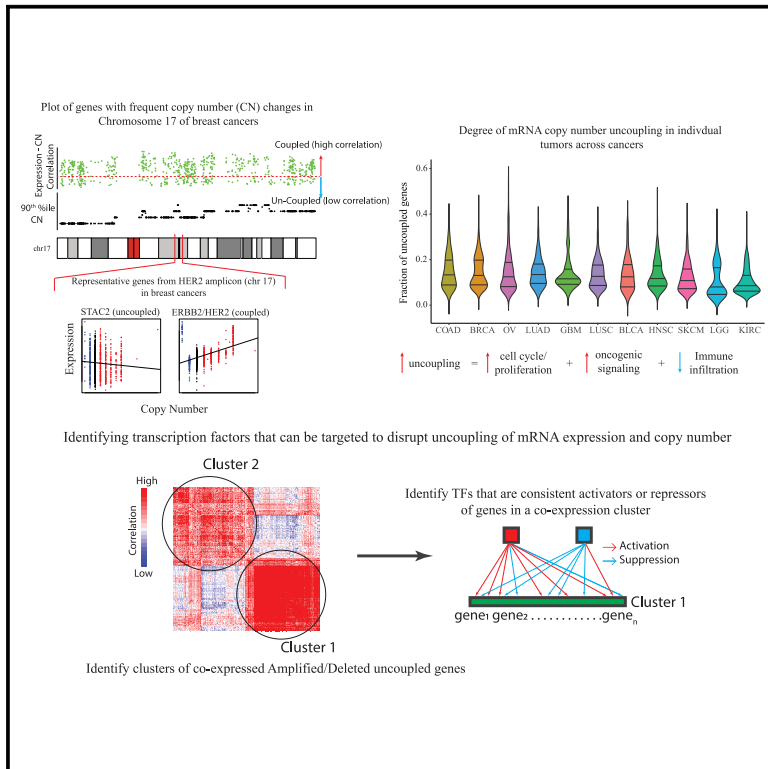


Uncoupling of gene expression from copy number presents therapeutic opportunities in aneuploid cancers

Graphical abstract



Authors

Vakul Mohanty, Fang Wang, Gordon B. Mills, CTD2 Research Network, Ken Chen

Correspondence

kchen3@mdanderson.org

In brief

Mohanty et al. show that uncoupling of mRNA expression from copy number may be an important mechanism to maintain tumor fitness in aneuploid cancers, orchestrated by complex regulatory changes. These events present a potential therapeutic avenue and can be targeted by perturbing appropriate regulatory molecules, such as transcription factors.

Highlights

- Gene expression is frequently uncoupled from copy number in cancers
- Uncoupling orchestrated by complex regulatory changes improves tumor fitness
- Targeting transcription factors can reestablish coupling and reduce tumor fitness



Article

Uncoupling of gene expression from copy number presents therapeutic opportunities in aneuploid cancers

Vakul Mohanty,¹ Fang Wang,^{1,3} Gordon B. Mills,² CTD2 Research Network, and Ken Chen^{1,4,*}¹Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA²Cell, Developmental and Cancer Biology, Knight Cancer Institute, Oregon Health University, Portland, OR 97201, USA³Present address: Precision Medicine Institute, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou 510080, China⁴Lead contact*Correspondence: kchen3@mdanderson.org<https://doi.org/10.1016/j.xcrm.2021.100349>**SUMMARY**

Uncoupling of mRNA expression from copy number (UECN) might be a strategy for cancer cells to tolerate high degree of aneuploidy. To test the extent and role of UECN across cancers, we perform integrative multiomic analysis of The Cancer Genome Atlas (TCGA) dataset, encompassing ~5,000 individual tumors. We find UECN is common in cancers and is associated with increased oncogenic signaling, proliferation, and immune suppression. UECN appears to be orchestrated by complex regulatory changes, with transcription factors (TFs) playing a prominent role. To further dissect the regulatory mechanisms, we develop a systems-biology approach to identify candidate TFs, which could serve as targets to disrupt UECN and reduce tumor fitness. Applying our approach to TCGA data, we identify 21 putative targets, 42.8% of which are validated by independent sources. Together, our study indicates that UECN is likely an important mechanism in development of aneuploid tumors and might be therapeutically targetable.

INTRODUCTION

Aneuploidy is commonly observed in cancers.^{1,2} Structural changes induced by aneuploidy are thought to increase tumor fitness by targeting driver genes, i.e., increasing the copy number (CN) of oncogenes and decreasing those of tumor suppressors, respectively. However, as these structural changes impact large chromosomal segments, they often result in collateral CN changes of hundreds to thousands of genes.¹⁻⁴ These genes are thought to be passenger events with minimal contribution to tumor phenotype.³ Many of the co-altered genes appear to have uncoupling of mRNA expression from copy number^{5,6} (UECN), suggesting that their CN changes are likely deleterious to tumor fitness. Although the landscapes of UECN have been previously surveyed,⁵ the roles UECN plays to modulate the phenotype of individual tumors remain poorly understood.

Gene expression is regulated by a number of factors including gene CN, changes in transcription factor (TF) activity, DNA methylation, and microRNAs (miRNAs), to name a few. These factors have different effect sizes on gene expression. However, their combinatorial effects, which may lead to UECN, have been underexplored. As a consequence of dissociation between gene CN and mRNA expression, UECN provides a powerful approach to identify genes that are deleterious to tumor pathophysiology. In UECN, the effects of CN on gene expression are overridden by changes in other regulatory factors.

Fitness screens using large small interfering RNA (siRNA)/CRISPR libraries have identified genes that are context essential for maintaining tumor cell viability,⁷⁻⁹ such as dependency of MYC-driven tumors on spliceosome¹⁰ and HER2+ breast cancers on phosphatidylinositol 3-kinase (PI3K)/mTOR signaling.¹¹ Unfortunately, exploring the functional impact of UECN through fitness screens can be hampered by multiple issues. First, amplified genes that are detrimental to tumor fitness and are silenced are unlikely to be identified by knockdown (KD) or knockout (KO) screens. Capturing their influence on fitness would require overexpression (OE). Second, current whole-genome screens are geared toward cell intrinsic viability and proliferation and cannot realistically examine cell-extrinsic factors such as the host immune system.

Dysregulation of TF activity is widely observed across cancers,¹² and targeting TFs has gained traction with the development of novel modalities to modulate TF activity.¹² Computational tools such as VIPER¹³ allow inference of protein level activity of TFs from RNA data, which provides more ubiquitous proteome coverage than currently available assays such as reverse phase protein arrays (RPPAs) and mass spectrometry. Applications of these approaches have facilitated characterization of aberrant activity of driver TFs and led to identification of likely therapeutic targets. We propose to extend this concept by identifying TF targets that can offset the effects of UECN. These TFs can be perturbed to reestablish CN-dependent expression of specific genes or gene sets, reconstituting the



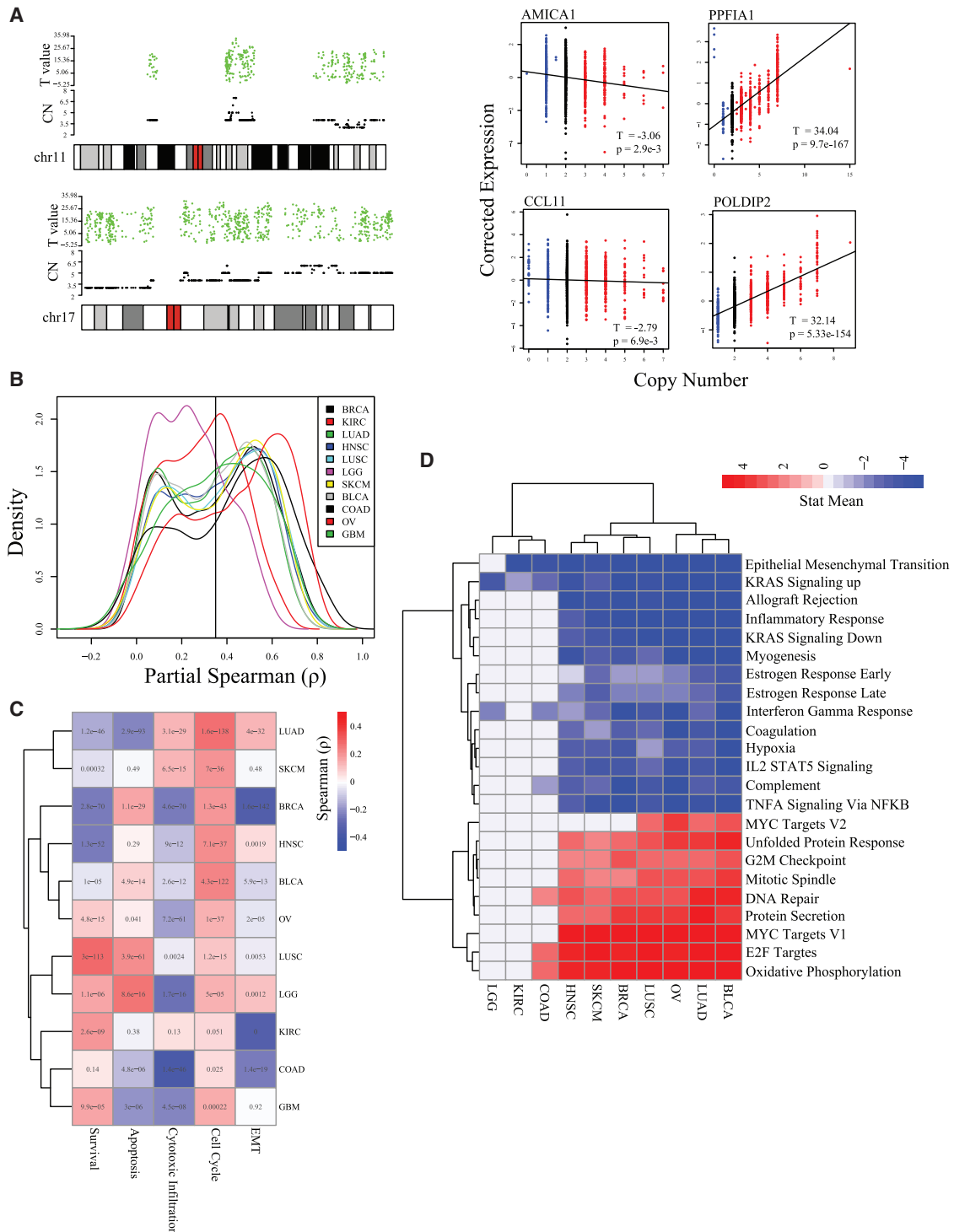


Figure 1. Functional characterization of CpD and nCpD genes

(A) Chromosomal plot of frequently amplified/deleted genes in breast cancer on chromosome (chr) 11 and chr17. The 90th percentile copy number (CN) for all genes is plotted in black, T_{CN} (see STAR Methods) is plotted in green (left), and purity-corrected mRNA expression is plotted against absolute CN for genes with highest and lowest T_{CN} on chr11 (*AMICA1* and *PPFIA1*) and chr17 (*CCL11* and *POLDIP2*). Regression T_{CN} and p values are reported.

(B) Density plot of partial Spearman correlation coefficient (ρ) between gene expression and CN of genes frequently amplified/deleted in each cancer. Note the bimodal distribution ρ in most cancers.

(legend continued on next page)

tumor toxic nature of those CN effects. For instance, genes in the antigen presentation complex are frequently amplified, but silenced in melanomas.⁵ By targeting the appropriate transcriptional regulators, these genes can be re-expressed in a CN-dependent manner that, in turn, is likely to result in improved anti-tumor immunity. However, TFs generally regulate large numbers of genes; thus, the inherent challenge is to identify TFs that primarily regulate uncoupled genes but also have minimal off-target effects. We seek to address these challenges by creating network biological models from multiomics data to predict phenotypic effects of perturbing specific TFs.

In this study, we examined multiomics data from more than 5,000 samples across 11 cancer types in The Cancer Genome Atlas (TCGA) as well as the complete CCLE (Cancer Cell Line Encyclopedia) compendium.¹⁴ We present a comprehensive functional and phenotypic characterization of the impact that UECN has on tumor phenotype, by quantifying uncoupling events in individual tumor samples across cancer types. Furthermore, we use machine learning to construct cancer-specific regulatory networks to elucidate regulatory mechanisms that govern UECN. We present an analytical framework to leverage these regulatory networks to identify regulatory perturbations that reverse tumor-toxic UECNs to identify therapeutic targets for aneuploid cancers. Finally, we present TF targets that we identified and validated *in silico* based on their association with patient survival and response to immune checkpoint therapy.

