UCLA
UCLA Previously Published Works

Title
Predicting master transcription factors from pan-cancer expression data

Permalink
https://escholarship.org/uc/item/6370g8qg

Journal
Science Advances, 7(48)

ISSN
2375-2548

Authors
Reddy, Jessica
Fonseca, Marcos AS
Corona, Rosario I
et al.

Publication Date
2021-11-26

DOI
10.1126/sciadv.abf6123

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

CANCER

Predicting master transcription factors from pan-cancer expression data

Jessica Reddy[1,2]‡, Marcos A. S. Fonseca[1,2]†, Rosario I. Corona[1,2,3]†, Robbin Nameki[1,2], Felipe Segato Dezem[1,2,3], Isaac A. Klein[4,5]§, Heidi Chang[1,2], Daniele Chaves-Moreira[6], Lena K. Afeyan[4,7], Tathiane M. Malta[8]‖, Xianzhi Lin[1,2], Forough Abbasi[1,2,3], Alba Font-Tello[9], Thais Sabedot[8], Paloma Cejas[9], Norma Rodríguez-Malavé[3], Ji-Heui Seo[5,9], De-Chen Lin[10], Ursula Matulonis[11], Beth Y. Karlan[1,2,12], Simon A. Gayther[3], Bogdan Pasaniuc[13,14,15,16], Alexander Gusev[9,17], Houtan Noushmehr[8], Henry Long[9], Matthew L. Freedman[5,9,18], Ronny Drapkin[6], Richard A. Young[4,19], Brian J. Abraham[20]*¶, Kate Lawrenson[1,2,3]*¶

Copyright © 2021 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution NonCommercial License 4.0 (CC BY-NC).

Critical developmental "master transcription factors" (MTFs) can be subverted during tumorigenesis to control oncogenic transcriptional programs. Current approaches to identifying MTFs rely on ChIP-seq data, which is unavailable for many cancers. We developed the CaCTS (Cancer Core Transcription factor Specificity) algorithm to prioritize candidate MTFs using pan-cancer RNA sequencing data. CaCTS identified candidate MTFs across 34 tumor types and 140 subtypes including predictions for cancer types/subtypes for which MTFs are unknown, including e.g. PAX8, SOX17, and MECOM as candidates in ovarian cancer (OvCa). In OvCa cells, consistent with known MTF properties, these factors are required for viability, lie proximal to superenhancers, co-occupy regulatory elements globally, co-bind loci encoding OvCa biomarkers, and are sensitive to pharmacologic inhibition of transcription. Our predictions of MTFs, especially for tumor types with limited understanding of transcriptional drivers, pave the way to therapeutic targeting of MTFs in a broad spectrum of cancers.

INTRODUCTION

Accumulating evidence indicates that tumor cells are driven by small sets of transcription factors (TFs) that control global gene expression programs (*1*–*3*). Often, tumor-driving master TFs (MTFs) are developmental regulators that are aberrantly expressed and functionally co-opted to regulate tumor cell states. For example, regulators of T cell development—TAL1, GATA3, RUNX1, and MYB—are highly expressed and coregulate oncogenic programs in T cell acute lymphoblastic leukemias (*1*, *2*). In addition, developmental regulators MYCN, HAND2, ISL1, PHOX2B, GATA3, and TBX2 have been identified as MTFs in neuroblastoma (*3*). MTFs are typically only expressed in a limited number of cell types, consistent with their potent role in establishing gene expression programs driving distinctive cell identities (*4*, *5*). MTFs are a class of promising therapeutic targets as they are selective essentialities in cancer cells because of a phenomenon termed transcriptional oncogene addiction (*6*).

Although TFs are notoriously difficult to directly target with small molecules, several MTFs have been shown to be highly sensitive to chemical inhibition of general transcriptional regulators, including those that target bromodomain (BRD)–containing proteins and transcriptional cyclin-dependent kinases 7/12/13 (CDK7/12/13) (*7*–*9*). The expression of tumor cell MTFs is often driven by large clusters of enhancers, termed superenhancers (SEs) or stretch enhancers (*10*, *11*). The exquisite sensitivity of these factors to chemical perturbation of BRDs and transcriptional CDKs is hypothesized to result from disruption of continuous, high-level transcription driven by SEs, compounded by short TF transcript half-lives and autoregulation. Together, studies on MTFs and transcriptional inhibition in tumor cells demonstrate the importance of identifying these critical factors and studying their responses to small molecules targeting general regulators of transcription.

MTFs are thought to form core transcriptional regulatory circuitries by co-occupying genomic sites, particularly at SEs, and coregulating the expression of MTF genes and other genes critical for cell identity (*1*, *3*, *12*). Presently, the main approaches to identifying MTFs computationally model these circuitries by identifying autoregulatory, interconnected networks of TFs (*13*). These analyses require performing chromatin immunoprecipitation sequencing (ChIP-seq)

[1]Women's Cancer Research Program at the Samuel Oschin Comprehensive Cancer Center, Cedars-Sinai Medical Center, Los Angeles, CA, USA. [2]Division of Gynecologic Oncology, Department of Obstetrics and Gynecology, Cedars-Sinai Medical Center, Los Angeles, CA, USA. [3]Center for Bioinformatics and Functional Genomics, Cedars-Sinai Medical Center, Los Angeles, CA, USA. [4]Whitehead Institute for Biomedical Research, Cambridge, MA, USA. [5]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, USA. [6]Penn Ovarian Cancer Research Center, University of Pennsylvania, Philadelphia, PA, USA. [7]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA. [8]Henry Ford Hospital, Detroit, MI, USA. [9]Center for Functional Cancer Epigenetics, Dana-Farber Cancer Institute, Boston, MA, USA. [10]Department of Medicine, Cedars-Sinai Medical Center, Los Angeles, CA, USA. [11]Division of Gynecologic Oncology, Dana Farber Cancer Institute, Boston, MA, USA. [12]Cancer Population Genetics, Jonsson Comprehensive Cancer Center, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA. [13]Bioinformatics Interdepartmental Program, University of California, Los Angeles, Los Angeles, CA, USA. [14]Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA. [15]Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA. [16]Department of Computational Medicine, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA. [17]McGraw/Patterson Center for Population Sciences, Dana-Farber Cancer Institute, Boston, MA, USA. [18]The Eli and Edythe L. Broad Institute, Cambridge, MA, USA. [19]Department of Biology, M.I.T., Cambridge, MA, USA. [20]Department of Computational Biology, St. Jude Children's Research Hospital, Memphis, TN, USA.
*Corresponding author. Email: brian.abraham@stjude.org (B.J.A.); kate.lawrenson@cshs.org (K.L.)
†These authors contributed equally to this work.
‡Present address: Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, MA, USA.
§Present address: Dewpoint Therapeutics, Boston, MA, USA.
‖Present address: University of Sao Paulo, Ribeirao Preto, Brazil.
¶These authors jointly directed the study.

experiments to map enhancers and identifying SE-associated TFs whose predicted binding motifs are enriched at SEs, both upstream and regulating other MTFs (*14*). These approaches have been shown to recover experimentally validated MTFs in various tumor types but are limited to settings with large numbers of generally homogenous cells (*12*, *15*, *16*). Other approaches such as motif enrichment in accessible regions identified by assay for transposase-accessible chromatin using sequencing (ATAC-seq) are not only possible but also suffer from low dynamic ranges that render SE identification impracticable and the imprecision in motif occurrences between members of the same TF family.

Obtaining adequate amounts of primary tumor cells for ChIP-seq experiments can be technically challenging. RNA sequencing (RNA-seq) experiments, however, require less starting material, and, accordingly, RNA-seq data from primary tumor samples are currently more widely available, especially for rare tumor types and subtypes. We therefore developed an approach to predicting tumor MTFs using RNA-seq data from The Cancer Genome Atlas (TCGA). This approach is called the Cancer Core Transcription factor Specificity (CaCTS) algorithm, and it attempts to determine MTFs by identifying those TFs exhibiting high levels of absolute expression in combination with tumor-specific expression compared to a background dataset that contains a diverse set of cancer types. A conceptually similar approach has been previously applied to gene expression microarray profiles of normal tissues (*5*) and subtype determination in specific tumor types (*17*). We find that candidate MTFs identified through the CaCTS algorithm have many expected qualities of MTFs, such as SE association and high levels of essentiality, indicating that our approach is an orthogonal metric to existing attempts to predict MTFs. This unique resource represents a collection of candidate MTFs for 34 tumor types and 140 molecular and histologic subtypes that can be directly explored for therapeutic potential.

RESULTS

The CaCTS algorithm identifies factors with features attributed to tumor cell MTFs

Many known tumor cell MTFs are developmental regulators that exhibit cell type–specific expression (*1*–*3*). We hypothesized that similarly potent MTFs could be retrieved in poorly studied lineages by identifying sets of TFs that exhibit tumor type–specific RNA expression. We developed the CaCTS algorithm, which uses an entropy-based measure of Jensen-Shannon divergence (JSD) (*5*) to identify factors highly expressed in a given tumor type relative to a diverse collection of cancers, applying the approach to a pan-cancer RNA-seq dataset from TCGA. This dataset contains 9691 patient samples representing 34 tumor types (Fig. 1A and table S1) (*18*). We calculated the average expression levels of 1578 TFs in 34 tumor types/major subtypes (table S2). The specificity of expression of each TF, or "CaCTS score," was calculated by comparing its expression level in the query tumor type to that in the remaining 33 tumor types. A high CaCTS score is therefore assigned to factors with high-level expression in the query tumor type as compared to the remaining background datasets (for example, TF_1 depicted in Fig. 1A compared to TF_2, which represents a factor that is ubiquitously expressed across the cohort). The output of the CaCTS algorithm is a list of all TFs ranked by CaCTS scores in each of the 34 tumor types (table S3).

Association with SEs is a feature of MTFs (*8*, *19*–*21*). To provide a measure of confidence in our CaCTS predictions, we tested for SE association in a few systems where the relevant data were publicly available. We curated a list of 17 experimentally validated MTFs in B cell lymphoma [lymphoid neoplasm diffuse large B cell lymphoma (DLBC)] (*8*), melanoma [skin cutaneous melanoma (SKCM)] (*20*), esophageal squamous carcinoma (ESSC) (*19*), esophageal adenocarcinoma (ESAD) (*22*), and breast cancer (BRCA) (*23*, *24*) (table S4), which served as positive controls in the development of our algorithm. By analyzing publicly available H3K27ac ChIP-seq data, we confirmed that 15 (of 17) positive controls are associated with SEs in the relevant tissue types (Fig. 1B and fig. S1, A and B). Fifteen (of 17) positive controls were scored highly by the CaCTS algorithm (within the top 5% in the relevant tumor type) (Fig. 1C, fig. S1C, and table S4). These findings demonstrate that the CaCTS algorithm efficiently recovers known MTFs and suggests a reasonable, empirical cutoff for determining high-confidence MTFs, i.e., top 5% in a given sample.

We next sought to determine whether SE-associated TFs are highly ranked by the CaCTS algorithm in general. We collected 92 publicly available and in-house H3K27ac ChIP-seq datasets, collecting an average of 4.5 (range 1 to 25) samples to represent 20 of the 34 tumor types (table S5). Fourteen of these were from primary tumors, while only cell line data were available for bladder urothelial carcinoma (BLCA), cervical squamous cell carcinoma (CESC), lung adenocarcinoma (LUAD), sarcoma (SARC), ESAD, and ESSC. We were unable to retrieve H3K27ac ChIP-seq datasets for adrenocortical carcinoma (ACC), cholangiocarcinoma (CHOL), head and neck squamous cell carcinoma (HNSC), kidney chromophobe (KICH), kidney renal papillary cell carcinoma (KIRP), lung squamous cell carcinoma (LUSC), mesothelioma (MESO), stomach adenocarcinoma (STAD), testicular germ cell tumor (TGCT), pheochromocytoma and paraganglioma (PCPG), thyroid carcinoma (THCA), thymoma (THYM), uterine carcinosarcoma (UCS), and uveal melanoma (UVM), so these tumor types were not included in this analysis. In each dataset, we performed peak calling, SE identification, and gene set enrichment analysis (GSEA) to determine whether factors with high CaCTS scores tend to be enriched for SE-associated TFs. The majority (81 of 92, 88%) of comparisons between SE and CaCTS ranks showed significant enrichment ($P_{GSEA} < 0.05$) (Fig. 1D and fig. S1D), demonstrating that TFs with high CaCTS ranks are enriched for SE-associated factors in these 20 tumor types and suggesting that bona fide MTFs are among the high CaCTS scoring factors in the 14 tumor types where enhancer data were not available.

Similar to the CaCTS score, most (13 of 17) known MTFs were among the top 5% of expressed factors in the relevant tumor type (table S4) (*8*, *19*, *20*). Therefore, to arrive at a collection of candidate MTFs, we retrieved factors with high CaCTS scores (top 5%; CaCTS rank ≤ 79) that were also highly expressed (within the top 5% of expression; expression rank ≤ 79) for each of the 34 tumor types (table S6 and fig. S1E). This filter removed 2170 instances of TFs with lower-level (outside of the 5% top-ranked TFs), tissue-restricted patterns of expression, as these factors are less likely to be master regulators. The final list contained a total of 273 candidate factors with an average of 8.0 candidate factors per tumor type (range 3 to 31). We note that 224 highly expressed factors with low specificity scores (outside of the top 5% JSD rank) represent candidate ubiquitous MTFs that may play vital roles across multiple cancer types (table S7). Similarly, highly specific but lowly expressed (outside of the top 5% by expression) genes may also be MTFs, but less is known of

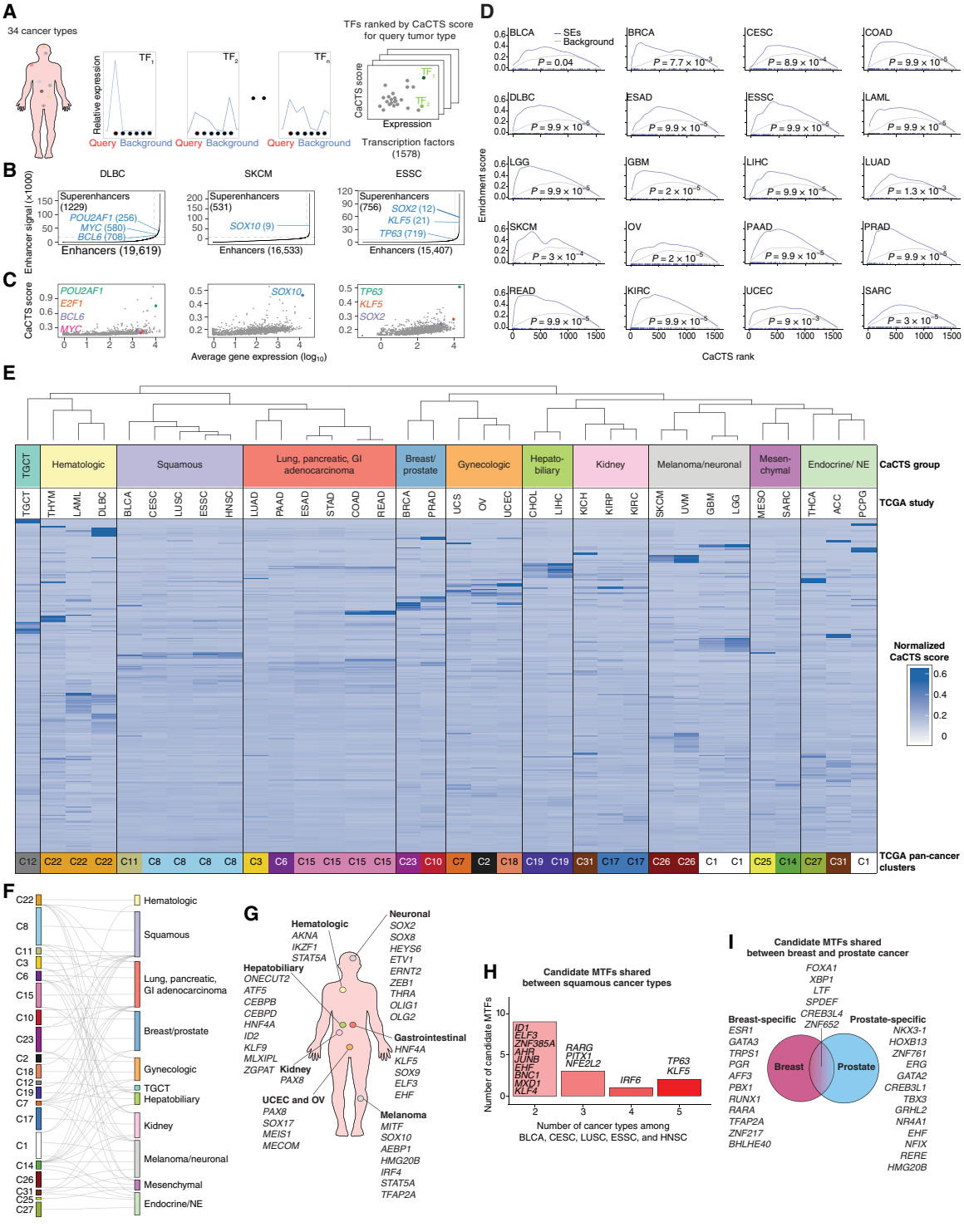

Fig. 1. A multicancer compendium of candidate MTFs. (**A**) Schematic of the CaCTS algorithm. (**B**) Positive control factors—*POU2AF1*, MYC, and BCL6 in DLBC; SOX10 in SKCM and SOX2; KLF5 and TP63 in ESSC—coincide with large SEs in the relevant tumor tissues. (**C**) These factors are highly expressed and have high CaCTS scores. (**D**) Across 20 tumor types, candidate MTFs with high CaCTS scores are significantly enriched at SEs in the corresponding tumor types. Gene set enrichment *P* values are shown, and analyses were performed with 10,000 permutations of randomly selected factors. (**E**) Unsupervised hierarchical clustering based on candidate MTF CaCTS scores across 34 tumor types, using the Spearman method and complete distance parameters. A height cutoff equal to 0.63 defines 11 CaCTS clusters. TGCT, testicular germ cell tumor; GI, gastro-intestinal; NE, neuroendocrine. (**F**) Sankey plot showing 20 pan-cancer clusters that correspond to the 11 CaCTS clusters. We identified the TCGA cluster assigned in 50% or more of the tumors in each of our 11 clusters. (**G**) Factors are often shared across tumors derived from organ systems with a shared development lineage. (**H**) Squamous tumors from diverse organ sites share keratinocyte differentiation TFs as candidates. (**I**) Breast and prostate adenocarcinomas share six candidate MTFs.

these TFs and their capacity to regulate the tens of thousands of genes within the overall expression program.

We asked how TFs were nominated by CaCTS compared to those identified by two alternative methods: first, a bootstrapping method to assess the statistical significance of the CaCTS scores of our nominated MTFs and, second, a classical differential expression test. First, for the bootstrapping approach, we generated 1000 permutations of cancer type assignments to each sample, breaking the relationship of TFs to cancer type. Then, we recalculated the average TF expression and CaCTS score for each randomized cancer type. Using these 1000 iterations, we generated an empirical distribution for the CaCTS scores for each TF for each cancer type and calculated the statistical significance (*P* value) of the observed TF CaCTS score for a given cancer type. The same absolute expression filter was applied (expression rank ≤ 79). We observed that most (514 of 516) CaCTS candidate MTFs have a CaCTS score with high statistical significance [false discovery rate (FDR) = 5%], while the remaining two (*FOSB* for MESO and *ELF3* for CHOL) are marginally significant (FDR = 10%; fig. S1F). We also observed that the statistical significance is correlated with the CaCTS score, with high CaCTS scores corresponding to the lowest *P* values (fig. S1G). For the differential expression test, we contrasted the TF expression in each tumor type individually to the average expression across the other 33 tumor types, identifying a total of 2686 candidates (FDR ≤ 5%; table S8). As with the CaCTS score and bootstrapping, there are TFs with statistically significant *P* values but low average expression, and, so, a cutoff for high expression is still required to discriminate candidate MTFs with both high differential and absolute expression. Four hundred seventy-four CaCTS candidates were identified by this method (table S8). We noted that both the bootstrapping and differential expression analyses nominated around five times more candidates than CaCTS, indicating that a major advantage of the CaCTS approach comes from the more manageable number of credible candidates prioritized.

