

ATARIS: Computational quantification of gene suppression phenotypes from multisample RNAi screens

Diane D. Shao,^{1,2,6} Aviad Tsherniak,^{1,6} Shuba Gopal,¹ Barbara A. Weir,¹ Pablo Tamayo,¹ Nicolas Stransky,^{1,7} Steven E. Schumacher,^{1,3} Travis I. Zack,^{1,3,4} Rameen Beroukhi,^{1,2,3,5} Levi A. Garraway,^{1,2,5} Adam A. Margolin,^{1,8} David E. Root,¹ William C. Hahn,^{1,2,5,9} and Jill P. Mesirov^{1,9}

¹Broad Institute of Harvard and M.I.T., Cambridge, Massachusetts 02142, USA; ²Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts 02215, USA; ³Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, Massachusetts 02215, USA; ⁴Program in Biophysics, Harvard University, Boston, Massachusetts 02115, USA; ⁵Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts 02115, USA

Genome-scale RNAi libraries enable the systematic interrogation of gene function. However, the interpretation of RNAi screens is complicated by the observation that RNAi reagents designed to suppress the mRNA transcripts of the same gene often produce a spectrum of phenotypic outcomes due to differential on-target gene suppression or perturbation of off-target transcripts. Here we present a computational method, Analytic Technique for Assessment of RNAi by Similarity (ATARIS), that takes advantage of patterns in RNAi data across multiple samples in order to enrich for RNAi reagents whose phenotypic effects relate to suppression of their intended targets. By summarizing only such reagent effects for each gene, ATARIS produces quantitative, gene-level phenotype values, which provide an intuitive measure of the effect of gene suppression in each sample. This method is robust for data sets that contain as few as 10 samples and can be used to analyze screens of any number of targeted genes. We used this analytic approach to interrogate RNAi data derived from screening more than 100 human cancer cell lines and identified *HNF1B* as a transforming oncogene required for the survival of cancer cells that harbor *HNF1B* amplifications. ATARIS is publicly available at <http://broadinstitute.org/ataris>.

[Supplemental material is available for this article.]

RNAi screening is a powerful approach that facilitates the systematic assessment of the effect of gene suppression on cell phenotypes such as cell death or the activity of a signaling pathway. The development and availability of genome-scale RNAi libraries provide the tools to identify new pathway components and context-specific cancer dependencies (Kittler and Pelletier 2008; Luo et al. 2008; Hirsch 2010; Cheung et al. 2011). Technological and analytical advances will provide further opportunities for the application and interpretation of functional screens.

For screens in mammalian cells, a short interfering RNA (siRNA) is introduced into cells either directly as a duplex or by expression of a short hairpin RNA (shRNA) that is processed into active siRNA. This siRNA is designed to specifically degrade mRNA transcripts of complementary sequence to reduce the expression of gene products (Elbashir et al. 2001; Root et al. 2006). In practice, these reagents exhibit a variable degree of suppression of the targeted gene, and may also suppress genes other than the intended target (Jackson et al. 2003, 2006; Birmingham et al. 2006). Here we refer to a reagent's phenotypic effects resulting from suppression

of unintended genes as *off-target effects*. Analytical approaches to identify specific types of off-target effects in siRNA (e.g., seed sequence similarity) have been previously developed (Marine et al. 2012; Sigoillot et al. 2012). However, we currently lack the ability to systematically characterize both the on-target and off-target effects of siRNAs.

To identify candidate genes that produce a desired phenotype based on imperfect reagents, multiple distinct RNAi reagents targeting each gene are often screened (Cullen 2006; Echeverri et al. 2006). Analyzing data from multiple reagents per gene has the potential to (1) increase the power to detect candidate genes, and (2) decrease false-positive rates. For example, the "frequency approach" considers a gene a candidate in a sample if several of its reagents induce a desired effect. This effect is usually measured by deviation from the experimental, or de facto, negative control effects (Müller et al. 2005; Bard et al. 2006; Chung et al. 2008). A variation of this approach is to assign a gene score by using a simple function, such as the average, of a few reagents with the most desired effects (Marcotte et al. 2012). More recent methods for scoring genes in individual samples, such as "redundant siRNA activity" (RSA) (König et al. 2007) and "strictly standardized mean difference" (SSMD) (Zhang et al. 2007, 2011) have further decreased false-positive rates. For each sample, they consider the phenotypes produced by reagents for all of the screened genes simultaneously (RSA) or by all reagents for each gene separately (SSMD).

As RNAi screens are being performed in increasing numbers of samples (Collinet et al. 2010; Brough et al. 2011; Cheung et al.

⁶These authors contributed equally to this work.
Present addresses: ⁷Blueprint Medicines, Cambridge, MA 02142, USA;
⁸Sage Bionetworks, Seattle, WA 98109, USA.

⁹Corresponding authors
E-mail mesirov@broad.mit.edu
E-mail William_Hahn@dfci.harvard.edu

Article published online before print. Article, supplemental material, and publication date are at <http://www.genome.org/cgi/doi/10.1101/gr.143586.112>.

2011; Marcotte et al. 2012), a common analytical approach has been to segregate samples into two predefined classes in order to identify genes with differential effects. By summarizing data within each class, aberrant reagent effects in individual samples are less likely to impact the final result. The “second best” method assigns scores to genes based on each gene’s second most differentially scoring reagent between classes (Cheung et al. 2011), requiring—similar to the “frequency approach”—that favorable genes have at least two high-scoring reagents. Alternatively, RNAi Gene Enrichment Ranking (RIGER) ranks all the reagents by their differential effects and generates a gene-level score for each gene based on the rank distribution of its reagents (Luo et al. 2008; Barbie et al. 2009), analogous to RSA. However, the requirement of two predefined classes can limit full inter-rogation of the data.

Currently, RNAi analysis methods do not attempt to assess the performance of individual reagents. Thus, there is an opportunity to further improve analysis of RNAi data by harnessing the statistical information across many samples to identify and avoid data from problematic reagents when determining gene-level effects. An analogous approach is used by dChip (Li and Hung Wong 2001) and RMA (Irizarry et al. 2003), two widely used methods for mRNA abundance quantification in microarray data. Given a set of samples, these algorithms quantify a probe set’s overall abundance level in each sample from a set of multiple, distinct, complementary probes. In the case of RNAi data, one must also consider additional factors such as off-target effects thought to exist for a subset of reagents, the dramatically greater biological variability, and the possibility of multiple phenotypic effects for a single gene (e.g., due to different levels of on-target gene suppression).

Here we introduce ATARiS (Analytic Technique for Assessment of RNAi by Similarity), a novel computational approach to the quantification of gene-specific suppression phenotypes. ATARiS uses patterns in the data from multisample RNAi screens to estimate the performance of individual RNAi reagents targeting each gene and generates a per-gene value for each sample that quantifies the phenotypic effect of gene suppression. We used data from two recent large-scale shRNA screens of 102 and 72 cancer cell lines (Cheung et al. 2011; Marcotte et al. 2012), respectively, to demonstrate the performance of ATARiS. We integrated ATARiS-generated gene phenotype values with gene copy-number and gene expression data to uncover novel cancer dependencies, including the identification of a novel amplified oncogene, *HNF1B*. We are making ATARiS publicly available (<http://broadinstitute.org/ataris>) in hopes of aiding current RNAi screening efforts.

Results

ATARiS overview

ATARiS is a computational method to assess gene suppression effects in each sample of multisample RNAi screens that include at least two RNAi reagents (siRNA or shRNA) designed to target each gene. Our method uses only data from reagents determined to have primarily on-target effects, discarding data from reagents with off-target effects. To identify on-target reagents, we noted that in an RNAi library, reagents are designed to target distinct sequences. Thus, it is unlikely that any two reagents—including those targeting the same gene—will suppress the same set of off-target genes. We therefore concluded that when RNAi reagents designed to target the same gene behave similarly across the screened samples, the observed effects are likely due to suppres-

sion of the intended gene rather than off-target suppression. For each gene in a screen, ATARiS identifies sets of reagents with similar behavior across all samples in order to produce two types of results:

1. A *gene solution* that summarizes the observed effects produced by identified on-target reagents into quantitative values across all screened samples (the value for an individual sample is called a *phenotype value*). We account for potential multiple phenotypic outcomes after suppression of a given gene, possibly due to different degrees of gene suppression, by allowing for multiple solutions composed of disjoint sets of consistent reagents.
2. A *consistency score* for each RNAi reagent that represents the confidence that its observed phenotypic effects are the result of on-target gene suppression. ATARiS assigns higher consistency scores to reagents whose profiles (i.e., the observed effect of that reagent in every screened sample) exhibit higher correlation to a larger number of reagent profiles within the same solution.

We give a general description of ATARiS here (see also Supplemental Fig. 1) and provide technical details in Methods. Figure 1A–D summarizes the different approaches used by ATARiS and current RNAi analysis methods.

First, to construct a *gene solution* for a given gene G , ATARiS considers the observed data of all RNAi reagents designed to target G . The data for each reagent r are median-centered, because we are interested in reagents whose *relative* effects across the samples are similar. For each sample s , ATARiS calculates a value c_s that summarizes the effects produced by all the reagents targeting G in s . We refer to the vector \mathbf{c} of all c_s values as the *consensus profile*. To estimate c_s , ATARiS models each data point $x_{r,s}$ (i.e., the observed effect induced by reagent r in sample s) as a product of two unknown quantities: e_r , representing the relative magnitude of the effects of reagent r , and c_s . ATARiS estimates the values for e_r and c_s by minimizing an L_1 -norm objective function using the method of alternating minimizations (Csiszar and Tusnady 1984). An L_1 norm makes the optimization more robust to outliers, which are common in this type of data.

Next, ATARiS iteratively refines the considered set of reagents by evaluating the similarity of each reagent profile to the consensus profile. If, for any reagent profile, the significance of the Spearman correlation (estimated using an empirical null distribution) is lower than a predefined threshold, the reagent whose profile is least similar to the consensus profile is discarded from further analysis. ATARiS then repeats the process of computing a consensus profile and discarding the most dissimilar reagent until either only one reagent remains—in which case no gene solution is generated—or until all remaining reagents have profiles significantly similar to the consensus profile. The consensus profile \mathbf{c} for the retained reagents is then used as a *gene solution* for that gene, and we refer to its elements c_s as the gene’s phenotype values. The entire process is then repeated for any remaining reagents not yet contributing to a solution until no more solutions are found. A greedy approach to refinement, rather than an exhaustive one, allows scaling to larger numbers of reagents per gene.