RESULTS

Disconnect between gene CN and expression is ubiquitous in cancers

Aneuploidy results in frequent CN changes of hundreds of genes in the tumor genome (Figure S1A). These changes can be focal (Figure 1A, top) as well as affecting large chromosomal regions (Figures 1A, bottom; and S1B).² These changes are thought to be selected because they carry driver genes. However, they also result in the collateral co-amplification/deletion of surrounding genes (Figures 1A and S1C). Conventionally, changes in CN are thought to result in proportional changes in expression of the gene. By contrast, we find that genes within the same amplicon show a great degree of heterogeneity in terms of the correlation between expression and CN (see STAR Methods; Figures 1A and S1B). To assess this phenomenon at a global level for frequently amplified and deleted genes (see STAR Methods), we used two complementary approaches: (1) partial correlation between CN and expression while controlling for tumor purity along with the top 20 expression PCs (principle components) in each cancer; and (2) regressing expression against CN while controlling for tumor purity and the top 20 expression PCs, where the strength of association between expression and CN is defined by the Spearman correlation coefficient (ρ_{CN}) and the

T-value corresponding to the CN term in the regression model (T_{CN}), respectively. Only genes that were consistently expressed in each tumor (90th percentile expression > 30 normalized counts) were used in the analysis. The density distribution of both terms (ρ_{CN} and T_{CN}) has a bimodal distribution in multiple cancers (Figures 1B and S1D, left). This suggests that while the expression of numerous genes is coupled with their CN (CpD, coupled genes), the expression of a sizable proportion of genes is uncoupled from their CN (nCpD, uncoupled genes) (Figure S1A; see STAR Methods). We also localized each nCpD gene to recurrent focal/regional CNVs (CN variants) and chromosome arm level CNVs (see STAR Methods). On average, 40% of nCpD genes were associated with arm-level CN changes; in contrast, 0.3% of nCpD genes was found associated with focal/regional recurrent CNVs (Figure S1E).

Tumor purity due to the presence of normal cells in the tumor micro-environment could confound detection of UECN due to dilution of CN and expression changes by normal cells. In the above-mentioned analysis with bulk TCGA tumors, we explicitly controlled for purity. We further performed CN-expression correlation analysis in CCLE¹⁴ cell lines to nullify effects of tumor purity, where we observed a similar bimodal distribution of T_{CN} (see STAR Methods; Figure S1F), suggesting that UECN is not an artifact of tumor purity. Other factors that can confound identification of UECN are (1) a gene may have low or no expression in the tumor's corresponding normal tissue, which could lead to false identification of UECN in the tumor; and (2) genes whose expression changes with differentiation status can also confound identification of UECN. For instance, an amplified gene, expressed in differentiated tissues, might be turned off not because its expression is deleterious to the tumor but because samples carrying amplifications of this gene are less differentiated. To address these issues, we excluded genes with low expression in normal tissue corresponding to each tumor (median expression <5 transcripts per million [TPM]) and genes whose expression changes as a function of stemness (see STAR Methods) and performed expression-CN regression analysis. We were able to recapture bimodal distribution of expression-CN association in several cancers (Figure S1D, middle), with numerous genes showing low dependence of expression on CN. In some cancers, such as colon adenocarcinoma (COAD) and ovarian cancer (OV), the bimodal distribution isn't as striking as previously observed (Figure S1D, left). We are however still able to identify nCpD genes in these cancers (Figure S1D, middle). These findings indicate that UECN is a real phenomenon observed in cancers and not an artifact of tumor purity, differentiation, or low expression of genes.

Each cancer type in TCGA consists of various distinct molecular subtypes. It is likely that the uncoupling we observe could be driven by specific tumor subtypes. To test whether this is the case, we performed CN-expression correlation analysis for

(C) Heatmap of ρ between T_{CN} and strength of association of gene expression with phenotype (see STAR Methods) for frequently amplified genes in each cancer type. Note: as -Ve Cox Z scores are associated with improved survival, these Z scores are multiplied by -1 to maintain consistency with other phenotype scores. Text in the box is the corresponding p value.

(D) GSEA of amplified CpD and nCpD genes using T_{CN} (see STAR Methods); significant pathways are identified at $q < 0.05$. Blue indicates enrichment of the pathway in nCpD genes, while red indicates enrichment of the pathway in CpD genes. Note pathways identified in at least 4 cancers are plotted here to show consistent trends.

frequently amplified and deleted genes in each cancer subtype with at least 50 samples and 1,000 genes with frequent CN changes (see STAR Methods). Similar to our observations mentioned above, T_{CN} in all tested subtypes has a wide distribution, covering both +Ve (positive) and -Ve (negative) T_{CN} values, and many cancer subtypes also show a bimodal distribution (Figure S1D, right), suggesting that the observed UECN is not introduced by having multiple subtypes but by tumor intrinsic biology.

UECN is a mechanism of maintaining tumor fitness

Structural changes often affect large chromosomal segments, resulting in collateral CN changes of numerous genes (Figures 1 and S1B). It is likely that a subset of these CN changes can be detrimental to tumor fitness. UECN of such genes would therefore negate effects of toxic CN changes and allow tumor cells to tolerate aneuploidy.

The chromosome 6p is commonly amplified in breast adenocarcinoma (BRCA), resulting in co-amplification of numerous genes including *TAP1* and *XPO5*, providing an instance of targeted uncoupling of expression from CN (Figures S1B and S1G). *TAP1* is involved in major histocompatibility complex class I (MHC class 1)-mediated antigen presentation, playing a critical role in anti-cancer immunity.¹⁵ Elevated *TAP1* is therefore associated with increased immune infiltration and cytotoxicity (Figures S1H and S1I). Amplification of *TAP1* would therefore be detrimental to tumor fitness, and its expression is thus weakly correlated with its CN (Figure S1B). By contrast, *XPO5*, which plays a role in the pathogenesis of multiple cancers,^{16–18} shows a strong coupling of CN and expression (Figure S1B).

To systematically characterize the association of nCpD genes with fitness, we looked at the correlation between T_{CN} and phenotype scores (survival, apoptosis, cytotoxic immune infiltration, cell cycle, and epithelial-to-mesenchymal transition [EMT]; see STAR Methods) of amplified and deleted genes, respectively. A positive correlation indicates CpD genes are associated with the phenotype, while a negative correlation indicates nCpD genes are associated with the phenotype. In cases of amplified genes, we found that T_{CN} shows a consistent positive correlation with cell cycle and EMT, in contrast to a negative correlation with patient cytotoxic immune infiltration and to a lesser extent with survival and apoptosis (Figure 1C). This suggests that while expression of amplified CpDs is associated with pro-tumor phenotypes, nCpDs are associated with anti-tumor phenotypes (Figure 1C). These trends were predominantly reversed in the cases of deleted genes (Figure S1J). These phenotypic associations were well conserved when the analysis was repeated after excluding genes not expressed in normal tissue corresponding to the tumor and genes whose expression changes as a function of stemness (Figures S1J and S1K). Looking at pathways enriched for CpD and nCpD genes (see STAR Methods; Figure 1D), we found that amplified CpD genes across cancers were associated with pro-oncogenic pathways such as MYC signaling,¹⁹ unfolded protein response (UPR) response,²⁰ and oxidative phosphorylation.²¹ By contrast, nCpD genes were enriched for pathways associated with immune response across cancers (Figure 1D). In addition, functional analysis in CCLE cell lines, tumor subtypes, and tumor types, while excluding genes that can confound identification of UECN, showed similar patterns of

function enrichment in amplified CpDs and nCpDs (Figures S1M–S1O). These data suggest that CN changes in nCpD genes are likely to be tumor toxic; therefore, their expression is actively uncoupled from their CN.

Assessing the impact of UECN at the level of individual samples

To quantify the degree of uncoupling (D_{UC}) at the level of individual samples, we first identified the top 100 CpD genes in each cancer. These genes were used to build a linear model to quantify how gene expression is expected to change for a single altered CN in each tumor (see STAR Methods). This model was then applied to all frequently amplified and deleted genes to identify samples in which their expression was uncoupled from CN, i.e., the actual expression was lower or greater than expected expression in the case of amplifications and deletions, respectively (see STAR Methods). nCpD genes were uncoupled in more samples relative to CpD genes (Figure S2A), suggesting that our approach of defining uncoupled genes at the level of individual samples could recapture CpD and nCpD genes identified by population-level analysis. Figure S2B illustrates uncoupled samples in *PSMD12* (CpD) and *STAC2* (nCpD) genes often amplified with *ERBB2* in BRCA.⁵

We defined the D_{UC} in a tumor as the ratio of the number of uncoupled genes to the number of genes with CN changes. COAD showed the highest median D_{UC} , while it was lowest in kidney renal clear cell carcinoma (KIRC; Figure 2A). The distribution of D_{UC} was fairly similar across cancers, although some cancers such as low grade glioma (LGG) and BRCA show distinct groups with high and low D_{UC} (Figure 2A). We also analyzed how D_{UC} varied between subtypes in each cancer (see STAR Methods). The distribution of D_{UC} across subtypes was similar in some cancer types, e.g., skin cutaneous melanoma (SKCM), OV, and glioblastoma multiforme (GBM), or distinctive in other cancer types, e.g., BRCA, head and neck squamous cell (HNSC) carcinoma, KIRC, and LGG (Figure S2C). For example, luminal A BRCA have much lower D_{UC} levels relative to other subtypes. It is known that luminal A BRCA are less aggressive and have better clinical prognosis.²² Furthermore, luminal A BRCA have a lower incidence of CN aberrations and are more likely to carry driver mutations than other breast cancer subtypes. Another example is basal HNSC tumors, which have higher D_{UC} than atypical and mesenchymal HNSC tumors. Basal tumors constitute most tumors that carry co-amplification of 11q13/q22.²³

In 7 of 11 cancers tested, D_{UC} showed a strong positive correlation with aneuploidy index (Figure 2B) from Davoli et al.²⁴ This was to be expected. With increased aneuploidy, the likelihood of tumor-toxic CN changes increases, resulting in greater compensatory uncoupling. Intriguingly, both brain cancers analyzed (GBM and LGG) show low correlation between D_{UC} and aneuploidy. As nCpD genes are predominantly associated with immune signaling, it is likely that these CN changes aren't tumor toxic in brain tumors because of low immune infiltration,²⁵ thus negating the requirement for UECN in aneuploidy brain tumors. In a subset of cancers—LUAD, OV, HNSC, and lung squamous cell (LUSC) carcinoma— D_{UC} also showed a significant negative correlation with immune signature score (Figure 2B).²⁴ We further correlated D_{UC} with sample-level phenotypic score (for

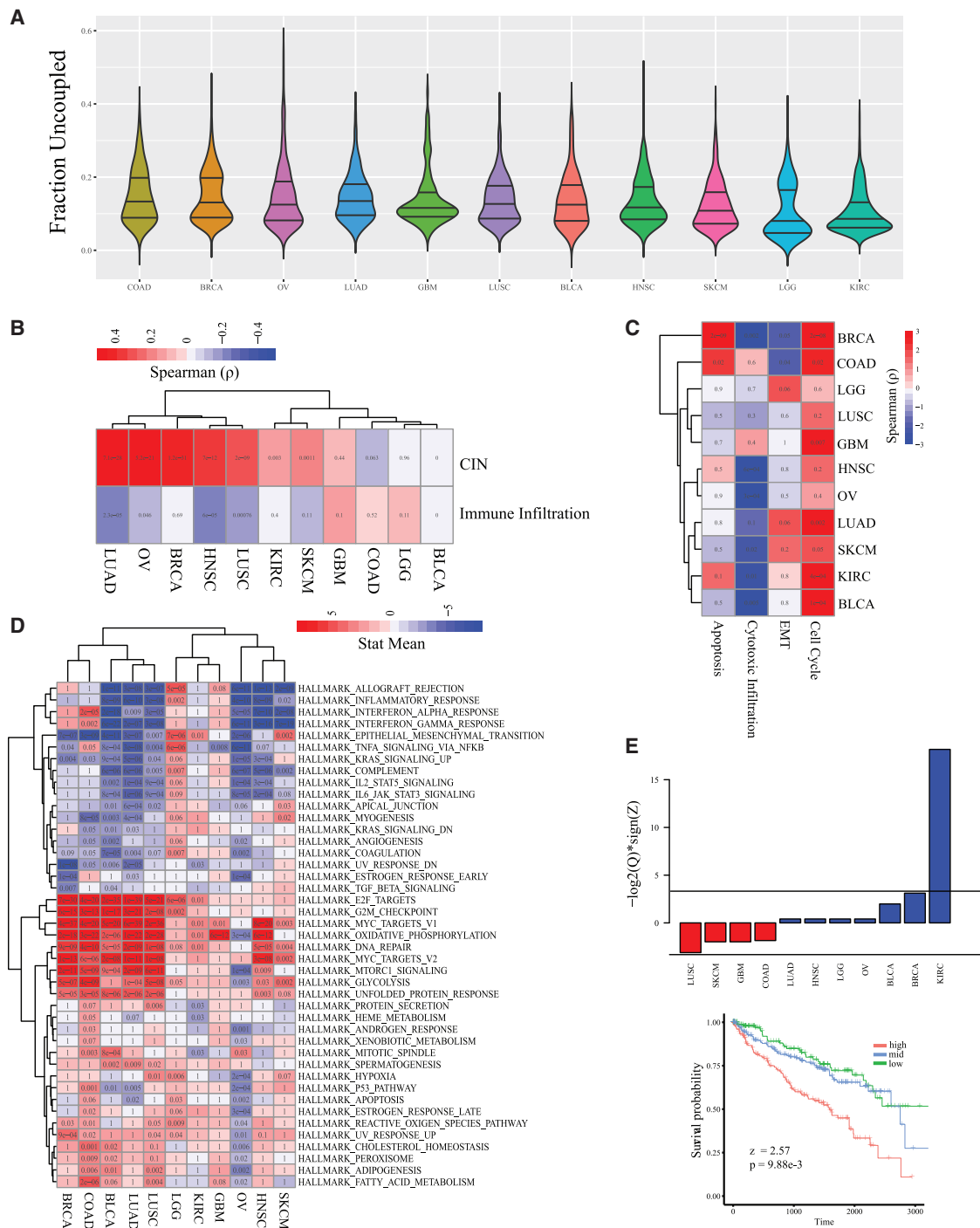


Figure 2. Quantification of uncoupling at the level of individual samples

(A) Distribution D_{UC} across cancers sorted by median D_{UC} . The lines in the violin plots indicate 25th, 50th, and 75th quantiles.

