ANALYSIS

Genome-wide identification of post-translational modulators of transcription factor activity in human B cells

Kai Wang[1,2,5,6], Masumichi Saito[3,5,6], Brygida C Bisikirska[2], Mariano J Alvarez[2], Wei Keat Lim[1,2,5], Presha Rajbhandari[2], Qiong Shen[3], Ilya Nemenman[2,5], Katia Basso[3], Adam A Margolin[1,2,5], Ulf Klein[3], Riccardo Dalla-Favera[3,4] & Andrea Califano[1–3]

The ability of a transcription factor (TF) to regulate its targets is modulated by a variety of genetic and epigenetic mechanisms, resulting in highly context-dependent regulatory networks. However, high-throughput methods for the identification of proteins that affect TF activity are still largely unavailable. Here we introduce an algorithm, modulator inference by network dynamics (MINDy), for the genome-wide identification of post-translational modulators of TF activity within a specific cellular context. When used to dissect the regulation of MYC activity in human B lymphocytes, the approach inferred novel modulators of MYC function, which act by distinct mechanisms, including protein turnover, transcription complex formation and selective enzyme recruitment. MINDy is generally applicable to study the post-translational modulation of mammalian TFs in any cellular context. As such it can be used to dissect context-specific signaling pathways and combinatorial transcriptional regulation.

Reverse engineering of cellular networks in prokaryotes and lower eukaryotes[1–3], as well as in mammals[4–6], has started to unravel the remarkable complexity of transcriptional programs. These programs, however, may change substantially as a function of the availability of proteins affecting their post-translational modification of transcription factors, such as phosphorylation, acetylation and ubiquitination enzymes[7], as well as of those participating in transcription complexes (cofactors), thus making cellular networks highly context dependent.

Although the large-scale reprogramming of the cell's transcriptional logic has been studied in yeast[8,9], the identification of genes that affect these events remains elusive. Indeed, in contrast to methods such as those based on genome-wide chromatin immunoprecipitation (ChIP-chip or ChIP-Seq) or reverse engineering algorithms for the analysis of transcriptional networks[4,10], only one experimentally validated algorithm exists for the dissection of signaling networks in a mammalian context[11], which inferred substrates of 73 kinases. Here we propose and experimentally validate MINDy, a gene expression profile method for the systematic identification of genes that modulate the transcriptional program of a transcription factor at the post-translational level—that is, genes encoding proteins that affect the TF's activity without changing the abundance of its mRNA. These proteins may post-translationally modify the TF (for example, kinases), affect its cellular localization or turnover, be its cognate partners in transcriptional complexes or compete for its DNA binding sites. They may also include proteins that do not physically interact with the TF, such as those in its upstream signaling pathways.

RESULTS

The MINDy algorithm

MINDy interrogates a large gene expression profile dataset to identify 'candidate modulator' genes whose expression strongly correlates with changes in a TF's transcriptional activity. As shown in **Supplementary Note 1** (see also **Supplementary Figs. 1–3**), this can be efficiently accomplished by computing an information-theoretic measure known as the conditional mutual information, $I[TF;t \mid M]$[12], between the expression profile of a TF and one of its putative targets (t), given the expression of a modulator gene (M). Accurate estimation of the conditional mutual information requires exceedingly large datasets. Thus MINDy infers candidate modulators using a related yet simpler estimator, which we denote as ΔI. Briefly, the estimator assesses the statistical significance of the difference in mutual information between the TF and a target in two subsets—the top and bottom 35% of samples in which the modulator is most and least expressed. The 35% parameter was determined empirically as the one optimizing the identification of proteins in the B-cell receptor signaling pathway as modulators of MYC (Online Methods).

A schematic representation of the MINDy algorithm is provided in **Figure 1a**. MINDy takes four inputs: a gene expression profile dataset, a TF of interest, a list of potential modulator genes ($M_1, M_2, \ldots$) and a list of potential TF targets ($t_1, t_2, \ldots$). The ΔI estimator requires that the expression of the modulator and of the TF be statistically independent ('independence constraint') and that the modulator expression have sufficient range ('range constraint'). Appropriate statistical tests for these

constraints are discussed in the Online Methods. Candidate modulators may include all genes satisfying these constraints (unbiased analysis) or may be filtered by additional criteria—for example, their molecular functions. Each possible (*TF*, *M*, *t*) triplet is then independently tested using the ΔI estimator. False positives are controlled using appropriate statistical thresholds (Online Methods).

A positive or negative mode of action is determined, depending on whether the TF-target mutual information increases or decreases as a function of the modulator abundance (**Fig. 1a**). The mode of action, however, is not necessarily equivalent to the modulator's biological activity as a TF activator or antagonist. For instance, a modulator may be such a strong TF activator that the TF-target kinetics becomes saturated even

when the TF is slightly activated. In that case, TF-target mutual information will actually decrease as a function of the modulator. Details on the conditional mutual information analysis and on how to assess both the mode of action and the biological activity of a modulator are provided in the Online Methods.

