

CRISPR screens identify gene targets and drug repositioning opportunities at breast cancer risk loci

Joseph Rosenbluh (✉ sefi.rosenbluh@monash.edu)

Monash University <https://orcid.org/0000-0001-9815-8049>

Natasha Tuano

Monash University

Jonathan Beesley

QIMR Berghofer Medical Research Institute

Murray Manning

Monash University

Wei Shi

QIMR Berghofer Medical Research Institute

Luis Malaver-Ortega

Monash University

Jacob Paynter

Monash University

Debra Black

QIMR Berghofer Medical Research Institute

Andrew Civitaresse

QIMR Berghofer Medical Research Institute

Karen McCue

QIMR Berghofer Medical Research Institute

Aaron Hatzipantelis

QIMR Berghofer Medical Research Institute

Kristine Hillman

QIMR Berghofer Medical Research Institute <https://orcid.org/0000-0003-3039-4437>

Susanne Kaufmann

QIMR Berghofer Medical Research Institute

Haran Sivakumaran

QIMR Berghofer Medical Research Institute <https://orcid.org/0000-0002-4691-4989>

Jose Polo

Monash University <https://orcid.org/0000-0002-2531-778X>

Roger Reddel

Children's Medical Research Institute <https://orcid.org/0000-0002-6302-6107>

Vimla Band

University of Nebraska Medical Center

Juliet French

QIMR Berghofer Medical Research Institute <https://orcid.org/0000-0002-9770-0198>

Stacey Edwards

QIMR Berghofer Medical Research Institute <https://orcid.org/0000-0001-7428-4139>

David Powell

Monash University <https://orcid.org/0000-0002-6639-6250>

Georgia Chenevix-Trench

Article

Keywords:

Posted Date: January 7th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1223972/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Title: CRISPR screens identify gene targets and drug repositioning opportunities at breast cancer risk loci

One-Sentence Summary: CRISPR activation and suppression screens identify gene targets of non-coding variants at loci associated with breast cancer risk.

Authors: Natasha K Tuano¹, Jonathan Beesley², Murray Manning^{1,7}, Wei Shi², Luis F Malaver-Ortega⁷, Jacob Paynter^{1,3}, Debra Black², Andrew Civitarese², Karen McCue², Aaron Hatzipantelis², Kristine Hillman², Susanne Kaufmann², Haran Sivakumaran², Jose M Polo³, Roger Reddel⁴, Vimla Band⁵, Juliet D French², Stacey L Edwards², David Powell⁶, Georgia Chenevix-Trench^{2,*}, Joseph Rosenbluh^{1,7,*}

Affiliations: ¹Cancer Research Program and Department of Biochemistry and Molecular Biology, Biomedicine Discovery Institute, Monash University, Clayton, Victoria, Australia

²Cancer Program, QIMR Berghofer Medical Research Institute, Brisbane, Australia

³Development and Stem Cells Program, Monash Biomedicine Discovery Institute, Clayton, Victoria, Australia.

⁴Cancer Research Unit, Children's Medical Research Institute, Faculty of Medicine and Health, University of Sydney, Westmead, NSW, Australia

⁵Department of Genetics, Cell Biology and Anatomy, University of Nebraska Medical Center, Omaha, Nebraska, USA.

⁶Bioinformatics Platform, Monash University, Clayton, VIC, Australia.

⁷Functional Genomics Platform, Monash University, Clayton, VIC, Australia.

*Correspondence: Georgia.Trench@gimrberghofer.edu.au (G.C.T),
sefi.rosenbluh@monash.edu (J.R).

Abstract:

Genome-wide association studies (GWAS) have identified >200 loci associated with breast cancer (BC) risk. The majority of candidate causal variants (CCVs) are in non-coding regions and likely modulate cancer risk by regulating gene expression. However, pinpointing the exact target of the association and identifying the phenotype it mediates is a major challenge in the interpretation and translation of GWAS. Here, we used pooled CRISPR activation and suppression screens to evaluate predicted GWAS target genes, and to define the cancer phenotypes they mediate. We measured proliferation in 2D, 3D, and in immune-deficient mice, as well as the effect on DNA repair. We performed 60 CRISPR screens and identified 21 genes predicted with high confidence to be GWAS targets that drive a cancer phenotype by driving a proliferation or DNA damage response in breast cells. We validated the regulation of a subset of these genes by BC-risk variants, and show the utility of expression profiling for drug repurposing. We provide a platform for identifying gene targets of risk variants, and present a blueprint of interventions for BC risk reduction and treatment.

Introduction

Genetic evidence that implicates a gene in disease etiology is a strong indicator that drugs targeting the encoded protein will be effective as therapies or for risk reduction^{1,2}. Indeed, one of the most commonly used drugs for primary and secondary prevention of BC is tamoxifen, an antagonist of the estrogen receptor (ER) which is encoded by *ESR1*, one of the known targets of a BC risk locus¹. GWAS have identified >200 loci associated with BC risk, most of which are associated with both ER+ and ER- BC². These loci represent a valuable source for identifying drug targets³⁻⁵, but translation of these findings to actionable mechanisms requires first identifying the target gene of the association. However, since most CCVs are in non-coding regions (Fig. 1A), and are presumed to act through regulatory mechanisms, identification of the target gene is challenging.

We recently developed a heuristic scoring system called INQUISIT (Integrated expression quantitative trait and in silico prediction of GWAS targets) to rank the predicted target genes at BC risk loci^{3,4}. INQUISIT treats any CCV as potentially able to regulate distal genes and awards points to each gene based on: 1) chromatin interaction data from Capture Hi-C and ChIA-PET experiments; 2) enhancer annotations based on computational methods designed to infer target genes from genomic data; 3) expression quantitative trait loci (eQTL) analysis of genes within 2 Mb of either side of each CCV when the risk and eQTL signals co-localize; 4) integration of transcription factor ChIP-seq data for specific proteins in breast cells shown to be positive predictors of BC CCVs. The intersection of CCVs, enhancers and these transcription factor binding sites resulted in up-weighting of the associated gene (see example of target gene rankings by INQUISIT at a BC risk locus in Extended Data Fig. 1). Promoter variants were assessed for overlap with chromatin signatures characteristic of transcription start sites (TSS) in breast cell lines and primary tissue, as well as putative functional transcription factor binding sites, gene expression data and eQTLs. Intragenic variants were evaluated for consequences of coding and splicing changes. We designated INQUISIT predictions with the strongest supporting evidence as Level 1, and Level 3 as the lowest. At 205 fine-mapped risk signals (having omitted one with >2,000 CCVs), INQUISIT identifies 1-10 Level 1 targets per signal at 114 signals (184 genes). For 76 signals, INQUISIT predicts 678 Level 2 genes and for 15 signals INQUISIT does not predict any gene targets (Supplementary Table 1).

Computational algorithms can only predict candidate genes that are associated with a particular locus. Pinpointing the gene target and identifying the phenotypes it mediates, requires experimental data, which is a daunting task for so many BC risk loci. High-throughput chromatin interaction capture methods such as HiChIP have been instrumental in identifying genes that are potentially regulated by distal elements^{6,7}. However, these experiments can be difficult to interpret due to complex 3D organization at gene-dense regions, and because at many loci there are a large number of CCVs over a large genomic interval. Furthermore, 3D chromatin structure may not be consistent

between the relevant primary tissues and cell lines which are typically used in these experiments⁸⁻¹⁰. A complementary approach is to use a phenotypic readout to identify putative GWAS target genes which mediate a cancer phenotype. We hypothesize that genes implicated by GWAS, with strong *in silico* supporting evidence, will influence a quantifiable cancer phenotype which will enable us to nominate the most likely BC risk genes. Pooled CRISPR screens are extensively used to identify genes related to a particular phenotype^{11,12} but have not been used to characterize GWAS target genes. Here, we used large-scale pooled CRISPR activation and suppression screens to identify genes which mediate proliferation *in vitro*, tumor formation *in vivo* and DNA damage response, in order to define gene targets at BC risk loci.

Results

Selection of genes for functional CRISPR screens. We selected genes using the following approaches: 1) 184 high-confidence INQUISIT Level 1 (INQ_1) target genes³, 2) 678 INQUISIT Level 2 (INQ_2) target genes, 3) 371 genes identified by Transcriptome Wide Association Studies (TWAS) and expression quantitative trait loci (eQTL) studies of BC-risk (referred to as TWAS genes)¹³⁻¹⁶, 4) 605 'background' genes, including low confidence INQUISIT targets and genes predicted only in an early INQUISIT version, and genes within 2Mb of 15 risk signals at which INQUISIT did not predict any targets. Genes predicted by both INQUISIT and TWAS were categorized as INQ_1 for Level 1 predictions, and, as TWAS for Level 2 predictions (Fig. 1B and Supplementary Table 1). For each of these genes we designed five single guide (sg)RNAs. In addition, we included 1,000 negative controls and 960 sgRNAs targeting 193 core essential genes, as well as 16 known tumor-suppressor genes and oncogenes (Supplementary Table 2).

BC risk genes that induce a proliferation phenotype in 2D and 3D cultures. Impaired proliferation is a hallmark of cancer¹⁷. We used CRISPR screens, in six immortalized mammary epithelial cell lines, to suppress (CRISPRko or CRISPRi) or overexpress (CRISPRa) candidate BC-risk genes and identify putative tumor-suppressors and oncogenes (Fig. 1C). Expression and ATAC-Seq profiling indicated that these six cell lines represent breast cells with either a luminal progenitor signature (K5+/K19+¹⁸, K5+/K19-¹⁸), a mesenchymal signature (B80-T17¹⁹, mesHMLE²⁰) or a more epithelial like signature (B80-T5¹⁹, HMLE²¹) (Extended Data Figs. 2A-C).

Following library infection, cells were plated in 2D or 3D conditions and sgRNA abundance was quantified after 21 days (Supplementary Table 3). Negative controls had no proliferation effect and as expected, suppression of core-essential genes had a negative impact in CRISPRko and CRISPRi screens but no effect in CRISPRa screens (Extended Data Fig. 3A). Known tumor-suppressors had a positive proliferation impact in

CRISPRko and CRISPRi screens and known oncogenes increased proliferation in CRISPRa screens (Extended Data Fig. 3A).