(B) Heatmap of ρ of D_{UC} with degree of aneuploidy (chromosomal instability) and immune signature scores. Text in each cell is the p value.

(C) Heatmap of ρ between D_{UC} and phenotype scores. Text in heatmap is p value.

(D) Differentially expressed pathways in samples with high D_{UC} relative to those with low D_{UC} (see STAR Methods). Red indicates overexpression and blue indicates suppression. Text in each cell is the q-value, with significance at $q < 0.1$.

(E) The barplot is of $-\log_2(q\text{-value}) \times \text{sign}(Z)$ from univariate survival analysis. Significance is at $q < 0.1$ indicated by horizontal lines (left). Kaplan-meier plots for KIRC, reported p value, and Z score are from a multivariate survival analysis.

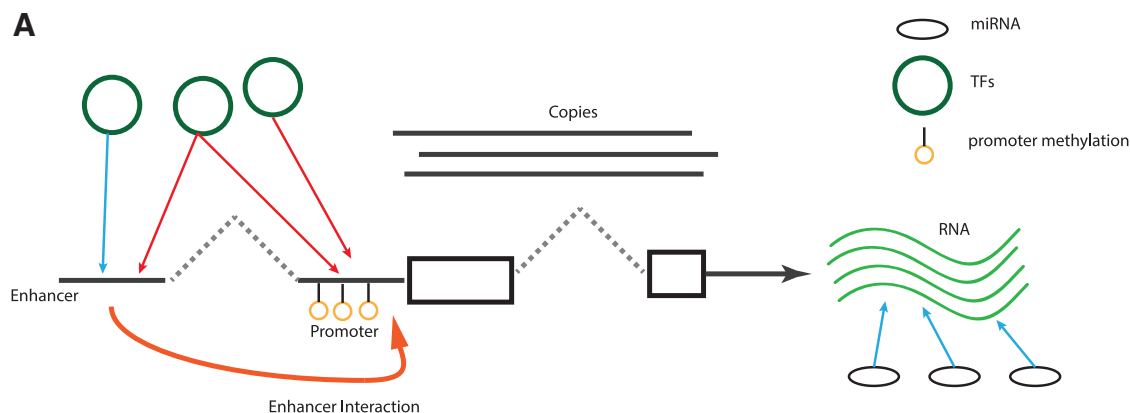


Figure 3. Pictorial representation of factors that can influence expression of a gene

Gene expression is modeled as function of CN, promoter methylation, transcription factor (TF) binding at promoter and active enhancer regions, and miRNA binding.

apoptosis, cytotoxic immune infiltration, EMT, and cell cycle) and found that D_{UC} was strongly associated with increased cell-cycle signature and decreased cytotoxic activity in tumors (Figure 2C). These phenotypic associations were further reinforced by differential expression and GSEA (gene set enrichment analysis), by comparing samples with high and low D_{UC} (see STAR Methods). Samples with high D_{UC} were associated with increased activity of oncogenic pathways (MTORC1 signaling, glycolysis,²⁶ UPR response, and MYC signaling, among others), while the downregulated pathways consisted of innate and adaptive immune response pathways (Figure 2D). We also quantified the association between D_{UC} and clinical survival. KIRC was the only cancer where mono-variant survival analysis revealed an association with survival ($q < 0.1$; Figure 2E, top). Subsequent multivariate analysis showed that high D_{UC} in KIRC was associated with poor clinical survival (Figure 2E, bottom).

The strong positive correlation observed between D_{UC} and aneuploidy indicates that large structural changes that influence the CN of a large number of genes are frequently accompanied by increased gene expression and CN uncoupling, likely to negate an increase in the number of tumor-toxic CN changes and maintain tumor fitness. This is characterized by the increased activity of oncogenic pathways and immune-suppressive phenotypes that we observe in tumor samples with high D_{UC} (Figures 2C and 2D).

UECN is mediated by complex regulatory changes

To elucidate the regulatory mechanisms underlying UECN, we built transcription regulatory networks in six cancers (BRCA, HNSC, LUAD, LUSC, LGG, and SKCM). For each expressed gene in a cancer, we modeled its expression as a function of its CN, promoter methylation, regulating miRNAs, and TFs that have binding sites in the promoter and enhancers. The model also accounted for the binding strength of miRNAs and TFs as well as enhancer activity. It was fit using an elastic net and merged to construct cancer-specific regulatory networks (Figure 3; see STAR Methods). We compared regulatory strength, here the β -values of regulators obtained from the regression models, between nCpD and CpD genes and found (1) CN β -values were

lower for nCpD genes, suggesting their expression is relatively insensitive to CN changes (Figure S3A) as indicated by our correlation analysis (Figures 1B and S1D); (2) promoter methylation β -values of amplified nCpD genes are more negative than those of amplified CpD genes, suggesting the nCpD genes are more strongly suppressed by promoter methylation (Figure S3B); (3) TF regulators of nCpD genes show higher β -values, although the difference is modest (Figure S3C); and (4) no significant differences appeared between β -values of miRNAs regulating nCpD and CpD genes (Figure S3D).

DNA methylation,^{27,28} TFs,¹² and miRNA²⁹ are often dysregulated in cancers and mediate downstream changes in the tumor transcriptome. To assess the role of these regulatory factors in UECN, for each uncoupled gene, we compared (1) promoter methylation, (2) top activating TF/miRNA, and (3) top inactivating TF/miRNA levels between uncoupled and coupled samples. We found that in the context of amplified nCpD genes, methylation and negative regulators are generally higher in uncoupled samples, while activating regulators are suppressed. The trends are reversed in the context of deleted nCpD genes (Figure 4A; Table S1). This observation indicates that UECN is mediated by a complex orchestration of changes in epigenetic (promoter methylation), transcriptional, and post-transcriptional regulation. This is illustrated in the case of the nCpD gene *TAP1* in BRCA (Figure S3E). Comparing the expression of regulators of *TAP1* between uncoupled and coupled samples, we find significant differential expression of 5 TFs (Figure S3F), with the four activating TFs showing lower expression in the uncoupled samples, while the deactivating TFs showed OE (Figure S3G).

To systematically quantify the extent to which methylation, TFs, and miRNAs contribute to UECN, we compared promoter methylation and expression of regulatory TFs and miRNAs between uncoupled and coupled samples for every nCpD gene and determined whether their activity levels were significantly changed in a direction that could explain uncoupling based in a regulatory context (see STAR Methods). For instance, in the case of an amplified uncoupled gene, samples where the gene's expression is uncoupled from its CN could have increased promoter methylation or decreased expression of an activating

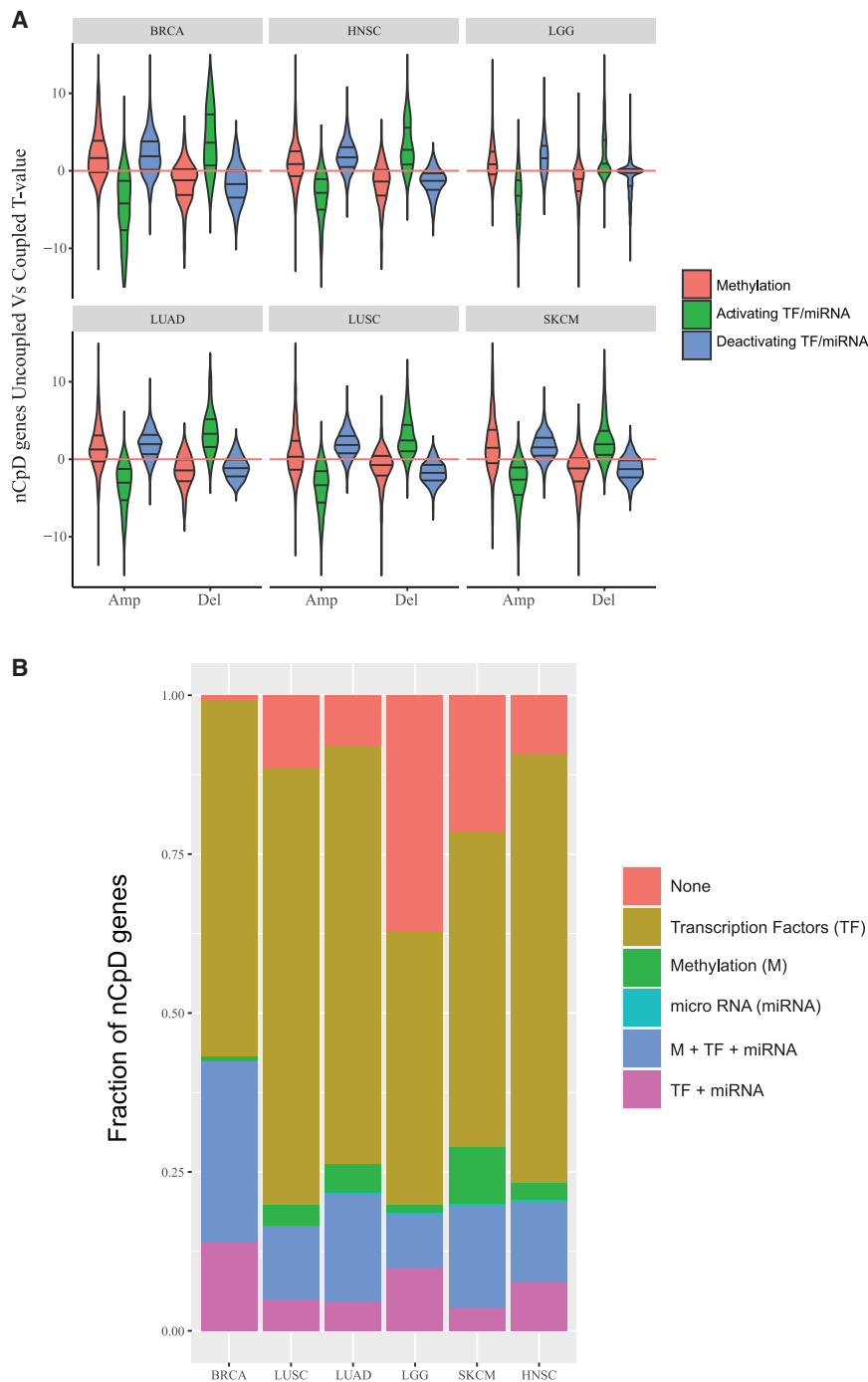


Figure 4. Regulatory trends associated with gene CN and expression uncoupling

(A) Violin plots of T-values comparing levels of promoter methylation and top regulators of amplified and deleted nCpD genes between coupled and uncoupled samples across 6 cancers. The lines in the violin plots indicate 25th, 50th, and 75th quantiles. The red line marks T-value = 0. Pairwise Tukey’s test q-values are in Table S1.

(B) Barplot of fraction of nCpD genes in each cancer where uncoupling can be explained by a regulator or a combination of regulators.

the dominant factor being regulation by TFs. On average, 81.5% of nCpD genes had at least one TF differentially expressed between uncoupled and coupled samples in a direction that could explain uncoupling.