For illustrative purposes, we show a simple synthetic network (**Fig. 1b**; Online Methods), which explicitly models two post-translational modulation events (activation by phosphorylation and cofactor binding) differentially affecting a TF's regulatory logic. Rather than representing a realistic case, this model is only a conceptual tool to illustrate two alternative regulatory programs of a TF depending on its modulators (**Fig. 1c**).
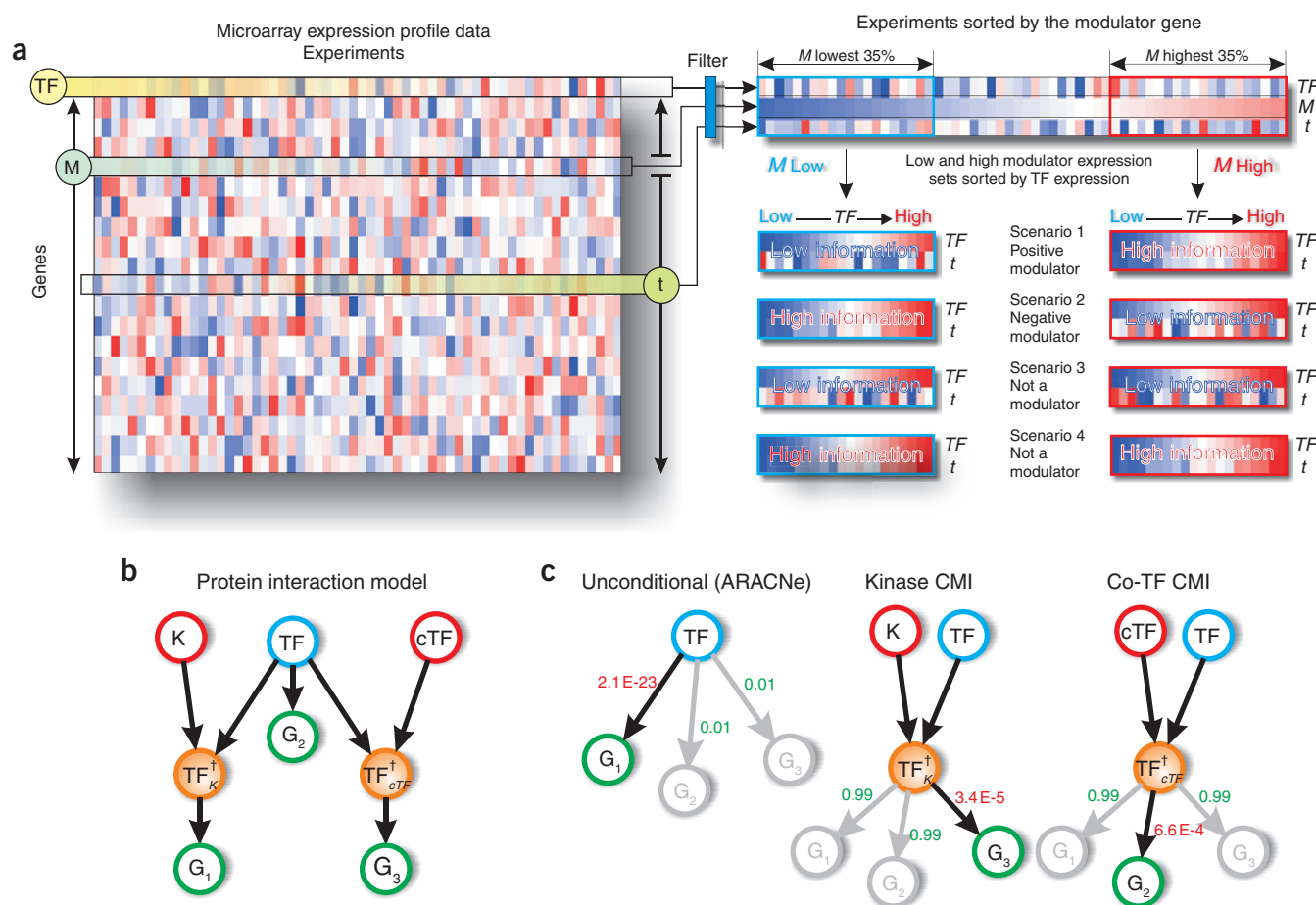
Figure 1 MINDy algorithm. (**a**) A gene expression profile dataset is represented as a matrix, where columns indicate different samples and rows indicate different genes. Given a transcription factor of interest (*TF*), a modulator (M) and a target (*t*) to be tested are chosen among the remaining genes. Some modulator-target combinations may be eliminated *a priori* based on functional or statistical constraints (blue rectangle). For instance, one may want to consider only ubiquitin ligases as candidate modulators. All the samples are then sorted according to the expression of the selected modulator *M*. The set of samples (for example, 35%) with the lowest and highest expression of the modulator are then selected. These sample sets are labeled *M* low and *M* high. In each of the two sample sets, samples are finally sorted according to the *TF* expression. Three cases are possible when comparing the TF-target correlation in *M* high versus *M* low. In Scenario 1 (positive inferred modulator), there is a significant increase in mutual information (that is, more correlation in the M-high set than in the M-low set). In Scenario 2 (negative inferred modulator), there is a significant decrease in mutual information (that is, less correlation in the M-high set than in the M-low set). In Scenarios 3 and 4 (not a modulator), no significant mutual information change is observed. (**b**) Synthetic network analysis using MINDy. TF, the transcription factor of interest; K, a protein kinase; cTF, a co-transcription factor; $TF_c^\dagger$, TF activated through phosphorylation by K; $TF_{cTF}^\dagger$, transcriptionally active complex formed between TF and cTF. G_1, G_2 and G_3 are three downstream targets of TF (Online Methods). (**c**) Networks reconstructed by ARACNe and MINDy. ARACNe (left) is based on pairwise mutual information and therefore captures only the unconditional interaction between TF and G_1. In contrast, MINDy infers three-way interactions using the conditional mutual information. When K is used as candidate modulator (middle), it correctly identified the conditional interaction between TF and G_3. Similarly, it also correctly inferred cTF as the modulator of TF-G_2 interaction (right). Interactions are labeled with their respective *P*-values. Interactions that are statistically significant, after correction for multiple testing, are shown in black with red *P* values. CMI, conditional mutual information.

MINDy-based identification of MYC modulators

We applied MINDy to the genome-wide identification of modulators of the MYC protein, using a previously assembled collection of 254 gene expression profiles[13,14] representing 17 distinct cellular phenotypes derived from normal and neoplastic human B lymphocytes (Online Methods). MYC is a basic helix-loop-helix/leucine zipper (bHLH/Zip) TF controlling many cellular processes, including cell growth, differentiation, apoptosis and DNA replication[15,16]. It is implicated in the pathogenesis of several human cancers[17] and can either activate or repress a large number of targets, depending on the cellular context (reviewed in ref. 18).

To study MINDy's robustness and generality, we tested its performance using different sets of candidate MYC targets. First, we used 340 literature-validated targets from the MYC database[19] (DB targets). Then, to also test whether the algorithm may generalize to TFs whose targets are not characterized in the literature, we used 197 MYC targets inferred by ARACNe[4], our previously described method for inferring the direct targets of a TF (AR targets). Finally, we considered all genes in the gene expression profile data as candidate targets (ALL targets), representing cases when literature or computationally inferred TF targets are not available (**Supplementary Note 2**).

Table 1 MYC modulators inferred by MINDy among signaling proteins and TFs

Signaling proteins					Transcription factors					
Modulator[a]	No. of MYC targets affected by modulator	MINDy-predicted mode of action[b]	Biological activity[c]	Evidence[d]	Modulator[a]	No. of MYC targets affected by modulator	MINDy-predicted mode of action[b]	Biological activity[c]	Evidence[d]	Binding-site enrichment analysis P-value[e]
CSNK2A1	178	↑	+	Direct[S18, S19]	SMAD3	159	↓	−	Direct[S20]	−
HCK	111	↓	−	Pathway[S19, S21]	AHR	156	↓	+	Direct[S22]	−
PPAP2B	109	↓	−		CREM	125	↓	+		1×10^{-36}
SAT	99	↓	−		DDIT3	96	↓	−		0.04
MAP4K4	83	↓	+	Pathway[S23, S24]	DRAP1	92	↑	−		−
DUSP2	82	↓	−	Pathway[S25, S26]	ZKSCAN1	87	↑	−		−
CSNK1D	80	↓	−		NR4A1	82	↓	−		−
PPM1A	80	↓	+		BHLHB2	80	↓	−	Validated	9×10^{-58}
GCAT	78	↑	+		ATF3	77	↓	+		1×10^{-17}
TRIO	74	↓	−		NR4A2	72	↓	−		−
STK38	60	↑	−	Validated	UBTF	71	↑	−		−
PRKCI	58	↑	+		NFKB2	66	↓	−		0.70 (ns)
CDKN1A	58	↓	−	Direct[S27]	HOXB7	65	↑	−		−
MTMR6	52	↓	+		BACH1	64	↓	−		5×10^{-9}
PRKACB	50	↓	+		SOX5	64	↓	+		0.66 (ns)
NEK9	47	↑	−		FOS	57	↓	−		2×10^{-3}
MYST1	45	↑	−		ARNT	55	↑	−	Direct[S22]	−
MAPK13	41	↑	−	Pathway[S28, S29]	IRF1	55	↓	−		6×10^{-8}
OXSR1	41	↓	+		ETV5	54	↓	+		−
DUSP4	40	↓	−		TCF12	51	↑	−		2×10^{-3}
MAP2K3	39	↓	−	Pathway[S30]	SMAD2	50	↑	−	Direct[S20]	−
FYN	38	↓	−	Pathway[S19, S31]	NFATC4	46	↑	−		−
PPP4R1	35	↓	+		E2F5	45	↑	−	Direct[S32]	−
MAPK1	35	↑	−	Direct[22, S33]	JUN	45	↓	−		1×10^{-7}
MAP4K1	34	↓	u	Pathway[S34]	CUTL1	39	↓	+		−
CSNK1E	33	↑	−		ESR2	38	↓	+		−
NEK7	32	↑	−		ZNF354A	37	↑	−		−
CSNK2A2	31	↑	+	Direct[S18]	MAF	37	↓	+		−
TRIB2	30	↓	+		SMARCB1	37	↑	−	Direct[S35]	−
HDAC1	8	↑	+	Validated	BRD8	34	↑	+		−
					DBP	34	↑	−		6×10^{-3}
					CBFA2T3	33	↑	−		−
					ESR1	31	↑	−		0.04
					RELB	31	↓	−		−
					TFEC	30	↓	u		0.92 (ns)
					MEF2B	14	↑	+	Validated	

[a]Shown are modulators that affect ≥ 30 MYC targets or that were experimentally validated in this work. [b]Mutual information increase (↑) or mutual information decrease (↓). [c]Activator (+), antagonist (−) or undetermined (u). [d]Direct, physically interacts with MYC protein or MYC binding sites; pathway, literature-validated role in pathway-mediated modulation of MYC activity (for example, B-cell receptor and mitogen-activated protein kinase pathways); validated, experimentally validated in this study. References S18–S35 are given in the **Supplementary References.** [e]ns, not statistically significant ($P > 0.05$).