On average, an individual factor passed the significance thresholds for being a candidate MTF for 1.9 tumor types (fig. S2A), consistent with our expectation that cancer MTFs will be enriched for developmental regulators with tissue-specific patterns of expression. Unsupervised hierarchical clustering of tumors based on CaCTS scores of the 273 TFs ever nominated as a candidate MTF in our tumor collection identified 11 clusters (Fig. 1E). MTFs identified in TGCTs exhibited the highest average CaCTS scores (TGCT average CaCTS score = 0.91 ± 0.73; compared to an average of 0.36 ± 0.25 across the whole cohort) and included known TGCT markers *NANOG* (ranked number 1; CaCTS score = 2.4) and *POU5F1* (ranked number 5; CaCTS score = 1.51). Most tumors cluster by organ site, which is expected because the expression of lineage-specific factors should be similar among tumors originating from related sites. For example, we identified a cluster consisting of ectoderm-derived adenocarcinomas from the lung, pancreas, esophagus, stomach, colon, and rectum (Fig. 1E). Our clusters were largely consistent with those defined by unsupervised consensus clustering based on the expression of ~15,000 genes in 10,165 tumor samples by TCGA (Fig. 1F and fig. S2B) (*25*), suggesting that our predictions largely distill global expression programs and might thus contain key drivers of those expression programs.

There were some notable differences between TCGA pan-cancer clusters and our own. In our clustering, urothelial BLCA clusters with squamous tumors, while lung adenocarcinoma (LUAD) and pancreatic adenocarcinoma (PAAD) now cluster with other gastrointestinal solid tumors. This argues that common factors or related factors may manifest in different expression programs depending on cellular context.

A subset of candidate MTFs are shared among tumors of similar anatomic or functional state

A subset of 62 MTFs was shared among three or more tumor types (fig. S2A), so we sought to further examine commonalities across the union set of candidate MTFs (Fig. 1, E and G to I). In a cluster of predominantly squamous tumors, two factors, *TP63* and *KLF5*, were common candidates among five tumors from diverse anatomic sites, bladder, cervix, lung, esophagus, and head and neck (Fig. 1, E and H). Similarly, six factors were shared between breast and prostate cancer, both derived from hormone-responsive organs (*FOXA1*, *XBP1*, *LTF*, *SPDEF*, *CREB3L4*, and *ZNF652*) (Fig. 1I). *FOXA1* is a critical player in breast and prostate cancer risk and somatic development (*26*, *27*); *XBP1* is involved in the unfolded protein response and MYC signaling in both cancer types (*28*); and *SPDEF* may function not only as a tumor suppressor gene in prostate cancer but also as an oncogene in breast cancer, where it regulates expression of lineage-specific genes in mammary luminal epithelial cells (*29*, *30*). *LTF*, *CREB3L4*, and *ZNF652* have been less well studied and warrant further investigation as BRCA and prostate adenocarcinoma (PRAD) candidate MTFs. While breast and prostate cancer closely clustered with gynecologic tumors arising from other hormone-responsive organs, these tumor types shared more candidate factors with each other than with ovarian serous cystadenocarcinoma (OV), UCS, and uterine corpus endometrial carcinoma (UCEC). This is consistent with germline susceptibility studies that report greater pleiotropy between prostate and breast cancer risk than between prostate and ovary (*31*, *32*). These functional groupings in the clustering (Fig. 1E) were less expected and suggest that shared MTFs across diverse tissue types are responsible for specific cellular functions in different cell identity contexts and differentiation states.

Candidate MTFs represent tumor cell dependencies

Tumor cell MTFs are generally required for cellular viability (*3*, *6*, *17*), and, so, we determined the extent to which loss of function of our candidate factors affect tumor cell viability by examining CRISPR-Cas9 screening data from The Cancer Dependency Map Consortium (DepMap; depmap.org) (*33*). CRISPR knockout screening data were available for 20 of the 34 tumor types in our study, with a mean of 21.7 cell lines per tumor type (range 1 to 31), and 434 cell lines in total. Dependency scores (CERES) are normalized such that −1 corresponds to the median effects of pan-essential genes (*33*). Dependency scores for the candidate MTFs in relevant cell line models are depicted in Fig. 2 (A to T). Eighty-two percent of (14 of 17) positive-control MTFs show modest dependency (minimum CERES score ≤ −0.4) and 47% (8 of 17) show high levels of dependency (minimum CERES score ≤ −1.0) in the relevant cancer type (fig. S3A). In these 20 tumor types, we found that 38.6% (152 of 394) of candidate MTFs showed at least modest dependency (Fig. 2U). Conversely, when we consider each factor individually, 63% (248 of 394) of candidate MTFs exhibit a minimum CERES score below the median CERES score across all tumor types (fig. S3B). This included previously unidentified factors such as *FOXM1* in ESSC (minimum CERES score = −0.85), previously implicated in this cancer type (*34*) although not as an MTF, and *PREB* in liver hepatocellular carcinoma (LIHC) (minimum CERES score = −0.99), a factor that regulates

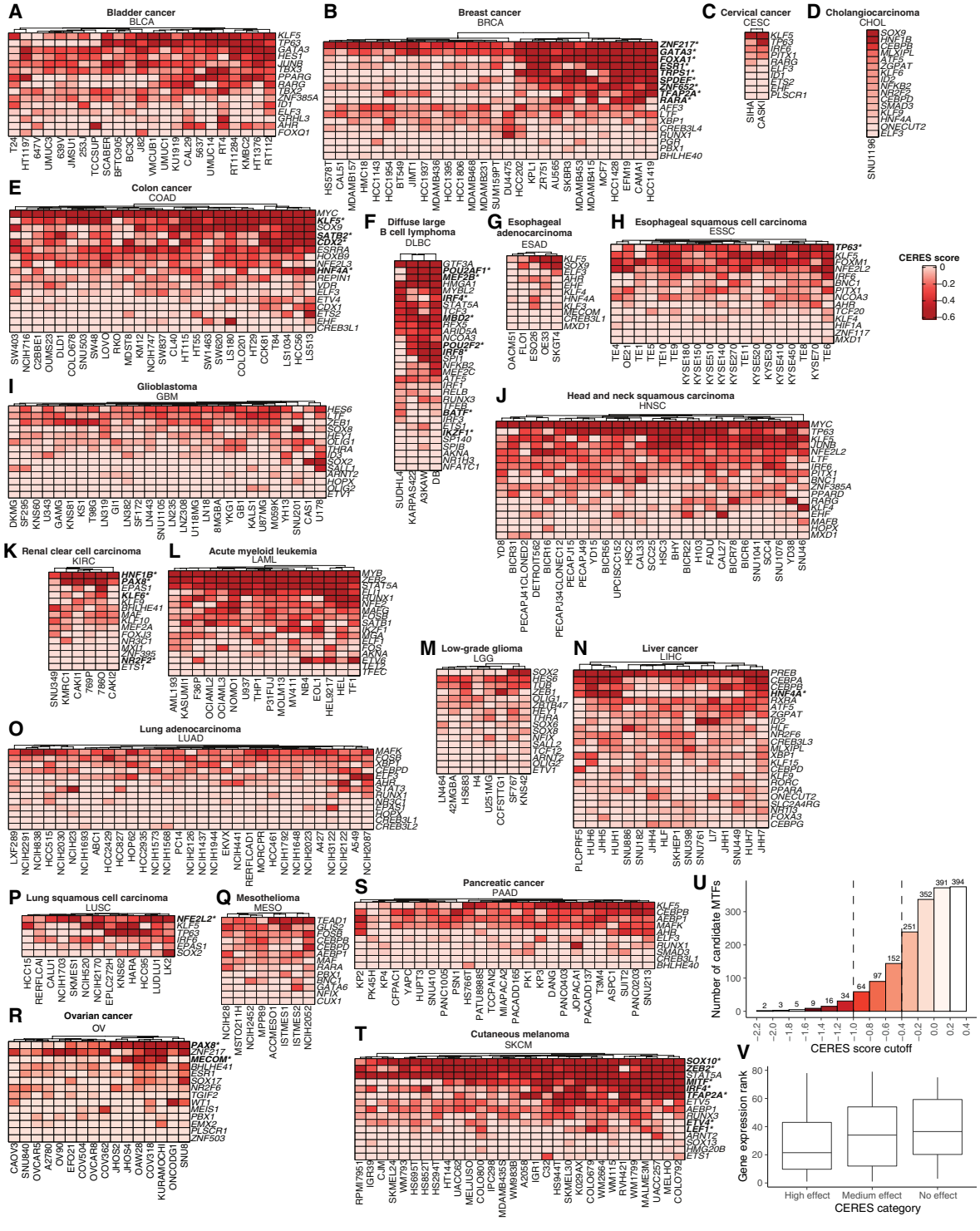

Fig. 2. Candidate MTFs are often essential genes. Dependency scores for MTF knockouts. Darker red color denotes higher levels of essentiality. Unsupervised hierarchical clustering was used to arrange cell lines (columns) and MTFs (rows). (**A** to **T**) Dependencies for MTF candidates across 20 tumor types. For nine tumor types (DLBC, SKCM, ESSC, BRCA, COAD, KIRC, LIHC, LUSC, and OV), lineage-specific dependency data were available. Factors that are lineage-specific dependencies in the relevant tumor type are indicated by bold fonts and an asterisk. Data were curated from depmap.org. (**U**) Cumulative number of CaCTS candidate MTFs with minimum dependency (CERES) scores of ≤−2.0 to 0.4. (**V**) Expression rank for CaCTS candidate MTFs stratified by dependency category: high effect, minimum CERES score < −0.4; medium effect, minimum CERES score between −0.4 and 0; no effect, minimum CERES score > 0.

prolactin expression and glucose homeostasis in the liver (*35*). We note that *PREB* has an average normalized expression level of 2907.3 (expression rank = 41), while *CEBPG* or *FOXA3* knockouts have more modest effect (minimum CERES scores of −0.162 and −0.144, respectively) and lower average normalized gene expression (2542.3 and 1860.6, expression ranks of 49 and 74, respectively) than *PREB*. Overall, gene expression levels of candidate MTFs with high dependencies tended to be higher than those with medium or low dependency scores ($P = 0.002$ for both tests, one-sided t test; Fig. 2V).

Lineage-specific dependency data were available for nine tumor types [DLBC, SKCM, ESSC, BRCA, COAD (colon adenocarcinoma), KIRC (kidney renal clear cell carcinoma), LIHC, LUSC, and OV]; 4 to 53% of CaCTS candidates in each cancer type were also lineage-specific TF dependencies in the relevant tumor cell lines. This was particularly notable for BRCA, where 9 of 17 candidates (53%) were dependencies specifically enriched in breast cancer (Fig. 2B). In general, we found that CaCTS candidate MTFs have more impressive CERES scores than other TFs or non-TF genes in 44% (12 of 27; one-sided Mann-Whitney U test, $P < 0.05$) of cancer types, where dependency data were available.

Candidate MTFs are targets of somatic mutations in cancer

We tested whether candidate MTFs were enriched in somatic mutations in cancer. We calculated the mutation rate across all coding exons and identified significantly frequently mutated genes ($P < 0.05$) in all available tumor types, using data for somatic single-nucleotide variants (SNVs) from the PanCancer Analysis of Whole Genomes (PCAWG) project (https://icgc.org). Data were available for 21 (of 34) cancer types represented in Fig. 1E. We compared the mutation frequencies of candidate MTFs to TFs with comparably high levels of expression (within the top 5% of all TFs; expression rank ≤ 79) but low CaCTS scores (CaCTS rank > 79) (table S7). Overall, candidate MTFs were significantly more likely to be mutated than noncandidate MTFs that were highly expressed in the same tumor type ($P = 1.0 \times 10^{-4}$, Pearson's chi-square test; Fig. 3A), with this effect particularly evident for nine tumor types, including BRCA, KICH, LUSC, and PAAD (fig. S4). Frequently mutated MTFs (Table 1) included factors previously known to be somatically mutated in a tumor-specific manner, *FOXA1* in BRCA and PRAD (*36*, *37*) and *NFE2L2* in squamous lung tumors (*38*). *TRPS1*, a known essential gene and lineage-specific factor in breast cancer, was mutated in 11 of 195 breast cancer cases ($P = 4.5 \times 10^{-3}$); other previously unidentified factors with significant mutation frequencies included *TBX3* in bladder tumors (mutated in 5 of the 23 cases; $P = 1.8 \times 10^{-3}$) and *PAX8* in UCEC (mutated in 4 of 44 cases; $P = 0.01$). Similarly, candidate MTFs are also more likely to coincide with regions of copy number gain than noncandidate, highly expressed TFs ($P = 8.8 \times 10^{-4}$, Pearson's chi-square test) with 11% (32 of 292) of candidate MTFs with a median copy number gain [median copy number variation (CNV) > 0], compared to only 5.4% (69 of 1269) of noncandidate, highly expressed TFs with median copy number gain (Fig. 3B). *TRPS1* shows high levels of copy number amplification (median CNV = 2), with 78% (42 of 54) breast cancer samples showing amplification (CNV = 1, $n = 12$) or high levels of amplification (CNV = 2, $n = 30$). *ZNF217* is also amplified in breast cancer, with 51.85% (28 of 54) samples showing amplification (CNV = 1, $n = 20$) or high amplification (CNV = 2, $n = 8$). This suggests that our predicted MTFs might become genetically overexpressed and establish oncogenic regulatory circuits.

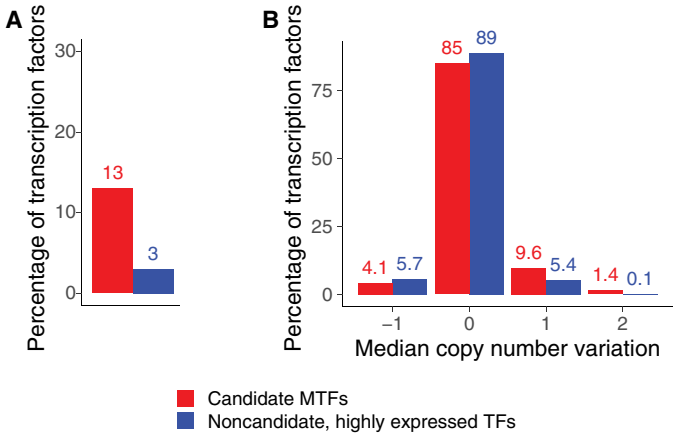

Fig. 3. Somatic mutations in candidate MTFs. (**A**) Percentage of TFs mutated in 21 tumor types comparing CaCTS candidate MTFs (red series) to highly expressed TFs (within the top 5% of all TFs) with low CaCTS scores (blue series; table S7). (**B**) Histogram of median CNV for candidate MTFs (red series) and noncandidate, highly expressed TFs (blue series; table S7).

Subtype-specific MTF predictions reveal subtype-specific regulators

Many tumors characterized by a shared anatomic origin can be stratified into molecular and/or histologic subtypes with markedly different prognoses and responses to therapy using standard diagnostic tools and practices. Subtype annotation was available for 7115 of the 9691 tumors in our original dataset (tables S1 and S9). CaCTS scores were consistent when we used 9691 samples or the subset of 7115 with subtype annotation (fig. S5A). We stratified tumors into subtypes based on molecular features (expression, methylation, coding mutations, or copy number alterations) except in three instances where we used histologic classifications: KICH, SARC, and UCS, as no alternative subclassification strategy has been proposed for these tumor types. To select the most clinically relevant subtype classifications for the remaining tumors, we used TCGAbiolinks (*39*) and primary publications from TCGA (*40*–*53*). Subtype annotations were not available for DLBC and ESSC. Table S9 details on how the 34 tumor groups were stratified into a total of 140 molecular and histologic subtypes.

To use the CaCTS algorithm for the identification of subtype-specific MTFs, we queried the average expression of 1578 TFs in each subtype (table S10) against a background dataset that contained the distribution of TF expression in the other tumor subtypes. To prevent redundancy in the dataset diminishing our sensitivity, samples of the same major tumor type were excluded from the background dataset (see Methods). The CaCTS algorithm identified a total of 439 different candidates across the 140 tumor subtypes; this included all candidates identified in our initial analyses and 166 (38%) factors that were only identified in the subtype-stratified analyses (table S11). For each tumor subtype, we calculated the Jaccard distance to determine how much the candidate MTFs for subtypes deviated from the parent tumor type (mean Jaccard distance = 0.47, SD = 0.22, range = [0.00, 1.00]) (table S12). Subtype stratification had a major impact on the candidates identified for several tumor subtypes, with malignant peripheral nerve sheath tumors ($n = 5$) having the most divergent collection of candidates compared to the parent tumor type (SARC), with zero factors in common. Other subtypes with

Table 1. Somatic mutations of candidate MTFs. MTFs with a significant burden of SNVs across coding exons, compared to all coding exons in the genome (listed in order of significance).

Tumor type	TF	Sample size (*n*)	Mutated samples (*n*)	SNVs (*n*)	*P* value	Adjusted *P* value
LUSC	*NFE2L2*	47	22	25	3.33×10^{-16}	2.00×10^{-15}
PRAD	*FOXA1*	275	11	11	5.13×10^{-9}	9.74×10^{-8}
KIRC	*MAF*	143	8	8	8.25×10^{-4}	0.01
KICH	*FOXI1*	43	2	2	1.46×10^{-3}	0.02
BLCA	*TBX3*	23	5	5	1.82×10^{-3}	0.03
BRCA	*TRPS1*	195	11	13	4.48×10^{-3}	0.04
BRCA	*FOXA1*	195	6	6	5.08×10^{-3}	0.04
READ	*SOX9*	52	9	11	6.15×10^{-3}	0.11
COAD	*SOX9*	52	9	11	6.15×10^{-3}	0.14
UCEC	*MSX1*	44	3	3	8.61×10^{-3}	0.09
PAAD	*CREB3L1*	234	5	5	0.01	0.07
PAAD	*BHLHE40*	234	5	5	0.01	0.07
KICH	*MECOM*	43	2	2	0.01	0.11
UCEC	*PAX8*	44	4	5	0.01	0.09
BLCA	*ELF3*	23	4	5	0.01	0.09
BLCA	*ID1*	23	2	2	0.01	0.09
READ	*MYC*	52	7	10	0.01	0.19
COAD	*MYC*	52	7	10	0.01	0.23
ESAD	*AHR*	97	7	7	0.01	0.32
BRCA	*XBP1*	195	4	7	0.01	0.18
LIHC	*XBP1*	324	6	6	0.01	0.47
LIHC	*RORC*	324	7	8	0.01	0.47
BRCA	*GATA3*	195	4	4	0.01	0.20

large Jaccard distances compared to the parent tumor group included basal-type breast cancers ($n = 189$, Jaccard distance = 0.93) (Fig. 4A), BLCA subgroup 3 (squamous enriched; $n = 31$, Jaccard distance = 0.79) and 4 ($n = 15$, Jaccard distance = 0.95) (Fig. 4B), and the adenocarcinoma-enriched CESC group 3 ($n = 41$, Jaccard distance = 0.81) (Fig. 4C). We compared candidate MTFs for the two predominant subtypes of breast cancer, luminal A (55% of BRCA cases) and basal (30% of cases). Only two factors (*TRPS1* and *LTF*) were shared between these two subtypes, reinforcing existing evidence that these two tumor types represent different cell states and cells of origin (*54*). Hierarchical clustering of predicted MTFs based on dependency scores across a union set of luminal A and basal BRCA candidate factors clearly divided cell lines by subtype (Fig. 4D). Luminal A–specific candidates were not dependencies in basal-type cell lines, and vice versa. We detected known subtype-specific factors including *GATA3* in the luminal A subtype (*27*) and *FOXC1* and *SOX9* in triple-negative breast cancer (TNBC) (which is enriched in the basal subtype) (*9*). Luminal A candidates largely overlapped with the candidates identified in the initial CaCTS analyses, which is to be expected as this comprised more than 50% of the BRCA cohort. We identified four additional candidates for luminal A breast cancer, *AEBP1*, *MYB*, *SREBF1*, and *TBX3*. *TBX3* is recurrently mutated in luminal A tumors (*27*) but has not been implicated as an MTF, and the other factors also represent previously unknown candidate MTFs for this tumor type. Previously unidentified candidates for basal BRCA included *NFIB* (*55*), which has been implicated in epigenetic reprogramming during small cell lung cancer metastasis (*56*), and *CREB3L2*, which is commonly fused to *FUS* in low-grade fibromyxoid SARCs (*57*), but has not been studied in the context of basal-type breast cancer.