Targeting cancers by reestablishing expression coupling of silenced tumor-toxic gene CN changes

Since transcriptional control mediated much of the uncoupling (Figure 4B), we hypothesized that by targeting appropriate TFs, we can reverse UECN by reestablishing expression coupling of tumor-toxic gene CN changes. This should result in a flux of signaling—likely strengthened by the CN changes—that is detrimental to tumor fitness. To identify such TFs, we propose the following analytical framework (see STAR Methods; Figures S4A–S4C). We first identify clusters of co-expressed amplified/deleted nCpDs separately using whole-genome co-expression networks analysis (WGCNA)³⁰ and their association with phenotypes of interest (see STAR Methods). The scores are merged into a single phenotype score (PhenoClus_c: sum of correlation coefficients of anti-tumor phenotypes minus the correlation coefficients of pro-tumor phenotypes) for every cluster “C.” We next identify TFs that preferentially activate or suppress

TF. These regulatory changes could indicate that reduced/maintained expression of the gene, despite an increase in CN, is a result of epigenetic silencing or inactivation of transcriptional activators. We next quantified the fraction of nCpD genes in cancer where uncoupling could be explained by changes in promoter methylation levels, TFs, and miRNA expression, or a combination of these regulatory factors (Figure 4B). On average, across cancers, we were able to identify at least one regulatory factor that could mediate uncoupling for 85.2% of nCpD genes, with

these clusters by defining a regulatory score for each TF cluster pair ($R_{score(TF,C)}$; see STAR Methods). The score weights up TFs that are strong regulators of genes in cluster and regulates them in the same direction. It also weights up TFs likely to mediate uncoupling of these genes and strongly regulates genes that are more central in PPI (protein-protein interaction) networks. Finally, we model the phenotypic impact of perturbing identified TFs using an insulated heat diffusion model.³¹ This allows us to identify and exclude TFs that, when targeted, can result in unwanted

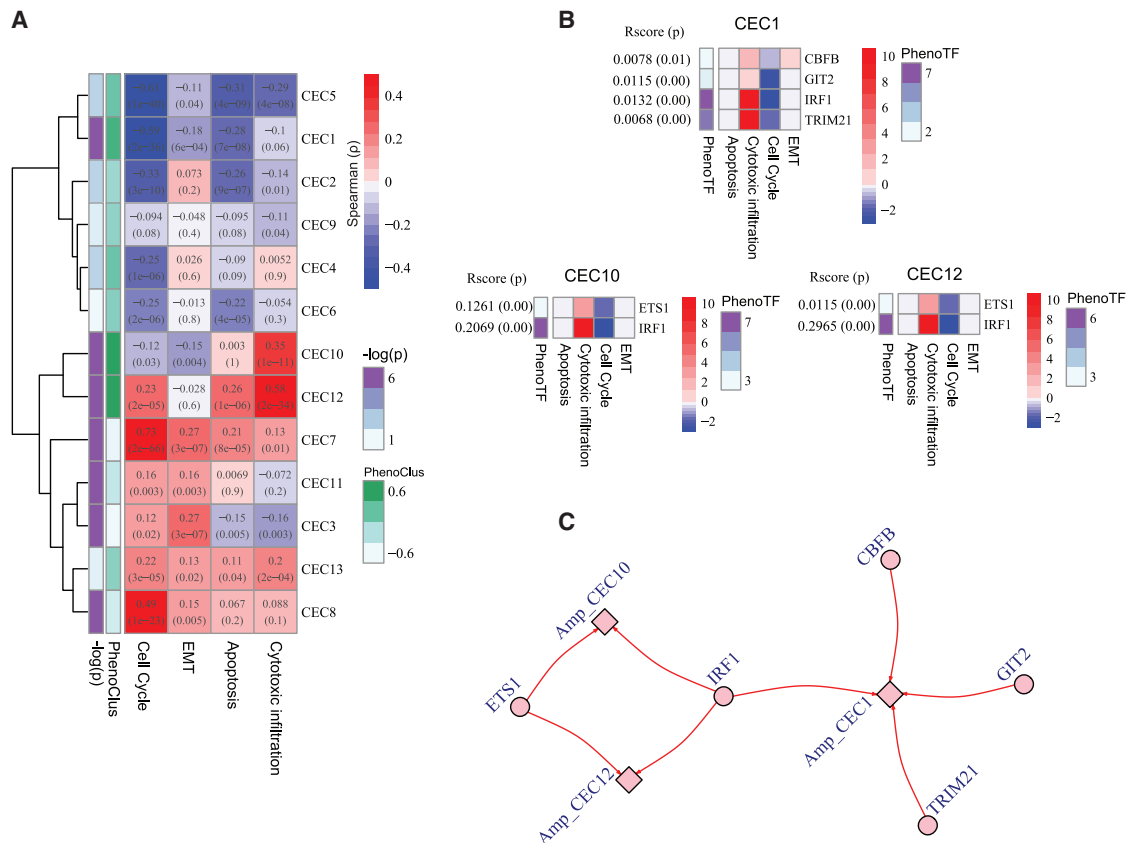


Figure 5. Identifying clusters of co-expressed nCpD genes and TFs that can be perturbed to reestablish their gene CN and expression coupling in LUAD

(A) Heatmap showing phenotypic association of clusters of co-expressed amplified nCpDs. Text in each cell shows the ρ and p value between the eigen gene expression of the cluster with sample level phenotypic measures. Clusters of interest are identified at $p < 0.05$ and PhenoClus > 0 (see STAR Methods). Three clusters, CEC1, CEC10, and CEC12, satisfy these criteria.

(B) Heatmaps of phenotypic impact of perturbing TF that activate clusters of interest ($Rscore > 0$) inferred using insulated heat diffusion. Red indicates increase in the phenotype, and blue indicates reduction in the phenotype.

(C) A network representation of phenotypically interesting CECs (from A) and TFs that regulate them (from B). Red arrow, activator; blue arrow, suppressor. AMP, a cluster of amplified nCpDs; color of the cluster indicates its PhenoClus score (red > 0 and blue < 0).

phenotypic consequences. Briefly, for each phenotype-TF pair, a weighted sum is computed across all genes that are significantly associated with the phenotype ($q < 0.01$; see STAR Methods). Individual phenotype scores were merged into a single phenotype score for a TF ($PhenoTF_{tf}$) capturing the net pro- or anti-tumor phenotypic effect of its perturbation. The score prioritizes TFs that strongly regulate phenotype-associated genes in a direction that elicits a consistent phenotypic change.

The analytical framework first identifies clusters of amplified/deleted nCpD genes that are phenotypically disadvantageous/advantageous to tumor fitness, respectively. In the context of amplification, the goal is to re-express the cluster of nCpD genes by either overexpressing an activating TF or knocking out a repressive TF, with the trends being reversed in the context of deleted nCpD genes. Perturbing a TF would, however, affect various other genes that are directly and indirectly regulated by the TF in addition to the intended nCpD genes, which could result in unwanted phenotypic consequences. We therefore modeled the flow of such perturbations in the regulatory network

to identify all genes that might be directly or indirectly affected and used them to define a net phenotypic impact the perturbation might have. Ideally TFs that are targeted for OE should promote anti-cancer phenotypes (apoptosis and immune infiltration), while not resulting in increased cell cycle or metastasis. These phenotypic trends are reversed in the case of KO candidates (see STAR Methods; Figure S4D).

We applied this analytical technique to 6 cancer types in TCGA (BRCA, HNSC, LGG, LUAD, LUSC, and SKCM) and identified 21 TFs as putative targets (Figure S4E; Table S2).

For instance, in LUAD we identified 3 clusters of amplified nCpD genes—CEC (co-expression cluster) 1 (suppression of cell cycle and EMT), CEC12 (increase in apoptosis and cytotoxic infiltration), and CEC10 (increased cytotoxic infiltration and decreased EMT and cell cycle)—that were associated with anti-tumor phenotype ($PhenoClus_C > 0$ and $p \leq 0.05$; Figure 5A). These clusters of genes are regulated by 5 TFs (Figures 5B and 5C). CEC10 and CEC12, which show strong association with increased cyto-infiltration (Figure 5A), show enrichment

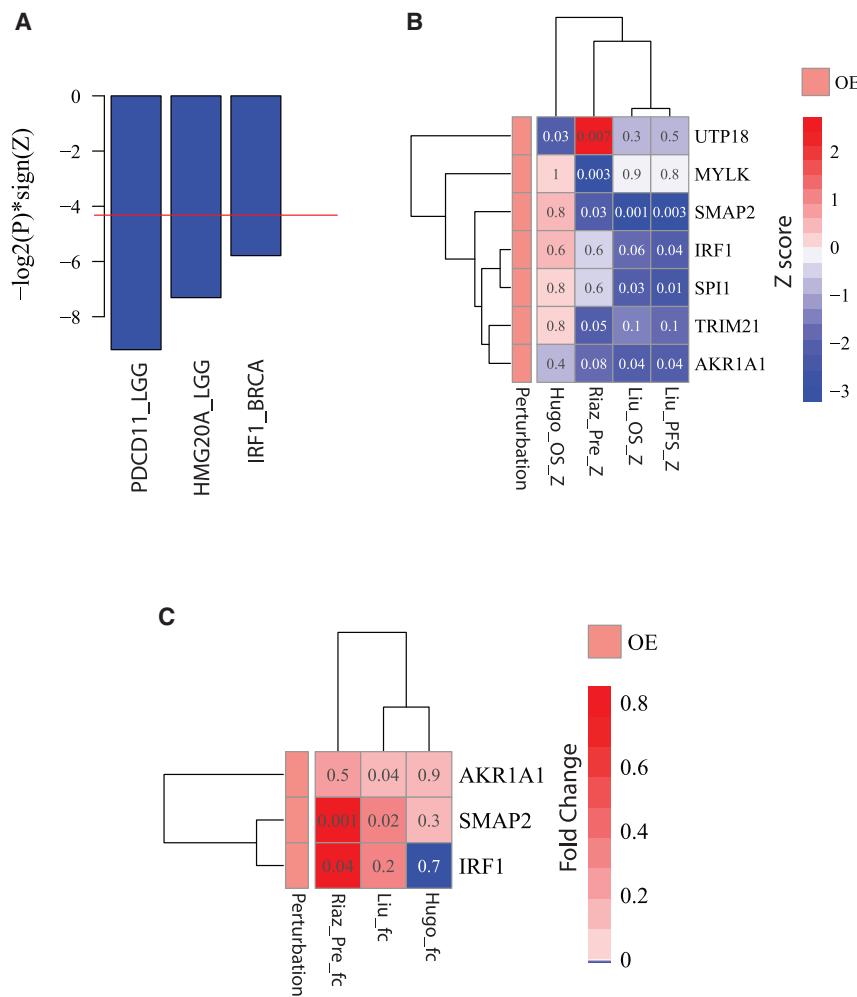


Figure 6. In silico validation of targets identified

(A) Barplot of $-\log_2(P) * \text{sign}(Z)$ of validated targets, where P and Z are Cox regression p value and Z score, respectively, quantifying association of gene expression with survival in indicated cancer. (B) Heatmap of survival Z score (Cox regression) quantifying the association of the gene's expression with progression-free survival (PFS) and overall survival (OS) in three immunotherapy cohorts. Text in each cell is the p value. (C) Heatmap of log₂-fold change in expression in responders versus non-responders in same cohorts as (B). Text in each cell is the p value. Note: for all figures, significance is called at p value < 0.05.

IRF1 and TRIM21 are TFs identified as targets when OE might reestablish expression-CN coupling of nCpD gene CECs in multiple cancers (Figure S4E). IRF1 is generally associated with activation of CECs associated with increased cytotoxic infiltration (Figure 5, LUAD; Figure S4F, HNSC, LUSC, and BRCA), while TRIM21 activates CECs associated with increased cytotoxic infiltration and suppression of EMT (Figure 5, LUAD; Figure S4F, HNSC). TRIM21 and IRF1 are positive regulators of immune pathways and strongly activate genes in the antigen presentation pathway (Figures S5A, S5C, and S5E). Upregulation of IRF1 or TRIM21 is thus likely to activate expression of these pro-cytotoxic clusters and reestablish expression-CN coupling, resulting in improved anti-tumor immunity. Consistent with these observations, expression of

($q < 0.05$) for genes associated with interferon (IFN) signaling, interleukin signaling, T cell receptor (TCR) signaling, and other pathways associated with immune response (Figures S4G and S4H). Interestingly, CEC10, which is associated with decreased cell-cycle progression, is also enriched for negative regulators of the *PI3K/AKT* signaling network, which is a regulator of cell-cycle progression and oncogenic transformation (Figure S4G).³² Our model proposes that CEC10 and CEC12 are activated by two TFs—IRF1 and ETS1. Furthermore, our perturbation model suggests that OE of *IRF1* and *ETS1* is likely to have an anti-tumor effect due to increased cytotoxic immune infiltration (Figure 5B). GSEA of the regulatory footprint of IRF1 and ETS1 across cancers suggests they are positive regulators of immune-associated pathways (Figures S5A and S5B). CEC1, associated with suppression of cell cycle, EMT, and immune infiltration, is activated by 4 TFs including IRF1 and TRIM21. Regulatory footprint of TRIM21 like IRF1 suggests it is a positive regulator of immune-related pathways (Figure S5C). Interestingly, TRIM21 also suppresses EMT-related genes in LUAD, consistent with the negative association of CEC1 with EMT and the negative correlation of *TRIM21* expression with EMT in LUAD ($p = 0.013$; Figure S5H).

IRF1 and *TRIM21* was associated with increased expression of the cytolytic markers *GZMA* and *PRF1* (Figure S5F). The expression of *IRF1* and *TRIM21* was also associated with increased infiltration of CD8 T cells, natural killer (NK) cells, and M1 macrophages, while resulting in exclusion of anti-inflammatory M2 macrophages and inactive NK and CD4 T cells (Figure S5G). These data suggest that OE of *TRIM21* and *IRF1* in aneuploid tumors can reestablish expression CN coupling of immune-associated genes, inducing immune signaling and cytotoxic immune infiltration into the tumor.