MINDy identifies known MYC modulators

From a pool of 3,131 genes satisfying both independence and range constraints and using DB targets, MINDy inferred 662 MYC modulators (**Supplementary Table 1**) at a false-discovery rate (FDR) of 4.5 $\times 10^{-3}$ (Online Methods). Gene Ontology (GO) "molecular function" enrichment analysis revealed that the 20 most enriched categories, by Fisher's exact test (FET), include protein kinase activity ($P = 0.002$), TF binding ($P = 0.003$), acetyltransferase activity ($P = 0.004$) and phosphoprotein phosphatase activity ($P = 0.016$). Thus, inferred modulators were enriched in categories known to include effectors of MYC activity[20–22] (Online Methods and **Supplementary Table 2**).

To test whether MINDy could recapitulate literature-based MYC modulators, we assembled an unbiased set of 233 MYC modulators, including both proteins physically interacting with MYC and indirect modulators (**Supplementary Table 3**), using the Ingenuity software (Ingenuity Systems). The assumption is that physical interactors are likely to affect MYC protein function, although there may be false positives. Though not exhaustive, this provides an independently established literature-based dataset to assess algorithm performance (see **Supplementary Note 3** for details on inclusion criteria).

From this set, 150 genes were excluded as not represented on the chip, not expressed in B cells or not satisfying the range or independence constraints. Of the remaining 83, 29 (35%) were inferred as MYC modulators by the algorithm ($P = 0.0078$ by FET). This suggests that the algorithm is effective in recapitulating known MYC modulators (especially as Ingenuity modulators may not be B-cell specific) with recall comparable to that of high-throughput assays for protein-protein interactions, which on average detect 20% of known interactions[23]. We note that 54/83 proteins were reported by Ingenuity as MYC modulators not supported by a direct physical interaction. Of these, MINDy identified 18 (33.3%, $P = 0.041$ by FET), suggesting that MINDy is useful in identifying both physically interacting and indirect TF-modulators and
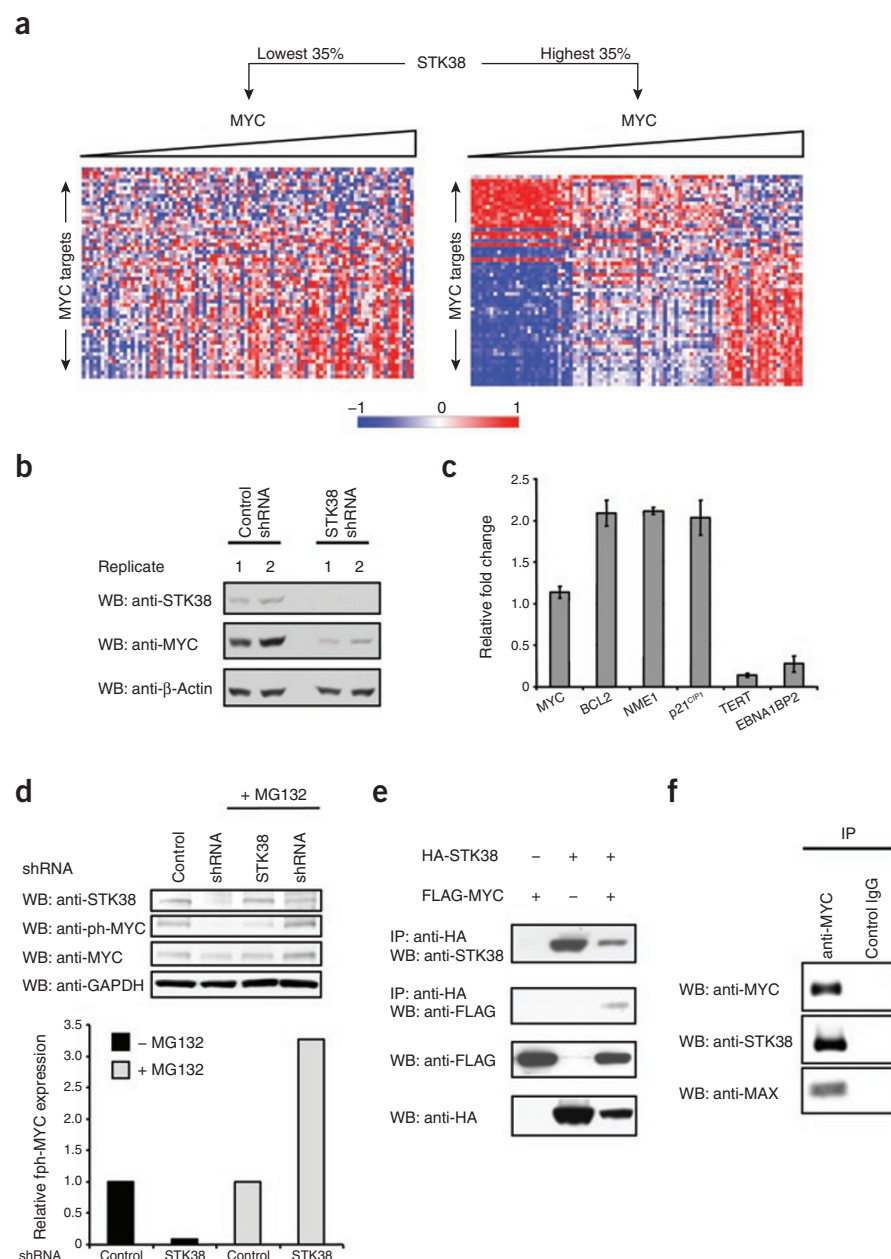
Figure 2 STK38 regulates the stability of MYC protein. (**a**) Visualization of MINDy output. Gene expression profiles are displayed with genes on rows and samples on columns. Expression values for each gene are rank transformed, median centered and rescaled between [−1, 1]. Samples were partitioned based on the expression levels of STK38 and, within each partition, sorted by the expression levels of MYC. (**b**) ST486 cells were infected with lentivirus expressing shRNA specific to STK38 and collected 96 h after infection. Whole cell extracts were analyzed by western blotting using anti-MYC, anti-STK38/NDR1 and anti-β-actin antibodies. Representative results from two of five independent experiments are shown. (**c**) qRT-PCR analysis of MYC and MYC target gene expression after silencing of STK38 in ST486 cells. Relative expression fold change between STK38 shRNA– and control shRNA– transduced cells was normalized to that of β-actin (*ACTB*) housekeeping gene. Bars represent the mean ± s.e.m. of three different samples. (**d**) STK38 mediates MYC phosphorylation. Top graph shows accumulation of phosphorylated MYC in STK38-silenced cells in the presence of proteasomal degradation inhibitor MG132 (48 h after infection). ST486 cells were infected with lentivirus expressing STK38 shRNA or control shRNA. The cells were treated with or without MG132 for 4 h before harvesting. Whole cell extracts were analyzed by western blotting using anti-STK38, anti-phospho-MYC (Thr58/Ser62), anti-MYC and anti-GAPDH antibodies. Bottom plot provides detailed densitometry histogram. Relative ph-MYC expressions (the ratio of ph-MYC to total MYC) are normalized to those of cells transduced with control shRNA. (**e**) HA-STK38 expression vector was transiently transfected into 293T cells with FLAG-MYC expression vector. At 24 h after transfection, immunoprecipitation was performed using the anti-HA agarose beads. Whole cell extracts and immunoprecipitated proteins were analyzed by western blotting with anti-FLAG and anti-HA antibodies. (**f**) Nuclear extracts from Ramos cells were immunoprecipitated with anti-MYC antibody or mouse IgG as a control. The precipitates were analyzed by western blotting with anti-STK38/NDR1, anti-MYC and anti-MAX antibodies.

that methods relying only on known molecular interactions would miss the vast majority of TF modulators. Indeed, based on Ingenuity almost twice as many modulators (18 versus 11) were found by MINDy among proteins not known to interact directly with MYC, compared to those having a physical protein-protein interaction.