We set the threshold for functional genes with a magnitude of effect $\text{Log}_2[\text{Fold Change}] > 1$ and a significance $-\text{Log}_{10}[\text{FDR}] > 1$ in at least one cell line. Oncogenes were defined as genes that upon overexpression increased proliferation in 2D or 3D cultures. Tumor-suppressors were defined as genes that increase proliferation upon suppression in 2D or 3D cultures or genes that inhibit proliferation in 2D cultures ($\text{Log}_2[\text{Fold Change}] < -1$) upon overexpression. Importantly, we only used overexpression to further support a gene as a tumor-suppressor and not as a stand-alone criterion (Fig. 1D and Extended Data Figs. 3B-G).

We identified 41 candidate BC-risk genes that mediate a proliferation phenotype in 2D or 3D cultures (Fig. 1D and Extended Data Figs. 3B-G). Our results demonstrate the utility of using multiple assays, cell lines and perturbation methods. We found high consistency between CRISPRi and CRISPRko screens (Extended Data Fig. 4A). The inconsistencies detected are likely due to CRISPRi bidirectional promoter off-target effects^{22,23}. For example, *ATXN7* scored as a strong tumor-suppressor using CRISPRko in 2D and 3D assays but did not score with CRISPRi (Extended Data Fig. 4B). This is because *ATXN7* and *THOC7* share a bidirectional promoter (Extended Data Fig. 4C) and *THOC7* is a common cell essential gene (Extended Data Figs. 4D, E). Thus, CRISPRi sgRNAs targeting the *ATXN7* promoter also inhibit *THOC7*, resulting in cell death. We found high correlation between 2D and 3D proliferation changes (Extended Data Figs. 5A-C). Interestingly, some genes showed the opposite effect in 2D and 3D cultures suggesting a function in mediating cell motility. For example, *CFL1* scored as a potent tumor-suppressor in 3D cultures but had no effect in 2D cultures (Extended Data Fig. 5D). This is consistent with the known function of *CFL1* as a regulator of actin filament polymerization and cell motility^{24,25}.

Validation of 2D and 3D proliferation hits. To validate these observations in a singleton experiment, we infected all six cell lines with individual sgRNAs targeting INQUISIT Level 1 hits that scored in the above screens. Using western blot analysis, we confirmed that these sgRNAs reduced (for tumor-suppressor genes) or increased (for oncogenes) expression of the target protein (Extended Data Fig. 6A, B). Following infection, cells were plated on 2D (Extended Data Figs. 7A, B) or 3D (Extended Data Figs. 7C, D) conditions and proliferation was measured by monitoring cell growth. Consistent with our screening results we were able to validate the 2D and 3D proliferation effects in at least one cell line. For *CREBBP* and *CFL1* we found a cell line specific effect (Extended Data Figs. 7A, B). This is consistent with reports showing that *CREBBP* can act as a tumor-suppressor or an oncogene in a cell type specific manner^{26,27}.

In summary, our approach identified 41 candidate BC risk genes that mediate a proliferation phenotype in 2D or 3D cultures, including 15 that were predicted with high confidence by INQUISIT. These 15 include well annotated tumor-suppressor genes and oncogenes (e.g. *TGFBR2* and *MYC* as well as genes that have never been previously linked to cancer in general, or to increased BC risk (including *ADCY3*, *ATXN7*, *CFL1* and *LPAR2*). Together, these results demonstrate the ability of systematic CRISPR screens to define genes associated with BC risk that drive a proliferation phenotype.

Identification of BC-risk genes that promote tumor formation in mice. To identify candidate BC-risk genes that play a role in tumor formation *in vivo*, we used a mouse xenograft model (Fig. 2A). We found that the immortalized mammary cell lines required a constitutively active form of MEK1 (MEKDD) and an additional oncogenic insult in order to support tumor formation (Extended Data Fig. 8A). Following library transduction, cells were injected to the flanks of immune deficient mice. Tumors were harvested 6-8 weeks post injection, and sgRNA abundance was quantified (Extended Data Figs. 8B-D and Supplementary Tables 3, 4). Five positive controls showed dramatically increased sgRNA abundance and ten INQUISIT-predicted genes (five at Level 1) scored in this assay, suggesting these are potent drivers of BC-risk (Fig. 2B). We validated INQUISIT Level 1 CRISPRko hits in B80-T5-MEKDD cells (Extended Data Figs. 8E, F).

We found that *DUSP4*, *CTD-2278I10.4* and *VPS45* scored only *in vivo* and had no effect *in vitro*. *DUSP4*, is an INQUISIT Level 1 hit which we have previously shown to be down-regulated by CCVs at 8p12²⁸. Since our *in vivo* screens used MEKDD expressing cells, we explored whether *DUSP4* is a context-dependent tumor-suppressor. Following *DUSP4* knockout in B80-T5 or B80-T5-MEKDD, we measured proliferation in 3D cultures. Consistent with the *in vivo* screen, *DUSP4* knockout only showed increased proliferation in the presence of MEKDD (Fig. 2C). To investigate *DUSP4*'s mechanism of action, we assessed *DUSP4* levels following MEKDD expression and found that MEKDD induced *DUSP4* expression, suggesting a negative feedback loop (Fig. 2D). However, consistent with previous results²⁹, we did not observe any change in pERK levels following suppression of *DUSP4* (Extended Data Figs. 9A, B). We did observe decreased pJNK and pp38 in *DUSP4* knockouts (Extended Data Figs. 9A, B), confirming previous observations^{30,31} and suggesting that downregulation of pJNK via *DUSP4* mediates its tumor suppressive activities. To further validate these observations, we used selumetinib and trametinib, two potent MEK inhibitors. We found that MEK inhibitors reversed the increased *DUSP4* protein levels (Fig. 2E and Extended Data Fig. 9C) as well as *DUSP4* induced proliferation (Fig. 2F) suggesting MEK inhibitors as a therapeutic strategy in BC with downregulated *DUSP4* expression. This might be particularly relevant in triple negative BC because *DUSP4* is deleted in about 50% BC, most often in this aggressive subtype.

Together, our *in vivo* and *in vitro* proliferation screens identify 44 predicted BC risk genes (including 16 INQUISIT Level 1 genes) that can drive a proliferation phenotype in 2D, 3D cultures or *in vivo* (Fig. 1D and Fig. 2B). We found a strong correlation in phenotypes between the different cell lines (Extended Data Fig. 9D), indicating that even if a particular gene did not pass our threshold it is likely to be a near hit in other cell lines.

Identification of genes that regulate the DNA damage response. DNA damage is a hallmark of cancer in general, and in particular is deregulated in BC³². To identify BC-risk associated genes that regulate the DNA damage response we used a PARP inhibitor synthetic lethality screen³³. Following sgRNA infection, cells are treated with olaparib, a potent PARP1 inhibitor, and only cells harboring a sgRNA that deregulates the homologous recombination DNA damage repair pathway are rendered sensitive to olaparib (Fig. 3A, B). Using this approach, we screened for genes that upon suppression (CRISPRko or CRISPRi) or overexpression (CRISPRa) regulate the DNA repair pathway (Fig. 3C and Extended Data Figs. 10A-F). We identified 28 candidate BC risk genes that regulate the DNA damage response, including 9 that were predicted with high confidence by INQUISIT. As expected, Gene Set Enrichment Analysis (GSEA) of hits (not including positive controls and background genes) showed enrichment for genes involved in the DNA repair pathway and cell cycle (Fig. 3D). This is consistent with the two types of cellular stresses known to be synthetic lethal with mutations in the DNA repair pathway³³. Indeed, four of the INQUISIT Level 1 scoring genes (*MYC*, *NF1*, *NRIP1* and *CREBBP*) also scored in the above described proliferation screens (Fig. 1D). To validate these results in a singleton experiment we used B80-T5 cells and showed that CRISPR mediated suppression of known DNA damage related genes (*ATM*), as well as newly identified synthetic lethal genes (*SIVA1* and *CMTR2*), had a dramatic effect on PARP inhibitor sensitivity (Fig. 3E and Extended Data Fig. 10G). These experiments define a set of 28 likely BC risk genes that drive a DNA repair phenotype, including nine INQUISIT Level 1 genes.

We tested the enrichment of hits from all screens amongst all genes predicted by various computational and statistical methods. We observed a trend for INQUISIT Level 1 genes to be over-represented across most screen modalities (Fig. 3F), suggesting that high-confidence INQUISIT predictions represent probable candidate genes at disease-associated loci. Importantly, the observed enrichment did not appear to be confounded by a gene's status as a known driver (a source of upweighting in the INQUISIT pipeline), since this consistent effect was observed when the genes which scored Level 1 on the basis of being known BC driver genes (*CASZ1*, *CCNE1*, *CREBBP*, and *NF1*) had been downgraded to INQUISIT Level 2. Taken together, these experiments define a set of 21 INQUISIT Level 1 likely BC risk genes, and an additional 46 genes that drive a proliferation or DDR phenotype in breast cells.

HiChIP and CRISPRqtl validate distal regulation between BC risk loci and genes that score in functional screens. Variants identified by GWAS commonly affect tissue-specific distal enhancers. We have previously shown that many CCVs regulate the expression of target genes through chromatin looping^{3,34}. To confirm this in the normal breast cells we used in the current screens, we performed HiChIP on B80-T5 and K5+/K19+ cells. For 18 of the 21 INQUISIT Level 1 hits identified in the screens we found chromatin interactions with regions containing BC risk variants (Fig. 4A and Supplementary Table 5). One of the genes for which we did not identify chromatin interactions was *BRCA2*, which contains a coding CCV. The interactions were particularly strong for *ATF7IP*. The risk signal at this locus comprises 18 SNPs, seven of which lie within a candidate enhancer region marked by open chromatin and H3K27ac histone marks (Fig. 4B). We used luciferase reporter assays to test whether variants within these enhancers altered *ATF7IP* promoter activity. Addition of the *ATF7IP* putative regulatory element (PRE) containing the protective allele to the *ATF7IP* promoter had a 9-fold increase ($p < 0.0001$) in luciferase activity (Fig. 4C). This increase in luciferase activity was reduced by 50% ($p < 0.001$) following introduction of the PRE containing the risk-associated allele. Furthermore, we found that introduction of a variant at rs11055880 (PRE mutant 1) had the same effect as the entire risk associated allele while introduction of rs16909788 and rs17221259 had no effect (PRE mutant 2) on luciferase activity (Fig. 4C). Overall, this effect is consistent with BC risk-associated variation at this locus reducing expression of the putative tumor-suppressor gene, *ATF7IP*. Since luciferase assays require expression of an exogenous construct and may not fully recapitulate the native chromosome structure, we validated these results using a CRISPRi approach. Previous studies showed that sgRNAs targeting enhancers are effective in suppressing the expression of the target gene³⁵. Using four *ATF7IP* enhancer-targeting sgRNAs we found a 50% reduction in *ATF7IP* expression (Fig. 4D), further demonstrating this BC-associated enhancer as a regulator of *ATF7IP* expression.