Another interesting target regulating CECs of nCpD genes in LUSC is *SNAI2* (Figure S4F) where it is an activator of the amplified nCpD genes in CEC2. *SNAI2* is a EMT gene, and its increased expression induces EMT,³³ consistent with the regulatory footprint of *SNAI2* and positive correlation of its expression with EMT scores (Figures S5D and S5I). Interestingly, in our analysis, OE of *SNAI2* in LUSC is predicted to suppress the cell cycle (Figure S4F); in agreement, we find *SNAI2* expression is inversely correlated with proliferation in LUSC (Figure S5J).

In silico validations of predicted targets

To validate the putative targets, we make use of three complementary *in silico* approaches (Figure 6): (1) association with

survival outcome in corresponding TCGA tumor type, (2) association with overall survival (OS) and progression-free survival (PFS) in immunotherapy cohorts,^{34–36} and (3) differential expression in responders relative to non-responders in immunotherapy cohorts.^{34–36} Positive examples are, in the context of amplified nCpDs, OE candidates associated with improved survival or overexpressed in responders, while these trends are reversed for KO candidates (Figure S4D). These associations are expected to be reversed for targets identified in the context of deleted nCpDs (Figure S4D). Using this approach, 14.2% (3/21) of identified targets show expected survival trends in corresponding cancers (Figure 6A); 33.3% (7/21) were associated with OS/PFS in immune therapy cohorts consistent with the context they were identified in (Figure 6B); and 14.2% (3/21) of targets regulating amplified nCpD gene clusters were overexpressed in responders (Figure 6C). In total, 42.8% (9/21) of the putative targets satisfied our validation criteria. We further used a randomization-based approach to compute an empirical p values for our validation rate. We found that the validation rate (42.8%) is unlikely to be obtained by chance ($p = 0.045$; see STAR Methods and Figure S6A).

Most of the targets were identified in the context of anti-cancer immunity (Figures 6B and 6C). *IRF1*, an OE target identified in LUAD, LUSC, BRCA, and HNSC (Figure S4E), plays a prominent role in anti-cancer immunity by promoting cytolytic immune cell infiltration (Figures S5E–S5G). *IRF1* expression is also associated with improved prognosis in BRCA (Figure 6A) and also predicts improved response to anti-PD1 immunotherapy (Figures 6B and 6C). *IRF1* is downstream of $IFN\gamma$, a cytokine critical for innate and adaptive immunity, and mediates its signaling.³⁷ Interestingly, *IRF1* is frequently deleted in patients that do not respond to CTLA4 inhibition.³⁸ Furthermore, *IRF1* is a strong transcriptional activator of the MHC class I antigen presentation complex (Figure S5E), which plays a critical role in anti-tumor immunity by presenting neo-epitopes and stimulating T cell-mediated tumor killing.^{39,40} *TRIM21* is another OE candidate (in LUAD and HNSC). *TRIM21* also predicted response to anti-CTLA4 immunotherapy (Figure 6B). Although the role of *TRIM21* in anti-tumor immunity isn't clear, it is known to play a role in anti-viral immune responses.⁴¹ Furthermore, *TRIM21* has been shown to be associated with improved survival in breast cancers⁴² and to suppress metastasis in breast cancers by increased ubiquitination and proteosomal degradation of Snail.⁴³ *SPI1* identified in LUSC regulates a pro-inflammatory gene cluster (Figure S4F) and has been found to regulate immune-associated networks.⁴⁴

DISCUSSION

Aneuploidy is a common hallmark of cancer and is associated with advanced stage and poor prognosis.¹ In contrast to malignant contexts, aneuploidy is poorly tolerated by normal cells.⁴⁵ How cancer cells tolerate and thrive in an aneuploidy context and whether aneuploidy represents a therapeutic opportunity are largely unclear due to both analytical and experimental challenges.¹ To help fill this knowledge gap, we performed comprehensive, multiomics analysis of 11 cancers in TCGA. We found extensive UECN in genes whose CNs were affected by aneu-

ploidy. Functional analysis of nCpD genes suggests CN changes in these genes are tumor toxic (Figure 1). In individual tumors, we find UECN increases with aneuploidy and suppresses pathways detrimental to tumor fitness (Figure 2). Regulatory analysis of nCpD genes suggests UECN is mediated by complex regulatory changes (Figure 4). Consequently, targeting these regulators can potentially reestablish CN expression coupling of nCpD genes, resulting in a flux of signaling detrimental to tumor fitness (Figure S4D). We therefore identified TFs that regulate clusters of co-expressed nCpDs that could potentially be targeted to negate UECN (Figures S4A–S4C).

Using elastic nets, we modeled gene expressed as a function of various regulators (Figure 3). This approach allows us to quantify the gene modulation direction and magnitude. Elastic nets use a mixture of L1 and L2 norm to fit the model; it sets non-significant terms to 0, while capturing subtle regulatory effects. This allowed us to build weighted and directed regulatory networks vital for perturbation modeling (Figure S4C). We selected TFs that can reverse UECN based on how they regulate clusters of co-expressed nCpD. However, to account for off-target effect, we modeled the net effect of perturbing a TF and its likely phenotypic consequences using heat diffusion. We focused on sets of co-expressed nCpDs rather than single genes, as this facilitates the identification of TFs that regulate a set of functionally similar genes that are likely to have similar expression patterns in the tumor. Thus, a single TF can be perturbed to cover numerous contexts, in which different sets of nCpD genes in a CEC may have their expression uncoupled from CN.

Aneuploidy has been linked with decreases in tumor infiltrating lymphocytes, resulting in immunologically cold tumors and poor responses to immune checkpoint therapy.²⁴ These observations seem to contradict several studies that inducing aneuploidy in cell lines elicits anti-tumor immunity *in vivo*.^{46–48} Interestingly, aneuploidy CT26 murine cancer cells passaged under immune selection developed an immune evasive phenotype by epigenetic silencing of genes associated with antigen presentation.⁴⁹ Interestingly, we find that aneuploidy can be paradoxically associated with the amplification of genes in multiple immune-associated pathways (Figure 1D). They are, however, frequently nCpD genes; hence, their tumor-toxic effects are negated. We also find that these pathways are suppressed in samples with increased D_{UC} , suggesting that UECN allows tumors to bypass aneuploidy-induced immune response and attain an immune-evasive phenotype, while maintaining oncogenic signaling (Figure 2D). Interestingly, in addition to epigenetic silencing of genes in the antigen presentation pathway in aneuploid tumors, as shown by Tripathi et al.,⁴⁹ we find TFs such as IRF1 and TRIM21 regulate clusters of co-expressed genes that are strongly associated with immune infiltration and cytotoxicity (Figures 5 and S4G and S4H). Perturbation modeling of IRF1 and TRIM21 suggests their OE could turn immunologically cold aneuploid tumors into inflammatory tumors and improve response to immunotherapy. Although TFs have traditionally been difficult to target, some small-molecule inhibitors have been developed to target them.^{50,51} Development of new approaches to target TFs allows for effectively targeting TFs.⁵² These, together with our analysis of regulatory changes mediating UECN, suggest that targeting appropriate TFs may be a

feasible approach to treat chromosomally unstable (high D_{UC}) tumors, especially in the context of cancer therapies that modulate the immune system.

For *in silico* validation of targets, we looked at (1) association with survival in corresponding cancer type and (2) differential expression and survival impact in the context of checkpoint therapy. We make use of 3 checkpoint therapy datasets^{34–36} covering both PD1 and CTLA4 inhibition. While KO screens,^{7,8} which are routinely applied for such target gene validation can identify genes associated with proliferation and viability through siRNA or guide RNA (gRNA) dropout, as they are currently designed they do not characterize EMT or anti-tumor immunity, due to the lack of a micro-environment. Furthermore, KO screens would not identify OE candidates. By utilizing clinical and immunotherapy data, we circumvent some of these limitations and validate 42.8% of predicted targets. Intriguingly, only 3 of 21 targets showed survival advantage in TCGA in contrast to a third of the targets showing survival advantages in immunotherapy datasets. This indicates that the survival advantage we expect to see is more likely to be observed in patients who have received immune-modulating agents. However, given the small sample size of immunotherapy datasets, it's likely our validation analysis is underpowered. Detailed experimental examination of these targets in appropriate contexts is required to capture more true positive targets.

In addition to immune-associated genes, amplified nCpD genes are also enriched for genes associated with apoptosis. Intriguingly, cell lines with low activity of the apoptotic pathway and extensive amplifications in apoptotic genes show nominally greater sensitivity to azacitidine ($p = 0.0812$), a DNA demethylation agent (see Figures S6B and S6C; Table S3). This is consistent with our observation that amplified nCpD genes are epigenetically silenced (Figure 4). This association was not observed in cell lines that don't have extensive amplifications in apoptotic genes (Figure S6D; Table S3). These observations further indicate that reversing UECN could have therapeutic potential in cancers.

In conclusion, aneuploidy can result in CN changes that are toxic to the tumor. The tumor therefore uncouples the CN of these genes from their expression to maintain tumor fitness. UECN also seems to be a prominent mechanism by which aneuploid tumors evade the host immune system. These toxic CN compensations are mediated by complex changes in epigenetic and transcriptional regulation that can be targeted to the detriment of the tumor, presenting a novel approach by which aneuploid tumors can be targeted.

Limitations of study

Malignant cells in tumors are extensively heterogeneous with distinct genomic and transcriptomic profiles. Accounting for this clonal diversity and its impact on UECN is extremely difficult from bulk data. To study rare sub-clonal UECN events, which may be especially interesting in cases of acquired resistance to therapy, co-profiling of RNA and DNA at the level of single cells is required.

The precise mechanism of development of UECN also remains unknown. We hypothesize that certain selective pressures might cause CN changes in specific genes to be tumor toxic,

thus favoring tumor cells that neutralize these CN changes. This could consequently select for regulatory rewiring, silencing such tumor-toxic CN changes resulting in UECN. Identifying and studying these selective pressures will require development of experimental models where such evolution can be tracked over time. Further development of statistical tools that can call UECN events in single samples, coupled with longitudinal data starting from pre-tumorous lesions to full developed malignancies, might provide an alternative approach to study selective pressures that result in UECN.

STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability
 - Data and code availability
- METHOD DETAILS
 - Datasets
 - Associating gene expression with phenotype
 - Population level analysis of uncoupling and pathway analysis
 - Localizing nCpD genes to frequent CNV events
 - Analysis of CCLE data
 - Analysis controlling for stemness associate genes and genes with low expression in normal tissue
 - Analysis of tumor subtypes
 - Sample level analysis of uncoupling
 - Degree of uncoupling survival analysis
 - Building regulatory models
 - Identifying regulators that mediate uncoupling of expression from CN for individual genes
 - Co-expression modules of uncoupled genes
 - Predicting phenotypic consequences of gene perturbation using heat diffusion
 - Target selection
 - Immuno-therapy datasets analysis
 - Randomized testing to obtain empirical p value for fraction of TFs validated
 - Analysis of sensitivity to azacitidine
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.xcrm.2021.100349>.

ACKNOWLEDGMENTS

This work was supported by the National Cancer Institute's Office of Cancer Genomics Cancer Target Discovery and Development (CTD²) initiative. This work was supported in part by the NIH (U01 CA247760) and the CPRIT (RP180248) to K.C., the NCI Cancer Center Support Grant (P30 CA016672), and NCI U01 CA217842 to G.B.M. The results published here are based in part upon data generated by Cancer Target Discovery and Development

(CTD²) Network (<https://ocg.cancer.gov/programs/ctd2/data-portal>) established by the National Cancer Institute's Office of Cancer Genomics.

AUTHOR CONTRIBUTIONS

V.M. and K.C. designed the study. V.M. and F.W. performed the analysis. V.M., K.C., F.W., and G.B.M. helped analyze the results. All authors helped to write the manuscript. All authors have read and approved this paper.

DECLARATION OF INTERESTS

G.B.M. is a SAB/Consultant at AstraZeneca, Chrysalis Biotechnology, GSK, ImmunoMET, Ionis, Lilly, PDX Pharmaceuticals, Signalchem Lifesciences, Symphogen, Tarveda, Turbine, Zentalis Pharmaceuticals; has stock/options/financial from Catena Pharmaceuticals, ImmunoMet, SignalChem, Tarveda; has licensed technology with HRD assay to Myriad Genetics, DSP patents with Nanostring; has sponsored research with Nanostring Center of Excellence, Ionis (provision of tool compounds); and has clinical trials support (funding or in kind) from AstraZeneca, Genentech, GSK, Lilly.

