ACE: A Probabilistic Model for Characterizing Gene-Level Essentiality in CRISPR Screens

Elizabeth R. Hutton[1], **Christopher R. Vakoc**[1], **and Adam Siepel**[1, ✉]

[1]Cold Spring Harbor Laboratory, Cold Spring Harbor NY, USA

High-throughput knockout screens based on CRISPR-Cas9 are widely used to evaluate the essentiality of genes across a range of cell types. Here we introduce a probabilistic modeling framework, Analysis of CRISPR-based Essentiality (ACE), that enables new statistical tests for essentiality based on the raw sequence read counts from such screens. ACE estimates the essentiality of each gene using a flexible likelihood framework that accounts for multiple sources of variation in the CRISPR-Cas9 experimental process. In addition, the method can identify genes that differ in their degree of essentiality across samples using a likelihood ratio test. We show using simulations that ACE is competitive with the best available methods in predicting essentiality, and is especially useful for the identification of differential essentiality. Furthermore, by applying ACE to publicly available CRISPR-screen data, we are able to identify both known and previously overlooked candidates for genotype-specific essentiality, including RNA m^6-A methyltransferases that exhibit enhanced essentiality in the presence of inactivating *TP53* mutations. In summary, ACE provides improved quantification of essentiality specific to cancer subtypes, and a robust probabilistic framework for identifying genes responsive to therapeutic targeting.

CRISPR | essentiality | statistical modeling | machine learning

Correspondence: ***asiepel@cshl.edu***

Pathogenic mutations and genotype-specific cancer liabilities can now be tested at unprecedented scales, owing to CRISPR-Cas9 (Clustered Regularly Interspaced Short Palindromic Repeats — CRISPR-associated protein 9) technology and large panels of diverse cell lines (1–10). These technologies allow genetic vulnerabilities to be demonstrated through direct perturbation, rather than through indirect associations from population enrichment studies or gene expression levels. Furthermore, in comparison to previous screens based on RNA interference, CRISPR knockout screens offer substantially improved sensitivity and specificity, due to increased effectiveness at disrupting targeted elements and reduced off-target effects (11–13).

In a typical CRISPR-Cas9 screen (Fig. 1a), the artificially expressed Cas9 nuclease introduces a double-stranded break in native DNA, guided by homology with the separately infected single-strand guide RNA

(sgRNA). The cell then repairs the double-stranded break, typically inserting or deleting several nucleotides at the cut site in the process. The resulting genetic lesion often disrupts the function of the targeted region. The genetically modified cell is allowed to proliferate for a series of divisions, after which the effects on cell growth are measured by sequencing the sgRNA present in the final surviving population of cells. Lethal sgRNA constructs that have disrupted sites critical for normal cell growth will be underrepresented in the final population in contrast to 'nonessential' control sgRNAs. Moreover, differential essentiality can be detected from different levels of sgRNA depletion between samples. Hundreds of different cell lines have now been assayed in publicly available CRISPR knockout screens, offering an excellent opportunity to compare gene essentiality between panels of cancer cell lines according to their oncogenic mutations.

Within the general paradigm of a CRISPR-Cas9 screen, several variations of experimental design are possible. For example, many experiments evaluate the depletion of cells infected with lethal sgRNA by sequencing both the initial infected and the final population of cells (14–17). In contrast, other experiments streamline the process by sequencing only the initial pool of sgRNA constructs prior to infection (9, 18, 19), and use the sgRNA abundances in this 'master library' as a proxy for the initial frequencies. However, the sgRNA abundances in the master library vary by several orders of magnitude, and each biological replicate represents a separate infection into a new pool of cells, making this proxy imperfect. Overall, the variation in experimental design necessitates a flexible analytical framework that can robustly account for different sources of experimental error.

Through deliberate selection of the cell lines used, a CRISPR-Cas9 screen can reveal genes that are essential only in a specific oncogenic context, for example, in the presence of a loss-of-function mutation of an important tumor suppressor such as *TP53*. These context-dependent essential genes in cancer — sometimes called 'non-oncogene addictions' — are of particular interest because they may offer druggable therapeutic targets even when irreversible loss-of-function mutations to oncogenes are present (20). A better understanding of genotype-specific gene essentiality has the potential to enable patient-specific therapeutics that are tailored to mutations present at the time of diagnosis. Non-oncogene addiction has recently achieved prominence, for example, in the first round of chemical inhibitors for RNA epigenetic modifiers soon to enter phase I trials (21).

However, statistical tests for identifying such context-dependent essentiality require special care. Not only are they subject to reduced power because they require contrasting one subset of the data with another,

but they can yield spurious results if the 'test' and 'control' subsets are not adequately matched in other respects. Thus, rigorous testing frameworks are needed that efficiently make use of the data and control for such biases.

Many of the statistical tests that have been applied to CRISPR-Cas9 screens rely on summary statistics such as average read counts across samples, or log-fold changes of read counts between the initial and final populations of cells. However, these summary statistics are incomplete descriptions of the raw data, leading to inevitable loss of power. In addition, these tests can be difficult to adapt to alternative experimental designs. For example, methods such as BAGEL and JACKS rely upon the log-fold change of read abundances (1, 2, 8–10, 22), but these log-fold changes may be calculated with respect to read counts obtained from either the initial infected population of cells or the master library, with potentially substantial impacts on the power of the tests. Another issue is that some methods reduce the estimation of gene essentiality to a binary classification of 'essential' or 'nonessential' genes (2, 4, 23), enabling the identification of clearly essential genes but blurring the comparison of weak and strong effects. Finally, most methods do not test directly for differential essentiality within sample subtypes (see Supplementary Fig. S1 for a comparison of methods).

In this paper, we introduce the Analysis of CRISPR-based Essentiality (ACE), a flexible probabilistic method that directly estimates gene-level essentiality values by maximum likelihood, accommodates variable experimental designs, and enables rigorous likelihood-based tests of both absolute and differential essentiality. The ACE model describes each experimental phase of the CRISPR screen — the master library infection, the initial and final sequencing — as a separate probabilistic process forming a hierarchical Bayesian model. The modularity of our framework permits the analysis of a variety of CRISPR-screen data sets while accounting for differences in experimental design. We validate ACE using both simulations and published sets of essential and nonessential genes. We then apply the method to publicly available data from Achilles DepMap (9, 24) and demonstrate several compelling examples of differential essentiality in non-small cell lung cancer, including both known and novel cases.

Results

Model Overview. ACE makes use of a hierarchical Bayesian model design to integrate information at the sample, gene, and individual sgRNA levels, with separate parameters at each level (see Fig. 1). The model is structured to account separately for the major sources of uncertainty at each stage of the experimental

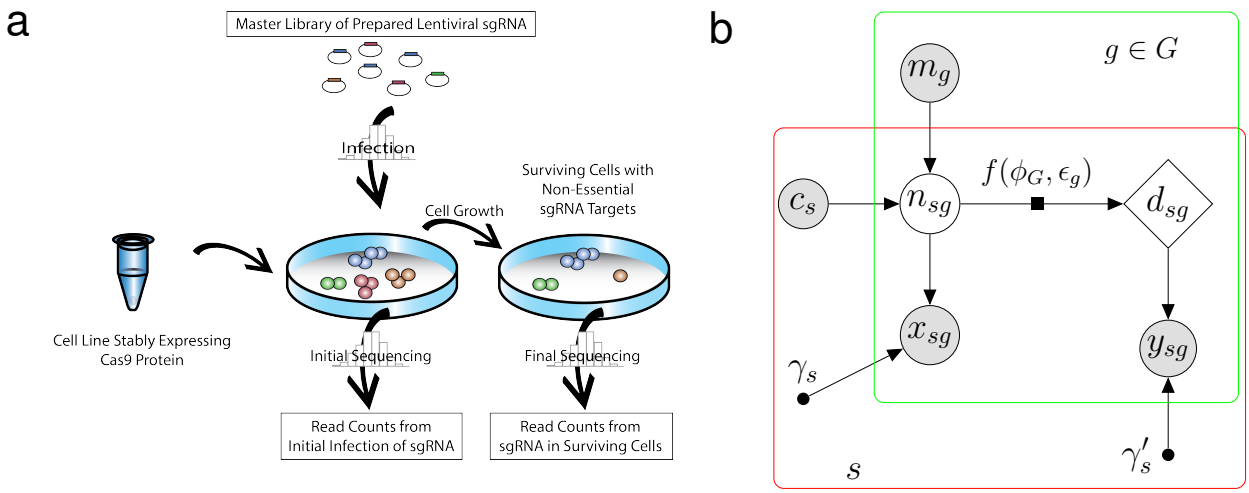


Fig. 1. CRISPR-Cas9 Negative Selection Screen and Probabilistic Graphical Model. (a) The initial pool of Cas9-expressing cells is infected with a lentiviral sgRNA master library. The initial sgRNA abundances are measured by sequencing either the newly infected population of cells, the master library, or both. Cells are allowed to grow, and final sgRNA abundances are obtained by sequencing the surviving cells. (b) Graphical model illustrating how information is pooled across samples (s) and sgRNAs targeting each gene ($g \in G$) to infer each gene's essentiality, ϕ_G. From top-left to bottom-right, m_g represents the fraction of the master library consisting of sgRNA g; c_s represents the number of cells in sample s (estimated by the user); n_{sg} represents the number of cells initially infected with sgRNA g (such that $n_{sg} \leq c_s$); d_{sg} represents the corresponding number of surviving cells after cell growth; and x_{sg} and y_{sg} represent the read counts obtained by sequencing g in the initial and surviving cells, respectively. The numbers of infected cells prior to (n_{sg}) and following (d_{sg}) growth are assumed to be related by the deterministic function $d_{sg} = f(n_{sg}; \phi_G, \epsilon_g) = n_{sg}(1 - \epsilon_g \phi_G)$ (indicated by factor-graph notation). The efficiency ϵ_g is determined by logistic regression and the gene essentiality values ϕ_G are estimated by maximum likelihood (see Methods). The prior distribution for n_{sg} and the sampling distributions for x_{sg} and y_{sg} are assumed to be Poisson. The scaling factors γ_s and γ'_s accommodate global properties of sequencing depth and cell growth, and are estimated in pre-processing (see Methods). Shaded nodes represent variables observed in the data, unfilled nodes represent latent variables, and smaller solid circles represent free parameters.

process. In addition, if samples are partitioned into 'test' and 'control' groups, ACE can estimate gene essentiality within each sample group separately, allowing for the detection of differential essentiality (discussed further below).

As illustrated in Fig. 1b, the structure of the ACE model reflects the experimental design of a CRISPR-Cas9 screen. The initial number of infected cells n_{sg}, for each sgRNA g and sample (replicate) s, is assumed to be Poisson-distributed with an expected value given by the relative frequency of guide g in the master library, m_g, scaled by the number of cells, c_s, infected in the assay. Estimates of both m_g and c_s are provided by the user. The final number of infected cells, d_{sg}, is then assumed to be given by a scaled version of the initial number, $d_{sg} = n_{sg}(1 - \epsilon_g \phi_G)$, where $\epsilon_g \in [0, 1]$ denotes the editing efficiency of sgRNA g and $\phi_G \leq 1$ denotes the essentiality of gene G. These initial and final numbers of infected cells are not observed directly (they are latent variables) but are indirectly sampled via the sequencing process, resulting in observed read counts of x_{sg} and y_{sg}, respectively, which are provided by the user as inputs to the inference procedure. We assume Poisson-distributed read counts x_{sg} and y_{sg}, and sum over possible values of the latent variable

n_{sg} in the likelihood calculation (see Methods). Notably, the combination of the Poisson prior and the Poisson sampling distribution adequately captures the observed overdispersion in the read count data (see Supplementary Fig. S2). Independence is assumed across samples and replicates. The free parameters of the model are estimated by numerical maximization of the likelihood function (see Methods). The ACE software reports estimates of not only the essentiality of each gene, ϕ_G, but also the efficiency of each sgRNA, ϵ_g.

A value of $\phi_G = 1$ represents complete essentiality (driving $d_{sg} = n_{sg}(1 - \epsilon_g \phi_G)$ to zero in the case of efficient editing, $\epsilon_g = 1$), whereas a value of $\phi_G = 0$ represents complete nonessentiality (allowing $d_{sg} = n_{sg}$ even when $\epsilon_g > 0$). Notably, however, our definition of ϕ_G also allows for negative values, reflecting increased cell growth after inactivation of gene G. Such an increase might occur, for example, if a tumor suppressor gene were deactivated by guide RNAs. In practice, negative estimates of ϕ_G are fairly frequently observed, either by random chance or from true increases in cell growth.

To pool information across sgRNAs and ensure identifiability of ϵ_g and ϕ_G, we determine ϵ_g by logistic regression based on a user-defined set of features, and treat the regression coefficients only as free parameters (Methods). The efficiency of a class of sgRNA has been shown to be strongly correlated with a number of genomic features, including local GC content, proximity to heterochromatin, number of base mismatches, and abundance of off-target sites (18, 25–28). However, for simplicity and efficiency, we have used only the GC content of the sgRNA (one of the strongest predictors of efficiency) in our initial implementation. Nevertheless, ACE can easily be extended to consider additional features.

The two sample-specific scaling factors, γ_s and γ'_s, determine the relationships between the numbers of infected cells and the expected sgRNA counts in the initial and final data sets, respectively. These parameters apply globally to all sgRNAs and all genes. Rather than including these parameters in the numerical optimization of the likelihood function, we find it adequate to estimate them in preprocessing based on a median-of-ratios normalization. Notably, γ_s is primarily a reflection of sequencing read depth, whereas γ'_s is determined by both read depth and sample-specific factors in cell growth (see Methods). In the case where the bulk distribution is strongly influenced by essential genes, as when a library has a large number of essential targets, γ'_s can alternatively be estimated from designated negative controls.

Differential Essentiality. In the case of differential essentiality, ACE compares two (log) likelihoods for each gene G: a likelihood maximized under the constraint that the essentiality, ϕ_G, is the same in designated

'test' and 'control' sample categories, and a likelihood that allows for different values of ϕ_G in these two categories. These sample categories can be selected by the user to examine differences, for example, in cell lines with and without mutations in *KRAS*, or lung- versus liver-derived sample panels. The two per-gene likelihoods are then compared in a likelihood ratio test, with significance assessed by comparison with an empirical null distribution based on simulated data (see Methods). In this way, ACE performs a rigorous test of the alternative hypothesis that gene G exhibits different degrees of essentiality in the test and control categories against a null hypothesis of equal essentiality.

Simulation Study. To evaluate the ability of ACE to infer sgRNA-, sample-, and gene-specific parameters, we devised a simulation method to generate artificial CRISPR-screen data sets. To ensure that our simulated data resemble data from real CRISPR screens as closely as possible, our simulator, called empiriCRISPR, mimics the variation in read counts observed in a reference experiment and makes minimal assumptions about the process by which the data are generated (see Methods and Supplementary Fig. S3). Importantly, empiriCRISPR is fully independent of the probabilistic model and parameter inference used in ACE.

Benchmarking of Absolute Essentiality Predictions. Using empiriCRISPR, we generated data sets of 5,250 genes in which the essentiality of each gene ranged from complete nonessentiality ($\phi_G = 0$) to complete essentiality ($\phi_G = 1$). To ensure a fair comparison with methods that rely on a median-based normalization method, we took care to include an adequate number of nonessential genes in each data set, simulating 3,150 genes with $\phi_G = 0$ (60% of all genes). The remaining 40% of genes consisted of 300 genes at each of several intermediate values of ϕ_G and 600 genes at $\phi_G = 0.99$ (see Fig. 2a). For each choice of simulation parameters, three replicates were simulated with four sgRNAs targeting each gene and an average of 300 reads per sgRNA. As a template data set for empiriCRISPR, we used a collection of previously published genome-wide CRISPR screens in which both the master library and initial infected sgRNA abundances were sequenced along with the depleted samples (16). The master library was generated through random selection of sgRNA from the template screen, followed by simulation of read counts according to the variation observed in the template.

We first compared ACE's estimates of essentiality with an estimator based on the average fold change (AFC) in read counts (see Methods), which has been used in several previous studies (17, 29) (see Fig. 2a). The ACE estimates are approximately unbiased and exhibit relatively low variance (see also Supplementary

Fig. S4). By contrast, the AFC-based estimates display a downward bias, particularly at low simulated values of ϕ_G, and show substantially greater variance. Notably, the simple AFC-based method often yields negative estimates of ϕ_G even when the true ϕ_G is positive.