Based on these results, we performed a systematic CRISPRi enhancer screen using the recently described CRISPRqtl approach³⁶. In CRISPRqtl, a pooled sgRNA library targeting putative enhancers is cloned in a vector that is compatible with single cell RNA-Seq (scRNA-Seq). Following transduction at high multiplicity of Infection (MOI=5), scRNA-Seq is used to detect sgRNA identity and global mRNA abundance. All cells expressing a particular sgRNA are aggregated and the sgRNA effect on expression of genes in cis (2Mbp from the sgRNA) is calculated (Fig. 4E). We generated a CRISPRqtl sgRNA library targeting 53 BC-associated genomic regions. We designed an sgRNA library targeting candidate enhancers identified using ATAC-seq. We segmented each of these regions to 1,000 base-pair blocks and selected 10 sgRNA/block using the CRISPick algorithm³⁷ (Supplementary Table 2). This included six loci that have an INQUISIT Level 1 prediction which scored in the above-described CRISPR proliferation screens, as positive controls. In addition, within these six regions, chromatin interactions were detected between these six genes and CCVs using H3K27ac-mediated HiChIP

(Supplementary Table 5). We included 50 sgRNAs targeting the TSS of 25 genes from³⁶ as additional positive controls, and 50 non-targeting sgRNAs as negative controls (Supplementary Table 3). Following transduction at MOI=5, two lanes of 10x chromium were used to collect single cells, and cDNA generated from these cells was sequenced. The CRISPR application in the Cell Ranger package³⁸ was used for deconvolution and alignment to the human genome. We detected a total of 13,334 cells with a mean of 16,953 reads/cell and a median unique molecular identifier (UMI) of 4,006 UMIs/cell. To reduce non-specific noise due to low level sgRNA detection we filtered out cells that did not have a minimum of 10 sgRNA UMIs/cell. We used the recently described SCEPTRE algorithm to identify the effects of enhancers on gene expression³⁹. SCEPTRE uses conditional resampling and avoids confounding issues associated with high-throughput expression profiling experiments. For each sgRNA we calculated its effect on gene expression (Z-Score) for every gene in a 2Mbp window from the sgRNA (Supplementary Table 6).

As expected, and consistent with previous reports³⁶, TSS targeting sgRNAs had a dramatic effect ($p < 0.0001$) on expression of their target gene (Fig. 4F), demonstrating the reliability of this approach. As predicted target genes, we considered sgRNA-gene pairs that showed a Z-score ≤ -1 (one SD below the mean) with at least three different sgRNAs within a targeted region. Using these criteria, we found 298 genes regulated by 50 fine mapped BC-associated risk signals (Supplementary Table 6). For all six positive control loci, CRISPRqtl, INQUISIT (at Level 1) and CRISPR functional screens identified the same target genes (*ATF7IP*, *ATXN7*, *MYC*, *RANBP9*, *NRIP1* and *TNFSF10*; Fig. 4G). As a CRISPRqtl example, we show sgRNA-gene pairs at the BC-risk locus at chr12:13913931-14913931. Consistent with our functional screen, HiChIP and luciferase assays, CRISPRqtl also found a strong interaction between this locus and *ATF7IP* (Figs. 4H, I), demonstrating the value of using a multi-assay approach to define GWAS targets.

In total, 22 of the 42 risk loci that had an INQUISIT Level 1 prediction showed the same hit by CRISPRqtl (Supplementary Table 6) demonstrating the reliability of this approach. Some of the inconsistencies between CRISPRqtl and INQUISIT are likely due to the cell lines we used. For example, the INQUISIT predicted target genes at chr5:779790-1797488 are *CLPTM1L* and *TERT*. Furthermore, using luciferase assays we have shown that CCVs in this region regulate *TERT* expression⁴⁰. However, since the cell line we used for CRISPRqtl has been engineered to overexpress *TERT*¹⁸ it is difficult to detect subtle enhancer regulation of the poorly expressed endogenous *TERT*. Similarly, INQUISIT identifies *ESR1*, *ARMT1* and *CCDC170* as target genes at chr6:151418856-152937016 and we have shown the effect of CCVs in this region using luciferase assays in ER+ cell lines⁴¹. However, here we used an ER- cell line which exhibits very low basal expression of these genes so altered changes in gene expression may be undetectable.

For nine of the 12 fine mapped regions where INQUISIT did not find any Level 1 candidate genes, CRISPRqtl identified 26 potential targets (Supplementary Table 6). However, for

three regions (chr8:75730301-76917937, chr3:4242276-5242276 and chr22:41538786-42538786) neither CRISPRqtl nor INQUISIT identified targets, suggesting these risk regions may have a different mechanism of action. Overall, these results show that regulation of gene expression through chromatin interactions is the most likely mechanism of action for these risk loci, and demonstrate functional CRISPR screens as a highly reliable strategy for defining targets of GWAS hits.

Survival analysis in cancer patients confirms directionality and cancer association of BC risk genes. To assess the clinical relevance of the hits we identified in our CRISPR screens, we evaluated their association with survival following a BC diagnosis. Of the 20 INQUISIT Level 1 coding genes that scored in our functional screens we found that 14 (70%) genes were significantly (Bonferroni-corrected significance threshold $p < 0.0008$) correlated with relapse-free survival from overall, ER+ or ER- BC (Extended Data Fig. 11). Furthermore, 9 of the 10 genes identified in the proliferation screens showed a directionality that is consistent to that we observed in our functional screens, with high levels of tumor suppressor gene expression associated with better outcome. These observations suggest that using functional CRISPR screens of candidate breast cancer risk genes is a robust method for identifying genes that have clinical significance.

Expression profiling identifies candidate drugs that target BC risk genes. Our work defines 21 INQUISIT Level 1 genes, at 19 BC risk loci, that mediate proliferation in 2D or 3D cultures, tumor growth *in vivo* and/or DNA damage phenotype. Translating these findings, to drugs that could be used for prevention or treatment, requires identifying strategies to target their gene products. Since many of the genes we identified are not known to be associated with cancer, and do not have known inhibitors, we used expression profiling, following CRISPR-mediated gene suppression (CROPSeq⁴²) to identify known drugs that could be repurposed. We transduced K5+/K19+ cells with a pooled CRISPRko sgRNA library containing 200 sgRNAs that target known cancer genes and negative controls, as well as all INQUISIT Level 1 hits (Supplementary Table 2). Although some of these genes scored as oncogenes, in this experiment we used gene suppression in order to find similar or opposing signatures in the cMAP database, as previously described⁴³. Following transduction at MOI=0.1 (ensuring one sgRNA/cell) chromium 10x was used to isolate 15,181 cells (~75 cells/sgrNA). sgRNA enrichment was performed as previously described⁴⁴, and sequencing reads were deconvoluted and aligned to the human genome using Cell Ranger³⁸. We found that cells with a threshold of 10 sgRNA UMIs had mostly a single sgRNA (Extended Data Fig. 12A).

For quality control, we assessed the levels of target gene suppression (Extended Data Fig. 12B). Some target genes did not show good suppression, but this could be attributed to low detection rate. Specifically, only transcripts that were detected with expression

levels >0.1 showed good target suppression (Extended Data Fig. 12C). However, GSEA analysis confirmed that even low expressing genes showed the expected expression signature. For example, sgRNAs targeting *RPTOR*, a known component of the PI3K pathway, induced an MTORC expression signature (Extended Data Fig. 12D), demonstrating the reliability of this dataset.

We normalized expression values in this dataset by comparing gene expression following sgRNA suppression to gene expression in cells expressing a control (non-targeting) sgRNA as previously described⁴⁵. By averaging expression levels of the three sgRNAs targeting a particular gene we calculated a CROPSeq gene score for every gene (Supplementary Table 7). To evaluate the ability of this dataset to find connections between components of known signaling pathways, we used unsupervised clustering and found clusters containing components of known signaling pathways (Fig. 5A). For example, *APC* and *CSNK1A1*, two known negative regulators of the WNT signaling pathway formed a tight cluster and both showed high up regulation of *AXIN2*, a well-known WNT target gene (Extended Data Fig. 12E). Similarly, components of the SWI/SNF complex (*ARID1A*, *SMARCB1*, *SMARCC1* and *SMARCC2*) formed a cluster that also contains *YAP1* and *TAZ* (*WWTR1*), which have recently been shown to regulate the SWI/SNF complex⁴⁶. Interestingly, *CREBBP* at risk locus chr16:3606788-4606788, which scored in 2D, 3D and *in vivo* screens as a tumor-suppressor gene, was strongly connected with the same cluster suggesting that *CREBBP* regulates the SWI/SNF complex. This is consistent with previous reports of *CREBBP* function^{47,48}. Furthermore, in agreement with the results described above, we found that *DUSP4* has an expression signature similar to components of the MAPK signaling pathway. This cluster contains known regulators of the MAPK pathway (*MAPK1*, *PTEN* and *NF1*) and the cell cycle (*CDKN1B*, *CDKN1A* and *CDKN2A*) (Fig. 5A).

We used this dataset to query the cMAP drug repurposing hub⁴³. For each gene we selected the top and bottom 150 expressed genes (using Z-scored expression profiles). We identified compounds with similar and opposing signatures to each of these genes (Supplementary Table 8). Perturbagen classes (PCLs) are a collection of molecules and give a more robust connectivity score⁴⁹. We found known compounds that were correctly connected to a gene knockout. For example, we found that knockout of regulators of the mTOR pathway, *MTOR*, *RPTOR* and *PIK3CA*, was positively connected to the mTOR compound PCL, and that *PTEN* knockout, a negative regulator of the pathway, had a negative connection to the mTOR compound signature (Fig. 5B). By comparing signatures of the novel BC risk genes, we identified drugs that could be potentially used as inhibitors for 12 of them (Figs. 5C-E). To validate these observations, we used K19+/K5+ cells expressing a *CREBBP* sgRNA. Consistent with the above cMAP prediction, knockout of *CREBBP* sensitized cells to vorinostat, a potent HDAC inhibitor (Fig. 5F).

