

Contents

1	Supplementary Note	1
1.1	Testing the incremental value of the different algorithms implemented in VIPER . . .	1
1.2	Comparison of VIPER with previous methods	2
1.3	Unbiased validation of VIPER-inferred protein activity using genetic perturbations .	3
1.4	Protein activity changes following pharmacologic perturbations	4
1.5	Comparison of VIPER results with Reverse Phase Protein Array data	5
2	Supplementary tables	7
3	Supplementary figures	13

1 Supplementary Note

1.1 Testing the incremental value of the different algorithms implemented in VIPER

To assess the incremental value of the additional refinements, we started with a naïve implementation of the algorithm that assessed enrichment of target genes against a gene expression signature (GES) ranked by absolute differential expression (1-tail approach). This can only assess the absolute change in protein activity but not its sign (i.e., activity increase or reduction). Significant activity changes were assessed for 4 of the 6 silenced proteins, two of which (*BCL6* and *MEF2B*) were inferred among the 10 most differentially active ones (Supplementary Fig. 3a and Supplementary Table 2).

To differentiate between activity increase and decrease, we integrated independent enrichment analyses of predicted positive (Spearman’s correlation coefficient (SCC) ≥ 0) and negative (SCC < 0) targets (2-tail analysis). This correctly inferred significantly decreased activity for all silenced proteins ($p < 0.05$) and showed improved accuracy and sensitivity for most assays, compared to 1-tail analysis (Supplementary Fig. 3a and 4 and Supplementary Table 2). However, the probabilistic mode of regulation model (3-tail analysis, see Methods section) outperformed both the 1-tail and the 2-tail approaches across all assays (Fig. 1e, Supplementary Fig. 3a and 4, and Supplementary Table 2). All six silenced TFs were inferred among the 10 most significant, with *FOXM1*, *MYB*, *BCL6* (Ly7), *STAT3*, *MEF2B*, and *BCL6* (Pfeiffer), ranking 1st, 1st, 1st, 1th, 5th, and 9th, respectively (Supplementary Fig. 3a and Supplementary Table 2).

Incorporation of the Interaction Confidence (IC) weight in the 3-tail analysis did not further improve accuracy, as there was virtually no margin for improvement (Fig. 1e and Supplementary

Fig. 3a). However, IC weight improved the accuracy of most 2-tail analysis results (Fig. 1e, Supplementary Fig. 3a and Supplementary Table 2), suggesting that IC weight provides independent information and improves algorithm performance. Based on these results we selected the 3-tail with IC correction (3T/IC) as the best performing method.

Detailed analysis of these results revealed that proteins whose regulon overlaps those of silenced TFs have higher enrichment than expected by chance. For instance, *MYBL1*, which had the most significant overlap with *MEF2B* (by Fisher’s Exact Test, FET), was the 2nd most significant TF following *MEF2B* silencing (see Supplementary Table 9 for a list of TFs with overlapping programs). These observations suggest that differential activity predictions may often be the result of significant regulon overlap with the bona fide differentially-active protein. Indeed, the Pleiotropy Correction (PC) analysis significantly improved specificity ($p < 0.02$, by paired U-test, Fig. 1e and Supplementary Fig. 3a, and Supplementary Table 2).

1.2 Comparison of VIPER with previous methods

We tested the Fisher Exact Test (1-tail FET) and its extension to explicitly account for the Mode of Regulation of a target gene (2-tail FET), as originally implemented in our Master Regulator Analysis (MRA) algorithm [4]. The latter accounts independently for targets that are either activated ($SCC \geq 0$) or repressed ($SCC < 0$) by the regulator, (see Methods section for details). We also compared the VIPER results to our previously described MAster Regulator INference algorithm (MARINA)[6], which computes enrichment based on 1-tail and 2-tail GSEA[8, 7]. Since MRA and MARINA require multiple samples ($N \geq 6$)[4, 6], these comparisons are limited to the multiple-sample version of VIPER (msVIPER).

The FET methods produced good accuracy for some of the experiments, but failed to capture the change in *FOXMI* and *STAT3* protein activity after their coding genes were silenced (Supplementary Fig. 3a and supplementary Table 2). This lack of consistency across all experiments could be related to the use of small, discrete gene lists by FET, which produces enrichments that are often not robust with respect to threshold selection (Supplementary Fig. 16). Despite the fact that GSEA eliminates the issue of threshold selection, it only partially improved the results previously obtained by FET (Supplementary Fig. 3a and Supplementary Table 2). More importantly, both FET and GSEA-based approaches showed a reduced accuracy, and in the case of GSEA also very poor specificity when compared to VIPER (Supplementary Fig. 3a and Supplementary Table 2).

We then tested the performance of VIPER when using tissue context-independent regulons assembled from experimentally supported interactions. For this we leveraged the ChIP-based ChEA and ENCODE databases ([5] and <http://amp.pharm.mssm.edu/Enrichr/#stats>), and inferred the MoR from tissue-matched expression profile data as described in methods. In agreement with

the context-specificity of most of the TF regulatory programs (Fig. 2c), we found a weaker performance of this analysis when compared against the ARACNe context-specific-based msVIPER analysis for all TFs but *FOXM1*, whose program seems to be more conserve across tissues (Supplementary Fig. 3c and Fig. 2c). *MEF2B* and *BCL6* could not be evaluated because their transcriptional program was not represented in the ChEA and ENCODE models, which included only 189 and 172 regulatory programs, respectively.

Finally, we compared msVIPER performance against the upstream regulator analysis module of Ingenuity Pathway Analysis (IPA). msVIPER outperformed IPA for all the tested regulators in our benchmark experiment. In fact, IPA inferred correctly a decrease in the knocked-down TF protein activity only for *FOXM1*, and *MEF2B* could not be evaluated since it was not represented in the IPA results (Supplementary Fig. 3c).

1.3 Unbiased validation of VIPER-inferred protein activity using genetic perturbations

To further benchmark the algorithm, we expanded our panel of gene knock-down data to 23 silencing experiments performed in breast carcinoma cells, covering 19 genes and 12 different cell lines whose profiles are available from Gene Expression Omnibus. For this analysis we used breast cancer specific regulons, inferred by ARACNe analysis of 1,037 TCGA breast carcinoma gene expression profiles (Table 1). VIPER analysis, using the full probabilistic model implemented by the aREA algorithm, detected a significant protein activity dysregulation for 20 of the 23 silencing experiments (87%, $p < 0.05$). The activity of 17 proteins was inferred as significantly decreased in response to coding gene knock-down, while 3 were inferred as significantly activated (Supplementary Fig. 6a). Use of 2-tail GSEA for VIPER analysis was consistently less sensitive and accurate than aREA, detecting 14 of the 23 assessed proteins (61%) as significantly dysregulated at $p < 0.05$ (Supplementary Fig. 6). Moreover, GSEA was dramatically more computationally demanding than aREA (6.7 min of computer time for the aREA implementation vs. 23 days and 6 hr of computer time for 2-tail GSEA implementation, measured in an 8 Gb RAM x86_64 1.2 GHz computer node).

We further expanded this analysis by leveraging gene expression profiles generated following shRNA-mediated silencing of 234 regulatory proteins in MCF7 cells, from the Library of Integrated Network-based Cellular Signatures (LINCS). LINCS represents a large repertoire of expression profiles following shRNA silencing of 3,680 genes. However, to ensure proper knock-down of the silenced gene, experiments were selected based on two criteria: (1) silenced genes had to be among the 978 experimentally assessed genes, such that their silencing could be assessed and (2) their expression had to be reduced by at least 2 standard deviations (SD), compared to the average across controls. $SD \geq 2$ emerged as a reasonable compromise between selecting assays with effective gene

silencing and having enough samples for a representative analysis. Since LINCS expression profiles are based on only 978 genes (i.e., <5% of a regulons genes, on average) by multiplexed Luminex technology (L1000), performance analysis on this dataset should be considered an extremely conservative lower bound. VIPER analysis detected a statistically significant protein activity decrease for 44 (50%, $p < 0.05$) of 87 silenced TFs (Supplementary Fig. 7a), while only 4 TFs were predicted as significantly activated following silencing (Supplementary Fig. 7a). Similarly, VIPER detected statistically significant protein activity decrease for 57 of the 147 silenced signaling proteins (39%, $p < 0.05$), while only 7 were predicted as significantly activated following silencing (Supplementary Fig. 7b). Interestingly, MoR can be incorrectly inferred for some genes, due to regulatory feedback loops that induce inverse correlation between gene expression and protein activity for a small number of proteins, more frequently among signal transduction ones. This is consistent with $\sim 10\%$ of silenced proteins being inferred with significantly increased activity. Thus MoR inversion may need to be experimentally evaluated within specific tissue contexts.

1.4 Protein activity changes following pharmacologic perturbations

Short-term perturbation with targeted inhibitors typically modulates protein activity, without affecting associated gene expression. We thus leveraged the MCF7 connectivity map (CMAP) dataset¹, which contains 3,095 gene expression profiles of MCF7 cells, following perturbation with 1,294 compounds. Among targeted TFs, the estrogen receptor (*ESR1*) has the highest number of samples ($n = 27$) and inhibitor diversity in this dataset, according to drugbank[10], including fulvestrant, tamoxifen and clomifene. We thus tested whether *ESR1* inhibition by these compounds was effectively recapitulated by VIPER analysis, using a breast cancer specific ARACNe network (Table 1). VIPER-inferred *ESR1* differential activity in samples treated with estrogen inhibitors was computed from their differential gene expression signature against matched DMSO-treated controls. P-values from replicated samples were integrated by the Stouffer’s method. VIPER inferred statistically significant, dose-dependent decrease in estrogen receptor protein activity for all three targeted inhibitors (Supplementary Fig. 8). To extend the analysis to signaling proteins, we evaluated the effect of sirolimus, an inhibitor of the *FKBP1A* and *MTOR* proteins, as the one with the highest number of treatment replicates ($n = 25$). Consistently, VIPER inferred significant protein activity decrease for both *FKBP1A* and *MTOR* (Supplementary Fig. 8). These results show that VIPER can effectively detect protein activity dysregulation in response to short term pharmacologic perturbations, and encouraged us to extend this analysis to the remaining profiled compounds, complementing in this way the MCF7-CMAP dataset by adding the protein activity layer.