After generating all gene solutions for gene G , ATARiS computes a consistency score for each of its reagents. The consistency score for reagent r is based on the negative \log_{10} of the integrated P -values of the Spearman correlation coefficients of r ’s profile to each of the other reagent profiles within the same solution. Thus, the consistency score may be interpreted as a P -value, i.e., a con-

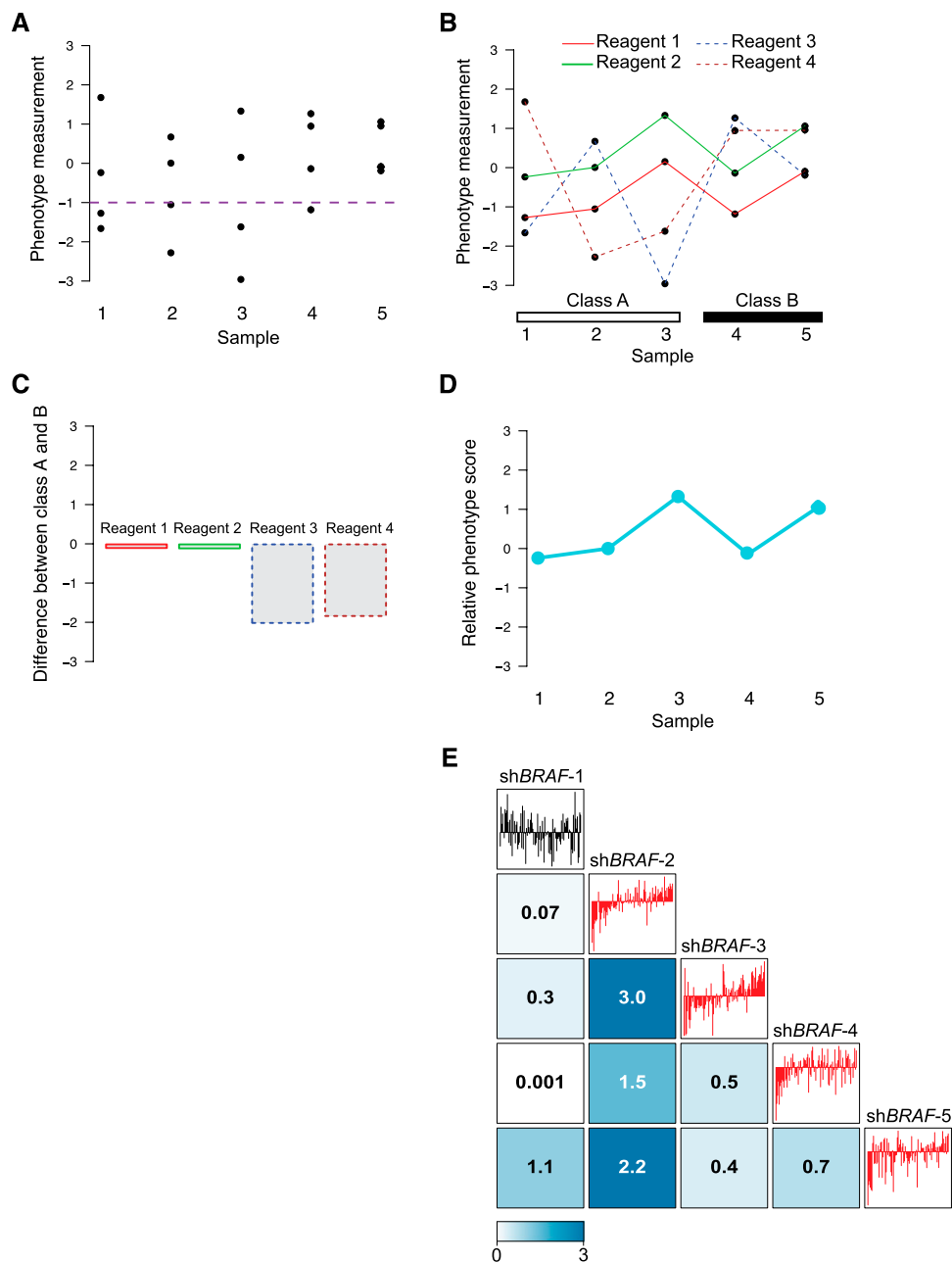


Figure 1. ATARIS accounts for patterns in RNAi reagent data in order to quantify the phenotypic effect of gene suppression in each sample. (A–D) Hypothetical phenotypic data from four RNAi reagents, all designed to target the same gene, in five independent samples from two classes, A and B. (A) Samples 1, 2, and 3 each have at least two reagents that score below a desired threshold (purple dotted line); thus, according to “frequency approach” methods, this gene may be a “hit” in those samples. (B) A line connecting each reagent’s effects across the samples reveals additional information. Specifically, we note that it is possible (as in this scenario) that different shRNAs drive the determination of hits in each sample when samples are each analyzed separately as in A. (C) For each reagent, the difference between its mean values in class A and class B is shown, reducing much of the noise from individual samples. Reagents 3 and 4 both show differential effects between the classes and would suggest that two-class-based analytic methods select this gene as a hit. (D) ATARIS phenotype values for each of the screened samples. Phenotype values represent relative gene-level effects in each individual sample by incorporating information from trends across all samples, favoring reagents that produce correlated effects (i.e., reagents 1 and 2 from B). If the user chooses to assess whether differential effects exist between classes A and B, this example would show no significant difference by avoiding uncorrelated reagents 3 and 4. (E) Real data from the Project Achilles data set for shRNAs targeting *BRAF*. Median-normalized screening data across 102 samples are displayed as barplots in sample order of ascending *BRAF* phenotype value. Boxed numbers display $-\log_{10}(P\text{-value})$ of the Spearman correlation coefficient for the two shRNAs labeled in the corresponding row and column. (Red) shRNAs with correlated effects that are incorporated into the *BRAF* gene solution.

sistency score of 1.3 corresponds to $-\log_{10}(P\text{-value of } 0.05)$. For RNAi reagents that do not participate in any solution, all reagents targeting *G* are considered in computing the consistency score.

Thus, even for a reagent that is excluded from a solution depending on the predefined threshold, ATARIS still provides an assessment of the confidence in its functional effects.

Application of ATARiS to data derived from multisample shRNA screens

To test and validate ATARiS, we primarily used the data produced by Project Achilles—a data set produced from massively parallel screening of 102 cancer cell lines with a genome-scale pooled shRNA library targeting more than 11,000 human genes with an average of five shRNAs per gene (Cheung et al. 2011). The final abundance of each shRNA after propagation of the cell line was determined with respect to the initial reference shRNA pool to assess cellular dependency on each shRNA's target (i.e., shRNAs that target essential genes will be depleted). See Methods for a full description of additional data pre-processing and normalization steps. The resulting data set is available as Supplemental Data 1.

The application of ATARiS to this data set yielded gene solutions for 7250 genes, and incorporated data from 49.5% of the screened shRNA reagents (Supplemental Data 2) when using a 0.15 significance threshold. With this threshold, we would expect ATARiS to generate solutions, on average, for 15% of the genes using randomly permuted data. Supplemental Figure 2 shows the distribution of the number of gene solutions identified for varying thresholds. In our data set, 6233 genes had one associated gene solution, 1017 genes had two or more solutions, and 3955 genes had no solutions (Supplemental Table 1). We illustrate the type of correlated reagent data that becomes incorporated into a gene solution by using an example gene *BRAF* (Fig. 1E). A consistency score was generated for every screened shRNA, including those that do not participate in any gene solution (Supplemental Data 3).

Influence of data set size and biological context on ATARiS results

Since most RNAi screens currently do not include as many samples as Project Achilles, we assessed the robustness of ATARiS on simulated cases where data from fewer samples were available. We generated 100 random subsets of the Project Achilles data set for each sample size of 10, 20, . . . , 100 cell lines. We used only the data for genes with solutions from analyzing the full 102-sample Achilles data set, which served as a reference data set for these analyses. For the subsets of size 10, 30, and 50 cell lines, we found that the median percentage of genes with solutions is 67%, 80%, and 85%, respectively (Fig. 2A). This increase corresponds to an overall increase in the number of genes with solutions as the sample size increases (Supplemental Fig. 3). In addition, we observed a substantial overlap between the shRNAs that are used for a gene solution in the reference data set and the shRNAs used in the subsets of size 10, 30, and 50 cell lines (median percentages 75%, 85%, and 89%, respectively) (Supplemental Fig. 4). The robustness of shRNA choice for gene solutions resulted in generally robust values for gene phenotype values even at small sample size. For example, ATARiS results from subsets of as few as 10 cell lines showed that >75% of gene phenotype values deviated from their respective gene phenotype values in the reference data set by less than one standard deviation (Fig. 2B). We observed similarly robust results when running ATARiS on data sets after the addition of random noise (Supplemental Fig. 5). We concluded that we can use ATARiS to identify consistent shRNAs, and thus to determine gene phenotype values, even for small numbers of screened samples.

We evaluated whether ATARiS results are robust when applied to samples screened independently of Project Achilles by (1) simulating independent screening data sets and (2) analyzing results from a set of independently performed genome-scale pooled shRNA viability screens (Marcotte et al. 2012). To simulate in-

dependent data sets, we generated 100 *pairs* of random disjoint subsets (i.e., no overlapping samples) from the full Achilles data set for each sample size of 10, 20, . . . , 50 cell lines. For pairs of subsets of size 10, 30, and 50 cell lines, the median percentage of shRNAs that were used to generate solutions in both disjoint subsets of samples are 71%, 80%, and 84%, respectively (Fig. 2C). Next, we analyzed the RNAi screening data from Marcotte et al. (2012), composed of 72 cancer cell lines screened using a comparable shRNA library (see Methods for details). ATARiS found relatively fewer gene solutions in this data (as a fraction of the number of genes targeted), consistent with its having fewer samples and higher homogeneity in cell lineages (Supplemental Table 3). For genes that have a solution in both data sets, we found that the shRNAs targeting those genes are more likely to participate in a solution in both data sets than in one data set but not in the other (odds ratio = 2.1; 95% confidence interval [1.96, 2.26]; P -value $< 2.2 \times 10^{-16}$; Fisher's exact test). Furthermore, the Pearson correlation coefficient for ATARiS shRNA consistency scores in the two data sets is 0.46 (95% confidence interval [0.45, 0.47]; P -value $< 2.2 \times 10^{-16}$). Together, these observations suggested that ATARiS produces robust results between independent screens.

To account for the fewer number of solutions when smaller subsets of cell lines are analyzed, we hypothesized that gene solutions can be identified primarily for genes whose suppression yields phenotypic variation across samples. To test this hypothesis, we determined the frequency of finding a *BRAF* solution when we apply ATARiS to sets of *BRAF* wild-type cell lines (which are expected to exhibit similar dependence on *BRAF*) versus when the set contains an equal number of *BRAF* wild-type and mutant cell lines (mutant lines are much more dependent on *BRAF* relative to wild type). In the latter case, a *BRAF* solution was found in 98% of runs using only 10 cell lines (five wild type, five mutant), whereas using as many as 26 wild-type cell lines alone (26 wild type, 0 mutant) yielded solutions in only 82% of runs (Fig. 2D). Thus, the reduced number of solutions at smaller sample size likely reflects the reduced overall heterogeneity among a few samples as compared with the full sample set.

Validation of shRNA consistency scores

ATARiS shRNA consistency scores are intended to reflect our confidence in the specificity of each reagent. However, validation is challenging since the currently accepted standards for evaluating reagent performance, i.e., immunoblotting and quantitative RT-PCR for on-target gene suppression, cannot assess off-target effects, whereas ATARiS consistency scores attempt to encompass both on-target and off-target aspects. Since a greater degree of target gene suppression does not necessarily amplify functional outcome, and effective on-target gene suppression does not equate to lack of off-target effects, we did not expect high correlation between immunoblotting results and ATARiS scores. We expected, however, that shRNAs with high consistency scores have some degree of detectable on-target gene suppression in order to produce correlated profiles.