Discussion

GWAS have been highly successful in identifying variants associated with BC risk. Although a major goal of these studies is to identify new strategies for cancer prevention or treatment, a major obstacle in translating these findings to meaningful biological insights is that most risk variants are non-coding and the gene targets of the associations are not clear. Following fine mapping to identify the CCVs for BC, prioritizing loci with relatively few CCVs, chromatin conformation capture (3C) and luciferase assays, have been performed at 16 BC risk loci implicating regulation of *TERT*⁴⁰, *CCND1*⁵⁰, *FGFR2*⁵¹, *IGFBP5*⁵², *MAP3K1*⁵³, *ESR1*, *RMND1* and *CCDC170*⁴¹, *KLF4*⁵⁴, *NRBF2*⁵⁵, *ABHD8*⁵⁶, *FGF10* and *MRPS30*⁵⁷, *KLHDC7A*, *PIDD1*, *CITED4*, *PRKRIP1* and *RASA4*⁴, *DUSP4*²⁸, *NTN4*⁵⁸, *TBX3*³⁴ and novel lncRNAs, *CUPID1* and *CUPID2*⁵⁹.

Identifying GWAS gene targets and evaluating functional mechanisms at all known BC-risk loci individually is challenging. Expression quantitative trait locus (eQTL) links a variant to a transcript and has been suggested as a powerful strategy to identify genes that are regulated by risk variants. Recent studies, however, demonstrate that for various reasons such as context dependent expression and sample size eQTL analysis is limited in its ability to identify target genes of risk variants⁶⁰. Here we show that functional CRISPR screens are a robust alternative for identification of target genes and the phenotypes they mediate. We used pooled CRISPR activation and suppression screens to simultaneously evaluate hundreds of putative GWAS target genes. This identified 21 genes predicted by INQUISIT with high-confidence to be GWAS targets, which mediate a cancer phenotype. Although about half of these are known BC driver genes³, the remainder were not previously implicated in BC biology. We identified an additional 46 genes that drive a proliferation or DNA damage response phenotype in breast cells, but most of these were predicted only with moderate confidence by INQUISIT to be target risk genes. Further studies will be needed to gain insights into their role in BC risk, and the mechanisms and pathways regulated by these genes.

This proof-of-principle experiment demonstrates the utility of functional screens in identifying GWAS targets. Future studies using a similar approach with other cancer-related readouts will likely identify other GWAS hits that regulate different phenotypes. CRISPRqtl identified 22 INQUISIT Level 1 genes that did not score in our functional screen. This includes *KLF4*, which has been previously shown to be regulated by BC risk CCVs⁵⁴, indicating these are high confidence hits. Since our proliferation and DNA repair CRISPR screens did not identify these genes, it is likely that these regulate other cancer phenotypes that we did not measure in the current study.

One of the strengths of our study was that we used four different phenotypic assays. Recent studies have shown the added utility of using 3D cell-based screens which more accurately measure cell proliferation than those carried out in 2D⁶¹. Similar to these studies, we also found that 3D proliferation assays gave a stronger and more robust signal. However, unlike genome wide 2D screens which failed to identify even known

tumor-suppressor genes⁶¹, here we show that in a smaller scale screen we are able to robustly identify known and new tumor-suppressors. This is likely due to the increased sensitivity that we achieved by increasing the number of cells infected with a given sgRNA (1000 cells/sgRNA as opposed to 300-500 cells/sgRNA typically used in whole genome screens¹²). Our results suggest that increasing the number of infected cells and sequencing reads increases sensitivity and enables robust detection of small proliferation changes. This should be considered in genome-wide gain of function CRISPR screens.

Another strength of our study was that we used six immortal mammary cell lines, each with different characteristics. Some genes scored in most cell lines, but the majority of genes only scored in a few. Although most genes did not pass our hit threshold in all cell lines, the phenotypes we observed were highly correlated between different cell lines (Extended Data Fig. 9D). However, some of the differences between cell lines following individual gene validation might be because the activity of the genes is context specific. These observations demonstrate the robustness of the screens and show the importance of using multiple cell lines and multiple assays when measuring the effect of gene perturbation on phenotypes.

Enrichment analysis showed that INQUISIT Level 1 genes were over-represented across all screen modalities in at least some cell lines, compared not only to background genes but also to INQUISIT Level 2 genes, predicted with moderate-confidence, and genes identified by TWAS or eQTL analyses, providing confidence in INQUISIT's ranking of putative target genes. Recently, several other algorithms that predict enhancer targets, including Activity By Contact (ABC) have been described^{4,62,63}. Of the functional genes detected in our screens, only 13 are predicted using the ABC method in breast derived samples. It is worth noting that, of the 21 INQUISIT Level 1 target genes that scored in our CRISPR screens, eight are potentially impacted by CCVs through splicing or coding changes which are not considered by ABC.

INQUISIT did not identify any target genes for 15 of the 205 BC risk signals. For these, we therefore included all genes within a 2Mb window centered on the risk signal (105 genes in total). Of these, only one gene (*JAZF1*) scored in our functional screens, and only with CRISPRi in two cell lines (hit rate of 1/105<1%), consistent with background detection levels. Using CRISPRqtl we identified potential targets for nine of these loci but for three we did not identify any targets. The CCVs at these loci may regulate genes in non-breast cell types, such as immune cells, or need specific stimuli; alternatively, the targets may be unannotated genes or non-coding RNAs. For example, we have recently identified several novel lncRNAs, unannotated in public databases, which are regulated by BC risk variants^{59,64}.

Using CRISPR activation and suppression screens we found 13 genes, predicted by TWAS/eQTL analyses, that induced a proliferation or DNA damage phenotype. We did not validate or further pursue these 13 genes because we did not find any enrichment of hits among this class of genes, and others have shown inconsistencies in the direction of

effect between TWAS findings and known Mendelian genes, including for BC⁶⁵. However, some of them might be genuine BC risk genes.

The most common mechanism for non-coding risk variants is through chromatin interactions to regulate gene expression. Although most studies of enhancer function have utilized chromatin confirmation experiments, two major factors limit the utility of chromatin structure studies: a) the majority of these studies are performed in cultured cells which may not recapitulate the native *in vivo* chromatin structure; b) enhancers frequently interact with multiple genes making functional interpretation challenging. Here, we use a combination of functional CRISPR screens, HiChIP and CRISPRqtl to identify chromatin interactions and phenotypes associated with BC-associated risk loci. Since the chromatin interaction experiments we performed were in cultured cell-lines, we cannot exclude the possibility that the chromatin interactions we found are different *in vivo*. In some cases, chromatin interactions are preserved *in vitro* and *in vivo* and we demonstrate the value of combining HiChIP and CRISPRqtl for identifying enhancer gene targets. HiChIP interacting regions can be large and it can be difficult to pinpoint at the exact region of association. In CRISPRqtl the effective window is relatively small (up to 500bp) and thus combining these approaches is likely to yield better resolution. Furthermore, for all genes that scored in our functional screen we could identify hits in both CRISPRqtl and HiChIP demonstrating the power of phenotypic screens in defining GWAS targets. In this study, we demonstrate the added value of functional phenotypic screens for identifying enhancer targets. Functional screens target the candidate genes rather than the CCV and thus a phenotype could be detected even if the chromatin interactions are not preserved in cultured cells, or if the genes are impacted by coding or splicing variants.

Although further studies into the mechanism of action of the novel BC genes we identified are necessary, we show the utility of high-throughput mRNA profiling in drug repurposing. Using the L1000 database we have identified drugs that are candidate inhibitors for 12 BC risk genes. Expression of most of these 12 genes was associated with relapse-free survival following a diagnosis of BC, suggesting that the inhibitors might be relevant for treatment, as well as potentially for risk reduction.

In summary, we demonstrate that pooled functional CRISPR screening is a cost-efficient, high-throughput and robust method for identifying genes that are associated with BC risk loci. Application and extensions of this approach will be important for harnessing the benefits of cancer GWAS, and for translating genomic findings to treatments.

References

1. King, E.A., Davis, J.W. & Degner, J.F. Are drug targets with genetic support twice as likely to be approved? Revised estimates of the impact of genetic support for drug mechanisms on the probability of drug approval. *PLoS Genet* **15**, e1008489 (2019).

2. Nelson, M.R. *et al.* The support of human genetic evidence for approved drug indications. *Nat Genet* **47**, 856-60 (2015).
3. Fachal, L. *et al.* Fine-mapping of 150 breast cancer risk regions identifies 191 likely target genes. *Nat Genet* **52**, 56-73 (2020).
4. Michailidou, K. *et al.* Association analysis identifies 65 new breast cancer risk loci. *Nature* **551**, 92-94 (2017).
5. Zhang, H. *et al.* Genome-wide association study identifies 32 novel breast cancer susceptibility loci from overall and subtype-specific analyses. *Nat Genet* **52**, 572-581 (2020).
6. Javierre, B.M. *et al.* Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. *Cell* **167**, 1369-1384 e19 (2016).
7. Mumbach, M.R. *et al.* Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements. *Nat Genet* **49**, 1602-1612 (2017).
8. Kribelbauer, J.F. *et al.* Context-Dependent Gene Regulation by Homeodomain Transcription Factor Complexes Revealed by Shape-Readout Deficient Proteins. *Mol Cell* **78**, 152-167 e11 (2020).
9. Schmitt, A.D. *et al.* A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome. *Cell Rep* **17**, 2042-2059 (2016).
10. Zhou, Y. *et al.* Temporal dynamic reorganization of 3D chromatin architecture in hormone-induced breast cancer and endocrine resistance. *Nat Commun* **10**, 1522 (2019).
11. Howard, T.P. *et al.* Functional Genomic Characterization of Cancer Genomes. *Cold Spring Harb Symp Quant Biol* **81**, 237-246 (2016).
12. Sanson, K.R. *et al.* Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities. *Nat Commun* **9**, 5416 (2018).
13. Guo, X. *et al.* A Comprehensive cis-eQTL Analysis Revealed Target Genes in Breast Cancer Susceptibility Loci Identified in Genome-wide Association Studies. *Am J Hum Genet* **102**, 890-903 (2018).
14. Wu, L. *et al.* A transcriptome-wide association study of 229,000 women identifies new candidate susceptibility genes for breast cancer. *Nat Genet* **50**, 968-978 (2018).
15. Barfield, R. *et al.* Transcriptome-wide association studies accounting for colocalization using Egger regression. *Genet Epidemiol* **42**, 418-433 (2018).
16. Ferreira, M.A. *et al.* Genome-wide association and transcriptome studies identify target genes and risk loci for breast cancer. *Nat Commun* **10**, 1741 (2019).
17. Hanahan, D. & Weinberg, R.A. Hallmarks of cancer: the next generation. *Cell* **144**, 646-74 (2011).
18. Zhao, X. *et al.* Telomerase-immortalized human mammary stem/progenitor cells with ability to self-renew and differentiate. *Proc Natl Acad Sci U S A* **107**, 14146-51 (2010).
19. Toouli, C.D. *et al.* Comparison of human mammary epithelial cells immortalized by simian virus 40 T-Antigen or by the telomerase catalytic subunit. *Oncogene* **21**, 128-39 (2002).
20. Mani, S.A. *et al.* The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell* **133**, 704-15 (2008).