Many methods do not provide a numerical estimate of the essentiality of a gene, but instead focus on a binary classification of genes into 'essential' and 'nonessential' categories. To compare ACE with these methods, we simulated data sets containing 300 genes at each of several values of ϕ_G, together with 3,150 nonessential genes. For each of these experiments, we then compared the performance of ACE in binary classification with BAGEL (1), JACKS (8), and a classifier based on simply thresholding the AFC statistic. We summarized the performance of all methods using Receiver Operating Characteristic (ROC) curves and the Area Under the ROC Curve (AUC) metric (Fig. 2b and (c)). We found that all methods performed similarly well at high values of ϕ_G, with the AFC-based method performing slightly worse, but in the harder scenarios where ϕ_G takes moderate to low values ($\phi_G < 0.5$), JACKS and ACE performed significantly better than the others. However, JACKS slightly outperformed ACE on this task, apparently because it makes particularly efficient use of the negative controls in defining a null distribution.

Benchmarking of Differential Essentiality Predictions. To test the detection of differential essentiality, we simulated CRISPR-screen data in two samples ('test' and 'control') that included 300 genes at each of several values of $\phi_{G_{test}}$ that were essential only in the 'test' sample (see Methods). To represent uniformly essential genes, both samples also included 300 genes at strong ($\phi_G = 0.99$) and 300 genes at moderate ($\phi_G = 0.5$) essentiality, as well as a large collection (3,150) of nonessential genes ($\phi_G = 0$). Similar to the previous tests, we compared the performance of ACE with BAGEL, JACKS, and a simple AFC-based method in the detection of the differentially essential genes. Because the other methods do not support direct tests for differential essentiality, we devised tests based on the results of two separate tests for essentiality. After some experimentation, we settled on using the ratio of p-values for the test and control panels for JACKS, and the absolute difference in the Bayes Factors for the test and control panels for BAGEL (see Methods).

We found that ACE's likelihood ratio test for differential essentiality did result in substantially better power than the competing methods (Fig. 3). All methods had reasonably good power when $\phi_{G_{test}}$ was large (≥ 0.6) with JACKS performing somewhat more poorly than the others; but at values of $\phi_{G_{test}} < 0.6$, ACE showed a clear advantage. Even when $\phi_{G_{test}} = 0.2$, ACE was able to detect 67% of differentially essential

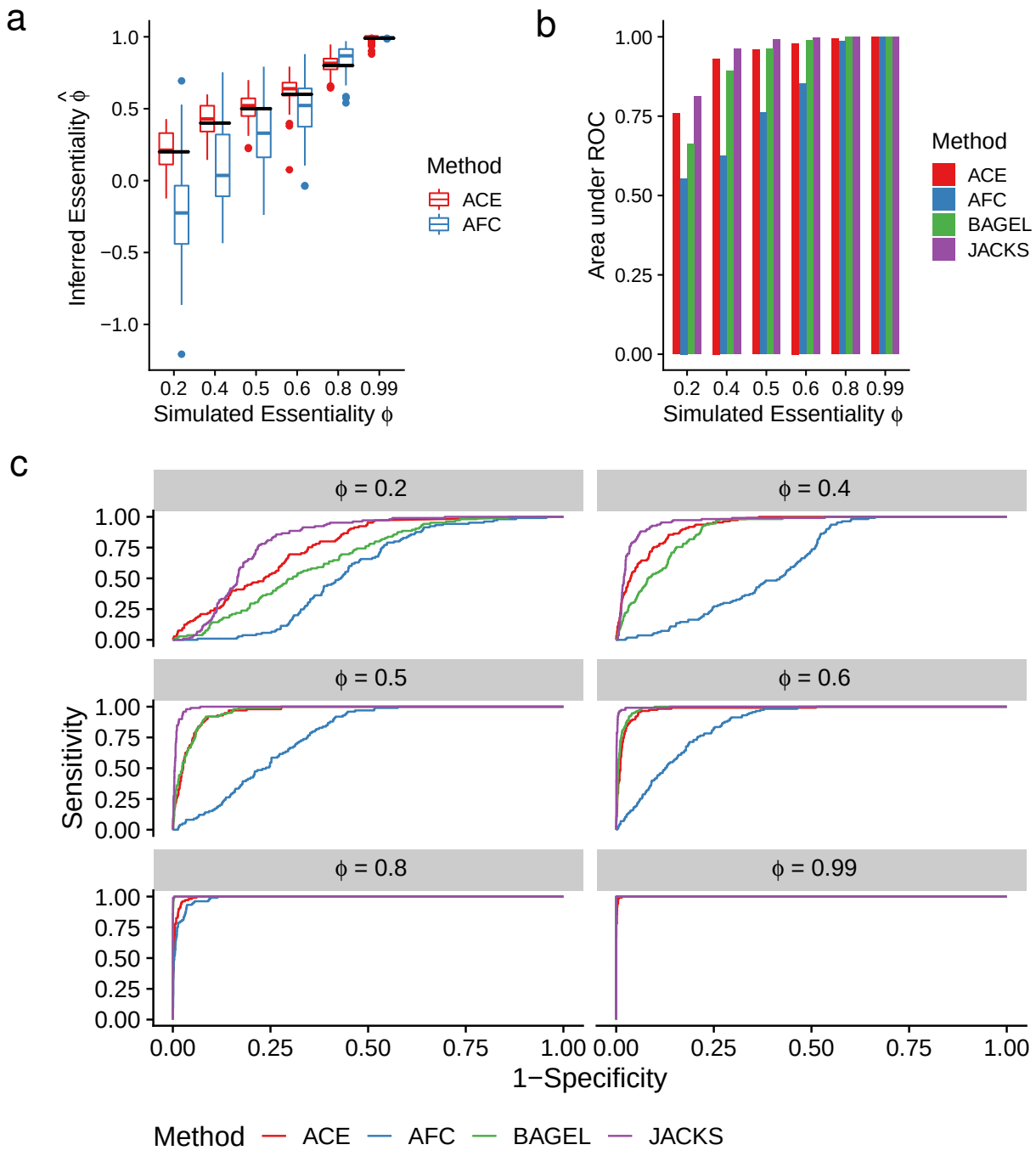


Fig. 2. Detection of Absolute Essentiality in Simulated Data. (a) Inferred essentiality values ($\hat{\phi}$) for 300 genes simulated in three replicates at each essentiality value shown (ϕ). Simulated data sets also included 3,150 nonessential genes ($\phi_G = 0$; not shown), 300 of which were provided to each method as negative controls. Black horizontal bars indicate true values ($\hat{\phi} = \phi$). (b) Performance in binary classification (essential vs. nonessential) of simulations of 300 'essential' genes at various values of ϕ_G and 300 'nonessential' genes. Performance reported as the Area under the Receiver Operating Characteristic (ROC) curve. Only initial and final read counts from the simulation were used for all methods. (c) Full ROC curves for each of the simulated essentiality values. 'ACE' — our probabilistic method (thresholded on log likelihood ratios), 'AFC' — method based on average fold changes in sgRNA abundance (thresholded on z-scores; see Methods). 'BAGEL' — Bayesian Analysis of Gene Essentiality (2) (thresholded on reported Bayes Factors); 'JACKS' — Joint analysis of CRISPR/Cas9 knockout Screens (8) (thresholded on reported p-values).

genes (Fig. 3b).

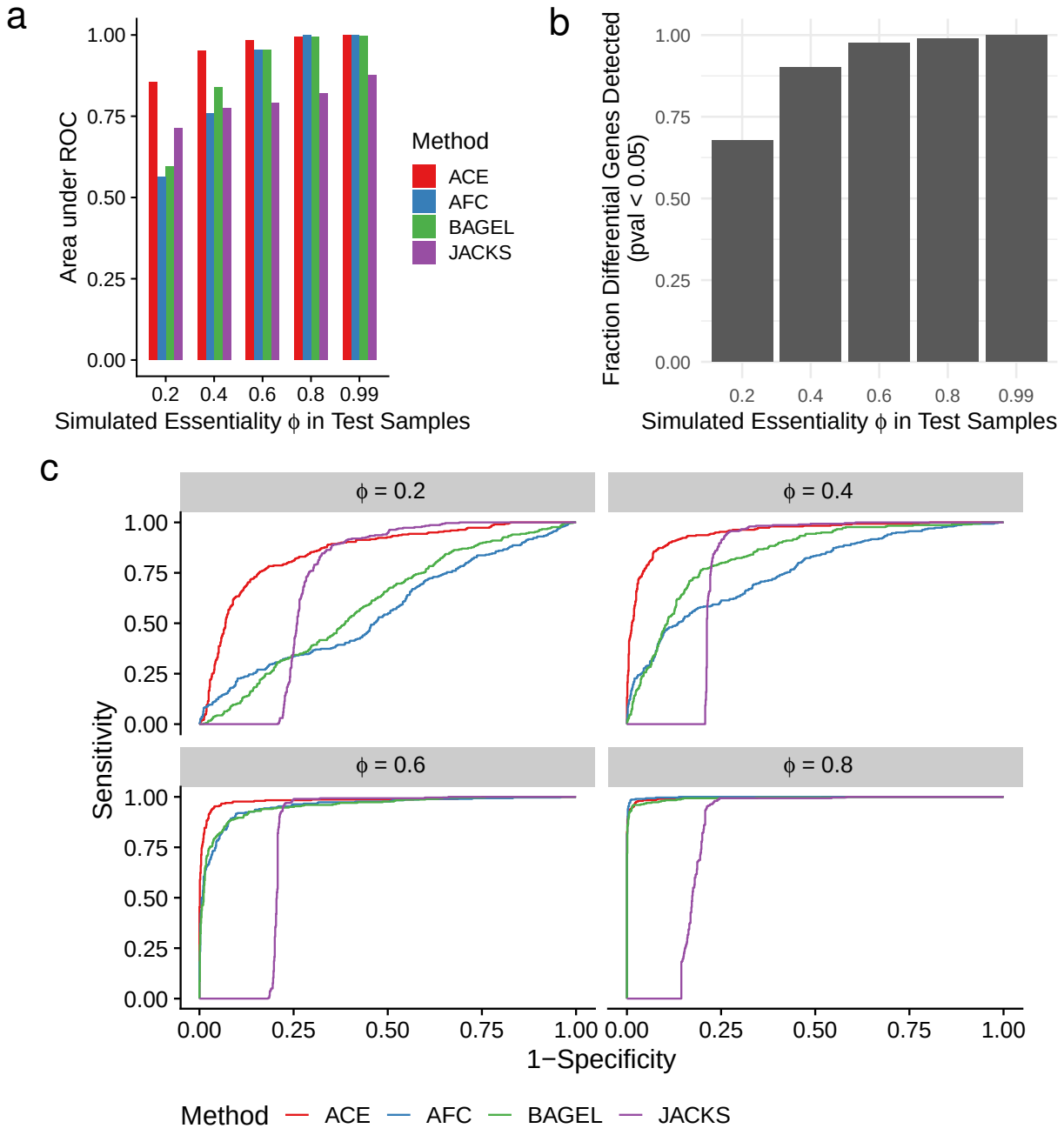


Fig. 3. Detection of Differential Essentiality in Simulated Data. (a) Performance in binary classification of differential essentiality for 300 genes that were 'essential' (at various levels of simulated essentiality ϕ) in 'test' samples and 'nonessential' ($\phi = 0$) in 'control' samples. Both samples also included 300 genes at strong ($\phi_G = 0.99$) and 300 genes at moderate ($\phi_G = 0.5$) essentiality, as well as a large collection (3,150) of nonessential genes ($\phi_G = 0$). 300 genes from each of these uniformly essential sets were used as negative controls. (b) Fractions of differentially essential genes correctly detected by ACE (at a Bonferroni-corrected empirical p < 0.05) at various values of simulated essentiality ϕ. (c) Full ROC curves for each of the simulated differential essentiality levels. In all cases, three replicates of the control and test samples were simulated for each value of ϕ. Methods are as described in Fig. 2.

Analysis of Real Data.

Identification of Essential Genes. We next sought to evaluate the performance of ACE in detecting previously identified essential genes from real data. We focused on 688 genes that have been classified as essential in at least two previous studies (1–3), based on both RNAi and CRISPR screens (19). For negative controls, we used 634 genes that are not widely expressed according to public RNA-seq data (see Methods). We used ACE to evaluate the essentiality of both gene sets using the CRISPR-screen data from Achilles DepMap. Notably, this data set includes sequence data representing only the sgRNA master library and the post-depletion sgRNA abundances. To ensure a similar genetic background between samples, we focused on 92 unique cell lines that all originate from lung adenocarcinoma, a subtype of non-small cell lung cancer (NSCLC).

Considering the many differences in methods and data sets, we found that ACE's characterization of these genes was fairly concordant with previous results. The genes previously identified as essential (positive controls) yielded both clearly elevated estimates of essentiality ($\hat{\phi}$) and clearly elevated log likelihood ratios (LLRs) in a test for non-zero essentiality relative to the negative controls (Fig. 4a). At a LLR threshold defined such that only 5% of the negative controls exceeded it (i.e., with empirical $p = 0.05$), 90.5% of the positive controls were predicted to be essential. In an ROC analysis based on the same data, ACE was able to distinguish the positive and negative controls with high accuracy (AUC = 0.97; see Supplementary Fig. S5).

Detection of Known Cases of Genotype-Specific Essentiality. One possible application of ACE's test for differential essentiality is to identify significant differences in essentiality between distinct genetic backgrounds. For example, cases of synthetic lethality may appear as genes that are essential in test samples in which another gene is mutated and nonessential in control samples in which that gene is intact. Conversely, other genes may be found to be essential in the presence of the wildtype but not the mutant genotype. To explore this possibility, we contrasted cell lines annotated by the Cancer Cell Line Encyclopedia (CCLE) with variants of two very well-known cancer-related genes, the tumor suppressor *TP53* and the proto-oncogene *KRAS* (see Methods) (30, 31).

We restricted ourselves to NSCLC samples in all cases, to avoid spurious signals unrelated to the mutations of interest. Moreover, we structured our test specifically to identify genes that were significantly more (rather than less) essential in the test samples than in the control samples. As in the previous analysis, we used data from Achilles DepMap (9, 24).

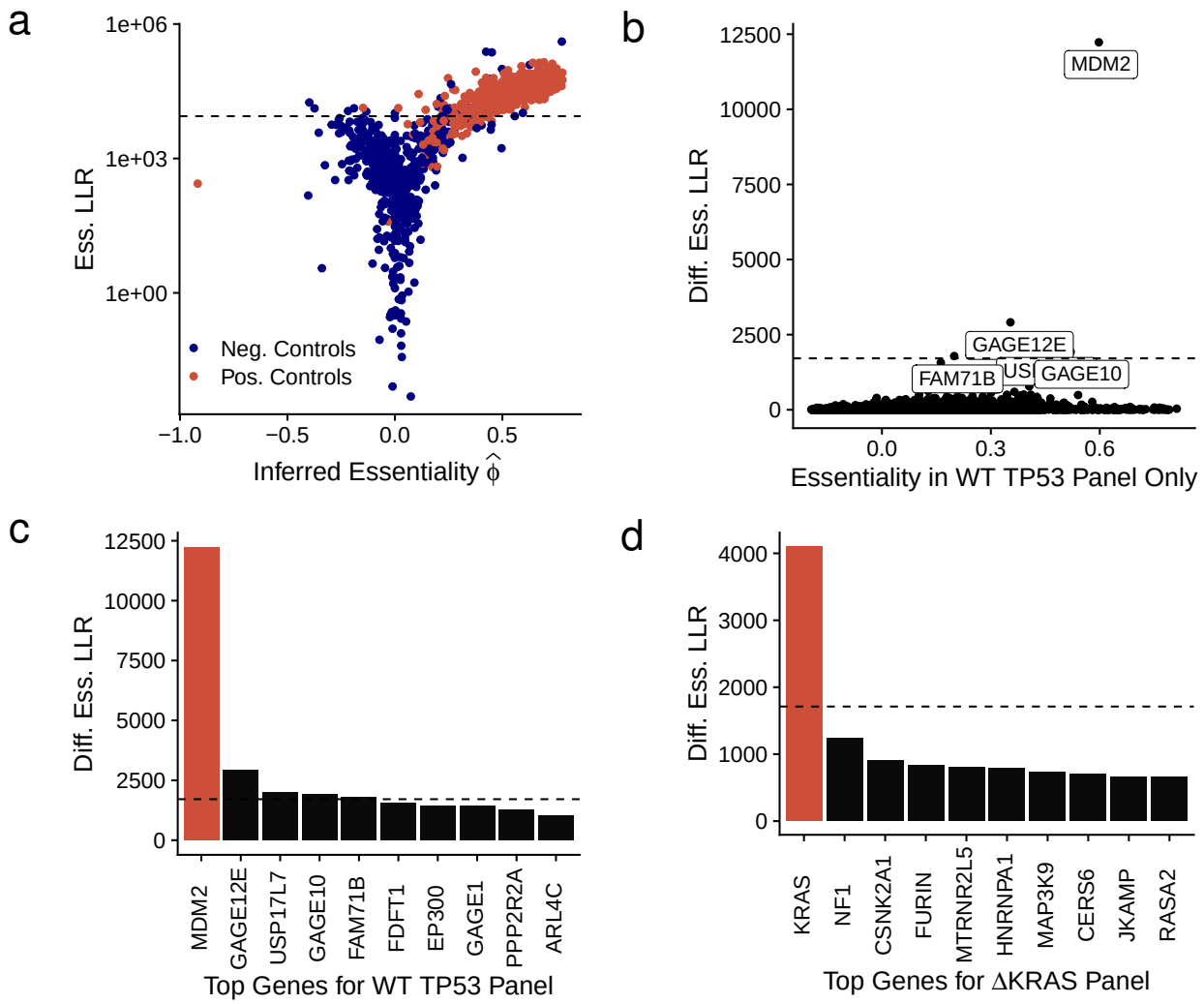


Fig. 4. Identification of known genotype-specific essentiality. (a) Estimates of essentiality per gene ($\hat{\phi}$; *x*-axis) and essentiality log likelihood ratios (Ess. LLR; y-axis) from ACE based on genome-wide CRISPR-screen data from the Achilles DepMap Project. Results are shown for 688 'essential' genes (Pos. Controls; *red*) identified in references (1–3) and 634 nonessential genes (Neg. Controls; *blue*; see Methods for details). The LLR metrics are relative to a null hypothesis of $\hat{\phi} = 0$. (b) Estimated essentiality per gene in wild-type *TP53* (ϕ; *x*-axis) vs. log likelihood ratio (LLR) for differential essentiality in wild-type vs. mutant *TP53* based on DepMap data for NSCLC Adenocarcinoma (9). Several high-scoring genes are highlighted with labels, including *MDM2* (see text). (c) Top 10 genes identified as differentially essential in NSCLC samples having wildtype *TP53*, ordered by differential essentiality log likelihood ratio (Diff. Ess. LLR). (d) Top 10 genes identified as differentially essential in NSCLC samples having mutant *KRAS*. In panels (b)–(d), the horizontal dashed lines indicate Bonferroni-corrected empirical p-values of 0.05. Genes highlighted in red in panels (c) and (d) are discussed in the text.

