Functional characterization of somatic mutations in cancer using network-based inference of protein activity

Mariano J Alvarez[1,2,9], Yao Shen[1,2,9], Federico M Giorgi[1], Alexander Lachmann[1], B Belinda Ding[3], B Hilda Ye[3] & Andrea Califano[1,4–8]

Identifying the multiple dysregulated oncoproteins that contribute to tumorigenesis in a given patient is crucial for developing personalized treatment plans. However, accurate inference of aberrant protein activity in biological samples is still challenging as genetic alterations are only partially predictive and direct measurements of protein activity are generally not feasible. To address this problem we introduce and experimentally validate a new algorithm, virtual inference of protein activity by enriched regulon analysis (VIPER), for accurate assessment of protein activity from gene expression data. We used VIPER to evaluate the functional relevance of genetic alterations in regulatory proteins across all samples in The Cancer Genome Atlas (TCGA). In addition to accurately infer aberrant protein activity induced by established mutations, we also identified a fraction of tumors with aberrant activity of druggable oncoproteins despite a lack of mutations, and vice versa. *In vitro* assays confirmed that VIPER-inferred protein activity outperformed mutational analysis in predicting sensitivity to targeted inhibitors.

Cancer initiation and progression are driven by aberrant activity of oncoproteins working in concert to regulate critical tumor hallmark programs[1]. Pharmacological inhibition of aberrantly activated oncoproteins can elicit oncogene dependency[2], which motivates the development and use of targeted inhibitors in precision cancer medicine. Activating genetic alterations have thus emerged as important candidate drug targets. Yet activating mutations represent only one of many possible ways to dysregulate the activity of an oncoprotein. Genetic and epigenetic events in cognate binding partners[3], competitive endogenous RNAs[4] and upstream regulators[5] can all

contribute to aberrant activity of oncoproteins. Thus, although cells with activating mutations in a specific oncogene are generally more sensitive to corresponding targeted inhibitors, cells lacking these mutations may also present equivalent sensitivity[6,7]. Conversely, an activating mutation is not guaranteed to induce aberrant protein activity, due to autoregulatory mechanisms and epigenetic allele silencing. A more universal and systematic methodology for the accurate and reproducible assessment of protein activity would complement our ability to identify targeted therapy responders based on mutational analysis, especially because most cancer patients have no actionable oncogene mutations[8].

While gene expression data are ubiquitous in cancer research[9–12], methods for the genome-wide assessment of protein activity are still elusive. Existing methods to measure protein abundance based on arrays[13] or mass spectrometry technologies[14] are still labor-intensive, costly, and either cover a small fraction of the proteomic landscape or require large amounts of tissue. More importantly, these methods provide only an indirect measure of protein activity, because the latter is determined by a complex cascade of events, including protein synthesis, degradation, post-translational modification, complex formation and subcellular localization[15] (**Fig. 1a**). It is ultimately unclear whether protein activity may be directly and systematically assessed by any individual assay.

We propose that the expression of the transcriptional targets of a protein, collectively referred to as its regulon, represent an optimal multiplexed reporter of its activity (**Fig. 1a**). Although this concept is not new and was initially proposed for transcription factors[16], it has not been successfully demonstrated in mammalian cells. There are currently no experimentally validated methods to accurately assess the activity of arbitrary proteins in individual samples based on the expression of their regulon genes. Reasons for this include a lack of accurate and context-specific protein regulon models, the largely pleiotropic nature of transcriptional regulation, and a lack of methodologies to assess statistical significance from single samples. This severely limits the ability to understand the functional effect of mutations on protein activity and to identify candidate responders to targeted inhibitors based on aberrant protein activity rather than mutations.

We have previously shown that regulon analysis, using the master regulator inference algorithm (MARINa), can help identify aberrantly activated tumor drivers[17–21]. However, this requires multiple samples representing the same tumor phenotype and cannot be used to assess aberrant protein activity from individual samples. To address this

challenge, we introduce a new regulatory-network based approach to infer protein activity from single gene expression profiles (VIPER; **Supplementary Table 1**). We first discuss development, optimization and validation of VIPER. Then we introduce a statistical framework to allow single-sample analysis, without loss of robustness or generality. Finally, we describe the use of VIPER to evaluate all non-silent somatic mutations in TCGA samples and report the aberrant activity of all oncogenes listed in the Catalogue Of Somatic Mutations In Cancer (COSMIC)[22] on an individual sample basis. VIPER can be used to systematically assess aberrant activity of oncoproteins for which high-affinity inhibitors are available, independent of their mutational state, thus establishing them as valuable therapeutic targets on an individual patient basis. The analysis is fully general and may be trivially extended to study the role of germ-line variants in dysregulating protein activity. We implemented VIPER as an R-system package available through Bioconductor.

RESULTS
The algorithm
VIPER infers protein activity by systematically analyzing expression of the protein's regulon, which is strongly tumor-context-dependent[20] (**Fig. 1b**). We used the algorithm for the reconstruction of accurate cellular networks (ARACNe[23]; Online Methods) to systematically infer regulons from tissue-specific gene expression data (**Fig. 1b** and **Table 1**). Although any algorithm or experimental assay providing accurate, tissue-specific assessments of protein regulons should be equally effective, we found that ARACNe outperformed competing algorithms that derive regulons from genome-wide chromatin immunoprecipitation (ChIP) databases, including ChIP enrichment analysis (ChEA)[24] and Encyclopedia of DNA Elements (ENCODE)[25] and

literature curated Ingenuity networks[26] (see below). We extended ARACNe to detect maximum information path targets (Online Methods), as originally proposed in ref. 21, to allow identification of regulons that report on the activity of proteins representing indirect regulators of transcriptional target expression, such as signaling proteins.

VIPER is based on a probabilistic framework that directly integrates target 'mode of regulation', that is, whether targets are activated or repressed (**Fig. 1b** and **Supplementary Figs. 1** and **2**), statistical confidence in regulator-target interactions (**Fig. 1b**) and target overlap between different regulators (pleiotropy) (**Fig. 1d**) to compute the enrichment of a protein's regulon in differentially expressed genes (Online Methods). Several methods exist for gene enrichment analysis, including the Fisher's exact test[27], T-profiler[28] and gene set enrichment analysis (GSEA)[27,29–31]. In all these methods, the contribution of individual genes to signature enrichment is binary (i.e., 0 or 1). In contrast, VIPER uses a fully probabilistic yet efficient enrichment analysis framework, supporting seamless integration of genes with different likelihoods of representing activated, repressed or undetermined targets, and probabilistic weighting of low vs. high-likelihood protein targets. To achieve this, we introduce analytic rank-based enrichment analysis (aREA) a statistical analysis based on the mean of ranks (**Fig. 1c** and Online Methods). Differential protein activity is quantitatively inferred as the normalized enrichment score computed by aREA.

Systematic assessment of VIPER's performance
We first tested VIPER's ability to correctly infer loss of protein activity following RNA interference (RNAi)-mediated silencing of a gene. *MEF2B*[32], *FOXM1* (ref. 17), *MYB*[17] and *BCL6*, were silenced in lymphoma cells and *STAT3* (ref. 18) in glioblastoma cells by RNAi-mediated
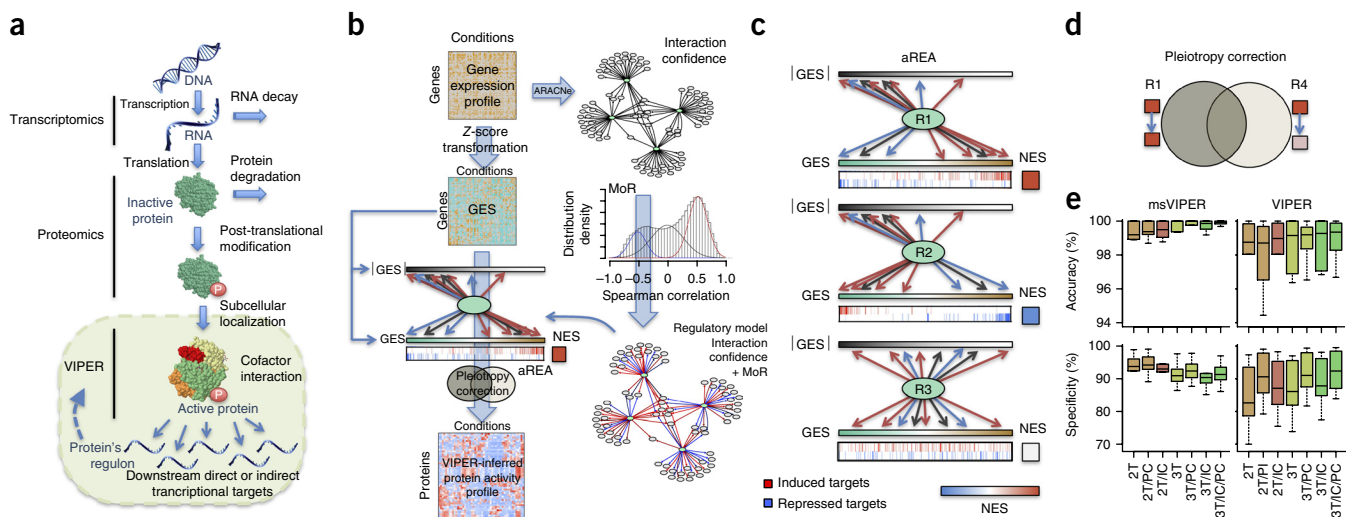

Figure 1 Overview of the VIPER method. (**a**) Schematic of molecular layers profiled: transcriptomics, used to measure steady-state mRNA levels; proteomics, used to quantify protein levels, including some defined post-translational isoforms; and VIPER, used to infer protein activity based on the protein's regulon, reflecting the abundance of the active protein isoform, including post-translational modifications, proper subcellular localization and interaction with co-factors. (**b**) VIPER workflow in which a regulatory model is generated from ARACNe-inferred context-specific interactome and mode of regulation (MoR) is computed from the correlation between regulator and target genes. Single-sample gene expression signatures (GES) are computed from genome-wide expression data, and transformed into regulatory protein activity profiles by the aREA algorithm. |GES|, absolute value of GES. NES, normalized enrichment score. (**c**) Three possible scenarios for the aREA analysis are increased, decreased or no change in activity for three regulatory proteins (R1, R2 and R3). GES and |GES| are indicated by color scale bars; induced and repressed target genes according to the regulatory model are indicated by blue and red vertical lines. (**d**) Pleiotropy correction, performed by evaluating whether the enrichment of a given regulon (R4) is driven by genes coregulated by a second regulator (R1). (**e**) Accuracy (relative rank for the silenced protein) and specificity (fraction of proteins inferred as differentially active at $P < 0.05$) for the six benchmark experiments (**Table 2**) with VIPER based on multiple-sample gene expression signatures (msVIPER) and single-sample gene expression signatures (VIPER). Colors indicate implementations of the aREA algorithm: two-tail (2T) and three-tail (3T), interaction confidence (IC) and pleiotropy correction (PC).

silencing (**Table 2**). We included multiple cell lines and distinct RNAi silencing protocols and profiling platforms to avoid bias associated with these variables. We used these data to benchmark different regulatory model attributes and enrichment methods.