21. Elenbaas, B. *et al.* Human breast cancer cells generated by oncogenic transformation of primary mammary epithelial cells. *Genes Dev* **15**, 50-65 (2001).
22. Davies, R. *et al.* CRISPRi enables isoform-specific loss-of-function screens and identification of gastric cancer-specific isoform dependencies. *Genome Biol* **22**, 47 (2021).
23. Rosenbluh, J. *et al.* Complementary information derived from CRISPR Cas9 mediated gene deletion and suppression. *Nat Commun* **8**, 15403 (2017).
24. Chen, L. *et al.* Identification of cofilin-1 as a novel mediator for the metastatic potentials and chemoresistance of the prostate cancer cells. *Eur J Pharmacol* **880**, 173100 (2020).
25. Li, X. *et al.* Hhex inhibits cell migration via regulating RHOA/CDC42-CFL1 axis in human lung cancer cells. *Cell Commun Signal* **19**, 80 (2021).
26. Jia, D. *et al.* Crebbp Loss Drives Small Cell Lung Cancer and Increases Sensitivity to HDAC Inhibition. *Cancer Discov* **8**, 1422-1437 (2018).
27. Hogg, S.J. *et al.* Targeting histone acetylation dynamics and oncogenic transcription by catalytic P300/CBP inhibition. *Mol Cell* **81**, 2183-2200 e13 (2021).
28. Glubb, D.M. *et al.* Candidate Causal Variants at the 8p12 Breast Cancer Risk Locus Regulate DUSP4. *Cancers (Basel)* **12**(2020).
29. Gupta, A. *et al.* Dual-specificity protein phosphatase DUSP4 regulates response to MEK inhibition in BRAF wild-type melanoma. *Br J Cancer* **122**, 506-516 (2020).
30. He, H., Du, Z., Lin, J., Wu, W. & Yu, Y. DUSP4 inhibits autophagic cell death in PTC by inhibiting JNK-BCL2-Beclin1 signaling. *Biochem Cell Biol*, 1-8 (2021).
31. Hijjiya, N. *et al.* Genomic Loss of DUSP4 Contributes to the Progression of Intraepithelial Neoplasm of Pancreas to Invasive Carcinoma. *Cancer Res* **76**, 2612-25 (2016).
32. Pilie, P.G., Tang, C., Mills, G.B. & Yap, T.A. State-of-the-art strategies for targeting the DNA damage response in cancer. *Nat Rev Clin Oncol* **16**, 81-104 (2019).
33. Olivieri, M. *et al.* A Genetic Map of the Response to DNA Damage in Human Cells. *Cell* **182**, 481-496 e21 (2020).
34. Beesley, J. *et al.* Chromatin interactome mapping at 139 independent breast cancer risk signals. *Genome Biol* **21**, 8 (2020).
35. Fulco, C.P. *et al.* Systematic mapping of functional enhancer-promoter connections with CRISPR interference. *Science* **354**, 769-773 (2016).
36. Gasperini, M. *et al.* A Genome-wide Framework for Mapping Gene Regulation via Cellular Genetic Screens. *Cell* **176**, 1516 (2019).
37. Doench, J.G. *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol* **34**, 184-191 (2016).
38. Zheng, G.X. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat Commun* **8**, 14049 (2017).
39. Barry, T., Wang, X., Morris, J.A., Roeder, K. & Katsevich, E. Conditional resampling improves calibration and sensitivity in single-cell CRISPR screen analysis. *bioRxiv*, 2020.08.13.250092 (2021).
40. Bojesen, S.E. *et al.* Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer. *Nat Genet* **45**, 371-84, 384e1-2 (2013).
41. Dunning, A.M. *et al.* Breast cancer risk variants at 6q25 display different phenotype associations and regulate ESR1, RMND1 and CCDC170. *Nat Genet* **48**, 374-86 (2016).

42. Datlinger, P. *et al.* Pooled CRISPR screening with single-cell transcriptome readout. *Nat Methods* **14**, 297-301 (2017).
43. Subramanian, A. *et al.* A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell* **171**, 1437-1452 e17 (2017).
44. Hill, A.J. *et al.* On the design of CRISPR-based single-cell molecular screens. *Nat Methods* **15**, 271-274 (2018).
45. Adamson, B. *et al.* A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. *Cell* **167**, 1867-1882 e21 (2016).
46. Chang, L. *et al.* The SWI/SNF complex is a mechanoregulated inhibitor of YAP and TAZ. *Nature* **563**, 265-269 (2018).
47. Alver, B.H. *et al.* The SWI/SNF chromatin remodelling complex is required for maintenance of lineage specific enhancers. *Nat Commun* **8**, 14648 (2017).
48. Mathies, L.D. *et al.* SWI/SNF complexes act through CBP-1 histone acetyltransferase to regulate acute functional tolerance to alcohol. *BMC Genomics* **21**, 646 (2020).
49. Corsello, S.M. *et al.* The Drug Repurposing Hub: a next-generation drug library and information resource. *Nat Med* **23**, 405-408 (2017).
50. French, J.D. *et al.* Functional variants at the 11q13 risk locus for breast cancer regulate cyclin D1 expression through long-range enhancers. *Am J Hum Genet* **92**, 489-503 (2013).
51. Meyer, K.B. *et al.* Fine-scale mapping of the FGFR2 breast cancer risk locus: putative functional variants differentially bind FOXA1 and E2F1. *Am J Hum Genet* **93**, 1046-60 (2013).
52. Baxter, J.S. *et al.* Functional annotation of the 2q35 breast cancer risk locus implicates a structural variant in influencing activity of a long-range enhancer element. *Am J Hum Genet* **108**, 1190-1203 (2021).
53. Glubb, D.M. *et al.* Fine-scale mapping of the 5q11.2 breast cancer locus reveals at least three independent risk variants regulating MAP3K1. *Am J Hum Genet* **96**, 5-20 (2015).
54. Orr, N. *et al.* Fine-mapping identifies two additional breast cancer susceptibility loci at 9q31.2. *Hum Mol Genet* **24**, 2966-84 (2015).
55. Darabi, H. *et al.* Polymorphisms in a Putative Enhancer at the 10q21.2 Breast Cancer Risk Locus Regulate NRBF2 Expression. *Am J Hum Genet* **97**, 22-34 (2015).
56. Lawrenson, K. *et al.* Functional mechanisms underlying pleiotropic risk alleles at the 19p13.1 breast-ovarian cancer susceptibility locus. *Nat Commun* **7**, 12675 (2016).
57. Ghossaini, M. *et al.* Evidence that the 5p12 Variant rs10941679 Confers Susceptibility to Estrogen-Receptor-Positive Breast Cancer through FGF10 and MRPS30 Regulation. *Am J Hum Genet* **99**, 903-911 (2016).
58. Beesley, J. *et al.* eQTL Colocalization Analyses Identify NTN4 as a Candidate Breast Cancer Risk Gene. *Am J Hum Genet* **107**, 778-787 (2020).
59. Betts, J.A. *et al.* Long Noncoding RNAs CUPID1 and CUPID2 Mediate Breast Cancer Risk at 11q13 by Modulating the Response to DNA Damage. *Am J Hum Genet* **101**, 255-266 (2017).
60. Giambartolomei, C. *et al.* H3K27ac HiChIP in prostate cell lines identifies risk genes for prostate cancer susceptibility. *Am J Hum Genet* (2021).

61. Han, K. *et al.* CRISPR screens in cancer spheroids identify 3D growth-specific vulnerabilities. *Nature* **580**, 136-141 (2020).
62. Boix, C.A., James, B.T., Park, Y.P., Meuleman, W. & Kellis, M. Regulatory genomic circuitry of human disease loci by integrative epigenomics. *Nature* **590**, 300-307 (2021).
63. Nasser, J. *et al.* Genome-wide enhancer maps link risk variants to disease genes. *Nature* **593**, 238-243 (2021).
64. Moradi Marjaneh, M. *et al.* Non-coding RNAs underlie genetic predisposition to breast cancer. *Genome Biol* **21**, 7 (2020).
65. Connally, N. *et al.* The missing link between genetic association and regulatory function. *medRxiv*, 2021.06.08.21258515 (2021).
66. Buenrostro, J.D., Wu, B., Chang, H.Y. & Greenleaf, W.J. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr Protoc Mol Biol* **109**, 21 29 1-21 29 9 (2015).
67. Rosenbluh, J. *et al.* Genetic and Proteomic Interrogation of Lower Confidence Candidate Genes Reveals Signaling Networks in beta-Catenin-Active Cancers. *Cell Syst* **3**, 302-316 e4 (2016).
68. Gyorffy, B. Survival analysis across the entire transcriptome identifies biomarkers with the highest prognostic power in breast cancer. *Comput Struct Biotechnol J* **19**, 4101-4109 (2021).
69. Tsherniak, A. *et al.* Defining a Cancer Dependency Map. *Cell* **170**, 564-576 e16 (2017).

Acknowledgments:

This work was supported by a DoD grant to J.R. and G.C.T (grant number: W81XWH1910116). J.R is supported by a Victoria cancer agency fellowship (grant number: MCRF20035). G.C.T. is an NHMRC Leadership Fellow. S.L.E is an NHMRC Senior Research Fellow (grant number: 1135932). J.D.F. is supported by a philanthropic donation from Isabel and Roderic Allpass.

We thank the Functional Genomics Platform, the Bioinformatics Platform and Micromon genomics platform at Monash University for help with CRISPR screens, data analysis and single cell experiments, and Professor Gail Risbridger and A/Prof. Renea Taylor, Monash University, for providing NSG mice.