In the case of *TP53*, we first defined the test samples as ones with 'wildtype' copies of *TP53*, with no annotated mutations, and the control samples as ones in which *TP53* has 'damaging' mutations (denoted here as Δ*TP53*; see Methods). This test for a wildtype-dependent increase in essentiality yielded a clear outlier: the gene *MDM2* (Fig. 4b and (c)). It is well known that cell lines with wildtype *TP53* require an intact copy of *MDM2* to negatively regulate the TP53 protein and allow for cell proliferation (32, 33). This result therefore provides support for the biological validity of our test for differential essentiality.

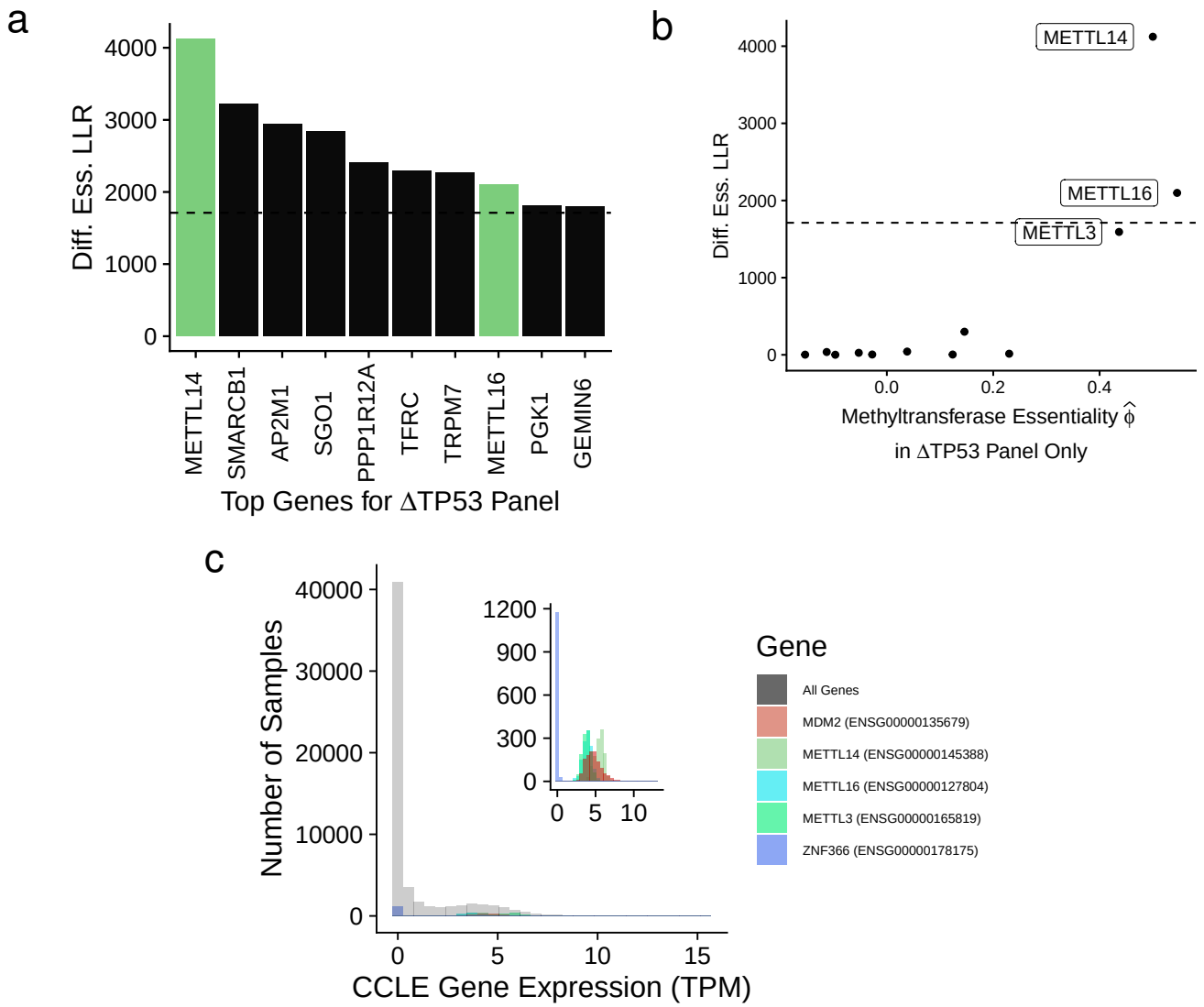
In the case of *KRAS*, which becomes oncogenic in the presence of gain-of-function mutations and is

known to drive aggressive tumorigenesis (34), we defined the samples in the opposite way: with test samples containing a missense mutation in *KRAS* (Δ*KRAS*), the majority of which are activating gain-of-function mutations (35), and control samples with no annotated SNP (WT*KRAS*). Samples containing wildtype *TP53* were additionally excluded to remove the signal from *MDM2*. As predicted, the *KRAS* gene itself emerged as having, by far, the strongest signal for a mutation-dependent increase in essentiality (Fig. 4d), further validating our method.

Test for Non-Oncogene Addictions in* TP53 *Mutants. Finally, we swapped the 'test' and 'control' samples for *TP53* — using Δ*TP53* samples for the 'test' and wildtype samples for the 'control' — with the goal of discovering novel non-oncogene addictions. As discussed above, such addictions may suggest new drug targets in cases of undruggable loss-of-function mutations. This test of increased essentiality in the presence of damaging *TP53* mutations identified several candidate genes (Fig. 5a and Supplementary Fig. S6).

Notably, the top candidates included two genes encoding RNA m^6-A methyltransferases, *METTL14*, and *METTL16*. A third gene, *METTL3*, encoding the heterodimeric partner of *METTL14*, also scored highly for differential essentiality (Fig. 5b). These three genes were assigned moderately high scores for differential essentiality according to the CERES method from the Achilles Project, but CERES did not rank any of them among its top 100 differentially essential genes (Fig. 5d). This difference in ranking between ACE and CERES possibly reflects ACE's improved power for detecting differential essentiality. All three methyltransferases are expressed at a moderate to high level across the *TP53*-mutant cell line panel, according to CCLE expression data (Fig. 5c and Supplementary Fig. S7) (24, 30, 31).

Several studies have noted CRISPR sgRNA cutting in multiple regions can reduce cell proliferation, regardless of the essentiality of the sites targeted. To ensure that the signatures of essentiality in *METTL3*, *METTL14*, and *METTL16* are not merely reflective of changes in copy number, we compared gene duplications between mutant and wildtype *TP53* sample panels (Supplementary Fig. S8). Copy number variation was similar between the two panels, suggesting that differences in gene essentiality are driven by biological activity of the genes themselves. In addition, flanking genomic regions of *METTL14* and *METTL16* do not contain known oncogenic drivers, eliminating the possibility that CRISPR-induced lesions are simultaneously influencing a nearby essential gene (Supplementary Fig. S9). Overall, our findings suggest that *METTL3*, *METTL14*, and *METTL16* are promising candidates for non-oncogene addiction in the absence of



d

Gene	ACE Essentiality in ΔTP53 Cell Lines	ACE p-Value (Empirical)	CERES Essentiality (Rescaled for Comparison)	CERES p-Value (Student's T-test; FDR adj.)	CERES Rank	ACE Rank
MDM2	0.598*	0.000	0.113	0.000	105*	1*
ZNF366	0.023	0.901	0.002	0.109	11129	16135
METTL3	0.437	0.001	0.051	0.000	923	14
METTL14	0.501	0.00	0.059	0.015	2456	1
METTL16	0.546	0.001	0.102	0.088	816	8

Fig. 5. RNA Methyltransferase Genes Associated with *TP53*. (a) Top 10 genes (by differential essentiality log-likelihood ratio) showing increased essentiality in NSCLC cell lines that harbor 'damaging' mutations in *TP53* (Δ*TP53*) according to the CCLE, suggesting possible non-oncogene addictions. The RNA m^6-A methyltransferases *METTL14* and *METTL16* (*green*) are of particular interest (see text). (b) Estimated essentiality of RNA methyltransferase genes in Δ*TP53* samples (*x*-axis) vs. log likelihood ratio in Δ*TP53* vs. wild-type *TP53* samples (*y*-axis). (c) Gene expression data from the CCLE, processed by the Achilles DepMap, for the Δ*TP53* cell line panel highlighting the differentially essential methyltransferase genes from panel (b). *MDM2* is shown as a positive control and nonessential *ZNF366* is shown as a negative control. The distribution of all genes is shown in gray. (d) Comparison of differential essentiality candidates and controls based on ACE and the CERES method from Project Achilles (see Methods for details). An asterisk indicates a result applies for values associated with *MDM2* only in the presence of wildtype *TP53*; other genes were tested for essentiality specific to cell lines with a damaging mutation in *TP53* (ΔTP53). In panels (a) and (b), the horizontal dashed lines indicate Bonferroni-corrected empirical p-values of 0.05.

functional *TP53*.

Discussion

In this paper, we have introduced a new probabilistic modeling and inference method, called ACE, for detecting essential genes and estimating levels of essentiality from CRISPR negative selection screens. As we have shown, ACE uses a hierarchical Bayesian model to account for the uncertainty associated with each major stage of a CRISPR screen and enable maximum-likelihood estimation of gene-level essentiality. It also supports likelihood ratio tests for both absolute and differential essentiality.

While the number of known essential genes is growing, simulations remain critical in providing an objective gold standard to evaluate the performance of computational methods. To test ACE and compare it with other available methods, we developed a new simulator, called empiriCRISPR. Our strategy with empiriCRISPR was not to generate simulated data according to a process-based model — which would tend not to account for all of the sources of noise in real data, and would be difficult to define without heavily overlapping the assumptions of ACE itself — but instead to mimic the empirical properties of real data sets as closely as possible. In this way, we were able to generate 'authentic-looking' data sets of arbitrary size with known sets of essential genes at known levels of essentiality. We showed that ACE performs quite well on these data sets despite their realistic 'noise' characteristics.

To our knowledge, ACE is the first method to support a direct likelihood-based test for differential essentiality between designated 'test' and 'control' sample panels. This test can be useful in a variety of applications, including tests for differences in essentiality in the presence of mutant and wild-type versions of known cancer genes. In comparison to indirect tests for differential essentiality — derived from separate analyses of 'test' and 'control' panels — ACE appears to offer improved power according to our simulation study (see Fig. 3). Notably, it is straightforward to define one-sided variants of this test to detect *increased* or *decreased* essentiality in the 'test' relative to 'control' panels. The framework also extends naturally to tests of more complex hypotheses (e.g., test essentiality > 0 and control essentiality ≤ 0). According to standard statistical theory, these will all be well-powered tests, provided they are applied to sufficiently abundant data for the asymptotic properties of likelihood ratio tests to become apparent.

As illustrated by the major revisions of 'core essential gene' sets over the past 10 years, the definition of 'essential genes' is inevitably somewhat subjective and arbitrary. In reality, essentiality occurs on a spectrum

and depends on the particular genetic background, cell-type, and conditions tested. As a result, candidate 'essential' genes require carefully designed experiments for validation, and researchers should be cautious about relying upon published sets of genes labeled 'essential'. Estimates of ϕ_G in ACE are also limited, in that they reflect an 'average' essentiality across all samples analyzed, so users must take care to choose a panel of samples that ensures a level of diversity appropriate for the biological question of interest. For example, a diverse collection of tissue samples or cell lines may be most suitable for the identification of universally essential genes, whereas a set of patient-derived tissue samples may be best for personalized tumor therapeutics. Notably, because it is a relative metric, differential essentiality may in some cases generalize better than absolute essentiality. We anticipate that findings of differential essentiality may be of interest to researchers exploring genetic vulnerabilities even in other tissues and cancer subtypes.

Among the genotype-dependent differentially essential genes identified by ACE, we found the RNA methyltransferase genes *METTL3*, *METTL14*, and *METTL16* to be particularly interesting. RNA m^6-A modifications have recently been found to be critical in the maintenance of leukemic cell growth (36). These genes have been previously implicated in other cancer types, and RNA m^6-A methyltransferases in general have been a topic of recent interest in the new field of cancer epitranscriptomics, with several phase I trials targeting METTL3 due to begin in 2021 (21). METTL3, in particular, was recently identified as interacting with *TP53* mRNA and inducing drug resistance in a biochemical assay in SW48 colon cancer cells (37). Another study found *METTL3* expression to be essential in NSCLC adenocarcinoma, although only two cell types were examined (38). *METTL3* mRNA knockdown with siRNA has also been shown, by M6A-seq, to result in an enrichment of methylation changes in genes in the p53-signaling pathway (39). Finally, patients with genetic alterations of m6A regulatory genes have been found to have an increase in *TP53* mutations, albeit only with sample sizes of $n = 5$ (40). Thus, there is abundant indirect evidence suggesting that these genes deserve further investigation as possible candidates for addiction in the absence of functional *TP53*. More generally, we anticipate that improved tests for differential essentiality, such as the ones introduced here, combined with rapidly growing CRISPR-screen data sets, will enable the detection of many new cases of genotype-dependent differences in essentiality, including both synthetic lethality and non-oncogene addiction.

Methods

Initial Infection. The initial abundance n_{sg} of each sgRNA g within sample s is assumed to be given by a Poisson distribution with mean $\mu_1 = c_s m_g$, where c_s is the total number of cells infected in the screen and m_g is the relative frequency of g in the master library:

$$P\Big(n_{sg}|c_s, m_g\Big) = \text{Pois}\Big(n_{sg}|\mu_1 = c_s m_g\Big). \tag{1}$$

The value m_g for each sgRNA g is estimated as the fraction of sequence reads corresponding to g in the sequence data for the master library. If multiple replicates are available, an average is taken. The parameter c_s is either provided by the user or a default value of 1000 is assumed.