We assessed three metrics: (i) the *P*-value-based rank of the silenced gene (accuracy), (ii) the total number of statistically significant regulators inferred by VIPER (specificity), and (iii) the overall *P* value of the silenced gene. The enrichment analysis methods

Table 1 Interactomes and the data sets used to reverse-engineer them

Tissue type	Data set			Interactome		
	Samples	Platform	Reference	Regulator	Targets	Interactions
B cell	254	HG-U95Av2	23	633 (TFs)	6,403	173,539
B cell	264	HG-U133plus2	34	1,223 (TFs)	13,007	327,837
Glioblastoma	176	HG-U133A	48	835 (TFs)	8,263	256,965
Bladder carcinoma	241	RNA-seq	TCGA	1,813 (TFs)	20,006	245,871
				666 (co-TFs)	18,739	181,730
				3,455 (Sig)	20,441	317,127
Breast carcinoma	1,037	RNA-seq	TCGA	1,813 (TFs)	20,428	249,501
				666 (co-TFs)	20,220	217,916
				3,455 (Sig)	20,515	366,924
Colon adenocarcinoma	434	RNA-seq	TCGA	1,813 (TFs)	20,462	294,725
				666 (co-TFs)	19,742	204,682
				3,456 (Sig)	20,492	369,870
Head and neck squamous cell carcinoma	424	RNA-seq	TCGA	1,813 (TFs)	20,452	319,799
				666 (co-TFs)	19,874	212,214
				3,456 (Sig)	20,520	395,966
Kidney renal clear cell carcinoma	506	RNA-seq	TCGA	1,813 (TFs)	20,474	355,932
				666 (co-TFs)	20,080	259,151
				3,456 (Sig)	20,522	429,651
Lung adenocarcinoma	488	RNA-seq	TCGA	1,813 (TFs)	20,405	341,285
				666 (co-TFs)	19,832	214,048
				3,456 (Sig)	20,528	472,933
Lung squamous cell carcinoma	482	RNA-seq	TCGA	1,813 (TFs)	20,426	342,737
				666 (co-TFs)	19,948	221,178
				3,453 (Sig)	20,498	397,774
Ovarian serous cystadenocarcinoma	262	RNA-seq	TCGA	1,813 (TFs)	20,261	247,063
				666 (co-TFs)	19,082	150,949
				3,456 (Sig)	20,459	334,906
Prostate adenocarcinoma	297	RNA-seq	TCGA	1,813 (TFs)	20,215	228,977
				666 (co-TFs)	19,599	180,315
				3,456 (Sig)	20,466	315,155
Rectum adenocarcinoma	163	RNA-seq	TCGA	1,810 (TFs)	18,506	236,899
				666 (co-TFs)	16,939	173,579
				3,455 (Sig)	19,773	332,088
Stomach adenocarcinoma	238	RNA-seq	TCGA	1,808 (TFs)	22,017	267,138
				661 (co-TFs)	20,984	194,782
				3,442 (Sig)	22,458	438,054
Thyroid carcinoma	498	RNA-seq	TCGA	1,813 (TFs)	20,478	333,725
				666 (co-TFs)	20,038	225,544
				3,369 (Sig)	20,511	408,356
Uterine corpus endometrial carcinoma	517	RNA-seq	TCGA	1,813 (TFs)	20,471	350,994
				666 (co-TFs)	20,190	237,518
				3,456 (Sig)	20,527	501,212
Glioblastoma multiforme	154	RNA-seq	TCGA	1,811 (TFs)	18,354	259,025
				660 (co-TFs)	16,655	157,230
				3,455 (Sig)	19,616	393,595
Low grade glioma	370	RNA-seq	TCGA	1,813 (TFs)	20,357	328,373
				666 (co-TFs)	19,558	228,634
				3,455 (Sig)	20,463	372,802
Skin cutaneous melanoma	374	RNA-seq	TCGA	1,813 (TFs)	20,475	281,486
				666 (co-TFs)	19,656	177,388
				3,453 (Sig)	20,501	418,136
Sarcoma	105	RNA-seq	TCGA	1,715 (TFs)	14,262	142,041
				620 (co-TFs)	10,920	72,486
				3,024 (Sig)	15,552	177,063

TFs, transcription factors; co-TFs, co-transcriptional regulators; Sig, signaling proteins.

Table 2 Benchmark experiments

Cell line	Knockdown gene	Technology	Replicates	Profile platform	DEG[a] at $P < 0.01$
P3HR1 (lymphoma)	MEF2B	shRNA[b]	5	HG-U95Av2	960
ST486 (lymphoma)	FOXM1	shRNA[b]	3	HG-U95Av2	276
	MYB	shRNA[b]	3	HG-U95Av2	469
OCI-Ly7 (lymphoma)	BCL6	siRNA[c]	3	HG-U133p2	646
Pfeiffer (lymphoma)	BCL6	siRNA[c]	3	HG-U133p2	1,311
SNB19 (glioma)	STAT3	siRNA[c]	6	Illumina HT12v3	501

[a]Differentially expressed genes. [b]Short hairpin RNA. [c]Small interfering RNA.

we tested were aREA, Fisher exact test (one-tail FET)[18] and one-tail GSEA. We also tested extensions of FET and GSEA to account for the mode of regulation of a target gene (two-tail FET and two-tail GSEA), which were previously implemented in our MARINa algorithm[17,18,20]. Use of a three-tail aREA (aREA-3T), accounting for target mode of regulation, confidence and pleiotropic regulation, systematically outperformed all other approaches (**Fig. 1e**, **Supplementary Figs. 3a** and **4**, **Supplementary Table 2** and **Supplementary Note**). Thus, we selected the aREA-3T method as the methodology of choice for the VIPER algorithm. The experimentally silenced proteins encoded by MYB, BCL6, STAT3, FOXM1, MEF2B and BCL6, were ranked as the 1st, 1st, 1st, 2nd, 3rd and 3rd most significantly inactivated proteins among all those tested, respectively (**Supplementary Fig. 3a** and **Supplementary Table 2**). The small number of additional transcription factors inferred by aREA was enriched in differentially expressed genes and thus likely represents downstream targets of the silenced regulators or RNAi off-target effects (**Supplementary Fig. 5**).

To evaluate suitability of ARACNe-inferred regulons for use in VIPER, we benchmarked VIPER performance with non-context-specific regulons, as assembled from ChIP-sequencing (ChIP-seq) data in ChEA[24] and in ENCODE[25]. We also benchmarked VIPER against the upstream regulator module of Ingenuity Pathway Analysis[26]. ARACNe-based VIPER outperformed these approaches (**Supplementary Fig. 3c** and **Supplementary Note**). The alternative methods/models correctly assessed protein activity decrease only for FOXM1 following its silencing. Among the five tested transcription factors, FOXM1 was the only one representing a core cell cycle regulator, whose regulon is strongly conserved across multiple tissue contexts (**Supplementary Fig. 3d**), thus not requiring use of context-specific regulatory models.

From each experiment we generated signatures using the control-sample-based Z transformation (Online Methods) to allow analysis of individual samples (**Table 2**). Results from single-sample analyses were virtually identical to those obtained with the multisample version of VIPER (**Fig. 1e**, **Supplementary Fig. 3b** and **Supplementary Table 3**), suggesting that single-sample analysis produces robust and highly reproducible results. We then performed several additional benchmarks to assess the specific improvements owing to the aREA probabilistic analysis, compared to GSEA, and to assess the overall ability of the algorithm to correctly identify proteins whose activity was modulated by RNAi and small-molecule perturbations, or whose abundance was quantified by reverse-phase protein arrays (**Supplementary Figs. 6–9**,
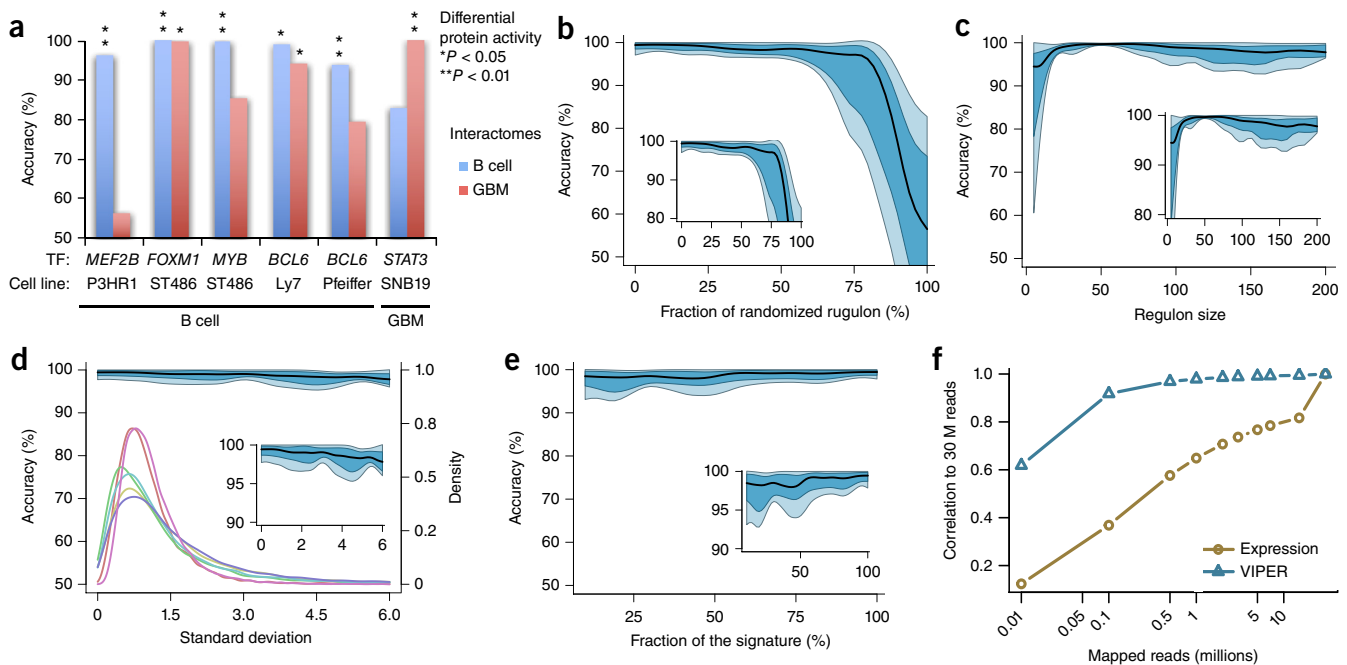