Author contributions:

Conceptualization, G.C.T and J.R; Methodology, G.C.T., J.R., N.T., J.B., M.M., W.S., L.M., J.P., D.B., A.C., K.M. A.H., K.H., S.K., H.S., J.M.P, J.F., S.E. Analysis, J.B., J.R., N.T., D.P. Resources, R.R., V.M. Writing-Original Draft, J.R., G.C.T, J.B. Writing - Review & Editing, all authors; Supervision, N.T., G.C.T, J.M.P, J.B. J.R. Funding Acquisition, J.R. and G.C.T

Competing interests:

The authors declare no competing interests.

Materials and Methods

Cell lines:

Human mammary epithelial cells (HMLE)²¹ used in this study was a gift from Prof. William Hahn (Dana Farber Cancer Institute), the B80 cell lines (B80-T17 and B80-T5) are in-vitro immortalized mammary cell lines previously described¹⁹. K5+/K19- and K5+/K19+ cell lines are immortalized progenitor mammary stem cells¹⁸. HMLE were induced to undergo epithelial to mesenchymal transition (EMT) to obtain a mesenchymal phenotype (mesHMLE) by culturing cells in DMEM:F12 media (1:1) supplemented with 10µg/ml insulin, 20ng/ml EGF, 0.5µg/ml hydrocortisone, 5µg/ml gentamycin, 5% FBS treated with 2.5ng/ml TGFβ1 for a minimum of 14 days²⁰. HMLE and B80-T17 were propagated in mammary epithelial growth medium (MEGM) (Sigma). B80-T5 were cultured in RPMI 1640 (Sigma) supplemented with 10% FBS, 1% Penicillin and Streptomycin and 1% Glutamine. K5+/K19- and K5+/K19- cells were maintained in DFCI medium containing: MEMα/Ham's F12 nutrient mixture (1:1, vol/vol) supplemented with 0.1M HEPES, 1µg/ml insulin, 1µg/ml hydrocortisone, 12.5ng/ml epidermal growth factor, 10µg/ml transferrin, 14.1µg/ml phosphoethanolamine, 0.545ng/ul β-Estradiol, 2mM glutamine, 2.6ng/ml sodium selenite, 1ng/ml cholera toxin, 6.5ng/ml triiodothyronine, 0.1 mM ethanolamine, 35µg/ml bovine pituitary extract, 10µg/ml gentamycin and 10µg/ml freshly prepared ascorbic acid. All cell lines were maintained in a humidified incubator at 37°C with 5%CO₂.

Generation of stable cell lines:

Lentiviral vector expressing a gene or sgRNA of interest, along with pMD2.G (Addgene#12259) and psPAX2 (Addgene#12260) were transfected into HEK293FT packaging cells (Thermo Fisher#R70007). Lentiviral supernatant was harvested after 48-hour incubation in DMEM containing 30% FBS and passed through a 0.45µm Milli-hex filter. For oncogenic potential: K5+/K19- and K5+/K19+ cells were transduced with pLENTI-Hygro-PGK-TP53-DD and selected using 100µg/ml hygromycin. For colony formation assays and in vivo assays: HMLE, mesHMLE, B80-T5, B80-T17 and K5+/K19+ were transduced with pLX311-GFP-MEKDD and selected for GFP using the BD Influx™ cell sorter. K5+/K19- was transduced with pRRLsin-SV40 T antigen-IRES-mCherry (Addgene#58993) and positive cells were sorted using the BD Influx™ cell sorter. For CRISPR screens and validations all cell lines were transduced with following lentiviral vectors: Lenti-Cas9-2A-Blast (Addgene#73310), Lenti-dCas9-KRAB-Blast (Addgene#89567) and Lenti-dCas9-VP64-Blast (Addgene #61425). Cells were selected and maintained in blasticidin (5µg/ml to 10µg/ml). For single gene perturbation, 3 sgRNAs

were cloned into BsmBI-digested lenti-Guide-Puro vector (Addgene#52963) for CRISPRko and pXPR502 vector (Addgene#96923) for CRISPRa. Cells were infected with sgRNAs, selected and maintained in puromycin (1µg/ml to 2µg/ml).

RNA-seq:

Transcriptome profiling was carried out using strand-specific TruSeq kit. Following RNA extraction (RNeasy, Qiagen) mRNA was enriched using polyT beads (Genewiz) and sequencing libraries were prepared using Illumina strand-specific TruSeq kit (Genewiz). Samples were sequenced on an Illumina HiSeq machine (PE 150bp). RNA-seq were aligned to Ensembl v70 gene models with STAR v2.7.1a. Duplicate reads were marked with PicardTools v2.19, then reads mapping to transcriptome using featureCounts in subread v1.6.0, count matrix generated using RSEM v1.3.1. Differential expression analysis was performed using DESeq2 in R v3.6.2.

ATAC-seq:

Profiling of regions of open chromatin using previously reported protocols⁶⁶. Duplicate libraries were prepared for each cell type and paired-end sequenced (150bp) generating a minimum of 40 filtered reads per library. Adapters trimmed using Cutadapt v1.13 and reads aligned to GRCh37 using Bowtie v2.2.9. Duplicates marked with Picard MarkDuplicates v2.19. Peaks were called using MACS2 and cell type-specific replicating peaks identified using BedTools.

HiChIP:

HiChIP libraries were generated with the Arima HiChIP kit using an antibody against H3K27ac (Active Motif AbFlex: 91193). Cells were counted using the Countess II automated cell counter (Thermo Scientific) and fixed with 2% formaldehyde using the Arima HiC+ Kit (Arima, A101020). 1e6 fixed cells were used in restriction enzyme digest, biotin end filling ligation reactions to the manufacturer's protocol. Libraries were prepared using the KAPA Kit (KAPA, KK2620), according to the Arima-HiC kit protocol. Libraries were indexed using the Swift Biosciences indexing kit then paired-end sequenced (150bp) with Illumina Novaseq 6000 to generate >500M raw reads per library. Individual replicate reads were processed with HiC-Pro (v 2.11.4) and aligned to hg19. Replicate samples for each cell type were quality controlled and checked for genome-wide signal correlation before merging with HiC-Pro. Enriched regions representing H3K27ac peaks were detected using MACS2. Chromatin loops were detected in each cell type-specific dataset using FitHiChIP v8.1 at 2 kb resolution limiting to 2 Mb interaction distance. Peak-to-peak and peak-to-nonpeak loops were used for background modelling and a $q < 0.01$ threshold set to determine significant interactions.

Generation of pooled sgRNA library:

sgRNA sequences in custom libraries are available in Supplementary Table 2. sgRNAs were designed using CRISPick algorithm³⁷. For each gene we chose top scoring 5 sgRNAs (based on CRISPick scores). Libraries were prepared as previously described^{22,23,67}. Briefly, oligonucleotide pools (CustomArray) contained the sgRNA sequence appended to BsmBI cutting sites and overhang sequences for PCR amplification. The final sequence obtained is: AGGCACTTGCTCGTACGACGCGTCTCACACCG[20nt spacer]GTTTCGAGACGTTAAGGTGCCGGGCCACAT. Following PCR amplification with Fwd: 5'-AGGCACTTGCTCGTACGACG-3', Rev: 5'-ATGTGGGCCCGGCACCTTAA-3' primers the PCR product was cloned via Golden Gate assembly into BsmBI-digested lentiGuide-Puro vector (Addgene# 52963) for CRISPRko and CRISPRi libraries and into pXRP502 (Addgene #96923) for CRISPRqtl and CROPSeq oligos were cloned into CROPseq-Guide-Puro (Addgene#86708). Ligated libraries were electroporated into NEB5 α electrocompetent cells (NEB), plasmid DNA was extracted using Qiagen Maxi Prep. For each library preparation, a 1000X representation was ensured.

2D and 3D proliferation screens:

Mammary cell lines (HMLE, mesHMLE, B80-T5, B80-T17, K5+/K19- and K5+/K19+ cells) stably expressing Cas9 (Addgene# 73310), KRAB-dCas9 (Addgene# 89567) or dCas9-VP64 (Addgene# 61425) were established. Cells were then transduced with either the CRISPRko, CRISPRi or CRISPRa Library at a MOI of 0.3 to obtain 1,000 cells/sgRNA. Twenty-four hours post infection cells were selected using Puromycin (2 μ g/ml) for 7 days. Cells were then subdivided to assay for 3D proliferation by plating cells in low attachment conditions (Corning#4615) or on 2D plates. To ensure sgRNA and Cas9/dCas9 expression, cells were maintained with puromycin and blasticidin throughout the screen. Twenty-one days post-infection cells were washed in PBS and genomic DNA was extracted using NucleoSpin Blood XL kit (Clontech). For colonies grown in low attachment conditions, genomic DNA was extracted using the DNeasy Kit (Qiagen).

Olaparib synthetic lethal screens:

Cell lines stably expressing Cas9 (Addgene#73310), KRAB-dCas9 (Addgene# 89567) or dCas9-VP64 (Addgene#61425), were transduced with the CRISPRko, CRISPRi or CRISPRa sgRNA libraries at a low MOI (0.3) at a coverage of 1,000 cells/sgRNA. Puromycin containing medium was added 24 hours post-infection and cells were allowed to undergo selection for 7 days. For all screens, following selection, cells were trypsinized and divided into two treatment groups: DMSO or Olaparib. HMLE, mesHMLE, B80-T17, K5+/K19- and K5+/K19+ cells were treated with 5 μ M of Olaparib and B80-T5 cells were treated with 2.5 μ M of Olaparib for 14 days. Media was replaced every 4 days with DMSO or Olaparib. Cells were harvested by centrifugation and genomic DNA was extracted using NucleoSpin Blood XL kit (Clontech).

In vivo screen:

HMLE-MEKDD, K5+/K19+-MEKDD and B80-T5-MEKDD cells expressing Cas9 or dCas9-VP64 were infected at MOI=0.3 with CRISPRko or CRISPRa libraries. Following puromycin selection (2µg/ml) for 7 days, 2e6 cells/site were subcutaneously injected into NSG mice at 3 sites/mouse. Tumor growth was measured using a digital caliper every 48 hours and monitored continuously until tumor volume reached 1cm³ (sum of all three sites). Tumor volume was calculated using the formula length (mm) × width (mm) × height (mm). Mice were sacrificed once tumors reached 1cm³. Cells were dissociated using Bead Ruptor machine and glass beads and DNA was extracted using DNeasy Kit (Qiagen).