Initial Read Counts. The sgRNA sequencing counts x_{sg} from the initial infected population n_{sg} are assumed to follow a Poisson distribution with mean $\mu_2 = \gamma_s n_{sg}$, where γ_s is a scaling parameter that captures the relationship between sequencing depth and the number of cells infected:

$$P\Big(x_{sg}|\gamma_s, n_{sg}\Big) = \text{Pois}\Big(x_{sg}|\mu_2 = \gamma_s n_{sg}\Big). \tag{2}$$

Final Read Counts. Following several rounds of cell division, the final abundance of each sgRNA g is assumed to be $d_{sg} = n_{sg}(1 - \epsilon_g \phi_G)$, where ϕ_G is the essentiality of the gene G that is targeted by g, and ϵ_g is the efficiency of g. This process is assumed to be deterministic; we assume that the number of infected cells is sufficiently large and sufficiently many cell divisions have occurred that stochastic effects can be ignored, and uncertainty in the value of d_{sg} can absorbed by the distributions for n_{sg} and y_{sg}. The final read counts y_{sg} are then assumed to follow a Poisson distribution with mean $\mu_3 = \gamma'_s d_{sg}$, where γ'_s is the final read count scaling factor:

$$\begin{aligned} P\Big(y_{sg}|d_{sg}, \gamma'_s, \epsilon_g \phi_G\Big) &= \text{Pois}\Big(y_{sg}|\mu_3 = \gamma'_s d_{sg}\Big) \\ &= \text{Pois}\Big(y_{sg}|\mu_3 = \gamma'_s n_{sg}(1 - \epsilon_g \phi_G)\Big). \end{aligned} \tag{3}$$

The scaling factor γ'_s not only captures the relationship between sequencing depth and the number of cells that remain after cell growth, but also can accommodate sample-specific differences in the growth rate of all

cells (independent of sgRNA).

Guide-Specific Effects. The sgRNA-specific efficiency ϵ_g is derived from a general feature vector by logistic regression:

$$\epsilon_g = \frac{\exp(\omega \cdot F_g)}{1 + \exp(\omega \cdot F_g)}, \tag{4}$$

where F_g is a feature vector and ω is a corresponding vector of real-valued coefficients. In this paper, we used a ten-dimensional vector F_g whose elements are indicators (via "one-hot" encoding) for the decile of GC content of each sgRNA's template guide. The corresponding 10-dimensional weight vector ω was treated as a free parameter and estimated by maximum likelihood. In this way, we effectively estimate a separate efficiency value for each GC decile, based on pooled information from all guides within that decile. As noted in the text, users may instead wish to use richer feature vectors such as those described in other recent studies (25, 41–45).

Estimation of Sample Scaling Parameters. Following the 'median-of-ratios' approach used by DESeq2 (5), the sample scaling parameters γ_s and γ'_s are not treated as free parameters in the maximization of the likelihood, but instead are set in a preprocessing step based on the raw read counts x_{sg} and y_{sg} for negative controls. Specifically, the scaling parameter is set equal to the median of the ratios of each sgRNA's read count to a guide-specific reference value, which in this case is the expected number of infected cells, $c_s m_g$:

$$\gamma_s = \underset{g \in \mathrm{G_{neg}}}{\mathrm{median}}\Big(\frac{x_{sg}}{c_s m_g}\Big), \qquad \gamma'_s = \underset{g \in \mathrm{G_{neg}}}{\mathrm{median}}\Big(\frac{y_{sg}}{c_s m_g}\Big), \tag{5}$$

where $g \in G_{\text{neg}}$ indicates all sgRNAs targeting genes in the designated set of negative controls. If no master library is available, the initial abundance of sgRNA $\bar{g}$ is approximated from the initial read counts as the average across samples:

$$m_{\bar{g}} \approx \frac{1}{S} \sum_s \frac{x_{s\bar{g}}}{\sum_g x_{sg}}, \tag{6}$$

where S is the total number of samples. The user may choose to use the median ratio across all genes rather than only the negative controls, if desired.

Full Model Likelihood. Substituting from equations 1–3, the full likelihood function can be expressed as:

$$L(\phi_{\cdot},\omega_{\cdot};x_{\cdot\cdot},y_{\cdot\cdot},m_{\cdot},c_{\cdot},\gamma_{\cdot},\gamma'_{\cdot}) = \prod_{G}\prod_{g\in G}\prod_{s}\sum_{n_{sg}=0}^{c_s}\text{Pois}\Big(n_{sg}|c_s m_g\Big)\text{Pois}\Big(x_{sg}|\gamma_s n_{sg}\Big) \times \text{Pois}\Big(y_{sg}|\gamma'_s n_{sg}(1-\epsilon_g\phi_G)\Big) \tag{7}$$

where ϵ_g is implicitly a function of ω as defined by equation 4. Here and below we use "dot" notation to indicate sets of all relevant indices (e.g., $x_{\cdot\cdot} = \{x_{sg} \mid \text{for all samples } s \text{ and sgRNAs } g\}$).

The likelihood function must sum over the number of infected cells n_{sg}, which cannot be directly observed. This summation step can be accelerated by an approximation that bundles summands together, for example, into groups of 10 cells. Weight vector ω is initialized to treat all sgRNA as 100% effective for the first optimization of ϕ_G.

When ω is held fixed, the full likelihood decomposes by gene, such that each ϕ_G can be estimated separately; and similarly, when the ϕ_G values are held fixed, the likelihood decomposes by the efficiency features f, owing to our "one-hot" design for the feature vector. Thus, we estimate each ϕ_G and component of ω_f iteratively, to enable parallelization by gene and feature category. Specifically, we repeatedly carry out the following optimizations until convergence (for each gene G and feature f, respectively):

$$\widehat{\phi_G} = \underset{\phi_G}{\text{argmax}}\Big\{L(\phi_G,\widehat{\phi_{\neg G}},\widehat{\omega};x_{\cdot\cdot},y_{\cdot\cdot},\theta)\Big\} \tag{8}$$
$$\widehat{\omega_f} = \underset{\omega_f}{\text{argmax}}\Big\{L(\widehat{\phi_{\cdot}},\omega_f,\widehat{\omega_{\neg f}};x_{\cdot\cdot},y_{\cdot\cdot},\theta)\Big\}$$

where $\phi_{\neg G}$ denotes the set $\{\phi_g \mid g \neq G\}$, $\omega_{\neg f}$ denotes the set $\{\omega_{f'} \mid f' \neq f\}$, and θ denotes the remaining parameter set.

Finally, to evaluate whether an estimated value of $\widehat{\phi_G}$ is significantly different from zero, we compare the component of the maximized likelihood function for that gene to the corresponding component when ϕ_G is held fixed at zero. Specifically, we compute a test statistic T as,

$$T = \ln\frac{L(\widehat{\phi_G})}{L(\phi_G = 0)}, \tag{9}$$

where $L(\phi_G)$ represents the component of the likelihood corresponding to gene G. We compute an empirical p-value for T based on the equivalent test statistics for the designated negative control genes.

Adaptation for Alternative Experimental Designs. The likelihood function can easily be adapted to settings in which data is unavailable for either the master library or the initial sgRNA abundances. In the case where the master library information is missing, we simply use a uniform prior for values of n_{sg}, $P(n_{sg}|c_s) = \frac{1}{c_s}$. When the initial infected sgRNA abundances are unavailable, we remove the corresponding portion of the likelihood function, effectively treating $x_{..}$ as missing data:

$$L(\phi_.,\omega_.; x_{..} = \varnothing, y_{..}, m_., c_., \gamma_., \gamma'_.) = \prod_G \prod_{g \in G} \prod_s \sum_{n_{sg}=0}^{c_s} \text{Pois}\Big(n_{sg}|c_s m_g\Big)\text{Pois}\Big(y_{sg}|\gamma'_s n_{sg}(1-\epsilon_g \phi_G)\Big) \quad \textbf{(10)}$$

Test for Differential Essentiality. To identify genes with significant differences in essentiality between a 'test' set of samples, S_t, and a 'control' set, S_c, we employ the following likelihood ratio test. Let $L(\phi_G)$ represent the component of the likelihood function relevant to gene G that has been maximized in the standard way, with one version of the ϕ_G parameter shared across both S_t and S_c. By contrast, let $L(\phi_G^t)L(\phi_G^c)$ represent the corresponding component of the likelihood optimized such that separate versions of the essentiality parameter, ϕ_G^t and ϕ_G^c, are used for samples S_t and S_c, respectively. The likelihood ratio test statistic $T'$is then given by,

$$T' = \ln \frac{L(\phi_G^t)L(\phi_G^c)}{L(\phi_G)}. \quad \textbf{(11)}$$

For this test, all samples are used to calculate the sample-specific parameters γ_s and γ'_s, and the sgRNA-specific parameter ϵ_g; only ϕ_G^t and ϕ_G^c are re-estimated to compute T'. As with the test for absolute essentiality, an empirical p-value is computed for T' based on the distribution of values computed for the negative controls. In some analyses, we further require that $\phi_G^t > \phi_G^c$ and otherwise set T' to 0.

Essential and Nonessential Controls. The set of essential genes used for validation was defined as those genes that were included in at least two of three previously published sets of essential genes (1–3). The set of nonessential genes was obtained from 604 total RNA-seq datasets for 128 cell lines, in vitro differentiated cells, primary cells, and tissues accessed from the Encyclopedia of DNA Elements (ENCODE) data portal on August 29, 2019 (see Supplemental Data Table 1 for accession info of all cell lines used). For the nonessential set, we chose genes expressed at ≤ 1 TPM in all cell lines.

Data from Cancer Dependency Map (DepMap). For our analysis of real data from the Cancer Dependency Map (DepMap) Consortium, we focused on genome-wide CRISPR-screen data from Project Achilles,

produced by the Broad Institute (9). This dataset uses the Avana library and includes 1374 publicly available screens at time of writing. Only the pre-infection master library and the post-depletion sgRNA abundances were sequenced, with one to four replicates per cell line available.

To retain a similar genetic background between sample panels, samples were compared from the same tissue type and lineage, specifically cell lines originating from lung Non-Small Cell Lung Cancer (NSCLC) Adenocarcinoma. The Achilles Project dataset includes 220 replicates of 92 unique samples of these cell lines. Following the Achilles DepMap convention (46), we assume a gene has lost function (indicated as Δ) in all CCLE cell lines annotated as containing mutations that introduce single nucleotide polymorphisms (SNPs), deletions, or insertions in the start codon; deletions or insertions that introduce a frameshift mutation; mutations that introduce a premature stop codon or a *de novo*, frameshifted start codon; or mutations that disrupt a splice site (30, 31).

ACE results were compared with CERES published p-values available for each Achilles DepMap sample. As in ref. (9), the CERES differential essentiality test statistic was calculated using a one-sided Student's t-test to compare the by-sample p-values of the test and control sample panels, and then was corrected by the rank-based false discovery rate.

empiriCRISPR Simulation of CRISPR Knockout Screens. empiriCRISPR simulates each gene using a series of gamma distributions to represent each stage of a CRISPR knockout screen (infection, initial and final sequencing). CRISPR knockout screens from (16) were used as the empiriCRISPR template, and initial master library abundances were selected by sampling the template master sgRNA library. empiriCRISPR adjusts the gamma distribution parameters by minimizing the adjusted mean squared error of the summary statistics of simulated data with the median of 5 summary statistics of the template screens, shown in Fig. S3. The abundance of each sgRNA construct was assumed to be independent from all other sgRNA used in the experiment. Unless otherwise noted, simulations used 300 genes per essentiality value, 4 sgRNA per gene, and 3 samples.

Package Software. ACE is implemented as an R package (R 3.5.0, (47)) using the *data.table* (48) and *R6* (49) packages. Optimization is performed using R's *optim* package with the 'Brent' method. Both the ACE and empiriCRISPR github repositories can be found at https://github.com/CshlSiepelLab.

Estimation of Essentiality from Mean Log-Fold Change. The 'average fold change' ('AFC') score for each gene G was calculated as follows:

$$\text{AFC} = 1 - \prod_{g=1}^{g \in G} \prod_{s=1}^{S} \left(\lambda_s \tau_{neg} \frac{(0.5 + y_{sg})}{(0.5 + x_{sg})} \right)^{\frac{1}{SG}}, \qquad \tau_{neg}^{-1} = \prod_{g=1}^{g \in G_{neg}} \prod_{s=1}^{S} \left(\lambda_s \frac{(0.5 + y_{sg})}{(0.5 + x_{sg})} \right)^{\frac{1}{SG_{neg}}}, \qquad \textbf{(12)}$$

where a pseudocount of 0.5 was added to both initial (x_{sg}) and final (y_{sg}) read counts to prevent undefined values. Each sample was normalized by λ_s to bring the total read count of each sample to a total of 10 million reads, and final fold changes were adjusted according to τ_{neg}, the mean fold change in all negative controls G_{neg}. Consistent with our ACE notation, an AFC value of zero has no impact on cell growth or proliferation.

Estimation of Essentiality using BAGEL. We applied the BAGEL (Bayesian Analysis of Gene Essentiality) method (1) using software version 0.91 (last modified 09/2015). Simulated data was analyzed using 4% of genes as positive and negative controls, with simulated essentiality scores of 0.99 (99% depletion) and 0.05 (5% depletion) respectively. Differential essentiality was calculated as the difference of Bayes Factors between separate analysis of test and control sample panels, as performed in (1). For the analysis of real data, we used all genes identified as essential or nonessential in our earlier analysis.

Estimation of Essentiality using JACKS. We applied the JACKS (Joint Analysis of CRISPR/Cas9 Knock-out Screens) method (8) version 0.2 (downloaded 11/2019). Differential essentiality was evaluated by running JACKS separately for the test and control sample panels, using the built-in hierarchical prior option and the same negative control genes provided to other methods for both real and simulated data. The test statistic for differential essentiality was indicated by the absolute difference in p-value between test and control panels, as used in (8).

Acknowledgments

We thank Professors Molly Hammell and Justin Kinney for advising on this project, Professor Yi-fei Huang for providing feedback on the mathematical design, and members of the Vakoc and Siepel labs for valuable advice and suggestions. Funding was provided by the Watson School of Biological Sciences William R. Miller Fellowship (to ERH), and the US National Institutes of Health grant R35-GM127070 (to AS). The

content is solely the responsibility of the authors and does not necessarily represent the official views of the US National Institutes of Health.