We validated consistency scores on a few selected genes—*BRAF*, *PIK3CA*, *KRAS*, and *MYC*—chosen for their importance in cancer, availability of reagents to assess the expression of these genes, and known functional effect of their shRNAs in a subset of Project Achilles cell lines (Cheung et al. 2011). We introduced individual shRNAs into the A549 cancer cell line and performed immunoblotting on cell lysates to determine changes at the pro-

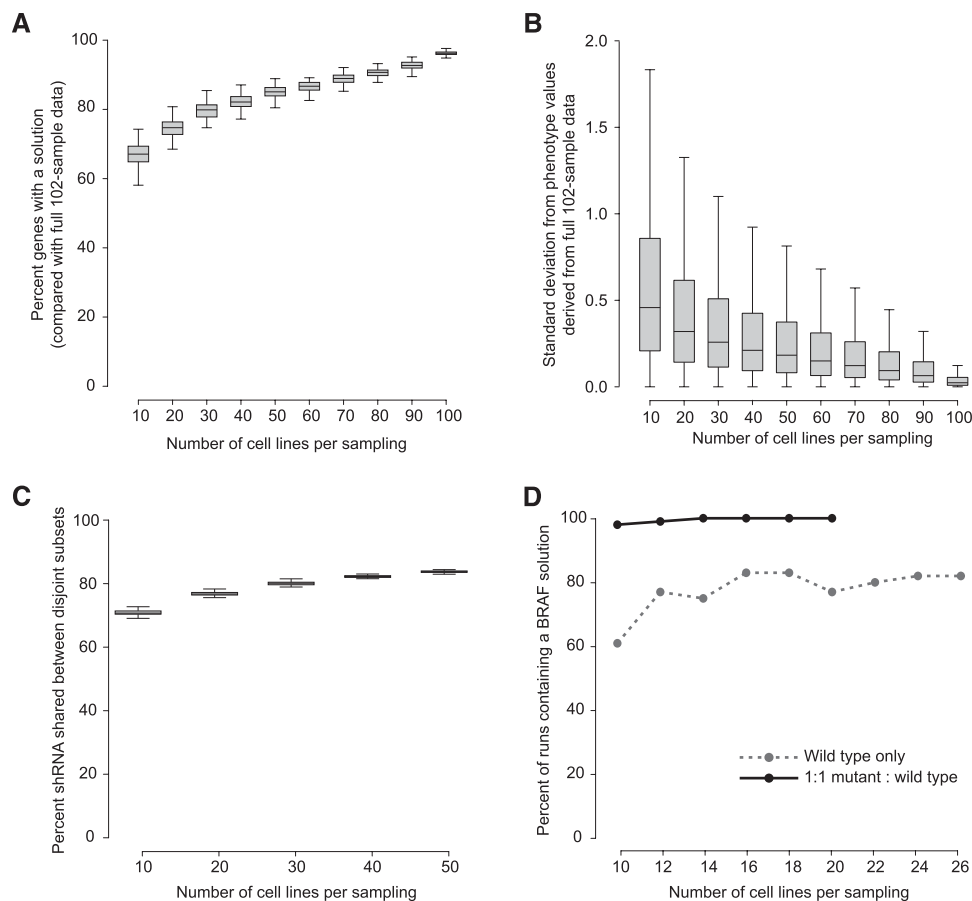


Figure 2. Influence of data set size and context on ATARiS results. (A,B) Robustness of ATARiS for data sets of smaller sample size. ATARiS results from 100 sets of randomly selected samples for each indicated sample size were compared with ATARiS results from the full 102-sample Achilles data set. We determined the percentage of genes with a solution in the Achilles data set results that are also represented in results from fewer samples (A). For 100 randomly selected genes, we also compared phenotype values in each sample to the values generated by ATARiS for the corresponding sample when using the full 102-sample data set (B). Standard deviations are based on phenotype values across all 102 samples for each gene independently. For each gene, differences between phenotype values from smaller data sets compared with the full data set are depicted in standard deviation units. (C) The robustness of shRNA selection by ATARiS is demonstrated by simulating independent screening data sets. One hundred pairs of *disjoint* sets of samples were randomly generated for the sample sizes indicated. Each set was independently analyzed by ATARiS. For each pair of sets, the overlap in shRNA used to generate solutions was determined. Boxplot displays the size of the overlap for each pair as a fraction of the average number of shRNAs used in the analysis of each set. P -value $< 2.2 \times 10^{-16}$ for all results, χ^2 test of independence. (D) ATARiS was used to analyze sets of samples that harbor either wild-type *BRAF* only, or an equal number of samples that harbor wild-type and mutant *BRAF*. One hundred randomly generated sets of samples were analyzed for each sampling size. The percentage of sets for which a *BRAF* solution was found is shown.

tein level. For *BRAF*, we observed that ATARiS consistency scores are high for shRNAs that reduce *BRAF* protein levels (Fig. 3A). We note that for *shBRAF-3* and *shBRAF-4*, which have similar consistency scores but different degrees of protein suppression, 40% protein suppression may be sufficient to produce functional effects, and the effects may not be enhanced by increased protein suppression. For *PIK3CA*, only two shRNAs (*shPIK3CA-1* and *shPIK3CA-2*) have high consistency scores and both result in increased suppression of *PIK3CA* protein levels (Fig. 3B). Our interpretation for the low consistency score of *shPIK3CA-3*, which effectively suppresses *PIK3CA* at the protein level, is that it may also have significant off-target effects. For *KRAS* and *MYC*, the effects of expressing individual shRNAs on protein levels also agreed with ATARiS consistency scores (Supplemental Fig. 6).

To test whether consistency scores reflect on-target gene suppression for many more genes, we compared ATARiS consistency scores to gene suppression assessed by qRT-PCR for 9050 of the shRNAs from the screening library (data not shown). We found

that shRNAs with significantly high consistency scores (corresponding to $FDR < 0.1$) suppress target gene mRNA levels to a greater degree than other shRNAs targeting the same gene (P -value = 0.003, χ^2 test) (Supplemental Fig. 7). We therefore concluded that genes with high consistency scores are likely to have a functionally relevant degree of gene suppression.

Gene phenotype values from the Achilles data set represent biological dependencies

We first validated individual ATARiS gene phenotype values, representing degree of dependency on each gene in the Achilles data set, by assessing whether they recapitulate known dependencies for the oncogenes *BRAF*, *PIK3CA*, and *KRAS*. For each oncogene, we calculated the area under the receiver operating characteristic curve (AUC) statistic (Mason and Graham 2002) to measure the degree to which the gene phenotype values discriminate between cell lines harboring a mutation versus those without mutation. We

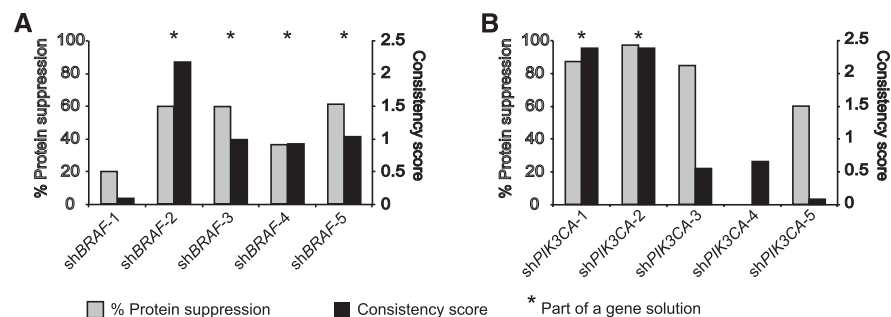


Figure 3. ATARIS consistency scores are associated with on-target gene suppression. Consistency scores computed by ATARIS and corresponding protein suppression levels by immunoblotting are shown for shRNAs targeting (A) *BRAF* and (B) *PIK3CA*. A higher consistency score represents greater confidence that the effects produced by the shRNA are due to suppression of the target gene. Immunoblotting for the effect of each shRNA compared with control shRNA was calculated based on quantification by ImageJ software. Shading of the axis labels corresponds to data bars of the same type. (*) Reagents used in the gene's ATARIS gene solution.

confirmed that cell lines harboring a mutation have significantly lower phenotype values for the respective gene, i.e., are more sensitive to gene suppression (Fig. 4A; P -value < 0.01; Mann-Whitney test). To show that our phenotype values may be meaningful for individual cell lines, we performed low-throughput viability assays on cell lines that span a range of *KRAS* phenotype values. We introduced two *KRAS*-specific shRNAs or a control shRNA into three *KRAS* wild-type and three *KRAS* mutant cell lines and measured cell proliferation/viability after 6 d using an ATP-luminescence assay. Indeed, the cell lines most sensitive to *KRAS* suppression were the ones that received the lowest *KRAS* phenotype values (Fig. 4B). Thus, we affirmed that ATARIS phenotype values reflect the relative effects of gene suppression between individual samples.

We reasoned that if ATARIS solutions are meaningful, then we should be able to “rediscover” the above oncogenic dependencies. Two-class comparisons between groups of cell lines with defined properties are currently a common application of this type of RNAi proliferation screen data, so we defined classes based on mutation status for each of *BRAF*, *PIK3CA*, and *KRAS* to identify differentially required genes (see Supplemental Data 4). For each analysis, we calculated the mean difference between mutant versus wild-type classes for each gene solution and estimated P -values from an empirically calculated null distribution by class permutation. *KRAS*, *BRAF*, and *PIK3CA* are each ranked first for being differentially required in their respective mutant class and remained significant after Benjamini-Hochberg adjustment for multiple hypothesis testing (q -value < 0.25) (Supplemental Table 2; Benjamini and Hochberg 1995). The fact that each class comparison yielded statistically meaningful results lends more validity to ATARIS phenotype values overall.

Finally, we sought to show that ATARIS phenotype values are valid for more than the specific oncogenes described above by defining classes using recurrent genomic alterations. Since commonly amplified or deleted regions in cancer are believed to include drivers that require unique cellular networks, we reasoned that more genes should be differentially essential when classes are defined by significant genomic alteration than when defined randomly. We defined significantly amplified and deleted peaks based on application of the genomic identification of significant targets in cancer (GISTIC) method (Beroukhi et al. 2007) to copy number data from the Cancer Cell Line Encyclopedia (CCLE)

(Barretina et al. 2012), a large collection of genomically annotated cancer cell lines, of which 76 were screened in Project Achilles (see Supplemental Data 5). For each peak present in at least six Achilles cell lines, we defined two classes based on the peak's presence or absence (101 total peaks) and calculated the difference in means between classes for every ATARIS gene solution to identify differentially essential genes (see Methods). Only 5% of analyses using randomly defined classes yielded more than four significantly differential genes, while 16% of analyses using GISTIC peaks do. We showed that significantly more essential genes were in classes defined by GISTIC peaks compared with random classes (P -value = 6×10^{-6} , Wilcoxon rank sum) (Fig. 4C; see also Supplemental Fig. 8a), supporting

the idea that ATARIS gene phenotype values likely reflect underlying biology.

Our analysis of the Achilles data results in multiple gene solutions for 9% of genes. To determine whether the supplementary solutions are also meaningful, we repeated the analysis using GISTIC peaks, as described above, after removing the first solution found for each gene (1030 solutions remaining). Indeed, we confirmed that even in this case, more significant solutions (FDR < 0.25) were identified when the analysis is performed using significant genomic alterations compared with random permutation of these alterations across samples (P -value = 0.0041; Wilcoxon rank sum) (see also Supplemental Fig. 8b), suggesting that the multiple solutions for each gene may have biological relevance.