Received: September 28, 2020

Revised: February 11, 2021

Accepted: June 22, 2021

Published: July 20, 2021

REFERENCES

- Ben-David, U., and Amon, A. (2020). Context is everything: aneuploidy in cancer. *Nat. Rev. Genet.* *21*, 44–62.
- Taylor, A.M., Shih, J., Ha, G., Gao, G.F., Zhang, X., Berger, A.C., Schumacher, S.E., Wang, C., Hu, H., Liu, J., et al. (2018). Genomic and Functional Approaches to Understanding Cancer Aneuploidy. *Cancer Cell* *33*, 676–689.e3.
- Zack, T.I., Schumacher, S.E., Carter, S.L., Cherniack, A.D., Saksena, G., Tabak, B., Lawrence, M.S., Zhsng, C.Z., Wala, J., Mermel, C.H., et al. (2013). Pan-cancer patterns of somatic copy number alteration. *Nat. Genet.* *45*, 1134–1140.
- Beroukhi, R., Mermel, C.H., Porter, D., Wei, G., Raychaudhuri, S., Donovan, J., Barretina, J., Boehm, J.S., Dobson, J., Urashima, M., et al. (2010). The landscape of somatic copy-number alteration across human cancers. *Nature* *463*, 899–905.
- Mohanty, V., Akmedova, O., and Komurov, K. (2016). Selective DNA methylation in cancers controls collateral damage induced by large structural variations. *Oncotarget* *8*, 71385–71392.
- Wang, F., Zhang, S., Kim, T.B., Lin, Y.Y., Iqbal, R., Wang, Z., Mohanty, V., Sircar, K., Karam, J.A., Wendl, M.C., et al. (2019). Integrated transcriptomic-genomic tool Texomer profiles cancer tissues. *Nat. Methods* *16*, 401–404.
- Tsherniak, A., Vazquez, F., Montgomery, P.G., Weir, B.A., Kryukov, G., Cowley, G.S., Gill, S., Harrington, W.F., Pantel, S., Krill-Burger, J.M., et al. (2017). Defining a Cancer Dependency Map. *Cell* *170*, 564–576.e16.
- Ghandi, M., Huang, F.W., Jané-Valbuena, J., Kryukov, G.V., Lo, C.C., McDonald, E.R., 3rd, Barretina, J., Gelfand, E.T., Bielski, C.M., Li, H., et al. (2019). Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature* *569*, 503–508.
- Hart, T., Chandrashekar, M., Aregger, M., Steinhart, Z., Brown, K.R., MacLeod, G., Mis, M., Zimmermann, M., Fradet-Turcotte, A., Sun, S., et al. (2015). High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities. *Cell* *163*, 1515–1526.
- Hsu, T.Y., Simon, L.M., Neill, N.J., Marcotte, R., Sayad, A., Bland, C.S., Echeverria, G.V., Sun, T., Kurley, S.J., Tyagi, S., et al. (2015). The spliceosome is a therapeutic vulnerability in MYC-driven cancer. *Nature* *525*, 384–388.
- Marcotte, R., Sayad, A., Brown, K.R., Sanchez-Garcia, F., Reimand, J., Haider, M., Virtanen, C., Bradner, J.E., Bader, G.D., Mills, G.B., et al. (2016). Functional Genomic Landscape of Human Breast Cancer Drivers, Vulnerabilities, and Resistance. *Cell* *164*, 293–309.
- Bhagwat, A.S., and Vakoc, C.R. (2015). Targeting Transcription Factors in Cancer. *Trends Cancer* *1*, 53–65.
- Alvarez, M.J., Shen, Y., Giorgi, F.M., Lachmann, A., Ding, B.B., Ye, B.H., and Califano, A. (2016). Functional characterization of somatic mutations in cancer using network-based inference of protein activity. *Nat. Genet.* *48*, 838–847.
- Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A.A., Kim, S., Wilson, C.J., Lehár, J., Kryukov, G.V., Sonkin, D., et al. (2012). The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* *483*, 603–607.
- Ling, A., Löfgren-Burström, A., Larsson, P., Li, X., Wikberg, M.L., Öberg, Å., Stenling, R., Edin, S., and Palmqvist, R. (2017). TAP1 down-regulation elicits immune escape and poor prognosis in colorectal cancer. *Oncolmunology* *6*, e1356143.
- Ott, C.A., Linck, L., Kremmer, E., Meister, G., and Bosserhoff, A.K. (2016). Induction of exportin-5 expression during melanoma development supports the cellular behavior of human malignant melanoma cells. *Oncotarget* *7*, 62292–62304.
- Leaderer, D., Hoffman, A.E., Zheng, T., Fu, A., Weidhaas, J., Paranjape, T., and Zhu, Y. (2011). Genetic and epigenetic association studies suggest a role of microRNA biogenesis gene exportin-5 (XPO5) in breast tumorigenesis. *Int. J. Mol. Epidemiol. Genet.* *2*, 9–18.
- Shigeyasu, K., Okugawa, Y., Toden, S., Boland, C.R., and Goel, A. (2017). Exportin-5 Functions as an Oncogene and a Potential Therapeutic Target in Colorectal Cancer. *Clin. Cancer Res.* *23*, 1312–1322.
- Gabay, M., Li, Y., and Felsner, D.W. (2014). MYC activation is a hallmark of cancer initiation and maintenance. *Cold Spring Harb. Perspect. Med.* *4*, a014241.
- Urra, H., Dufey, E., Avril, T., Chevet, E., and Hetz, C. (2016). Endoplasmic Reticulum Stress and the Hallmarks of Cancer. *Trends Cancer* *2*, 252–262.
- Ashton, T.M., McKenna, W.G., Kunz-Schughart, L.A., and Higgins, G.S. (2018). Oxidative Phosphorylation as an Emerging Target in Cancer Therapy. *Clin. Cancer Res.* *24*, 2482–2490.
- Kalimutho, M., Nones, K., Srihari, S., Duijff, P.H.G., Waddell, N., and Khanna, K.K. (2019). Patterns of Genomic Instability in Breast Cancer. *Trends Pharmacol. Sci.* *40*, 198–211.
- Cancer Genome Atlas Network (2015). Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature* *517*, 576–582.
- Davoli, T., Uno, H., Wooten, E.C., and Elledge, S.J. (2017). Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. *Science* *355*, eaaf8399.
- Sampson, J.H., Gunn, M.D., Fecci, P.E., and Ashley, D.M. (2020). Brain immunology and immunotherapy in brain tumours. *Nat. Rev. Cancer* *20*, 12–25.
- Mossmann, D., Park, S., and Hall, M.N. (2018). mTOR signalling and cellular metabolism are mutual determinants in cancer. *Nat. Rev. Cancer* *18*, 744–757.
- Jones, P.A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.* *13*, 484–492.
- Kulis, M., and Esteller, M. (2010). DNA methylation and cancer. *Adv. Genet.* *70*, 27–56.
- Croce, C.M. (2009). Causes and consequences of microRNA dysregulation in cancer. *Nat. Rev. Genet.* *10*, 704–714.
- Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* *9*, 559.
- Leiserson, M.D., Vandin, F., Wu, H.T., Dobson, J.R., Eldridge, J.V., Thomas, J.L., Papoutsaki, A., Kim, Y., Niu, B., McLellan, M., et al. (2015). Pan-cancer network analysis identifies combinations of rare

- somatic mutations across pathways and protein complexes. *Nat. Genet.* **47**, 106–114.
32. Chang, F., Lee, J.T., Navolanic, P.M., Steelman, L.S., Shelton, J.G., Blalock, W.L., Franklin, R.A., and McCubrey, J.A. (2003). Involvement of PI3K/Akt pathway in cell cycle progression, apoptosis, and neoplastic transformation: a target for cancer chemotherapy. *Leukemia* **17**, 590–603.
 33. Baum, B., Settleman, J., and Quinlan, M.P. (2008). Transitions between epithelial and mesenchymal states in development and disease. *Semin. Cell Dev. Biol.* **19**, 294–308.
 34. Liu, D., Schilling, B., Liu, D., Sucker, A., Livingstone, E., Jerby-Aron, L., Zimmer, L., Gutzmer, R., Satzger, I., Loquai, C., et al. (2019). Integrative molecular and clinical modeling of clinical outcomes to PD1 blockade in patients with metastatic melanoma. *Nat. Med.* **25**, 1916–1927.
 35. Hugo, W., Zaretsky, J.M., Sun, L., Song, C., Moreno, B.H., Hu-Lieskovan, S., Berent-Maoz, B., Pang, J., Chmielowski, B., Cherry, G., et al. (2016). Genomic and Transcriptomic Features of Response to Anti-PD-1 Therapy in Metastatic Melanoma. *Cell* **165**, 35–44.
 36. Riaz, N., Havel, J.J., Makarov, V., Desrichard, A., Urba, W.J., Sims, J.S., Hodi, F.S., Martín-Algarra, S., Mandal, R., Sharfman, W.H., et al. (2017). Tumor and Microenvironment Evolution during Immunotherapy with Nivolumab. *Cell* **171**, 934–949.e16.
 37. Dunn, G.P., Koebel, C.M., and Schreiber, R.D. (2006). Interferons, immunity and cancer immunoediting. *Nat. Rev. Immunol.* **6**, 836–848.
 38. Gao, J., Shi, L.Z., Zhao, H., Chen, J., Xiong, L., He, Q., Chen, T., Roszik, J., Bernatchez, C., Woodman, S.E., et al. (2016). Loss of IFN- γ Pathway Genes in Tumor Cells as a Mechanism of Resistance to Anti-CTLA-4 Therapy. *Cell* **167**, 397–404.e9.
 39. Hobart, M., Ramassar, V., Goes, N., Urmsen, J., and Halloran, P.F. (1996). The induction of class I and II major histocompatibility complex by allogeneic stimulation is dependent on the transcription factor interferon regulatory factor 1 (IRF-1): observations in IRF-1 knockout mice. *Transplantation* **62**, 1895–1901.
 40. Jongtsma, M.L.M., Guarda, G., and Spaapen, R.M. (2019). The regulatory network behind MHC class I expression. *Mol. Immunol.* **113**, 16–21.
 41. Foss, S., Bottermann, M., Jonsson, A., Sandlie, I., James, L.C., and Andersen, J.T. (2019). TRIM21-From Intracellular Immunity to Therapy. *Front. Immunol.* **10**, 2049.
 42. Zhou, W., Zhang, Y., Zhong, C., Hu, J., Hu, H., Zhou, D., and Cao, M. (2018). Decreased expression of TRIM21 indicates unfavorable outcome and promotes cell growth in breast cancer. *Cancer Manag. Res.* **10**, 3687–3696.
 43. Jin, Y., Zhang, Y., Li, B., Zhang, J., Dong, Z., Hu, X., and Wan, Y. (2019). TRIM21 mediates ubiquitination of Snail and modulates epithelial to mesenchymal transition in breast cancer cells. *Int. J. Biol. Macromol.* **124**, 846–853.
 44. Thorsson, V., Gibbs, D.L., Brown, S.D., Wolf, D., Bortone, D.S., Ou Yang, T.H., Porta-Pardo, E., Gao, G.F., Plaisier, C.L., Eddy, J.A., et al. (2018). The Immune Landscape of Cancer. *Immunity* **48**, 812–830.e14.
 45. Sheltzer, J.M., and Amon, A. (2011). The aneuploidy paradox: costs and benefits of an incorrect karyotype. *Trends Genet.* **27**, 446–453.
 46. Senovilla, L., Vitale, I., Martins, I., Tailler, M., Paillet, C., Michaud, M., Galluzzi, L., Adjemian, S., Kepp, O., Niso-Santano, M., et al. (2012). An immunosurveillance mechanism controls cancer cell ploidy. *Science* **337**, 1678–1684.
 47. Santaguida, S., Richardson, A., Iyer, D.R., M'Saad, O., Zasadil, L., Knouse, K.A., Wong, Y.L., Rhind, N., Desai, A., and Amon, A. (2017). Chromosome Mis-segregation Generates Cell-Cycle-Arrested Cells with Complex Karyotypes that Are Eliminated by the Immune System. *Dev. Cell* **41**, 638–651.e5.
 48. Mackenzie, K.J., Carroll, P., Martin, C.A., Murina, O., Fluteau, A., Simpson, D.J., Olova, N., Sutcliffe, H., Rainger, J.K., Leitch, A., et al. (2017). cGAS surveillance of micronuclei links genome instability to innate immunity. *Nature* **548**, 461–465.
 49. Tripathi, R., Modur, V., Senovilla, L., Kroemer, G., and Komurov, K. (2019). Suppression of tumor antigen presentation during aneuploid tumor evolution contributes to immune evasion. *Oncolmmunology* **8**, 1657374.
 50. Koehler, A.N. (2010). A complex task? Direct modulation of transcription factors with small molecules. *Curr. Opin. Chem. Biol.* **14**, 331–340.
 51. Arkin, M.R., Tang, Y., and Wells, J.A. (2014). Small-molecule inhibitors of protein-protein interactions: progressing toward the reality. *Chem. Biol.* **21**, 1102–1114.
 52. Bushweller, J.H. (2019). Targeting transcription factors in cancer - from undruggable to reality. *Nat. Rev. Cancer* **19**, 611–624.
 53. Aran, D., Sirota, M., and Butte, A.J. (2015). Systematic pan-cancer analysis of tumour purity. *Nat. Commun.* **6**, 8971.
 54. Newman, A.M., Liu, C.L., Green, M.R., Gentles, A.J., Feng, W., Xu, Y., Hoang, C.D., Diehn, M., and Alizadeh, A.A. (2015). Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **12**, 453–457.
 55. Abugessaisa, I., Noguchi, S., Carninci, P., and Kasukawa, T. (2017). The FANTOM5 Computation Ecosystem: Genomic Information Hub for Promoters and Active Enhancers. *Methods Mol. Biol.* **1611**, 199–217.
 56. Chen, H., Li, C., Peng, X., Zhou, Z., Weinstein, J.N., and Cancer Genome Atlas Research Network; and Liang, H. (2018). A Pan-Cancer Analysis of Enhancer Expression in Nearly 9000 Patient Samples. *Cell* **173**, 386–399.e12.
 57. Lewis, B.P., Burge, C.B., and Bartel, D.P. (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**, 15–20.
 58. GTEx Consortium (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330.
 59. Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* **1**, 417–425.
 60. Luo, W., Friedman, M.S., Shedden, K., Hankenson, K.D., and Woolf, P.J. (2009). GAGE: generally applicable gene set enrichment for pathway analysis. *BMC Bioinformatics* **10**, 161.
 61. Rees, M.G., Seashore-Ludlow, B., Cheah, J.H., Adams, D.J., Price, E.V., Gill, S., Javaid, S., Coletti, M.E., Jones, V.L., Bodycombe, N.E., et al. (2016). Correlating chemical sensitivity and basal gene expression reveals mechanism of action. *Nat. Chem. Biol.* **12**, 109–116.
 62. Akbani, R., Ng, P.K., Werner, H.M., Shahmoradgol, M., Zhang, F., Ju, Z., Liu, W., Yang, J.Y., Yoshihara, K., Li, J., et al. (2014). A pan-cancer proteomic perspective on The Cancer Genome Atlas. *Nat. Commun.* **5**, 3887.
 63. Forbes, S.A., Beare, D., Boutselakis, H., Bamford, S., Bindal, N., Tate, J., Cole, C.G., Ward, S., Dawson, E., Ponting, L., et al. (2017). COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res.* **45** (D1), D777–D783.
 64. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550.
 65. Weirauch, M.T., Yang, A., Albu, M., Cote, A.G., Montenegro-Montero, A., Drewe, P., Najafabadi, H.S., Lambert, S.A., Mann, I., Cook, K., et al. (2014). Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* **158**, 1431–1443.
 66. Tan, G., and Lenhard, B. (2016). TFBSTools: an R/bioconductor package for transcription factor binding site analysis. *Bioinformatics* **32**, 1555–1556.
 67. Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* **14**, 7.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
gage; RRID:SCR_017067	Bioconductor	https://bioconductor.org/packages/release/bioc/html/gage.html
ProliferativeIndex	R library	https://cran.r-project.org/web/packages/ProliferativeIndex/index.html
survival; RRID:SCR_021137	R library	https://cran.r-project.org/web/packages/survival/index.html
FindOverlap (GenomicRanges); RRID:SCR_000025	Bioconductor	https://bioconductor.org/packages/release/bioc/html/GenomicRanges.html
DEseq2; RRID:SCR_015687	Bioconductor	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
TFBSTools	Bioconductor	https://bioconductor.org/packages/release/bioc/html/TFBSTools.html
caret; RRID:SCR_021138	R library	https://cran.r-project.org/web/packages/caret/index.html
glmnet; RRID:SCR_015505	R library	https://cran.r-project.org/web/packages/glmnet/index.html
WGCNA; RRID:SCR_003302	R library	https://cran.r-project.org/web/packages/WGCNA/index.html
HotNet2	Leiserson et al. ³¹	https://github.com/raphael-group/hotnet2
GSVA; RRID:SCR_021058	Bioconductor	https://bioconductor.org/packages/release/bioc/html/GSVA.html