When we stratified molecular groups in bladder and cervical cancer, we found that squamous-enriched group BLCA.3 and CESC "keratin" groups CESC.C1 and CESC.C2 now cluster together with other squamous tumor types (LUSC, ESSC, and HNSC), likely because these subtypes have squamous differentiation TFs as candidates (*TP63*, *IRF6*, and *PITX1*). In contrast, adenocarcinoma subgroups from the same organs cluster more distantly (for example, CESC.C3 clusters with uterine and ovarian adenocarcinoma) (fig. S5B) and have distinct MTF candidates (Fig. 4C). Therefore, while squamous subgroups of bladder and cervix tumors cluster with squamous types from distant organs, adenocarcinomas share greater similarities with tumors derived from a similar developmental lineage.

PAX8 and *MECOM* were candidates in CESC.C3 and OV, so we hypothesized that there may be additional factors shared across gynecologic tumor subtypes. Six factors were common across two or more gynecologic tumor types, *PAX8*, *MECOM*, *SOX17*, *ESR1*, *MEIS1*,

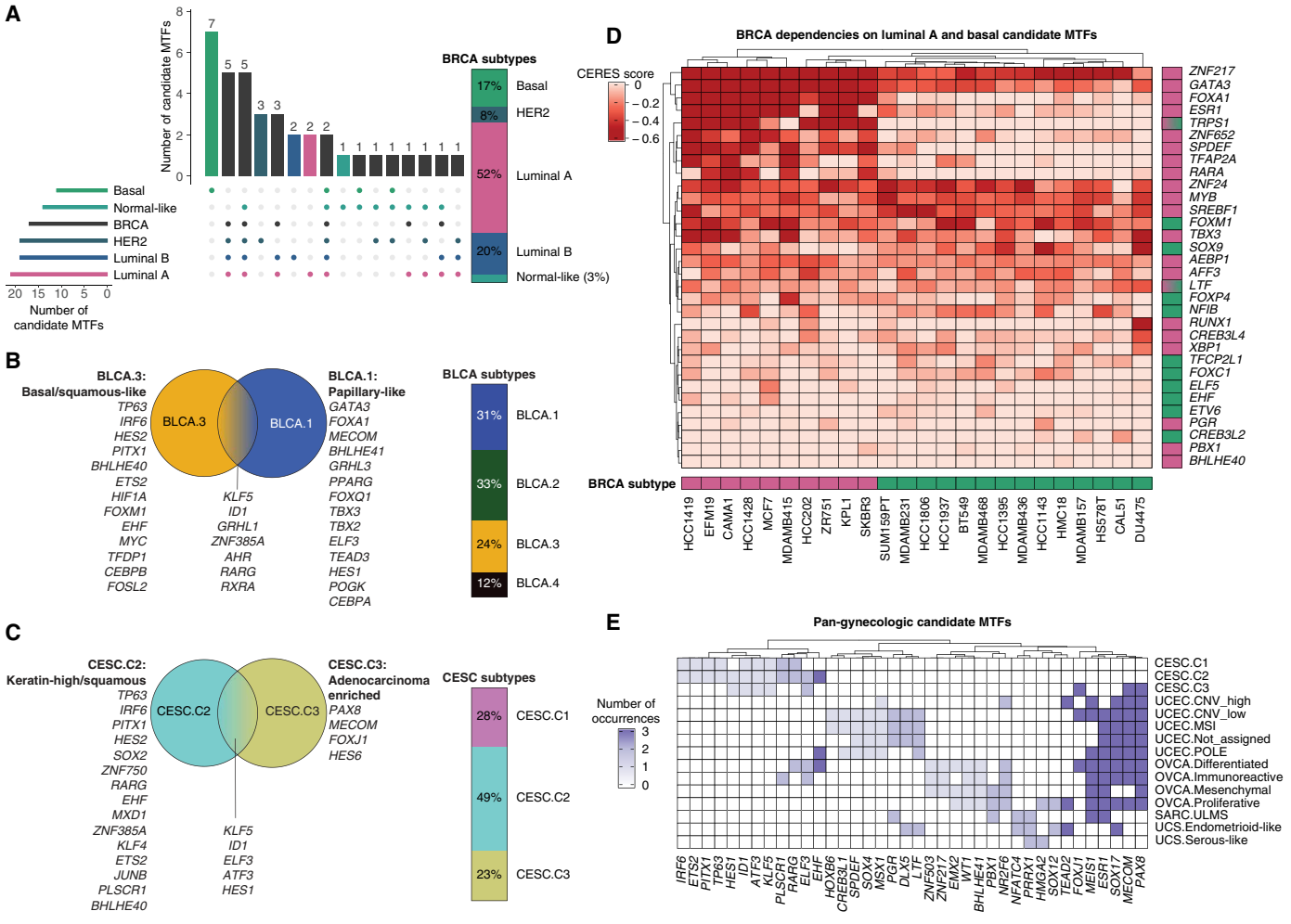

Fig. 4. Candidate MTFs across clinically relevant tumor subtypes. (**A**) Distinct sets of candidate MTFs were identified for tumor subtypes in breast invasive carcinoma (BRCA). UpSet plot shows the overlap of candidates across the groups. (**B**) Distinct sets of candidate MTFs were identified for squamous-enriched tumor subtypes in BLCA and (**C**) CESC. In (A) to (C), the stacked bar plots show the frequency of each subgroup within the overall parent group. (**D**) Unsupervised clustering of luminal and basal breast cancer cell lines based on dependencies on a union set of luminal A and basal candidate MTFs. Cell lines cluster by molecular subtype. Green boxes correspond to luminal A cell lines and luminal A candidate MTFs; fuchsia denotes basal/TNBC cell lines and candidate MTFs. (**E**) Candidate MTFs across pan-gynecologic tumor subtypes.

and *FOXJ1*. *PAX8*, *MECOM*, and *SOX17* were the top three shared candidate MTFs among this set of tumors (Fig. 4E). Molecular subgroups of ovarian and uterine carcinoma exhibited the greatest similarities, with CESC.C3, uterine leiomyosarcoma, and uterine endometrioid-like SARCs sharing 3, 2, and 2 candidate MTFs in common with the uterine/ovarian carcinoma metagroup, respectively. For molecular subtypes of OV, MTF candidates largely mirrored those identified for OV as a whole, with a few differences, the mesenchymal molecular subgroup of ovarian tumors uniquely lacked *SOX17* and *MECOM*, both present in the other OV subtypes; *FOXJ1*, *BCL6*, *EHF*, and *RARG* were candidates for differentiated tumors, but not the other subtypes; *ELF3* was a candidate MTF for the differentiated and immunoreactive OV subtypes only; and proliferative-type tumors had four unique candidates, *HMGA2*, *SOX12*, *TEAD2*, and *PLAGL2*. *HMGA2* is a known, subtype-specific marker for the proliferative subgroup of OV (*58*), and transcriptional enhanced associate domain (TEAD) family proteins likely cooperate with paired box 8 (PAX8) to regulate gene expression in models of ovarian cancer (*59*, *60*).

Using candidate MTFs to build core regulatory circuitry models in ovarian cancers

To validate the CaCTS algorithm, we tested our success at identifying MTFs for OvCa. The TCGA OV study consists exclusively of high-grade serous ovarian cancers (HGSOCs). HGSOCs are relatively rare tumors for which MTFs are currently unknown, and novel therapeutic targets are urgently needed because of the frequent late-stage diagnoses and high rates of recurrence. We identified 14 candidate MTFs for OV (listed in order of ranked CaCTS score): *WT1*, *EMX2*, *SOX17*, *MEIS1*, *BHLHE41*, *PAX8*, *ESR1*, *ZNF503*, *MECOM*, *TGIF2*, *NR2F6*, *PBX1*, *ZNF217*, and *PLSCR1* (Fig. 2R). PAX8 has not previously been characterized as an OvCa MTF but is a known lineage-specific dependency in this tumor type (*61*), and both PAX8 and WT1 are used as clinical biomarkers for serous ovarian carcinomas. We tested whether these factors show orthogonal characteristics of MTFs using ChIP-seq experiments and in vitro knockdown studies.

Ten of these factors were associated with SEs in at least 4 (of 12) primary HGSOC samples (Fig. 5, A and B, and fig. S6), and eight factors—*MEIS1*, *SOX17*, *PAX8*, *WT1*, *ZNF217*, *BHLHE41*, *MECOM*,

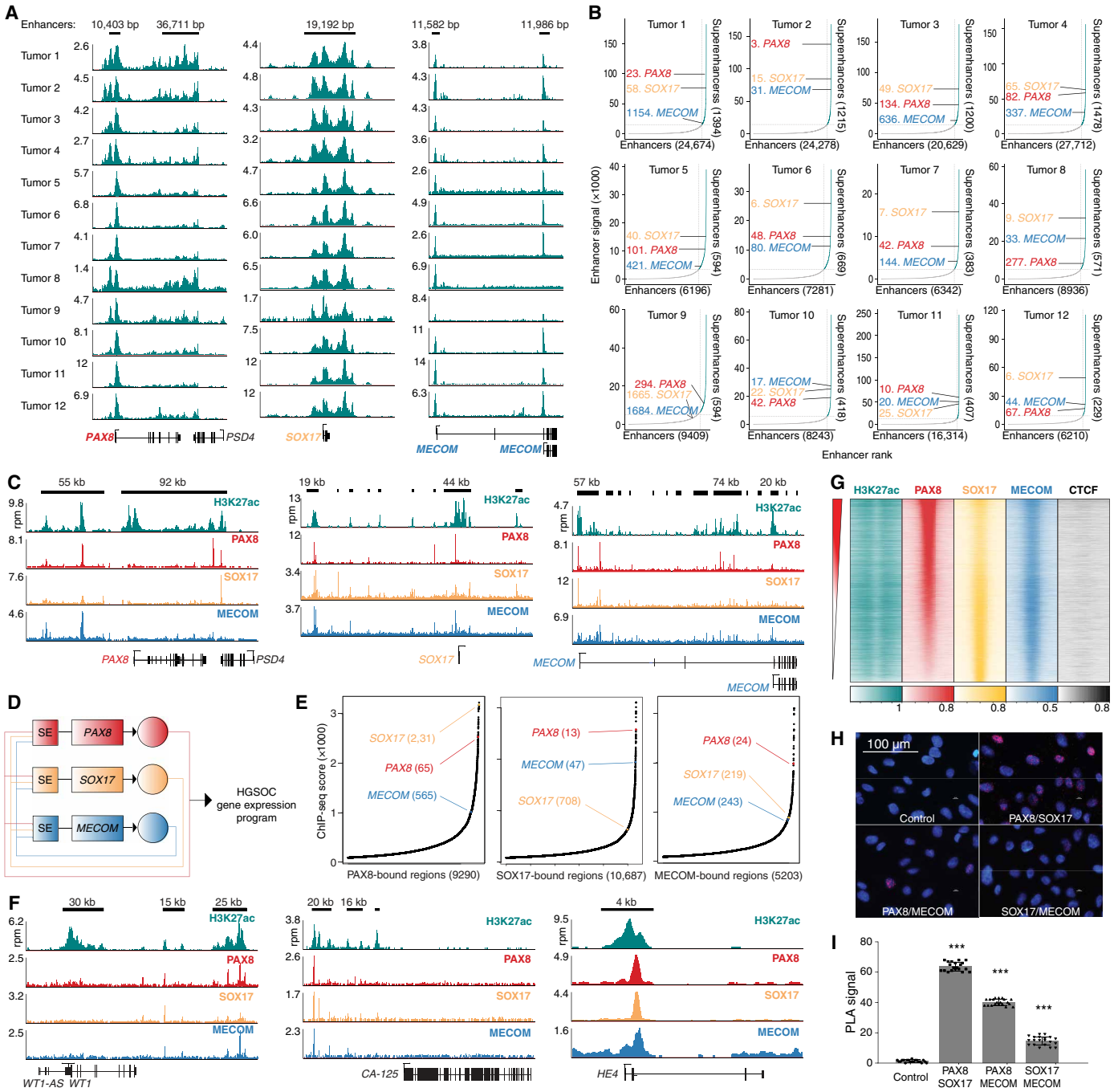

Fig. 5. PAX8, SOX17, and MECOM are candidate MTFs in HGSOC. (**A**) *PAX8*, *SOX17*, and *MECOM* coincide with SEs in primary ovarian tumors. H3K27ac ChIP-seq data were generated in 12 primary tumors, and data were normalized to counts per million (CPM) mapped reads. (**B**) SEs were identified in 12 primary HGSOCs; *PAX8*, *SOX17*, and *MECOM* are associated with SEs in most tumors. Stitched enhancers were ranked by H3K27ac ChIP-seq signal. (**C**) PAX8, SOX17, and MECOM occupy their own and each other's SEs. Enhancers are indicated as black bars. (**D**) Proposed core regulatory circuit for HGSOC based on TF co-occupancy. (**E**) PAX8, SOX17, and MECOM ChIP-seq peaks were ranked by CPM-normalized signal. Strong binding sites for each factor are detected proximal to their own and each other's gene loci. (**F**) PAX8, SOX17, and MECOM bind to enhancers of HGSOC clinical biomarkers. (**G**) Global cobinding of PAX8, SOX17, and MECOM. Each row is a PAX8 ChIP-seq peak. CPM-normalized ChIP-seq reads were plotted for a 4-kb window centered on each binding site. Rows are ordered by decreasing PAX8 signal. (**H**) PLA performed in Kuramochi cells. Each red dot represents a single interaction. Nuclei were counterstained blue with 4′,6-diamidino-2-phenylindole (blue). Scale bar, 100 µm. Image magnification, ×40. (**I**) Quantified PLA signal per cell. In (C) and (E) to (G), ChIP-seq data were generated in Kuramochi cells and normalized to CPM mapped reads. ***$P < 0.001$; Student's paired *t* test. Error bars indicate SD of the mean values from three independent experiments (performed with technical triplicates).

and *PBX1*—were all associated with SEs in at least 11 of the 12 tumors. Notably, *PAX8*, *SOX17*, and *MECOM* are marked by SEs in primary ovarian tumors (minimum SE rank for PAX8 = 3, range = 3 to 294; minimum SE rank for SRY-box transcription factor 17 (*SOX17*); italicize SOX17 as a gene = 6, range = 6 to 65; minimum SE rank for MDS1 and EVI1 complex locus (*MECOM*); italicize MECOM as a gene = 17, range = 17 to 636) (fig. S6). We therefore performed TF ChIP-seq for these three factors in a HGSOC cell line (Kuramochi). PAX8, SOX17, and MECOM proteins occupy regulatory elements at the *PAX8*, *SOX17*, and *MECOM* gene loci, consistent with their formation of a core regulatory circuit (Fig. 5, C and D). The 2nd and 31st top-ranked PAX8 binding peaks were at the *SOX17* gene locus (Fig. 5E). In addition, PAX8, SOX17, and MECOM bind at three clinical biomarkers of the disease: *WT1*, *MUC16* (CA125), and *HE4* (Fig. 5F) (*62*). Core regulatory circuit factors are expected to drive expression programs by co-occupying enhancers across the genome, and we observed a consistent colocalization of these factors genome wide (Fig. 5G). Proximity ligation assays (PLAs) detect protein-protein interactions in situ (*63*) and were performed to test whether PAX8, SOX17, and MECOM physically interact in the nucleus. PLAs performed in human ovarian cancer cells confirmed that these proteins can be part of the same complex (Fig. 5, H and I).

To study the collaboration between PAX8, SOX17, and MECOM in driving tumor cell survival, we analyzed tumor cell survival in their absence. Overall, 5 (of 14; 36%) OV MTF candidates showed moderate to high levels of essentiality in at least one HGSOC cell line (minimum CERES score of −0.4 or less) (Fig. 2R), particularly for PAX8 and MECOM where minimum CERES scores were −1.3 and −0.7, respectively. These two factors are also selective dependencies for OvCa (*61*). PAX8, SOX17, and MECOM dependency correlates with level of expression (Fig. 6A), consistent with a model in which these factors are playing oncogenic roles. In addition, *PAX8*, *SOX17*, and *MECOM* gene loci are amplified in 6, 11, and 36% HGSOCs, respectively, indicative of an oncogenic role for these genes (Fig. 6B). Using RNA interference, we depleted PAX8, SOX17, or MECOM and quantified protein abundance by Western blotting to identify evidence of cross-regulation. For each TF, RNA expression was reduced by 80 to 95% following treatment with the corresponding small interfering RNA (siRNA) pool (Fig. 6C and fig. S7A). PAX8 knockdown resulted in a 60 to 90% reduction of SOX17 and MECOM expression, SOX17 knockdown down-regulated MECOM by 90% but did not affect PAX8 expression, and MECOM knockdown did not affect PAX8 expression but did induce a 90% increase in SOX17 (Fig. 6C). These trends were consistent with gene expression patterns quantified at the RNA level (fig. S7B). These data suggest a transcriptional circuitry that involves both negative and positive cross-regulation (Fig. 6D). Following PAX8, SOX17, and MECOM knockdown, colony formation of HGSOC cells was reduced by 76% (SD = 3.1%; $P < 0.001$), 44% (SD = 30%; $P = 0.013$), and 51% (SD = 17.9%; $P = 0.004$), respectively, in comparison to cells treated with nontargeting siRNA (siNT1; Tukey's multiple comparison test, $n = 4$) (Fig. 6E). Comparisons to a second control (siNT2) showed growth patterns similar to those induced by siNT1 (Fig. 6E); therefore, we found a requirement for maintained PAX8, SOX17, and MECOM expression for HGSOC cell viability.

Targeting an MTF-driven oncogenic expression program in HGSOC

General transcription inhibitors are showing remarkable anticancer effects across multiple cancer types, which are thought to be largely because of their preferential activity toward MTFs (*3*, *9*). OVCAR4 HGSOC cells exhibit notably low median inhibitory concentration (IC_{50}) values of 45 nM for THZ1, 1.3 μM for JQ1, and 97 nM for THZ531, a covalent inhibitor of CDK12 and CDK13 (Fig. 6F) (*64*). Expression of *PAX8*, *SOX17*, and *MECOM* was strongly inhibited by these molecules (Fig. 6G and fig. S7C). *PAX8* and *SOX17* were among the 10% most-sensitive highly expressed transcripts in low-dose (50 nM) treatment with THZ1 (Fig. 6H). The potent inhibition of these factors with low-dose treatment is most relevant in terms of target engagement and the concentration range that selectivity is observed (*7*).

To catalog the genes regulated by our MTFs, we knocked down PAX8, SOX17, and MECOM and performed RNA-seq. PAX8 and SOX17 target genes were largely overlapping, with the most down-regulated genes enriched in pathways associated with cell cycle progression (Fig. 6I), including cell cycle regulators in the retinoblastoma (Rb) pathway such as known ovarian cancer oncogene *CCNE1* (Fig. 6J) (*58*). MECOM, on the other hand, did not regulate the same cell cycle pathways but converged with PAX8 and SOX17 to regulate extracellular organization. MECOM also independently down-regulated genes involved in regulation of insulin-like growth factor, posttranslational protein phosphorylation, and phase 2 conjugation of compounds. Last, we found that target genes of PAX8 and SOX17 phenocopy effects of low-dose THZ1 treatment, suggesting that these factors, at least in part, explain the anticancer effect of this drug in ovarian cancer cells (Fig. 6K). MECOM, however, did not show the same effect. Together, these results indicate that the efficacy of transcriptional inhibitors might be explained by their preferential targeting of the OvCa MTFs PAX8, SOX17, and MECOM.