ATARIS phenotype values enable novel approaches to biological discovery

In Figure 4C and Supplemental Table 2, we showed that ATARIS gene phenotype values could be effectively used for two-class comparisons, the focus of current analytic methods for multi-sample RNAi data. Additionally, ATARIS per-sample phenotype values expand the repertoire of downstream analyses from comparison between two classes to a range of additional possibilities. These include integrated analysis with other types of genomic data, e.g., gene expression, mutations, and genome copy number, which provide quantitative information for genes in each sample. We describe illustrative examples here.

Using phenotype values to identify genomic predictors of gene dependency

Genes that control the cell cycle G_1 restriction point are commonly altered in the cancer genome. Thus, we focused on ATARIS solutions for E2F transcription factors, well characterized in checkpoint regulation, to determine whether we could identify known (and unknown) genetic alterations related to E2F activation. We used an annotated sample feature list that includes significant amplification and deletion peaks, cell lineage, mutation, and copy number alterations of major oncogenes/tumor suppressors (see Methods; Supplemental Data 5). As expected, using the *E2F1* ATARIS solution, we found that *RB1* loss was one of the most highly associated features with *E2F1* dependence (Supplemental Fig. 9). In addition, when we examined *E2F3*, we found that *E2F3* dependence is significantly associated with an *E2F3*-containing

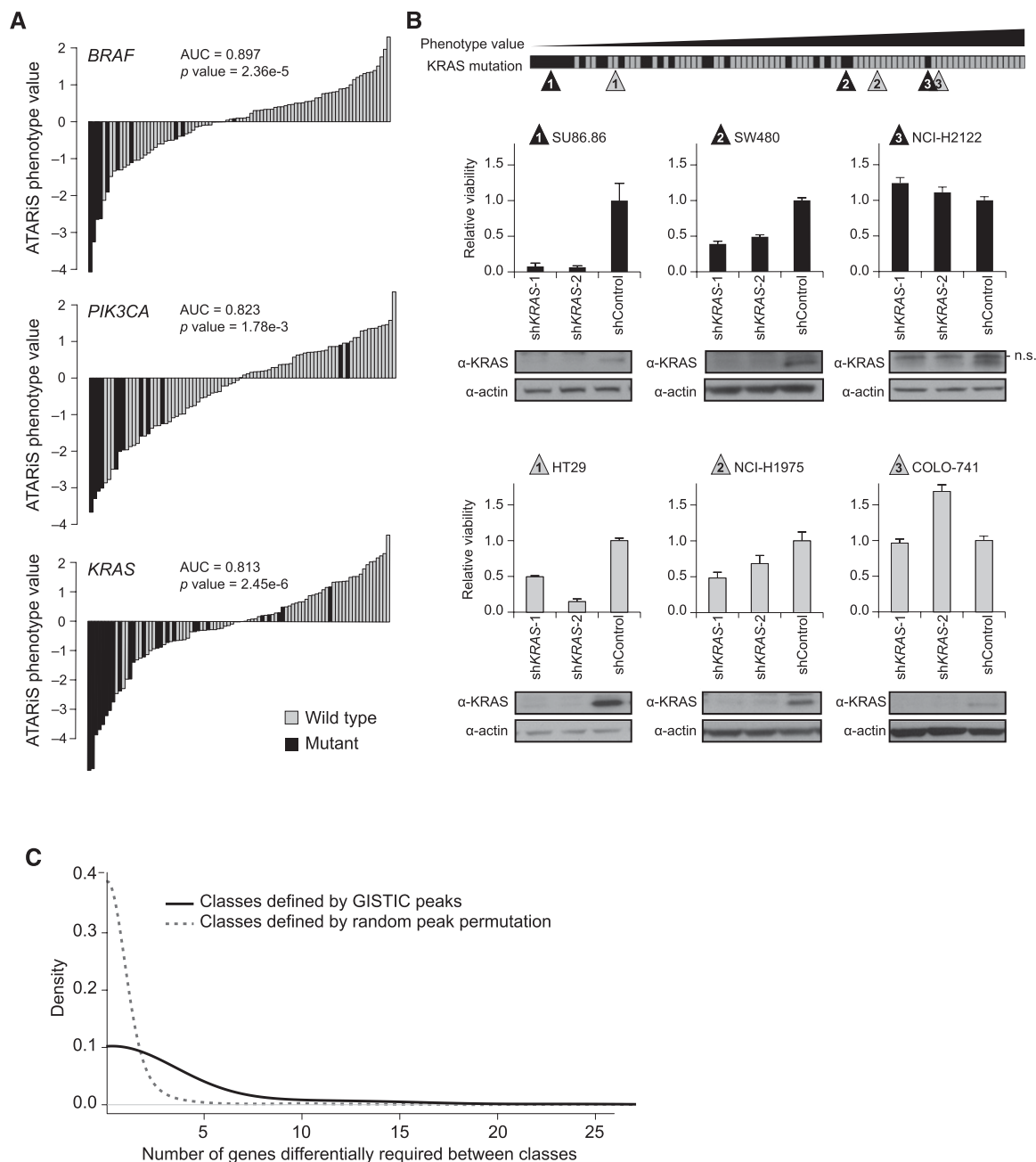


Figure 4. ATARiS gene phenotype values reflect biological dependencies. (A) Correspondence between gene mutation status and ATARiS phenotype values for *BRAF*, *PIK3CA*, and *KRAS*. Each vertical bar represents a single screened sample, colored by mutation status. In each plot, samples are ordered by increasing phenotype values. (AUC) Area under receiver operating characteristic curve. P -value, assessed by Mann-Whitney test. (B) Low-throughput validation of the relationship between gene phenotype scores and gene dependency. Six cell lines infected with sh*KRAS* were counted 4 d post-selection to determine cell number relative to infection with control shRNA. Immunoblots were performed using lysates from each sample collected at 2 d post-selection and stained using primary antibodies from Santa Cruz Biotechnology *KRAS* (sc-30) or actin (sc-1615). Immunoblot lanes correspond to bars in the graph directly above. Horizontal bar orders all cell lines with known *KRAS* mutation status in increasing order by ATARiS phenotype value, with validated samples marked by corresponding triangles. (Gray) *KRAS* wild-type; (black) *KRAS* mutant; (error bars) ± 1 SD ($n = 3$); (n.s.) non-specific band. (C) Genes differentially required in sample classes defined by recurrent amplification or deletion peaks. Recurrent genomic peaks were identified by GISTIC analysis across genomic data for samples from the Cancer Cell Line Encyclopedia. For each peak existing in at least six samples screened in Project Achilles ($n = 101$), two classes of samples were defined based on presence or absence of the peak. Genes that are differentially required in samples harboring the peak as compared with samples that do not (FDR < 0.25) were determined. The distribution of the number of significantly differential genes is shown. For comparison, the same analysis was performed using classes defined by random permutation of peak assignments.

amplification peak 6p22, a *MYC*-containing amplification peak at 8q24.1, as well as *RBI* copy-number loss (Fig. 5A)—all mechanisms that lead to *E2F3* activation (Dyson 1998; Leone et al. 2001;

Oeggerli et al. 2006). Similar analyses can be applied to other gene phenotype scores to elucidate genomic relationships with functional data.

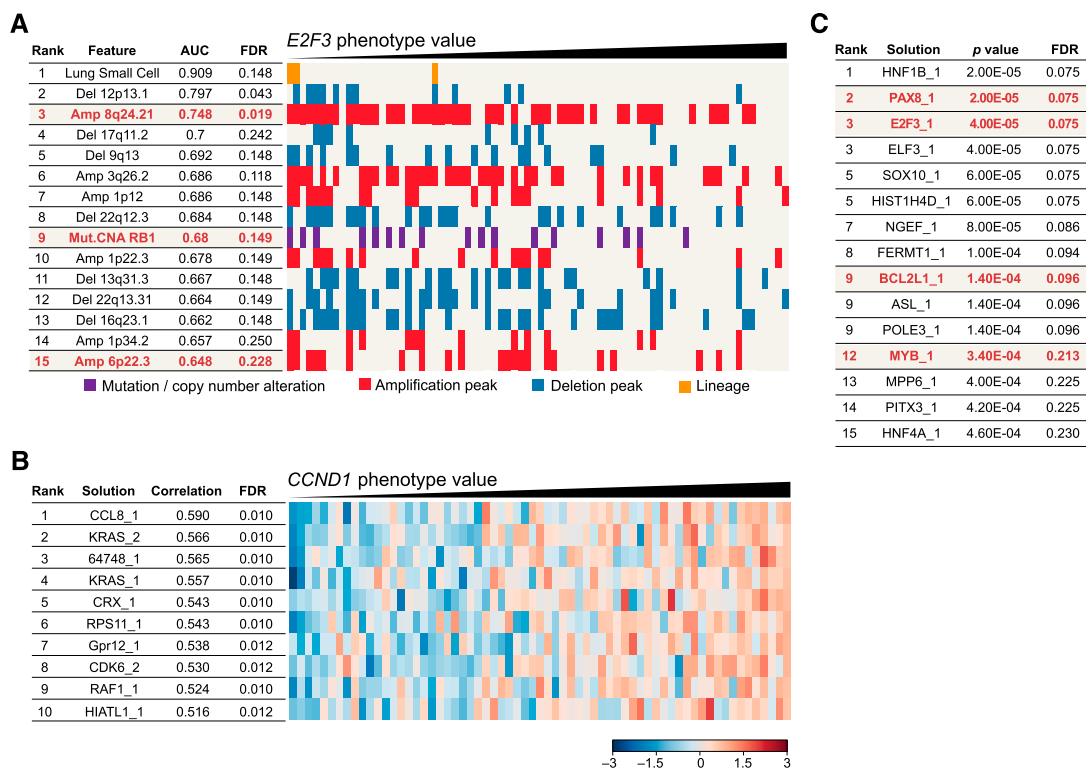


Figure 5. ATARIS phenotype values enable phenotype-based analyses for biological discovery. (A) Identifying genomic predictors of dependency on *E2F3*. Genomic features are shown ranked by their correspondence to *E2F3* phenotype values as measured by the area under receiver operating characteristic curve (AUC). (Amp) Amplification and (Del) deletion peaks, as determined by GISTIC. Columns correspond to individual cell lines. (Red) Genomic alterations pertinent to *E2F3*. (B) Correlations between gene phenotype value profiles to *CCND1* gene solution. Gene solutions are ranked by their similarity to the *CCND1* gene solution using the Pearson correlation coefficient. *P*-values were generated by permutation of sample labels. (C) Identifying significant cancer genes by integrating expression data and phenotype values. Gene solutions are ranked by increasing the Pearson correlation coefficient between the solution and expression data for the corresponding gene. Thus, genes that are essential in samples with high expression and less essential in samples with low expression are more negatively correlated, and receive higher ranks. *P*-values were calculated from a null distribution derived by permutation of sample labels. (Red) Previously reported gene dependencies in cancer. Numbers following gene names in B and C indicate gene solution number (see the Supplemental Data). (FDR) False discovery rate.

