas B, Weiss MB, Abukhdeir AM, Lauring J, Garay JP, Cosgrove D, Tamaki A, Konishi H, Konishi Y, et al. Knockin of mutant PIK3CA activates multiple oncogenic pathways. *Proc Natl Acad Sci U S A*. 2009;106(8):2835–40.
16. Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, Szczesniak MW, Gaffney DJ, Elo LL, Zhang X, Mortazavi A. A survey of best practices for RNA-seq data analysis. *Genome Biol*. 2016;17:13.
17. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods*. 2013;10(12):1213–8.
18. Bentsen M, Goymann P, Schultheis H, Klee K, Petrova A, Wiegandt R, Fust A, Preussner J, Kuenne C, Braun T, et al. ATAC-seq footprinting unravels kinetics of transcription factor binding during zygotic genome activation. *Nat Commun*. 2020;11(1):4267.
19. Galas DJ, Schmitz A. DNase footprinting: a simple method for the detection of protein-DNA binding specificity. *Nucleic Acids Res*. 1978;5(9):3157–70.
20. Doench JG, Fusi N, Sullender M, Hegde M, Vaimberg EW, Donovan KF, Smith I, Tothova Z, Wilen C, Orchard R, et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol*. 2016;34(2):184–91.
21. Hartenian E, Doench JG. Genetic screens and functional genomics using CRISPR/Cas9 technology. *FEBS J*. 2015;282(8):1383–93.
22. Boutros M, Ahringer J. The art and design of genetic screens: RNA interference. *Nat Rev Genet*. 2008;9(7):554–66.
23. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550.
24. Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst*. 2015;1(6):417–25.
25. Panigrahi A, O'Malley BW. Mechanisms of enhancer action: the known and the unknown. *Genome Biol*. 2021;22(1):108.
26. Castro-Mondragon JA, Riudavets-Puig R, Rauluseviciute I, Lemma RB, Turchi L, Blanc-Mathieu R, Lucas J, Boddie P, Khan A, Manosalva Perez N, et al. JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. *Nucleic Acids Res*. 2022;50(D1):D165–73.
27. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*. 2010;38(4):576–89.
28. Borreguero-Munoz N, Fletcher GC, Aguilar-Aragon M, Elbediwy A, Vincent-Mistiaen ZI, Thompson BJ. The Hippo pathway integrates PI3K-Akt signals with mechanical and polarity cues to control tissue growth. *PLoS Biol*. 2019;17(10): e3000509.
29. Chi M, Liu J, Mei C, Shi Y, Liu N, Jiang X, Liu C, Xue N, Hong H, Xie J, et al. TEAD4 functions as a prognostic biomarker and triggers EMT via PI3K/AKT pathway in bladder cancer. *J Exp Clin Cancer Res*. 2022;41(1):175.
30. Vierbuchen T, Ling E, Cowley CJ, Couch CH, Wang X, Harmin DA, Roberts CWM, Greenberg ME. AP-1 Transcription Factors and the BAF Complex Mediate Signal-Dependent Enhancer Selection. *Mol Cell*. 2017;68(6):1067–82 e1012.
31. Shaulian E, Karin M. AP-1 in cell proliferation and survival. *Oncogene*. 2001;20(19):2390–400.
32. Whitmarsh AJ, Davis RJ. Transcription factor AP-1 regulation by mitogen-activated protein kinase signal transduction pathways. *J Mol Med (Berl)*. 1996;74(10):589–607.
33. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, Bejerano G. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol*. 2010;28(5):495–501.
34. Tanigawa Y, Dyer ES, Bejerano G. WhichTF is functionally important in your open chromatin data? *PLoS Comput Biol*. 2022;18(8): e1010378.
35. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, Clark NR, Ma'ayan A. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*. 2013;14:128.
36. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res*. 2016;44(W1):W90–97.
37. Mi H, Muruganujan A, Thomas PD. PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Res*. 2013;41 (Database issue):377–86.
38. Thomas PD, Ebert D, Muruganujan A, Mushayama T, Albu LP, Mi H. PANTHER: Making genome-scale phylogenetics accessible to all. *Protein Sci*. 2022;31(1):8–22.
39. Zhou B, Lin W, Long Y, Yang Y, Zhang H, Wu K, Chu Q. Notch signaling pathway: architecture, disease, and therapeutics. *Signal Transduct Target Ther*. 2022;7(1):95.
40. Zhan T, Rindtorff N, Boutros M. Wnt signaling in cancer. *Oncogene*. 2017;36(11):1461–73.
41. Sanson KR, Hanna RE, Hegde M, Donovan KF, Strand C, Sullender ME, Vaimberg EW, Goodale A, Root DE, Piccioni F, Doench JG. Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities. *Nat Commun*. 2018;9(1):5416.
42. Li W, Xu H, Xiao T, Cong L, Love MI, Zhang F, Irizarry RA, Liu JS, Brown M, Liu XS. MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol*. 2014;15(12):554.
43. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*. 2012;486(7403):346–52.
44. Tsherniak A, Vazquez F, Montgomery PG, Weir BA, Kryukov G, Cowley GS, Gill S, Harrington WF, Pantel S, Krill-Burger JM, et al. Defining a Cancer Dependency Map. *Cell*. 2017;170(3):564–76 e516.
45. Li X, Luo OJ, Wang P, Zheng M, Wang D, Piecuch E, Zhu JJ, Tian SZ, Tang Z, Li G, Ruan Y. Long-read ChIA-PET for base-pair-resolution mapping of haplotype-specific chromatin interactions. *Nat Protoc*. 2017;12(5):899–915.