RESOURCE AVAILABILITY

Lead contact

Please direct queries about the study to the Lead Contact Dr. Ken Chen (kchen3@mdanderson.org), Associate Professor, Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA.

Materials availability

This study did not generate any new unique reagents.

Data and code availability

All data used in the study is publically available and detailed in the Method details section. No unique code was generated in this study.

METHOD DETAILS

Datasets

mRNA expression (RNaseqV2; counts and log₂ normalized counts), miRNA expression (miRNaseq; log₂ RPM), protein expression (RPPA), DNA methylation (450K) and tumor subtype data for TCGA samples were obtained from UCSC Xena (<https://xenabrowser.net/datapages/>). TCGA clinical data were obtained from Firehose (<https://gdac.broadinstitute.org/>). Frequent focal/regional and chromosome arm CNVs are obtained from Firehose. TCGA purity measurements were obtained from Aran et al.⁵³ and the combined purity estimate was used. Absolute gene copy numbers and CIBERSORT⁵⁴ quantification for TCGA samples were obtained from Thorsson et al.⁴⁴ Expression and CN data for CCLE¹⁴ cell lines were obtained from (<https://depmap.org/portal/download/>). Enhancer locations and the gene promoters they interact with were obtained from FANTOM5⁵⁵, expression of these enhancers in TCGA samples were quantified from RNaseq BAM file obtained from GDC (<https://portal.gdc.cancer.gov/>), as described by Chen et al.⁵⁶. miRNA binding data were obtained from TargetScan⁵⁷ binding site predictions (http://www.targetscan.org/cgi-bin/targetscan/data_download.vert72.cgi). mRNA expression levels for immune-therapy samples were obtained from three studies.^{34–36} Gene

expression (TPM) in normal tissue was obtained from GTEx⁵⁸ using UCSC Xena. Hallmark pathway⁵⁹ definitions were used for GSEA, which was performed using *gage*.⁶⁰ Cell line drug sensitivity data was obtained from CTRPv2.⁶¹ Samples sizes of the omics and clinical datasets are detailed in Table S4.

Associating gene expression with phenotype

Immune infiltration, apoptosis, cell cycle and EMT scores were defined for each sample based on defined markers.⁶² Two was raised to the power of the expression of the markers and the average of this value for positive markers was divided by the average value for negative markers. Proliferation rate was inferred from RNaseq using the R library *ProliferativeIndex* (<https://cran.r-project.org/web/packages/ProliferativeIndex/vignettes/ProliferativeIndexVignette.html>). Gene expression was then regressed against phenotype scores in individual tumor types and a corresponding T-value is extracted. A positive T-value indicated that the gene's expression was positively associated with the phenotype.

Relationship of gene expression with survival was identified using cox-hazard regression implemented in the R library *survival*, while controlling for age and tumor stage. A positive cox Z score indicates increased expression of the gene is associated with poor survival while a negative Z score indicates improved survival.

Population level analysis of uncoupling and pathway analysis

Analysis was performed in each individual cancer type. First, genes with low expression level were excluded (90th quantile expression < 30 normalized reads). Top 5000 variably expressed genes in each cancer were used to perform PCA (Principle component analysis). Frequently amplified and deleted genes were then identified using absolute CN. Genes with CN = 1 in 25% of samples were considered to be frequently deleted (or deleted) while genes with CN = 3,4 in 25% samples and CN > 4 in 5% of samples were defined as frequently amplified (or amplified). Expression of frequently amplified/deleted genes was then regressed against CN while controlling for tumor purity and top 20 expression PCs. Regression T-value (T_{CN}) of the CN term was used to stratify genes. $T_{CN} < 4.5$ indicates nCpD gene, while $T_{CN} \geq 4.5$ indicate CpD gene (Figure S1D). COSMIC⁶³ consensus driver genes were excluded from nCpD and CpD genes.

Pathway analysis was carried out for amplified and deleted genes separately. T_{CN} was centered on 4.5 so that nCpD genes had negative scores and CpD genes positive scores. These scores were then used to perform GSEA to identify pathways enriched in CpD and nCpD genes ($q < 0.05$). Similar analysis with deleted genes didn't consistently (in at least four cancer) identify pathways enriched in CpD and nCpD genes.

Localizing nCpD genes to frequent CNV events

FindOverlaps was used to identify nCpD genes in recurrent CNVs. An amplified nCpD was linked to a recurrent CNV if 1. It shared genomic locations with the CNV, 2. CN of the gene and CNV correlated strongly ($q < 0.1$ and Spearman correlation > 0.5) and 3. The CNV was called amplification or both by GISTIC2. The same statistical cutoff was used in the case of deleted nCpD genes, but the GISTIC2 call for the event was either deletion or both.

Analysis of CCLE data

Genes with low expression (90th percentile expression < 1TPM) were excluded from the analysis. Frequently amplified and deleted genes were identified as genes with CN > 2 in 20% and CN < 2 in 10% of cell lines respectively and COSMIC⁶³ consensus genes were filtered out. For all frequently amplified/deleted genes, their expression was regressed against CN with tumor type as a covariant. T-values for the CN term were extracted, and genes with T-value < 4.5 were considered nCpD. For the functional analysis, T-values were centered on 4.5 and used for GSEA analysis, and significant pathways were identified at $q < 0.05$ (Figure S1M).

Analysis controlling for stemness associate genes and genes with low expression in normal tissue

Expression of all expressed genes in each cancer were regressed against stemness. Regression p values for the stemness term was corrected for multiple testing by FDR. Genes with $q < 0.01$ and absolute stemness T-value > 75th percentile of all absolute T-values in the cancer are excluded from the analysis. To exclude genes with low/no expression in corresponding normal tissues we calculated mean expression of genes in the corresponding GTEx tissue and excluded all genes with mean expression < 5TPM from further analysis. Expression – CN regression analysis, phenotypic association and pathways analysis was performed as described above for frequently amplified or deleted genes in each cancer after excluding genes that meet the exclusion criteria.

Analysis of tumor subtypes

For tumor subtypes with at least 50 samples and > 1000 frequently amplified and deleted genes, we computed T_{CN} for frequently amplified and deleted genes as described above. For functional analysis, T_{CN} was centered at 2.5 and the resultant vector was used for GSEA analysis. Significant pathways were identified at $q < 0.05$ (Figure S1N).

To compare degree of uncoupling (see below) between different cancer subtypes, cancers with at least two subtypes with minimally 50 samples and > 1000 frequently amplified and deleted genes were used. The degree of uncoupling was compared between all groups using ANOVA and pairwise post hoc testing was performed using Tukey test. Significance was defined at $q < 0.1$.

Sample level analysis of uncoupling

The analysis was carried out in each tumor type separately. Gene expression was corrected for purity by regressing expression of each gene against tumor purity, and the residual expression was extracted and Z-transformed. Expression and CN of the top 100 coupled genes, based on the T_{CN} , as discussed above, were coalesced and expression was regressed against CN. The regression coefficient (β_{CN}) defines the expected change in expression for a unit change in CN. For a gene g in sample j with copy number l , the expected expression was defined as:

$$Eg_j = Eg_2 + \beta_{CN} * (l - 2),$$

where Eg_2 and Sg_2 are mean and standard deviation of expression of g in samples with CN = 2. Ag_j is the observed expression of gene g in sample j . If g is amplified in ($l > 2$) and its expression was lower than estimated expression (i.e., $Ag_j < Eg_j - Sg_2$) it's expression was considered uncoupled from its CN. If g was deleted in ($l < 2$) and its expression is higher than estimated expression (i.e., $Ag_j > Eg_j - Sg_2$) it's expression was considered uncoupled from its CN (Figure S2B).

To assess the functional impact of uncoupling, the degree of uncoupling (D_{UC}) was computed for each sample as the ratio of number uncoupled genes to number amplified/deleted genes. For each tumor type, samples were sorted based on the degree of uncoupling, and differential expression analysis using DESeq2⁶⁴ was performed, comparing samples with high D_{UC} (top 30%) to those with low D_{UC} (bottom 30%). T statistic from the differential expression analysis was used to perform GSEA analysis, significant pathways were identified at $q < 0.1$.

Degree of uncoupling survival analysis

To quantify association of D_{UC} with survival we first performed univariate cox regression analysis to identify cancers in which D_{UC} was significantly associated with survival ($q < 0.1$). In cancers where significant association with survival was identified we performed multi-variant cox regression analysis while controlling for age at diagnosis, stage, chromosomal instability and tumor subtype.