Identifying functional relationships between gene phenotype scores

We evaluated the potential of using correlations between different gene solutions to yield functionally meaningful gene relationships. In particular, we focused on the ATARIS solution for cyclin D1 (*CCND1*) since this gene serves as a critical mediator between the mitogenic pathway and cell cycle progression. The genes whose solutions most significantly correlated to *CCND1* solution included a gene that encodes the *CCND1*-binding partner CDK6 (no solution was available for the other cyclin D-binding partner CDK4) and also members of the mitogenic pathway, including *KRAS* and *RAF1* (Fig. 5B; Liu et al. 1995; Musgrove et al. 2011). Thus, we showed that using statistical relationships between ATARIS phenotype values allowed us to assess functional gene networks.

Identifying novel cancer-associated genes by integrating data sets on a per-sample basis

Currently, identification of novel cancer genes using RNAi data primarily consists of intersecting candidate gene lists separately derived from RNAi analysis and copy-number or expression data (Garraway et al. 2005; Cheung et al. 2011). Gene expression data are a powerful tool that has been integrated with copy-number data in a sample-specific manner to identify cancer drivers (Bussey et al. 2006). Since ATARIS enables integration of functional data with other data sets on a per-sample basis, we reasoned that we

could identify cancer drivers by looking for genes that are essential in samples where the gene is highly expressed. Thus, we independently calculated the correlation between each gene solution and the corresponding gene's expression values across 83 Project Achilles cell lines for which expression microarrays are available (Supplemental Table 4). We noted that previously reported cancer dependencies or oncogenes such as *PAX8* (Cheung et al. 2011; Li et al. 2011), *BCL2L1* (Beroukhi et al. 2010), *E2F3* (Oeggerli et al. 2006), and *MYB* (Ramsay and Gonda 2008) are significantly essential in samples that express the gene highly (Fig. 5C; Supplemental Table 5). Furthermore, we evaluated whether oncogenes, a subset of genes that might be expected to be essential in highly expressed cell lines, are enriched in our results. We determined that the list of known amplified oncogenes reported in Beroukhi et al. (2010) was significantly over-represented toward the top of our list (*P*-value = 5.38×10^{-06} ; Wilcoxon rank sum). For comparison, as expected, known tumor suppressors were not enriched (*P*-value = 0.84). These results suggested that other statistically significant genes from this analysis might contribute to malignant transformation, in particular, the top gene, *HNF1B*.

Characterization of HNF1B dependency

We found that cell lines that express high levels of *HNF1B* required *HNF1B* expression for proliferation/survival (Fig. 5C). A common

mechanism for increased gene expression in cancer is genomic amplification; thus, we also analyzed which genes involved in recurrent, focal genomic amplifications specifically scored as dependent in those samples (see Supplemental Methods). *HNF1B* again ranked at the top of this analysis (Supplemental Table 6), suggesting that *HNF1B* was one target of this amplification. We note that *HNF1B* is amplified in 23% of all cancers (<http://broadinstitute.org/tumorscape>).

Characterization of *HNF1B* dependency was made straightforward by directly applying ATARiS results to reveal which shRNAs are on-target and, specifically, which samples show higher dependency. We confirmed that the two *HNF1B*-specific shRNAs receiving the highest consistency scores suppressed *HNF1B* levels as assessed by immunoblotting (Fig. 6A). Furthermore, exogenous expression of *HNF1B* in cells harboring a doxycycline-inducible *HNF1B* 3'-UTR-specific shRNA (sh*HNF1B*-1) abrogated the cell death induced by expressing the *HNF1B* 3'-UTR-specific shRNA alone (Fig. 6B), confirming that the observed shRNA effects were specific. We used a panel of cell lines to confirm that *HNF1B* protein expression was correlated to *HNF1B* phenotype values (Supplemental Fig. 10). Finally, we used cell lines that expressed high levels of *HNF1B* to confirm that they were indeed sensitive to *HNF1B* suppression by the two *HNF1B*-specific shRNAs as compared with control shRNA. For comparison, we showed that DLD-1 and an immortalized cell line, HA1E (Hahn et al. 1999), neither of which harbor amplifications involving *HNF1B* nor express high levels of the gene, are insensitive to *HNF1B* suppression (Fig. 6C). ATARiS phenotype values allowed us to identify two additional cell lines SLR-21 and 786-O that had low phenotype values for *HNF1B*, but for which we did not have corresponding copy-number data. We confirmed that these cell lines were also dependent on *HNF1B* and had corresponding genomic copy-number gain (Fig. 6C; Supplemental Fig. 11).

To determine whether *HNF1B* expression is essential for tumor maintenance in vivo, we performed xenograft experiments by implanting HT29 colon cancer cells subcutaneously after expression of control or *HNF1B*-specific shRNAs (Fig. 6D). In the initial 2 wk, the xenografts with suppressed *HNF1B* showed marked growth impairment ($n = 3$; $P < 0.01$, one-tailed Student's *t*-test). Four weeks post-injection, their growth increased, likely due to re-activation of *HNF1B* expression (Fig. 6D), suggesting that *HNF1B* expression was critical for growth.

Finally, we sought to determine whether expression of *HNF1B* transforms human cell lines. Specifically, we introduced *HNF1B* or *LacZ* cDNA into HA1EM cells, which are immortalized, non-tumorigenic human embryonic kidney cells that are transformed upon addition of oncogenes *AKT* or *IKKBE* (also known as *IKKE*; Boehm et al. 2007). Expression of *HNF1B* conferred the ability for anchorage-independent growth, a marker of cell transformation

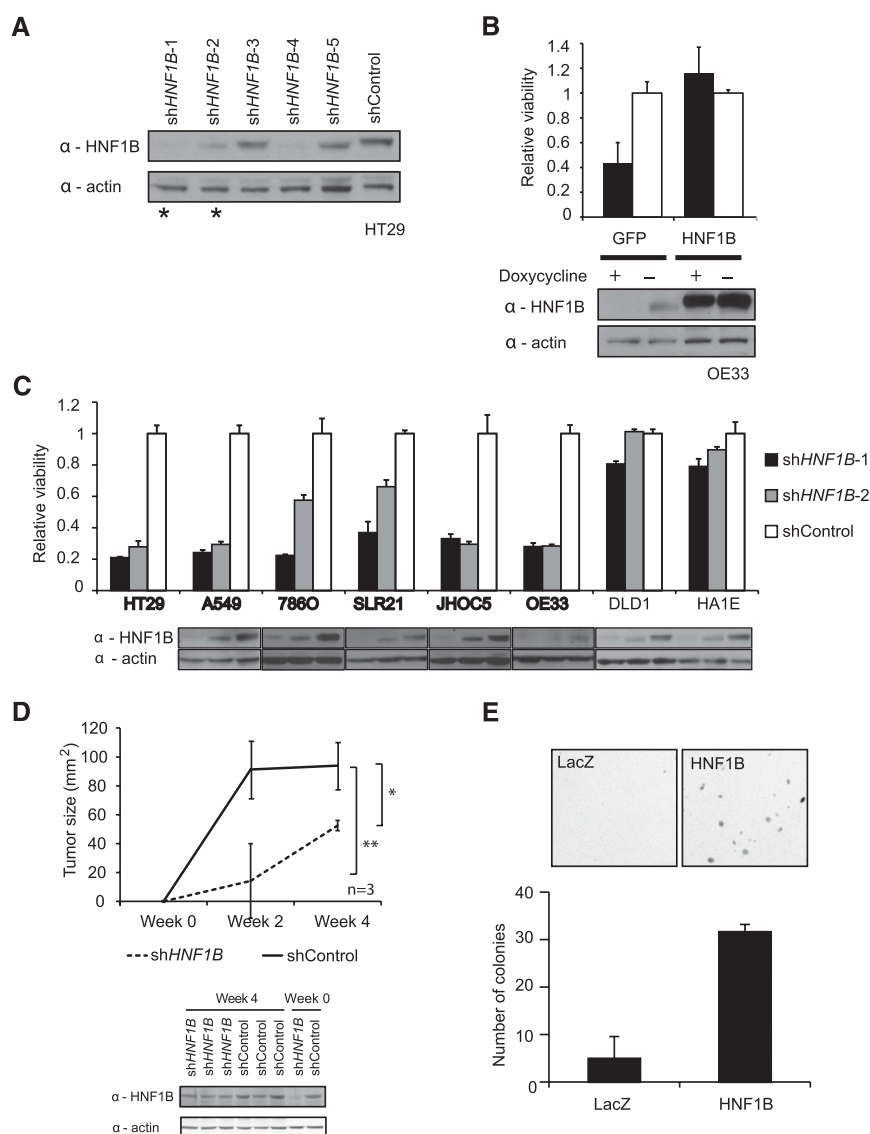


Figure 6. Characterizing the role of *HNF1B* in cancer. (A) Immunoblot of *HNF1B* after expression of five independent shRNAs designed to target *HNF1B*. (*) The two shRNAs incorporated into the ATARiS solution, which also have the highest consistency scores. (B) Cell viability upon exogenous expression of *HNF1B* or *GFP* in an *HNF1B*-sensitive cell line OE33 with stable integration of doxycycline-inducible expression of sh*HNF1B*-1. Each bar in the graph corresponds to the immunoblot lane directly below. (C) Relative viability of a panel of cell lines upon suppression of control or two *HNF1B*-specific shRNAs. Cell lines with high levels of *HNF1B* are shown in bold text. Each bar in the graph corresponds to the immunoblot lane directly below. (D) *HNF1B*-sensitive cell line HT29 expressing sh*HNF1B*-1 or shControl was implanted subcutaneously into immunocompromised mice. Sh*HNF1B*-1 was used for all experiments since it has potent effects and is specific for *HNF1B*, as shown in panels A and B. Tumor volume was monitored biweekly, and lysates were collected pre-implantation and from tumors at 4 wk. (*) P -value < 0.05 ; (**) P -value < 0.01 (one-tailed Student's *t*-test). (E) *HNF1B* or *LacZ* was expressed in HA1EM cells and anchorage-independent growth was determined. Representative photos shown after 6 wk. (Error bars) ± 1 SD ($n = 3$).

(Fig. 6E). Together, these observations—that *HNF1B* is amplified in human cancers, transforms immortalized cells, and is essential for those cancer cell lines that harbor increased *HNF1B* copy number—provide strong evidence that *HNF1B* is an oncogene.

Discussion

One key advance of ATARiS lies in the ability to distinguish reagents with on-target effects and reject reagents with significant off-target effects by mining patterns across multisample screens. ATARiS reagent consistency scores may be interpreted as a *P*-value that estimates the confidence in each reagent and thus enables the use of ATARiS in the selection of reagents for validation studies and as an aid in the design and refinement of RNAi libraries. Based on our work with these shRNAs, we anticipate that we lack more than one effective shRNA for a fraction of targeted genes. ATARiS will allow us to interrogate these situations and to develop improved libraries in the future. For example, one might create additional shRNA reagents for genes that lack solutions or create sublibraries only containing shRNAs involved in ATARiS solutions. As RNAi libraries include more reagents per gene, and as screens include more samples, the ability of ATARiS to correctly identify on-target reagents will also improve.