¹<http://www.broadinstitute.org/cmap/>

To maximize the reliability of the results, we only included perturbations performed at least in duplicate and for which we could verify a significant correlation between the gene expression signatures (FDR < 0.05, Spearman’s correlation analysis). Briefly, we computed the mean correlation for each sample $k \in P$, where P is a set of replicated perturbation conditions, as the mean Pearson’s correlation coefficient between all sample pairs $k \times j | j \in P$. The correlation was computed between the rank-transformed signatures downloaded from the CMAP web page¹. Statistical significance was estimated by comparison against the empirical distribution of correlation coefficients obtained between each rank-transformed signature and all remaining non-matching drug perturbation signatures, i.e. $k \times j, \forall k, j | k \in P, j \notin P$.

We then used VIPER, together with a breast carcinoma context specific interactome (Table 1), to transform 573 gene expression signatures satisfying the reproducibility condition (see previous paragraph), into inferred protein activity signatures. The mean and standard deviation across replicated samples is reported in Supplementary Table 7. This represents an unbiased portrait for the effect of 166 unique perturbation conditions, encompassing 156 distinct small molecule compounds, on the activity of 2,956 regulatory proteins.

1.5 Comparison of VIPER results with Reverse Phase Protein Array data

Finally, to benchmark VIPER using a gold standard for which both gene expression and protein abundance were experimentally measured, we leveraged sample-matched RNAseq and RPPA data for 4,417 tumor samples, across 17 tumor types. RPPA arrays monitor an average of 135 proteins and 60 phospho-specific isoforms per tumor type (Supplementary Table 4). Protein regulons were inferred by ARACNe analysis of the corresponding gene expression profile datasets (see Table 1). VIPER-inferred activity significantly correlated with RPPA-based protein abundance for 875 of the 1,359 tumor specific protein abundance profiles (64.4%, $p < 0.05$, Supplementary Table 5). While similar correlation between gene expression and protein abundance was also observed (Supplementary Table 5), the latter had much larger variance at the individual sample level (see main text, Fig. 3b).

To use the RPPA data to estimate the changes in protein activity, associated with post-translational protein modifications, we measured the ratio between the RPPA-measured abundance of 443 individual isoforms and their total protein abundance. Overall, protein activity may depend on either total protein abundance or on the abundance of specific, differentially active isoforms. To distinguish between these two contributions, we estimated both global VIPER activity, as well as the residual post-translational VIPER activity (i.e., the component of activity that cannot be accounted for by differential expression), by removing the transcriptional variance component (RPT-activity, see Methods). Thus, by definition, RPT-activity is statistically independent of gene expression and

should thus account for the purely post-translational contribution to protein activity. Remarkably, when taken together, global and RPT-activity were predictive for the abundance of 105 protein isoforms (24%, $p < 0.05$, Spearman's correlation analysis), which significantly outperformed the 38 isoforms (8.6%) predicted by mRNA expression ($p = 8 \times 10^{-10}$ by X^2 test). Individually, RPT activity was predictive for 77 isoforms (17.4%, $p = 7 \times 10^{-5}$), of which only 19 were also predicted by global activity, while global activity was predictive for 47 isoforms (10.6%), suggesting that global- and RPT-activity effectively account for mostly complementary effects (Supplementary table 6). Since not all post-translational modified isoforms have differential protein activity (Fig. 1a), not all isoform specific antibodies provide accurate RPPA measurements, and most isoforms present little abundance variability in TCGA cohorts, this represents a substantial fraction ($> 24\%$) of RPPA monitored proteins. Overall, of 105 VIPER correlated isoforms, 74 (70.5%) were undetectable by differential expression, while only 7 of the isoforms captured by differential expression were missed by VIPER (Supplementary Fig. 9).

2 Supplementary tables

Supplementary Table 1: List of acronyms used in the manuscript.

Acronym	Definition
aREA	analytic Rank-based Enrichment Analysis
aREA-3T	3-tail aREA analysis
CDF	Cumulative Distribution Function
CMAP	Connectivity MAP
COSMIC	Catalogue Of Somatic Mutations In Cancer
ES	Enrichment Score
FET	Fisher's Exact Test
GES	Gene Expression Signature
GSEA	Gene Set Enrichment Analysis
IC	Interaction Confidence
LINCS	Library of Integrated Network-based Cellular Signatures
MARINa	Master Regulator Inference algorithm
MoR	Mode of Regulation
MPS	Mutant Phenotype Score
NES	Normalized Enrichment Score
NSSM	Non-Silent Somatic Mutations
PC	Pleiotropy Correction
PDE	Pleiotropy Differential Score
RPPA	Reverse Phase Protein Arrays
RPT	Residual Post-Translational
SCC	Spearman's Correlation Coefficient
TCGA	The Cancer Genome Atlas
TF	Transcription Factor
VIPER	Virtual Inference of Protein-activity by Enriched Regulon analysis
WT	Wild Type

Supplementary Table 2: Accuracy and specificity of Fisher’s Exact Test (FET), Gene Set Enrichment Analysis (GSEA) and msVIPER for the detection of a reduction in protein activity after coding gene silencing. The table lists the accuracy (rank for the silenced gene), specificity (number of significant regulators at $p < 0.05$) and silenced gene p -value inferred by 1-tail (1T) and 2-tail (2T) FET and GSEA, and by the 1-tail, 2-tail and 3-tail implementations of msVIPER, including Interaction Confidence (IC) analysis and Pleiotropy Correction (PC).

		FET		GSEA		msVIPER							
		1T	2T	1T	2T	1T	2T	2T/IC	2T/PC	3T	3T/IC	3T/PC	3T/IC/PC
MEF2B P3HR1	Accuracy	34	11	143	16	6	5	3	6	5	4	6	3
	Specificity	179	43	271	98	132	88	95	70	87	95	79	89
	p-value	4.77E-07	1.24E-08	0.00147	0.00418	1.23E-12	0.000271	0.000127	0.000624	0.000355	0.000157	7.00E-04	0.000164
FOXM1 ST486	Accuracy	240	17	241.5	1	328.5	1	1	1	1	2	1	2
	Specificity	58	3	235	12	88	8	10	7	16	21	15	20
	p-value	0.346	0.145	0.0528	0.00584	0.434	0.005	0.0035	0.0025	0.00116	0.00165	0.0055	0.004
MYB ST486	Accuracy	7	2	117	3	43	1	1	1	1	1	1	1
	Specificity	97	6	245	47	116	26	36	22	49	54	38	42
	p-value	0.000261	8.31E-05	0.00462	0.00429	0.000194	0.003	0.00248	0.003	0.00029	0.00141	1.06E-06	0.000271
BCL6 Ly7	Accuracy	3	1	97	16	12	13	9.5	7.5	1	1	1	1
	Specificity	133	13	403	98	191	94	95	88	130	128	116	114
	p-value	3.39E-07	0.00057	0.00244	0.00834	1.27E-07	0.011	0.0075	0.0085	0.000153	0.000185	0.000111	0.000358
BCL6 Pfeiffer	Accuracy	1	11	78	18	6	14.5	16	17	9	11	4	3
	Specificity	216	25	422	141	197	82	98	74	133	139	119	127
	p-value	4.52E-11	0.0123	0.00164	0.00751	8.90E-14	0.009	0.0095	0.022	0.00177	0.00202	0.00131	0.00209
STAT3 SNB19	Accuracy	774	247	702	31	258	10	9	7	1	2	1	1
	Specificity	76	0	304	68	111	50	54	48	60	75	47	69
	p-value	0.911	0.499	0.495	0.018	0.209	0.01	0.005	0.0125	0.000661	0.00101	0.000463	0.000658

Supplementary Table 3: Accuracy and specificity of VIPER for the detection of a reduction in protein activity after coding gene silencing. The table lists the accuracy (rank for the silenced gene), specificity (number of significant regulators at $p < 0.05$) and silenced gene p -value inferred by the 1-tail (1T), 2-tail (2T) and 3-tail (3T) implementations of VIPER, including Interaction Confidence (IC) analysis and Pleiotropy Correction (PC).

		1T	2T	2T/IC	2T/PC	3T	3T/IC	3T/PC	3T/IC/PC
MEF2B P3HR1	Accuracy	27	13	13	23	24	21	23	22
	Specificity	142	130	114	84	104	97	80	83
	p-value	5.65E-05	8.35E-23	4.01E-21	2.88E-08	8.14E-15	3.03E-16	3.04E-07	1.14E-07
FOXM1 ST486	Accuracy	289	1	1	1	1	1	1	1
	Specificity	41	18	12	7	20	13	5	4
	p-value	0.384	4.30E-10	2.28E-10	6.19E-06	1.50E-10	3.36E-11	7.63E-05	5.57E-05
MYB ST486	Accuracy	23	2	2	3	4	4	5	2
	Specificity	69	92	65	37	74	62	37	35
	p-value	0.00968	1.72E-07	9.16E-07	0.000742	1.92E-05	7.46E-06	0.00515	0.00176
BCL6 Ly7	Accuracy	9	25	25	18	16	13	13	15
	Specificity	106	262	193	176	222	181	151	122
	p-value	0.000584	0.000119	0.00036	0.00103	5.52E-05	5.79E-05	0.00157	0.00262
BCL6 Pfeiffer	Accuracy	8	112	87	69	39	37	21	22
	Specificity	280	368	301	255	321	277	225	198
	p-value	1.03E-06	0.000102	0.000285	0.00218	1.24E-06	4.21E-06	4.98E-05	0.000117
STAT3 SNB19	Accuracy	767	6	2	11	1	1	4	2
	Specificity	17	55	40	19	38	33	18	14
	p-value	0.813	0.00136	0.000394	0.0307	6.91E-05	0.00022	0.0163	0.0122

Supplementary Table 4: Number of profiled samples, and profiled proteins and isoforms per sample in the RPPA dataset from TCGA.

	BLCA	BRCA	COAD	GBM	HNSC	KIRC	LGG	LUAD	LUSC	OV	PRAD	READ	SARC	SKCM	STAD	THCA	UCEC
Samples	127	410	331	214	212	454	260	181	195	412	164	130	227	206	264	430	200
Proteins	138	112	131	131	129	120	147	129	135	130	147	131	150	142	147	148	128
Isoforms	63	51	59	59	56	56	64	56	59	55	64	59	63	61	64	64	57

Supplementary Table 5: Number of RPPA profiled proteins and significant associations at the transcripts (mRNA expression) and VIPER-inferred global protein activity (G-activity) levels ($p < 0.05$, Spearman's correlation analysis).