To focus our analysis on specific molecular functions, we restricted candidate modulators to 542 signaling proteins—including protein kinases, phosphatases, acetyltransferases and deacetylases—and to 598 TFs (Online Methods). Among these, MINDy identified 91 signaling proteins (**Supplementary Table 4**) and 99 TFs (**Supplementary Table 5**), respectively, as MYC modulators (FDR = 0.0053). For each modulator, virtually all of the ΔI tests inferred the same mode of action (see columns "T+" and "T-" in **Supplementary Tables 4** and **5**) and fewer than 15% of the modulators had an ambiguous biological activity.

To assess a lower bound on the fraction of true positives among inferred modulators (that is, the precision), we performed manual literature curation. Because of the labor-intensive nature of this step, we considered only 29 signaling proteins and 35 TFs affecting more than 30 MYC targets. Among the former (**Table 1**), 11 appear as MYC modulators in published reports (precision = 37.9%). Similarly, among the latter (**Table 1**), 6 appear in published reports as MYC cofactors or antagonists (precision = 17.1%). For TFs with informative binding profiles in TRANSFAC[24], we tested whether their binding sites were enriched in promoters of the modulated targets (Online Methods). Fourteen of 35 TFs had appropriate binding profiles, and of these, 11 were highly enriched. Overall, 17 of 35 TFs (precision = 48.6%) either were literature-validated or had enriched binding sites (see **Table 1** and **Supplementary Table 6**). This suggests that MINDy's precision may approach that of experimental assays[23] when all modulators will be experimentally tested.

We then compared the performance of MINDy using literature-based targets (DB targets) to that with targets computationally inferred by ARACNe (AR targets). Overlap between the two target sets was highly significant ($P = 2.89 \times 10^{-18}$) but relatively small (17%). Nonetheless, overlap between MINDy-inferred modulators, when using the two target sets, was almost complete: 93.8% ($P = 3.10 \times 10^{-27}$) among signaling proteins and 95.3% ($P = 4.56 \times 10^{-288}$) among TFs, respectively (**Supplementary Note 2** and **Supplementary Tables 7** and **8**). This suggests that the method is highly robust with respect to target selection and can be effectively generalized to TFs whose targets are not known from the literature but can be inferred computationally.

Experimental validation

We selected four candidates among signaling genes and co-TFs for experimental validation, including a kinase (STK38), two TFs (BHLHB2 and MEF2B) and a deacetylating enzyme (HDAC1). These genes were selected based on the availability of reagents allowing the thorough validation of their activity and on the diversity of the possible mechanisms by which they may modulate MYC activity. As no single B-cell line expressed more than two of the four tested modulators, appropriate lines were selected for the silencing experiments among those with the highest modulator expression.

STK38 mediates MYC phosphorylation and protein stability

As a first candidate, we validated STK38, a serine/threonine kinase[25] inferred by MINDy as a strong positive modulator of MYC activity, affecting 60 targets (**Fig. 2a**, **Supplementary Table 4**). We silenced STK38 by lentiviral vector–mediated short hairpin RNA (shRNA) expression in ST486 cells. Although quantitative reverse transcription PCR (qRT-PCR) analysis showed that MYC mRNA concentration was unchanged after STK38 silencing, MYC protein levels were significantly affected

(**Fig. 2b**), suggestive of a post-translational modulation effect.

We proceeded to test two MYC targets in MINDy predictions, BCL2 and NME1, which are normally repressed by MYC[26,27]. Consistently with what MINDy predicted, both genes were upregulated after STK38
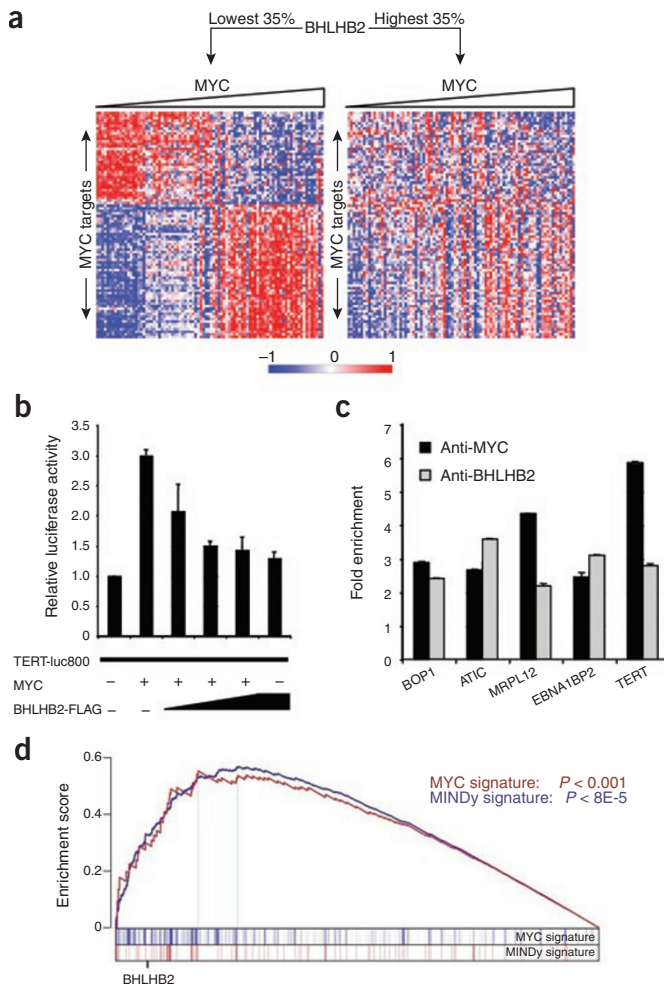
Figure 3 BHLHB2 antagonizes MYC activity in B cells. (**a**) Visualization of MINDy output. See **Figure 2a** for interpretation of this graph. (**b**) A TERT promoter–driven luciferase reporter construct, TERT-Luc800, was transiently co-transfected into 293T cells with MYC plasmid and three different doses of BHLHB2-FLAG plasmids. pRL-TK plasmid was used as internal control to monitor the transfection efficiency. The luciferase activities were measured 48 h after transfection and normalized against *Renilla* activity used as control (see Online Methods). Each experiment was done in duplicate, and data are shown as the mean ± s.e.m. of three independent experiments. (**c**) qChIP assays on ODH-III cells were performed in parallel using equivalent number of cells. Chromatin was immunoprecipitated with anti-MYC and anti-BHLHB2/DEC1 antibodies or with an irrelevant antibody (rabbit IgG) as a control. The precipitated DNA fragments were assessed by quantitative PCR, and data are shown as mean ± s.e.m. (**d**) GSEA enrichment analysis for BHLHB2 silencing. Gene expression profiles were generated for five samples of ODH-III cells transduced with BHLHB2-specific or control shRNA using a lentiviral vector. The x axis represents all probes on the microarray, ranked by their absolute differential expression in the cell transduced with the BHLHB2 shRNA versus the control shRNA. Most differentially expressed genes are toward the left. Sorting was based on the value of $-\log_{10}$ of their differential expression P values. Blue vertical bars represent all 340 genes in the MYC target signature, whereas the red bars represent the 80 MINDy-inferred, BHLHB2-modulated subset signature. Color intensity is proportional to the density of the bars. Rank of BHLHB2 among differentially expressed genes is shown by a tick mark.

silencing (**Fig. 2c**). Additionally, to test whether STK38-mediated MYC modulation is target specific, we tested three additional MYC targets not inferred by MINDy. The first two, TERT and EBNA1BP2, which are known to be activated by MYC[4,28], were downregulated following STK38

silencing, while the third one, p21[CIP1], which is known to be repressed by MYC[29], was upregulated (**Fig. 2c**). These results confirm the observed target-independent downregulation of MYC at the protein level.

Furthermore, STK38-mediated modulation of MYC affected its phosphorylation. Immunoblot analysis of MYC protein in ST486 cells, in the presence of an inhibitor of proteasomal degradation (MG132), showed accumulation of phosphorylated MYC in cells following STK38 silencing compared to cells treated with control shRNA, suggesting that STK38 mediates MYC phosphorylation (**Fig. 2d**).