Figure 2 Effect of network and signature quality on VIPER results. (**a–c**) Effect of network quality on VIPER accuracy (rank position of the silenced gene) when using a non-tissue-matched interactome, by computing protein activity with a B-cell interactome (B cell) or glioma interactome (GBM) (**a**); when the network was degraded by partially randomizing the regulons (**b**); or when the regulon size was progressively reduced (**c**). Accuracy is shown across the six benchmark experiments as a bar plot (**a**) or by the median (black line), IQR (blue) and the lowest and highest data points still inside 1.5 times the IQR away from the quartiles (light blue), resembling a box-and-whiskers plots (continuous boxplots; **b**,**c**). (**d**) VIPER accuracy (continuous boxplot) for progressive signature degradation obtained by addition of Gaussian noise. Probability density plots show the distribution of gene expression variance for the six benchmark data sets (density). (**e**) VIPER accuracy (continuous boxplot) for reduced signature coverage obtained by randomly removing genes. (**f**) Average correlation between 30 M mapped-reads-based gene expression (expression) or VIPER-inferred protein activity signatures and the corresponding signatures computed from lower-depth RNA-seq. Signatures were obtained from 100 breast carcinoma samples profiled by TCGA. Insets show magnification.

Figure 3 Reproducibility of VIPER results. (**a**) Distribution of correlation coefficients computed between all possible pairs of gene expression signatures (yellow) or VIPER protein activity signatures (cyan) for samples of the same B cell phenotype, including normal (indicated by asterisks: GC, germinal center reaction; M, memory and N, peripheral blood B cell) and pathologic (B-CLL, B cell chronic lymphocytic leukemia; BL, Burkitt lymphoma; HCL, hairy cell leukemia; PEL, primary effusion lymphoma; MCL, mantle cell lymphoma; FL, follicular lymphoma) phenotypes. The number of samples per phenotype is indicated on top. (**b**) Probability density for the relative rank position of the most upregulated gene (mRNA), relatively abundant protein (RPPA) or activated protein (VIPER), identified in each profiled basal breast carcinoma sample, across all the remaining profiled samples. The horizontal line and number under it indicates the distribution mode. (**c**) Probability density for the relative rank position of the top ten most upregulated genes (yellow) or VIPER-inferred activated proteins (cyan), identified from fresh-frozen samples on the corresponding FFPE samples.
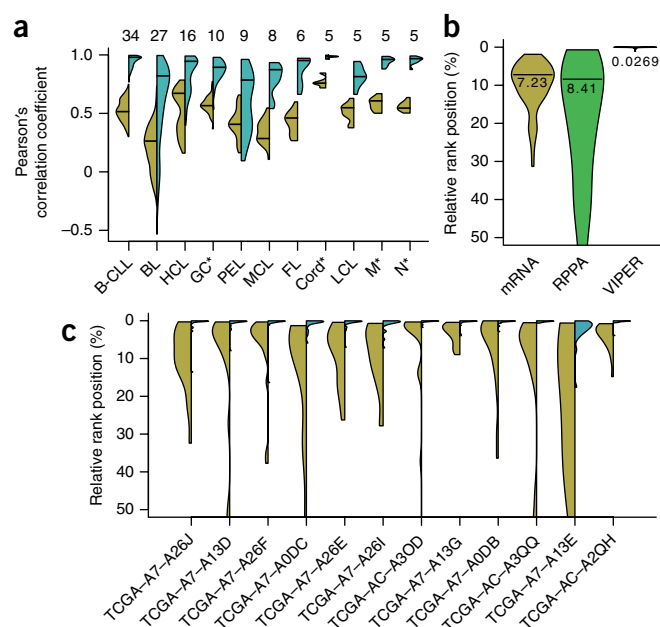
Supplementary Tables 4–6 and **Supplementary Note**). Based on our benchmarking results, we generated a comprehensive map of protein activity dysregulation in response to short-term pharmacologic perturbations. We selected 166 compounds in CMAP[33] that induced reproducible perturbation profiles across replicates (FDR < 0.05, **Supplementary Note**) and report their effect on the activity of 2,956 regulatory proteins in **Supplementary Table 7**.

Algorithm robustness

Poor reproducibility across biological replicates is a critical reason why gene expression analysis has not been broadly adopted in clinical tests. We thus rigorously assessed the reproducibility of the VIPER inferences as a result of multiple sources of technical and biological noise (**Fig. 2**).

Regulons were degraded by progressively randomizing regulatory interactions while maintaining network topology. Although VIPER's performance depends on availability of tissue-specific regulons (**Fig. 2a**), it tolerates a high fraction of false positive interactions, with noticeable performance degradation observed only when >60% of regulon interactions were randomized (**Fig. 2b**). Assuming ~30% false positive rate by ARACNe[34,35], this suggests that as long as >28% of genes in a regulon represent bona fide regulatory interactions, protein differential activity can be accurately inferred.

VIPER assessment of protein activity was robust to reduced regulon representation, as confirmed by the analysis of the library of integrated network-based cellular signatures (LINCS) data (**Supplementary Fig. 7** and **Supplementary Note**). Progressive target removal starting with those with lowest mutual information further increased accuracy, with optimal accuracy achieved at $n = 50$ targets and only modest degradation down to $n = 25$ targets (**Fig. 2c**). Regulons of fewer than 25 targets showed a dramatic decrease in accuracy (**Fig. 2c**).

VIPER was also highly insensitive to gene expression signature degradation, as seen by adding zero-centered Gaussian noise with increasing variance (comparable to benchmark data sets variance) (**Fig. 2d**). This makes it well-suited for assessment of protein activity from noisy single-sample gene expression profiles, where the variance of VIPER-inferred activity is much smaller than the variance of gene expression (**Fig. 3a,b** and **Supplementary Fig. 10**). For instance, considering a B cell phenotype, VIPER-based protein activity signatures were significantly more correlated than gene expression signatures ($P < 10^{-15}$, Wilcoxon signed-rank test; **Fig. 3a** and **Supplementary Fig. 10a**). Addition of Gaussian noise decreased expression-based sample-sample correlation with only a minimal effect on VIPER-inferred activity correlation (**Supplementary Fig. 10b**). VIPER activity was highly resilient to reduced transcriptome representation, showing

minimal accuracy decrease when up to 90% of the genes in the signature were removed from the analysis (**Fig. 2e**) or when RNA-seq profiles where subsampled from 30 million (M) reads to 0.5 M reads (**Fig. 2f**), making VIPER appropriate for the analysis of low-depth RNA-seq profiles. This was further evidenced when comparing protein activity profiles inferred from fresh-frozen vs. matched formalin-fixed paraffin-embedded (FFPE) samples (**Fig. 3c** and **Supplementary Fig. 10c**). The reproducibility of the results from FFPE samples represents a critical prerequisite for precision medicine applications.

To assess the effect of biological variability, we computed VIPER activity signatures for 173 TCGA basal breast carcinomas. VIPER-inferred activity signatures were significantly more correlated across samples ($P < 10^{-15}$ by Wilcoxon signed-rank test for the correlation coefficients, **Supplementary Fig. 10d**) and top-ranking aberrantly activated proteins were more conserved across samples based on differential activity than on differential expression of the associated gene (**Fig. 3b**). Overall sample-to-sample variance was reduced more than 250-fold compared to gene expression (**Fig. 3b**). Thus, VIPER-inferred differentially activated proteins are much more conserved than differentially expressed genes or differentially abundant proteins (based on RPPA measurements) across different samples representing the same tumor subtype (**Fig. 3b**).

Functionalizing the somatic mutational landscape of cancer

Based on these benchmarks, we used VIPER to systematically test the effect of recurrent mutations on corresponding protein activity. We considered a pan-cancer set of 3,912 TCGA samples, representing 14 tumor types, for which exome data are available (**Supplementary Table 8**). We first computed the VIPER-inferred activity of each transcription factor and signaling protein in each of the analyzed samples and tested whether samples harboring recurrent mutations were enriched in those with high VIPER-inferred differential activity of the affected protein. From 150 recurrently mutated genes in COSMIC, we selected 89 that were mutated in at least 10 samples in at least one tumor type and for which a matching regulatory model was available (**Supplementary Table 8**). This identified a total of 342 pairs (for example, *EGFR* in glioblastoma multiforme, GBM) where a specific oncoprotein could be tested in a specific tumor cohort.