ATARiS gene phenotype values are an inherently different metric from previous gene scores for RNAi. Existing methods determine gene candidates in a manner that is dependent on a user-defined desired phenotype, whereas ATARiS aims to summarize the data available for each gene in an unbiased way. For example, with a “frequency approach,” it is theoretically possible for a single gene in a sample to be a candidate for both a positive and a negative phenotype, whereas the ATARiS phenotype value provides a single metric for that gene. For methods such as RSA and SSMD, gene scores are influenced by the distribution of reagents toward or away from a *desired* phenotype. In comparison, ATARiS attempts to determine the best subset of reagents that describe the actual gene-level effect. Furthermore, it incorporates information across all screened samples instead of using data from each sample independently. Another major difference between ATARiS and previous work is that gene phenotype values are *relative* to the samples screened instead of absolute, as is the case in RSA and SSMD. Finally, unlike RNAi analysis methods for two-class comparisons such as RIGER (Barbie et al. 2009) and “second best” (Cheung et al. 2011), which are primarily used to determine a single value representing each gene’s differential effect across classes, ATARiS phenotype values describe the effect of each gene in each individual sample.

ATARiS is analogous to the approaches used by methods such as RMA (Irizarry et al. 2003) and dChip (Li and Hung Wong 2001) for microarray data analysis in that gene scores are evaluated by incorporating multiple probes/reagents and excluding problematic ones. It is similar to dChip specifically in that a multiplicative model is fit to the set of probes/reagents. However, ATARiS differs from both methods in order to account for the unique attributes of RNAi data. For example, while most microarray probes are assumed to generally agree, the majority of RNAi reagents do not. Supplemental Figure 12 shows how the correlation coefficients between data from shRNA pairs targeting the same gene are only marginally higher than the coefficients from random shRNA pairs. Thus, ATARiS implements an empirical null distribution to determine correlations that are significantly above background. Another difference is that ATARiS considers multiple solutions for each gene, because varying degrees of gene suppression by distinct reagents may produce different effect profiles across samples.

We identified *HNF1B* as an oncogene by examining the correlation between each gene’s expression and ATARiS gene solution in a sample-specific manner. On the other hand, when we examined the correlation between measurements of each individual shRNA and corresponding gene expression values, we found *HNF1B* shRNAs spread throughout the ranked results: one ranked in the top 10 shRNAs, two in the top 500, and two ranked considerably lower. In the same way that using coregulated sets of genes, rather than individual genes, can increase the signal in transcription profiling data (Subramanian et al. 2005), ATARiS uses multiple shRNAs, enriching for on-target effects, to increase the signal from individual RNAi reagents. Furthermore, experimental validation of *HNF1B* was made straightforward by using ATARiS consistency scores to predict the shRNA reagents driving the cellular phenotype and by using gene phenotype scores to identify specific samples to examine.

HNF1B is located near the known oncogene *ERBB2*. However, our observations indicate that *HNF1B* independently induces cell transformation. It remains possible that *HNF1B* may cooperate with *ERBB2* to drive transformation in a manner analogous to what has been observed for *YAP1* and *CIAP1*, which reside in a single amplicon in hepatocellular cancer (Zender et al. 2006). *HNF1B* has previously been described as an essential gene in ovarian clear cell carcinoma, where it is highly expressed (Tsuchiya et al. 2003), and genome-wide association studies have associated SNPs in the *HNF1B* locus with risk for prostate and endometrial cancers (Schumacher et al. 2011; Spurdle et al. 2011), although *HNF1B* may also be epigenetically inactivated in certain contexts (Terasawa et al. 2006). Developmentally, *HNF1B* is required for visceral endoderm formation (Barbacci et al. 1999) and proper development of the genitourinary tract (Ryffel 2001; Bellanne-Chantelot et al. 2005), but appears to be dispensable in adult tissue (Verdeguer et al. 2009), making it a reasonable candidate for therapeutic targeting.

Although high-throughput shRNA viability screens are discussed here, ATARiS can be applied to any screen in which multiple, redundant reagents produce different observed outcomes in multiple samples. For example, ATARiS would apply to a screen that uses siRNA reagents or measures a phenotype other than viability. A similar approach can potentially be applied to small molecule screening where multiple target-specific compounds are assayed across different samples. In addition, since ATARiS analyzes the data of each gene independently, it can be effectively applied to screens that target a small number of genes as long as multiple samples are screened. One such example is validation screens, where screeners have prior expectation that the reagents screened will produce an effect. In contrast, methods such as RIGER (Barbie et al. 2009) construct a null distribution from all screened reagents, requiring many reagents to be screened, while the majority of them may have no effect on the measured phenotype.

Despite the fact that parallel screens continue to grow in size, we recognize that not all data sets will be as large as Achilles. When the sample size is small, the number of genes with solutions decreases. This is due to the loss of genetic heterogeneity between samples and reduced statistical power to discriminate true effects from noise. The user-defined significance threshold allows for tailoring to specific applications. In our analyses, we chose a relatively lax threshold (0.15) such that greater numbers of gene solutions will be available for analysis. Because our significance calculations are based on an empirical null distribution, a more stringent threshold will provide increased confidence in the solutions generated (i.e., lower false-positive rate) (see also Supplemental Fig. 2). Furthermore, the compatibility of results from independent

data sets suggests that investigators screening single or few samples may use ATARIS results (such as consistency scores) derived from larger data sets to improve their ability to assess reagent performance and gene effects in their screened samples.

We hope that by providing a foundation for interpreting RNAi gene suppression effects as quantifiable values in individual samples, we will aid functional genomics in reaching its full potential. We illustrated several analytic methods that are enabled by ATARIS and note that investigators have already begun to map phenotype-based gene networks (Amit et al. 2009; Horn et al. 2011). Nevertheless, much remains to be explored. ATARIS results from the Project Achilles and Marcotte et al. (2012) data sets will be useful for deeper analysis, but we also believe that the application of ATARIS to other screening data sets, large and small, will yield novel insights. ATARIS is available online at <http://broadinstitute.org/ataris>.

Methods

Statistical modeling

Given the measurements of phenotypic effects produced by a set of RNAi reagents designed to target the same gene G , ATARIS generates a consensus profile that represents the effect of suppressing G in each screened sample relative to the other samples. Let n denote the number of screened samples and p denote the number of reagents targeting G for which the measurements are given. Let \mathbf{X} denote a $p \times n$ matrix with each element x_{ij} representing the observed phenotypic effect produced by reagent i in sample j . Because we are only interested in finding the *relative* effects of gene suppression, we median-center each row of \mathbf{X} to obtain $\mathbf{X}^* = \mathbf{X} - \mu \mathbf{1}_n^T$, where μ is a vector of length p such that $\mu_i = \text{median}(x_{i,*})$ and $\mathbf{1}_n$ is a vector of 1's of length n .

Let \mathbf{c} denote a vector of length n representing the consensus profile for \mathbf{X}^* and let \mathbf{e} denote a vector of length p consisting of a relative effect size for each RNAi reagent. ATARIS models each measurement x_{ij}^* as a product of its corresponding (unknown) relative effect size e_i and phenotypic effect c_j , such that an approximation for \mathbf{X}^* is given by $\hat{\mathbf{X}}^* = \mathbf{e}\mathbf{c}^T$, and we set $\max(\mathbf{e}) = 1$ for identifiability. We can then formulate the problem of finding the values for \mathbf{e} and \mathbf{c} as the following optimization problem:

$$\begin{aligned} & \text{minimize}_{\mathbf{e}, \mathbf{c}} \|\mathbf{X}^* - \mathbf{e}\mathbf{c}^T\|_1 \text{ subject to } \max(\mathbf{e}) = 1, \\ & \text{where } \|\mathbf{A}\|_1 = \sum_i \sum_j |a_{ij}|. \end{aligned}$$

This criterion, which can also be seen as a rank-1 matrix factorization problem, although not convex, is bilinear in \mathbf{c} and \mathbf{e} (i.e., with \mathbf{c} fixed, it is linear in \mathbf{e} , and vice versa). To optimize it, we use the following iterative algorithm of alternating minimizations (Csiszar and Tusnady 1984).

We begin by initializing \mathbf{c} with the mean values of \mathbf{X}^* in each sample:

$$c_j \leftarrow \frac{1}{p} \sum_i x_{ij}^* \text{ for } j \in 1, \dots, n.$$

We then update \mathbf{e} and \mathbf{c} repeatedly until convergence:

$$\mathbf{e} \leftarrow \arg \min_{\mathbf{e}} \|\mathbf{X}^* - \mathbf{e}\mathbf{c}^T\|_1$$

$$\mathbf{c} \leftarrow \arg \min_{\mathbf{c}} \|\mathbf{X}^* - \mathbf{e}\mathbf{c}^T\|_1.$$

The elements of \mathbf{e} and \mathbf{c} are updated in an element-wise manner, i.e.,

$$e_i \leftarrow \arg \min_{\hat{e}} \sum_j |x_{ij}^* - \hat{e}c_j| \text{ for } i \in 1, \dots, p$$

and similarly

$$c_j \leftarrow \arg \min_{\hat{c}} \sum_i |x_{ij}^* - e_i\hat{c}| \text{ for } j \in 1, \dots, n.$$

Each such assignment can be viewed as a problem of finding a weighted median, which can be solved efficiently.

We cease iterating when a decrease of <1% in $\|\mathbf{X}^* - \mathbf{e}\mathbf{c}^T\|_1$ is observed. For the Achilles and Marcotte et al. (2012) data sets, we found that convergence almost always occurs after fewer than 20 iterations. Because this optimization problem is not convex, we are not guaranteed to find a global minimum. To test the performance of the optimization in practice, we ran it multiple times with random initialization values and found that the variations in the parameters estimated are minimal. Finally, to identify the solution set

$$\mathbf{e} \leftarrow \frac{1}{\max(\mathbf{e})} \cdot \mathbf{e} \text{ and } \mathbf{c} \leftarrow \max(\mathbf{e}) \cdot \mathbf{c}.$$

Refinement of RNAi reagent subset

For each gene, ATARIS tries to identify subsets of its RNAi reagents that produce similar effects across the screened samples. Given a set of reagents R_G targeting gene G , we iteratively refine R_G until we identify a subset $\hat{R}_G \subseteq R_G$ that consists of reagents whose profiles (i.e., effects across the samples) are all similar to the consensus profile computed for R_G . We then consider \hat{R}_G to be a consistent set and use its consensus profile as a gene solution, as described in the main text. We begin by computing a consensus profile for the reagent set $R_G^* = R_G$. We then evaluate the following criteria to determine whether R_G^* is a consistent set of reagents:

1. For each reagent $r \in R_G^*$, the Spearman correlation coefficient ρ_r between the reagent profile $(x_{r,1}, x_{r,2}, \dots, x_{r,n})$ and the consensus profile \mathbf{c} must be greater than the 85th percentile of the corresponding Spearman correlation coefficients similarly generated from data of random reagent sets of size $|R_G^*|$. (Note: This 0.15 significance threshold can be adjusted depending on the user's desired confidence and properties of the data. See also Supplemental Fig. 2 for an analysis of the influence of this threshold on the number of solutions found.)
2. All the reagents in R_G^* must have a relative effect size e_r of at least 0.3, i.e., $e_r \geq 0.3, \forall r \in R_G^*$. We therefore favor reagents whose effects have comparable magnitudes, avoiding the inclusion of reagents whose effects are mainly due to noise (assuming that noise magnitudes are similar across reagents).