	BLCA	BRCA	COAD	GBM	HNSC	KIRC	LGG	LUAD	LUSC	OV	PRAD	READ	SARC	SKCM	STAD	THCA	UCEC	TOTAL
RPPA profiled	81	66	85	84	76	74	88	75	83	81	85	84	58	84	87	86	82	1359
mRNA expression	58	60	66	58	56	64	61	58	64	70	52	57	43	65	67	57	64	1020
G-activity	52	52	55	49	47	58	57	51	62	68	37	39	38	62	54	46	48	875

Supplementary Table 6: Number of RPPA profiled protein isoforms and significant associations at the transcripts (mRNA expression), VIPER-inferred global protein activity (G-activity), residual post-translational VIPER-inferred activity (RPT-activity) and their integration (Integrated activity) with the protein isoform levels at $p < 0.05$ by Spearman’s correlation analysis.

	BLCA	BRCA	COAD	GBM	HNSC	KIRC	LGG	LUAD	LUSC	OV	PRAD	READ	SARC	SKCM	STAD	THCA	UCEC	TOTAL
RPPA profiled	30	23	28	28	21	22	29	21	28	25	29	28	19	28	29	29	26	443
mRNA expression	1	1	3	1	2	2	4	3	1	2	4	2	1	1	4	3	3	38
G-activity	1	1	4	5	2	3	4	3	2	2	6	0	2	1	3	3	5	47
RPT-activity	4	6	8	3	4	5	4	3	5	3	6	3	3	4	3	4	9	77
Integrated activity	5	7	10	7	5	6	7	5	6	5	9	3	5	5	5	5	10	105

Supplementary Table 7: VIPER-inferred protein activity for 2,956 regulatory proteins (rows) after perturbation of MCF7 cells with 156 small molecule compounds (columns). Protein activity predictions were based on the Connectivity Map dataset[9] and a breast carcinoma regulatory network inferred from 1,037 RNAseq profiles from TCGA. The table lists the VIPER-inferred relative protein activity as normalized enrichment score values. The first two columns indicate the geneID and gene symbol corresponding to the evaluated regulatory protein. The first row lists the small molecule compound, molar concentration and time of exposure.

See Supplementary -Table.7.xlsx file

Supplementary Table 8: Number of samples harboring non-silent somatic mutations in COSMIC genes.

See Supplementary-Table.8.xlsx file

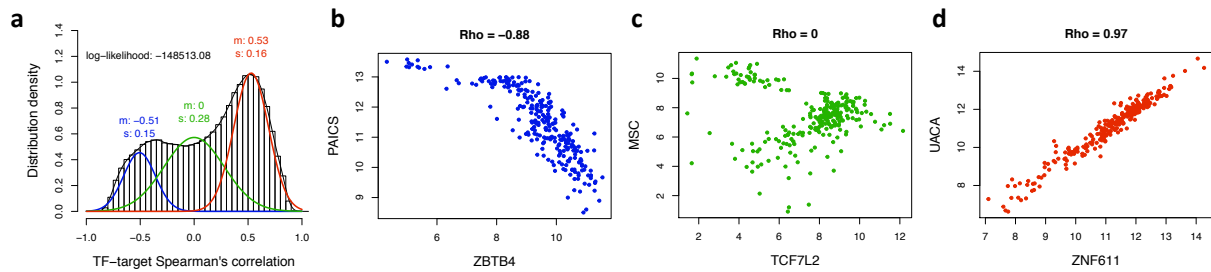
Supplementary Table 9: Regulon overlap analysis.

Listed are the regulons showing a significant enrichment on the gene expression signature ($p < 0.05$) and a significant overlap with the knocked-down TF regulon. The numbers indicate the odds ratio and Fisher's exact test p-value when testing that the overlap between the knocked-down and listed regulon is larger than the overlap expected by chance.

MEF2B P3HR1	FOXM1 ST486	MYB ST486	BCL6 Ly7	BCL6 Pfeifer	STAT3 SNB19
MYBL1 4.22 (9.66e-10)	ILF3 2.32 (5.11e-06)	FOXM1 2.31 (0.000785)	CUX1 11.63 (1.74e-39)	ZEB2 7.19 (1.32e-55)	IRF1 3.54 (8.49e-12)
BCL6 2.73 (1.99e-06)	BCL6 1.96 (0.000132)	PLAGL1 2.17 (0.00246)	ZFP64 5.71 (7.89e-32)	HHEX 5.49 (1.26e-46)	ZNF529 3.12 (8.59e-10)
CUX1 2.67 (1.88e-05)	STAT5A 1.95 (0.000267)		IKZF2 4.14 (1.74e-30)	BACH2 6.55 (1.43e-45)	HLX 2.9 (1.15e-09)
BACH1 2.83 (2e-05)	KLF10 2.1 (0.000479)		MYBL1 6.11 (1.05e-23)	ZNF828 6.65 (4.82e-45)	GATAD1 2.9 (5.23e-08)
ESR2 3.06 (4.42e-05)			MEF2B 5.35 (2.33e-20)	TGIF1 6.19 (7.38e-40)	ATF5 3.09 (1.73e-07)
KLF9 2.53 (5.97e-05)			ZBTB32 5.58 (2.31e-17)	CUX1 11.63 (1.74e-39)	MAZ 2.8 (2.3e-07)
MORC3 2.62 (6.35e-05)			LHX2 5.28 (1.4e-16)	IKZF1 7.18 (1.36e-33)	IRF7 2.75 (2.4e-07)
CLOCK 2.71 (7.25e-05)			SCML1 2.61 (3.03e-16)	IKZF2 4.14 (1.74e-30)	BCL3 2.26 (8.7e-06)
ZMYND11 2.22 (0.000139)			HOXA5 4.65 (5.35e-16)	NOTCH2 4.6 (8.8e-30)	ZNF248 2.21 (1.13e-05)
E2F5 2.69 (0.000153)			MTA3 4.78 (5.64e-12)	ZNF74 7.17 (1.11e-28)	TEAD3 2.47 (1.37e-05)
CREB3L2 2.16 (0.000583)			DDIT3 5.08 (2.56e-11)	LYL1 5.4 (4.87e-24)	CAMTA1 2.04 (2.14e-05)
PTTG1 1.97 (0.000899)			ETV6 3.64 (1.45e-10)	MYBL1 6.11 (1.05e-23)	ZNF142 2.23 (2.14e-05)
ZEB2 2.2 (0.0013)			SMAD2 4.49 (3.3e-10)	ZBTB32 5.58 (2.31e-17)	TAF5L 2.47 (9.13e-05)
ZNF248 2.25 (0.0021)			SCMH1 3.85 (1.44e-09)	TFEC 4.53 (9.76e-17)	ZNF3 2.33 (0.000129)
ETV6 2.33 (0.00238)			HOXA1 3.98 (3.99e-09)	E2F7 4.4 (5.1e-16)	ZNF365 1.9 (0.000252)
IRF5 2.43 (0.00384)			ZNF318 3.73 (4.96e-09)	BCL11A 3.8 (2.23e-15)	ZNF638 2.26 (0.000266)
MYBL2 2.14 (0.0053)			ZNF354A 3.65 (3.67e-08)	IRF8 4.31 (2.27e-14)	JUNB 2.39 (0.000343)
TADA3 2.03 (0.00693)			BATF3 2.86 (5.86e-08)	SP140 3.36 (6.51e-14)	CEBPD 2.15 (0.000423)
SRF 1.99 (0.00835)			HDAC1 2.8 (7.64e-07)	IRF4 3.53 (2.43e-13)	MSRB2 1.75 (0.000501)
CSDA 1.91 (0.00837)			POU2F2 2.79 (9.92e-06)	MTA3 4.78 (5.64e-12)	LASS2 1.97 (0.0014)
			WHSC1 2.4 (1.57e-05)	CREB3L2 3.32 (1.15e-11)	NFYA 1.88 (0.00305)

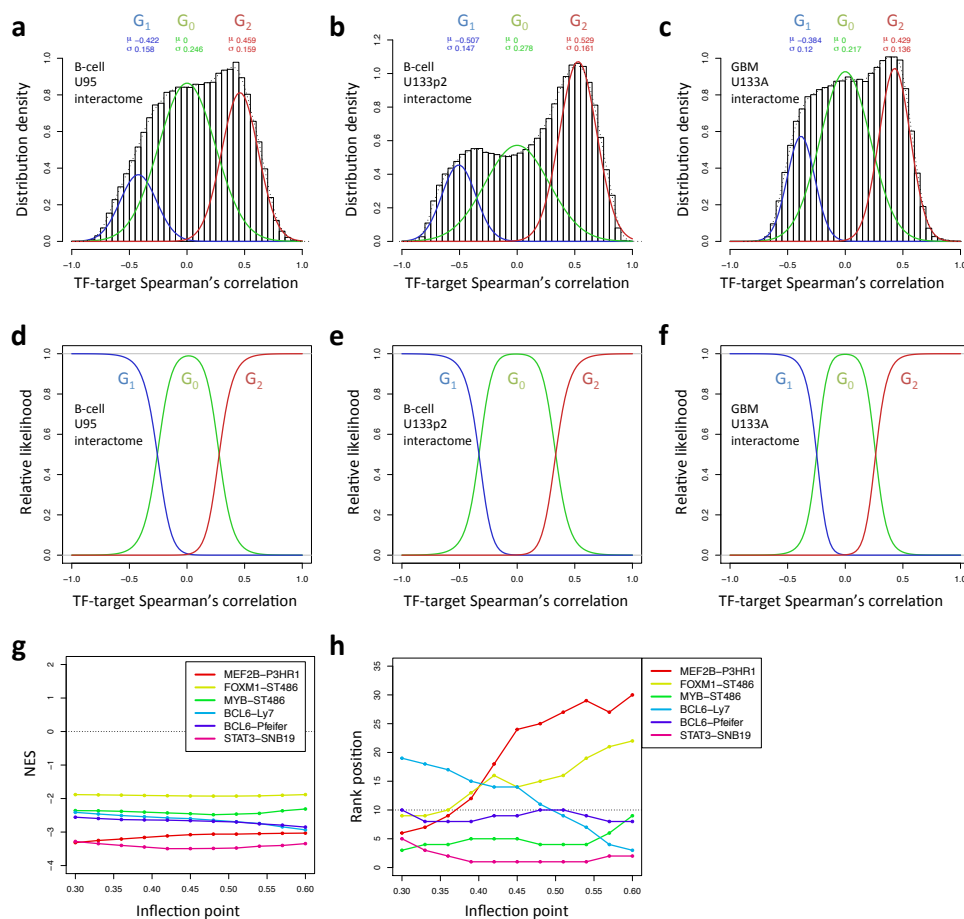
MEF2B P3HR1	FOXM1 ST486	MYB ST486	BCL6 Ly7	BCL6 Pfeifer	STAT3 SNB19
			IRF1 2.25 (4.98e-05)	MEIS2 4.81 (1.99e-11)	MLL4 1.83 (0.00368)
			EGR1 2.64 (8.89e-05)	NFIA 3.29 (2.35e-11)	RELB 1.86 (0.00397)
			ZNF385B 2.47 (7e-04)	AHR 3.2 (5.61e-11)	TCF20 1.9 (0.00511)
			ILF3 1.96 (0.000702)	SMAD2 4.49 (3.3e-10)	ESRRG 1.88 (0.00777)
			LITAF 1.73 (0.000985)	HOXA1 3.98 (3.99e-09)	YBX1 1.7 (0.00905)
			XBP1 2.35 (0.0012)	BATF3 2.86 (5.86e-08)	AHR 1.78 (0.00982)
			TSHZ1 2.12 (0.0014)	ZNF215 3.75 (6.34e-08)	
			HOXA10 2.31 (0.00145)	ESR2 6.57 (1.28e-07)	
			TAF4B 2.11 (0.00147)	HDAC1 2.8 (7.64e-07)	
			ZNF362 1.85 (0.00225)	ARID3A 4.27 (9.61e-07)	
			ZNF775 1.93 (0.00352)	SATB1 2.45 (1.43e-06)	
			TCF4 2.15 (0.00385)	ZHX2 2.8 (6.57e-06)	
			KDM1A 2.05 (0.00735)	XBP1 2.35 (0.0012)	
				TSHZ1 2.12 (0.0014)	
				ETS2 2.17 (0.00171)	