Figure 4 Detecting changes in protein activity induced by nonsilent somatic mutations. (**a**–**c**) Changes associated with protein activity only (**a**), protein activity and mRNA expression (**b**), and mRNA expression only (**c**) for GBM, COAD, breast carcinoma (BRCA), lung squamous carcinoma (LUSC), head and neck squamous carcinoma (HNSC), stomach adenocarcinoma (STAD), lung adenocarcinoma (LUAD), kidney renal clear cell carcinoma (KIRC), uterine corpus endometrial carcinoma (UCEC), bladder carcinoma (BLCA), and prostate adenocarcinoma (PRAD). The complete list of evaluated proteins is available in **Supplementary Figure 11**. For each indicated gene harboring nonsilent somatic mutations, the proportion of mutated samples from that tumor type is indicated. Violin plots indicate the distribution density for the mutated samples on all samples rank-sorted by mRNA expression (yellow) and VIPER-inferred protein activity (cyan); background color gradient indicates both expression and VIPER-inferred protein activity signatures: downregulated genes and inactivated proteins (blue) and overexpressed genes and activated proteins (red). Bar plots show significance for the association computed by the aREA algorithm. Blue and red bars indicate enrichment of the mutated samples among low expression or protein activity, and among high levels of expression or protein activity, respectively.

As protein activity may depend on either total protein abundance or on the abundance of specific, differentially active isoforms, we estimated both global VIPER activity and the residual post-translational (RPT) VIPER activity (the component of activity that cannot be accounted for by differential expression) by removing the transcriptional variance component (Online Methods). By definition, RPT activity is statistically independent of gene expression and should account for the purely post-translational contribution to protein activity. Almost 30% of subtype-specific variation-harboring proteins (92/342) were associated with statistically significant differential protein activity, as assessed by VIPER ($P < 0.05$): 65/342 (19%) by global activity analysis and 51/342 (15%) by RPT activity analysis, respectively (**Supplementary Fig. 11**). This included the vast majority of established oncogenes and tumor suppressors (**Fig. 4** and **Supplementary Fig. 11a,b**), suggesting that this integrative analysis provides an effective means to capture mutation-dependent dysregulation of oncogene and tumor suppressor activity (**Supplementary Fig. 11**). VIPER-inferred RPT activity effectively eliminates the effect of feedback loops on the corresponding gene's expression, thus identifying mutations resulting only in post-translational effects (**Supplementary Fig. 11a,b**). We observed that 45% of mutations associated with VIPER-inferred differential activity (41/92) induced no significant differential expression of the corresponding gene (**Fig. 4a** and **Supplementary Fig. 11a**), including mutations in established oncogenes and tumor suppressors, such as *TP53*, *PTEN*, *NFE2L2*, *ARID1A*, *CARD11*, *BRCA2*, *CTNNB1*, *MLH1*, *VHL* and *SMAD4*, among others (**Fig. 4a** and **Supplementary Fig. 11a**).

To assess whether a pharmacologically targetable protein may be aberrantly activated in a tumor sample, independent of the sample's mutational state, we define a sample's mutant phenotype score (MPS). This represents the probability of observing mutations in samples with equal or higher total VIPER activity (**Supplementary Fig. 12**). This is computed as the fraction of mutated vs. wild-type (WT) samples for the specific protein and tumor type. We thus ranked samples based on their MPS for each of the 92 protein/tumor-type pairs for which mutated samples were enriched in differentially activated proteins based on our previous analysis (Online Methods). Although the majority of mutated samples had a high MPS, a few had a low MPS, comparable to WT samples, suggesting nonfunctional mutations, or subclonal mutations or regulatory compensation of their effect (**Fig. 5a** and **Supplementary Fig. 12**), including samples harboring activating
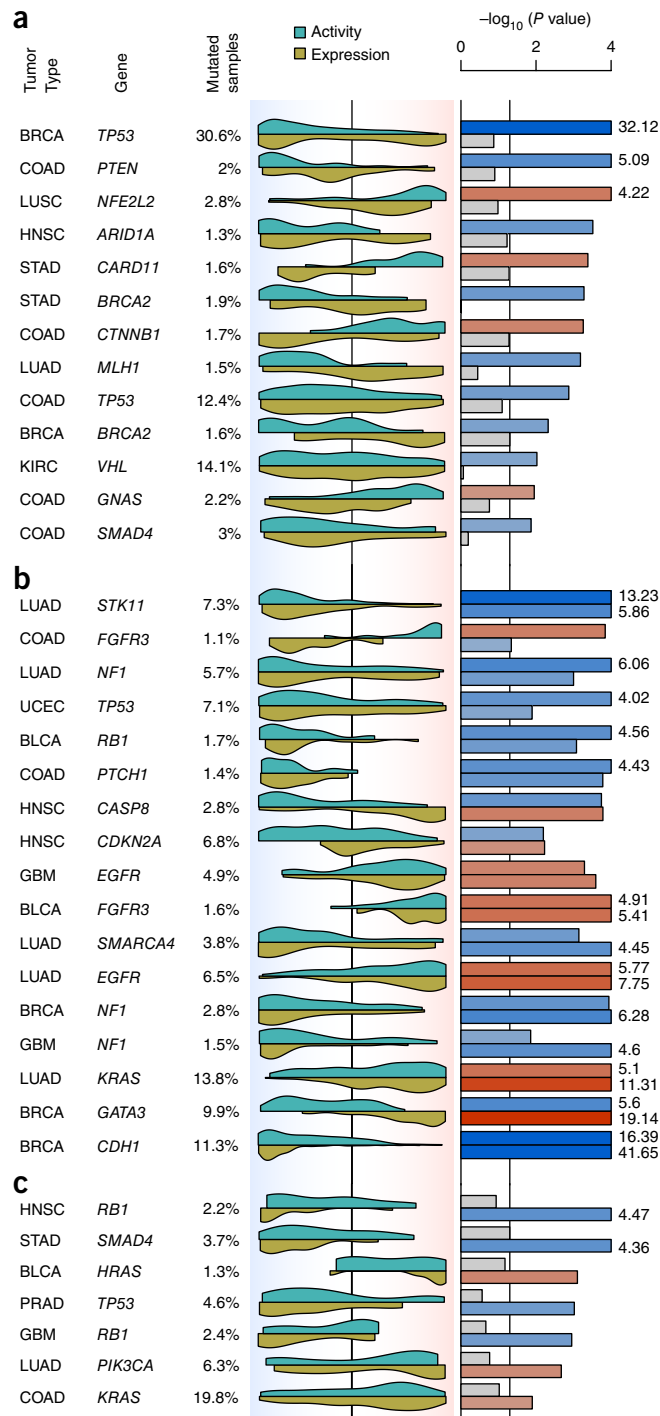
mutations in actionable proteins, such as those encoded by *EGFR*, *ERBB2*, *BRAF* and *PI3K*, with MPSs ≤ −0.5 (i.e., threefold more likely to have WT activity) (**Fig. 5a**), suggesting subpar response to targeted inhibitors. Many WT samples had MPSs ≥ 0.5 (i.e., threefold more likely to have mutated activity) (**Fig. 5a**), suggesting they may respond to targeted inhibitors.

Validating drug sensitivity

To assess whether the MPS is a good predictor of drug sensitivity, we performed *EGFR*-specific MPS analysis of 79 lung adenocarcinoma cell lines, for which gene expression profiles, *EGFR* status and chemosensitivity to *EGFR* inhibitors were available from the Cancer Cell Line

Figure 5 Mutant phenotype score and its association with drug sensitivity. (**a**) Probability density for the nonmutated (salmon) and mutated (green) samples based on MPS for six actionable mutations (complete list in **Supplementary Fig. 12**). Right plots show MPS (*y* axis) for all samples rank-sorted by MPS; green vertical lines indicate the mutated samples. MPS-defined WT and mutant phenotypes (likelihood ratio > 3) are highlighted in salmon and green. (**b**) MPS analysis for *EGFR* on lung carcinoma cell lines. Scatter plots show drug sensitivity, quantified by the area under the titration curves (AUC), for EGFR-targeting drugs as a function of MPS (expressed as likelihood ratio). Cell lines resembling an *EGFR* mutated and WT phenotypes are highlighted in green (likelihood ratio > 3) and salmon boxes, respectively. Green dots indicate cell lines harboring nonsilent mutations. Solid and doted horizontal lines indicate the mean and 2.33 s.d. over the mean of the chemoresistant cell lines, respectively. The association between drug sensitivity and MPS is shown on top of each plot by the Pearson's correlation coefficient (*R*) and associated *P* value. Violin plots show the probability density for drug sensitivity (AUC) of the cell lines showing an *EGFR* WT (green) or mutant (brown) phenotype according to MPS; horizontal lines indicate distribution means, which were contrasted by Student *t*-test (*P* values in insets).

Encyclopedia[7], including saracatinib (AZD0530), erlotinib and lapatinib. Of the cell lines with low *EGFR* MPS (< −0.5) that yet harbored *EGFR* mutations, 0/2, 1/2 and 1/2 were sensitive to AZD0530, erlotinib and lapatinib, respectively. Conversely, 5/6, 5/6 and 4/6 of those with MPS > 0.5, were sensitive to those drugs, respectively (**Fig. 5b**), suggesting a strong association between MPS and chemosensitivity in *EGFR*-mutated cell lines. Moreover, considering only *EGFR* WT cell lines, the fraction responding to *EGFR* inhibitors was higher among those with MPS > 0.5 (50% vs. 33% for AZD0530, 43% vs. 33% for erlotinib and 36% vs. 27% for lapatinib, respectively) compared to those with MPS < −0.5 (**Fig. 5b**). MPS was significantly associated with chemosensitivity, regardless of *EGFR* mutation status, by Pearson correlation analysis ($P < 10^{-5}$ for each of the three drugs; **Fig. 5b**), and by comparing sensitivity of cells with MPS > 0.5 and MPS < −0.5 by Student's *t*-test ($P < 0.01$ and $P < 0.05$ for AZD0530 and erlotinib, respectively **Fig. 5b**).