Bibliography

1. Traver Hart and Jason Moffat. BAGEL: a computational framework for identifying essential genes from pooled library screens. *BMC Bioinformatics*, 17(1):164, 2016. ISSN 1471-2105. doi: 10.1186/s12859-016-1015-8.
2. Traver Hart, Megha Chandrashekhar, Michael Aregger, Zachary Steinhart, Kevin R. Brown, Graham MacLeod, Monika Mis, Michal Zimmermann, Amelie Fradet-Turcotte, Song Sun, Patricia Mero, Peter Dirks, Sachdev Sidhu, Frederick P. Roth, Olivia S. Rissland, Daniel Durocher, Stephane Angers, and Jason Moffat. High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities. *Cell*, 163(6):1515–1526, 2015. ISSN 10974172. doi: 10.1016/j.cell.2015.11.015.
3. Traver Hart, Amy Hin Yan Tong, Katie Chan, Jolanda Van Leeuwen, Ashwin Seetharaman, Michael Aregger, Megha Chandrashekhar, Nicole Hustedt, Sahil Seth, Avery Noonan, Andrea Habsid, Olga Sizova, Lyudmila Nedyalkova, Ryan Climie, Leanne Tworzyanski, Keith Lawson, Maria Augusta Sartori, Sabriyeh Alibeh, David Tieu, Sanna Masud, Patricia Mero, Alexander Weiss, Kevin R. Brown, Matej Usaj, Maximilian Billmann, Mahfuzur Rahman, Michael Constanzo, Chad L. Myers, Brenda J. Andrews, Charles Boone, Daniel Durocher, and Jason Moffat. Evaluation and Design of Genome-Wide CRISPR/SpCas9 Knockout Screens. *G3:Genes|Genomes|Genetics*, 7(8):2719–2727, 2017. ISSN 2160-1836. doi: 10.1534/g3.117.041277.
4. Wei Li, Han Xu, Tengfei Xiao, Le Cong, Michael I Love, Feng Zhang, Rafael a Irizarry, Jun S Liu, Myles Brown, and X Shirley Liu. MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biology*, 15(12):554, 2014. ISSN 1465-6906. doi: 10.1186/s13059-014-0554-4.
5. Michael I Love, Wolfgang Huber, and Simon Anders. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12):550, 2014. ISSN 1474-760X. doi: 10.1186/s13059-014-0550-8.
6. Zhiyin Dai, Julie M. Sheridan, Linden J. Gearing, Darcy L. Moore, Shian Su, Sam Wormald, Stephen Wilcox, Liam O'Connor, Ross A. Dickins, Marnie E. Blewitt, and Matthew E. Ritchie. edgeR: a versatile tool for the analysis of shRNA-seq and CRISPR-Cas9 genetic screens. *F1000Research*, 3:95, oct 2014. ISSN 2046-1402. doi: 10.12688/f1000research.3928.2.
7. Jungsik Noh and Beibei Chen. sgRSEA: Enrichment Analysis of CRISPR/Cas9 Knockout Screen Data, 2015.
8. Felicity Allen, Fiona Behan, Anton Khodak, Francesco Iorio, Kosuke Yusa, Mathew Garnett, and Leopold Parts. JACKS: joint analysis of CRISPR/Cas9 knockout screens. *Genome research*, 29(3): 464–471, jan 2019. ISSN 1549-5469. doi: 10.1101/gr.238923.118.
9. Robin M. Meyers, Jordan G. Bryan, James M. McFarland, Barbara A. Weir, Ann E. Sizemore, Han Xu, Neekesh V. Dharia, Phillip G. Montgomery, Glenn S. Cowley, Sasha Pantel, Amy Goodale, Yenarae Lee, Levi D. Ali, Guozhi Jiang, Rakela Lubonja, William F. Harrington, Matthew Strickland, Ting Wu, Derek C. Hawes, Victor A. Zhivich, Meghan R. Wyatt, Zohra Kalani, Jaime J. Chang, Michael Okamoto, Kimberly Stegmaier, Todd R. Golub, Jesse S. Boehm, Francisca Vazquez, David E. Root, William C. Hahn, and Aviad Tsherniak. Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. *Nature Genetics*, 49(12):1779–1784, 2017. ISSN 15461718. doi: 10.1038/ng.3984.
10. Jiyang Yu, Jose Silva, and Andrea Califano. ScreenBEAM: A novel meta-analysis algorithm for functional genomics screens via Bayesian hierarchical modeling. *Bioinformatics*, 32(2):260–267, 2015. ISSN 14602059. doi: 10.1093/bioinformatics/btv556.
11. David W Morgens, Richard M Deans, Amy Li, and Michael C Bassik. Systematic comparison of CRISPR/Cas9 and RNAi screens for essential genes SUPPLEMENT. *Nature Biotechnology*, 34(6): 634–636, 2016. ISSN 1087-0156. doi: 10.1038/nbt.3567.
12. Bastiaan Evers, Katarzyna Jastrzebski, Jeroen P.M. Heijmans, Wipawadee Grernrum, Roderick L. Beijersbergen, and Rene Bernards. CRISPR knockout screening outperforms shRNA and CRISPRi in identifying essential genes. *Nature Biotechnology*, 34(6):631–633, 2016. ISSN 15461696. doi: 10.1038/nbt.3536.
13. Fumitaka Inoue, Martin Kircher, Beth Martin, Gregory M. Cooper, Daniela M. Witten, Michael T. McManus, Nadav Ahituv, and Jay Shendure. A systematic comparison reveals substantial differences in chromosomal versus episomal encoding of enhancer activity. *Genome Research*, 27(1):38–52, jan 2017. ISSN 15495469. doi: 10.1101/gr.212092.116.
14. T. Wang, K. Birsoy, N. W. Hughes, K. M. Krupczak, Y. Post, J. J. Wei, E. S. Lander, and D. M. Sabatini. Identification and characterization of essential genes in the human genome. *Science*, 350(6264): 1096–1101, 2015. ISSN 0036-8075. doi: 10.1126/science.aac7041.
15. Tim Wang, Haiyan Yu, Nicholas W Hughes, Walter W Chen, Eric S Lander, and David M Sabatini. Gene Essentiality Profiling Reveals Gene Networks and Synthetic Lethal Interactions with Oncogenic Ras. *Cell*, 168(5):1–14, 2017. ISSN 0092-8674. doi: 10.1016/j.cell.2017.01.013.
16. Timothy D. Martin, Danielle R. Cook, Mei Yuk Choi, Mamie Z. Li, Kevin M. Haigis, and Stephen J. Elledge. A Role for Mitochondrial Translation in Promotion of Viability in K-Ras Mutant Cells. *Cell Reports*, 20(2):427–438, 2017. ISSN 22111247. doi: 10.1016/j.celrep.2017.06.061.
17. Yusuke Tarumoto, Bin Lu, Tim D.D. Somerville, Yu Han Huang, Joseph P. Milazzo, Xiaoli S. Wu, Olaf Klingbeil, Osama El Demerdash, Junwei Shi, and Christopher R. Vakoc. LKB1, Salt-Inducible Kinases, and MEF2C Are Linked Dependencies in Acute Myeloid Leukemia. *Molecular Cell*, 69(6):1017–1027.e6, 2018. ISSN 10974164. doi: 10.1016/j.molcel.2018.02.011.
18. Andrew J. Aguirre, Robin M. Meyers, Barbara A. Weir, Francisca Vazquez, Cheng Zhong Zhang, Uri Ben-David, April Cook, Gavin Ha, William F. Harrington, Mihir B. Doshi, Maria Kost-Alimova, Stanley Gill, Han Xu, Levi D. Ali, Guozhi Jiang, Sasha Pantel, Yenarae Lee, Amy Goodale, Andrew D. Cherniack, Coyin Oh, Gregory Kryukov, Glenn S. Cowley, Levi A. Garraway, Kimberly Stegmaier, Charles W. Roberts, Todd R. Golub, Matthew Meyerson, David E. Root, Aviad Tsherniak, and William C. Hahn. Genomic copy number dictates a gene-independent cell response to CRISPR/Cas9 targeting. *Cancer Discovery*, 6(8):914–929, 2016. ISSN 21598290. doi: 10.1158/2159-8290.CD-16-0154.
19. Konstantinos Tzelepis, Hiroko Koike-Yusa, Etienne De Braekeleer, Yilong Li, Emmanouil Metzakopian, Oliver M. Dovey, Annalisa Mupo, Vera Grinkevich, Meng Li, Milena Mazan, Malgorzata Gozdecka, Shuhei Ohnishi, Jonathan Cooper, Miten Patel, Thomas McKerrell, Bin Chen, Ana Filipa Domingues, Paolo Gallipoli, Sarah Teichmann, Hannes Ponstingl, Ultan McDermott, Julio Saez-Rodriguez, Brian J.P. Huntly, Francesco Iorio, Cristina Pina, George S. Vassiliou, and Kosuke Yusa. A CRISPR Dropout Screen Identifies Genetic Vulnerabilities and Therapeutic Targets in Acute Myeloid Leukemia.

Cell Reports, 17(4):1193–1205, 2016. ISSN 22111247. doi: 10.1016/j.celrep.2016.09.079.

20. Remco Nagel, Ekaterina A Semenova, and Anton Berns. Drugging the addict: non-oncogene addiction as a target for cancer therapy. *EMBO reports*, 17(11):1516–1531, nov 2016. ISSN 1469-221X. doi: 10.15252/embr.201643030.
21. Megan Cully. Chemical inhibitors make their RNA epigenetic mark, nov 2019. ISSN 14741784.
22. Gaoxiang Jia, Xinlei Wang, and Guanghua Xiao. A permutation-based non-parametric analysis of CRISPR screen data. *BMC Genomics*, 18(1):545, dec 2017. ISSN 1471-2164. doi: 10.1186/s12864-017-3938-5.
23. Philipp N. Spahn, Tyler Bath, Ryan J. Weiss, Jihoon Kim, Jeffrey D. Esko, Nathan E. Lewis, and Olivier Harismendy. PinAPL-Py: A comprehensive web-application for the analysis of CRISPR/Cas9 screens. *Scientific Reports*, 7(1), dec 2017. ISSN 20452322. doi: 10.1038/s41598-017-16193-9.
24. Juliana Leung. The Cancer Dependency Map Consortium, 2019.
25. John G Doench, Nicolo Fusi, Meagan Sullender, Mudra Hegde, Emma W Vaimberg, Katherine F Donovan, Ian Smith, Zuzana Tothova, Craig Wilen, Robert Orchard, Herbert W Virgin, Jennifer Listgarten, and David E Root. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature Biotechnology*, 34(2):184–191, 2016. ISSN 1087-0156. doi: 10.1038/nbt.3437.
26. Luca Pinello, Matthew C. Canver, Megan D. Hoban, Stuart H. Orkin, Donald B. Kohn, Daniel E. Bauer, and Guo Cheng Yuan. Analyzing CRISPR genome-editing experiments with CRISPResso. *Nature Biotechnology*, 34(7):695–697, 2016. ISSN 15461696. doi: 10.1038/nbt.3583.
27. Martin Jinek, Fuguo Jiang, David W Taylor, Samuel H Sternberg, Emine Kaya, Enbo Ma, Carolin Anders, Michael Hauer, Kaihong Zhou, Steven Lin, Matias Kaplan, Anthony T Iavarone, Emmanuelle Charpentier, Eva Nogales, and Jennifer a Doudna. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science (New York, N.Y.)*, 343(6176):1247997, mar 2014. ISSN 1095-9203. doi: 10.1126/science.1247997.
28. Robert L. Unckless, Andrew G. Clark, and Philipp W. Messer. Evolution of Resistance Against CRISPR/Cas9 Gene Drive. *Genetics*, 205(2):827–841, feb 2017. ISSN 0016-6731. doi: 10.1534/GENETICS.116.197285.
29. Edwin H Yau, Indrasena Reddy Kummetha, Gianluigi Lichinchi, Rachel Tang, Yunlin Zhang, and Tariq M Rana. Genome-Wide CRISPR Screen for Essential Cell Growth Mediators in Mutant KRAS Colorectal Cancers. *Cancer research*, 77(22):6330–6339, sep 2017. ISSN 1538-7445. doi: 10.1158/0008-5472.CAN-17-2043.
30. Jordi Barretina, Giordano Caponigro, Nicolas Stransky, Kavitha Venkatesan, Adam A. Margolin, Sungjoon Kim, Christopher J. Wilson, Joseph Lehár, Gregory V. Kryukov, Dmitriy Sonkin, Anupama Reddy, Manway Liu, Lauren Murray, Michael F. Berger, John E. Monahan, Paula Morais, Jodi Meltzer, Adam Korejwa, Judit Jané-Valbuena, Felipa A. Mapa, Joseph Thibault, Eva Bric-Furlong, Pichai Raman, Aaron Shipway, Ingo H. Engels, Jill Cheng, Guoying K. Yu, Jianjun Yu, Peter Aspesi, Melanie De Silva, Kalpana Jagtap, Michael D. Jones, Li Wang, Charles Hatton, Emanuele Palescandolo, Supriya Gupta, Scott Mahan, Carrie Sougnez, Robert C. Onofrio, Ted Liefeld, Laura MacConaill, Wendy Winckler, Michael Reich, Nanxin Li, Jill P. Mesirov, Stacey B. Gabriel, Gad Getz, Kristin Ardlie, Vivien Chan, Vic E. Myer, Barbara L. Weber, Jeff Porter, Markus Warmuth, Peter Finan, Jennifer L. Harris, Matthew Meyerson, Todd R. Golub, Michael P. Morrissey, William R. Sellers, Robert Schlegel, and Levi A. Garraway. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*, 483(7391):603–607, mar 2012. ISSN 00280836. doi: 10.1038/nature11003.
31. Mahmoud Ghandi, Franklin W. Huang, Judit Jané-Valbuena, Gregory V. Kryukov, Christopher C. Lo, E. Robert McDonald, Jordi Barretina, Ellen T. Gelfand, Craig M. Bielski, Haoxin Li, Kevin Hu, Alexander Y. Andreev-Drakhlin, Jaegil Kim, Julian M. Hess, Brian J. Haas, François Aguet, Barbara A. Weir, Michael V. Rothberg, Brenton R. Paolella, Michael S. Lawrence, Rehan Akbani, Yiling Lu, Hong L. Tiv, Prafulla C. Gokhale, Antoine de Weck, Ali Amin Mansour, Coyin Oh, Juliann Shih, Kevin Hadi, Yanay Rosen, Jonathan Bistline, Kavitha Venkatesan, Anupama Reddy, Dmitriy Sonkin, Manway Liu, Joseph Lehar, Joshua M. Korn, Dale A. Porter, Michael D. Jones, Javad Golji, Giordano Caponigro, Jordan E. Taylor, Caitlin M. Dunning, Amanda L. Creech, Allison C. Warren, James M. McFarland, Mahdi Zamanighomi, Audrey Kauffmann, Nicolas Stransky, Marcin Imielinski, Yosef E. Maruvka, Andrew D. Cherniack, Aviad Tsherniak, Francisca Vazquez, Jacob D. Jaffe, Andrew A. Lane, David M. Weinstock, Cory M. Johannessen, Michael P. Morrissey, Frank Stegmeier, Robert Schlegel, William C. Hahn, Gad Getz, Gordon B. Mills, Jesse S. Boehm, Todd R. Golub, Levi A. Garraway, and William R. Sellers. Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature*, 569(7757):503–508, may 2019. ISSN 14764687. doi: 10.1038/s41586-019-1186-3.
32. Rebecca A. Frum and Steven R. Grossman. Mechanisms of mutant p53 stabilization in cancer. *Sub-Cellular Biochemistry*, 85:187–197, 2014. ISSN 03060225. doi: 10.1007/978-94-017-9211-0_10.
33. Martin F. Lavin and N. Gueven. The complexity of p53 stabilization and activation, jun 2006. ISSN 13509047.
34. Sebastian Mueller, Thomas Engleitner, Roman Maresch, Magdalena Zukowska, Sebastian Lange, Thorsten Kaltenbacher, Björn Konukiewitz, Rupert Öllinger, Maximilian Zwiebel, Alex Strong, Hsi Yu Yen, Ruby Banerjee, Sandra Louzada, Beiyuan Fu, Barbara Seidler, Juliana Götzfried, Kathleen Schuck, Zonera Hassan, Andreas Arbeiter, Nina Schönhuber, Sabine Klein, Christian Veltkamp, Mathias Friedrich, Lena Rad, Maxim Barenboim, Christoph Ziegenhain, Julia Hess, Oliver M. Dovey, Stefan Eser, Swati Parekh, Fernando Constantino-Casas, Jorge De La Rosa, Marta I. Sierra, Mario Fraga, Julia Mayerle, Günter Klöppel, Juan Cadiñanos, Pentao Liu, George Vassiliou, Wilko Weichert, Katja Steiger, Wolfgang Enard, Roland M. Schmid, Fengtang Yang, Kristian Unger, Günter Schneider, Ignacio Varela, Allan Bradley, DIeter Saur, and Roland Rad. Evolutionary routes and KRAS dosage define pancreatic cancer phenotypes. *Nature*, 554(7690):62–68, 2018. ISSN 14764687. doi: 10.1038/nature25459.
35. Francisco Sanchez-Vega, Marco Mina, Joshua Armenia, Walid K. Chatila, Augustin Luna, Konnor C. La, Sofia Dimitriadoy, David L. Liu, Havish S. Kantheti, Sadegh Saghafinia, Debyani Chakravarty, Foysal Daian, Qingsong Gao, Matthew H. Bailey, Wen Wei Liang, Steven M. Foltz, Ilya Shmulevich, Li Ding, Zachary Heins, Angelica Ochoa, Benjamin Gross, Jianjiong Gao, Hongxin Zhang, Ritika Kundra, Cyriac Kandoth, Istemi Bahceci, Leonard Dervishi, Ugur Dogrusoz, Wanding Zhou, Hui Shen, Peter W. Laird, Gregory P. Way, Casey S. Greene, Han Liang, Yonghong Xiao, Chen Wang, Antonio Iavarone, Alice H. Berger, Trever G. Bivona, Alexander J. Lazar, Gary D. Hammer, Thomas Giordano, Lawrence N. Kwong, Grant McArthur, Chenfei Huang, Aaron D. Tward, Mitchell J. Frederick, Frank McCormick, Matthew Meyerson, Samantha J. Caesar-Johnson, John A. Demchok, Ina Felau, Melpomeni Kasapi, Martin L. Ferguson, Carolyn M. Hutter, Heidi J. Sofia, Roy Tarnuzzer, Zhining Wang, Liming Yang, Jean C. Zenklusen, Jiashan (Julia) Zhang, Sudha Chudamani, Jia Liu, Laxmi Lolla, Rashi Naresh, Todd Pihl, Qiang Sun, Yunhu Wan, Ye Wu, Juok Cho, Timothy DeFreitas, Scott Frazer, Nils Gehlenborg, Gad Getz, David I. Heiman, Jaegil Kim, Michael S. Lawrence, Pei Lin, Sam Meier, Michael S. Noble, Gordon Saksena, Doug Voet, Hailei Zhang, Brady Bernard, Nyasha Chambwe, Varsha Dhankani, Theo Knijnenburg, Roger Kramer, Kalle Leinonen, Yuexin Liu, Michael Miller, Sheila Reynolds, Ilya Shmulevich, Vesteinn Thorsson, Wei Zhang, Rehan Akbani, Bradley M. Broom, Apurva M. Hegde, Zhenlin Ju, Rupa S. Kanchi, Anil Korkut, Jun Li, Han Liang, Shiyun Ling, Wenbin Liu, Yiling Lu, Gordon B. Mills, Kwok Shing Ng, Arvind Rao, Michael Ryan, Jing Wang, John N. Weinstein, Jiexin Zhang, Adam Abeshouse, Joshua Armenia, Debyani Chakravarty, Walid K. Chatila, Ino de Bruijn, Jianjiong Gao, Benjamin E. Gross, Zachary J. Heins, Ritika Kundra, Konnor La, Marc Ladanyi, Augustin