3 Supplementary figures



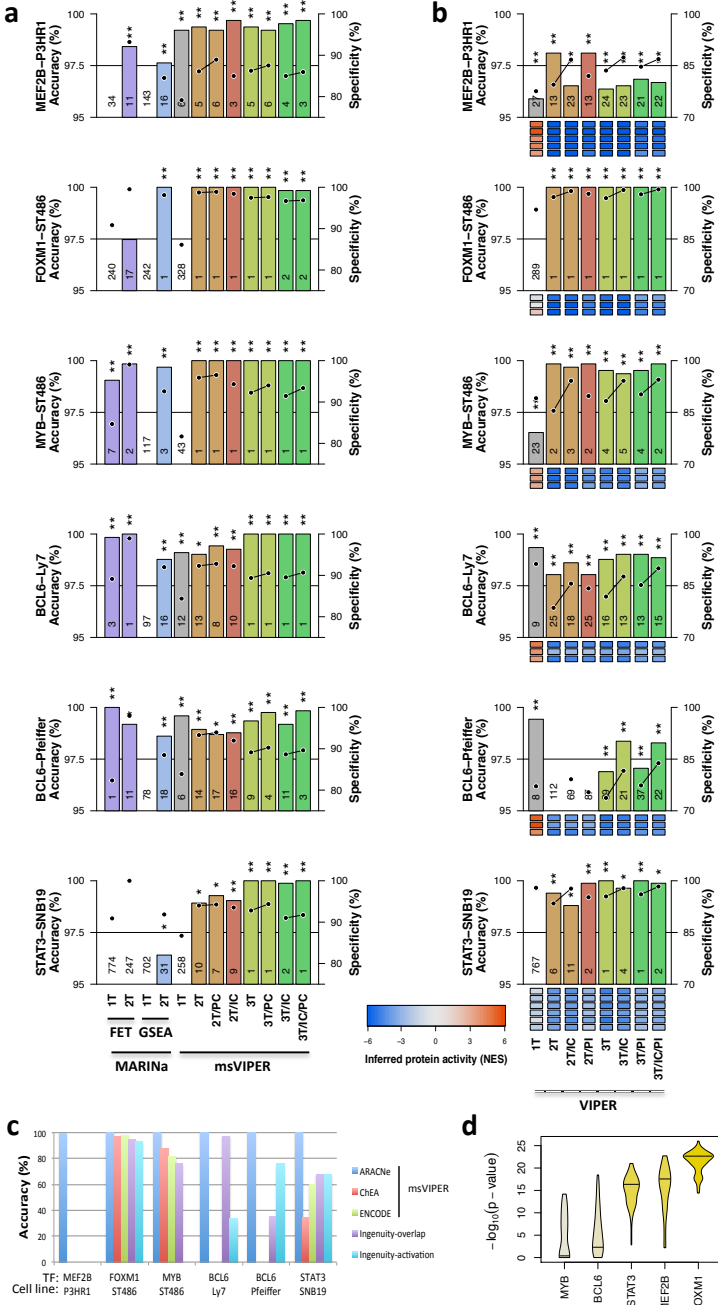
Supplementary Figure 1: TF Mode of Regulation for the B-cell U133-based interactome.

(a) TF-target Spearman's correlation coefficient distribution fitted to 3-Gaussian models mixture using the mixtools package for R[3]. Shown is the estimated mean (m) and standard deviation (s) for each distribution and the final log-likelihood for the fit. (b-d) Scatter-plots for TF (*x-axis*) and target genes (*y-axis*) showing the most negative (b), weakest (c) and most positive (d) Spearman's correlation coefficient.

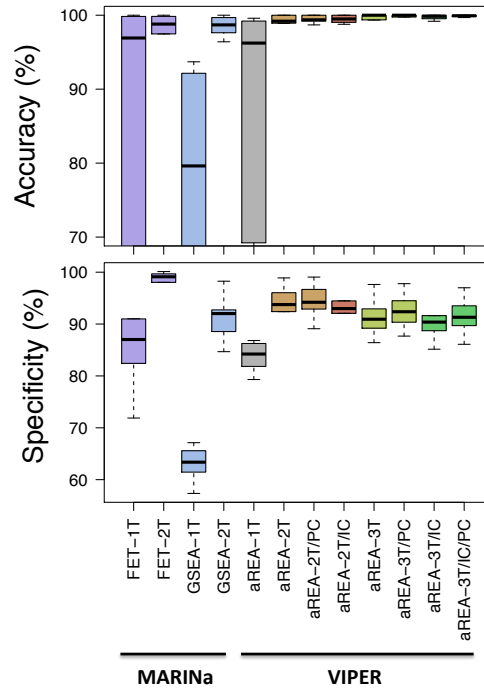


Supplementary Figure 2: Inferring the Mode of Regulation (MoR).

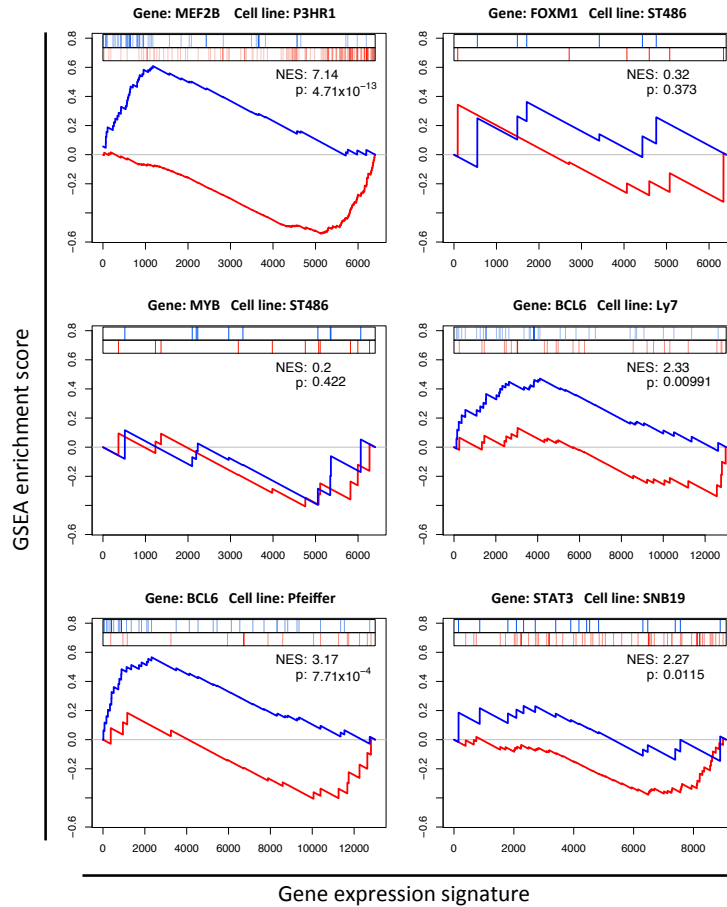
(a-c) Histogram and distribution density (dotted line) of the TF-target Spearman's correlation coefficient for the B-cell U95, U133plus2 and GBM U133A interactomes, respectively. Shown are also three Gaussians distributions that were fitted to the data (G_1 for the repressed targets, G_2 for the induced targets, and G_0 for the targets which more cannot be determined), whose parameters are shown in the figure. (d-f) Proportion of G_1 (blue), G_0 (green) and G_2 (red) gaussian distributions relative to all three distributions for each interactome. (g and h) Effect of the 'mean' parameter in G_1 and G_2 on the VIPER-inferred relative protein activity shown as NES (g) and the rank position of the silenced TF (h). Each line represents the result from a different benchmarking experiment.