Assessing the role of site-specific mutations

In the previous analysis, all mutations in a gene were considered equivalent. We next tested whether VIPER could also be used to assess differential activity associated with mutations at specific protein sites. This could be instrumental in elucidating the functional effect of rare or private mutations. Specifically, we tested whether different mutations in the same gene (for example, p.Gly12Val vs. p.Gly12Asp changes for the *KRAS* product) may produce quantitatively distinct effects on protein activity. We assessed all mutations affecting COSMIC genes that were detected in at least two samples of the same tumor type, based on four quantitative measurements: (i) their VIPER-inferred global activity, (ii) their VIPER-inferred RPT activity, (iii) their differential gene expression, and (iv) their MPS (for mutations affecting at least 10 samples). In total, we analyzed 648 locus-specific mutations in 49 distinct genes, across 12 tumor types (**Supplementary Fig. 13**). In **Figure 6** we summarized the cases with adequate statistical power. Careful examination showed that functional impact of these mutations was both variant-specific (for example, *KRAS*: p.Gly12Val vs. p.Gly12Asp in colon adenocarcinoma (COAD); **Fig. 6a**) and tumor specific (for example, *KRAS*: p.Gly12Ala in COAD vs. lung adenocarcinoma (LUAD); **Fig. 6a**). In addition, although some mutations induce effects equivalent to differential expression, others produce exquisitely post-translational effects that can only be predicted by RTP activity (for example, *KRAS*: p.Gly12Val in LUAD vs. p.Gly13Asp in COAD; **Fig. 6a** and **Supplementary Fig. 13**).
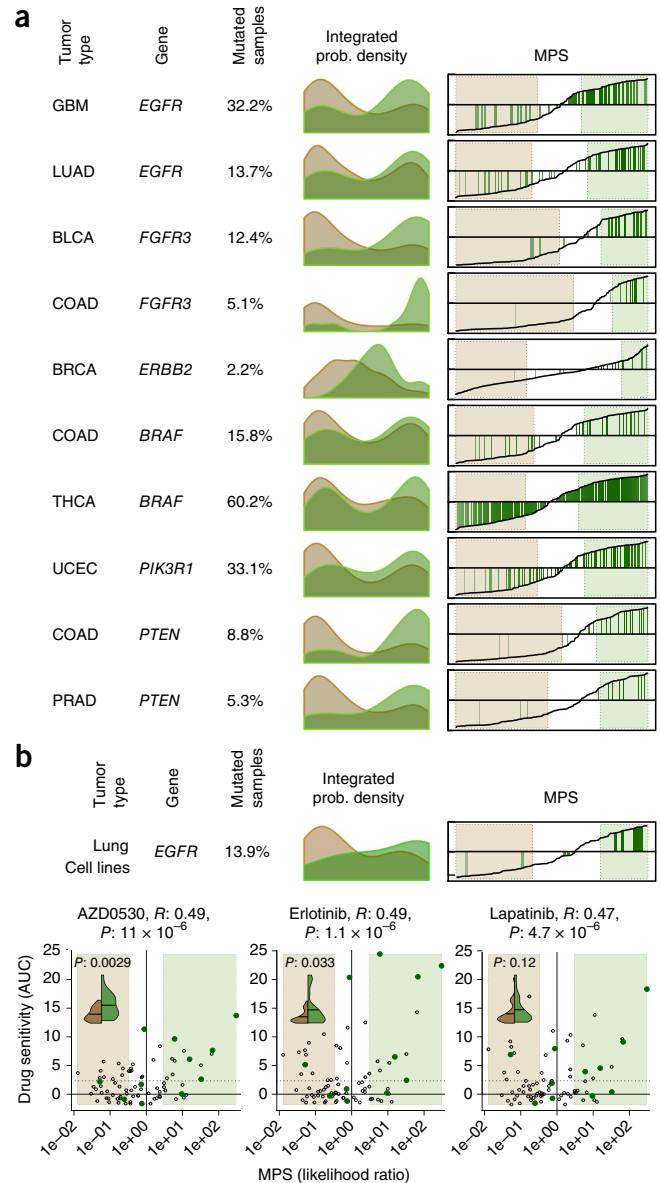
Although different mutations may have similar impact on protein activity (for example, all *TP53* functional variants were associated with reduction in inferred TP53 protein activity), their effects on gene expression were highly heterogeneous. For instance, nonsense and frame-shift mutations in *TP53* invariably reduced mRNA levels (**Fig. 6a**), likely due to nonsense and nonstop-mediated mRNA decay[36]. In contrast, missense mutations were consistently associated with increased mRNA levels, likely due to feedback loops attempting to compensate for mutation-induced loss of TP53 protein activity (**Fig. 6a**)[37]. Such dichotomy in *TP53* somatic variant effect may explain the lack of association between mutations and gene expression, when all variants are considered together (**Fig. 4a**).

To compensate for the lack of statistical power due to the potentially small number of samples harboring locus-specific mutations (**Supplementary Fig. 13**), we performed integrated analysis across all tumor types. We accounted for heterogeneity among tumor types by aggregating the samples at the protein activity level, originally inferred using tissue-matched interactomes. This yielded a pan-cancer repertoire of functionally relevant somatic variants, based on the analysis of 3,343 samples across 12 tumor types, for which we
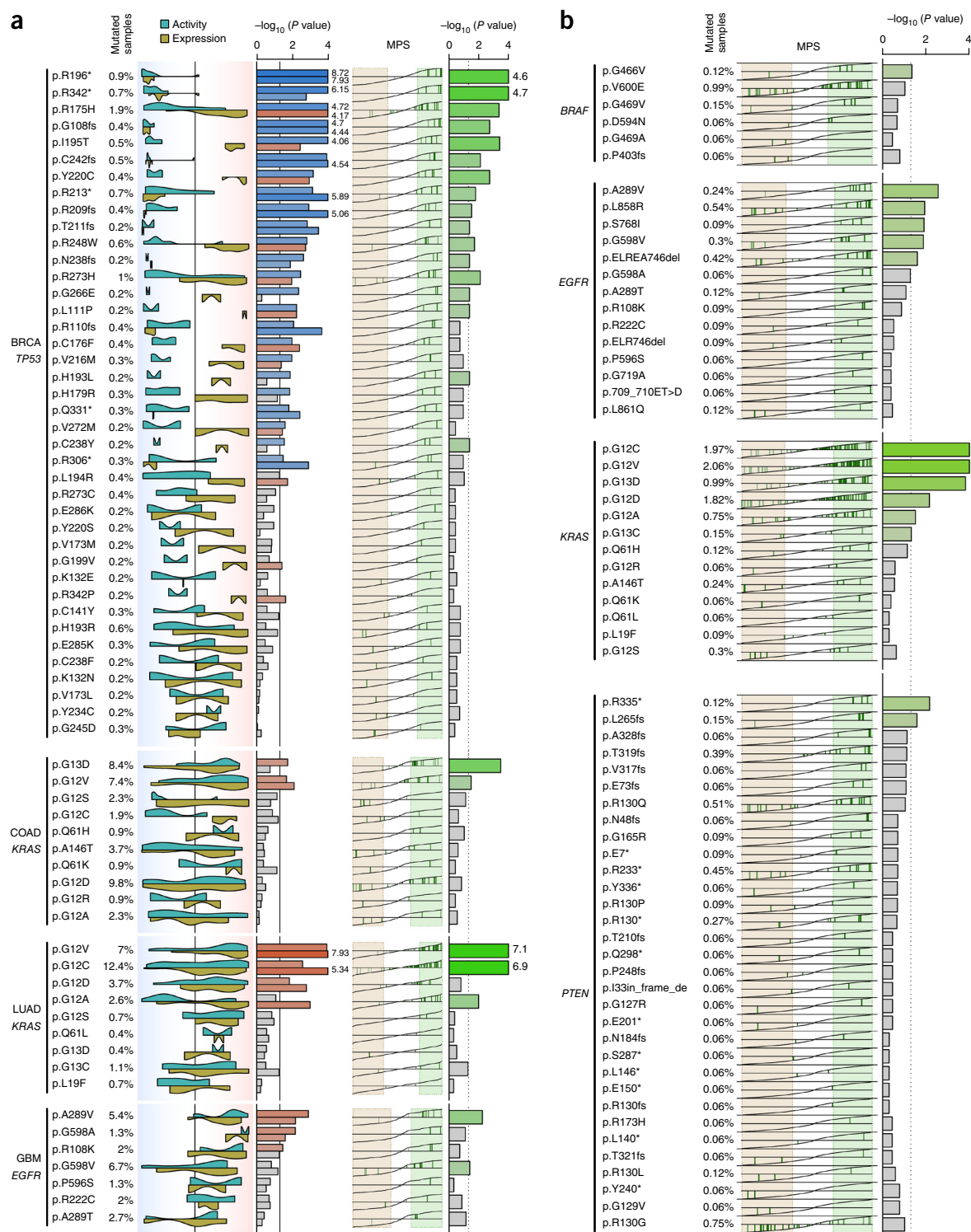
Figure 6 Effect of specific nonsilent somatic mutation variants on VIPER-inferred protein activity. (**a**) Association of nonsilent somatic mutation variants with VIPER-inferred protein activity and mRNA expression. Violin plots indicate the probability density for the mutated samples on all samples rank-sorted by coding gene mRNA levels (yellow) or VIPER-inferred protein activity (cyan). Background color gradient indicates both expression and VIPER-inferred protein activity signatures from decreased (blue) to increased (orange). Statistical level for the association, as estimated by aREA (bar plot), with color indicating association with increased (red) or decreased (blue) expression or protein activity. Rightmost bar plot shows the significance level for the association of mutation variants and the MPS-defined mutant phenotype (likelihood ratio > 3, light-green box). The MPS-defined WT phenotype (likelihood ratio > 3) is indicated by the light-salmon box. Missense mutations are indicated as p.*XnY* where *X* stands for 1-letter amino acid in position *n* that was mutated to *Y*. *, nonsense mutations; frameshift mutations are indicated as p.*Xn*fs. Vertical lines crossing the bars indicate the *P*-value threshold of 0.05. (**b**) Effect of nonsilent variants integrated across different tumor types. MPS was integrated for all 12 tumor types (3,343 samples) and is shown as the *x* axis in the left side of the plot, while the enrichment of each variant among the samples with at least threefold likelihood of mutation vs. the WT samples (likelihood-ratio > 3), is indicated as −log$_{10}$(*P*) by the bar plots. Dashed line indicates the *P*-value threshold of 0.05.

report the statistical association between each locus-specific mutation and its MPS, as well as the pan-cancer VIPER *P* value (**Fig. 6b** and **Supplementary Fig. 14**).