Building regulatory models

To build regulatory models, we first identified static miRNA and transcription binding sites. Human transcription factor PWMs (position weight matrices) were obtained from CIS-BP.⁶⁵ The Bioconductor package TFBSTools⁶⁶ was used to scan promoters (TSS-500 to TSS+1500) and FAMTOM5 enhancers⁵⁵. At each promoter/enhancer, the TFMPvalue() function was used to compute a p value for the binding sites. Predicted binding sites with p value $< 1e-4$ were retained. The binding score from TFBSTools was used as a surrogate for TF binding strength. miRNA binding was obtained by multiplying the absolute context score from TargetScan⁵⁷ by ten.

mRNAs (90th %ile expression < 30 normalized reads), miRNAs (90th %ile expression < 1) and enhancers (expressed (> 1 RPM (read per million)) in $< 10\%$ of samples) with low expression were excluded. mRNA, miRNA, enhancer expression and methylation M values were corrected for purity by regressing against tumor purity estimates and extracting the residual. These corrected values were used for regulatory modeling. For each gene, a linear regression model was constructed to explain its expression, which consists of: CN, promoter methylation, promoter and enhancer binding TFs and miRNAs. The model also took into account transcription factor and miRNA binding affinity and enhancer expression. The linear model was given by the following equation:

$$E_g \sim CN_g + M_g + \sum_{i \in pTF} \mu_{g,i} E_i + \sum_{i \in Enh_g} \sum_{j \in iTF} eE_j \delta_{i,j} E_j + \sum_{l \in miRNA_g} \alpha_{g,l} mE_l,$$

Where E_g , CN_g and M_g are the expression, CN and methylation of gene g , respectively. pTF is the set of TFs with a binding site in the promoter region of g . $\mu_{g,i}$ is the binding strength of a TF i ($i \in pTF$) in the promoter of g . If multiple binding sites are present, the average binding score was used. Enh_g is the set of enhancers associated with g . For an enhancer i in Enh_g , iTF is the set of TFs with a binding site in i and eE_j is the expression of the enhancer. $\delta_{i,j}$ is the binding score of a TF j ($j \in iTF$) in enhancer i . If multiple binding sites exist, their average binding score is used. $miRNA_g$ is the set of miRNA with a binding site in g . If l is a miRNA in $miRNA_g$, $\alpha_{g,l}$ is the average binding affinity of all binding sites of l in g and mE_l is the expression level of the miRNA l . The model was trained using the R package caret (<https://topepo.github.io/caret/>), using elastic nets (glmnet: <https://cran.r-project.org/web/packages/glmnet/index.html>) with 10-fold cross validation. 80% of the data was used for training and 20% for testing. Low quality regulatory models (test $R^2 < 0.4$) were discarded. Regression coefficients from the model quantify how each regulatory factor (transcription factors, miRNAs, CN and methylation) impact expression of the gene in question. The regression coefficients for TFs and miRNAs were used to construct a weighted directed transcriptional regulatory network (Note: methylation regression coefficient β_M , regression coefficient of transcription factor or miRNA β_{TF}).

The regulatory footprint of a TF is defined as the genes directly regulated by the TF. Weights corresponding to these interactions from the adjacency matrix are used to perform GSEA to define pathways regulated by the TF, significant pathways are defined at $q < 0.1$.

Identifying regulators that mediate uncoupling of expression from CN for individual genes

For every amplified or deleted gene in a cancer, purity corrected methylation levels were compared between uncoupled and coupled amplified (CN > 2) and deleted (CN < 2) samples respectively. A p value was obtained using a Wilcoxon signed-rank test and a T-value

using a t test, FDR (false discovery rate) correction was performed on the p values. Methylation is considered to mediate uncoupling of a gene's expression if 1) regulatory models predict that methylation suppresses expression of the gene ($\beta_M < 0$), and 2) there was a significant change in promoter methylation of the gene between uncoupled and coupled samples in the appropriate direction depending on context ($q < 0.01$ and T-value $> 0 / < 0$ for amplified and deleted genes, respectively).

A similar approach was used for TFs and miRNAs. For a gene of interest, the purity corrected expression of regulatory molecules (TFs and miRNAs) were compared between coupled and uncoupled samples using a Wilcox signed-rank test to compute a p value and t test for t-value, and p values were corrected for multiple testing using FDR. A regulatory factor was considered to facilitate uncoupling if 1) it shows significant differential expression between coupled and uncoupled samples ($q < 0.01$) and 2) its direction of regulation and change in expression explains the uncoupling (Amplified genes: $\beta_{TF} * t\text{-value} < 0$ or Deleted genes: $\beta_{TF} * t\text{-value} > 0$).

Co-expression modules of uncoupled genes

Co-expressed modules of amplified or deleted nCpD genes were identified using the WGCNA,³⁰ using purity corrected expression. The `blockwiseModules()` function was run on the expression data with `corType = "bicor"` to construct a signed network and identify module (size $> = 10$) of co-expressed genes. The clusters were further given an association score for 4 phenotypes (immune infiltration, apoptosis, cell cycle and EMT) by correlating the eigen-gene vector with sample phenotype scores using Spearman correlation and a p value was computed using the function `corPvalueFisher()`. Functional enrichment of the cluster was performed using `enrichPathway()`, with significantly enriched pathways identified at $q < 0.05$ after a FDR correction.

Predicting phenotypic consequences of gene perturbation using heat diffusion

To model the phenotypic effect of perturbing a TF, we made use of insulated heat diffusion described in Hotnet2.³¹ Briefly, the original adjacency matrix (W) of the regulatory network is used to generate two additional matrices, A : weights were converted to absolute values and normalized and I : an unweighted adjacency matrix. I was used to calculate shortest paths between TF's and all genes in the network. To model the effect that a unit positive perturbation of a TF has on other genes, we used an insulated heat diffusion model where β is the insulation parameter governing the amount of heat a node holds on to, and F gives the degree to which the perturbation spreads to other nodes.

$$F = \beta(I - (1 - \beta)A)^{-1}$$

The optimal value of β was determined by running the diffusion model on β values between 0 and 1 in increments of 0.05. The optimal β values maximizes heat in neighboring nodes compared to non-neighbors. To access the direction of impact on a target node (n) when a TF (s) is perturbed, we used the product of the signs of edge weights in W along the shortest path ($shpath(s,n)$) between s and n . IF is the matrix of these values, where each entry was defined as:

$$IF_{n,s} = \prod_{ij \in shpath(s,n)} A_{ij} / |A_{ij}|$$

The impact of perturbing a TF (P_{tf}) depends on the genes it influences, defined by F and IF and the individual phenotypic associations of these genes defined by the regression coefficient β_{Pheno} (see above). Note that only genes significantly associated with the phenotype ($q < 0.01$) were used for the analysis and were denoted by the set G . The equation below describes how P_{tf} was computed. This score is the weighted sum across gene whose expression is associated with the phenotype ($q < 0.01$), where the terms are weighted by 1) the amount of perturbation flowing to the gene, 2) the strength and direction of association of the gene's expression with the phenotype (β -value) and 3) direction of impact of the perturbation on the gene's expression.

$$P_{tf} = \sum_{g \in G} F(tf, g) IF(tf, g) \beta_{Pheno_g}$$

Target selection

To identify potential TFs, we first identified clusters of amplified/deleted nCpD genes using WGCNA (see above). For each cluster, TF's were picked based on 3 criteria, briefly:

1. Regulatory score is the weighted sum across genes in the cluster C regulated by the TF, the weights account for 1) whether the TF is likely to mediate uncoupling of the gene (differentially expressed between coupled and uncoupled samples in direction to explain uncoupling of the regulated gene), 2) strength of the regulatory interaction and 3) importance of the regulated gene in signaling networks. Regulatory score weights up TF's that consistently regulate a cluster of genes in the same direction and strongly regulate central genes in PPI (protein-protein interactions) networks. The score also weighs up TF's whose expression is significantly different between uncoupled and coupled samples ($q < 0.01$). For a cluster C with genes g and a transcription factor tf , the score is defined as

$$Rscore_{(tf,C)} = \sum_{i \in g} -1^a T(tf, i) A(tf, i) EigenC(i),$$

where a is 1 for amplified nCpD clusters and 0 for deleted nCpD clusters, T is the t-value matrix with TF's as rows and genes as columns. Each entry is the t-value for expression of a TF compared between uncoupled and coupled samples of a gene. A Wilcox

signed-rank test is used to calculate a p value for each entry, which is in turn used for FDR. Entries with $q > 0.01$ are set to t-value/absolute(t-value). A is the adjacency matrix of the TF-gene regulator network and $\text{EigenC}(i)$ gives the eigenvector centrality of i in the PPI network. An empirical p value was computed for $Rscore$ by generating 100 random adjacency matrices by randomly swapping input node of two randomly picked edges. This procedure was repeated 0.4 times the total number of edges to generate a single random matrix. A vector of 100 random $Rscores$ ($rRscore$) was generated for each of the random adjacency matrices and the p values ($RscoreP$) was defined as:

$$\text{If } Rscore > 0 : RscoreP = \frac{\text{length}(rRscore > Rscore)}{\text{length}(rRscore)} ; \text{ or}$$

$$\text{If } Rscore < 0 : RscoreP = \frac{\text{length}(rRscore < Rscore)}{\text{length}(rRscore)}$$

2. TF phenotype scores $PhenoTF_{tf} = (P_{tf(\text{apoptosis})} + P_{tf(\text{infiltration})}) - (P_{tf(\text{cell cycle})} + P_{tf(\text{EMT})}) \cdot P_{tf(S)}$ were computed for every phenotype 'S' using heat diffusion, as described above. A positive $PhenoTF_{tf}$ score indicates that an increasing expression of the TF increases apoptosis/infiltration or decreased cell cycle/EMT. The score if Z-transformed for all TF's regulating a cluster and TFs with absolute $PhenoTF_{tf} > 1.5$ were considered for further analysis. Note in each cancer if a phenotype has < 5% of genes significantly associated with the phenotype ($q < 0.01$) the phenotype was excluded to reduce noise.

3. Individual phenotype scores for a cluster 'C' (see **WGCNA section**) were combined as $PhenoClus_C = (P_{C(\text{apoptosis})} + P_{C(\text{infiltration})}) - (P_{C(\text{cell cycle})} + P_{C(\text{EMT})}) \cdot P_{C(i)}$ – denote individual score of phenotype i . A p value was computed by randomizing the phenotype score of samples and re-computing $PhenoClus$ (and $P_{C(i)}$). For each of the thousand randomizations, p value was calculated similar to the $Rscore$ (see above).

Targets were selected based on $Rscore_{(tf,c)}$, $PhenoTF_{tf}$ and $PhenoClus_C$. In the case of cluster of amplified nCpDs $PhenoClus_C > 0$ and $p \leq 0.05$ were selected. We then selected TF's associated with these cluster such that $Rscore_{(tf,c)} * PhenoTF_{tf} > 0$ and $RscoreP_{tf} \leq 0.01$. In case of cluster of deleted nCPDs $PhenoClus_C < 0$ and $p \leq 0.05$ and $Rscore_{(tf,c)} * PhenoTF_{tf} < 0$ and $RscoreP_{tf} \leq 0.01$, see [Figure S4D](#) for pictorial representation. Additionally, TF's that regulate at least 10% of the genes in the cluster were selected.

Immuno-therapy datasets analysis

Gene expression and clinical data were obtained from three immunotherapy datasets Liu et al.,³⁴ Hugo et al.,³⁵ and Riaz et al.³⁶ Liu et al.³⁴ and Hugo et al.³⁵ had normalized expression data while counts were obtained for Riaz et al.³⁶ The datasets were analyzed as follows: 1. Liu et al.³⁴ For all predicted targets, identified association of expression with overall and progression free survival was quantified using Cox regression, while controlling for gender and tumor stage. Differential expression of these genes in responders relative to non-responders was quantified using Wilcoxon signed-rank. 2. Hugo et al.³⁵ was analyzed in the same way as Liu et al.,³⁴ with the exception that only over-all survival was quantified. 3. Riaz et al.³⁶ Differential expression of genes between responders and non-responders was performed using DESeq2⁶⁴ and target TFs were extracted. Survival analysis was performed as described above, controlling for stage, cohort and sub-type of the cancer. All survival analysis was performed using the R library *survival*. Significance of differential expression and survival analysis was defined at $p \leq 0.05$.

Randomized testing to obtain empirical p value for fraction of TFs validated

To quantify a degree of significance for the fraction of putative TF targets validated by the *in silico* approach above, we made use of a randomization based approach. A random set of TFs were constructed after excluding TFs identified above. These TFs were then randomly assigned to the 6 cancers types and the validation analysis was repeated to identify the fraction of genes validated. This analysis was repeated 1000 times to obtain a distribution of fraction of genes validated (X), p value was defined as the number of values in $X > 0.428$ divided by 1000.

Analysis of sensitivity to azacitidine

We calculated average CN across genes in the hallmark apoptosis pathway and it's activity (ssGSEA implemented in GSEA⁶⁷) in CCLE cell lines ([Figure S6B](#)).¹⁴ Apoptotic pathway activity was factorized into two levels, low (\leq median) and high ($>$ median). We then quantified the association between response to azacitidine and activity of apoptotic pathway using the linear model:

$$AUC \sim Pa * \text{tissue_of_origin}$$

where AUC is area under the drug response curve of azacitidine obtained from CTRPv2,⁶¹ Pa is activity level of apoptotic pathway and tissue_of_origine is the normal tissue of origin of the cancer cell line. aCN is the average CN of apoptotic genes in the cell lines. Lower AUC values indicate greater sensitivity to the drug. The analysis was carried out in non-amplified cell lines ($aCN \leq 2.5$) and amplified cell lines separately ($aCN > 2.5$) ([Figures S6C and S6D](#)). The regression statistics are reported in [Table S3](#).

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis was performed in R. linear association between continuous variables (or continuous and nominal variables like CN) we quantified using linear regression or Spearman correlation. Difference between continuous distributions were tested using 1. Two sided Wilcoxon test/ t test in case of two variables and 2. ANOVA in cases with more than two variables. Survival analysis was performed using COX-regression analysis implemented in the R library *survival*. Correction for false discovery was performed with 1. FDR using the function *p.adjust()* or 2. By randomization when indicated. Significance is defined at $p < 0.05$ and $q < 0.1$ (in case of correction for multiple testing) unless specified otherwise.