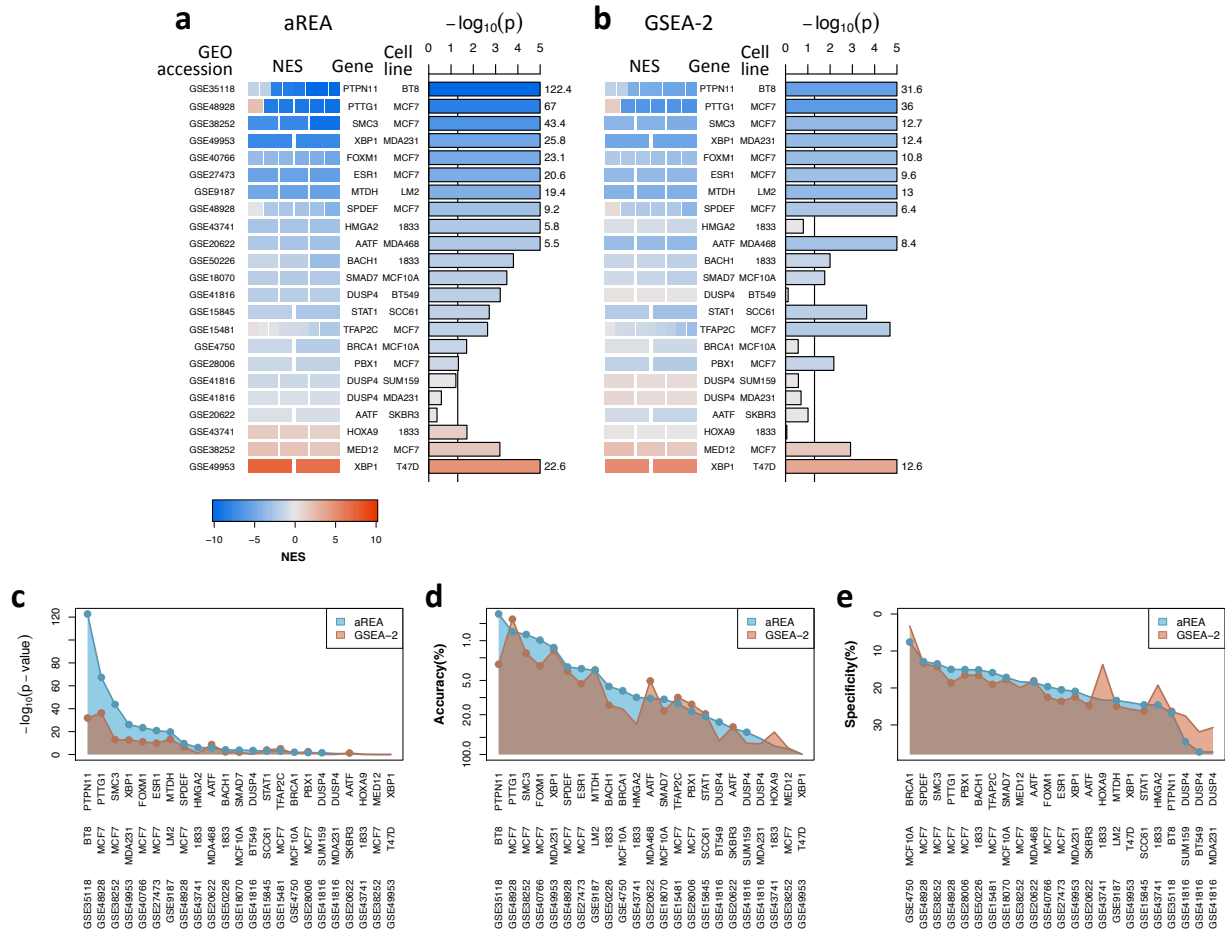
Supplementary Figure 3: Accuracy and specificity of VIPER for GES computed from multiple samples (msVIPER, a) and for single-samples VIPER (b). Barplot showing the accuracy (relative rank for the silenced gene) and dots showing the specificity (fraction of significant regulators at $p < 0.05$) for six benchmark experiments. The numbers in the bars indicate the rank position for the protein coded by the silenced genes. Panel (a) shows also results obtained by 1-tail and 2-tail versions of FET and GSEA, as previously implemented in MARiNa. VIPER results shown include the 1-tail (1T), 2-tail (2T) and 3-tail (3T) implementations of the aREA algorithm, including Interaction Confidence (IC) analysis and Pleiotropy Correction (PC). Different bar colors were used to highlight the use of different algorithms. Colored boxes under the bars in panel (b) show the single-sample estimation of relative protein activity. * $p < 0.05$, ** $p < 0.01$, estimated by permutation analysis as described in methods. (c) Accuracy of msVIPER based on alternative regulatory models (ARACNe, ChEA and ENCODE), and of Ingenuity upstream regulator analysis, for the six benchmark experiments. (d) Regulon functional conservation across 17 tissue context specific networks reverse-engineered by ARACNe from TCGA data (Table 2). Regulaon conservation was computed as described in Aytes et.al[1] and expressed as $-\log_{10}(p\text{-value})$. Shown is the density distribution for all 136 network pairs.



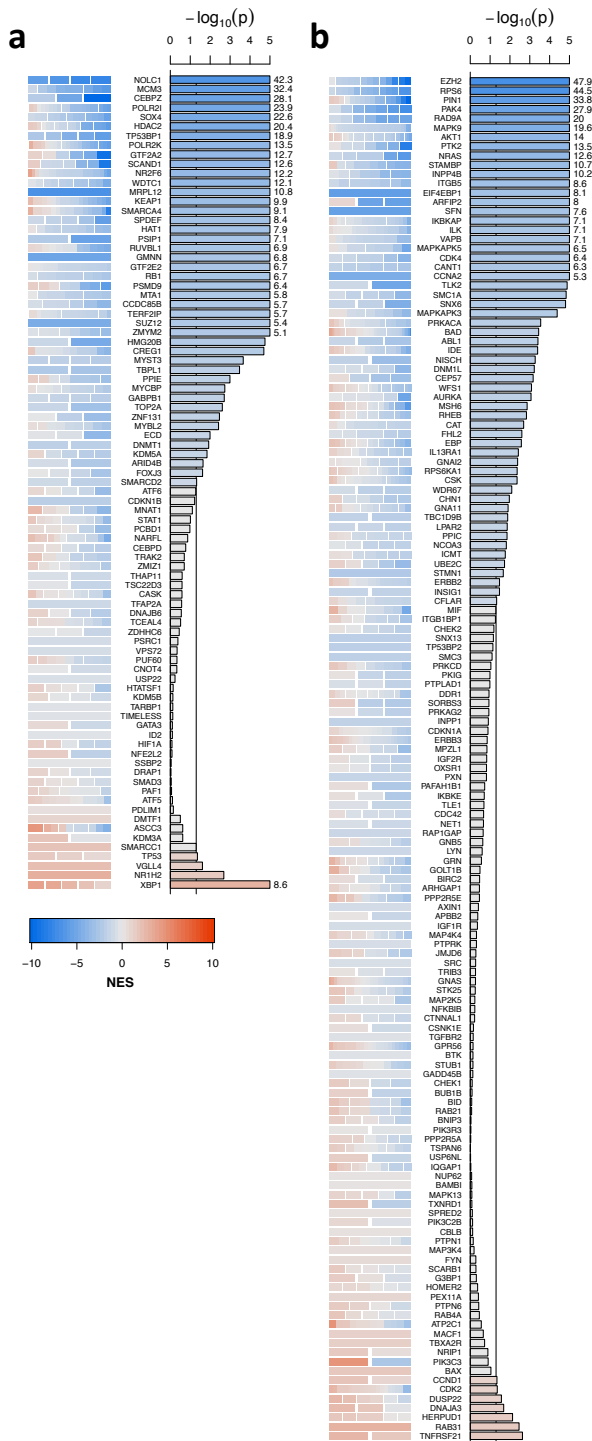
Supplementary Figure 4: Accuracy and specificity of protein activity inferred by different algorithms, including 1-tail (1T) and 2-tail (2T) Fisher’s Exact Test (FET), 1-tail and 2-tail Gene Set Enrichment Analysis (GSEA), and the 1-tail, 2-tail, 3-tail implementations of aREA, with Interaction Confidence (IC) and Pleiotropy Correction (PC). Boxplots showing the accuracy (relative rank for the silenced gene) and specificity (fraction of significant regulators at $p < 0.05$) for six benchmark experiments (see Table 2). Different colors were used to highlight the different algorithms. Related to Figure 1e.



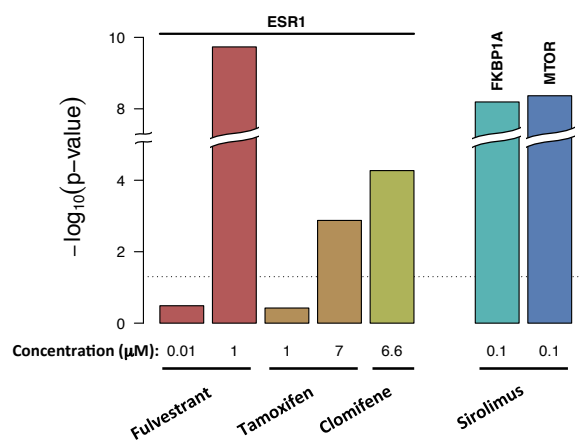
Supplementary Figure 5: Gene Set Enrichment Analysis for the regulators identified by VIPER as differentially active ($p < 0.05$) in each benchmark experiment, on the corresponding experiment gene expression signature. The silenced gene and cell line is indicated on top of each plot. The horizontal axis represents the profiled genes sorted from the most down-regulated (on the left) to the most up-regulated (on the right). Only genes represented in the regulatory network were used for this analysis, including 6,403 genes for P3HR1 and ST486, 13,007 for Ly7 and Pfeiffer, and 8,263 genes for SNB19. The vertical axes indicate the GSEA enrichment score for the regulators showing a decreased (blue) or increased (red) VIPER-inferred protein activity.