46. Tang Z, Luo OJ, Li X, Zheng M, Zhu JJ, Szalaj P, Trzaskoma P, Magalska A, Wlodarczyk J, Ruszczycycki B, et al. CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. *Cell*. 2015;163(7):1611–27.
47. Zhang H, Ahearn TU, Lecarpentier J, Barnes D, Beesley J, Qi G, Jiang X, O'Mara TA, Zhao N, Bolla MK, et al. Genome-wide association study identifies 32 novel breast cancer susceptibility loci from overall and subtype-specific analyses. *Nat Genet*. 2020;52(6):572–81.
48. Beaver JA, Gustin JP, Yi KH, Rajpurohit A, Thomas M, Gilbert SF, Rosen DM, Ho Park B, Lauring J. PIK3CA and AKT1 mutations have distinct effects on sensitivity to targeted pathway inhibitors in an isogenic luminal breast cancer model system. *Clin Cancer Res*. 2013;19(19):5413–22.
49. Zhou M, Zou X, Cheng K, Zhong S, Su Y, Wu T, Tao Y, Cong L, Yan B, Jiang Y. The role of cell-penetrating peptides in potential anti-cancer therapy. *Clin Transl Med*. 2022;12(5): e822.
50. Alves AC, Ribeiro D, Nunes C, Reis S. Biophysics in cancer: The relevance of drug-membrane interaction studies. *Biochim Biophys Acta*. 2016;1858(9):2231–44.
51. Liu J, Kang R, Tang D. The KRAS-G12C inhibitor: activity and resistance. *Cancer Gene Ther*. 2022;29(7):875–8.
52. Wu X, Renuse S, Sahasrabudhe NA, Zahari MS, Chaerkady R, Kim MS, Nirujogi RS, Mohseni M, Kumar P, Raju R, et al. Activation of diverse signalling pathways by oncogenic PIK3CA mutations. *Nat Commun*. 2014;5:4961.
53. Young CD, Zimmerman LJ, Hoshino D, Formisano L, Hanker AB, Gatzka ML, Morrison MM, Moore PD, Whitwell CA, Dave B, et al. Activating PIK3CA Mutations Induce an Epidermal Growth Factor Receptor (EGFR)/Extracellular Signal-regulated Kinase (ERK) Paracrine Signaling Axis in Basal-like Breast Cancer. *Mol Cell Proteomics*. 2015;14(7):1959–76.

54. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet*. 2011;17:10–2.
55. Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010.
56. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15–21.
57. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H. Twelve years of SAMtools and BCFtools. *Gigascience*. 2021;10(2):1–4.
58. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;30(7):923–30.
59. Frankish A, Diekhans M, Ferreira AM, Johnson R, Jungreis I, Loveland J, Mudge JM, Sisu C, Wright J, Armstrong J, et al. GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res*. 2019;47(D1):D766–73.
60. Dalton WB, Helmenstine E, Walsh N, Gondek LP, Kelkar DS, Read A, Natrajan R, Christenson ES, Roman B, Das S, et al. Hotspot SF3B1 mutations induce metabolic reprogramming and vulnerability to serine deprivation. *J Clin Invest*. 2019;129(11):4708–23.
61. Zabransky DJ, Yankaskas CL, Cochran RL, Wong HY, Croessmann S, Chu D, Kavuri SM, Red Brewer M, Rosen DM, Dalton WB, et al. HER2 missense mutations have distinct effects on oncogenic signaling and migration. *Proc Natl Acad Sci U S A*. 2015;112(45):E6205–6214.
62. Korotkevich G, Sukhov V, Budin N, Shpak B, Artyomov MN, Sergushichev A: Fast gene set enrichment analysis. *BioRxiv*. 2021
63. Barnett KR, Decato BE, Scott TJ, Hansen TJ, Chen B, Attalla J, Smith AD, Hodges E. ATAC-Me Captures Prolonged DNA Methylation of Dynamic Chromatin Accessibility Loci during Cell Fate Transitions. *Mol Cell*. 2020;77(6):1350–64 e1356.
64. Bushnell B, Rood J, Singer E. BBMerge - Accurate paired shotgun read merging via overlap. *PLoS ONE*. 2017;12(10): e0185056.
65. Wang Q, Li M, Wu T, Zhan L, Li L, Chen M, Xie W, Xie Z, Hu E, Xu S, Yu G. Exploring Epigenomic Datasets by ChIPseeker. *Curr Protoc*. 2022;2(10): e585.
66. Sanjana NE, Shalem O, Zhang F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat Methods*. 2014;11(8):783–4.
67. Li W, Koster J, Xu H, Chen CH, Xiao T, Liu JS, Brown M, Liu XS. Quality control, modeling, and visualization of CRISPR screens with MAGeCK-VISPR. *Genome Biol*. 2015;16:281.
68. Carithers LJ, Ardlie K, Barcus M, Branton PA, Britton A, Buia SA, Compton CC, DeLuca DS, Peter-Demchok J, Gelfand ET, et al. A Novel Approach to High-Quality Postmortem Tissue Procurement: The GTEx Project. *Biopreserv Biobank*. 2015;13(5):311–9.
69. Consortium G. The Genotype-Tissue Expression (GTEx) project. *Nat Genet*. 2013;45(6):580–5.
70. Consortium G. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science*. 2020;369(6509):1318–30.
71. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. Verlag New York: Springer; 2016.
72. Kramer NE, Davis ES, Wenger CD, Deoudes EM, Parker SM, Love MI, Phanstiel DH. Plotgardener: cultivating precise multi-panel figures in R. *Bioinformatics*. 2022;38(7):2042–5.
73. Ramirez F, Ryan DP, Gruning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dundar F, Manke T. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res*. 2016;44(W1):W160–165.
74. Kolde R. Pheatmap: pretty heatmaps. R package version. 2012;1(2):726.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.