Luna, Moriah G. Nissan, Angelica Ochoa, Sarah M. Phillips, Ed Reznik, Francisco Sanchez-Vega, Chris Sander, Nikolaus Schultz, Robert Sheridan, S. Onur Sumer, Yichao Sun, Barry S. Taylor, Jioajiao Wang, Hongxin Zhang, Pavana Anur, Myron Peto, Paul Spellman, Christopher Benz, Joshua M. Stuart, Christopher K. Wong, Christina Yau, D. Neil Hayes, Joel S. Parker, Matthew D. Wilkerson, Adrian Ally, Miruna Balasundaram, Reanne Bowlby, Denise Brooks, Rebecca Carlsen, Eric Chuah, Noreen Dhalla, Robert Holt, Steven J.M. Jones, Katayoon Kasaian, Darlene Lee, Yussanne Ma, Marco A. Marra, Michael Mayo, Richard A. Moore, Andrew J. Mungall, Karen Mungall, A. Gordon Robertson, Sara Sadeghi, Jacqueline E. Schein, Payal Sipahimalani, Angela Tam, Nina Thiessen, Kane Tse, Tina Wong, Ashton C. Berger, Rameen Beroukhim, Andrew D. Cherniack, Carrie Cibulskis, Stacey B. Gabriel, Galen F. Gao, Gavin Ha, Matthew Meyerson, Steven E. Schumacher, Juliann Shih, Melanie H. Kucherlapati, Raju S. Kucherlapati, Stephen Baylin, Leslie Cope, Ludmila Danilova, Moiz S. Bootwalla, Phillip H. Lai, Dennis T. Maglinte, David J. Van Den Berg, Daniel J. Weisenberger, J. Todd Auman, Saianand Balu, Tom Bodenheimer, Cheng Fan, Katherine A. Hoadley, Alan P. Hoyle, Stuart R. Jefferys, Corbin D. Jones, Shaowu Meng, Piotr A. Mieczkowski, Lisle E. Mose, Amy H. Perou, Charles M. Perou, Jeffrey Roach, Yan Shi, Janae V. Simons, Tara Skelly, Matthew G. Soloway, Donghui Tan, Umadevi Veluvolu, Huihui Fan, Toshinori Hinoue, Peter W. Laird, Hui Shen, Wanding Zhou, Michelle Bellair, Kyle Chang, Kyle Covington, Chad J. Creighton, Huyen Dinh, Harsha Vardhan Doddapaneni, Lawrence A. Donehower, Jennifer Drummond, Richard A. Gibbs, Robert Glenn, Walker Hale, Yi Han, Jianhong Hu, Viktoriya Korchina, Sandra Lee, Lora Lewis, Wei Li, Xiuping Liu, Margaret Morgan, Donna Morton, Donna Muzny, Jireh Santibanez, Margi Sheth, Eve Shinbrot, Linghua Wang, Min Wang, David A. Wheeler, Liu Xi, Fengmei Zhao, Julian Hess, Elizabeth L. Appelbaum, Matthew Bailey, Matthew G. Cordes, Li Ding, Catrina C. Fronick, Lucinda A. Fulton, Robert S. Fulton, Cyriac Kandoth, Elaine R. Mardis, Michael D. McLellan, Christopher A. Miller, Heather K. Schmidt, Richard K. Wilson, Daniel Crain, Erin Curley, Johanna Gardner, Kevin Lau, David Mallery, Scott Morris, Joseph Paulauskis, Robert Penny, Candace Shelton, Troy Shelton, Mark Sherman, Eric Thompson, Peggy Yena, Jay Bowen, Julie M. Gastier-Foster, Mark Gerken, Kristen M. Leraas, Tara M. Lichtenberg, Nilsa C. Ramirez, Lisa Wise, Erik Zmuda, Niall Corcoran, Tony Costello, Christopher Hovens, Andre L. Carvalho, Ana C. de Carvalho, José H. Fregnani, Adhemar Longatto-Filho, Rui M. Reis, Cristovam Scapulatempo-Neto, Henrique C.S. Silveira, Daniel O. Vidal, Andrew Burnette, Jennifer Eschbacher, Beth Hermes, Ardene Noss, Rosy Singh, Matthew L. Anderson, Patricia D. Castro, Michael Ittmann, David Huntsman, Bernard Kohl, Xuan Le, Richard Thorp, Chris Andry, Elizabeth R. Duffy, Vladimir Lyadov, Oxana Paklina, Galiya Setdikova, Alexey Shabunin, Mikhail Tavobilov, Christopher McPherson, Ronald Warnick, Ross Berkowitz, Daniel Cramer, Colleen Feltmate, Neil Horowitz, Adam Kibel, Michael Muto, Chandrajit P. Raut, Andrei Malykh, Jill S. Barnholtz-Sloan, Wendi Barrett, Karen Devine, Jordonna Fulop, Quinn T. Ostrom, Kristen Shimmel, Yingli Wolinsky, Andrew E. Sloan, Agostino De Rose, Felice Giuliante, Marc Goodman, Beth Y. Karlan, Curt H. Hagedorn, John Eckman, Jodi Harr, Jerome Myers, Kelinda Tucker, Leigh Anne Zach, Brenda Deyarmin, Hai Hu, Leonid Kvecher, Caroline Larson, Richard J. Mural, Stella Somiari, Ales Vicha, Tomas Zelinka, Joseph Bennett, Mary Iacocca, Brenda Rabeno, Patricia Swanson, Mathieu Latour, Louis Lacombe, Bernard Têtu, Alain Bergeron, Mary McGraw, Susan M. Staugaitis, John Chabot, Hanina Hibshoosh, Antonia Sepulveda, Tao Su, Timothy Wang, Olga Potapova, Olga Voronina, Laurence Desjardins, Odette Mariani, Sergio Roman-Roman, Xavier Sastre, Marc Henri Stern, Feixiong Cheng, Sabina Signoretti, Andrew Berchuck, Darell Bigner, Eric Lipp, Jeffrey Marks, Shannon McCall, Roger McLendon, Angeles Secord, Alexis Sharp, Madhusmita Behera, Daniel J. Brat, Amy Chen, Keith Delman, Seth Force, Fadlo Khuri, Kelly Magliocca, Shishir Maithel, Jeffrey J. Olson, Taofeek Owonikoko, Alan Pickens, Suresh Ramalingam, Dong M. Shin, Gabriel Sica, Erwin G. Van Meir, Hongzheng Zhang, Wil Eijckenboom, Ad Gillis, Esther Korpershoek, Leendert Looijenga, Wolter Oosterhuis, Hans Stoop, Kim E. van Kessel, Ellen C. Zwarthoff, Chiara Calatozzolo, Lucia Cuppini, Stefania Cuzzubbo, Francesco DiMeco, Gaetano Finocchiaro, Luca Mattei, Alessandro Perin, Bianca Pollo, Chu Chen, John Houck, Pawadee Lohavanichbutr, Arndt Hartmann, Christine Stoehr, Robert Stoehr, Helge Taubert, Sven Wach, Bernd Wullich, Witold Kycler, Dawid Murawa, Maciej Wiznerowicz, Ki Chung, W. Jeffrey Edenfield, Julie Martin, Eric Baudin, Glenn Bubley, Raphael Bueno, Assunta De Rienzo, William G. Richards, Steven Kalkanis, Tom Mikkelsen, Houtan Noushmehr, Lisa Scarpace, Nicolas Girard, Marta Aymerich, Elias Campo, Eva Giné, Armando López Guillermo, Nguyen Van Bang, Phan Thi Hanh, Bui Duc Phu, Yufang Tang, Howard Colman, Kimberley Evason, Peter R. Dottino, John A. Martignetti, Hani Gabra, Hartmut Juhl, Teniola Akeredolu, Serghei Stepa, Dave Hoon, Keunsoo Ahn, Koo Jeong Kang, Felix Beuschlein, Anne Breggia, Michael Birrer, Debra Bell, Mitesh Borad, Alan H. Bryce, Erik Castle, Vishal Chandan, John Cheville, John A. Copland, Michael Farnell, Thomas Flotte, Nasra Giama, Thai Ho, Michael Kendrick, Jean Pierre Kocher, Karla Kopp, Catherine Moser, David Nagorney, Daniel O'Brien, Brian Patrick O'Neill, Tushar Patel, Gloria Petersen, Florencia Que, Michael Rivera, Lewis Roberts, Robert Smallridge, Thomas Smyrk, Melissa Stanton, R. Houston Thompson, Michael Torbenson, Ju Dong Yang, Lizhi Zhang, Fadi Brimo, Jaffer A. Ajani, Ana Maria Angulo Gonzalez, Carmen Behrens, Jolanta Bondaruk, Russell Broaddus, Bogdan Czerniak, Bita Esmaeli, Junya Fujimoto, Jeffrey Gershenwald, Charles Guo, Alexander J. Lazar, Christopher Logothetis, Funda Meric-Bernstam, Cesar Moran, Lois Ramondetta, David Rice, Anil Sood, Pheroze Tamboli, Timothy Thompson, Patricia Troncoso, Anne Tsao, Ignacio Wistuba, Candace Carter, Lauren Haydu, Peter Hersey, Valerie Jakrot, Hojabr Kakavand, Richard Kefford, Kenneth Lee, Georgina Long, Graham Mann, Michael Quinn, Robyn Saw, Richard Scolyer, Kerwin Shannon, Andrew Spillane, Jonathan Stretch, Maria Synott, John Thompson, James Wilmott, Hikmat Al-Ahmadie, Timothy A. Chan, Ronald Ghossein, Anuradha Gopalan, Douglas A. Levine, Victor Reuter, Samuel Singer, Bhuvanesh Singh, Nguyen Viet Tien, Thomas Broudy, Cyrus Mirsaidi, Praveen Nair, Paul Drwiega, Judy Miller, Jennifer Smith, Howard Zaren, Joong Won Park, Nguyen Phi Hung, Electron Kebebew, W. Marston Linehan, Adam R. Metwalli, Karel Pacak, Peter A. Pinto, Mark Schiffman, Laura S. Schmidt, Cathy D. Vocke, Nicolas Wentzensen, Robert Worrell, Hannah Yang, Marc Moncrieff, Chandra Goparaju, Jonathan Melamed, Harvey Pass, Natalia Botnariuc, Irina Caraman, Mircea Cernat, Inga Chemencedji, Adrian Clipca, Serghei Doruc, Ghenadie Gorincioi, Sergiu Mura, Maria Pirtac, Irina Stancul, Diana Tcaciuc, Monique Albert, Iakovina Alexopoulou, Angel Arnaout, John Bartlett, Jay Engel, Sebastien Gilbert, Jeremy Parfitt, Harman Sekhon, George Thomas, Doris M. Rassl, Robert C. Rintoul, Carlo Bifulco, Raina Tamakawa, Walter Urba, Nicholas Hayward, Henri Timmers, Anna Antenucci, Francesco Facciolo, Gianluca Grazi, Mirella Marino, Roberta Merola, Ronald de Krijger, Anne Paule Gimenez-Roqueplo, Alain Piché, Simone Chevalier, Ginette McKercher, Kivanc Birsoy, Gene Barnett, Cathy Brewer, Carol Farver, Theresa Naska, Nathan A. Pennell, Daniel Raymond, Cathy Schilero, Kathy Smolenski, Felicia Williams, Carl Morrison, Jeffrey A. Borgia, Michael J. Liptay, Mark Pool, Christopher W. Seder, Kerstin Junker, Larsson Omberg, Mikhail Dinkin, George Manikhas, Domenico Alvaro, Maria Consiglia Bragazzi, Vincenzo Cardinale, Guido Carpino, Eugenio Gaudio, David Chesla, Sandra Cottingham, Michael Dubina, Fedor Moiseenko, Renumathy Dhanasekaran, Karl Friedrich Becker, Klaus Peter Janssen, Julia Slotta-Huspenina, Mohamed H. Abdel-Rahman, Dina Aziz, Sue Bell, Colleen M. Cebulla, Amy Davis, Rebecca Duell, J. Bradley Elder, Joe Hilty, Bahavna Kumar, James Lang, Norman L. Lehman, Randy Mandt, Phuong Nguyen, Robert Pilarski, Karan Rai, Lynn Schoenfield, Kelly Senecal, Paul Wakely, Paul Hansen, Ronald Lechan, James Powers, Arthur Tischler, William E. Grizzle, Katherine C. Sexton, Alison Kastl, Joel Henderson, Sima Porten, Jens Waldmann, Martin Fassnacht, Sylvia L. Asa, Dirk Schadendorf, Marta Couce, Markus Graefen, Hartwig Huland, Guido Sauter, Thorsten Schlomm, Ronald Simon, Pierre Tennstedt, Oluwole Olabode, Mark Nelson, Oliver Bathe, Peter R. Carroll, June M. Chan, Philip Disaia, Pat Glenn, Robin K. Kelley, Charles N. Landen, Joanna Phillips, Michael Prados, Jeffry Simko, Karen Smith-McCune, Scott VandenBerg, Kevin Roggin, Ashley Fehrenbach, Ady Kendler, Suzanne Sifri, Ruth Steele, Antonio Jimeno, Francis Carey, Ian Forgie, Massimo Mannelli, Michael Carney, Brenda Hernandez, Benito Campos, Christel Herold-Mende, Christin Jungk, Andreas Unterberg, Andreas von Deimling, Aaron Bossler, Joseph Galbraith, Laura Jacobus, Michael Knudson, Tina Knutson, Deqin Ma, Mohammed Milhem, Rita Sigmund, Andrew K. Godwin, Rashna Madan, Howard G. Rosenthal, Clement Adebamowo, Sally N. Adebamowo, Alex Boussioutas, David Beer, Thomas Giordano, Anne Marie Mes-Masson, Fred Saad, Therese Bocklage, Lisa Landrum, Robert Mannel, Kathleen Moore, Katherine Moxley, Russel Postier, Joan Walker, Rosemary Zuna, Michael Feldman, Federico Valdivieso, Rajiv Dhir, James Luketich, Edna M.Mora Pinero, Mario Quintero-Aguilo, Carlos Gilberto Carlotti,

Jose Sebastião Dos Santos, Rafael Kemp, Ajith Sankarankuty, Daniela Tirapelli, James Catto, Kathy Agnew, Elizabeth Swisher, Jenette Creaney, Bruce Robinson, Carl Simon Shelley, Eryn M. Godwin, Sara Kendall, Cassaundra Shipman, Carol Bradford, Thomas Carey, Andrea Haddad, Jeffey Moyer, Lisa Peterson, Mark Prince, Laura Rozek, Gregory Wolf, Rayleen Bowman, Kwun M. Fong, Ian Yang, Robert Korst, W. Kimryn Rathmell, J. Leigh Fantacone-Campbell, Jeffrey A. Hooke, Albert J. Kovatich, Craig D. Shriver, John DiPersio, Bettina Drake, Ramaswamy Govindan, Sharon Heath, Timothy Ley, Brian Van Tine, Peter Westervelt, Mark A. Rubin, Jung Il Lee, Natália D. Aredes, Armaz Mariamidze, Eliezer M. Van Allen, Andrew D. Cherniack, Giovanni Ciriello, Chris Sander, and Nikolaus Schultz. Oncogenic Signaling Pathways in The Cancer Genome Atlas. *Cell*, 173(2):321–337.e10, 2018. ISSN 10974172. doi: 10.1016/j.cell.2018.03.035.