DISCUSSION

Core regulatory circuitries potentially represent a universal vulnerability in tumor cells, and, consequently, they are likely to represent critical therapeutic opportunities for many cancer types. Given recent developments in targeting core regulatory circuitries through the use of general transcription inhibitors and targeted protein degradation strategies, this approach to nominating candidate master regulators based solely on RNA-seq data is timely, particularly for tumor types where limited access to tumor specimens prohibits the generation of the ChIP-seq data typically required to identify candidate MTFs. The candidate MTFs recovered by the CaCTS algorithm include both known and previously unkown MTFs, which are especially valuable for tumor types in which transcriptional circuits are poorly characterized. As a proof of concept, we performed functional validation and confirmed MTF features for PAX8, SOX17, and MECOM in HGSOC cells, demonstrating that the CaCTS predictions recover previously unidentified critical regulators. In the DepMap cell line dependency data, we observed notable clustering patterns within each tumor type, with factors and cell lines clustering by codependencies for many tumor types, demonstrating the success of the CaCTS approach to identifying transcriptional circuitries. Cell lines that show the greatest dependence on tumor-specific MTFs may be the models that most faithfully maintain the critical features of primary tumors and are therefore superior models to use for translational studies. For example, in HGSOC, Kuramochi, OAW28, and ONCODG1 exhibited similar dependencies on candidate MTFs and have all been prioritized as cell lines models that faithfully recapitulate molecular hallmarks of HGSOC (*65*).

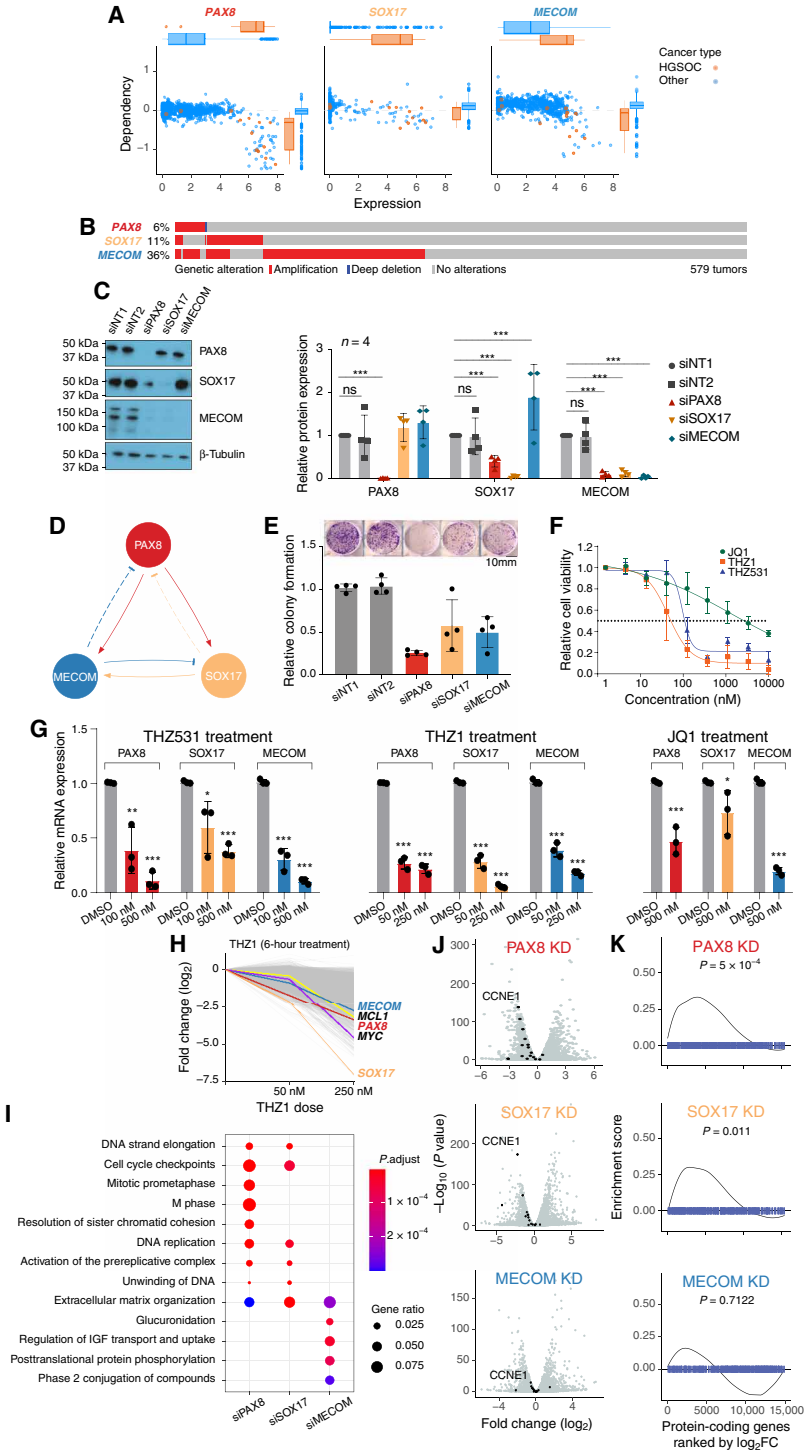

Fig. 6. PAX8, SOX17, and MECOM are functional dependencies in HGSOC. (**A**) PAX8, SOX17, and MECOM are selective dependencies and [their gene loci] are commonly amplified in HGSOC tumors (**B**). (**C**) Western blot, with quantification, to confirm knockdown of PAX8, SOX17, and MECOM in OVCAR4 cells. (**D**) PAX8, SOX17, and MECOM coregulation based on normalized quantification of Western blot signal intensities after knockdown. Linewidths denote the percentage of up-regulation (arrows) or down-regulation (flat ends). Solid lines ($P < 0.05$), dashed lines (not significant) (four independent experiments). (**E**) Anchorage-independent growth assays. (**F**) Dose response curves for THZ1, THZ531, and JQ1 treatment. Nonlinear fit curves are shown. Data are representative of three independent experiments. (**G**) TF expression following a 6-hour drug treatment. (**H**) *PAX8* and *SOX17* are among the most sensitive genes in THZ1-treated Kuramochi cells. (**I**) Ontology analyses for transcripts down-regulated following siPAX8, siSOX17, and siMECOM treatment. IGF, insulin-like growth factor. (**J**) Retinoblastoma (Rb) pathway genes are among the most significantly down-regulated transcripts following PAX8, SOX17, or MECOM knockdown. (**K**) GSEA of the top 500 down-regulated genes following PAX8 and SOX17 knockdown, compared to a ranked list of THZ1-responsive genes. (C to G) *$P < 0.05$, **$P < 0.01$, and ***$P < 0.001$; Student's paired *t* test; ns, not significant. Error bars indicate one SD of the mean values from three independent experiments (performed with technical triplicates).

We predicted candidate factors for major tumor groups, as well as clinically relevant molecular and histologic subtypes. This analysis revealed some common factors across diverse origins, a phenomenon most clearly illustrated by squamous tumors, where tumors across diverse anatomic sites shared three common factors (*TP63*, *KLF5*, and *IRF6*). A further three factors were shared by three sites, *RARG* (BLCA, CESC, and HNSC), *PITX1* (CESC, ESSC, and HNSC), and *NFE2L2* (ESSC, HNSC, and LUSC). In some instances, squamous tumors have both squamous and organ-specific candidate MTFs, such as *ID1* and *KLF5* in cervix and *AHR* and *KLF5* in esophageal tumors, suggestive of dual circuitries with cooperating factors dictating lineage-specific and functional programs.

Using CaCTS, we were able to make MTF predictions for rare and common tumor types that lack publicly available SE or functional dependency data. H3K27ac ChIP-seq data from cell lines, which may not closely recapitulate the epigenetic signatures of disease, are generally more available than from primary tumors, adding an obstacle to existing approaches. Dependency data are limited or absent for 10 tumor types and also generally rely on lines. Overall, around half of the tumor types represented in our study currently have limited publicly available data to predict MTFs, and, for six tumor types, neither H3K27ac ChIP-seq nor dependency data are available—ACCs, KICH tumors, KIRPs, TGCTs, THYM, and UCSs—and so, MTF prediction is not possible on the basis of current methods. Discussion of the MTFs identified for these tumor types are included as Supplementary Note.

Cancer MTFs represent attractive therapeutic targets, given their essential role in governing cell state and the tendency for cancer cells to become highly dependent on maintained high-level expression of a select handful of MTFs (*6*). Consistent with this, many candidate MTFs identified by CaCTS were lineage-specific essentialities in the relevant tumor type. General transcriptional inhibitors, which show preferential activity toward MTFs, offer an efficient approach to anticancer treatment, rather than developing drugs to target each individual MTF; several such agents are now in clinical trials. Another approach to targeting MTFs is more directly with protein degradation strategies, which are currently in preclinical stages of development. Functional validation of the OvCa candidate MTFs—*PAX8*, *SOX17*, and *MECOM*—revealed that each was sensitive to general transcription inhibition, with *PAX8* and *SOX17* particularly vulnerable to CDK7 inhibition with THZ1. This suggests that PAX8 and SOX17 contribute to the antiproliferative effects of this drug, in addition to its previously studied effects on MYC and MCL1 (*66*).

MTFs evince multiple characteristics, including selective dependency and association with SEs. We leveraged high, lineage-restricted expression to identify candidate MTFs, as this is another known characteristic of some known MTFs. While much previous work has shown the roles of MTFs in activating cell type–specific gene expression patterns, other work illustrates that MTFs can both activate genes and/or suppress them in collaboration with coactivators or repressive protein complexes. While MECOM exhibited some properties of MTFs, including functional dependency, genome-wide cobinding, and sensitivity to transcription inhibitors, MECOM was revealed as a negative regulator of *SOX17* expression. Further work is needed to determine whether this is a direct result of loss of MECOM expression, given the time needed to achieve sufficient down-regulation in our system, whether MECOM depletion results in a selection of a subpopulation of high SOX17 expressors, or whether MECOM loss induces an alternate cell identity program.

As PAX8, SOX17, and MECOM are predicted to be MTFs in other tumor types and subtypes, the results in HGSOC models may be applicable to these other tumor types, including gynecologic tumors most similar to OV (UCEC and CESC.C3) and some non-gynecologic tumors, thyroid and kidney carcinomas, ESADs, KICH tumors, and STADs. Early down-regulated genes following MTF depletion included members of the Rb pathway, indicating a mechanism of action for anticancer activity of THZ1 in ovarian cancer models. Critically, the known ovarian cancer oncogene *CCNE1* was down-regulated following MTF depletion. Rb pathway alterations predict responses of patient-derived xenograft models to SY-1365 (*67*), a covalent CDK7 inhibitor currently in clinical trials for advanced breast and ovarian cancer (NCT03134638). While we validated three factors in OvCa, additional MTFs may contribute to the transcriptional circuitry of HGSOC. One such factor is *WT1*, an ovarian cancer biomarker, whose SE was cobound by PAX8, SOX17, and MECOM. Another candidate, *MEIS1*, was the highest-ranking SE-associated MTF for this tumor type. Cells dependent on PAX8 also tended to be dependent on ZNF217, a factor implicated in breast cancer (*23*). We noted for OV and most tumor types we examined that the tumor MTF predictions were very similar to those previously predicted for normal tissues of the same organ (*5*), suggesting that a major mechanism of tumorigenesis involves aberrant reinforcement of developmental transcriptional programs (*20*) or that normal MTF activities become perturbed to acquire oncogenic properties during cancer development (*59*). MTFs predicted by CaCTS are more likely to contain coding somatic mutations or coincide with copy number gain than TFs with comparably high levels of expression but lacking lineage-restricted patterns of expression, suggesting that somatic mutations that bestow pro-oncogenic properties on developmental MTFs are selected for during cancer development. Multicancer MTFs, which can cooperate with lineage-specific factors during tumorigenesis, may be dysregulated by different mechanisms in different systems.

We present a prioritized collection of candidate MTFs for 34 major tumor types and 140 tumor subtypes, which will be enriched for bone fide MTFs. We note that the tumor specificity metric provided by JSD statistic within the CaCTS workflow served to reduce the number of candidate factors to around fivefold compared to approaches based on differential expression of TFs in one tumor type compared to all other tumor types grouped together. Nonetheless, the prioritized candidates likely contain false positive, and so, for users of this resource, we recommend integration of complementary data, where available, such as dependency data and SE landscapes, based on H3K27ac ChIP-seq and other epigenetic data, dependency data, and motif-based circuitry mapping (*14*) to inform the design of functional validation experiments. We caution that not all cancer MTFs will fulfill all the canonical MTF criteria, for example, *ZNF217* has lower levels of expression in basal-type breast tumors (expression rank = 232; 15th percentile) but high levels of dependency (minimum dependency in basal BRCA cell lines = −1.12). Similarly, the androgen receptor was not prioritized as a candidate for prostate cancer because of its expression rank (rank = 371 of 1578 TFs; 24th percentile), although it is known that this factor plays a critical role in transcriptional regulation and development of prostate cancers (*26*). This could be due to poor correlation of transcript expression with functional protein for some factors or by confounding factors, in this case, a signal-responsive ligand. An additional caveat to this approach is that nomination of candidates is directly related to the

composition of the background dataset. TF expression across several of the tumor types was quite similar (for example, in the kidney tumor types), so it is possible that additional tumor type–specific factors exist but were lowly ranked. Despite this, known regulators of kidney tumor types were efficiently retrieved with this analysis, suggesting that our background dataset was sufficiently heterogeneous to minimize this issue. In addition, our analyses were restricted to the pan-cancer dataset compiled by TCGA, so some tumor types are not represented. Metastatic and treatment-resistant tumor states, where new therapies are most urgently needed, are largely absent. Last, because of the limited availability of bone fide positive and negative controls, we were unable to statistically test the positive and negative predictive values for the method. In closing, we present a timely and valuable resource of candidate MTFs for tumor types where transcriptional circuitries are currently unknown and leverage this resource to identify PAX8, SOX17, and MECOM as master regulators for aggressive HGSOCs.

METHODS

Computational methods

Identification of positive-control MTFs

We curated a list of positive-control MTFs using the following criteria: (i) SE association; (ii) cobinding to the genome with another MTF; (iii) MTF gene expression is highly sensitive to treatment with general transcriptional inhibitors JQ1, THZ1, or THZ531; (iv) evidence of cell death or decrease in proliferation upon MTF depletion; and (v) convergence of MTF target genes and genes regulated by general transcription inhibitors. We searched for criteria related to these terms to identify positive-control MTFs for any of the 34 major tumor types published since 2010.

The CaCTS algorithm

PanCancer TCGA RNA sequence level 3 normalized data were downloaded from the Genomic Data Commons (GDC) Data Portal using TCGAbiolinks functions GDCquery, GDCdownload, and GDCprepare and imported into R (www.r-project.org) for analysis (*39*). Table S1 contains the tumor IDs for all the samples included in our analysis. After exclusion of recurrent, metastatic, and nontumor tissues, a total of 9691 samples across 34 tumor types were available. Sample annotations were curated from TCGA publications (*40*, *42*–*53*, *68*) and TCGAbiolinks (*39*) (www.bioconductor.org/packages/devel/bioc/vignettes/TCGAbiolinks/inst/doc/subtypes.html). Tumor types/subtypes included were ACC, BLCA, BRCA, CESC, CHOL, COAD, DLBC, ESCA (esophageal carcinoma), GBM (glioblastoma multiforme), HNSC, KICH, KIRC, KIRP, LAML (acute myeloid leukemia), LGG (brain lower-grade glioma), LIHC, LUAD, LUSC, MESO, OV, PAAD, PCPG, PRAD, READ (rectum adenocarcinoma), SARC, SKCM, STAD, TGCT, THCA, THYM, UCEC, UCS, and UVM. For the main analyses, we preserved all grouping defined by TCGA, apart from for ESCA, which we divided into ESAD and ESSC.

Published lists of TF were retrieved from Saint-André *et al.* (*13*) (1253 TFs) and Lambert *et al.* (*69*) (1639 TFs). Merging both lists created a catalog of 1671 unique TFs, of which 1578 were expressed in the pan-cancer dataset. The 93 TFs not detected in this dataset are listed in table S13. To calculate a JSD score for each TF for each tumor type, we shifted the normalized expression values such that the new minimum normalized expression value is equal to 0.

The CaCTS score ($\mathrm{CaCTS}_{i,j}$) of gene i in cancer type j is a measure of gene expression specificity. It is calculated as follows

$$\mathrm{CaCTS}_{i,j} = -\log_{10} \mathrm{JSD}(\hat{x}_i, \hat{u}_j)$$

where

$$\hat{x}_i = \frac{x_i}{|x_i|}$$

$x_i = (x_{i,k})$ is the ordered vector of normalized gene expression of gene i and length n [n is the number of cancer types, $k \in \{1, n\}$], $\hat{u}_j = (u_{j,k})$ is the idealized cancer type–specific gene expression for cancer type j represented by unit vector of length n such as $u_{j,k} = 1$, if $k = j$, and $u_{j,k} = 0$ otherwise.

The JSD measures the similarity between two probability distributions, here used to measure the similarity between two unit vectors $\hat{x}_i$ and $\hat{u}_j$. The JSD was calculated using the R package jsd (version 0.1). Practically, none of the values in vectors $\hat{x}_i$ and $\hat{u}_j$ should be equal to zero; therefore, we substituted any zeros for 0.1^{-17} in vectors x_i and u_j and then get the unit vectors $\hat{x}_i$ and $\hat{u}_j$ by dividing over $|x_i|$ and $|u_j|$, respectively. The final candidate MTF list for a given cancer type was defined by considering the intersection of the 5% most highly expressed TFs (expression rank ≤ 79) in said tumor type and the TFs in the top 5% when ranked by the CaCTS score.

Bootstrapping and t test methods for the statistical assessment of the CaCTS score

To estimate the statistical significance of the CaCTS score for a particular cancer type, we generated 1000 permutations of sample/cancer type associations. With these random permutations, we then calculated the average expression per cancer type and the CaCTS scores. Using these CaCTS scores, we obtained the parameters of the normal distribution (sample mean and sample SD) to then calculate the statistical significance (P value) of the observed CaCTS score (i.e., the real calculated CaCTS score). We observed that the CaCTS score distributions are likely normal and somewhat smooth; therefore, we can reasonably assume normality and that our high number of randomizations (10,000) generates a representative approximation of the null distributions. To perform differential expression tests, TF expression in each tumor type was contrasted to the average expression of each TF across the remaining 33 tumor types. One-sided t tests were used to derive P values for each contrast.

Identification of SE-associated genes

We collected publicly available H3K27ac ChIP-seq data from the Gene Expression Omnibus (www.ncbi.nlm.nih.gov/geo/) using search term "H3K27ac," "TCGA study abbreviation," or "tumor type," e.g., "BRCA" or "breast cancer." We prioritized data generated for primary tumor tissues and only included data for cell lines when primary tumor data were not available. For OV, ESAD, PRAD, KIRC, and GBM, we used in-house tumor tissue H3K27ac ChIP-seq data. Data were processed using ENCODE pipeline version v1.2.0 and v1.1.7. ENCODE performed the alignment using bwa (version bwa-0.7.15-r1140) (*70*) and peak calling the Model-based Analysis of ChIP-Seq (MACS) (version 2.2.4, MACS2) (*71*). We used the total IP reads, "Normalized Strand Cross-correlation coefficient" (NSC), "Relative Strand Cross-correlation coefficient" (RSC), and "Fraction of reads in peaks" (FRiP) quality metrics to select the samples with high quality. Higher values of NSC indicate more enrichment, values less than 1.5 are relatively low NSC scores, and the minimum

possible value is 1 (no enrichment); for RSC, the minimum possible value is 0 (no signal), highly enriched experiments have values greater than 1, and values much less than 1 may indicate low quality; and FRiP is the fraction of mapped reads that fall into the called peak regions. Adjusted from ENCODE guidelines, we only included samples that passed the following quality control thresholds: total IP reads > 15 million, NSC > 1.05, RSC > 0.8, and FRiP > 0.1. The full curated list of H3K27ac ChIP-seq datasets used in this study can be found in table S5. SE calls were obtained using the rank ordering of SE2 (ROSE2) algorithm for all tumor types, except prostate and kidney, where ROSE was used (*4*). For ROSE2, we aligned to genome build hg19, with the following parameters: stitching distance -s 12500 and distance from TSS to exclude -t 2500. We selected SEs assigned to known TFs (*13*, *69*).