DISCUSSION

Precision cancer medicine currently relies on the identification of actionable mutations. These can be reproducibly identified from whole-genome and exome analysis of tumor tissue and have demonstrated clinical relevance. However, only ~25% of adult cancer patients present with potentially actionable mutations[8]. Thus, methodologies, such as VIPER, for inferring aberrant protein activity, independent of mutational state, may complement and greatly extend available genomic approaches. Indeed, genetic mutations are neither necessary nor sufficient to induce aberrant activity and tumor essentiality of protein isoforms. An increasing catalog of non-oncogene dependencies has emerged in recent years[5,18,20,21,38,39], whose aberrant activity depends on indirect genetic alterations, such as those in upstream pathways and cognate binding proteins. It is not surprising that many tumor cells respond to inhibitors targeting established oncoproteins, such as *EGFR*, even in the absence of activating mutations, as shown by large-scale dose-response studies in the cancer cell line encyclopedia[6,7] and by recent analysis of pathways upstream of functional tumor drivers[5].

VIPER has three critical roles. First, it helps elucidate aberrant protein activity resulting either from direct or pathway-mediated mutations. Second, it can help prioritize the functional relevance of rare and private nonsynonymous mutations as hypomorph, hypermorph or neutral events. Systematic analysis of TCGA cohorts showed that 27% of nonsynonymous mutations induced aberrant VIPER-inferred protein activity. This is a substantial fraction, especially considering that not all mutations substantially affect protein activity on canonical targets, including those resulting in entirely new protein functions (neomorphs), and that mutation clonality was not accounted for in these studies. Third, VIPER can help distinguish between transcriptionally and post-translationally mediated mutational effects (**Figs. 4a–c** and **6**).

Systematic VIPER of TCGA samples (**Fig. 5a**) showed that although genetic alterations strongly co-segregated with aberrant VIPER-inferred oncoprotein activity, many WT samples had VIPER-inferred activity comparable to and even greater than those harboring actionable mutations. This is critically relevant for alterations in pharmacologically actionable oncogenes, such as *BRAF*, *EGFR*, *ERBB2* and *FGFR3*, among others, suggesting that VIPER may be used to identify additional patients who may benefit from targeted therapy. Similarly, VIPER identified samples with actionable mutations presenting no aberrant activity of the corresponding oncoprotein. Validation of the predictive value of VIPER-inferred activity to infer targeted inhibitor response, using the cancer cell line encyclopedia, suggests that the algorithm may provide valuable insight in precision cancer medicine.

Several approaches have been proposed to estimate pathway activity[40,41], co-regulation of gene expression modules[42] or activity of selected proteins[43] from gene expression signatures. These, however, do not predict activity of arbitrary proteins, lack tumor specificity and cannot be used to analyze individual samples. Other approaches developed for yeast[44] and other model organisms[44–47] have never been extended to mammalian cells. Earlier attempts based on transcription factor targets inferred from promoter sequence analysis[16] or from proprietary, literature-based networks[26] have not been systematically validated. As a result, with the exception of VIPER, to our knowledge there are currently no validated methods to systematically predict the activity of all signal transduction and transcription factors proteins in individual samples.

VIPER leverages protein regulons reverse-engineered from primary tumor sample data to quantitatively assess differential protein activity in individual samples, without any manual annotation or curated gene sets. Critically, VIPER's performance is extremely robust and resilient to signature noise, regulon subsampling and sample quality. Indeed, VIPER accurately inferred protein activity for ~50% of all regulatory proteins using <1,000 genes from LINCS perturbational signatures (**Supplementary Fig. 7**). Furthermore, inference of differentially active proteins from fresh-frozen or FFPE samples from the same tissue was highly correlated, even though correlation of the corresponding gene expression data was low. VIPER predictions were remarkably reproducible across samples belonging to the same molecular tumor subtype. This is critically important for precision medicine applications.

Tissue specificity of protein-target is a critical element of our analysis. Genes with expression affected by changes in protein activity are highly context-specific[35], owing to lineage-specific chromatin remodeling, combinatorial regulation by multiple transcription factors and post-translational modification. This is supported by the fact that inference of protein activity using the incorrect regulatory model produced substantially degraded results (**Fig. 2a**).

VIPER constitutes only a partial contribution toward the ultimate goal of accurately measuring protein activity in mammalian samples. Yet our data suggest that improvements in the accuracy and coverage of regulatory models could further increase the quality and breadth of these predictions, thus helping determine which proteins drive key pathophysiological phenotypes. We illustrated the potential application of VIPER to mine existing data sets, including expression profiles in TCGA and LINCS. Finally, VIPER has the power to infer relative protein activity as an extra layer of information, providing additional evidence over classical genetics and functional genomics data to assess the effect of nonsilent mutations.

URLs. VIPER package for the R system is available from Bioconductor at http://www.bioconductor.org/packages/release/bioc/html/viper.html. VIPER package vignette, http://bioconductor.org/packages/release/bioc/vignettes/viper/inst/doc/viper.pdf; The Cancer Genome Atlas, http://cancergenome.nih.gov/; ChEA database, http://amp.pharm.mssm.edu/Enrichr/#stats; CMAP, https://www.broadinstitute.org/cmap/.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Affymetrix array data for the *BCL6* knockdown experiments described in the paper have been deposited at the Gene Expression Omnibus (GEO) under accession number GSE45838.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

generated by The Cancer Genome Atlas pilot project established by the NCI and NHGRI as of January 2011.

AUTHOR CONTRIBUTIONS

M.J.A. conceptualized and developed the algorithms, designed the experiments, analyzed the data and wrote the manuscript. Y.S., F.M.G. and A.L. analyzed the data. B.B.D. generated the *BCL6* knockdown experiment and expression profile. B.H.Y. designed the experiments used for benchmarking the algorithms with transcription factor knockdown assays. A.C. conceptualized the algorithm, directed the project, designed the experiments and wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the online version of the paper.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Hanahan, D. & Weinberg, R.A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).
2. Weinstein, I.B. Cancer. addiction to oncogenes—the Achilles heal of cancer. *Science* **297**, 63–64 (2002).
3. Wang, X., Haswell, J.R. & Roberts, C.W. Molecular pathways: SWI/SNF (BAF) complexes are frequently mutated in cancer—mechanisms and potential therapeutic insights. *Clin. Cancer Res.* **20**, 21–27 (2014).
4. Sumazin, P. *et al.* An extensive microRNA-mediated network of RNA–RNA interactions regulates established oncogenic pathways in glioblastoma. *Cell* **147**, 370–381 (2011).
5. Chen, J.C. *et al.* Identification of causal genetic drivers of human disease through systems-level analysis of regulatory networks. *Cell* **159**, 402–414 (2014).
6. Basu, A. *et al.* An interactive resource to identify cancer genetic and lineage dependencies targeted by small molecules. *Cell* **154**, 1151–1161 (2013).
7. Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603–607 (2012).
8. MacConaill, L.E. *et al.* Prospective enterprise-level molecular genotyping of a cohort of cancer patients. *J. Mol. Diagn.* **16**, 660–672 (2014).
9. Klein, U. *et al.* Transcriptional analysis of the B cell germinal center reaction. *Proc. Natl. Acad. Sci. USA* **100**, 2639–2644 (2003).
10. Alizadeh, A.A. *et al.* Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403**, 503–511 (2000).
11. Tothill, R.W. *et al.* Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome. *Clin. Cancer Res.* **14**, 5198–5208 (2008).
12. Creighton, C.J. *et al.* Residual breast cancers after conventional therapy display mesenchymal as well as tumor-initiating features. *Proc. Natl. Acad. Sci. USA* **106**, 13820–13825 (2009).
13. Wolf-Yadlin, A., Sevecka, M. & MacBeath, G. Dissecting protein function and signaling using protein microarrays. *Curr. Opin. Chem. Biol.* **13**, 398–405 (2009).
14. Božović, A. & Kulasingam, V. Quantitative mass spectrometry–based assay development and validation: from small molecules to proteins. *Clin. Biochem.* **46**, 444–455 (2013).
15. Rodríguez, J.A. Interplay between nuclear transport and ubiquitin/SUMO modifications in the regulation of cancer-related proteins. *Semin. Cancer Biol.* **27**, 11–19 (2014).
16. Rhodes, D.R. *et al.* Mining for regulatory programs in the cancer transcriptome. *Nat. Genet.* **37**, 579–583 (2005).
17. Lefebvre, C. *et al.* A human B-cell interactome identifies MYB and FOXM1 as master regulators of proliferation in germinal centers. *Mol. Syst. Biol.* **6**, 377 (2010).
18. Carro, M.S. *et al.* The transcriptional network for mesenchymal transformation of brain tumours. *Nature* **463**, 318–325 (2010).
19. Chudnovsky, Y. *et al.* ZFHX4 interacts with the NuRD core member CHD4 and regulates the glioblastoma tumor-initiating cell state. *Cell Rep.* **6**, 313–324 (2014).
20. Aytes, A. *et al.* Cross-species regulatory network analysis identifies a synergistic interaction between FOXM1 and CENPF that drives prostate cancer malignancy. *Cancer Cell* **25**, 638–651 (2014).
21. Piovan, E. *et al.* Direct reversal of glucocorticoid resistance by AKT inhibition in acute lymphoblastic leukemia. *Cancer Cell* **24**, 766–776 (2013).
22. Forbes, S.A. *et al.* COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res.* **39**, D945–D950 (2011).
23. Basso, K. *et al.* Reverse engineering of regulatory networks in human B cells. *Nat. Genet.* **37**, 382–390 (2005).
24. Lachmann, A. *et al.* ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X experiments. *Bioinformatics* **26**, 2438–2444 (2010).
25. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
26. Krämer, A., Green, J., Pollard, J. Jr. & Tugendreich, S. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics* **30**, 523–530 (2014).
27. Abatangelo, L. *et al.* Comparative study of gene set enrichment methods. *BMC Bioinformatics* **10**, 275 (2009).
28. Boorsma, A., Foat, B.C., Vis, D., Klis, F. & Bussemaker, H.J. T-profiler: scoring the activity of predefined groups of genes using gene expression data. *Nucleic Acids Res.* **33**, W592–W595 (2005).
29. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **102**, 15545–15550 (2005).
30. Jiang, Z. & Gentleman, R. Extensions to gene set enrichment. *Bioinformatics* **23**, 306–313 (2007).
31. Dinu, I. *et al.* Improving gene set analysis of microarray data by SAM-GS. *BMC Bioinformatics* **8**, 242 (2007).
32. Wang, K. *et al.* Genome-wide identification of post-translational modulators of transcription factor activity in human B cells. *Nat. Biotechnol.* **27**, 829–839 (2009).
33. Lamb, J. *et al.* The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* **313**, 1929–1935 (2006).
34. Basso, K. *et al.* Integrated biochemical and computational approach identifies BCL6 direct target genes controlling multiple pathways in normal germinal center B cells. *Blood* **115**, 975–984 (2010).
35. Kushwaha, R. *et al.* Interrogation of a context-specific transcription factor network identifies novel regulators of pluripotency. *Stem Cells* **33**, 367–377 (2015).
36. Adjibade, P. & Mazroui, R. Control of mRNA turnover: implication of cytoplasmic RNA granules. *Semin. Cell Dev. Biol.* **34**, 15–23 (2014).
37. Harris, S.L. & Levine, A.J. The p53 pathway: positive and negative feedback loops. *Oncogene* **24**, 2899–2908 (2005).
38. Luo, J., Solimini, N.L. & Elledge, S.J. Principles of cancer therapy: oncogene and non-oncogene addiction. *Cell* **136**, 823–837 (2009).
39. Compagno, M. *et al.* Mutations of multiple genes cause deregulation of NF-κB in diffuse large B-cell lymphoma. *Nature* **459**, 717–721 (2009).
40. Lee, E., Chuang, H.Y., Kim, J.W., Ideker, T. & Lee, D. Inferring pathway activity toward precise disease classification. *PLoS Comput. Biol.* **4**, e1000217 (2008).
41. Vaske, C.J. *et al.* Inference of patient-specific pathway activities from multi-dimensional cancer genomics data using PARADIGM. *Bioinformatics* **26**, i237–i245 (2010).
42. Segal, E. *et al.* Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nat. Genet.* **34**, 166–176 (2003).
43. Yörük, E., Ochs, M.F., Geman, D. & Younes, L. A comprehensive statistical model for cell signaling. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **8**, 592–606 (2011).
44. Boorsma, A., Lu, X.J., Zakrzewska, A., Klis, F.M. & Bussemaker, H.J. Inferring condition-specific modulation of transcription factor activity in yeast through regulon-based analysis of genomewide expression. *PLoS One* **3**, e3112 (2008).
45. Foat, B.C., Morozov, A.V. & Bussemaker, H.J. Statistical mechanical modeling of genome-wide transcription factor occupancy data by MatrixREDUCE. *Bioinformatics* **22**, e141–e149 (2006).
46. Kundaje, A. *et al.* Learning regulatory programs that accurately predict differential expression with MEDUSA. *Ann. NY Acad. Sci.* **1115**, 178–202 (2007).
47. di Bernardo, D. *et al.* Chemogenomic profiling on a genome-wide scale using reverse-engineered gene networks. *Nat. Biotechnol.* **23**, 377–383 (2005).
48. Phillips, H.S. *et al.* Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis. *Cancer Cell* **9**, 157–173 (2006).