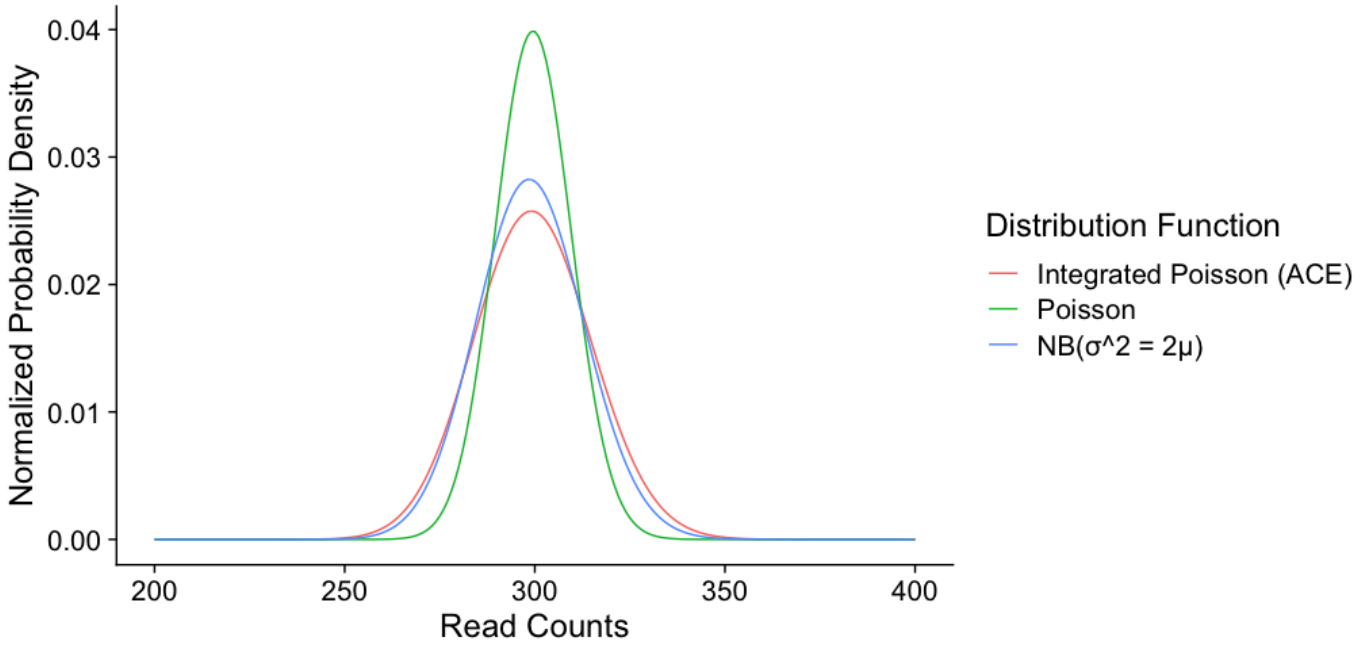
36. Isaia Barbieri, Konstantinos Tzelepis, Luca Pandolfini, Junwei Shi, Gonzalo Millán-Zambrano, Samuel C. Robson, Demetrios Aspris, Valentina Migliori, Andrew J. Bannister, Namshik Han, Etienne De Braekeleer, Hannes Ponstingl, Alan Hendrick, Christopher R. Vakoc, George S. Vassiliou, and Tony Kouzarides. Promoter-bound METTL3 maintains myeloid leukaemia by m6A-dependent translation control. *Nature*, 552(7683):126–131, nov 2017. ISSN 14764687. doi: 10.1038/nature24678.
37. Mohammad B. Uddin, Kartik R. Roy, Salman B. Hosain, Sachin K. Khiste, Ronald A. Hill, Seetharama D. Jois, Yunfeng Zhao, Alan J. Tackett, and Yong Yu Liu. An N6-methyladenosine at the transited codon 273 of p53 pre-mRNA promotes the expression of R273H mutant protein and drug resistance of cancer cells. *Biochemical Pharmacology*, 160:134–145, feb 2019. ISSN 18732968. doi: 10.1016/j.bcp.2018.12.014.
38. S Lin, J Choe, P Du, R Triboulet, and R I Gregory. The m6A Methyltransferase METTL3 Promotes Translation in Human Cancer Cells. *Mol. Cell*, 62:335–345, 2016.
39. Dan Dominissini, Sharon Moshitch-Moshkovitz, Schraga Schwartz, Mali Salmon-Divon, Lior Ungar, Sivan Osenberg, Karen Cesarkas, Jasmine Jacob-Hirsch, Ninette Amariglio, Martin Kupiec, Rotem Sorek, and Gideon Rechavi. Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature*, 485(7397):201–206, may 2012. ISSN 00280836. doi: 10.1038/nature11112.
40. C T Kwok, A D Marshall, J E J Rasko, and J J L Wong. Genetic alterations of m6A regulators predict poorer survival in acute myeloid leukemia. *J. Hematol. Oncol*, 10:39, 2017.
41. Robin Graf, Xun Li, Van Trung Chu, and Klaus Rajewsky. sgRNA Sequence Motifs Blocking Efficient CRISPR/Cas9-Mediated Gene Editing. *Cell Reports*, 26(5):1098–1103.e3, jan 2019. ISSN 2211-1247. doi: 10.1016/J.CELREP.2019.01.024.
42. Santiago Gisler, Joana P. Gonçalves, Waseem Akhtar, Johann de Jong, Alexey V. Pindyurin, Lodewyk F. A. Wessels, and Maarten van Lohuizen. Multiplexed Cas9 targeting reveals genomic location effects and gRNA-based staggered breaks influencing mutation efficiency. *Nature Communications*, 10(1):1598, dec 2019. ISSN 2041-1723. doi: 10.1038/s41467-019-09551-w.
43. Sebald AN Verkuijl and Marianne G Rots. The influence of eukaryotic chromatin state on CRISPR–Cas9 editing efficiencies. *Current Opinion in Biotechnology*, 55:68–73, feb 2019. ISSN 0958-1669. doi: 10.1016/J.COPBIO.2018.07.005.
44. Jennifer Listgarten, Michael Weinstein, Benjamin P. Kleinstiver, Alexander A. Sousa, J. Keith Joung, Jake Crawford, Kevin Gao, Luong Hoang, Melih Elibol, John G. Doench, and Nicolo Fusi. Prediction of off-target activities for the end-to-end design of CRISPR guide RNAs. *Nature Biomedical Engineering*, 2(1):38–47, jan 2018. ISSN 2157-846X. doi: 10.1038/s41551-017-0178-6.
45. Belen Gutierrez, Jérôme Wong Ng, Lun Cui, Christophe Becavin, and David Bikard. Genome-wide CRISPR-Cas9 screen in E. coli identifies design rules for efficient targeting. *bioRxiv*, page 308148, apr 2018. doi: 10.1101/308148.
46. DepMap FAQ, 2020.
47. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2018.
48. Matt Dowle and Arun Srinivasan. *data.table: Extension of 'data.frame'*, 2019.
49. Winston Chang. *R6: Classes with Reference Semantics*, 2017.

Supplementary Section 1: Supplemental Figures

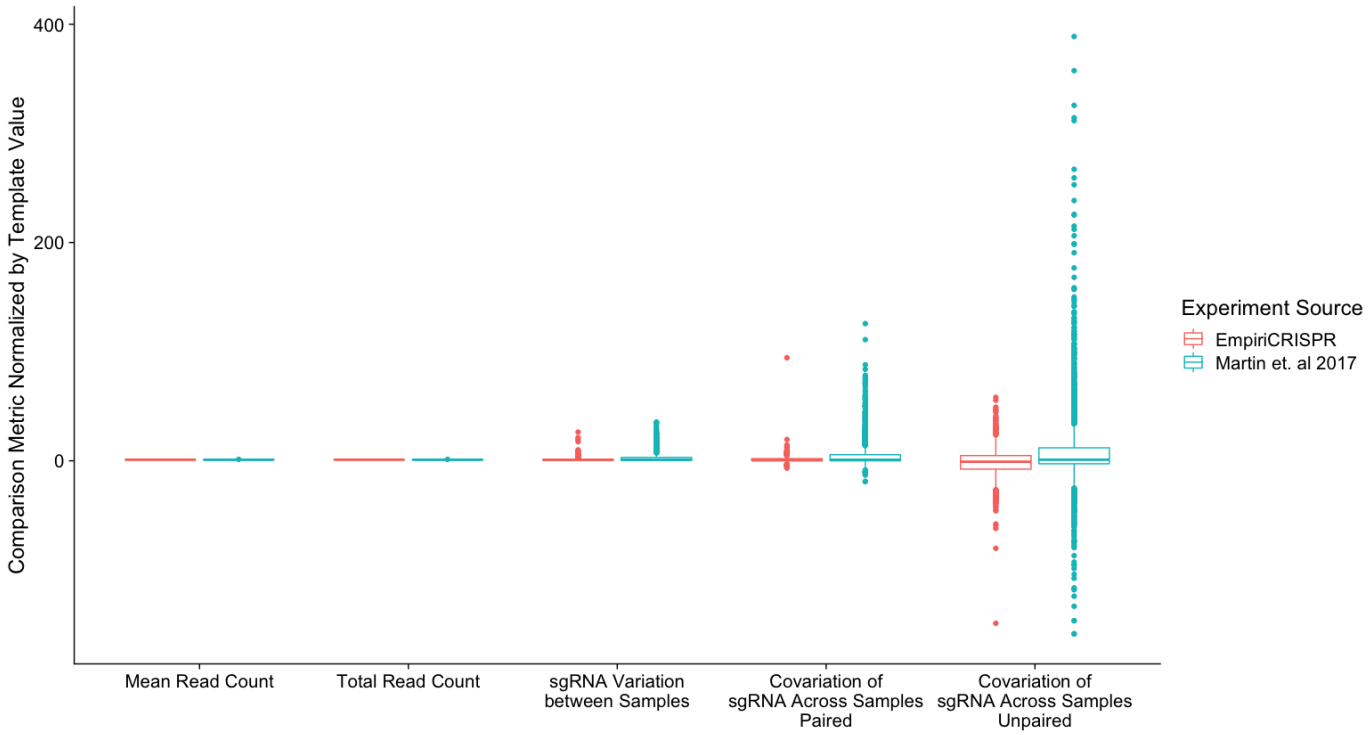
Method	Directly Models Count Data	Estimates Essentiality	Tests for Differences Between Subtypes
ACE	✓	✓	✓
Average Fold Change (AFC)	✗	✓	✗
BAGEL (Hart et al. 2015)	✗	✗	✓
JACKS (Allen et al. 2019)	✗	✓	✗
MAGeCK-RRA (Li et al. 2014)	✗	✗	✗
CERES (Meyers et al. 2017)	✗	✓	✓
ScreenBEAM (Yu et al. 2015)	✗	✓	✗
CB2 (Jeong et al. 2019)	✓	✗	✗
PinAPL-Py (Spahn et al. 2017)	✓	✗	✗
MAGeCK-MLE (Li et al. 2015)	✓ *	✓	✗

*depleted read counts only

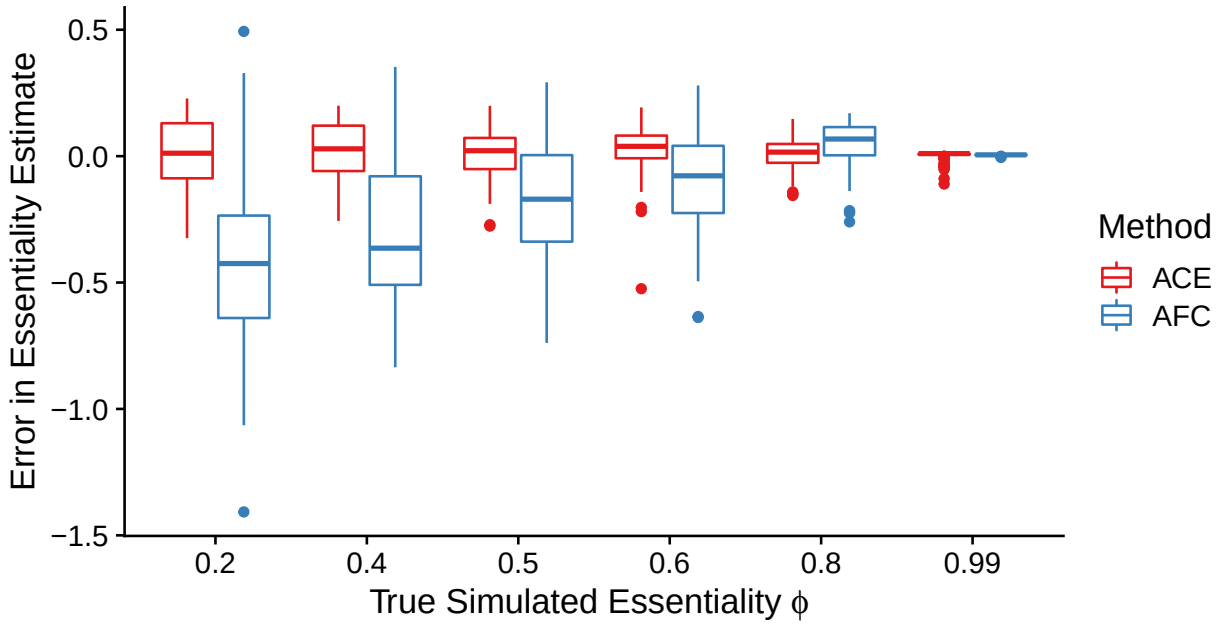
Supplementary Figure S1. Methods for Identifying Differential Essentiality in CRISPR Screens. Many methods have been developed to analyze CRISPR screening data, but none are designed to test for differential essentiality between sample subtypes as a part of their inference framework. Some, such as BAGEL and CERES, test for significant differences in essentiality after parameter inference, but do not test whether the data is best described by a single essentiality parameter ('Tests for Differences Between Subtypes' column). Directly modeling count data, as opposed to relying upon summary statistics such as log-fold change, enables ACE to adapt to variations in experimental design, and leverage information from all available data ('Directly Models Count Data' column). Providing a quantification of essentiality, as opposed to fitting a binary essential/nonessential classification, enables researchers to more intuitively estimate phenotype ('Estimates Essentiality' column).



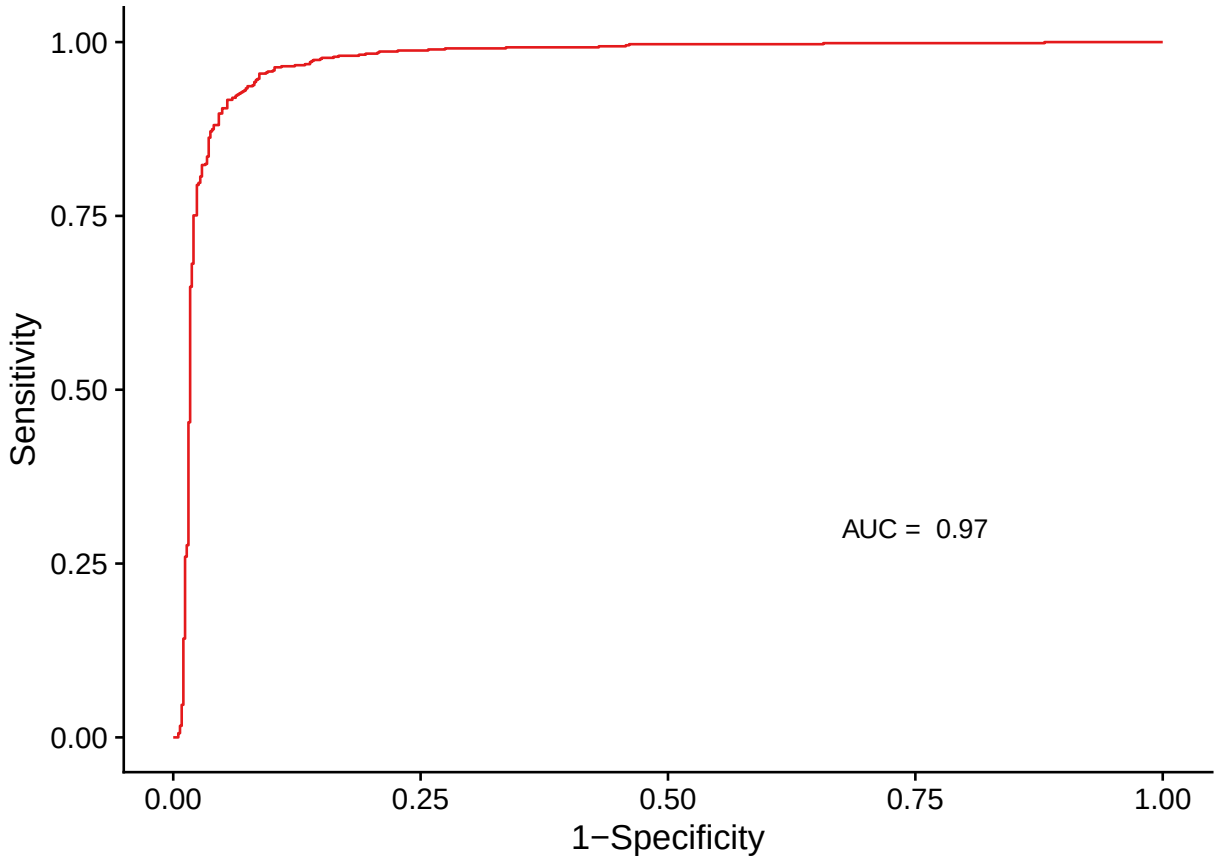
Supplementary Figure S2. Illustration of Dispersion of ACE Distribution Function. Read count data is known to have a high dispersion similar to a Negative Binomial distribution (5). The values shown here reflect the probability density of different values of read counts in a CRISPR screen, as described by a Poisson distribution with point estimates of the mean (green), a Negative Binomial distribution with point estimates of the mean and a variance equal to twice the mean (blue). The ACE model integrates over all possible numbers of infected sgRNA, resulting in a probability density (red) similar to the Negative Binomial.