Assessing enrichment of SE-associated genes in CaCTS ranked lists

We implemented the GSEA using the R package fgsea (version 1.10) (*72*) to evaluate the enrichment of SE-associated genes with the list of TFs ranked by CaCTS scores for each tumor type. We applied the fgsea function with the parameter nperm equal to 10,000. Numbers of SEs detected in each tumor are listed in table S5. Some samples did not have SE-defining datasets available; so, given the biological similarities of GBM and LGG, GBM SEs were used as a proxy to evaluate LGG MTFs predicted by CaCTS. Similarly, GSEA was performed on the READ candidates using COAD SEs.

Hierarchical clustering using CaCTS score

Clustering was performed using the Spearman method and complete distance parameters. We selected a height cutoff equal to 0.63 to define the clusters, which resulted in 11 groups (in the 34 tumor group analyses). To compare our clusters with groups defined by TCGA (*25*), CaCTS clusters were matched to the TCGA cluster to which 50% (or more) of the samples were assigned.

Analyses of CaCTS TF dependencies

For the dependency analysis, we manually searched the Cancer Dependency Map Project database (DepMap Achilles 19Q1 public release; https://figshare.com/articles/DepMap_Achilles_19Q1_Public/7655150) (*33*) for cell lines that rightfully correspond to the 34 tumor types. We found dependency data for 20 of 34 (58.8%) cancer types and retrieved 434 cell lines across these, with LUAD having the largest number of lines (n = 31) and CHOL having the lowest (n = 1). We performed hierarchical clustering of cell lines and the CaCTS TFs for the corresponding tumor type (method = ward.d2, distance = euclidean). We also calculated the percentage of tumor types with at least one predicted candidate with a CERES score of <−0.4, <−0.6, <−0.8, or <−1 in at least 50% of the cell lines for that tumor type. Search terms used for each tumor type (for primary disease and subtype) were as follows: ACC: "Adrenal Cancer"; BLCA: "Bladder Cancer"; LGG: "Brain Cancer", and then filtered by "Astrocytoma," Astrocytoma, anaplastic," "Glioma, Neuroglioma," "Oligodendroglioma," and "Oligodendroglioma, anaplastic"; BRCA: "Breast Cancer"; CESC: "Cervical Cancer"; CHOL: "Bile Duct Cancer"; COAD: "Colon/Colorectal Cancer"; ESAD: "Esophageal Cancer" and then filtered by "Adenocarcinoma"; ESSC: "Esophageal Cancer" and then filtered by "Squamous Cell Carcinoma"; GBM: "Brain Cancer" and then filtered by "Glioblastoma"; HNSC: "Head and Neck Cancer"; KIRC: "Kidney Cancer" then filtered by "Renal Carcinoma, clear cell" and "Renal Adenocarcinoma, clear cell"; LIHC: "Liver Cancer"; LUAD: "Lung Cancer" and then filtered by "Non-Small Cell Lung Cancer (NSCLCL), Adenocarcinoma"; LUSC: "Lung Cancer" and then filtered by "Non-Small Cell Lung Cancer (NSCLCL), Squamous Cell Carcinoma"; DLBC: "Lymphoma" and then filtered by "Diffuse Large B-cell Lymphoma (DLBCL)"; MESO: "Lung Cancer" and then filtered by "Mesothelioma"; LAML: "Leukemia" and then filtered by "AML"; PAAD: "Pancreatic Cancer"; PRAD: "Prostate Cancer"; COAD: "Colon/Colorectal Cancer"; SARC: "Sarcoma" and then filtered by "Liposarcoma"; SKCM: "Skin Cancer" and then filtered by "Melanoma" and "Melanoma, amelanotic"; STAD: "Gastric Cancer"; TGCT: "Embryonal Cancer"; THCA: "Thyroid Cancer"; UCS: "Endometrial/Uterine Cancer" and then filtered by "Endometrial Stromal Sarcoma"; UCEC: "Endometrial/Uterine Cancer" and then filtered by "Uterine/Endometrial Adenocarcinoma," "Endometrial Carcinoma," and "Endometrial Adenocarcinoma"; UVM: "Eye Cancer"; and PCPG: "Neuroblastoma". For OV and BRCA subtypes, we manually curated cell lines for inclusion as follows: OV (OAW28, COV318, KURAMOCHI, SNU8, ONCODG, JHOS4, JHOS2, OVCAR8, COV504, COV362, OV90, X59M, OVCAR5, CAOV3, EFO21, and A2780); BRCA (luminal A) (EFM19, HCC1428, CAMA1, HCC1419, MCF7, MDAMB415, SKBR3, ZR751, KPL1, HCC202, and HMC18); BRCA (luminal B) (EFM19, HCC1428, CAMA1, HCC1419, MCF7, MDAMB415, SKBR3, ZR751, KPL1, HCC202, and HMC18); BRCA (basal/TNBC) (MDAMB468, HCC1806, HCC1395, MDAMB436, MDAMB231, SUM159PT, BT549, HCC1937, MDAMB157, CAL51, HS578T, HCC1143, DU4475, and HMC18); and BRCA (HER2) (AU565, MDAMB453, JIMT1, and HCC1954).

Tumor subtype MTF predictions

To predict candidate MTFs specific for each of the 140 tumor subtypes (table S9), we implemented the same workflow developed for the 34 TCGA cancer types; however, instead of adding all samples for a query cancer type, we selected one subtype at once to be the query group and all other cancer subtypes to be the background. For example, considering the four molecular subgroups for ovarian cancer, we select "proliferative" samples as query and all other 136 cancer subtypes as background, leaving out other molecular subgroups for OV, i.e., mesenchymal, differentiated, and immunoreactive.

Somatic mutation analyses

We used coding SNVs from 2715 tumors from the PCAWG project (https://icgc.org). We removed all SNVs that fall into regions of low mappability (wgEncodeDacMapabilityConsensusExcludable.bed). To identify frequently mutated genes, we calculated a background mutation rate for each sample. Let $X_i (X_i \in [0, n])$ be a random variable that represents the number of samples with at least one mutation in the ith gene (where n is the total number of samples of a given tumor type), and then X_i follows a Poisson binomial distribution with a vector of probabilities $p = [1 - (1 - p_k)^{ni}]_k$, where n_i is the size of the coding sequence of the ith gene in base pairs and p_k is the global background rate of sample k ($k \in [1, n]$) empirically estimated by the ratio of the total number of SNVs in sample k (n_k) over the total coverage of all exons (in base pairs) (n_{cov})

$$p_k = \frac{n_k}{n_{\text{cov}}}$$

To determine the statistical significance of the observed number of mutated samples in the ith gene, we calculated the probability of having at least s_i samples mutated, i.e., P value$_i = P(X_i \geq s_i)$. P values were adjusted using the Benjamini-Hochberg method.

Oncoplots of PAX8, SOX17, and MECOM genetic aberrations

Data were obtained from cBioPortal (*73*). Data included 579 patients/samples with OV, from the study TCGA Provisional.

Experimental methods

H3K27ac ChIP-seq of primary HGSOC tissues

All tissues used were collected with informed consent and the approval of the institutional review boards of the University of Southern California, Cedars-Sinai Medical Center (CSMC), the Whitehead Institute for Biomedical Research (WIBR), and the Dana-Farber Cancer Institute. All specimens profiled were primary, chemotherapy-naïve HGSOCs. Tumors 1 to 4 were profiled at CSMC and have been previously described (*74*). Briefly, 5-mm punches of optimal cutting temperature compound (OCT)–embedded, pathology-reviewed tumor specimens were taken from epithelial enriched regions. Tissues were subjected to ChIP-seq using an anti-H3K27ac antibody (C15410196, Diagenode, Denville, NJ). Tumors 5 to 12 were profiled at WIBR. Thirty-micrometer sections of frozen tissues, with a >90% tumor enrichment, were washed with phosphate-buffered saline (PBS) and cross-linked with 1% formaldehyde for 10 min and quenched with 0.125 M glycine for 5 min at room temperature. Cross-linked material was resuspended in 1 ml of lysis buffer [0.1%SDS, 1× Triton X-100, 10 mM tris-HCl (pH 8), 1 mM EDTA (pH 8), 0.1% sodium deoxycholate, 0.25% sarkosyl, 0.3 mM NaCl, 1× protease inhibitor cocktail, and 5 mM sodium butyrate] and sonicated for 20 min with a Covaris E220 instrument (10% duty cycle, 175 peak incident power, 200 cycles per burst, and 1 ml of AFA Fiber milliTUBEs). Eight micrograms of soluble chromatin was immunoprecipitated with 10 μg of H3K27ac (C15410196, lot #a1723-0041d, Diagenode) antibody. ChIP-seq libraries were constructed using an Accel-NGS 2S DNA library kit from Swift BioSciences. Fragments of the desired size were enriched using AMPure XP beads (Beckman Coulter). Thirty-six–base pair (bp) paired-end reads were sequenced on a NextSeq instrument (Illumina). Data were processed using the ENCODE pipeline, as described above, with SEs identified using ROSE.

Cell culture

OVCAR4 and Kuramochi cell line models were selected as they closely recapitulate the molecular features of human HGSOC (*65*). Cells were cultured in RPMI 1640 supplemented with 10% fetal bovine serum, 1× nonessential amino acid (NEAA) cell culture supplement , insulin (11.4 μg/ml), and 1× penicillin/streptomycin and maintained at 37°C with 5% CO_2. Cells were passaged with 0.05% trypsin using standard cell culture procedures. Cells were confirmed to be negative for *Mycoplasma* and were authenticated by profiling of short tandem repeats using the Promega PowerPlex 16HS assay, performed at the University of Arizona Genomics Core (table S14).

RNA interference and colony formation assays

OVCAR4 cells were reverse transfected with nontargeting [Dharmacon ON-TARGETplus Non-targeting Control Pool (NT1) and a second custom control pool containing the following: D-0012-03, D-001210-04, and D-001210-05 (NT2)] or pooled *PAX8*, *SOX17*, and *MECOM* oligonucleotides (L-003778-00-0005, L-013028-01-0010, and L-006530-02-0005, Dharmacon) by incubating 120 nM of each siRNA pool in Opti-MEM I (Thermo Fisher Scientific) for 5 min, which was then combined with a mix of Opti-MEM I and Lipofectamine RNAiMAX (Thermo Fisher Scientific) and incubated for 20 min at room temperature. The transfection reagent mix was then combined with 300,000 cells and seeded in a 60-mm dish. Medium was replenished after 24 hours, and transfected cells were used for analysis or assays 48 hours later. For colony formation assays, transfected cells were trypsinized and counted, and 1000 cells per condition were seeded in six-well plates, in triplicate. Media were replenished once per week, and after 14 days, the cells were washed with 1× PBS (Thermo Fisher Scientific) three times and fixed with 10% formalin (McKesson) for 20 min. Plates were then washed with water and stained with 0.1% crystal violet for 30 min. Excess crystal violet was washed with water, and colonies were counted manually.

Western blotting

Cells were lysed with 100 μl of cell lysis buffer per 1 million cells [10 mM Hepes (pH 7.5) by KOH, 300 mM NaCl, 0.1% NP-40, 5 mM EGTA, with aprotinin (10 μg/ml), leupeptin (10 μg/ml), 1× protease inhibitor cocktail (Roche), 1× PhosSTOP Protease Inhibitor Cocktail (Roche), 1× phenylmethylsulfonyl fluoride (Sigma-Aldrich), and Supraise-in (Ambion)] at 4°C for 1 hour. Lysed samples were then centrifuged at 12,000*g* for 10 min at 4°C, and supernatants were collected. Thirty micrograms of whole-cell extracts were treated with sample buffer and boiled at 95°C for 5 min. Samples were separated via SDS–polyacrylamide gel electrophoresis (Bio-Rad) and transferred to a nitrocellulose membrane with the Trans-Blot Turbo system (Bio-Rad) per the manufacturer's instructions. Membranes were blocked with StartingBlock (Thermo Fisher Scientific) blocking buffer for 60 min at room temperature, followed by incubation with primary antibodies to detect PAX8 (1:1000 dilution; 32440, Novus), SOX17 (1:2000; ab224637, Abcam), MECOM (1:1000; C50E12, Cell Signaling Technology), or β-tubulin (1:2000; D3U1W, Cell Signaling Technology). Primary antibody incubations were performed in blocking buffer overnight at 4°C. Samples were then washed with tris-buffered saline with 0.1% Tween® 20 Detergent (TBT-T) three times for 10 min each and incubated in secondary antibody (1:10,000; ab6721 or ab6789, Abcam) for 1 hour, followed by three 10-min TBS-T washes. Membranes were developed using the Piece ECL Western Substrate (Thermo Fisher Scientific) following the manufacturer's protocol.

RNA-seq and data analysis

OVCAR4 cells were transfected in triplicate with the two siRNA control pools and target gene siRNA pools described above. RNA and protein were harvested 72 hours after transfection. Protein lysates were used to verify knockdown using Western blotting, as described above. Cells were washed with cold PBS, collected by scraping, and RNA extraction was performed using the NucleoSpin RNA Plus Kit (Macherey-Nagel) per the manufacturer's protocol. Extracted RNA samples were used for polyadenylate nonstranded library preparation and 150-bp paired-end sequencing at 40 million reads using the DNBseq next-generation sequencing platform (RNA-seq performed by BGI). Reads were filtered and aligned using STAR-2.5.1b (ref_genome_hg38_gencodev26), and a gene-level read count matrix was generated using featureCounts (subread-1.6.3-source). Differential gene expression analyses were then performed using the R package DESeq2 (version 1.24.0). Differentially expressed genes were selected using an absolute $\log_2$ fold change of ≥ 1 and an adjusted P value of ≤ 0.01. Pathway analyses were performed using the R package ReactomePA (*75*).

Proximity ligation assay

To perform the PLA, we used the Duolink Technology (DUO92101, Sigma-Aldrich). Kuramochi cells were grown for 24 hours on a 96-well imaging plate (0030741030, Eppendorf). Cells were fixed in 4% paraformaldehyde for 15 min, permeabilized with 0.25% Triton X-100 for 15 min, and blocked with 1% bovine serum albumin in PBS containing 0.1% Tween 20 for 30 min. Primary antibodies against PAX8 (1:250 dilution; NBP2-29903, Novus), SOX17 (1:250 dilution; 81778S, Cell Signaling Technology), and MECOM (1:250 dilution; 23201-1-AP, ProteinTech) were incubated overnight

at 4°C. After three 5-min washes with TBS-T, PLA probes were incubated overnight at 4°C. Detection was performed using Duolink RED detection reagents as recommended by the manufacturer. Samples were air-dried and covered with Duolink mounting medium with 4′,6-diamidino-2-phenylindole and then imaged using a Nikon Eclipse Ti inverted microscope under ×40 magnification.

ChIP-seq of PAX8, SOX17, MECOM, and CTCF in Kuramochi cells

Kuramochi cells were grown to 80% confluence, cross-linked with 1% formaldehyde in PBS for 15 min, pelleted, and flash-frozen. One hundred million cells were used per ChIP with 10 μg of each antibody, PAX8 (catalog number 59019, lot 1, Cell Signaling Technology), SOX17 (catalog number AF1924, lot KG0818071, R&D Systems), and MECOM (catalog number 2593, lot 4, Cell Signaling Technology). Sonications were performed with a Qsonica microtip sonicator with 4 min of total time (30-s on, 1-min off) with 18 to 21 W. The supernatant of the sonicated lysates was incubated overnight at 4°C with the antibody and Invitrogen DynaI magnetic bead mix. After extensive washing, enriched chromatin was purified as follows: for PAX8, using a phenol:chloroform:isoamyl alcohol extraction (*76*), and for SOX17 and MECOM, bead:antibody:chromatin complexes were resuspended in elution buffer, placed at 65°C for 45 min with intermittent vortexing and spun down. Ribonuclease A (RNase A) was added to the supernatant, and samples were incubated at 65°C for 3.5 hours before a proteinase K digest at 42°C for 1 hour. DNA was then purified using polymerase chain reaction (PCR) column purification. CCCTC-binding factor (CTCF) ChIP-seq was performed as previously described (*77*) using 25 μg of chromatin and 5 μg of anti-CTCF antibody (catalog number 61311, Active Motif), and chromatin was sheared to 100- to 300-bp fragments using a Covaris E220 evolution focused ultrasonicator. ChIP libraries were constructed using the KAPA Hyper Library Preparation kit, quantified, and sequenced on an Illumina NextSeq 500 sequencer. Reads were aligned to the hg19 version of the human reference genome using bowtie v1.2 (*78*) with parameters -k 2 -m 2 –best and -l set to the read length. WIG files for display were made using MACS v1.4 (*71*) with parameters -w -S –space = 50 –nomodel –shiftsize = 200. Regions statistically enriched in reads were identified using MACS v1.4 with corresponding input control and parameters -p 1e-9 –keep-dup = auto. Regions for the colocalization heatmap were constructed by collapsing regions enriched in PAX8, SOX17, and MECOM using bedtools merge (*79*) and creating 4-kb regions centered on the center of the collapsed regions. Read coverage was quantified for heatmap analysis using bamToGFF (https://github.com/BradnerLab/pipeline) with parameters -m 100 -r using a mapped read bam with non-PCR duplicate reads created with samtools rmdup (*80*). Heatmaps were ranked using read coverage quantified in 1-kb windows centered on the middle of each collapsed region using bamToGFF with parameters -m 1 -r.

THZ531, THZ1, and JQ1 dose response curves

OVCAR4 cells were plated in 96-well plates at 5000 cells per well. After 24 hours, THZ531 (ApexBio), THZ1 (Selleck Chemicals), and JQ1 (Tocris) or vehicle [dimethyl sulfoxide (DMSO)] were added in 1:3 dilutions starting from 10,000 to 1.5 nM in triplicate and incubated for 72 hours at 37°C. Cell numbers were quantified using the Promega Cell Titer Glo reagent. Signals were then normalized to the lowest dose, and IC_{50} values were calculated with GraphPad Prism.

THZ531, THZ1, and JQ1 drug treatment for RNA

OVCAR4 cells (400,000) were seeded in 60-mm dishes 24 hours before experiment. Cells were treated with either low-dose THZ1 (50 nM), high-dose THZ1 (250 nM), low-dose THZ531 (100 nM), high-dose THZ531 (500 nM), JQ1 (500 nM), or DMSO (500 nM) for 6 hours. Plates were then washed with ice-cold PBS once, followed by RNA extraction with the NucleoSpin RNA Mini Kit (Macherey-Nagel), followed by quantitative PCR. Relative expression was measured and normalized using the average expression of *GAPDH* and *ACTB* and then normalized to DMSO control.

THZ531, THZ1, and JQ1 global transcriptome data

RNA-seq data from THZ1-treated Kuramochi cells were obtained from GSE116282 (*66*). Data were filtered to remove lowly transcripts [Reads Rer Kilobase of transcript, per Million mapped reads (RPKM) > 1] to select protein-coding genes, and expression of the top 10% of genes, at 50 and 250 nM, was plotted. Gene set enrichment was conducted using an ascending-ordered ranked list of $\log_2$FC after target gene knockdown and gene sets of THZ1 differentially expressed genes.

Quantitative PCR

Total RNA was converted to complementary DNA (cDNA) using random primers (Promega) and the Moloney Murine Leukemia Virus (M-MLV) Reverse Transcriptase RNase H (Promega) as per the manufacturer's instructions. cDNA was then amplified by the QuantStudio 12K Flex Real-Time PCR system (Thermo Fisher Scientific) with the Taqman Universal Master Mix with UNG (Applied Biosystems) along with the following probes: GAPDH, TUBB, and/or ACTB as housekeeping genes (Hs02786624_g1, Hs00742828_s1, and/or Hs01060665_g1), as well as PAX8, SOX17, and MECOM (Hs00247586_m1, Hs00751752_s1, and Hs00602795_m1).