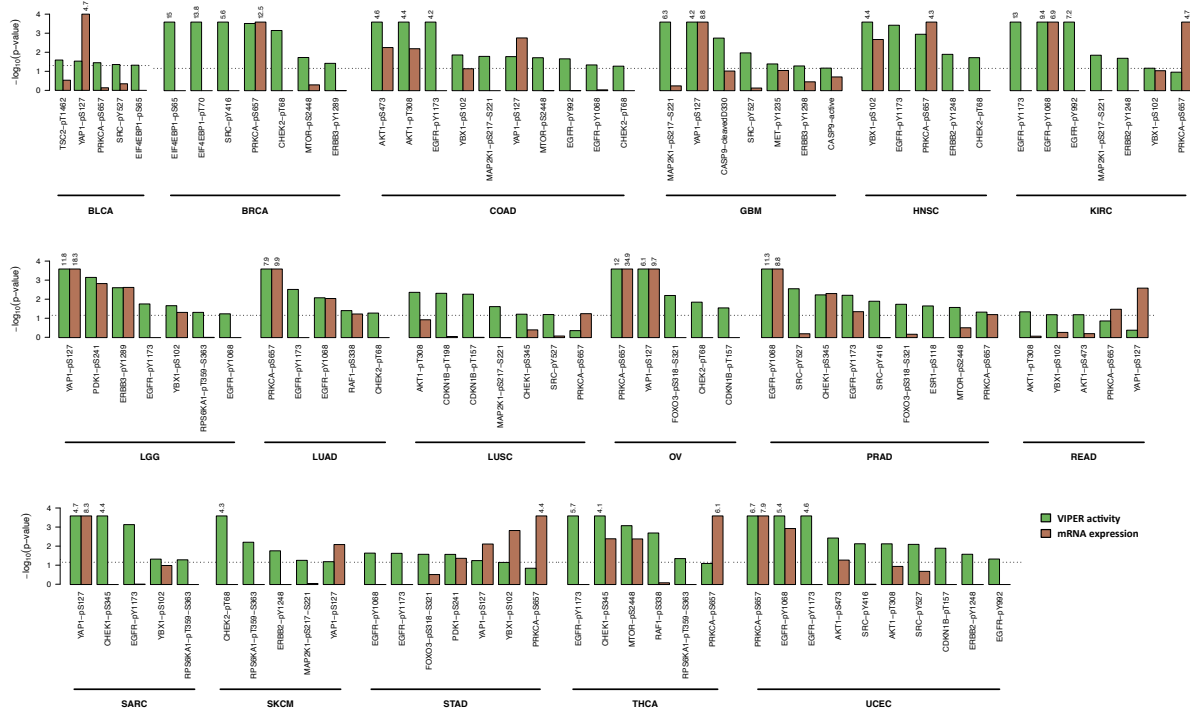
Supplementary Figure 6: Detecting changes in protein activity after genetic perturbations. (a and b) Heatmaps showing the VIPER-inferred change in TF protein activity based on aREA (a) and 2-tail GSEA (b) enrichment methods. Displayed results correspond to silencing experiments in breast carcinoma cells. Statistical significance was estimated by Stouffer's integration of the single-sample NES (i.e. $\frac{1}{\sqrt{k}} \sum_{i=1}^k NES_i$). The vertical black line crossing the bar-plot indicates the significance threshold at $p = 0.05$. Bars showing a statistically significant change in protein activity at $p < 0.05$ are highlighted in blue (decreased protein activity, i.e. $NES < 0$) and red (increased protein activity, $NES > 0$). Values higher than the axis scale are indicated to the right of each bar. (c - e) VIPER analysis of 23 silencing experiments in breast carcinoma cells, using aREA (blue) or 2-tail GSEA (red) as gene enrichment methods. (c) Statistical significance for protein activity decrease expressed as $-\log_{10}(p\text{-value})$, (d) Accuracy expressed as rank position percentage of the evaluated regulators, and (e) Specificity, expressed as proportion (%) of regulators inferred as differentially active.



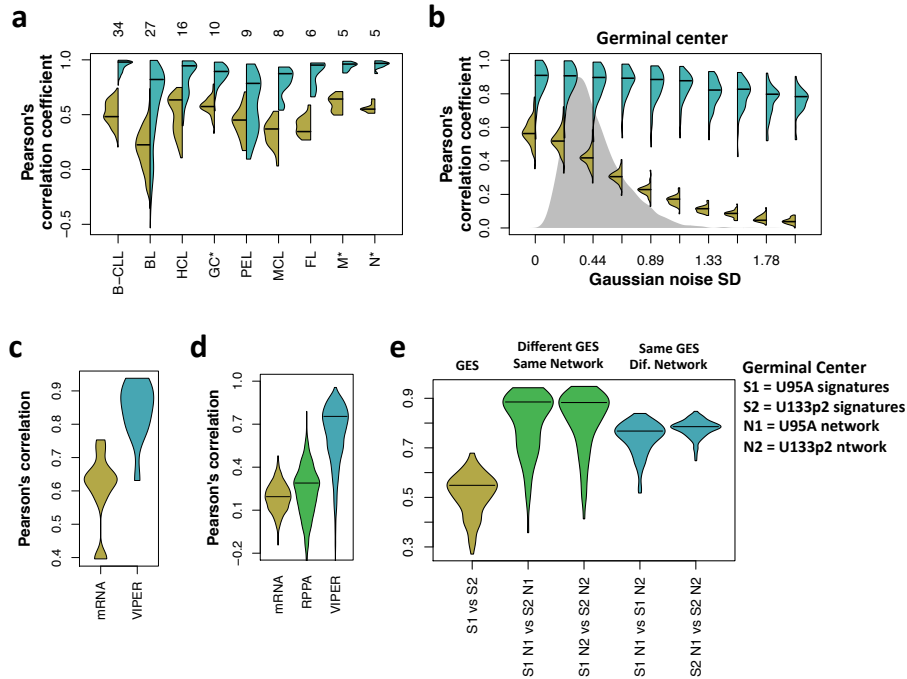
Supplementary Figure 7: Detecting changes in protein activity after genetic perturbations. Heatmaps showing the VIPER-inferred change in TF protein activity (a) and signaling protein activity (b) after coding gene expression knock-down. Displayed results correspond to silencing experiments in MCF7 breast carcinoma cells from LINCS showing a reduction in mRNA level of at least 2 standard deviations of the control samples. Statistical significance was estimated by Stouffer's integration of the single-sample NES (i.e. $\frac{1}{\sqrt{k}} \sum_{i=1}^k NES_i$). The vertical black line crossing the bar-plot indicates the significance threshold of $p = 0.05$. Bars showing a statistically significant change in protein activity at $p < 0.05$ are highlighted in blue (decreased protein activity, i.e. $NES < 0$) and red (increased protein activity, $NES > 0$). Values higher than the axis scale are indicated to the right of each bar.



Supplementary Figure 8: Detecting changes in protein activity after pharmacologic perturbations. Barplot showing the statistical significance for the change in protein activity inferred by VIPER after pharmacological perturbation with fulvestrant, tamoxifen and clomifene (targeting *ESR1*), and sirolimus (targeting *FKBP1A* and *MTOR*). The horizontal dotted line indicated the threshold at $p = 0.05$. A dose-dependent response can be observed for fulvestrant and clomifene (concentration indicated in μM units for each bar).



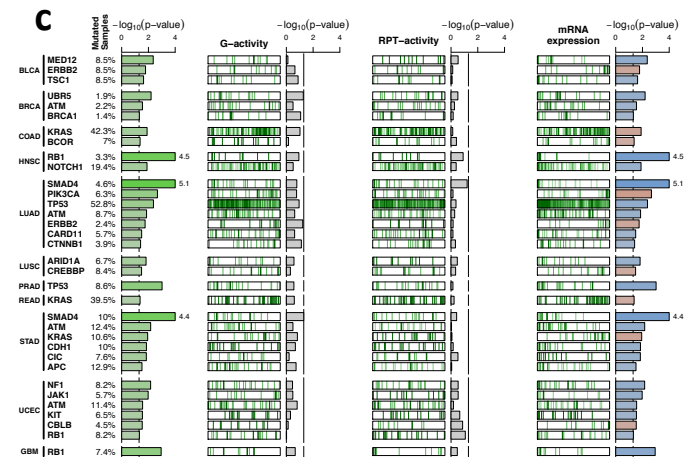
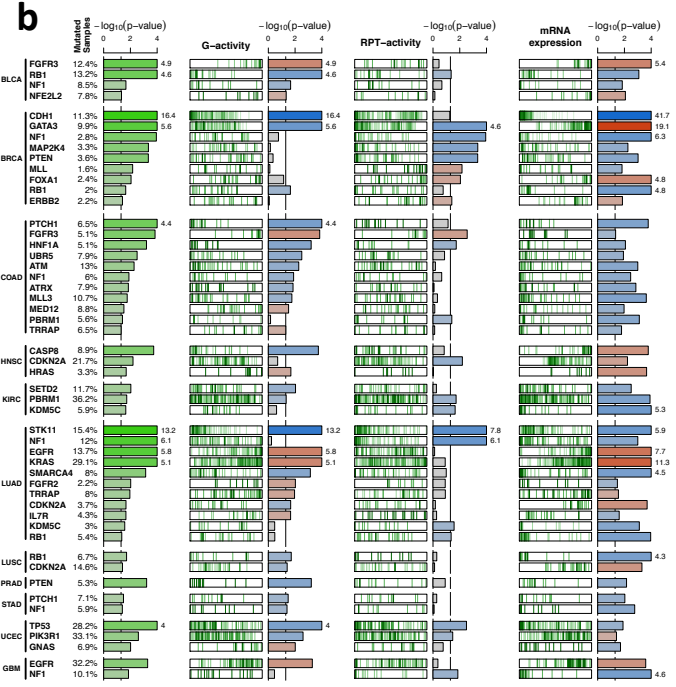
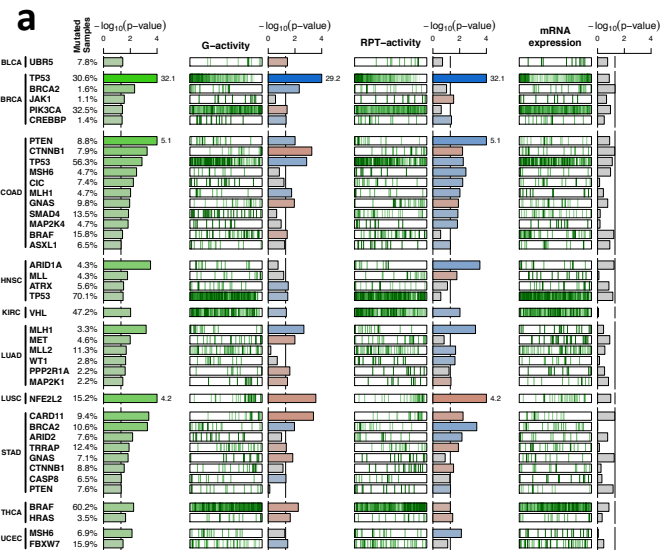
Supplementary Figure 9: Association between VIPER-inferred protein activity and protein isoform abundance. Bars show the significance level as $-\log_{10}(\text{p-value})$ for the Spearman's correlation between specific protein isoform abundance and either VIPER activity (green bars) or coding gene mRNA levels (red bars). VIPER activity shows the maximum association (correlation) between specific protein isoform abundance and either global (G-activity) or residual post-translational VIPER-inferred protein activity (RPT-activity).