Library preparation, sequencing and analysis:

High-throughput sequencing library was generated using one-step PCR to amplify the integrated sequence within the construct and the addition of a barcode as previously described^{22,23,67}. PCR products were then purified using AMPure beads and samples sequenced using HiSeq (Illumina). PoolQ (<https://portals.broadinstitute.org/gpp/public/software/poolq>) was used for deconvolution and alignment of sgRNA reads.

Crystal violet proliferation assay:

Cells were plated at 2,000 cells/well and allowed to propagate until confluent. Media was aspirated and washed twice in PBS followed by fixation in 10% formalin for 10 minutes at room temperature. Formalin was removed and 0.5% (w/v) of crystal violet solution (Sigma) was added and incubated for 20 minutes at room temperature. Plates were washed in dH₂O and imaged. For quantification 10% acetic acid was added to each well and incubated at room temperature for 30 minutes. The crystal violet solution was quantified by measuring the OD at 590nm using the PHERAstar (BMG).

3D proliferation assays:

Cells were plated at 8000 cells/well in a 24-well low attachment plate (Corning). Colonies were allowed to form for 21 days. Images were taken at 4X magnification using an EVOS M5000 microscope (Thermo). Quantification of colonies was done by adding Cell-Titer-Glo Reagent (Promega) to wells, followed by a 10-minute incubation at room temperature on a shaker. Cell lysates were transferred to a 96-well white plate and luminescence measured using the PHERAstar (BMG).

sgRNAs for validation screens:

All sgRNA sequences used in validation screens are in Supplementary Table 9

Western blot:

Cells/tissue were harvested, washed in PBS and resuspended in RIPA buffer (CST-9806) containing proteinase inhibitors (Roche) and quantified using the Pierce BCA Protein Assay Kit (Thermo Fisher). Protein lysates diluted in 4 X Laemmli Sample Buffer (Bio-Rad 161-0747) were loaded onto Bio-Rad 4-20% precast gels. Following electrophoresis, proteins were transferred to a pre-activated PVDF membrane using the Trans-Blot®Turbo™ Transfer System and visualized using ECL (Bio-Rad Chemidoc). Antibodies used in this study are DUSP4 (CST#5149), Cyclin D1 (CST#2978), Cyclin E1 (CST#4129), ATF7IP (Sigma#16578), ATF7IP (Sigma#HPA023505), ADCY3 (Abcam#ab199157), ATXN7 (Invitrogen#PAI-749), CREBBP (CST#7389), SAPK/JNK (CST#9252), LPAR2 (Abcam#ab135980), NF1 (Bethyl#A300-140A-M), Phospho-p38 MAPK (Thr180/Tyr182) (CST#4511), p38 MAPK (D13E1) (CST#8690), p44/42 MAPK (Erk1/2) (CST#4695), Phospho-p44/42 MAPK (Erk1/2) (Thr202/Tyr204) (CST#4370), Phospho-SAPK/JNK (Thr183/Tyr185) (CST#4668), TRAIL (CST#3219), RIP140 (Santa Cruz#sc518071), GAPDH (Santa Cruz#sc32233), c-Myc (CST#5605), TGFBR2 (Santa Cruz#sc17792), CASZ1 (Santa Cruz#sc398303), CFL1 (Abcam#ab42824).

Animals:

The Monash University Animal Ethics Committee approved all animal use in this study (AEC – approval number 2020-24197-49078). For these experiments, 5-7 week old female NSG mice were purchased from Australian Research Laboratories (WA, Australia) or were kindly gifted from Professor Gail Risbridger and A/Prof. Renea Taylor (Monash University).

Validation of in vivo screens:

B80-T5-MEKDD cells stably expressing Cas9 were infected with lentiviruses containing sgRNA's targeting AAVS1 (control), ATF7IP, DUSP4, TGFBR2, CREBBP. Twenty-four hours post-infection, cells underwent puromycin selection for 7 days and expanded. Cells were trypsinized, washed twice in PBS and injected into NSG mice subcutaneously under isoflurane anesthesia. For each sgRNA, we injected 2e6 cells/site, 3 sites per mouse. Tumor growth was measured using a digital caliper every 48 hours and monitored continuously until tumor volume reached 1cm³ (sum of all three sites). Tumor volume was calculated using the formula length (mm) × width (mm) × height (mm). Mice were sacrificed once tumors reached 1cm³.

Enrichment analysis:

For each method of target gene selection, enrichment was calculated as the fraction of positive hits that were selected by that method, divided by the fraction of the total number of genes in the screen that were included based on that method: (Nhits method/Nhits total)/(Ngenes method/Ngenes total). The calculation was performed for each combination of assay, phenotype and cell line.

Survival analysis:

We used KMplotter⁶⁸ to evaluate associations between relapse free survival following diagnosis of overall BC (and separately ER+ and ER- BC) and expression of the genes.

CRISPRqtl:

CRISPRqtl was done as previously described³⁶. Briefly, K5+/K19+ cells stably expressing KRAB-dCas9 were infected with the CRISPRqtl library at MOI=5. 24h post infection, cells were selected with puromycin (2µg/ml) and cultured for 10 days. Cells were trypsinized washed with PBS and resuspended in PBS to reach a concentration of 1,200 cells/ml. Single-cell suspensions were loaded on two lanes of a 10X Genomics Chromium Controller and Chromium Next GEM Single Cell 3' GEM. We detected a total of 13,334 cells with a mean of 16,953 reads/cell and a median unique molecular identifier (UMI) of 4,006 UMIs/cell. Library and Gel Bead Kit v3.1 (10X Genomics cat #1000121), per manufacturer's instructions (CG000204 Rev D) with the following modifications and variables. A single sample was loaded in two wells of the Next Gem Chip G, overloaded at 150% of the recommended cell input volume, with the corresponding volume of dH2O deducted at Step 1.2b (using the Cell Suspension Volume Calculator Table; p26). At Step 2.2d, cDNA was generated using 11 cycles of PCR. Samples were recombined 1:1 before Step 3.1. Prior to enzymatic shearing, 10% of the cDNA was used for sgRNA PCR enrichment. Specifically, A three-step nested PCR was used for gRNA enrichment⁴⁴.

PCR 1: 5 ng of 10x cDNA was amplified using NEBNext high fidelity 2x PCR mix (NEB # M0541) and the following primers: Rxn1_Fwd: TTTCCCATGATTCCTTCATATTTGC, Rxn1_Rev: AACTCTTTCCCTACACGACG. Cycling conditions: 98°C for 30s, 14x (98°C for 10s, 50°C for 10s, 72°C for 20s), 72°C for 2min. PCR product was gel purified using the Qiagen MinElute Gel extraction kit (Qiagen # 28604).

PCR 2: 5ng of PCR 1 was amplified using NEBNext high fidelity 2x PCR mix (NEB # M0541) and the following primers: Rxn2_Fwd: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTTGTGGAAAGGACGAAACAC, Rxn2_Rev:

AATGATACGGCGACCACCGAGATCTACTACTCTTTCCCTACACGACGCTC. Cycling conditions: 98°C for 30s, 7x (98°C for 10s, 64°C for 10s, 72°C for 15s), 72°C for 2min. PCR product was gel purified using the Qiagen MinElute Gel extraction kit (Qiagene # 28604).

PCR 3: 5ng of PCR 1 was amplified using NEBNext high fidelity 2x PCR mix (NEB # M0541) and the following primers: Rxn3_Fwd: CAAGCAGAAGACGGCATAACGAGATGACAGCATGTGACTGGAGTTCAGACGT, Rxn2_Rev (see PCR_2). Cycling conditions: 98°C for 30s, 11x (98°C for 10s, 64°C for 10s, 72°C for 15s), 72°C for 2min. PCR product was gel purified using the Qiagen MinElute Gel extraction kit (Qiagene # 28604) and then purified using AMPure beads (Beckman Coulter # A63881). cDNA and PCR product were pooled at a 1:10 ration and

sequenced on two lanes of an MGISEq machine (Genewiz) using 150 PE-cycles (total of 569e6 reads).

CROPSeq:

CROPSeq⁴⁵ analysis was similar to CRISPRqtl with the following modifications. CROPSeq library was transduced into K5+/K19+ cells expressing WT Cas9 at a MOI of 0.1 ensuring 1 sgRNA/cell. Single cell isolation and library preparation was exactly as described for CRISPRqtl but only one 10x chromium lane was used (15,181 cell isolated) and sequencing was done on one MGISEq lane.

L1000 analysis

For each gene knockout we use the Z-score matrix (Supplementary Table 7) to define the up and down regulated genes. Top 150 up/down regulated genes were used as an input for the CMap for reference perturbagen signatures (<https://clue.io/query>)⁴³. We used version 1 of the CMap signature database for this analysis and collected both individual compound (PERT) and PCL for each of these signatures.

Data and materials availability:

All unique reagents generated in this study will be made available upon request. An agreement with our Institute's Materials Transfer Agreement (MTA) may be required. Further information and requests for resources and reagents should be directed to and will be fulfilled by Joseph Rosenbluh (sefi.rosenbluh@monash.edu).

RNA-Seq and ATAC-Seq data generated in this study will be made available at GEO. All CRISPR functional screening raw and analyzed data is available in the supplementary data of this paper.

Figure Legends

Figure 1: CRISPR activation and suppression screens identify BC risk genes that regulate proliferation in 2D, 3D. (A) Pie chart showing locations of CCVs at 205 BC risk signals identified by GWAS. (B) Classes of genes selected for functional CRISPR screens. INQ_1 – high-confidence INQUISIT predictions; INQ_2 - moderate-confidence INQUISIT predictions; TWAS – identified by transcriptome wide association studies and eQTL studies. (C) Experimental approach. (D) Summary of results from 2D and 3D proliferation screens.

Figure 2: CRISPR activation and suppression screens identify BC risk genes that regulate tumor growth in mice. (A) Experimental approach. (B) Summary of hits from the *in vivo* screens. (C) Proliferation in 3D cultures of WT or MEKDD expressing B80-T5 cells following suppression of *DUSP4* expression. (D) *DUSP4* protein levels following expression of MEKDD. (E) Phosphorylation of ERK, JNK and p38 following MEK inhibition. (F) The effect of MEK inhibitors on proliferation in *DUSP4* ko cells.