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at https://science.org/doi/10.1126/sciadv.abf6123

View/request a protocol for this paper from *Bio-protocol*.

REFERENCES AND NOTES

1. T. Sanda, L. N. Lawton, M. I. Barrasa, Z. P. Fan, H. Kohlhammer, A. Gutierrez, W. Ma, J. Tatarek, Y. Ahn, M. A. Kelliher, C. H. M. Jamieson, L. M. Staudt, R. A. Young, A. T. Look, Core transcriptional regulatory circuit controlled by the TAL1 complex in human T cell acute lymphoblastic leukemia. *Cancer Cell* **22**, 209–221 (2012).
2. M. R. Mansour, B. J. Abraham, L. Anders, A. Berezovskaya, A. Gutierrez, A. D. Durbin, J. Etchin, L. Lawton, S. E. Sallan, L. B. Silverman, M. L. Loh, S. P. Hunger, T. Sanda, R. A. Young, A. T. Look, Oncogene regulation. An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element. *Science* **346**, 1373–1377 (2014).
3. A. D. Durbin, M. W. Zimmerman, N. V. Dharia, B. J. Abraham, A. B. Iniguez, N. Weichert-Leahey, S. He, J. M. Krill-Burger, D. E. Root, F. Vazquez, A. Tsherniak, W. C. Hahn, T. R. Golub, R. A. Young, A. T. Look, K. Stegmaier, Selective gene dependencies in MYCN-amplified neuroblastoma include the core transcriptional regulatory circuitry. *Nat. Genet.* **50**, 1240–1246 (2018).
4. W. A. Whyte, D. A. Orlando, D. Hnisz, B. J. Abraham, C. Y. Lin, M. H. Kagey, P. B. Rahl, T. I. Lee, R. A. Young, Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013).
5. A. C. D'Alessio, Z. P. Fan, K. J. Wert, P. Baranov, M. A. Cohen, J. S. Saini, E. Cohick, C. Charniga, D. Dadon, N. M. Hannett, M. J. Young, S. Temple, R. Jaenisch, T. I. Lee, R. A. Young, A systematic approach to identify candidate transcription factors that control cell identity. *Stem Cell Rep.* **5**, 763–775 (2015).
6. J. E. Bradner, D. Hnisz, R. A. Young, Transcriptional addiction in cancer. *Cell* **168**, 629–643 (2017).
7. N. Kwiatkowski, T. Zhang, P. B. Rahl, B. J. Abraham, J. Reddy, S. B. Ficarro, A. Dastur, A. Amzallag, S. Ramaswamy, B. Tesar, C. E. Jenkins, N. M. Hannett, D. McMillin, T. Sanda, T. Sim, N. D. Kim, T. Look, C. S. Mitsiades, A. P. Weng, J. R. Brown, N. S. Gray, Targeting transcription regulation in cancer with a covalent CDK7 inhibitor. *Nature* **511**, 616–620 (2014).
8. B. Chapuy, M. R. McKeown, C. Y. Lin, S. Monti, M. G. M. Roemer, J. Qi, P. B. Rahl, H. H. Sun, K. T. Yeda, J. G. Doench, E. Reichert, A. L. Kung, S. J. Rodig, R. A. Young, M. A. Shipp, J. E. Bradner, Discovery and characterization of super-enhancer-associated dependencies in diffuse large B cell lymphoma. *Cancer Cell* **24**, 777–790 (2013).
9. Y. Wang, T. Zhang, N. Kwiatkowski, B. J. Abraham, T. I. Lee, S. Xie, H. Yuzugullu, T. Von, H. Li, Z. Lin, D. G. Stover, E. Lim, Z. C. Wang, J. D. Iglehart, R. A. Young, N. S. Gray, J. J. Zhao,

CDK7-dependent transcriptional addiction in triple-negative breast cancer. *Cell* **163**, 174–186 (2015).

10. J. Lovén, H. A. Hoke, C. Y. Lin, A. Lau, D. A. Orlando, C. R. Vakoc, J. E. Bradner, T. I. Lee, R. A. Young, Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* **153**, 320–334 (2013).
11. S. C. J. Parker, M. L. Stitzel, D. L. Taylor, J. M. Orozco, M. R. Erdos, J. A. Akiyama, K. L. van Bueren, P. S. Chines, N. Narisu; NISC Comparative Sequencing Program, B. L. Black, A. Visel, L. A. Pennacchio, F. S. Collins, Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 17921–17926 (2013).
12. L. Chen, M. Huang, J. Plummer, J. Pan, Y. Y. Jiang, Q. Yang, T. C. Silva, N. Gull, S. Chen, L. W. Ding, O. An, H. Yang, Y. Cheng, J. W. Said, N. Doan, W. N. Dinjens, K. M. Waters, R. Tuli, S. A. Gayther, S. J. Klempner, B. P. Berman, S. J. Meltzer, D.-C. Lin, H. P. Koeffler, Master transcription factors form interconnected circuitry and orchestrate transcriptional networks in oesophageal adenocarcinoma. *Gut* **69**, 630–640 (2020).
13. V. Saint-André, A. J. Federation, C. Y. Lin, B. J. Abraham, J. Reddy, T. I. Lee, J. E. Bradner, R. A. Young, Models of human core transcriptional regulatory circuitries. *Genome Res.* **26**, 385–396 (2016).
14. A. J. Federation, D. R. Polaski, C. J. Ott, A. Fan, C. Y. Lin, J. E. Bradner, Identification of candidate master transcription factors within enhancer-centric transcriptional regulatory networks. bioRxiv 345413 [**Preprint**]. 12 June 2018. https://doi.org/10.1101/345413.
15. C. J. Ott, A. J. Federation, L. S. Schwartz, S. Kasar, J. L. Klitgaard, R. Lenci, Q. Li, M. Lawlor, S. M. Fernandes, A. Souza, D. Polaski, D. Gadi, M. L. Freedman, J. R. Brown, J. E. Bradner, Enhancer architecture and essential core regulatory circuitry of chronic lymphocytic leukemia. *Cancer Cell* **34**, 982–995.e7 (2018).
16. C. Y. Lin, S. Erkek, Y. Tong, L. Yin, A. J. Federation, M. Zapatka, P. Haldipur, D. Kawauchi, T. Risch, H.-J. Warnatz, B. C. Worst, B. Ju, B. A. Orr, R. Zeid, D. R. Polaski, M. Segura-Wang, S. M. Waszak, D. T. W. Jones, M. Kool, V. Hovestadt, I. Buchhalter, L. Sieber, P. Johann, L. Chavez, S. Gröschel, M. Ryzhova, A. Korshunov, W. Chen, V. V. Chizhikov, K. J. Millen, V. Amstislavskiy, H. Lehrach, M.-L. Yaspo, R. Eils, P. Lichter, J. O. Korbel, S. M. Pfister, J. E. Bradner, P. A. Northcott, Active medulloblastoma enhancers reveal subgroup-specific cellular origins. *Nature* **530**, 57–62 (2016).
17. B. E. Gryder, S. Pomella, C. Sayers, X. S. Wu, Y. Song, A. M. Chiarella, S. Bagchi, H.-C. Chou, R. S. Sinniah, A. Walton, X. Wen, R. Rota, N. A. Hathaway, K. Zhao, J. Chen, C. R. Vakoc, J. F. Shern, B. Z. Stanton, J. Khan, Histone hyperacetylation disrupts core gene regulatory architecture in rhabdomyosarcoma. *Nat. Genet.* **51**, 1714–1722 (2019).
18. T. M. Malta, A. Sokolov, A. J. Gentles, T. Burzykowski, L. Poisson, J. N. Weinstein, B. Kamińska, J. Huelsken, L. Omberg, O. Gevaert, A. Colaprico, P. Czerwińska, S. Mazurek, L. Mishra, H. Heyn, A. Krasnitz, A. K. Godwin, A. J. Lazar; Cancer Genome Atlas Research Network, J. M. Stuart, K. A. Hoadley, P. W. Laird, H. Noushmehr, M. Wiznerowicz, Machine learning identifies stemness features associated with oncogenic dedifferentiation. *Cell* **173**, 338–354.e15 (2018).
19. J. Yuan, Y.-Y. Jiang, A. Mayakonda, M. Huang, L.-W. Ding, H. Lin, F. Yu, Y. Lu, T. K. S. Loh, M. Chow, S. Savage, J. W. Tyner, D.-C. Lin, H. P. Koeffler, Super-enhancers promote transcriptional dysregulation in nasopharyngeal carcinoma. *Cancer Res.* **77**, 6614–6626 (2017).
20. P. Eliades, B. J. Abraham, Z. Ji, D. M. Miller, C. L. Christensen, N. Kwiatkowski, R. Kumar, C. N. Njauw, M. Taylor, B. Miao, T. Zhang, K.-K. Wong, N. S. Gray, R. A. Young, H. Tsao, High MITF expression is associated with super-enhancers and suppressed by CDK7 inhibition in melanoma. *J. Invest. Dermatol.* **138**, 1582–1590 (2018).
21. S. Shang, J. Yang, A. A. Jazaeri, A. J. Duval, T. Tufan, N. Lopes Fischer, M. Benamar, F. Guessous, I. Lee, R. M. Campbell, P. J. Ebert, T. Abbas, C. N. Landen, A. Difeo, P. C. Scacheri, M. Adli, Chemotherapy-induced distal enhancers drive transcriptional programs to maintain the chemoresistant state in ovarian cancer. *Cancer Res.* **79**, 4599–4611 (2019).
22. Y.-Y. Jiang, Y. Jiang, C.-Q. Li, Y. Zhang, P. Dakle, H. Kaur, J.-W. Deng, R. Y.-T. Lin, L. Han, J.-J. Xie, A. Mayakonda, M. Hazawa, L. Xu, Y. Li, L. Aswad, M. Jeitany, D. Kanojia, X.-Y. Guan, M. J. Fullwood, D.-C. Lin, H. P. Koeffler, TP63, SOX2 and KLF5 establish core regulatory circuitry and construct cancer specific epigenome in esophageal squamous cell carcinoma. bioRxiv 825372 [**Preprint**]. 6 November 2019. https://doi.org/10.1101/825372.
23. S. Frietze, H. O'Geen, L. E. Littlepage, C. Simion, C. A. Sweeney, P. J. Farnham, S. R. Krig, Global analysis of ZNF217 chromatin occupancy in the breast cancer cell genome reveals an association with ERalpha. *BMC Genomics* **15**, 520 (2014).
24. R. M. Witwicki, M. B. Ekram, X. Qiu, M. Janiszewska, S. Shu, M. Kwon, A. Trinh, E. Frias, N. Ramadan, G. Hoffman, K. Yu, Y. Xie, G. McAllister, R. McDonald, J. Golji, M. Schlabach, A. deWeck, N. Keen, H. M. Chan, D. Ruddy, T. Rejtar, S. Sovath, S. Silver, W. R. Sellers, Z. Jagani, M. D. Hogarty, C. Roberts, M. Brown, K. Stegmaier, H. Long, R. A. Shivdasani, D. Pellman, K. Polyak, TRPS1 is a lineage-specific transcriptional dependency in breast cancer. *Cell Rep.* **25**, 1255–1267.e5 (2018).
25. K. A. Hoadley, C. Yau, T. Hinoue, D. M. Wolf, A. J. Lazar, E. Drill, R. Shen, A. M. Taylor, A. D. Cherniack, V. Thorsson, R. Akbani, R. Bowlby, C. K. Wong, M. Wiznerowicz, F. Sanchez-Vega, A. G. Robertson, B. G. Schneider, M. S. Lawrence, H. Noushmehr, T. M. Malta; Cancer Genome Atlas Network, J. M. Stuart, C. C. Benz, P. W. Laird, Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer. *Cell* **173**, 291–304.e6 (2018).
26. M. M. Pomerantz, F. Li, D. Y. Takeda, R. Lenci, A. Chonkar, M. Chabot, P. Cejas, F. Vazquez, J. Cook, R. A. Shivdasani, M. Bowden, R. Lis, W. C. Hahn, P. W. Kantoff, M. Brown, M. Loda, H. W. Long, M. L. Freedman, The androgen receptor cistrome is extensively reprogrammed in human prostate tumorigenesis. *Nat. Genet.* **47**, 1346–1351 (2015).
27. G. Ciriello, M. L. Gatza, A. H. Beck, M. D. Wilkerson, S. K. Rhie, A. Pastore, H. Zhang, M. McLellan, C. Yau, C. Kandoth, R. Bowlby, H. Shen, S. Hayat, R. Fieldhouse, S. C. Lester, G. M. K. Tse, R. E. Factor, L. C. Collins, K. H. Allison, Y.-Y. Chen, K. Jensen, N. B. Johnson, S. Oesterreich, G. B. Mills, A. D. Cherniack, G. Robertson, C. Benz, C. Sander, P. W. Laird, K. A. Hoadley, T. A. King; TCGA Research Network, C. M. Perou, Comprehensive molecular portraits of invasive lobular breast cancer. *Cell* **163**, 506–519 (2015).
28. N. Zhao, J. Cao, L. Xu, Q. Tang, L. E. Dobrolecki, X. Lv, M. Talukdar, Y. Lu, X. Wang, D. Z. Hu, Q. Shi, Y. Xiang, Y. Wang, X. Liu, W. Bu, Y. Jiang, M. Li, Y. Gong, Z. Sun, H. Ying, B. Yuan, X. Lin, X.-H. Feng, S. M. Hartig, F. Li, H. Shen, Y. Chen, L. Han, Q. Zeng, J. B. Patterson, B. A. Kaipparettu, N. Putluri, F. Sicheri, J. M. Rosen, M. T. Lewis, X. Chen, Pharmacological targeting of MYC-regulated IRE1/XBP1 pathway suppresses MYC-driven breast cancer. *J. Clin. Invest.* **128**, 1283–1299 (2018).
29. G. Buchwalter, M. M. Hickey, A. Cromer, L. M. Selfors, R. N. Gunawardane, J. Frishman, R. Jeselsohn, E. Lim, D. Chi, X. Fu, R. Schiff, M. Brown, J. S. Brugge, PDEF promotes luminal differentiation and acts as a survival factor for ER-positive breast cancer cells. *Cancer Cell* **23**, 753–767 (2013).
30. X.-H. Cheng, M. Black, V. Ustiyan, T. Le, L. Fulford, A. Sridharan, M. Medvedovic, V. V. Kalinichenko, J. A. Whitsett, T. V. Kalin, SPDEF inhibits prostate carcinogenesis by disrupting a positive feedback loop in regulation of the Foxm1 oncogene. *PLOS Genet.* **10**, e1004656 (2014).
31. S. P. Kar, J. Beesley, A. Amin Al Olama, K. Michailidou, J. Tyrer, Z. Kote-Jarai, K. Lawrenson, S. Lindstrom, S. J. Ramus, D. J. Thompson; ABCTB Investigators, A. S. Kibel, A. Dansonka-Mieszkowska, A. Michael, A. K. Dieffenbach, A. Gentry-Maharaj, A. S. Whittemore, A. Wolk, A. Monteiro, A. Peixoto, A. Kierzek, A. Cox, A. Rudolph, A. Gonzalez-Neira, A. H. Wu, A. Lindblom, A. Swerdlow; AOCS Study Group & Australian Cancer Study (Ovarian Cancer); APCB Bio Resource, A. Ziogas, A. B. Ekici, B. Burwinkel, B. Y. Karlan, B. G. Nordestgaard, C. Blomqvist, C. Phelan, C. M. Lean, C. L. Pearce, C. Vachon, C. Cybulski, C. Slavov, C. Stegmaier, C. Maier, C. B. Ambrosone, C. K. Høgdall, C. C. Teerlink, D. Kang, D. C. Tessier, D. J. Schaid, D. O. Stram, D. W. Cramer, D. E. Neal, D. Eccles, D. Flesch-Janys, Digna R. Velez Edwards, D. Wokozorczyk, D. A. Levine, D. Yannoukakos, E. J. Sawyer, E. V. Bandera, E. M. Poole, E. L. Goode, E. Khusnutdinova, E. Høgdall, F. Song, F. Bruinsma, F. Heitz, F. Modugno, F. C. Hamdy, F. Wiklund, G. G. Giles, H. Olsson, H. Wildiers, H.-U. Ulmer, H. Pandha, H. A. Risch, H. Darabi, H. B. Salvesen, H. Nevanlinna, H. Gronberg, H. Brenner, H. Brauch, H. Anton-Culver, H. Song, H.-Y. Lim, I. M. Neish, I. Campbell, I. Vergote, J. Gronwald, J. Lubiński, J. L. Stanford, J. Benítez, J. A. Doherty, J. B. Permuth, J. Chang-Claude, J. L. Donovan, J. Dennis, J. M. Schildkraut, J. Schleutker, J. L. Hopper, J. Kupryjanczyk, J. Y. Park, J. Figueroa, J. A. Clements, J. A. Knight, J. Peto, J. M. Cunningham, J. Pow-Sang, J. Batra, K. Czene, K. H. Lu, K. Herkommer, K.-T. Khaw; kConFab Investigators, K. Matsuo, K. Muir, K. Offitt, K. Chen, K. B. Moysich, K. Aittomäki, K. Odunsi, L. A. Kiemeney, L. F. A. G. Massuger, L. M. Fitzgerald, L. S. Cook, L. Cannon-Albright, M. J. Hooning, M. C. Pike, M. K. Bolla, M. Luedeke, M. R. Teixeira, M. T. Goodman, M. K. Schmidt, M. Riggan, M. Aly, M. A. Rossing, M. W. Beckmann, M. Moisse, M. Sanderson, M. C. Southey, M. Jones, M. Lush, M. A. T. Hildebrandt, M.-F. Hou, M. J. Schoemaker, M. Garcia-Closas, N. Bogdanova, N. Rahman; NBCS Investigators, N. D. Le, N. Orr, N. Wentzensen, N. Pashayan, P. Peterlongo, P. Guénel, P. Brennan, P. Paulo, P. M. Webb, P. Broberg, P. A. Fasching, P. Devilee, Q. Wang, Q. Cai, Q. Li, R. Kaneva, R. Butzow, R. K. Kopperud, R. K. Schmutzler, R. A. Stephenson, R. J. Mac Innis, R. N. Hoover, R. Winqvist, R. Ness, R. L. Milne, R. C. Travis, S. Benlloch, S. H. Olson, Shannon K. Mc Donnell, S. S. Tworoger, S. Maia, S. Berndt, S. C. Lee, S.-H. Teo, S. N. Thibodeau, S. E. Bojesen, S. M. Gapstur, S. K. Kjær, T. Pejovic, T. L. J. Tammela; GENICA Network; PRACTICAL consortium, T. Dörk, T. Brüning, T. Wahlfors, T. J. Key, T. L. Edwards, U. Menon, U. Hamann, V. Mitev, V.-M. Kosma, V. W. Setiawan, V. Kristensen, V. Arndt, W. Vogel, W. Zheng, W. Sieh, W. J. Blot, W. Kluzniak, X.-O. Shu, Y.-T. Gao, F. Schumacher, M. L. Freedman, A. Berchuck, A. M. Dunning, J. Simard, C. A. Haiman, A. Spurdle, T. A. Sellers, D. J. Hunter, B. E. Henderson, P. Kraft, S. J. Chanock, F. J. Couch, P. Hall, S. A. Gayther, D. F. Easton, G. Chenevix-Trench, R. Eeles, P. D. P. Pharoah, D. Lambrechts, Genome-wide meta-analyses of breast, ovarian, and prostate cancer association studies identify multiple new susceptibility loci shared by at least two cancer types. *Cancer Discov.* **6**, 1052–1067 (2016).
32. X. Jiang, H. K. Finucane, F. R. Schumacher, S. L. Schmit, J. P. Tyrer, Y. Han, K. Michailidou, C. Lesseur, K. B. Kuchenbaecker, J. Dennis, D. V. Conti, G. Casey, M. M. Gaudet, J. R. Huyghe, D. Albanes, M. C. Aldrich, A. S. Andrew, I. L. Andrulis, H. Anton-Culver,