Supplementary Figure 10: Reproducibility of single-sample gene expression, protein abundance and VIPER protein activity signatures. (a) Violin plot showing the distribution of correlation coefficients computed between gene expression signatures (yellow) or VIPER protein activity signatures (cyan) for samples of the same B cell phenotype, including normal (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. This analysis corresponds to the one shown in Figure 2g but limiting the expression signatures to the regulators represented in the VIPER analysis. The number of samples per phenotype is indicated on top of the figure. (b) Violin plot for the correlation between all possible pairs of GC B cell single-sample gene expression (yellow) and VIPER protein activity (cyan) signatures, after adding different levels of Gaussian noise to the expression profiles (indicated in the x-axis in standard deviation (SD) units). The grey probability density plot shows the distribution for the variance across samples in the original data. (c) Probability density for the correlation coefficient computed between FF and FFPE derived expression (yellow) and VIPER-inferred protein activity (cyan) signatures. (d) Violin plot showing the probability density for the correlation coefficients computed between all possible pairs of gene expression (yellow), RPPA protein abundance (green) and VIPER-inferred protein activity (cyan) signatures, corresponding to basal-subtype breast carcinoma tumors profiled by TCGA. (e) Correlation of germinal center B cell gene expression signatures between two datasets (yellow), the corresponding VIPER-inferred protein activity signatures between two datasets (green), or between two different B cell context specific networks (cyan). The horizontal line in the violin plots indicates the major mode of the distribution. See Table 2 for information about the datasets and networks.

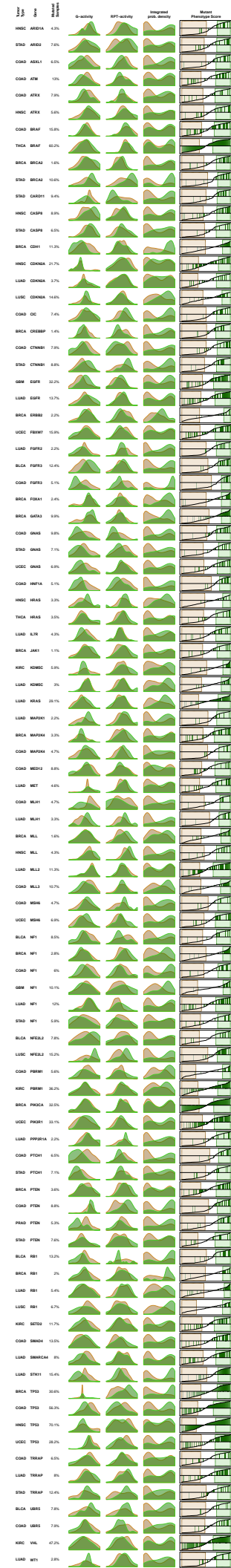
See figure in next page

Supplementary Figure 11: Detecting changes in protein activity induced by non-silent somatic mutations. Shown are all the genes listed in the Catalog Of Somatic Mutations in Cancer (COSMIC) where mutations are associated with: (a) protein activity but not mRNA expression, (b) inferred protein activity and mRNA expression, and (c) mRNA expression but not protein activity. The green bars indicate the integrated statistical significance for the effect of mutations on coding gene expression or protein activity. Each group of green enrichment plot and red/blue bar-plot indicates the enrichment of samples harborig NSSM on the VIPER-inferred global activity (G-activity) and residual post-translational activity (RPT-activity) for the coded protein, and differential gene expression, as indicated in the plot. The samples for each tumor type were rank sorted according to G-activity (left enrichment plot), RPT-activity (center enrichment plot) and gene expression (right enrichment plot), and the samples harboring NSSM were indicated by the green vertical lines. The significance level for the association is shown as $-\log_{10}(p)$ (barplot), with significant associations ($p < 0.05$) shown by red bars, for mutations associated with high activity or expression, and blue bars for mutations associated with low expression or activity. The value for genes associated at $p < 10^{-4}$ is shown beside the bars. Tumor type, gene name and proportion of mutated samples are indicated in the plot.



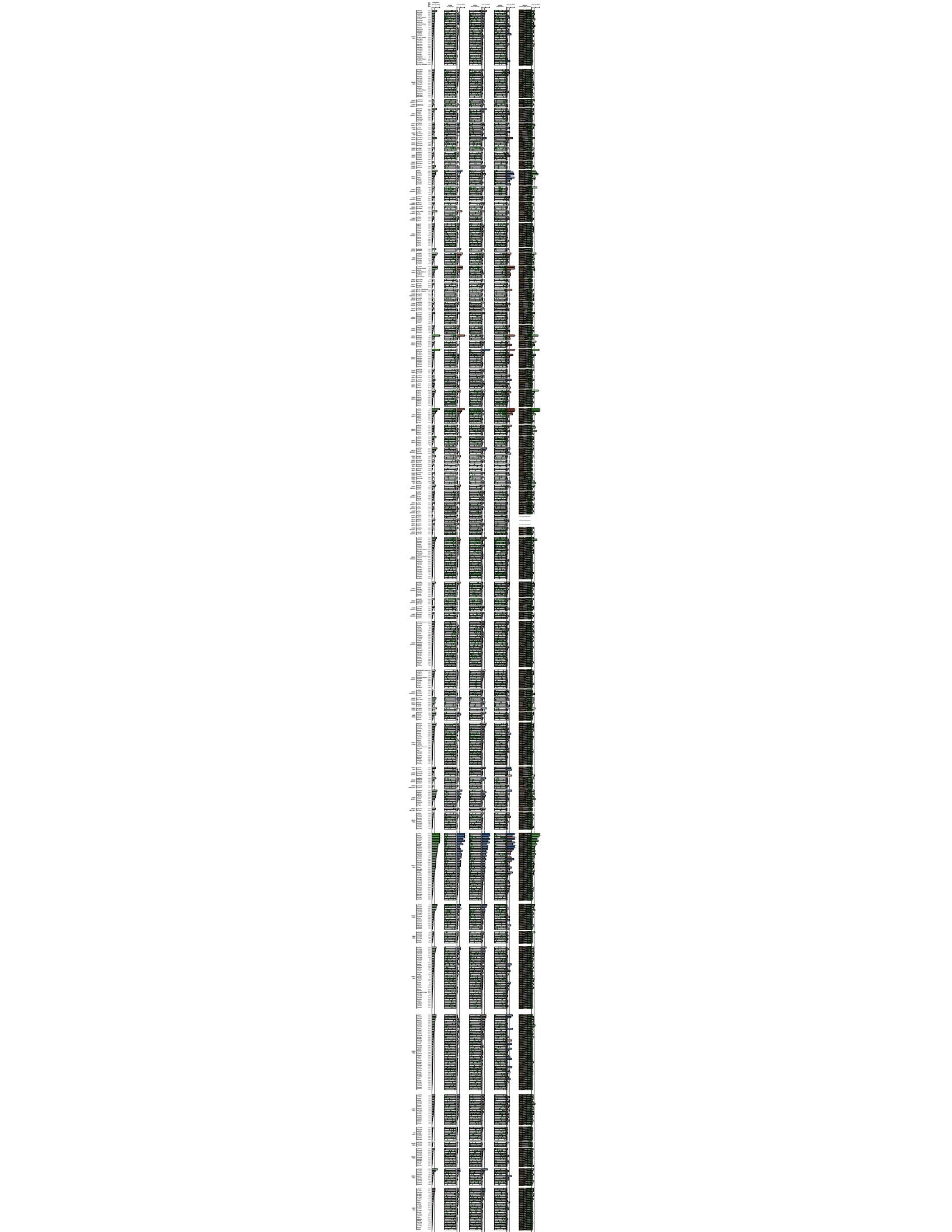
See figure in next page

Supplementary Figure 12: Mutant Phenotype Score (MPS). This figure lists all genes showing a significant association of mutations with either global activity (G-activity) or residual postranslational activity (RPT-activity). Each row shows the tumor type, gene and proportion of mutated samples, histograms for the probability density estimation of wt (salmon) and mutated (green) samples for each of the traits: VIPER-inferred G-activity and VIPER-inferred RPT-activity. The Integrated probability density histograms show the distributions of wt and mutated samples for the computed MPS. The rightmost plots show the MPS values (*y-axis*) for the samples rank-sorted by MPS (*x-axis*), with the mutated samples indicated by green vertical lines. The light salmon and green boxes highlight the MPS range corresponding to a likelihood ratio > 3 for wt and mutated phenotypes, respectively.