ONLINE METHODS

Regulatory networks. The regulatory networks were reverse engineered by ARACNe[49] from 20 different data sets: two B-cell context data sets profiled on Affymetrix HG-U95Av2 and HG-U133plus2 platforms, respectively; a high-grade glioma data set profiled on Affymetrix HG-U133A arrays; and 17 human cancer tissue data sets profiled by RNA-seq from TCGA (**Table 1**). The Affymetrix platform data sets were summarized by MAS5 (affy R package[50,51]) using probe clusters generated by the 'cleaner' algorithm[52]. Cleaner generates 'informative' probe-clusters by analyzing the correlation structure between probes mapping to the same gene and discarding noncorrelated probes, which might represent poorly hybridizing or cross-hybridizing probes[52]. The RNA-seq level 3 data were downloaded from TCGA data portal, raw counts were normalized to account for different library size, and the variance was stabilized by fitting the dispersion to a negative-binomial distribution as implemented in the DESeq R package[53] (Bioconductor[54]). ARACNe was run with 100 bootstrap iterations using all probe clusters mapping to a set of 1,813 transcription factors (genes annotated in Gene Ontology molecular function database (GO)[55] as GO:0003700, 'transcription factor activity', or as GO:0004677, 'DNA binding', and GO:0030528, 'transcription regulator activity', or as GO:0004677 and GO: 0045449, 'regulation of transcription'), 969 transcriptional cofactors (a manually curated list, not overlapping with the transcription factor list, built upon genes annotated as GO:0003712, 'transcription cofactor activity', or GO:0030528 or GO:0045449) or 3,370 signaling pathway related genes (annotated in GO Biological Process database as GO:0007165 'signal transduction' and in GO cellular component database as GO:0005622, 'intracellular', or GO:0005886, 'plasma membrane') as candidate regulators. Parameters were set to 0 DPI (data processing inequality) tolerance and MI (mutual information) P-value threshold of 10^{-8}.

The regulatory networks based on ChIP experimental evidence were assembled from ChEA and ENCODE data. The mode of regulation was computed based on the correlation between transcription factor and target gene expression as described below.

Benchmarking experiments. We used gene expression profile data after *MEF2B*[32], *FOXM1* (ref. 17), *MYB*[17] (GSE17172) and *BCL6* (GSE45838) silencing in human B cells, and *STAT3* silencing in the human glioma cell line SNB19 (ref. 18; GSE19114, **Table 2**). *BCL6* knockdown experiments were performed in OCI-Ly7 and Pfeiffer GCB-DLBCL cell lines. Both cell lines were maintained in 10% FBS supplemented IMDM and transiently transfected with either a *BCL6*-specific or a nontarget control siRNA oligo in triplicate as described previously[56]. Total RNA was isolated 48 h after transfection, time at which knockdown of BCL6 protein was observed (**Supplementary Fig. 15a**), and gene expression was profiled on H-GU133plus2 Affymetrix gene chips following the manufacturer protocol (Affymetrix Inc.). All experiments showed a significant reduction at the mRNA level for the silenced gene as quantified by expression profile (**Supplementary Fig. 15b**). Gene expression signatures were obtained by Student's *t*-test analysis of the gene expression profiles; see **Table 2**.

VIPER. The VIPER algorithm tests for regulon enrichment on gene expression signatures. The gene expression signature is first obtained by comparing two groups of samples representing distinctive phenotypes or treatments. Any method that generates a quantitative measurement of difference between the groups can be used (fold change, Student's *t*-test, Mann-Whitney *U* test, etc.). Alternatively, single-sample-based gene expression signatures can be obtained by comparing the expression levels of each feature in each sample against a set of reference samples by any suitable method, including for example Student's *t*-test, Z-score transformation or fold change; or relative to the average expression level across all samples when clear reference samples are not available. Then we compute the enrichment of each regulon on the gene expression signature using different implementations of aREA (see below). Finally, we estimate the significance, including *P* value and normalized enrichment score, by comparing each regulon enrichment score to a null model generated by randomly and uniformly permuting the samples 1,000 times. Alternatively, when the number of samples is not enough to support permutation with reposition (at least five samples per group is required), permutation of the genes in the gene expression signature or its analytic approximation can be used (see below).

Fisher's exact test. We tested whether the overlap between the subset of genes that were differentially expressed following RNAi-mediated silencing of each gene ($P < 0.01$) and the genes in its regulon was statistically significant by Fisher's exact test (FET). The classical FET method considers equally all differentially expressed genes, regardless of whether they are up- or downregulated and hence, FET cannot infer whether the regulator activity is increased or decreased by the perturbation. To address this issue, we modified the FET approach to compute independently the enrichment of activated and repressed targets of a regulator (positive and negative parts of its regulon) on up- and downregulated genes, respectively. Specifically, the genes in each regulon were divided into two subsets: (i) transcriptionally activated (R+) and (ii) transcriptionally repressed (R−) targets. We used the sign of the Spearman's correlation between the mRNA expression level for the regulator and each of the genes in its regulon to classify them as part of R+ or R−. This correlation analysis was performed on the same data set used to infer the network by ARACNe. Then, FET analysis was performed independently for R+ and R− on the two tails of each gene expression signature. Regulators with an increase in activity would thus show enrichment of R+ targets in overexpressed genes and of R− targets in underexpressed genes, respectively. The opposite would be the case for regulators with a decrease in activity. The use of discrete gene lists by FET produces enrichments that are often not robust with respect to threshold selection (**Supplementary Fig. 16**).

Gene set enrichment analysis. One-tail GSEA was implemented as described[29]. For two-tail GSEA, we divided the query regulon into two subsets: a positive subset containing the genes predicted to be transcriptionally activated by the regulator (R+), and a negative subset encompassing the target genes predicted to be repressed by the regulator (R−). The target genes were classified as being part of the R+ or R− subsets depending on whether their mRNA levels were positively or negatively correlated with the regulator mRNA levels (Spearman's correlation). The gene expression signature was then sorted from the most upregulated to the most downregulated gene (signature A) and the rank positions for R+ were computed. The rank positions for R− were then computed from the gene expression signature, but this time sorted from the most downregulated to the most upregulated gene (signature B). The enrichment score was computed as described[29], using the computed rank positions for the R+ and R− subsets, but taking the weighting score values only from signature A.