A. C. Antoniou, N. N. Antonenkova, S. M. Arnold, K. J. Aronson, B. K. Arun, E. V. Bandera, R. B. Barkardottir, D. R. Barnes, J. Batra, M. W. Beckmann, J. Benitez, S. Benlloch, A. Berchuck, S. I. Berndt, H. Bickeböller, S. A. Bien, C. Blomqvist, S. Boccia, N. V. Bogdanova, S. E. Bojesen, M. K. Bolla, H. Brauch, H. Brenner, J. D. Brenton, M. N. Brook, J. Brunet, H. Brunnström, D. D. Buchanan, B. Burwinkel, R. Butzow, G. Cadoni, T. Caldés, M. A. Caligo, I. Campbell, P. T. Campbell, G. Cancel-Tassin, L. Cannon-Albright, D. Campa, N. Caporaso, A. L. Carvalho, A. T. Chan, J. Chang-Claude, S. J. Chanock, C. Chen, D. C. Christiani, K. B. M. Claes, F. Claessens, J. Clements, J. M. Collée, M. C. Correa, F. J. Couch, A. Cox, J. M. Cunningham, C. Cybulski, K. Czene, M. B. Daly, A. deFazio, P. Devilee, O. Diez, M. Gago-Dominguez, J. L. Donovan, T. Dörk, E. J. Duell, A. M. Dunning, M. Dwek, D. M. Eccles, C. K. Edlund, D. R. V. Edwards, C. Ellberg, D. G. Evans, P. A. Fasching, R. L. Ferris, T. Liloglou, J. C. Figueiredo, O. Fletcher, R. T. Fortner, F. Fostira, S. Franceschi, E. Friedman, S. J. Gallinger, P. A. Ganz, J. Garber, J. A. García-Sáenz, S. A. Gayther, G. G. Giles, A. K. Godwin, M. S. Goldberg, D. E. Goldgar, E. L. Goode, M. T. Goodman, G. Goodman, K. Grankvist, M. H. Greene, H. Gronberg, J. Gronwald, P. Guénel, N. Håkansson, P. Hall, U. Hamann, F. C. Hamdy, R. J. Hamilton, J. Hampe, A. Haugen, F. Heitz, R. Herrero, P. Hillemanns, M. Hoffmeister, E. Høgdall, Y. C. Hong, J. L. Hopper, R. Houlston, P. J. Hulick, D. J. Hunter, D. G. Huntsman, G. Idos, E. N. Imyanitov, S. A. Ingles, C. Isaacs, A. Jakubowska, P. James, M. A. Jenkins, M. Johansson, M. Johansson, E. M. John, A. D. Joshi, R. Kaneva, B. Y. Karlan, L. E. Kelemen, T. Kühl, K. T. Khaw, E. Khusnutdinova, A. S. Kibel, L. A. Kiemeney, J. Kim, S. K. Kjaer, J. A. Knight, M. Kogevinas, Z. Kote-Jarai, S. Koutros, V. N. Kristensen, J. Kupryjanczyk, M. Lacko, S. Lam, D. Lambrechts, M. T. Landi, P. Lazarus, N. D. le, E. Lee, F. Lejbkowicz, H. J. Lenz, G. Leslie, D. Lessel, J. Lester, D. A. Levine, L. Li, C. I. Li, A. Lindblom, N. M. Lindor, G. Liu, F. Loupakis, J. Lubiński, L. Maehle, C. Maier, A. Mannermaa, L. L. Marchand, S. Margolin, T. May, L. McGuffog, A. Meindl, P. Middha, A. Miller, R. L. Milne, R. J. MacInnis, F. Modugno, M. Montagna, V. Moreno, K. B. Moysich, L. Mucci, K. Muir, A. M. Mulligan, K. L. Nathanson, D. E. Neal, A. R. Ness, S. L. Neuhausen, H. Nevanlinna, P. A. Newcomb, L. F. Newcomb, F. C. Nielsen, L. Nikitina-Zake, B. G. Nordestgaard, R. L. Nussbaum, K. Offit, E. Olah, A. A. A. Olama, O. I. Olopade, A. F. Olshan, H. Olsson, A. Osorio, H. Pandha, J. Y. Park, N. Pashayan, M. T. Parsons, T. Pejovic, K. L. Penney, W. H. M. Peters, C. M. Phelan, A. I. Phipps, D. Plaseska-Karanfilska, M. Pring, D. Prokofyeva, P. Radice, K. Stefansson, S. J. Ramus, L. Raskin, G. Rennert, H. S. Rennert, E. J. van Rensburg, M. J. Riggan, H. A. Risch, A. Risch, M. J. Roobol, B. S. Rosenstein, M. A. Rossing, K. de Ruyck, E. Saloustros, D. P. Sandler, E. J. Sawyer, M. B. Schabath, J. Schleutker, M. K. Schmidt, V. W. Setiawan, H. Shen, E. M. Siegel, W. Sieh, C. F. Singer, M. L. Slattery, K. D. Sorensen, M. C. Southey, A. B. Spurdle, J. L. Stanford, V. L. Stevens, S. Stintzing, J. Stone, K. Sundfeldt, R. Sutphen, A. J. Swerdlow, E. H. Tajara, C. M. Tangen, A. Tardon, J. A. Taylor, M. D. Teare, M. R. Teixeira, M. B. Terry, K. L. Terry, S. N. Thibodeau, M. Thomassen, L. Bjørge, M. Tischkowitz, A. E. Toland, D. Torres, P. A. Townsend, R. C. Travis, N. Tung, S. S. Tworoger, C. M. Ulrich, N. Usmani, C. M. Vachon, E. van Nieuwenhuysen, A. Vega, M. E. Aguado-Barrera, Q. Wang, P. M. Webb, C. R. Weinberg, S. Weinstein, M. C. Weissler, J. N. Weitzel, C. M. L. West, E. White, A. S. Whittemore, H. E. Wichmann, F. Wiklund, R. Winqvist, A. Wolk, P. Woll, M. Woods, A. H. Wu, X. Wu, D. Yannoukakos, W. Zheng, S. Zienolddiny, A. Ziogas, K. K. Zorn, J. M. Lane, R. Saxena, D. Thomas, R. J. Hung, B. Diergaarde, J. McKay, U. Peters, L. Hsu, M. García-Closas, R. A. Eeles, G. Chenevix-Trench, P. J. Brennan, C. A. Haiman, J. Simard, D. F. Easton, S. B. Gruber, P. D. P. Pharoah, A. L. Price, B. Pasaniuc, C. I. Amos, P. Kraft, S. Lindström, Shared heritability and functional enrichment across six solid cancers. *Nat. Commun.* **10**, 431 (2019).

33. R. M. Meyers, J. G. Bryan, J. M. McFarland, B. A. Weir, A. E. Sizemore, H. Xu, N. V. Dharia, P. G. Montgomery, G. S. Cowley, S. Pantel, A. Goodale, Y. Lee, L. D. Ali, G. Jiang, R. Lubonja, W. F. Harrington, M. Strickland, T. Wu, D. C. Hawes, V. A. Zhivich, M. R. Wyatt, Z. Kalani, J. J. Chang, M. Okamoto, K. Stegmaier, T. R. Golub, J. S. Boehm, F. Vazquez, D. E. Root, W. C. Hahn, A. Tsherniak, Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. *Nat. Genet.* **49**, 1779–1784 (2017).
34. A. Takata, S. Takiguchi, K. Okada, T. Takahashi, Y. Kurokawa, M. Yamasaki, H. Miyata, K. Nakajima, M. Mori, Y. Doki, Clinicopathological and prognostic significance of FOXM1 expression in esophageal squamous cell carcinoma. *Anticancer Res* **34**, 2427–2432 (2014).
35. J.-M. Park, M.-Y. Kim, T.-H. Kim, D.-K. Min, G. E. Yang, Y.-H. Ahn, Prolactin regulatory element-binding (PREB) protein regulates hepatic glucose homeostasis. *Biochim. Biophys. Acta Mol. Basis Dis.* **1864**, 2097–2107 (2018).
36. M. Annala, S. Taavitsainen, G. Vandekerkhove, J. V. W. Bacon, K. Beja, K. N. Chi, M. Nykter, A. W. Wyatt, Frequent mutation of the FOXA1 untranslated region in prostate cancer. *Commun. Biol.* **1**, 122 (2018).
37. J. L. L. Robinson, K. A. Holmes, J. S. Carroll, FOXA1 mutations in hormone-dependent cancers. *Front. Oncol.* **3**, 20 (2013).
38. R. Frank, M. Scheffler, S. Merkelbach-Bruse, M. A. Ihle, A. Kron, M. Rauer, F. Ueckeroth, K. König, S. Michels, R. Fischer, A. Eisert, J. Fassunke, C. Heydt, M. Serke, Y.-D. Ko, U. Gerigk, T. Geist, B. Kaminsky, L. C. Heukamp, M. Clement-Ziza, R. Büttner, J. Wolf, Clinical and pathological characteristics of *KEAP1*- and *NFE2L2*-mutated non-small cell lung carcinoma (NSCLC). *Clin. Cancer Res.* **24**, 3087–3096 (2018).
39. A. Colaprico, T. C. Silva, C. Olsen, L. Garofano, C. Cava, D. Garolini, T. S. Sabedot, T. M. Malta, S. M. Pagnotta, I. Castiglioni, M. Ceccarelli, G. Bontempi, H. Noushmehr, TCGAbiolinks: An R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.* **44**, e71 (2016).
40. Cancer Genome Atlas Research Network; Albert Einstein College of Medicine; Analytical Biological Services; Barretos Cancer Hospital; Baylor College of Medicine; Beckman Research Institute of City of Hope; Buck Institute for Research on Aging; Canada's Michael Smith Genome Sciences Centre; Harvard Medical School; Helen F. Graham Cancer Center & Research Institute at Christiana Care Health Services; HudsonAlpha Institute for Biotechnology; ILSbio. LLC; Indiana University School of Medicine; Institute of Human Virology; Institute for Systems Biology; International Genomics Consortium; Leidos Biomedical; Massachusetts General Hospital; McDonnell Genome Institute at Washington University; Medical College of Wisconsin; Medical University of South Carolina; Memorial Sloan Kettering Cancer Center; Montefiore Medical Center; NantOmics; National Cancer Institute; National Hospital, Abuja, Nigeria; National Human Genome Research Institute; National Institute of Environmental Health Sciences; National Institute on Deafness & Other Communication Disorders; Ontario Tumour Bank, London Health Sciences Centre; Ontario Tumour Bank, Ontario Institute for Cancer Research; Ontario Tumour Bank, The Ottawa Hospital; Oregon Health & Science University; Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center; SRA International; St Joseph's Candler Health System; Eli & Edythe L. Broad Institute of Massachusetts Institute of Technology & Harvard University; Research Institute at Nationwide Children's Hospital; Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins University; University of Bergen; University of Texas MD Anderson Cancer Center; University of Abuja Teaching Hospital; University of Alabama at Birmingham; University of California, Irvine; University of California Santa Cruz; University of Kansas Medical Center; University of Lausanne; University of New Mexico Health Sciences Center; University of North Carolina at Chapel Hill; University of Oklahoma Health Sciences Center; University of Pittsburgh; University of São Paulo, Ribeir ão Preto Medical School; University of Southern California; University of Washington; University of Wisconsin School of Medicine & Public Health; Van Andel Research Institute; Washington University in St Louis, Integrated genomic and molecular characterization of cervical cancer. *Nature* **543**, 378–384 (2017).
41. F. Farshidfar, S. Zheng, M. C. Gingras, Y. Newton, J. Shih, A. G. Robertson, T. Hinoue, K. A. Hoadley, E. A. Gibb, J. Roszik, K. R. Covington, C. C. Wu, E. Shinbrot, N. Stransky, A. Hegde, J. D. Yang, E. Reznik, S. Sadeghi, C. S. Pedamallu, A. I. Ojesina, J. M. Hess, J. T. Auman, S. K. Rhie, R. Bowlby, M. J. Borad; Cancer Genome Atlas Network, A. X. Zhu, J. M. Stuart, C. Sander, R. Akbani, A. D. Cherniack, V. Deshpande, T. Mounajjed, W. C. Foo, M. S. Torbenson, D. E. Kleiner, P. W. Laird, D. A. Wheeler, A. J. McRee, O. F. Bathe, A. B. Andersen, N. Bardeesy, L. R. Roberts, L. N. Kwong, Integrative genomic analysis of cholangiocarcinoma identifies distinct IDH-mutant molecular profiles. *Cell Rep.* **18**, 2780–2794 (2017).
42. C. F. Davis, C. J. Ricketts, M. Wang, L. Yang, A. D. Cherniack, H. Shen, C. Buhay, H. Kang, S. C. Kim, C. C. Fahey, K. E. Hacker, G. Bhanot, D. A. Gordenin, A. Chu, P. H. Gunaratne, M. Biehl, S. Seth, B. A. Kaipparettu, C. A. Bristow, L. A. Donehower, E. M. Wallen, A. B. Smith, S. K. Tickoo, P. Tamboli, V. Reuter, L. S. Schmidt, J. J. Hsieh, T. K. Choueiri, A. A. Hakimi; Cancer Genome Atlas Research Network, L. Chin, M. Meyerson, R. Kucherlapati, W.-Y. Park, A. G. Robertson, P. W. Laird, E. P. Henske, D. J. Kwiatkowski, P. J. Park, M. Morgan, B. Shuch, D. Muzny, D. A. Wheeler, W. M. Linehan, R. A. Gibbs, W. K. Rathmell, C. J. Creighton, The somatic genomic landscape of chromophobe renal cell carcinoma. *Cancer Cell* **26**, 319–330 (2014).
43. Cancer Genome Atlas Research Network, Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature* **499**, 43–49 (2013).
44. Cancer Genome Atlas Research Network, W. M. Linehan, P. T. Spellman, C. J. Ricketts, C. J. Creighton, S. S. Fei, C. Davis, D. A. Wheeler, B. A. Murray, L. Schmidt, C. D. Vocke, M. Peto, A. A. M. Al Mamun, E. Shinbrot, A. Sethi, S. Brooks, W. K. Rathmell, A. N. Brooks, K. A. Hoadley, A. G. Robertson, D. Brooks, R. Bowlby, S. Sadeghi, H. Shen, D. J. Weisenberger, M. Bootwalla, S. B. Baylin, P. W. Laird, A. D. Cherniack, G. Saksena, S. Haake, J. Li, H. Liang, Y. Lu, G. B. Mills, R. Akbani, M. D. M. Leiserson, B. J. Raphael, P. Anur, D. Bottaro, L. Albiges, N. Barnabas, T. K. Choueiri, B. Czerniak, A. K. Godwin, A. A. Hakimi, T. H. Ho, J. Hsieh, M. Ittmann, W. Y. Kim, B. Krishnan, M. J. Merino, K. R. M. Shaw, V. E. Reuter, E. Reznik, C. S. Shelley, B. Shuch, S. Signoretti, R. Srinivasan, P. Tamboli, G. Thomas, S. Tickoo, K. Burnett, D. Crain, J. Gardner, K. Lau, D. Mallery, S. Morris, J. D. Paulauskis, R. J. Penny, C. Shelton, W. T. Shelton, M. Sherman, E. Thompson, P. Yena, M. T. Avedon, J. Bowen, J. M. Gastier-Foster, M. Gerken, K. M. Leraas, T. M. Lichtenberg, N. C. Ramirez, T. Santos, L. Wise, E. Zmuda, J. A. Demchok, I. Felau, C. M. Hutter, M. Sheth, H. J. Sofia, R. Tarnuzzer, Z. Wang, L. Yang, J. C. Zenklusen, J. Zhang, B. Ayala, J. Baboud, S. Chudamani, J. Liu, L. Lolla, R. Naresh, T. Pihl, Q. Sun, Y. Wan, Y. Wu, A. Ally, M. Balasundaram, S. Balu, R. Beroukhim, T. Bodenheimer, C. Buhay, Y. S. N. Butterfield, R. Carlsen, S. L. Carter, H. Chao, E. Chuah, A. Clarke, K. R. Covington, M. Dahdouli, N. Dewal, N. Dhalla, H. V. Doddapaneni, J. A. Drummond, S. B. Gabriel, R. A. Gibbs, R. Guin, W. Hale, A. Hawes, D. N. Hayes, R. A. Holt, A. P. Hoyle, S. R. Jefferys, S. J. M. Jones, C. D. Jones, D. Kalra, C. Kovar, L. Lewis, J. Li, Y. Ma, M. A. Marra, M. Mayo,

S. Meng, M. Meyerson, P. A. Mieczkowski, R. A. Moore, D. Morton, L. E. Mose, A. J. Mungall, D. Muzny, J. S. Parker, C. M. Perou, J. Roach, J. E. Schein, S. E. Schumacher, Y. Shi, J. V. Simons, P. Sipahimalani, T. Skelly, M. G. Soloway, C. Sougnez, A. Tam, D. Tan, N. Thiessen, U. Veluvolu, M. Wang, M. D. Wilkerson, T. Wong, J. Wu, L. Xi, J. Zhou, J. Bedford, F. Chen, Y. Fu, M. Gerstein, D. Haussler, K. Kasaian, P. Lai, S. Ling, A. Radenbaugh, D. Van Den Berg, J. N. Weinstein, J. Zhu, M. Albert, I. Alexopoulou, J. J. Andersen, J. T. Auman, J. Bartlett, S. Bastacky, J. Bergsten, M. L. Blute, L. Boice, R. J. Bollag, J. Boyd, E. Castle, Y.-B. Chen, J. C. Cheville, E. Curley, B. Davies, A. De Volk, R. Dhir, L. Dike, J. Eckman, J. Engel, J. Harr, R. Hrebinko, M. Huang, L. Huelsenbeck-Dill, M. Iacocca, B. Jacobs, M. Lobis, J. K. Maranchie, S. M. Meekin, J. Myers, J. Nelson, J. Parfitt, A. Parwani, N. Petrelli, B. Rabeno, S. Roy, A. L. Salner, J. Slaton, M. Stanton, R. H. Thompson, L. Thorne, K. Tucker, P. M. Weinberger, C. Winemiller, L. A. Zach, R. Zuna, Comprehensive molecular characterization of papillary renal-cell carcinoma. *N. Engl. J. Med.* **374**, 135–145 (2016).