Finally, coimmunoprecipitation (co-IP) of epitope-tagged STK38 (hemagglutinin (HA)-STK38) and MYC (FLAG-MYC), in 293T cells transfected with vectors expressing both proteins, showed that STK38 and MYC interacts at the protein level (**Fig. 2e**). Immunoprecipitation of endogenous MYC using specific antibodies in Ramos cells confirmed that the two proteins can interact physiologically in native cells (**Fig. 2f**). These results suggest that STK38 modulates MYC activity by directly affecting MYC protein stability.

BHLHB2 is a MYC antagonist

MINDy infers this TF as a negative modulator of MYC activity, affecting the regulation of 80 targets (**Fig. 3a**, **Supplementary Table 5**). Indeed, BHLHB2 is a TF able to bind to E-boxes through its bHLH domain, and it has been proposed to act as a transcriptional repressor[30], but it has not been validated as a MYC antagonist. Thus, we tested whether BHLHB2 could antagonize MYC-mediated transcriptional activation of its target genes by first investigating whether BHLHB2 could affect the transcriptional activation of TERT, a well-characterized MYC target[28].

Transient co-transfection in 293T cells of a reporter gene driven by a segment of the human TERT promoter region, carrying two E-boxes (TERT-Luc800[28]), and vectors encoding MYC or BHLHB2 showed that BHLHB2 represses MYC-mediated transcriptional activation on TERT in a dose-dependent manner. This effect is MYC dependent, as the basal transcriptional activity of the reporter gene is actually moderately increased by BHLHB2 (1.2-fold, **Fig. 3b**). Thus BHLHB2 represses MYC-mediated regulation but is not a direct repressor of the TERT promoter.

We next analyzed whether endogenous BHLHB2 molecules are physiologically bound to E-boxes within the promoter region of MYC target genes *in vivo* by quantitative chromatin immunoprecipitation (qChIP) assays in B-cells (ODH-III) using antibodies against MYC and BHLHB2. The results showed that the promoters of BOP1, ATIC, MRPL12, EBNA1BP2 and TERT were bound by both MYC and BHLHB2 (**Fig. 3c**).

To establish the functional significance of BHLHB2-mediated modulation of MYC transcriptional activity, we examined whether shRNA-mediated inhibition of BHLHB2 could affect the response of the 340 canonical MYC target genes used in the MINDy analysis or, more specifically, of the MYC target genes modulated by BHLHB2 as inferred by MINDy. The latter signature was used in case the effect was highly target specific and thus only a subset of MYC targets might be affected by BHLHB2 silencing. To this end, ODH-III cells were transduced with lentiviral vectors expressing BHLHB2-specific or control shRNAs. Western blot analysis showed that BHLHB2 was effectively silenced, while MYC levels were not affected (**Supplementary Fig. 4**).

We then performed gene expression profile analysis to assess the effect of BHLHB2 silencing on the expression of MYC targets. MYC is known to both positively and negatively regulate its targets[31]. Thus, without prior knowledge of MYC's specific activity on each target, we used gene set enrichment analysis (GSEA)[32] to assess whether MYC target signature genes are more differentially expressed than non–MYC target genes following BHLHB2 silencing (Online Methods). The analysis confirmed a
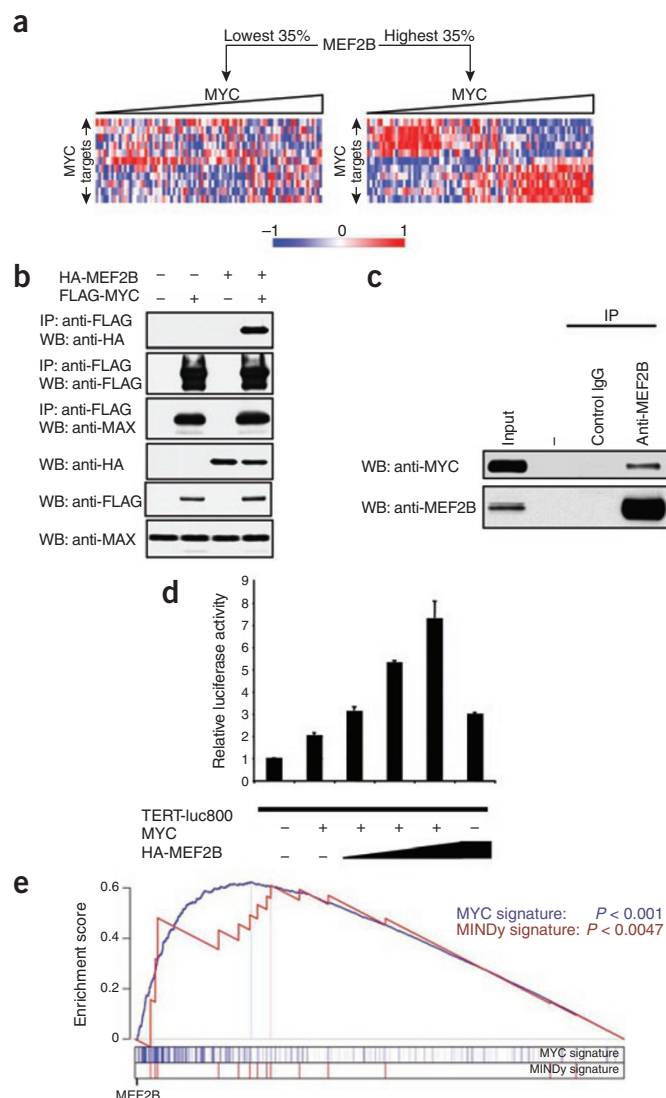
Figure 4 MEF2B enhances MYC activity via protein-protein interaction. (**a**) Visualization of MINDy output. See **Figure 2a** for interpretation of this graph. (**b**) HA-MEF2B expression vector was transiently transfected into 293T cells with FLAG-MYC expression vector. At 48 h after transfection, immunoprecipitation was performed using the anti-FLAG agarose beads (M2). Whole cell extracts and immunoprecipitation eluates were analyzed by western blotting with anti-FLAG, anti-HA and anti-MAX antibodies. (**c**) Nuclear extracts from P3HR1 cells were immunoprecipitated with anti-MEF2B serum or with rabbit serum as a control. The precipitates were analyzed by western blotting with anti-MEF2B and anti-MYC antibodies. (**d**) A TERT promoter–driven luciferase reporter construct, TERT-800Luc, was transiently co-transfected into 293T cells with MYC plasmid and three different amounts of HA-MEF2B plasmids. pRL-TK plasmid was used as internal control to monitor the transfection efficiency. The luciferase activities were measured 48 h after transfection and normalized against the *Renilla* activity. Each experiment was done in duplicate, and data are shown as the mean ± s.e.m. of three independent experiments. (**e**) GSEA enrichment analysis for MEF2B silencing. Gene expression profiles were generated for five samples of P3HR1 cells transduced with MEF2B-specific or control shRNA using a lentiviral vector. See **Figure 3d** for further interpretation of this graph.

highly significant enrichment of canonical MYC targets within the differentially expressed genes ($P < 0.001$) (**Fig. 3d**). Among the 80 MINDy inferred BHLHB2-modulated MYC targets, 30 were among the most differentially expressed genes. This constitutes approximately a twofold increase over their enrichment in non–differentially expressed genes ($P = 8 \times 10^{-5}$ by FET). MYC mRNA and protein levels were not affected (data not shown and **Supplementary Fig. 4**), indicating a post-translational effect. These results validate BHLHB2 as an antagonist of MYC activity.

MEF2B is a positive modulator of *MYC* activity

This TF was inferred as a positive modulator of MYC activity, affecting 14 MYC targets (**Fig. 4a**, **Supplementary Table 5**). MEF2B is a member of the MEF TF family, which interacts with the myogenic bHLH proteins MyoD and E12 to activate gene transcription through direct binding to E-boxes on target promoters[33]. Details of the validation assays are provided in **Supplementary Note 4**.