If either criterion is not fulfilled, we remove one reagent from the set R_G^* as follows:

1. If any reagent $r \in R_G^*$ does not satisfy criterion (2), we discard the one with the lowest effect magnitude e_r .
2. Otherwise, we discard the reagent $r \in R_G^*$ with the lowest Spearman correlation coefficient between its profile and the consensus profile \mathbf{c} .

The refinement process is repeated until R_G^* is consistent or until it consists of only one reagent, in which case, we conclude that there is no solution to the set R_G of reagents. Our refinement algorithm is greedy so that it is scalable and can be used to analyze RNAi screens performed using reagent libraries that have a large number of reagents per gene.

Assignment of consistency scores

We determine a consistency score for each RNAi reagent of a given gene G based on its similarity to other reagents targeting G . For a reagent r that is part of a consistent set \hat{R}_G (and hence was used to generate a gene solution), we determine the similarity of its profile $(x_{r,1}, x_{r,2}, \dots, x_{r,n})$ to the profiles of all other reagents in \hat{R}_G by computing the corresponding Spearman correlation coefficients. We estimate a P -value for each correlation coefficient based on an empirical null distribution of Spearman correlation coefficients of random pairs of reagent profiles. We combine the P -values associated with r into a single significance estimate, $p\text{-value}_r^*$, using Stouffer's method (Stouffer et al. 1949; Whitlock 2005). The consistency score of r is defined as $-\log_{10}(p\text{-value}_r^*)$.

For reagents that were not used to generate a gene solution, we proceed as above to estimate a consistency score and a P -value but use the set R_G of all reagents that target gene G .

Analysis of the Project Achilles data set

Genome-scale pooled shRNA screens to identify genes essential for proliferation in 102 cancer cell lines were performed using a lentivirally delivered pool of 54,020 shRNAs targeting 11,217 genes (Cheung et al. 2011). Each cell line was infected in quadruplicates and propagated for at least 16 population doublings. The abundance of shRNA constructs was measured by microarray hybridization and raw .CEL files from custom Affymetrix barcode arrays were processed with a modified version of dCHIP software. ShRNAs that had an overlap of >3 nucleotides to other screened shRNAs were removed ($n = 679$). The \log_2 fold change in shRNA abundances for each cell line at the conclusion of the screening relative to the initial plasmid DNA reference pool was calculated (Cheung et al. 2011). The \log_2 fold change data were then normalized by a robust Z -score normalization (i.e., centering around the median and scaling by the Median Absolute Deviation). The median value was used to collapse data from replicates. The resulting data set is provided as Supplemental Data 1. Each data point represents the abundance of one shRNA construct within one cell line as compared with the initial abundance of that shRNA construct in the initial plasmid DNA pool. The ATARiS analysis ignored data for shRNAs targeting non-human genes ($n = 4$) and genes targeted by only one shRNA ($n = 8$).

Analysis of the Marcotte et al. data set

Marcotte et al. (2012) performed genome-wide pooled shRNA screens to identify genes essential for cancer cell survival and proliferation in 72 breast, pancreatic, and ovarian cancer cell lines. They used a lentiviral shRNA library targeting ~16,000 genes with 78,432 shRNAs, of which 50,981 shRNAs were also used in the Project Achilles screens. We obtained shRNA-level shARP (shRNA Activity Ranking Profile) scores for all the shRNAs and cell lines screened through the COLT-Cancer database (Koh et al. 2012) and considered them to represent the observed phenotypic effects. We computed a robust Z -score for each cell line separately and discarded data for two cell lines (OVCA1369_TR, HPDE) that showed aberrant score distributions. We ran ATARiS on the normalized values using the same parameters used for the analysis of the Achilles data set. ATARiS found gene solutions for 8406 (54.4%) of the genes using data from 29,731 (39.2%) of the shRNAs (Supplemental Table 3; Supplemental Data 6, 7).

Cell culture

All cancer cell lines were cultured in RPMI-1640 (Invitrogen) supplemented with 10% FBS (Sigma-Aldrich) and 1% streptomycin

and penicillin. HA1E and HA1EM immortalized lines were cultured in alpha-MEM (Invitrogen) supplemented with 10% FBS.

Lentiviral infection

Lentivirus containing shRNAs targeting *BRAF*, *PIK3CA*, *KRAS*, *MYC*, *HNF1B* and controls targeting *GFP* or *LacZ* for validation of ATARiS consistency scores were purchased directly from The RNAi Consortium (Root et al. 2006) for gene suppression validation studies. Lentivirus for *KRAS* and *HNF1B* dependency experiments were produced as previously described (Barbie et al. 2009). See Supplemental Table 6 for detailed shRNA identities. Cells were infected in media containing 8 $\mu\text{g}/\text{mL}$ Polybrene and a 1:10 dilution of virus. Infected cells were selected with 2 $\mu\text{g}/\text{mL}$ puromycin for 48 h.

Low-throughput assessment of cell viability

Cells were replated at 50,000 cells/well post-infection and post-selection in triplicate in 12-well plates. Wells were counted 4 d later by ViaCell.

Immunoblotting

Cell lysates collected 72 h post-infection were run on 4%–12% Bis-Tris gel (Invitrogen NuPAGE) and transferred to nitrocellulose membrane for immunoblotting. Primary antibodies were obtained from Santa Cruz (*KRAS* sc-30, *BRAF* sc-5284, *MYC* sc-764, *HNF1B* sc-7411, β -actin sc-1615) and Cell Signaling (*PI3 Kinase 110 alpha* #4255). Immunoblots for *BRAF* and *PIK3CA* protein were visualized by infrared imaging (LI-COR). Quantification was performed by ImageJ software (<http://rsb.info.nih.gov/ij>).

Two-class comparisons

For each two-class comparison, ATARiS gene phenotype values were used to calculate a mean for each class for each gene solution. The difference of means between the classes was used as a scoring metric, and P -values were estimated based on a null distribution generated by 50,000 class label permutations. Q -values were generated by the Benjamini-Hochberg method (Benjamini and Hochberg 1995).

Annotation of cell line genomic features

We constructed a matrix of genomic features for cell lines that had matched genomic data from the Cancer Cell Line Encyclopedia (CCLE) (Barretina et al. 2012). As previously described, features include mutational status, tumor tissue lineage, regions of recurrent copy-number gain or loss (derived from GISTIC), and combined gene mutation and copy number amplification (for oncogenes) or combined mutation and copy number deletion (for tumor suppressors). GISTIC regions were assessed across all available CCLE cell lines, of which 76 were screened in Achilles. Amplification and deletion of specific genes were defined by relative log fold copy-number value >0.25 or <-0.25, respectively. All data are represented as binary values, with 1 representing the presence of the indicated feature in the sample. Refer to Supplemental Data 5 for the full feature matrix.

Anchorage-independent growth assay

HA1EM cells infected with lentiviral expression plasmid pLX-304 with desired genes were selected for 5 d in 10 $\mu\text{g}/\text{mL}$ blasticidin. Cells were seeded in triplicate at 2.5×10^4 cells per well in 0.4% top

agar (Difco) in six-well plates. The bottom agar was 0.6% agar (Difco) supplemented with 20% FBS. Macroscopic images were collected of each well, and colonies were counted using CellProfiler (<http://cellprofiler.org>).

Xenograft assay

HT29 infected with lentiviral plasmid PLKO.1 sh*HNF1B*-1 or shControl was expanded for 4 d before subcutaneous implantation into immunocompromised mice (Taconic, CrTac:NCr-Foxn1nu). Two million cells were implanted into each of three sites per mouse. Tumor growth was monitored every 2 wk by digital caliper measurement of tumor diameter. The approximate cross-sectional area was calculated.

Data access

ATARiS can be run online on user-provided data through the GenePattern computational genomics suite (Reich et al. 2006) accessible on the ATARiS website (<http://broadinstitute.org/ataris>). The website also includes all data sets used to obtain the results described in this manuscript.

Competing interest statement

W.C.H. and R.B. are consultants for Novartis Pharmaceuticals.

Acknowledgments

We thank the Broad Institute Project Achilles investigators for the data from RNAi pooled screening, and the Broad-Novartis Cancer Cell Line Encyclopedia investigators for providing cell lines for validation. We thank D. Nijhawan, J. Rosenbluh, D. Barbie, and F. Verdeguer for helpful discussions. We thank L. Solomon and L. Gaffney for assistance with figure design. This work was supported by grants from the NIH/NCI (RC2 CA148268, P01 CA050661, U54 CA112962, R01 CA109467), the H.L. Snyder Medical Foundation, and the Ivy Research Foundation.

Author contributions: A.T., D.D.S., S.G., B.A.W., A.A.M., P.T., D.E.R., W.C.H., and J.P.M. conceptualized the method. W.C.H. and J.P.M. directed the work. A.T. implemented the method. D.D.S. performed validation experiments. A.T., D.D.S., B.A.W., S.G., N.S., S.E.S., T.I.Z., and R.B. performed computational analyses. N.S., L.A.G., S.G., and D.E.R. contributed data. A.T., D.D.S., W.C.H., and J.P.M. wrote the manuscript. B.A.W., P.T., A.A.M., N.S., and D.E.R. edited the manuscript.