Figure 3: Olaparib synthetic lethal screens identify BC-risk genes that regulate the DNA repair pathway. (A) Experimental approach. (B) B80-T5 cells infected with control or *BRCA1*-targeting sgRNAs were treated with olaparib for 7 days. (C) Summary of hits. Known DNA repair genes were annotated based on³³. (D) GSEA pathway enrichment analysis of hits (not including background positive control genes). (E) Validation of selected hits in a singleton experiment using a crystal violet readout. (F) Enrichment analysis of hits from CRISPR screens from the classes of genes selected for the screens. INQ_1 – high-confidence INQUISIT predictions (genes scored as INQ_1 due to their driver status were downgraded to INQ_2 for this analysis); INQ_2 - moderate-confidence INQUISIT predictions; TWAS – identified by transcriptome wide association studies or eQTL studies.

Figure 4: Chromatin conformation assays confirm interactions between BC risk loci and genes that score in functional screens. (A) Summary of HiChIP chromatin interactions observed between BC risk loci and genes that scored in functional screens, where color scale signifies scaled levels of chromatin interaction scores and count. (B) Example of chromatin interactions between CCVs and *ATF7IP*. (C) The regulatory element carrying the protective alleles of CCVs rs16909788, rs17221259, rs11055880 increase *ATF7IP* promoter activity. Constructs containing all three SNPs were tested using luciferase reporter assays. PRE mutant 1 contains the protective haplotype with rs11055880 altered to the risk allele. PRE mutant 2 contains the risk haplotype with rs16909788 and rs17221259 altered to protective alleles. Bars show mean luciferase

intensity relative to promoter activity and error bars represent 95% confidence intervals. P-values were determined by two-way ANOVA followed by Dunnett's multiple comparisons test (**** $p < 0.0001$). (D) *ATF7IP* expression was measured in K5+/K19- cells 21 days post infection with CRISPRi sgRNAs targeting the *ATF7IP* CCV-containing enhancer. (E) Strategy used for CRISPRqtl experiment. (F) Z-Scores from CRISPRqtl screen of sgRNAs targeting 25 known TSSs. (G) Gene targets identified by CRISPRqtl or INQUISIT for six regions included in CRISPRqtl screen that contain genes that scored in CRISPR functional screens. (H) Example of CRISPRqtl results at the chr12:1391331-14913931 locus. sgRNA-gene pairs with a Z-Score < -1 are shown. Colored lines represent genes that scored as hits (>3 sgRNAs with a Z-Score < -1). Green - *GPRC5A*. Blue - *EMP1*, Red - *ATF7IP*. (I) Zoom in on *ATF7IP* sgRNA-gene pairs showing consistency between HiChIP and CRISPRqtl.

Figure 5: CROPSeq expression profiling identifies drugs targeting BC associated gene targets. (A) Unsupervised hierarchical clustering using CROPSeq expression profiles. (B) PCLs connected to control genes. (C) PCLs connected to genes that scored as tumor suppressor genes. (D) PCLs connected to genes that scored as oncogenes. (E) Summary of BC risk loci and INQUISIT Level 1 genes. (F) Validation of HDAC inhibitors in K5+/K19+ cells following *CREBBP* ko.

Extended Data Figure legends

Extended Data Figure 1: Example of genomic features used in INQUISIT to predict gene targets.

Extended Data Figure 2: Characterization of immortalized cell lines used in this study. (A) PCA analysis using RNA-Seq or (B) ATAC-Seq or data for cell lines used in this study. (C) Top 200 variable genes identified in RNA-Seq. Genes that are part of the Luminal Progenitor (LumProg) or mesenchymal (MASC) gene signatures are highlighted demonstrating that K5+/K19+, K5+/K19- and mesHMLE are more mesenchymal.

Extended Data Figure 3: Identification of genes that upon suppression or activation promote 2D or 3D growth. (A) Distribution of positive and negative controls in 2D proliferation screens. Plots showing genes that score in 2D or 3D proliferation screens in: (B) K5+/K19- (C) K5+/K19+ (D) B80-T5 (E) B80-T17 (F) HMLE (G) mesHMLE.

Extended Data Figure 4: Comparison between CRISPRko and CRISPRi screens. (A) Correlation between proliferation changes observed in 2D proliferation screens for the

indicated cell lines. (B) Proliferation changes in 6 cell lines following CRISPRko or CRISPRi mediated suppression of *ATXN7*. (C) Genomic view of *ATXN7* showing the shared promoter of *ATXN7* and *THOC7* (D) Proliferation changes in 6 cell lines following CRISPRko or CRISPRi mediated suppression of *THOC7*. (E) Dependency score (CRES scores) in 796 cell lines for *ATXN7* and *THOC7* from DepMap⁶⁹.

Extended Data Figure 5: Comparison between 2D and 3D screens. Correlation between 2D and 3D proliferation assays in (A) CRISPRko (B) CRISPRi or (C) CRISPRa. (D) Proliferation changes in 2D and 3D cultures mediated by CRISPRko of *CFL1*.

Extended Data Figure 6: Validation of sgRNAs used to suppress or activate INQUISIT Level 1 hits. (A) Western blot analysis of candidate tumor-suppressor genes using CRISPRko. (B) Western blot analysis of candidate oncogenes using CRISPRa.

Extended Data Figure 7: Validation of INQUISIT Level 1 hits that induce a 2D or 3D proliferation phenotype. Following transduction with sgRNAs targeting INQUISIT Level 1 hits (3 sgRNAs/gene) using either CRISPRko (A, C) or CRISPRa (B, D) proliferation was measured in 2D (A, B) or 3D (C, D). Results are displayed as an average +/- SD of three sgRNAs.

Extended Data Figure 8: Identification of BC-risk genes that upon suppression or activation promote growth in immune deficient mice. (A) 3D proliferation of the indicated cell lines with or without MEKDD expression. Hits from CRISPRko and CRISPRa in vivo screens in (B) K5+/K19+_MEKDD (C) HMLE_MEKDD (D) B80-T5_MEKDD cells. INQUISIT Level 1 predicted genes that scored are labelled with the gene name. (E) Validation of INQUISIT Level 1 hits in B80-T5_MEKDD cells. Each time point is an average +/- SD of three sgRNAs in six mouse tumors. (F) Representative tumors from in-vivo validation in B80-T5-MEKDD cells.

Extended Data Figure 9: DUSP4 is a tumor-suppressor gene that is regulated by MEK1 expression and regulates phosphorylation of pJNK and pp38. Protein levels of DUSP4 and pERK following expression of MEKDD. DUSP4, pERK, pp38 and pJNK protein levels in K5+/K19+ or K5+/K19+_MEKDD cell lines cultured in (A) 2D or (B) in vivo. (C) DUSP4, pERK, levels in B80-T5 or B80-T5_MEKDD cells treated for 1h with 10nM of trametinib or 100nM of selumetinib. (D) Correlation between different in vivo and in vitro proliferation assays.

Extended Data Figure 10: Identification of genes that upon suppression or activation modulate the DNA damage response. Plots showing genes that score in olaparib synthetic lethal screen (A) K5+/K19- (B) K5+/K19+ (C) mesHMLE (D) HMLE (E) B80-T17 (F) B80-T5. (G) Quantification of crystal violet proliferation assays.

Extended Data Figure 11: Survival analysis of BC patients correlated with expression levels of hits. Survival analysis was performed using KMplotter5 to evaluate associations between relapse free survival following diagnosis of BC and expression of the genes. Results are shown for genes that met the Bonferroni-corrected significance threshold ($p < 0.0008$). All the survival curves are for overall BC except for F, H, L and P which are for ER+ BC.

Extended Data Figure 12: CROPSeq identifies signatures and opportunities for drug repurposing. (A) Threshold of sgRNA UMIs in CROPSeq. Using a threshold of 10 most cells contain only one sgRNA. (B) Expression of target gene following CRISPRko mediated suppression in CROPSeq. (C) Target gene expression in CROPSeq for genes with low (normalized expression < 0.1) or high (normalized expression > 0.1) expression. (D) GSEA analysis identifies mTORC signature following CRISPRko mediated suppression of RPTOR in CROPSeq. (E) *AXIN2* expression in CROPSeq showing that negative regulators of the WNT signaling pathway activate *AXIN2* expression.

Supplementary Tables

Supplementary Table 1: BC risk signals identified in GWAS and INQUISIT gene predictions. Signals from BC GWAS and INQUISIT gene prediction for these signals.

Supplementary Table 2: sgRNA sequences. Sequences of sgRNAs used in various CRISPR screens in this study.

Supplementary Table 3: Raw counts from different CRISPR screens. Raw sequencing reads from various CRISPR screens.

Supplementary Table 4: MAGeCK analysis from CRISPR screens. MAGeCK analysis for identification of enriched and depleted genes in CRISPR screens.

Supplementary Table 5: HiChIP interactions in K5+/K19+ and BRE80-T5 cells. Chromatin interactions between BC risk signals and genes at a 2Mb window.

Supplementary Table 6: CRISPRqtl identifies genes regulated by breast cancer risk enhancers. Chromatin interactions between BC risk signals and genes identified using CRISPRqtl.

Supplementary Table 7: Gene Z-scores obtained following CRISPRko deletion of the indicated gene. Each score is the mean of three sgRNAs. mRNA expression profiles following CRISPR deletion.

Supplementary Table 8: L1000 analysis identifies opportunities for drug repurposing. Drugs identified in L1000 analysis that correlate or anti-correlate with expression signatures of genes that score as hits in functional screens.

Supplementary Table 9: Sequences of sgRNAs used in validation assays.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryTable1.xlsx](#)
- [ExtendedDataFigure1.jpg](#)
- [ExtendedDataFigure10.jpg](#)
- [ExtendedDataFigure11.jpg](#)
- [ExtendedDataFigure12.jpg](#)
- [ExtendedDataFigure2.jpg](#)
- [ExtendedDataFigure3.jpg](#)
- [ExtendedDataFigure4.jpg](#)
- [SupplementaryTable9.xlsx](#)
- [ExtendedDataFigure5.jpg](#)
- [ExtendedDataFigure6.jpg](#)
- [SupplementaryTable5.xlsx](#)
- [ExtendedDataFigure7.jpg](#)
- [SupplementaryTable8.xlsx](#)
- [ExtendedDataFigure8.jpg](#)
- [SupplementaryTable6.xlsx](#)
- [ExtendedDataFigure9.jpg](#)
- [SupplementaryTable4.xlsx](#)
- [SupplementaryTable7.xlsx](#)
- [SupplementaryTable3.xlsx](#)
- [SupplementaryTable2.xlsx](#)