Briefly, similar to our observations with BHLHB2, we showed that (i) MEF2B physically interacts with MYC both exogenously in 293T cells (**Fig. 4b**) and endogenously in P3HR1 and Ramos cells (**Fig. 4c**); (ii) MYC and MEF2B can synergistically activate a TERT reporter gene (**Fig. 4d**); and (iii) genes differentially expressed following shRNA-mediated silencing of MEF2B were highly enriched in MYC targets by GSEA ($P < 0.001$, **Fig. 4e**), whereas MYC expression was not affected (**Supplementary Fig. 5b**).

## HDAC1 may deacetylate MYC and repress MYC targets

Finally, MINDy identified the histone deacetylase and well-known transcriptional co-repressor HDAC1 (refs. 34,35) as a modulator of MYC transcriptional activity on eight MYC targets (**Fig. 5a**, **Supplementary Table 4**). Experiments (see **Supplementary Note 5** for details) confirmed that (i) HDAC1 and MYC can interact *in vivo*, both exogenously in 293T cells (**Fig. 5b**) as also reported in ref. 36 and endogenously in Ramos and P3HR1 cells (**Fig. 5c**); (ii) genes differentially expressed following shRNA-mediated silencing of HDAC1 were highly enriched in MYC targets by GSEA ($P < 0.001$, **Fig. 5d**), whereas MYC expression was not affected (**Supplementary Fig. 6a**); (iii) HDAC1 can deacetylate MYC *in vitro*, following its CBP-mediated acetylation (**Fig. 5e**), which has been shown to increase MYC's activity as a transcriptional activator; and (iv) as indicated by qChIP assays with anti-MYC and anti-HDAC1 specific antibodies, both MYC and HDAC1 bind to the promoters of p21^CIP1 and CR2, which are repressed by MYC in B-cells (**Fig. 5f**).

These results suggest that MYC may recruit HDAC1 to repress transcription of specific target genes. Taken together, our data demonstrate

that HDAC1 may modulate MYC activity both by deacetylation of the MYC protein and by transcriptional repression of selected targets.

Extension to other *TFs*

To validate MINDy on a broader range of TFs, we used the algorithm to infer all TFs modulated by BHLHB2, MEF2B, HDAC1 and STK38, for
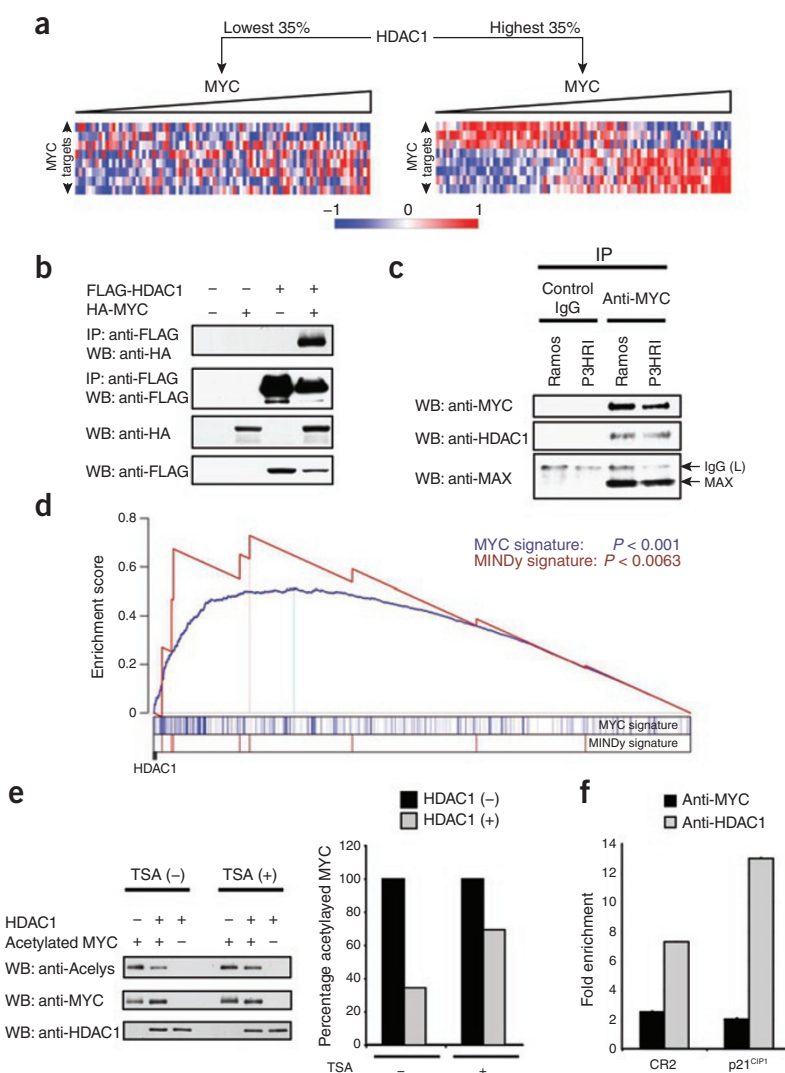
Figure 5 MYC selectively recruits HDAC1 to its targets as co-repressor. (**a**) Visualization of MINDy output. See **Figure 2a** for interpretation of this graph. (**b**) FLAG-HDAC1 expression vector was transiently transfected into 293T cells with HA-MYC expression vector. At 48 h after transfection, immunoprecipitation was performed using the anti-FLAG agarose beads (M2). Whole cell extracts and immunoprecipitation eluates were analyzed by western blotting with anti-FLAG and anti-HA antibodies. (**c**) Nuclear extracts from Ramos and P3HR1 cells were immunoprecipitated with anti-MYC antibody or rabbit IgG as a control. The precipitates were analyzed by western blotting with anti-HDAC1, anti-MYC and anti-MAX antibodies. (**d**) GSEA enrichment analysis for HDAC1 silencing. Gene expression profiles were generated for five samples of P3HR1 cells transduced with HDAC1-specific or control shRNA using a lentiviral vector. See **Figure 3d** for further interpretation this graph. (**e**) FLAG-MYC and CBP-HA expression vectors were transiently transfected into 293T cells to purify acetylated-MYC protein. HDAC1, which was also purified from transient transfected 293T cells, was incubated with acetylated-MYC protein *in vitro*. The signal intensities of acetylated-MYC and total amount of MYC were measured with the imageQuant 5.2 software. The densitometry histogram on the right shows the percent of acetylated-MYC normalized by total amount of MYC. (**f**) qChIP assays on P3HR1 cells were performed in parallel using equivalent number of cells. Chromatin was immunoprecipitated with anti-MYC and anti-HDAC1 antibodies or an irrelevant antibody (rabbit IgG) as a control. The precipitated DNA fragments were assessed by quantitative PCR with CR2 and p21^CIP1 primers and are shown as mean ± s.e.m.

which we had already collected gene expression profile data following shRNA-mediated silencing (**Supplementary Note 6**). Specifically, we tested (by GSEA) whether their MINDy-inferred targets were enriched in differentially expressed genes following lentivirus-mediated shRNA silencing of the corresponding modulators. Seventy-five percent of the TFs inferred by MINDy as modulated by any of the four modulators with support from more than 100 targets, could be experimentally validated by the analysis (33 out of 44, $P \leq 5.1 \times 10^{-34}$) (**Supplementary Table 9**). Furthermore, as one may expect, validation rates increased—from 51% (87 out of 171) to 61% (59 out of 96) to 75% (33 out of 44)—when the minimum number of MINDy-inferred targets supporting the modulator was increased from 25 to 50 to 100, respectively. In general, these results are consistent with the MYC analysis and suggest that MINDy is broadly applicable to the analysis of TFs other than MYC.

DISCUSSION

We have introduced MINDy, a new method for the identification of context-specific, post-translational modulators of TF activity. Literature-based and experimental validation suggests that MINDy can recapitulate a large fraction of known MYC modulators and infer novel, context-specific modulators of both MYC and other TFs.