Analytic rank-based enrichment analysis. aREA tests for a global shift in the positions of each regulon genes when projected on the rank-sorted gene expression signature. Following up on the work in refs. 57,58, we used the mean of the quantile-transformed rank positions as test statistic (enrichment score). The enrichment score is computed twice: first by a one-tail approach, based on the absolute value of the gene expression signature (i.e., genes are rank-sorted from the less invariant between groups to the most differentially expressed, regardless of the direction of change); and then by a two-tail approach, where the positions of the genes whose expression is repressed by the regulator (R−) are inverted in the gene expression signature before computing the enrichment score. The one-tail and two-tail enrichment score estimates are integrated while weighting their contribution based on the estimated mode of regulation through a procedure we call 'three-tail' approach (see below). The contribution of each target gene from a given regulon to the enrichment score is also weighted based on the regulator-target gene interaction confidence (see below). Finally, the statistical significance for the enrichment score is estimated by comparison to a null model generated by permuting the samples uniformly at random or by an analytic approach equivalent to shuffle the genes in the signatures uniformly at random. The arithmetic mean-based enrichment score has several desirable properties, both at the algebraic level, by making the weighted contribution of the targets to the enrichment score trivial to formulate, as well as at the computational level. Regarding this last point, given the linear nature of the mean-based enrichment score, its computation across the elevated number of permutations required to generate the null model can be performed very efficiently by matrix operations. Moreover, the use of the arithmetic mean as enrichment score allows for analytical approaches to estimate its statistical significance, which is equivalent to shuffle the genes in the signatures uniformly at random. We note, however, that the null hypotheses tested by these two alternative approaches are not equivalent. In the case

of sample shuffling, we test whether the calculated enrichment score for a given gene expression signature (i.e., for gene expression profiles associated with the phenotypes) is significantly higher than the one we can obtain when there is no association between the phenotype and the gene expression profile. Conversely, gene shuffling (or its analytic approximation) tests whether the enrichment score is higher than the one we can obtain when the set of genes to test is uniformly distributed in the gene expression signature. Gene shuffling can be approximated analytically as follows: according to the central limit theorem, the mean of a sufficiently large number of independent random variables will be approximately normally distributed. The enrichment score of our null hypothesis fulfill this condition, and we ensure a mean of zero and variance equal to one for the enrichment score under the null hypothesis by applying a quantile transformation based on the normal distribution to the rank-transformed gene expression signature before computing the enrichment score. Then, under the null hypothesis, the enrichment score will be normally distributed with mean equals zero and variance $1/n$, where n is the regulon size. This definition can be generalized, when the weighted mean is used, by

$$\sigma^2 = \sum_{i=1}^{n} w_i^2$$

where w_i is the weight for target i.

Mode of regulation. The mode of regulation (MoR) is determined based on the Spearman's correlation coefficient (SCC) between the regulator and the target expression, computed from the data set used to reverse engineer the network. However, for complex non-monotonic dependencies (for example, for context-specific rewiring[59–61]), assessing the MoR may not be trivial. To address this issue, we first model the SCC probability density for all regulator-target interactions in the network using a three-Gaussian mixture (**Supplementary Fig. 1**), representing (i) clearly repressed targets (MoR−), (ii) clearly activated targets (MoR+), and (iii) non-monotonically regulated targets for which the MoR cannot be reliably estimated (MoR^NM). The parameters for the three-Gaussian mixture model were estimated with the 'mixtools' R package[62]. Then, rather than defining MoR+ or MoR− targets based on the sign of the SCC, we associate each target with three weights (p_A, p_R, p_{NM}), representing the probability that, given its SCC, it may be activated, repressed, or non-monotonically regulated. These probabilities are computed as the relative likelihood of a given regulator-target interaction to be described by any of these three models and computed as the difference between the cumulative distribution for activation (CDF(G_2)) and the CDF for repression (CDF(G_1)), divided by the total CDF: CDF(G_1 upper-tail) + CDF(G_2 lower-tail) + CDF(G_0 lower-tail for Rho < 0 or G_0 upper-tail for Rho > 0) (**Supplementary Fig. 2a–f**).

The aREA-3T approach implemented in VIPER uses MoR to weight the contribution of the one-tail- and two-tail-based enrichment scores as: $ES = |MoR| \, ES_2 + (1-|MoR|) \, ES_1$, where ES_1 and ES_2 are the one-tail aREA and two-tail aREA estimations of the enrichment score (**Fig. 1c**). Such probabilistic formulation avoids selection of arbitrary thresholds for determining target MoR, reducing parameter choices and thus risk of data overfitting.

aREA-3T behaves remarkable robust to changes in the parameter estimates for the three-Gaussian mixed model. We scanned the 'mean' parameter space on a wide range, from −0.3 to −0.6 for G_1 and from 0.3 to 0.6 for G_2; and found a uniform response of aREA on the estimated normalized enrichment score and P values across all benchmarking experiments, with only the rank positions being slightly affected (**Supplementary Fig. 2g,h**).

Regulator-target confidence. We used the mutual information (MI) between regulator and target gene mRNA levels as inference of regulator-target interaction confidence. To compute a regulator-target interaction confidence score, we first generated a null set of interactions for each regulator by selecting target genes at random from all the profiled genes while excluding those in the actual regulon (i.e., ARACNe inferred). The number of target genes for the null regulon was chosen to match those in the actual regulon. Then we computed a CDF for the MI in the ARACNe regulons (CDF_1) and null regulons (CDF_2), and estimated the confidence score for a given regulator-target interaction (interaction confidence or IC) as the ratio: $IC = CDF_1 / (CDF_1 + CDF_2)$. IC was

used to weight the contribution of each target gene to the enrichment score (**Supplementary Fig. 17**).

Pleiotropy. Pleiotropic regulation of gene expression (genes regulated by several different transcription factors) can lead to false positive results if a non-active regulator shares a significant proportion of its regulon with a bona fide active regulator (**Fig. 1d** and **Supplementary Table 9**). To account for this effect, we extended the shadow analysis procedure originally described in ref. 17 to take full advantage of the probabilistic framework used by VIPER. Briefly, we first generated all possible pairs of regulators AB satisfying two conditions: (i) both A and B regulons are significantly enriched in the gene expression signature ($P < 0.05$), and (ii) they co-regulate ($A \cap B$) at least ten genes. Then we evaluate whether the regulons in each pair are enriched in the gene expression signature mostly due to the co-regulated genes. This is performed by computing the enrichment of the co-regulated genes ($A \cap B$) on a subset of the gene expression signature representing only the genes in A (pA) and in B (pB), where pA and pB represent the estimated P values for the enrichment computed by aREA. Then we compute the pleiotropy differential score as $PDE = \log_{10}(pB) - \log_{10}(pA)$. If $pA < pB$, we penalize the co-regulated genes for A by $PDE^{PI/NT}$, where pleiotropy index (PI) is a constant and NT is the number of test pairs involving the regulon A. Conversely, if $pA > pB$ we penalize the co-regulated genes for B by $|PDE|^{PI/NT}$. VIPER results showed in general to be robust to different values for the pleiotropy index (**Supplementary Fig. 18**). We set PI = 20 based on the benchmark data (**Table 2**), because it was a reasonable compromise between accuracy and specificity (**Supplementary Fig. 18**).

Availability. The VIPER algorithm is available as an R system package from Bioconductor. A detailed description of the package functionality and use-case examples can be found in the viper package vignette.

Residual post-translational RPT activity. We found a strong association between VIPER-inferred protein activity and the coding gene mRNA level (**Supplementary Fig. 19**). We estimated the variance in VIPER-inferred protein activity owing to the expression level of the coding gene by fitting a lineal model to the rank transformed data. Then, the residuals of such fit constitute the remaining variance in protein activity after removing the expression effect. By definition, this residual post-translational protein activity (RPT activity) and the expression level of the coding genes are decoupled.

Association of somatic mutations with protein activity. We estimated the association between nonsilent somatic mutations and three quantitative traits: (i) mutated gene mRNA levels, (ii) VIPER-inferred global protein activity (G activity), and (iii) VIPER-inferred residual post-translational RPT activity, by computing the enrichment of the mutated samples on each of the traits using the aREA algorithm. An integrated association was obtained by taking the maximum association (minimum P value) among these traits.

The mutant phenotype score was computed by integrating the relative likelihoods of mutation for a given G- and RPT-activity level. Distribution densities for the mutated and non-mutated (WT) samples, for genes mutated in at least ten samples, were estimated by a Gaussian kernel, and the probabilities, computed by the derived cumulative distribution functions, were used to compute the relative likelihood for each trait as follows:

$$RL(x) = \frac{p_M(x) - p_{wt}(x)}{p_M(x) + p_{wt}(x)}$$

where p_M and p_{wt} are the estimated probabilities for mutant and WT phenotypes at a given value x of the evaluated trait, either G or RPT activity. MPS is then defined as the maximum deviance from zero of RL among the two evaluated traits.

49. Margolin, A.A. *et al.* ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* **7** (suppl. 1), 1 S7 (2006).
50. R Development Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2012).

51. Gautier, L., Cope, L., Bolstad, B.M. & Irizarry, R.A. affy—analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* **20**, 307–315 (2004).
52. Alvarez, M.J., Sumazin, P., Rajbhandari, P. & Califano, A. Correlating measurements across samples improves accuracy of large-scale expression profile experiments. *Genome Biol.* **10**, R143 (2009).
53. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).
54. Dudoit, S., Gentleman, R.C. & Quackenbush, J. Open source software for the analysis of microarray data. *Biotechniques* (suppl.), 45–51 (2003).
55. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).
56. Ding, B.B. *et al.* Constitutively activated STAT3 promotes cell proliferation and survival in the activated B-cell subtype of diffuse large B-cell lymphomas. *Blood* **111**, 1515–1523 (2008).
57. Kim, S.Y. & Volsky, D.J. PAGE: parametric analysis of gene set enrichment. *BMC Bioinformatics* **6**, 144 (2005).
58. Tian, L. *et al.* Discovering statistically significant pathways in expression profiling studies. *Proc. Natl. Acad. Sci. USA* **102**, 13544–13549 (2005).
59. Park, S.Y., Gönen, M., Kim, H.J., Michor, F. & Polyak, K. Cellular and genetic diversity in the progression of *in situ* human breast carcinomas to an invasive phenotype. *J. Clin. Invest.* **120**, 636–644 (2010).
60. Park, C.C. *et al.* β1 integrin inhibitory antibody induces apoptosis of breast cancer cells, inhibits growth, and distinguishes malignant from normal phenotype in three dimensional cultures and *in vivo*. *Cancer Res.* **66**, 1526–1535 (2006).
61. Wang, K. *et al.* Dissecting the interface between signaling and transcriptional regulation in human B cells. *Pac. Symp. Biocomput.* **2009**, 264–275 (2009).
62. Benaglia, T., Chauveau, D., Hunter, D.R. & Young, D.S. mixtools: an R package for analyzing finite mixture models. *J. Stat. Softw.* **32**, 1–29 (2009).