45. Cancer Genome Atlas Research Network, Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N. Engl. J. Med.* **368**, 2059–2074 (2013).
46. J. Hmeljak, F. Sanchez-Vega, K. A. Hoadley, J. Shih, C. Stewart, D. Heiman, P. Tarpey, L. Danilova, E. Drill, E. A. Gibb, R. Bowlby, R. Kanchi, H. U. Osmanbeyoglu, Y. Sekido, J. Takeshita, Y. Newton, K. Graim, M. Gupta, C. M. Gay, L. Diao, D. L. Gibbs, V. Thorsson, L. Iype, H. Kantheti, D. T. Severson, G. Ravegnini, P. Desmeules, A. A. Jungbluth, W. D. Travis, S. Dacic, L. R. Chirieac, F. Galateau-Sallé, J. Fujimoto, A. N. Husain, H. C. Silveira, V. W. Rusch, R. C. Rintoul, H. Pass, H. Kindler, M. G. Zauderer, D. J. Kwiatkowski, R. Bueno, A. S. Tsao, J. Creaney, T. Lichtenberg, K. Leraas, J. Bowen; TCGA Research Network, I. Felau, J. C. Zenklusen, R. Akbani, A. D. Cherniack, L. A. Byers, M. S. Noble, J. A. Fletcher, A. G. Robertson, R. Shen, H. Aburatani, B. W. Robinson, P. Campbell, M. Ladanyi, Integrative molecular characterization of malignant pleural mesothelioma. *Cancer Discov.* **8**, 1548–1565 (2018).
47. Cancer Genome Atlas Research Network; Cancer Genome Atlas Research Network, Integrated genomic characterization of pancreatic ductal adenocarcinoma. *Cancer Cell* **32**, 185–203.e13 (2017).
48. L. Fishbein, I. Leshchiner, V. Walter, L. Danilova, A. G. Robertson, A. R. Johnson, T. M. Lichtenberg, B. A. Murray, H. K. Ghayee, T. Else, S. Ling, S. R. Jefferys, A. A. de Cubas, B. Wenz, E. Korpershoek, A. L. Amelio, L. Makowski, W. K. Rathmell, A.-P. Gimenez-Roqueplo, T. J. Giordano, S. L. Asa, A. S. Tischler; Cancer Genome Atlas Research Network, K. Pacak, K. L. Nathanson, M. D. Wilkerson, Comprehensive molecular characterization of pheochromocytoma and paraganglioma. *Cancer Cell* **31**, 181–193 (2017).
49. The Cancer Genome Atlas Research Network, Comprehensive and integrated genomic characterization of adult soft tissue sarcomas. *Cell* **171**, 950–965.e28 (2017).
50. H. Shen, J. Shih, D. P. Hollern, L. Wang, R. Bowlby, S. K. Tickoo, V. Thorsson, A. J. Mungall, Y. Newton, A. M. Hegde, J. Armenia, F. Sánchez-Vega, J. Pluta, L. C. Pyle, R. Mehra, V. E. Reuter, G. Godoy, J. Jones, C. S. Shelley, D. R. Feldman, D. O. Vidal, D. Lessel, T. Kulis, F. M. Cárcano, K. M. Leraas, T. M. Lichtenberg, D. Brooks, A. D. Cherniack, J. Cho, D. I. Heiman, K. Kasaian, M. Liu, M. S. Noble, L. Xi, H. Zhang, W. Zhou, J. C. Z. Klusen, C. M. Hutter, I. Felau, J. Zhang, N. Schultz, G. Getz, M. Meyerson, J. M. Stuart; Cancer Genome Atlas Research Network, R. Akbani, D. A. Wheeler, P. W. Laird, K. L. Nathanson, V. K. Cortessis, K. A. Hoadley, Integrated molecular characterization of testicular germ cell tumors. *Cell Rep.* **23**, 3392–3406 (2018).
51. H.-S. Lee, H.-J. Jang, R. Shah, D. Yoon, M. Hamaji, O. Wald, J.-S. Lee, D. J. Sugarbaker, B. M. Burt, Genomic analysis of thymic epithelial tumors identifies novel subtypes associated with distinct clinical features. *Clin. Cancer Res.* **23**, 4855–4864 (2017).
52. A. D. Cherniack, H. Shen, V. Walter, C. Stewart, B. A. Murray, R. Bowlby, X. Hu, S. Ling, R. A. Soslow, R. R. Broaddus, R. E. Zuna, G. Robertson, P. W. Laird, R. Kucherlapati, G. B. Mills; Cancer Genome Atlas Research Network, J. N. Weinstein, J. Zhang, R. Akbani, D. A. Levine, Integrated molecular characterization of uterine carcinosarcoma. *Cancer Cell* **31**, 411–423 (2017).
53. A. G. Robertson, J. Shih, C. Yau, E. A. Gibb, J. Oba, K. L. Mungall, J. M. Hess, V. Uzunangelov, V. Walter, L. Danilova, T. M. Lichtenberg, M. Kucherlapati, P. K. Kimes, M. Tang, A. Penson, O. Babur, R. Akbani, C. A. Bristow, K. A. Hoadley, L. Iype, S. E. Woodman, Integrative analysis identifies four molecular and clinical subsets in uveal melanoma. *Cancer Cell* **32**, 204–220.e15 (2017).
54. K. Polyak, Breast cancer: Origins and evolution. *J. Clin. Invest.* **117**, 3155–3163 (2007).
55. H.-G. Moon, K.-T. Hwang, J.-A. Kim, H. S. Kim, M.-J. Lee, E.-M. Jung, E. Ko, W. Han, D.-Y. Noh, NFIB is a potential target for estrogen receptor-negative breast cancers. *Mol. Oncol.* **5**, 538–544 (2011).
56. S. K. Denny, D. Yang, C.-H. Chuang, J. J. Brady, J. S. Lim, B. M. Grüner, S.-H. Chiou, A. N. Schep, J. Baral, C. Hamard, M. Antoine, M. Wislez, C. S. Kong, A. J. Connolly, K.-S. Park, J. Sage, W. J. Greenleaf, M. M. Winslow, Nfib promotes metastasis through a widespread increase in chromatin accessibility. *Cell* **166**, 328–342 (2016).
57. A. Matsuyama, M. Hisaoka, S. Shimajiri, T. Hayashi, T. Imamura, T. Ishida, M. Fukunaga, T. Fukuhara, H. Minato, T. Nakajima, S. Yonezawa, M. Kuroda, F. Yamasaki, S. Toyoshima, H. Hashimoto, Molecular detection of FUS-CREB3L2 fusion transcripts in low-grade fibromyxoid sarcoma using formalin-fixed, paraffin-embedded tissue specimens. *Am. J. Surg. Pathol.* **30**, 1077–1084 (2006).
58. Cancer Genome Atlas Research Network, Integrated genomic analyses of ovarian carcinoma. *Nature* **474**, 609–615 (2011).
59. K. M. Elias, M. M. Emori, T. Westerling, H. Long, A. Budina-Kolomets, F. Li, E. MacDuffie, M. R. Davis, A. Holman, B. Lawney, M. L. Freedman, J. Quackenbush, M. Brown, R. Drapkin, Epigenetic remodeling regulates transcriptional changes between ovarian cancer and benign precursors. *JCI Insight* **1**, e87988 (2016).
60. E. K. Adler, R. I. Corona, J. M. Lee, N. Rodriguez-Malave, P. Mhawech-Fauceglia, H. Sowter, D. J. Hazelett, K. Lawrenson, S. A. Gayther, The PAX8 cistrome in epithelial ovarian cancer. *Oncotarget* **8**, 108316–108332 (2017).
61. H. W. Cheung, G. S. Cowley, B. A. Weir, J. S. Boehm, S. Rusin, J. A. Scott, A. East, L. D. Ali, P. H. Lizotte, T. C. Wong, G. Jiang, J. Hsiao, C. H. Mermel, G. Getz, J. Barretina, S. Gopal, P. Tamayo, J. Gould, A. Tsherniak, N. Stransky, B. Luo, Y. Ren, R. Drapkin, S. N. Bhatia, J. P. Mesirov, L. A. Garraway, M. Meyerson, E. S. Lander, D. E. Root, W. C. Hahn, Systematic investigation of genetic vulnerabilities across cancer cell lines reveals lineage-specific dependencies in ovarian cancer. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 12372–12377 (2011).
62. M. Köbel, S. E. Kalloger, N. Boyd, S. McKinney, E. Mehl, C. Palmer, S. Leung, N. J. Bowen, D. N. Ionescu, A. Rajput, L. M. Prentice, D. Miller, J. Santos, K. Swenerton, C. B. Gilks, D. Huntsman, Ovarian carcinoma subtypes are different diseases: Implications for biomarker studies. *PLOS Med.* **5**, e232 (2008).
63. M. S. Alam, Proximity ligation assay (PLA). *Curr. Protoc. Immunol.* **123**, e58 (2018).
64. T. Zhang, N. Kwiatkowski, C. M. Olson, S. E. Dixon-Clarke, B. J. Abraham, A. K. Greifenberg, S. B. Ficarro, J. M. Elkins, Y. Liang, N. M. Hannett, T. Manz, M. Hao, B. Bartkowiak, A. L. Greenleaf, J. A. Marto, M. Geyer, A. N. Bullock, R. A. Young, N. S. Gray, Covalent targeting of remote cysteine residues to develop CDK12 and CDK13 inhibitors. *Nat. Chem. Biol.* **12**, 876–884 (2016).
65. S. Domcke, R. Sinha, D. A. Levine, C. Sander, N. Schultz, Evaluating cell lines as tumour models by comparison of genomic profiles. *Nat. Commun.* **4**, 2126 (2013).
66. M. Zeng, N. P. Kwiatkowski, T. Zhang, B. Nabet, M. Xu, Y. Liang, C. Quan, J. Wang, M. Hao, S. Palakurthi, S. Zhou, Q. Zeng, P. T. Kirschmeier, K. Meghani, A. L. Leggett, J. Qi, G. I. Shapiro, J. F. Liu, U. A. Matulonis, C. Y. Lin, P. A. Konstantinopoulos, N. S. Gray, Targeting MYC dependency in ovarian cancer through inhibition of CDK7 and CDK12/13. *eLife* **7**, e39030 (2018).
67. S. Hu, J. J. Marineau, N. Rajagopal, K. B. Hamman, Y. J. Choi, D. R. Schmidt, N. Ke, L. Johannessen, M. J. Bradley, D. A. Orlando, S. R. Alnemy, Y. Ren, S. Ciblat, D. K. Winter, A. Kabro, K. T. Sprott, J. G. Hodgson, C. C. Fritz, J. P. Carulli, E. di Tomaso, E. R. Olson, Discovery and characterization of SY-1365, a selective, covalent inhibitor of CDK7. *Cancer Res.* **79**, 3479–3491 (2019).
68. F. Farshidfar, S. Zheng, M.-C. Gingras, Y. Newton, J. Shih, A. G. Robertson, T. Hinoue, K. A. Hoadley, E. A. Gibb, J. Roszik, K. R. Covington, C.-C. Wu, E. Shinbrot, N. Stransky, A. Hegde, J. D. Yang, E. Reznik, S. Sadeghi, C. S. Pedamallu, A. I. Ojesina, J. M. Hess, J. T. Auman, S. K. Rhie, R. Bowlby, M. J. Borad; Cancer Genome Atlas Network, A. X. Zhu, J. M. Stuart, C. Sander, R. Akbani, A. D. Cherniack, V. Deshpande, T. Mounajjed, W. C. Foo, M. S. Torbenson, D. E. Kleiner, P. W. Laird, D. A. Wheeler, A. J. M. Ree, O. F. Bathe, J. B. Andersen, N. Bardeesy, L. R. Roberts, L. N. Kwong, Integrative genomic analysis of cholangiocarcinoma identifies distinct IDH-mutant molecular profiles. *Cell Rep.* **19**, 2878–2880 (2017).
69. S. A. Lambert, A. Jolma, L. F. Campitelli, P. K. Das, Y. Yin, M. Albu, X. Chen, J. Taipale, T. R. Hughes, M. T. Weirauch, The human transcription factors. *Cell* **175**, 598–599 (2018).
70. H. Li, R. Durbin, Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
71. Y. Zhang, T. Liu, C. A. Meyer, J. Eeckhoute, D. S. Johnson, B. E. Bernstein, C. Nusbaum, R. M. Myers, M. Brown, W. Li, X. S. Liu, Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
72. A. Sergushichev, An algorithm for fast preranked gene set enrichment analysis using cumulative statistic calculation. bioRxiv 060012 [**Preprint**]. 20 June 2016. https://doi.org/10.1101/060012.
73. E. Cerami, J. Gao, U. Dogrusoz, B. E. Gross, S. O. Sumer, B. A. Aksoy, A. Jacobsen, C. J. Byrne, M. L. Heuer, E. Larsson, Y. Antipin, B. Reva, A. P. Goldberg, C. Sander, N. Schultz, The cBio cancer genomics portal: An open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2**, 401–404 (2012).
74. R. I. Corona, J.-H. Seo, X. Lin, D. J. Hazelett, J. Reddy, F. Abassi, Y. G. Lin, P. Y. Mhawech-Fauceglia, J. Lester, S. P. Shah, D. G. Huntsman, A. Gusev, B. Y. Karlan, B. P. Berman, M. L. Freedman, S. A. Gayther, K. Lawrenson, Non-coding somatic mutations converge on the PAX8 pathway in epithelial ovarian cancer. bioRxiv 537886 [**Preprint**]. 1 February 2019. https://doi.org/10.1101/537886.
75. G. Yu, Q.-Y. He, ReactomePA: An R/Bioconductor package for reactome pathway analysis and visualization. *Mol. Biosyst.* **12**, 477–479 (2016).
76. A. Boija, I. A. Klein, B. R. Sabari, A. Dall'Agnese, E. L. Coffey, A. V. Zamudio, C. H. Li, K. Shrinivas, J. C. Manteiga, N. M. Hannett, B. J. Abraham, L. K. Afeyan, Y. E. Guo, J. K. Rimel, C. B. Fant, J. Schuijers, T. I. Lee, D. J. Taatjes, R. A. Young, Transcription factors activate

genes through the phase-separation capacity of their activation domains. *Cell* **175**, 1842–1855.e16 (2018).
77. D. Schmidt, M. D. Wilson, C. Spyrou, G. D. Brown, J. Hadfield, D. T. Odom, ChIP-seq: Using high-throughput sequencing to discover protein-DNA interactions. *Methods* **48**, 240–248 (2009).
78. B. Langmead, C. Trapnell, M. Pop, S. L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
79. A. R. Quinlan, I. M. Hall, BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
80. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin; 1000 Genome Project Data Processing Subgroup, The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
81. K. L. Parker, B. P. Schimmer, Steroidogenic factor 1: A key determinant of endocrine development and function. *Endocr. Rev.* **18**, 361–377 (1997).
82. S. Zheng, A. D. Cherniack, N. Dewal, R. A. Moffitt, L. Danilova, B. A. Murray, A. M. Lerario, T. Else, T. A. Knijnenburg, G. Ciriello, S. Kim, G. Assie, O. Morozova, R. Akbani, J. Shih, K. A. Hoadley, T. K. Choueiri, J. Waldmann, O. Mete, A. G. Robertson, H.-T. Wu, B. J. Raphael, L. Shao, M. Meyerson, M. J. Demeure, F. Beuschlein, A. J. Gill, S. B. Sidhu, M. Q. Almeida, M. C. B. V. Fragoso, L. M. Cope, E. Kebebew, M. A. Habra, T. G. Whitsett, K. J. Bussey, W. E. Rainey, S. L. Asa, J. Bertherat, M. Fassnacht, D. A. Wheeler; Cancer Genome Atlas Research Network, G. D. Hammer, T. J. Giordano, R. G. W. Verhaak, Comprehensive pan-genomic characterization of adrenocortical carcinoma. *Cancer Cell* **29**, 723–736 (2016).
83. D. Lindgren, P. Eriksson, K. Krawczyk, H. Nilsson, J. Hansson, S. Veerla, J. Sjölund, M. Höglund, M. E. Johansson, H. Axelson, Cell-type-specific gene programs of the normal human nephron define kidney cancer subtypes. *Cell Rep.* **20**, 1476–1489 (2017).
84. B. Porath, S. Livingston, E. L. Andres, A. M. Petrie, J. C. Wright, A. E. Woo, C. G. Carlton, R. Baybutt, G. B. Vanden Heuvel, Cux1 promotes cell proliferation and polycystic kidney disease progression in an ADPKD mouse model. *Am. J. Physiol. Renal Physiol.* **313**, F1050–F1059 (2017).
85. N. Naiman, K. Fujioka, M. Fujino, M. T. Valerius, S. S. Potter, A. P. McMahon, A. Kobayashi, Repression of interstitial identity in nephron progenitor cells by Pax2 establishes the nephron-interstitium boundary during kidney development. *Dev. Cell.* **41**, 349–365.e3 (2017).
86. R. Morizane, A. Q. Lam, B. S. Freedman, S. Kishi, M. T. Valerius, J. V. Bonventre, Nephron organoids derived from human pluripotent stem cells model kidney development and injury. *Nat. Biotechnol.* **33**, 1193–1200 (2015).
87. D. Su, S. Ellis, A. Napier, K. Lee, N. R. Manley, Hoxa3 and pax1 regulate epithelial cell death and proliferation during thymus and parathyroid organogenesis. *Dev. Biol.* **236**, 316–329 (2001).
88. G. Holländer, J. Gill, S. Zuklys, N. Iwanami, C. Liu, Y. Takahama, Cellular and molecular events during early thymus development. *Immunol. Rev.* **209**, 28–46 (2006).
89. R. Vishnubalaji, R. Elango, M. Al-Toub, M. Manikandan, A. Al-Rikabi, L. Harkness, N. Ditzel, M. Atteya, R. Hamam, M. Alfayez, A. Aldahmash, M. Kassem, N. M. Alajez, Neoplastic transformation of human mesenchymal stromal cells mediated via LIN28B. *Sci. Rep.* **9**, 8101 (2019).
90. O. H. Ocaña, R. Córcoles, A. Fabra, G. Moreno-Bueno, H. Acloque, S. Vega, A. Barrallo-Gimeno, A. Cano, M. A. Nieto, Metastatic colonization requires the repression of the epithelial-mesenchymal transition inducer Prrx1. *Cancer Cell* **22**, 709–724 (2012).

Acknowledgments

Funding: This project was supported by an Ovarian Cancer Research Fund Alliance Liz Tilberis Early Career Award (599175) (to K.L.), an American Cancer Society Research Scholars Grant (134005-RSG-19-135-01-DMC to K.L.), Ovarian Cancer Research Fund Alliance Program Project Development (373356) (to B.Y.K. and K.L.), the American Lebanese Syrian Associated Charities (to B.J.A.), a Southern California Clinical and Translational Science Institute Core Voucher (V148) (to K.L. and R.I.C.), and grants CA155258 and CA213333 (to R.A.Y.). The research described was supported in part by NIH/National Center for Advancing Translational Science (NCATS) UCLA CTSI grant number UL1TR001881. R.N. is supported in part by a Ruth L. Kirschstein Institutional National Research Service Award (T32) from the NIH (grant number 5 T32 GM 118288-2). I.A.K. was supported by an American Society of Clinical Oncology Young Investigator Award, an American Cancer Society Postdoctoral Fellowship, and an Ovarian Cancer Research Alliance Mentored Investigator Award. U.M. is funded by Ovarian Ca SPORE P50CA240243 and Breast Cancer Research Foundation FP50P50CA240243P50CA240243P50CA24. D.C.-M. was also supported by an Ovarian Cancer Research Alliance Mentored Investigator Award. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. This work used equipment and services provided at the Cedars-Sinai Biobank and Translational Research Core and the Applied Genomics, Computation and Translation (AGCT) Core. The results shown here are in whole or part based upon data generated by the TCGA Research Network: www.cancer.gov/tcga, and the ICGC's PanCancer Analysis of Whole Genomes (PCAWG; https://pcawg.icgc.org/pcawg) consortium. Ovarian tumor specimens were collected as by the Jean Richardson Gynecologic Tissue and Fluid Repository at the University of Southern California, as part of the Women's Cancer Program at CSMC, or as part of the Dana-Farber Cancer Institute ovarian tumor banking protocol. **Author contributions:** K.L. and J.R. directed the study. M.A.S.F., R.I.C., F.S.D., and J.R. led data analysts and data curation. T.M.M., H.N., T.S., A.F.-T., P.C., J.-H.S., A.G., M.L.F., D.-C.L., B.Y.K., U.M., I.A.K., L.K.A., B.J.A., and R.A.Y. provided data and/or samples. R.N., H.C., R.I.C., F.A., X.L., I.A.K., L.K.A., B.J.A., R.A.Y., D.C.-M., and R.D did the functional validation. K.L. conceived, designed, supervised, and implemented the project. J.R., K.L., M.A.S.F., and R.I.C. drafted the manuscript. All authors reviewed and approved the final draft. **Competing interests:** R.A.Y. is a founder and shareholder of Syros Pharmaceuticals, Camp4 Therapeutics, Omega Therapeutics, and Dewpoint Therapeutics. I.A.K. is an officer, employee, and shareholder of Dewpoint Therapeutics. B.J.A. is a shareholder in Syros Pharmaceuticals. U.M. reports conflicts of interest with Merck, AstraZeneca, Blueprint Medicines, NextCure, and Novartis. All other authors declare that they have no competing interests. **Data and materials availability:** Raw and processed H3K27ac and TF ChIP-seq data generated for this study have been deposited in GEO with accession code GSE152885. RNA-seq data from MTF knockdown (KD) experiments have been deposited in GEO with accession code GSE150443. The CaCTS algorithm was developed as an R package and has been deposited in GitHub (https://github.com/lawrenson-lab/CaCTS) and Zenodo (https://doi.org/10.5281/zenodo.5234007). All other data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials.

Submitted 12 November 2020
Accepted 5 October 2021
Published 24 November 2021
10.1126/sciadv.abf6123