For well-studied TFs, targets for the analysis may be selected from the literature or by performing genome-wide ChIP assays[37,38]. However, computationally inferred targets performed as well as or better than literature-based ones, likely due to their context-specific nature. About 269 TFs have more than 50 ARACNe-inferred targets using the B-cell profiles and may thus be effectively analyzed by MINDy.

Algorithm performance was remarkably robust to candidate target selection (DB targets versus AR targets). Additionally, the input data required by MINDy is relatively straightforward, requiring only the availability of a large gene expression profile dataset ($n \geq 200$ profiles) characterizing a sufficient variety of naturally occurring or experimentally perturbed cellular phenotypes. This suggests that MINDy can be used to analyze most TFs in a variety of cellular contexts.

Several limitations should be noted. First, candidate modulators that do not satisfy the range constraint cannot be tested by the method. These, however, primarily include either genes that are not expressed or genes whose availability is so tightly regulated (for example, housekeeping genes) that variability in the gene expression profiles is too limited to establish a low and a high range of expression. Second, candidate modulators that do not satisfy the independence constraint may not be tested using this approach. In practice, fewer than half (100/233 = 42.9%) of the Ingenuity modulators were in this category. This is not a theoretical limitation of the method but rather an assumption we use to increase its sensitivity; thus, if desired, a more general test may be used without relying on the assumption $I[TF; M] = 0$ (**Supplementary Note 1**). Additionally, transcriptional modulators of MYC can be directly inferred by ARACNe and do not require MINDy. Third, in the rare event that the regulatory program of a TF changes from activation to repression for specific targets, as a function of a modulator, this may not be detected by the algorithm because the mutual information may not change substantially. In this case the multi-information could be used instead of the conditional (**Supplementary Note 1**).

MINDy is able to discover large numbers of modulators of the same TF. In contrast, finding the optimal Bayesian network structure, assuming arbitrary interactions among TF modulators (that is, the TF parents in the network topology), would have time and memory requirements that are hyperexponential in the number of modulators. As a result, dissecting network topologies with large numbers (tens to hundreds) of upstream modulators, such as those of MYC, may be difficult. Other methods[39], although promising in a yeast context, have

not yet been extended to mammalian networks. Finally, comparison with a recently introduced algorithm, NetworKIN[11], shows that the latter is restricted to substrates of only 73 kinases, from 20 families, whereas MINDy can be used to dissect post-translational interactions of a much wider nature, including phosphorylation, acetylation, chromatin modification, formation of transcription complexes and binding-site antagonism, as we have demonstrated here for STK38, HDAC1, MEF2B and BHLHB2.

The ability to infer direct and upstream modulators of a desired TF's activity suggests that MINDy may provide highly specific pharmacological targets for the activation or repression of specific transcriptional programs, when modulators are restricted to druggable genes[40]. This could be valuable because TFs are generally considered difficult pharmacological targets.

Although preliminarily applied to identifying modulators of MYC for experimental validation purposes, MINDy has already provided biological insights. First, the results indicate that not all modulators can influence the program of a TF in a global fashion; they may rather influence specific subsets of the target genes. This observation suggests that additional levels of regulation can influence the relationships between modulators and the TFs they control in different cellular contexts or depending on different signals. Second, MINDy identified novel molecules that regulate the activity of the MYC protein. These mechanisms may be critically altered in tumors, thus modulating MYC's established oncogenic activity. Finally, MINDy is not limited to dissecting post-translational interactions and may be applied without modification to identify TFs that are directly modulated by microRNAs or indirectly by genetic and epigenetic alterations.

Algorithm availability. At manuscript publication, the source code and executables for MINDy will be made available under the Open Source licensing agreement. Additionally, MINDy will be incorporated into the geWorkbench package, which is distributed both by the US National Cancer Institute and by the National Center for Biomedical Computing at Columbia University (http://wiki.c2b2.columbia.edu/workbench/index.php/Home).

Note: Supplementary information is available on the Nature Biotechnology website.

METHODS
Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturebiotechnology/.

AUTHOR CONTRIBUTIONS
K.W. and A.C. designed the algorithm and the computational analysis procedures. I.N. and A.A.M. assisted in algorithm design and computational analysis. M.S. experimentally validated BHLHB2, MEF2B and HDAC1. B.C.B. experimentally validated STK38. M.J.A. and W.K.L. performed the analysis and validation of TFs other than MYC. I.N. and K.W. studied the information theoretic properties of the algorithm. P.R., Q.S., K.B. and U.K. assisted with the experimental validation and microarray expression profile generation. A.C., R.D.-F., M.S., BC.B., K.W. and K.B. designed the experimental assays. A.C., R.D.-F., K.W. and M.S. wrote the manuscript.

COMPETING INTERESTS STATEMENT

1. Faith, J.J. *et al.* Large-scale mapping and validation of *Escherichia coli* transcriptional regulation from a compendium of expression profiles. *PLoS Biol.* **5**, e8 (2007).
2. Friedman, N. Inferring cellular networks using probabilistic graphical models. *Science* **303**, 799–805 (2004).
3. Gardner, T.S., di Bernardo, D., Lorenz, D. & Collins, J.J. Inferring genetic networks and identifying compound mode of action via expression profiling. *Science* **301**, 102–105 (2003).
4. Basso, K. *et al.* Reverse engineering of regulatory networks in human B cells. *Nat. Genet.* **37**, 382–390 (2005).
5. Elkon, R., Linhart, C., Sharan, R., Shamir, R. & Shiloh, Y. Genome-wide in silico identification of transcriptional regulators controlling the cell cycle in human cells. *Genome Res.* **13**, 773–780 (2003).
6. Stuart, J.M., Segal, E., Koller, D. & Kim, S.K.A. Gene-coexpression network for global discovery of conserved genetic modules. *Science* **302**, 249–255 (2003).
7. Zeitlinger, J. *et al.* Program-specific distribution of a transcription factor dependent on partner transcription factor and MAPK signaling. *Cell* **113**, 395–404 (2003).
8. Luscombe, N.M. *et al.* Genomic analysis of regulatory network dynamics reveals large topological changes. *Nature* **431**, 308–312 (2004).
9. Segal, E. *et al.* Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nat. Genet.* **34**, 166–176 (2003).
10. Ren, B. *et al.* Genome-wide location and function of DNA binding proteins. *Science* **290**, 2306–2309 (2000).
11. Linding, R. *et al.* Systematic discovery of in vivo phosphorylation networks. *Cell* **129**, 1415–1426 (2007).
12. Cover, T.M. & Thomas, J. *Elements of Information Theory* edn. 2 (Wiley Interscience, Hoboken, New Jersey, USA, 2006).
13. Basso, K. *et al.* Reverse engineering of regulatory networks in human B cells. *Nat. Genet.* **37**, 382–390 (2005).
14. Mani, K.M. *et al.* A systems biology approach to prediction of oncogenes and molecular perturbation targets in B-cell lymphomas. *Mol. Syst. Biol.* **4**, 169 (2008).
15. Dang, C.V. *et al.* The c-Myc target gene network. *Semin. Cancer Biol.* **16**, 253–264 (2006).
16. Dominguez-Sola, D. *et al.* Non-transcriptional control of DNA replication by c-Myc. *Nature* **448**, 445–451 (2007).
17. Pelengaris, S., Khan, M. & Evan, G. c-MYC: more than just a matter of life and death. *Nat. Rev. Cancer* **2**, 764–776 (2002).
18. Zeller, K.I., Jegga, A.G., Aronow, B.J., O'Donnell, K.A. & Dang, C.V. An integrated database of genes responsive to the Myc oncogenic transcription factor: identification of direct genomic targets. *Genome Biol.* **4**, R69 (2003).
19. Zeller, K.I., Jegga, A., Aronow, B., O'Donnell, K. & Dang, C. An integrated database of genes responsive to the Myc oncogenic transcription factor: identification of direct genomic targets. *Genome Biol.* **4**, R69 (2003).
20. Levens, D.L. Reconstructing MYC. *Genes Dev.* **17**, 1071–1077 (2003).
21. Patel, J.H. *et al.* The c-MYC oncoprotein is a substrate of the acetyltransferases hGCN5/PCAF and TIP60. *Mol. Cell. Biol.* **24**, 10826–10834 (2004).
22. Sears, R. *et al.* Multiple Ras-dependent phosphorylation pathways regulate Myc protein stability. *Genes Dev.* **14**, 2501–2514 (2000).
23. Yu, H. *et al.* High-quality binary protein interaction map of the yeast interactome network. *Science* **322**, 104–110 (2008).
24. Matys, V. *et al.* TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res.* **31**, 374–378 (2003).
25. Tamaskovic, R., Bichsel, S.J. & Hemmings, B.A. NDR family of AGC kinases–essential regulators of the cell cycle and morphogenesis. *FEBS Lett.* **546**, 73–80 (2003).
26. Patel, J.H. & McMahon, S.B. BCL2 is a downstream effector of MIZ-1 essential for blocking c-MYC-induced apoptosis. *J. Biol. Chem.* **282**, 5–13 (2007).
27. Schuhmacher, M. *et al.* The transcriptional program of a human B cell line in response to Myc. *Nucleic Acids Res.* **29**, 397–406 (2001).
28. Wu, K.J. *et al.* Direct activation of TERT transcription by c-MYC. *Nat. Genet.* **21**, 220–224 (1999).
29. Seoane, J., Le, H.V. & Massague, J. Myc suppression of the p21^{Cip1} Cdk inhibitor influences the outcome of the p53 response to DNA damage. *Nature* **419**, 729–734 (2002).
30. St-Pierre, B., Flock, G., Zacksenhaus, E. & Egan, S.E. Stra13 homodimers repress transcription through class B E-box elements. *J. Biol. Chem.* **277**, 46544–46551 (2002).
31. Wanzel, M., Herold, S. & Eilers, M. Transcriptional repression by Myc. *Trends Cell Biol.* **13**, 146–150 (2003).
32. 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).
33. Molkentin, J.D. *et al.* MEF2B is a potent transactivator expressed in early myogenic lineages. *Mol. Cell. Biol.* **16**, 3814–3824 (1996).
34. Wade, P.A. Transcriptional control at regulatory checkpoints by histone deacetylases: molecular connections between cancer and chromatin. *Hum. Mol. Genet.* **10**, 693–698 (2001).
35. Satou, A., Taira, T., Iguchi-Ariga, S.M.M. & Ariga, H. A novel transrepression pathway of c-Myc. Recruitment of a transcriptional corepressor complex to c-Myc by MM-1, a c-Myc-binding protein. *J. Biol. Chem.* **276**, 46562–46567 (2001).
36. Matsuoka, Y., Fukamachi, K., Uehara, N., Tsuda, H. & Tsubura, A. Induction of a novel histone deacetylase 1/c-Myc/Mnt/Max complex formation is implicated in parity-induced refractoriness to mammary carcinogenesis. *Cancer Sci.* **99**, 309–315 (2008).
37. Margolin, A.A. *et al.* ChIP-on-chip significance analysis reveals large-scale binding and regulation by human transcription factor oncogenes. *Proc. Natl. Acad. Sci. USA* **106**, 244–249 (2009).
38. Robertson, G. *et al.* Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat. Methods* **4**, 651–657 (2007).
39. Steffen, M., Petti, A. & Aach, J., D'haeseleer, P. & Church, G. Automated modelling of signal transduction networks. *BMC Bioinformatics* **3**, 34 (2002).
40. Hopkins, A.L. & Groom, C.R. The druggable genome. *Nat. Rev. Drug Discov.* **1**, 727–730 (2002).