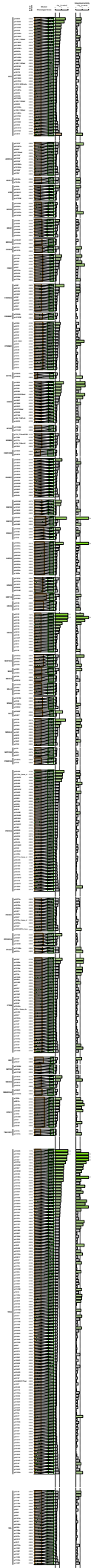
See figure in next page

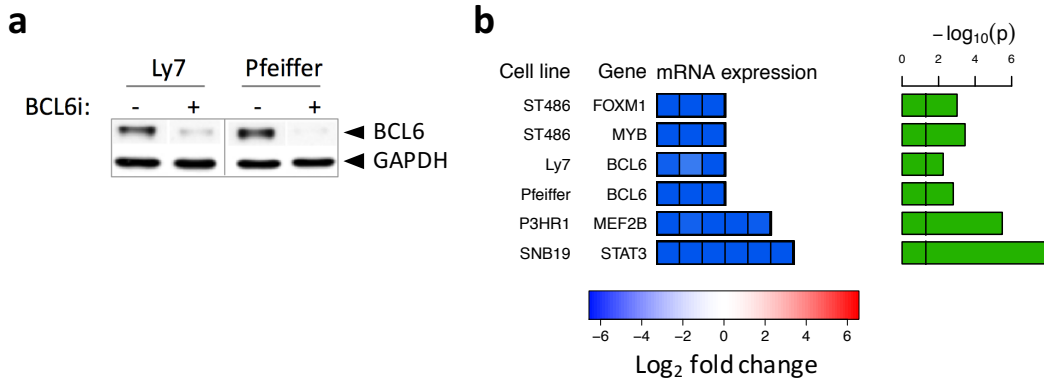
Supplementary Figure 13: Impact of specific NSSM variants on protein activity. Shown are all NSSM variants present in at least 2 samples in any of the 12 tumor types analyzed. The green barcode-like plots indicate the samples harboring each mutation when rank-sorted according to four quantitative traits: (1) VIPER-inferred G-activity (leftmost plot), (2) VIPER-inferred RPT-activity, (3) mutated gene mRNA expression levels, and (4) MPS (rightmost plot). The bars indicate the statistical significance, shown as $-\log_{10}(p)$, for the enrichment of the mutated samples on each of the four evaluated quantitative traits. The enrichment ‘side’ is indicated by the color of the bars, with over-expression or hyperactivity indicated by red bars, and under-expression or hypoactivity indicated by blue bars. The leftmost barplot, showing grey and green bars, indicates the statistical significance after integrating VIPER-inferred global activity (G-activity) and residual postranslational activity (RPT-activity). The rightmost barplot, showing grey, green and salmon bars, indicates the statistical significance for the enrichment of the mutated samples among the MPS-defined mutant phenotype (likelihood-ratio > 3 , indicated by the light-green box), or wt phenotype (likelihood-ratio > 3 , light-salmon box). The tumor type, gene name, mutation type and proportion of mutated samples are indicated in the plot. Missense mutations are indicated as p.*XnY* where *X* stands for the aminoacid in position *n* that was mutated to *Y*. Nonsense mutations are indicated by ‘*’ while frame shift mutations are indicated as p.*Xnfs*.



See figure in next page

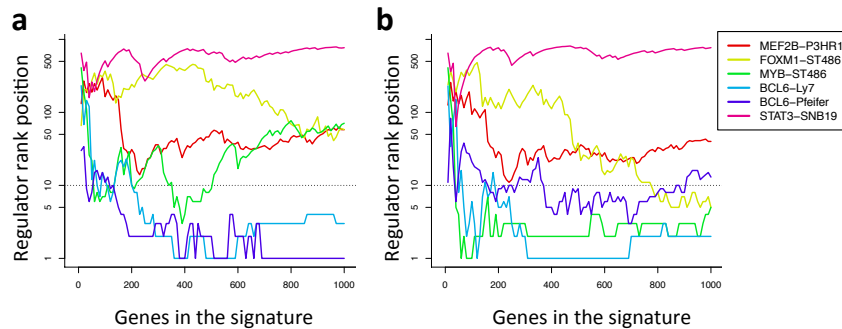
Supplementary Figure 14: Summary of the differential impact of NSSM variants on the coded protein activity. The leftmost plot shows the rank of mutated samples (vertical green lines) when all samples, across 12 tumor types, were rank sorted according to MPS. The center barplot shows the statistical significance, as $-\log_{10}(p)$, for the enrichment of the mutated samples among the MPS-defined mutant phenotype (likelihood-ratio > 3 , highlighted by a light-green box in the center plot), or wt phenotype (likelihood-ratio > 3 , light-salmon box). The rightmost barplot shows the association of each specific variant with either VIPER-inferred protein activity, conditional protein activity or mRNA levels, integrated across 12 tumor types. Bars indicate the statistical significance as $-\log_{10}(p)$ for each NSSM present in at least two samples. The gene name, mutation and proportion of samples harboring the mutation are shown in the plot.





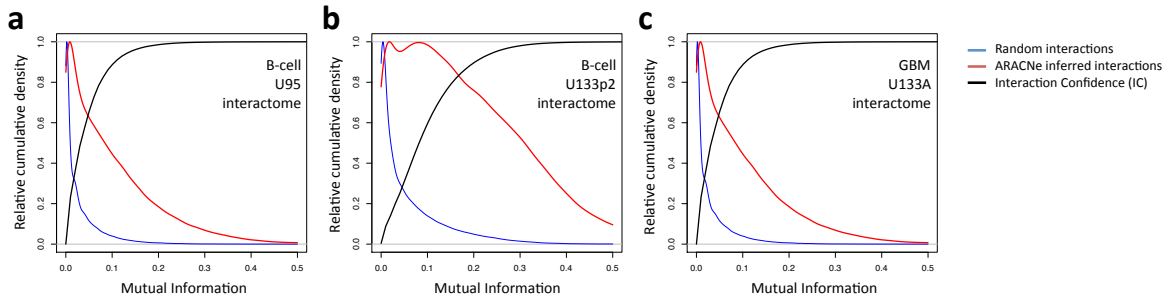
Supplementary Figure 15: Differential expression of the silenced genes.

(a) *BCL6* protein levels assessed by western-blot in OCI-Ly7 and Pfeiffer DLBCL cell lines 48h after *BCL6*-siRNA (+) or non-target control siRNA (-) transfection. *GAPDH* was used as loading control. Antibodies used were anti-*BCL6* N3 (Santa Cruz, sc-858), and *GAPDH* (sf-25778). (b) Heatmap showing the differential mRNA expression for each silenced TFs, and bar-plots indicating the statistical significance level. Shown are 4 experiments performed in B-cell lymphoma cell lines and 1 experiment in a glioma cell line (see Table 2). Statistical significance was estimated by Student's t-test. The vertical line crossing the bars indicate the significance threshold of $p = 0.05$.



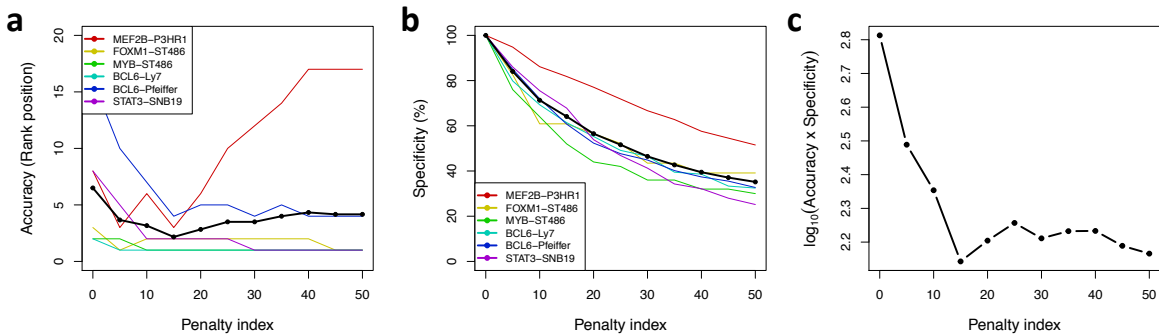
Supplementary Figure 16: The number of genes in the signature strongly impacts FET results.

Shown is the rank position in the differential activity signature (y -axis) inferred by 1-tail (a) and 2-tail FET (b) as a function of the number of genes considered as differentially expressed (x -axis).

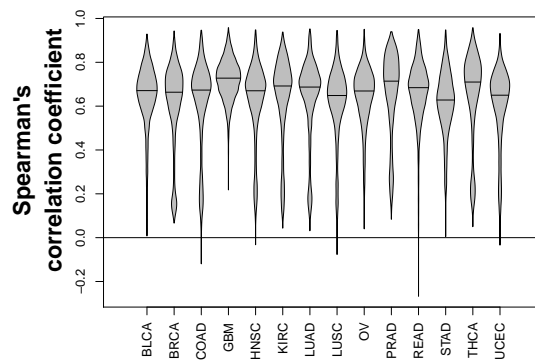


Supplementary Figure 17: Inferring the interaction confidence.

Interaction Confidence (y -axis) as a function of the interaction mutual information (x -axis) for the B-cell U95 (a), B-cell U133plus2 (b) and GBM U133A (c) interactomes. Shown are also kernel estimates for the distribution density of random interactions (blue line), and interactions inferred by ARACNe (red line). Both curves were scaled so their maximum value is one.



Supplementary Figure 18: Effect of the pleiotropy index parameter (PI) on VIPER results assessed with the benchmark data (see Table 2). (a) Accuracy of VIPER predictions, expressed as the rank position for the protein coded by the silenced gene, for varying values of PI. (b) Relative specificity expressed as the number of differentially active proteins inferred by VIPER for different values of PI, relative to the predictions obtained when no pleiotropy correction was applied (PI = 0). The results for different silencing experiments are shown by the color lines as indicated in panels (a) and (b). The black line shows the average across all experiments. (c) Integration of accuracy and specificity across all benchmark experiments.



Supplementary Figure 19: Correlation between VIPER-inferred protein activity and coding-gene mRNA level. Violin plot showing the density distribution of the Spearman's correlation coefficient for each tumor type.

References

- [1] Aytes, A. et al. (2014) Cross-Species Regulatory Network Analysis Identifies a Synergistic Interaction between FOXM1 and CENPF that Drives Prostate Cancer Malignancy. *Cancer Cell* 25, 638-51.
- [2] Anders,S. and Huber,W. (2010) Differential expression analysis for sequence count data. *Genome Biol.*, 11, R106.
- [3] Benaglia,T. et al. (2009) mixtools: An R Package for Analyzing Finite. *J. Stat. Softw.*, 32, 1-29.
- [4] Carro,M.S. et al. (2010) The transcriptional network for mesenchymal transformation of brain tumours. *Nature*, 463, 318-25.
- [5] Lachmann A, Xu H, Krishnan J, Berger SI, Mazloom AR, Maayan A. (2010) ChEA: Transcription Factor Regulation Inferred from Integrating Genome-Wide CHIP-X Experiments. *Bioinformatics*, 26, 2438-2444.
- [6] Lefebvre, C. et al. (2010) A human B-cell interactome identifies MYB and FOXM1 as master regulators of proliferation in germinal centers. *Mol. Syst. Biol.* 6, 377.
- [7] Lim,W.K. et al. (2009) Master regulators used as breast cancer metastasis classifier. *Pac. Symp. Biocomput.*, 504-15.
- [8] Subramanian,A. et al. (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.*, 102, 15545-50.
- [9] Lamb, J. et al. (2006) The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* 313, 1929-35.
- [10] Wishart, D.S. et al. (2008) DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res* 36, D901-6.