References

- Amit I, Garber M, Chevrier N, Leite AP, Donner Y, Eisenhaure T, Guttman M, Grenier JK, Li W, Zuk O, et al. 2009. Unbiased reconstruction of a mammalian transcriptional network mediating pathogen responses. *Science* **326**: 257–263.
- Barbacci E, Reber M, Ott MO, Breillat C, Huetz F, Cereghini S. 1999. Variant hepatocyte nuclear factor 1 is required for visceral endoderm specification. *Development* **126**: 4795–4805.
- Barbie D, Tamayo P, Boehm J, Kim S, Moody S, Dunn I, Schinzel A, Sandy P, Meylan E, Scholl C, et al. 2009. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* **462**: 108–112.
- Bard F, Casano L, Mallabiabarrena A, Wallace E, Saito K, Kitayama H, Guizzunti G, Hu Y, Wandler F, Dasgupta R, et al. 2006. Functional genomics reveals genes involved in protein secretion and Golgi organization. *Nature* **439**: 604–607.
- Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, Wilson CJ, Lehár J, Kryukov GV, Sonkin D, et al. 2012. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**: 603–607.
- Bellanne-Chantelot C, Clauin S, Chauveau D, Collin P, Daumont M, Douillard C, Dubois-Laforgue D, Dusselier L, Gautier JF, Jadoul M, et al. 2005. Large genomic rearrangements in the hepatocyte nuclear factor-1 (TCF2) gene are the most frequent cause of maturity-onset diabetes of the young type 5. *Diabetes* **54**: 3126–3132.
- Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc Ser B Methodol* **57**: 289–300.
- Beroukhim R, Getz G, Nghiemphu L, Barretina J, Hsueh T, Linhart D, Vivanco I, Lee JC, Huang JH, Alexander S, et al. 2007. Assessing the significance of chromosomal aberrations in cancer: Methodology and application to glioma. *Proc Natl Acad Sci* **104**: 20007–20012.
- Beroukhim R, Mermel CH, Porter D, Wei G, Raychaudhuri S, Donovan J, Barretina J, Boehm JS, Dobson J, Urashima M, et al. 2010. The landscape of somatic copy-number alteration across human cancers. *Nature* **463**: 899–905.
- Birmingham A, Anderson EM, Reynolds A, Ilsley-Tyree D, Leake D, Fedorov Y, Baskerville S, Maksimova E, Robinson K, Karpilow J, et al. 2006. 3' UTR seed matches, but not overall identity, are associated with RNAi off-targets. *Nat Methods* **3**: 199–204.
- Boehm J, Zhao J, Yao J, Kim S. 2007. Integrative genomic approaches identify IKBKE as a breast cancer oncogene. *Cell* **129**: 1065–1079.
- Brough R, Frankum JR, Sims D, Mackay A, Mendes-Pereira AM, Bajrami I, Costa-Cabral S, Rafiq R, Ahmad AS, Cerone MA, et al. 2011. Functional viability profiles of breast cancer. *Cancer Discov* **1**: 260–273.
- Bussey KJ, Chin K, Lababidi S, Reimers M, Reinhold WC, Kuo W-L, Gwadry F, Ajay, Kouros-Mehr H, Fridlyand J, et al. 2006. Integrating data on DNA copy number with gene expression levels and drug sensitivities in the NCI-60 cell line panel. *Mol Cancer Ther* **5**: 853–867.
- Cheung HW, Cowley GS, Weir BA, Boehm JS, Rusin S, Scott JA, East A, Ali LD, Lizotte PH, Wong TC, et al. 2011. Systematic investigation of genetic vulnerabilities across cancer cell lines reveals lineage-specific dependencies in ovarian cancer. *Proc Natl Acad Sci* **108**: 12372–12377.
- Chung N, Zhang XD, Kreamer A, Locco L, Kuan P-F, Bartz S, Linsley PS, Ferrer M, Strulovici B. 2008. Median absolute deviation to improve hit selection for genome-scale RNAi screens. *J Biomol Screen* **13**: 149–158.
- Collinet C, Stöter M, Bradshaw CR, Samusik N, Rink JC, Kenski D, Habermann B, Buchholz F, Henschel R, Mueller MS, et al. 2010. Systems survey of endocytosis by multiparametric image analysis. *Nature* **464**: 243–249.
- Csiszar I, Tuszynski G. 1984. Information geometry and alternating minimization procedures. *Statist Decisions* (Supplement) **1984**: 205–237. http://www.mit.edu/~6.454/www_fall_2002/shaas/Csiszar.pdf. <http://research.microsoft.com/apps/pubs/default.aspx?id=64832>.
- Cullen BR. 2006. Enhancing and confirming the specificity of RNAi experiments. *Nat Methods* **3**: 677–681.
- Dyson N. 1998. The regulation of E2F by pRB-family proteins. *Genes Dev* **12**: 2245–2262.
- Echeverri C, Beachy P, Baum B, Boutros M. 2006. Minimizing the risk of reporting false positives in large-scale RNAi screens. *Nat Methods* **3**: 777–779.
- Elbashir SM, Harborth J, Lendeckel W, Yalcin A, Weber K, Tuschl T. 2001. Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411**: 494–498.
- Garraway L, Widlund H, Rubin M, Getz G. 2005. Integrative genomic analyses identify MITF as a lineage survival oncogene amplified in malignant melanoma. *Nature* **436**: 117–122.
- Hahn WC, Counter CM, Lundberg AS, Beijersbergen RL, Brooks MW, Weinberg RA. 1999. Creation of human tumour cells with defined genetic elements. *Nature* **400**: 464–468.
- Hirsch AJ. 2010. The use of RNAi-based screens to identify host proteins involved in viral replication. *Future Microbiol* **5**: 303–311.
- Horn T, Sandmann T, Fischer B, Axelsson E. 2011. Mapping of signaling networks through synthetic genetic interaction analysis by RNAi. *Nat Methods* **8**: 341–346.
- Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP. 2003. Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res* **31**: e15.
- Jackson AL, Bartz SR, Schelter J, Kobayashi SV, Burchard J, Mao M, Li B, Cavet G, Linsley PS. 2003. Expression profiling reveals off-target gene regulation by RNAi. *Nat Biotechnol* **21**: 635–637.
- Jackson AL, Burchard J, Schelter J, Chau BN, Cleary M, Lim L, Linsley PS. 2006. Widespread siRNA “off-target” transcript silencing mediated by seed region sequence complementarity. *RNA* **12**: 1179–1187.
- Kittler R, Pelletier L. 2008. Systems biology of mammalian cell division. *Cell Cycle* **7**: 2123–2128.
- Koh JL, Brown KR, Sayad A, Kasimer D, Ketela T, Moffat J. 2012. COLT-Cancer: Functional genetic screening resource for essential genes in human cancer cell lines. *Nucleic Acids Res* **40**: D957–D963.

- König R, Chiang C-Y, Tu BP, Yan SF, Dejesus PD, Romero A, Bergauer T, Orth A, Krueger U, Zhou Y, et al. 2007. A probability-based approach for the analysis of large-scale RNAi screens. *Nat Methods* **4**: 847–849.
- Leone G, Sears R, Huang E, Rempel R, Nuckolls F, Park CH, Giangrande P, Wu L, Saavedra HI, Field SJ, et al. 2001. Myc requires distinct E2F activities to induce S phase and apoptosis. *Mol Cell* **8**: 105–113.
- Li C, Hung Wong W. 2001. Model-based analysis of oligonucleotide arrays: Model validation, design issues and standard error application. *Genome Biol* **2**: RESEARCH0032.
- Li CG, Nyman JE, Braithwaite AW, Eccles MR. 2011. PAX8 promotes tumor cell growth by transcriptionally regulating E2F1 and stabilizing RB protein. *Oncogene* **30**: 4824–4834.
- Liu JJ, Chao JR, Jiang MC, Ng SY, Yen JJ, Yang-Yen HF. 1995. Ras transformation results in an elevated level of cyclin D1 and acceleration of G₁ progression in NIH 3T3 cells. *Mol Cell Biol* **15**: 3654–3663.
- Luo B, Cheung HW, Subramanian A, Sharifnia T, Okamoto M, Yang X, Hinkle G, Boehm JS, Beroukhi R, Weir BA, et al. 2008. Highly parallel identification of essential genes in cancer cells. *Proc Natl Acad Sci* **105**: 20380–20385.
- Marcotte R, Brown KR, Suarez F, Sayad A, Karamboulas K, Krzyzanowski PM, Sircoulomb F, Medrano M, Fedyszyn Y, Koh JLY. 2012. Essential gene profiles in breast, pancreatic, and ovarian cancer cells. *Cancer Discov* **2**: 172–189.
- Marine S, Bahl A, Ferrer M, Buehler E. 2012. Common seed analysis to identify off-target effects in siRNA screens. *J Biomol Screen* **17**: 370–378.
- Mason SJ, Graham NE. 2002. Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: Statistical significance and interpretation. *Q J R Meteorol Soc* **128**: 2145–2166.
- Müller P, Kutenkeuler D, Gesellchen V, Zeidler MP, Boutros M. 2005. Identification of JAK/STAT signalling components by genome-wide RNA interference. *Nature* **436**: 871–875.
- Musgrove EA, Caldon CE, Barraclough J, Stone A, Sutherland RL. 2011. Cyclin D as a therapeutic target in cancer. *Nat Rev Cancer* **11**: 558–572.
- Oeggerli M, Schraml P, Ruiz C, Bloch M, Novotny H, Mirlacher M, Sauter G, Simon R. 2006. E2F3 is the main target gene of the 6p22 amplicon with high specificity for human bladder cancer. *Oncogene* **25**: 6538–6543.
- Ramsay RG, Gonda TJ. 2008. MYB function in normal and cancer cells. *Nat Rev Cancer* **8**: 523–534.
- Reich M, Liefeld T, Gould J, Lerner J, Tamayo P, Mesirov JP. 2006. GenePattern 2.0. *Nat Genet* **38**: 500–501.
- Root DE, Hacohen N, Hahn WC, Lander ES, Sabatini DM. 2006. Genome-scale loss-of-function screening with a lentiviral RNAi library. *Nat Methods* **3**: 715–719.
- Ryffel GU. 2001. Mutations in the human genes encoding the transcription factors of the hepatocyte nuclear factor (HNF)1 and HNF4 families: Functional and pathological consequences. *J Mol Endocrinol* **27**: 11–29.
- Schumacher FR, Berndt SI, Siddiq A, Jacobs KB, Wang Z, Lindstrom S, Stevens VL, Chen C, Mondul AM, Travis RC, et al. 2011. Genome-wide association study identifies new prostate cancer susceptibility loci. *Hum Mol Genet* **20**: 3867–3875.
- Sigoillot FD, Lyman S, Huckins JF, Adamson B. 2012. A bioinformatics method identifies prominent off-targeted transcripts in RNAi screens. *Nat Methods* **9**: 363–366.
- Spurdle AB, Thompson DJ, Ahmed S, Ferguson K, Healey CS, O'Mara T, Walker LC, Montgomery SB, Dermizakis ET, Fahey P, et al. 2011. Genome-wide association study identifies a common variant associated with risk of endometrial cancer. *Nat Genet* **43**: 451–454.
- Stouffer SA, Suchman EA, DeVinney LC, Star SA, Williams RMJ. 1949. *The American soldier, Vol. 1: Adjustment during army life*, 1st ed. Princeton University Press, Princeton.
- Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, et al. 2005. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. **102**: 15545–15550.
- Terasawa K, Toyota M, Sagae S, Ogi K, Suzuki H, Sonoda T, Akino K, Maruyama R, Nishikawa N, Imai K, et al. 2006. Epigenetic inactivation of TCF2 in ovarian cancer and various cancer cell lines. *Br J Cancer* **94**: 914–921.
- Tsuchiya A, Sakamoto M, Yasuda J, Chuma M, Ohta T, Ohki M, Yasugi T, Taketani Y, Hirohashi S. 2003. Expression profiling in ovarian clear cell carcinoma: Identification of hepatocyte nuclear factor-1 β as a molecular marker and a possible molecular target for therapy of ovarian clear cell carcinoma. *Am J Pathol* **163**: 2503–2512.
- Verdeguer F, Le Corre S, Fischer E, Callens C, Garbay S, Doyen A, Igarashi P, Terzi F, Pontoglio M. 2009. A mitotic transcriptional switch in polycystic kidney disease. *Nat Med* **16**: 106–110.
- Whitlock MC. 2005. Combining probability from independent tests: The weighted Z-method is superior to Fisher's approach. *J Evol Biol* **18**: 1368–1373.
- Zender L, Spector MS, Xue W, Flemming P, Cordon-Cardo C, Silke J, Fan S-T, Luk JM, Wigler M, Hannon GJ, et al. 2006. Identification and validation of oncogenes in liver cancer using an integrative oncogenomic approach. *Cell* **125**: 1253–1267.
- Zhang XD, Ferrer M, Espeseth AS, Marine SD, Stec EM, Crackower MA, Holder DJ, Heyse JF, Strulovici B. 2007. The use of strictly standardized mean difference for hit selection in primary RNA interference high-throughput screening experiments. *J Biomol Screen* **12**: 497–509.
- Zhang XD, Santini F, Lacson R, Marine SD, Wu Q, Benetti L, Yang R, McCampbell A, Berger JP, Toolan DM, et al. 2011. cSSMD: Assessing collective activity for addressing off-target effects in genome-scale RNA interference screens. *Bioinformatics* **27**: 2775–2781.

Received May 23, 2012; accepted in revised form December 18, 2012.