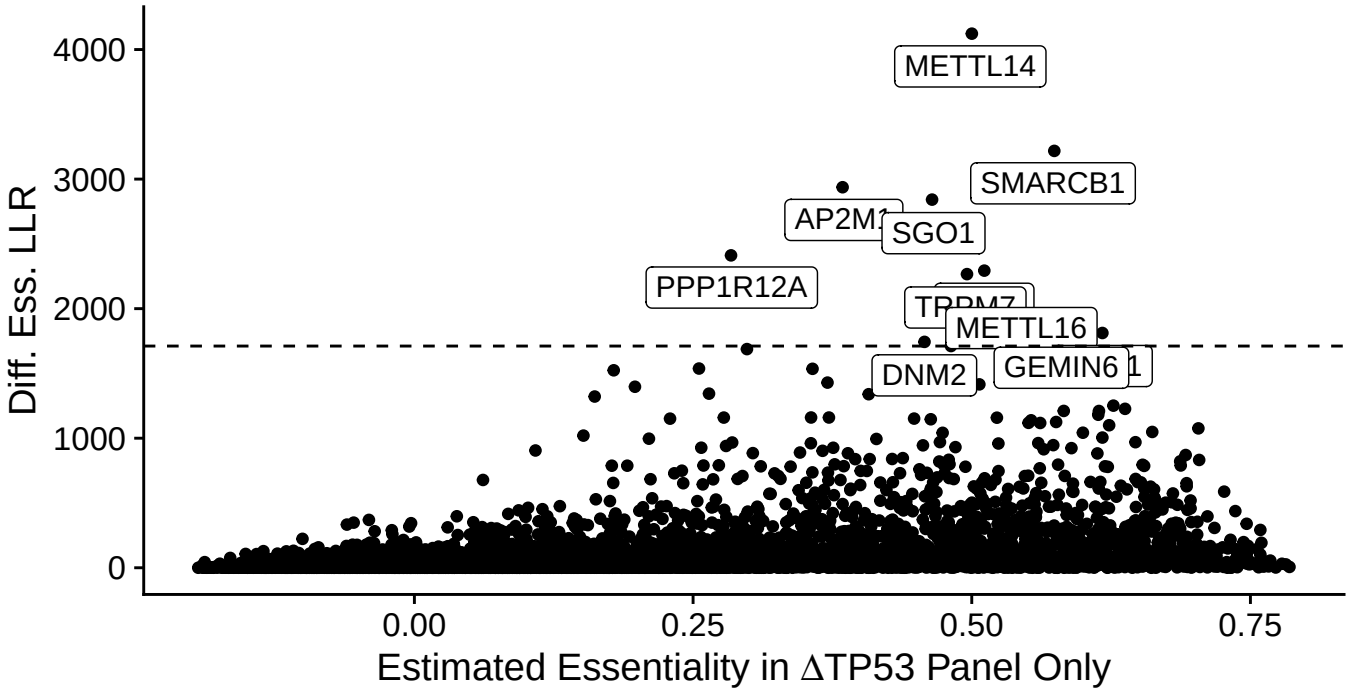
Supplementary Figure S3. Simulated Data for Evaluation of ACE Performance. The genome-wide CRISPR screen performed by Martin et. al 2017 (16) was used as our EmpiriCRISPR template, as this work sequenced the sgRNA master library used as well as the initial and final samples. The sequenced master library provided the initial sgRNA abundances in the simulation, and all metrics shown were calculated in the initial and depleted counts. Shown are approximately 2300 guides targeting globally nonessential genes, as identified using ENCODE expression data (see Methods), and a randomly selected matching subset of nonessential sgRNA from simulated read counts. Mean and total read counts were calculated within each sample, while variation and covariation metrics were calculated for each guide across samples. 'Paired' indicates covariance was calculated between initial and depleted samples, with each sgRNA paired with the depleted sample produced by the same master library transfection; 'Unpaired' have been paired with samples from a different transfection replicate.



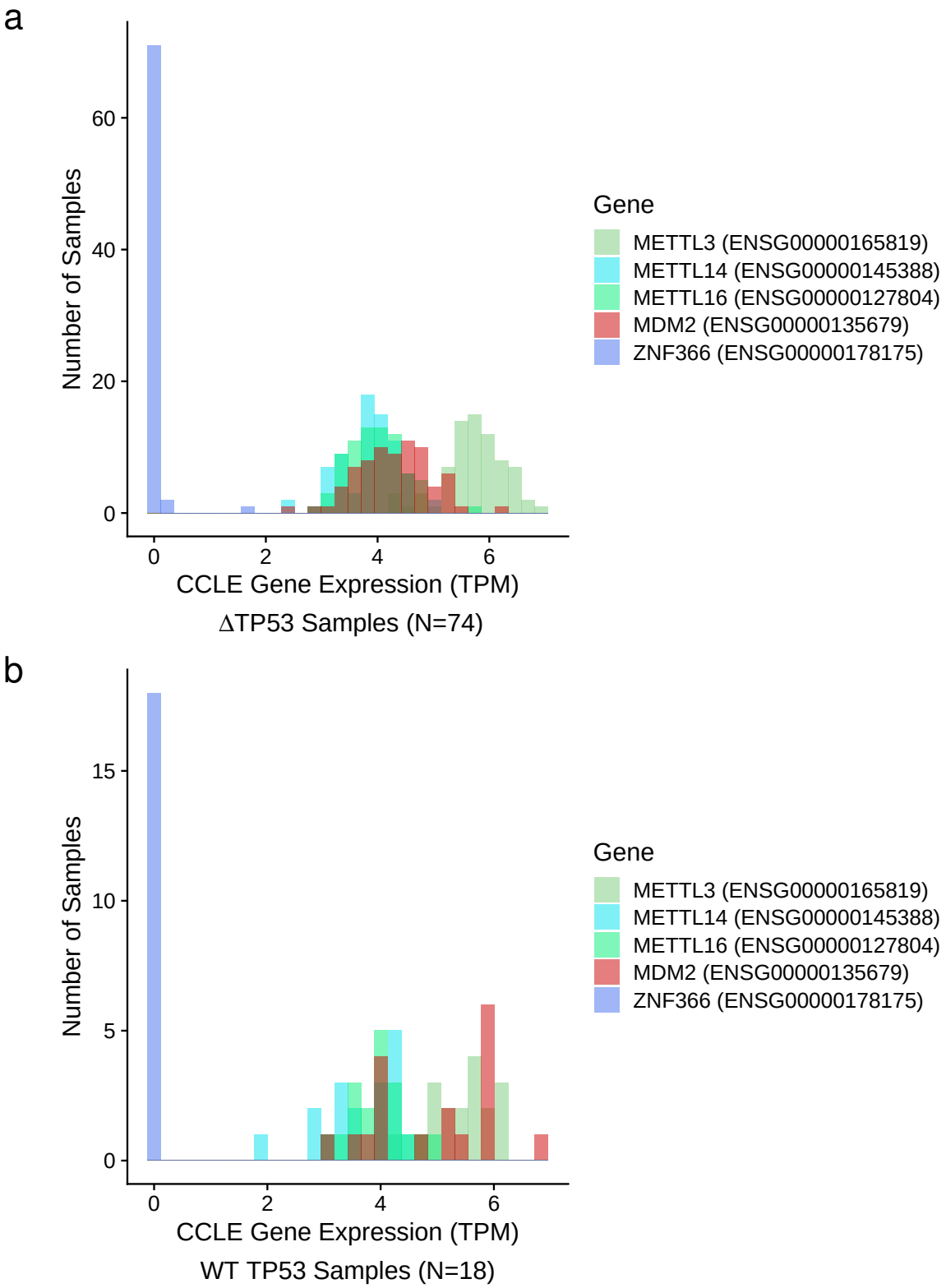
Supplementary Figure S4. Error in Estimation of Gene Essentiality. Gene essentiality ϕ_G was calculated using ACE and average fold change (AFC; see Methods) across three replicates of simulated count data. 300 genes were simulated at each essentiality level shown; results from an additional 3,150 nonessential genes are not shown.



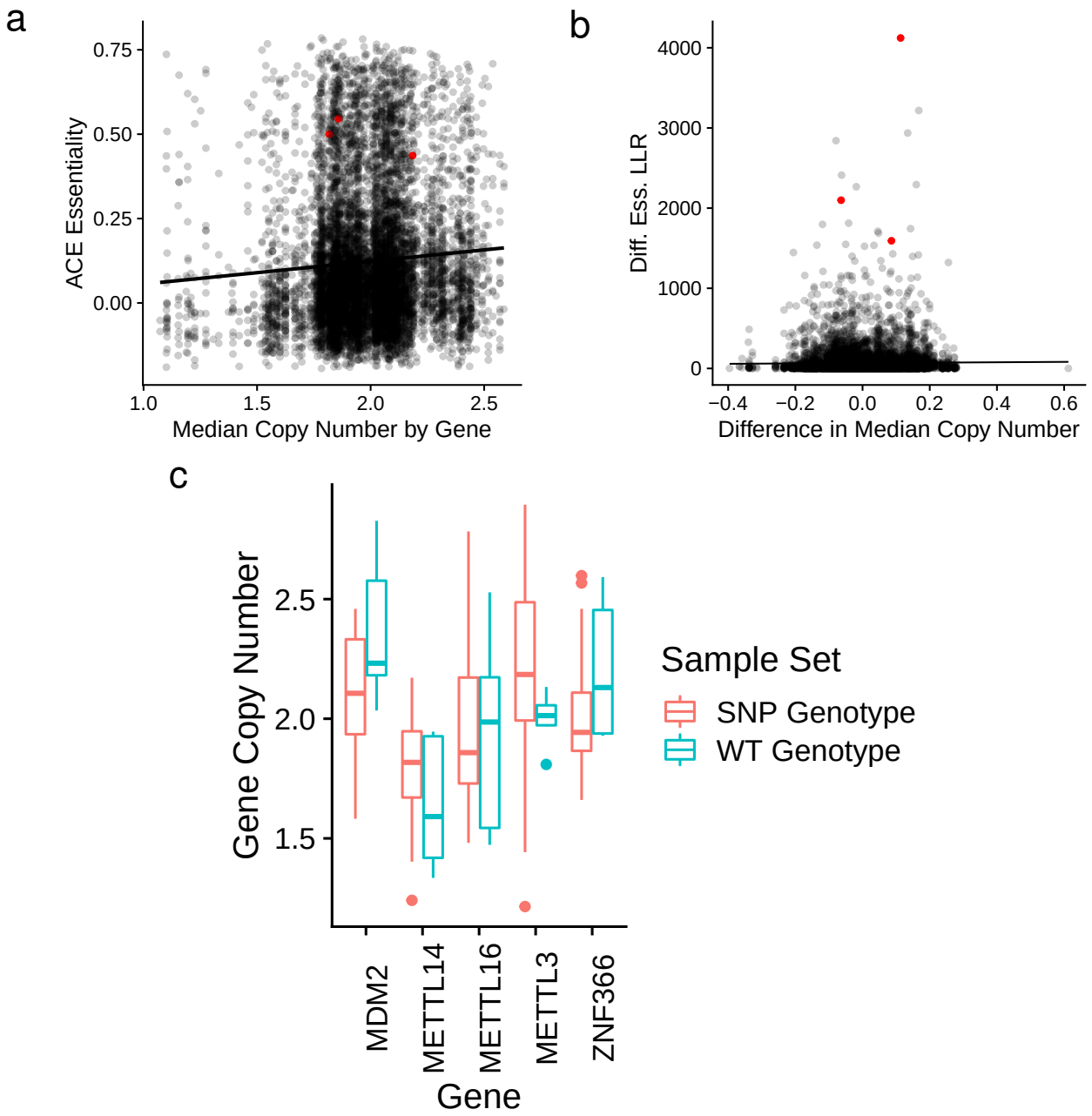
Supplementary Figure S5. ACE Classification of Established Essential and Nonessential Genes in Achilles DepMap CRISPR Screens. All 220 CRISPR screens performed in 92 different NSCLC Adenocarcinoma cell lines in the Achilles DepMap project (9) were used by ACE to estimate the likelihood of essentiality of 688 essential genes.



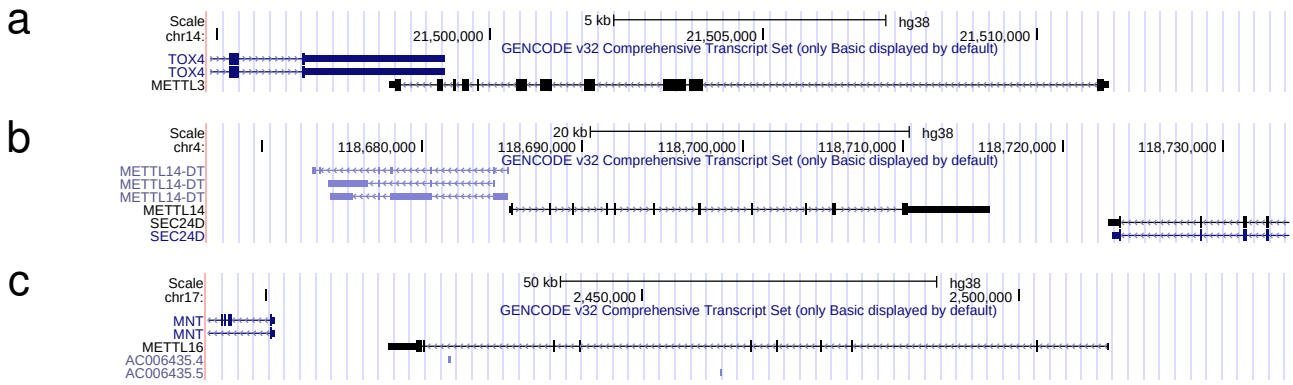
Supplementary Figure S6. Genes Essential in NSCLC Adenocarcinoma cell lines with a 'damaging' mutation in *TP53*. All genes evaluated by genome-wide CRISPR screens in NSCLC cell lines by the DepMap Project Achilles are shown. On the x-axis is the ACE estimate of essentiality in cell lines with a nonsense or frameshift mutation indicated as 'damaging' in *TP53* (ΔTP53 Panel; see Methods). On the y-axis is the test statistic indicating whether this essentiality is significantly different from cell lines with wildtype *TP53*, calculated with a log likelihood ratio test comparing a shared versus a unique essentiality between the two sample panels. The horizontal line indicates the Bonferroni-correct empirical p-value <0.05.



Supplementary Figure S7. Expression Levels of Differential Essentiality Candidates. CCLE reported expression levels of differential essentiality candidates *METTL3*, *METTL14*, and *METTL16* in cell lines used by the Achilles DepMap project with annotated nonfunctional (a) and presumed functional (b) variants of *TP53* (9, 30, 31). *MDM2* is shown as a positive control for gene essentiality; *ZNF366* as a negative control.



Supplementary Figure S8. ACE Essentiality Estimates Avoid Copy Number Sensitivity by Pooling Across Sample Panels. (a) The ACE estimated essentiality of each gene (y-axis), with the median copy number of that gene across all NSCLC Adenocarcinoma samples with a mutation in *TP53* used to infer the essentiality value (x-axis). (b) The differential essentiality test statistic of each gene between wildtype and Δ*TP53* cell lines, determined by ACE. On the x-axis is the difference in median copy number between the compared panels of cell lines. In red are *METTL3, METTL14* and *METTL16*. The black line indicates the best linear relationship given a gaussian error model. (c) Copy number variation across each sample panel for the three candidate differentially essential methyltransferases, with *MDM2* as a positive control, and *ZNF366* as a negative control for differential essentiality. The range in copy number largely overlaps for all genes in these sample panels.



Supplementary Figure S9. Differentially Essential Gene Candidates Independent of Flanking Genes. The majority of METTL3 (a), and all of METTL14 (b), and METTL16 (c) do not overlap with other coding genes. Image from the UCSC genome browser.