ONLINE METHODS

Gene expression profile dataset. We used 254 gene expression profiles previously generated by our labs for several studies of normal and tumor-related B-cell phenotypes using the Affymetrix HG-U95Av2 GeneChip System (approximately 12,600 probes)[13,14]. The gene expression profiles are available in GEO (series GSE2350, including samples GSM44075-44095, GSM44113-44302 and GSM65674-65716). Probe sets with expression mean $\mu < 50$ and s.d. $\sigma < 0.3\mu$ were considered uninformative and were therefore excluded, leaving 7,907 probe sets for the analysis. **Supplementary Table 10** summarizes the 17 B-cell phenotypes included in this study.

Synthetic network. The synthetic network models two post-transcriptional modifications of a TF, affecting its regulatory behavior (**Fig. 1b**). It includes the transcription factor (TF), an activating protein kinase (K), a cofactor (cTF) forming a transcriptionally active complex with the TF and three downstream TF targets. The full kinetic model is described in **Supplementary Table 11**. 250 synthetic expression profiles were generated from this model by (i) randomly sampling the TF, K and cTF mRNA concentration from independent normal distributions ($\mu = 4$ and $\sigma = 1$), (ii) simulating network dynamics until a steady state was reached and (iii) measuring the mRNA concentration of the represented species with a multiplicative Gaussian noise ($\mu = 0$ and $\sigma = 0.1\mu$).

Candidate modulators. The statistical test for the 'range constraint' is defined in **Supplementary Note 7**. The statistical significance test for the independence constraint is based on the mutual information, as described in ref. 41 (see also **Supplementary Notes 8** and **9**).

For the category-specific analysis, we further selected 542 signaling proteins (GO molecular function: "protein kinase activity," "phosphoprotein phosphatase activity," "acetyltransferase activity" and "deacetylase activity") and 598 TFs (GO molecular category: "transcription factor activity") as candidate modulators.

MINDy inference. Given a triplet (TF, M, t), with $t \neq TF$ and $t \neq M$, MINDy assesses whether the conditional mutual information, $I[TF;t \mid M]$, is constant as a function of M. Assuming that the conditional mutual information is a monotonic function of M (see **Supplementary Note 10**), this can be efficiently tested by measuring $\Delta I = I[TF;t \mid M \in L_m^+] - I[TF;t \mid M \in L_m^-] \neq 0$, where L_m^+ and L_m^- represent two subsets of the samples where M is respectively most and least expressed. For this dataset $L_m^\pm$ is chosen to be 35% of the samples by optimizing the effect of B-cell receptor pathway genes (which is known to modulate MYC activity[42]) on regulating canonical MYC target genes (**Supplementary Note 11** and **Supplementary Fig. 7**). Mutual information was computed using the Gaussian kernel estimator of ref. 41 (**Supplementary Note 8**). The P value corresponding to a specific ΔI is obtained by permutation tests (**Supplementary Note 1** and **Supplementary Fig. 3**), Bonferroni-corrected for the total number of tested target-modulator pairs. If a candidate target list is not provided, triplets are further pruned if there exists a third gene, x, such that $I[TF;x] \geq [TF;t]$ and $I[t;x] \geq [TF;t]$ in both $L_m^\pm$, showing an indirect relationship between TF and t mediated by x, as suggested by the data-processing inequality[41] (**Supplementary Note 12** and **Supplementary Fig. 8**).

Modulator minimum support. Once all (TF, M, t) triplets have been processed, modulators are selected based on their support—that is, the number of distinct TF targets they modulate. The minimum support is determined using a permutation test procedure where an identical MINDy run is performed on the same set of candidate modulators and candidate targets except that L_m^+ and L_m^- are chosen at random. This produces a permuted set of MINDy inferences, on the basis of which we can compute the modulator support under the null. The minimum support is then selected as the support that gives the smallest FDR (the ratio between the numbers of selected modulators in the permutated run versus the real run). For MINDy analysis based on DB targets, 15 was determined as the minimum support searching for modulators among all genes, and 7 if candidate modulators are restricted to only signaling proteins and TFs. For MINDy analysis based on AR targets, the minimum support is determined to be 8 when candidate modulators are selected among signaling proteins and TFs.

Modulator mode of action and biological activity. For each significant triplet, we define M as a positive (negative) modulator of the $TF \rightarrow t$ interaction if ΔI

> 0 ($\Delta I < 0$). This indicates only whether M increases or decreases the mutual information between TF and t, and does not necessarily correspond to the biological activity of the modulator (the TF activator or repressor). The latter can be assessed for each tested triplet as:

$$\begin{cases} \text{activator} & \text{if } \rho\,(\mu_t^+ - \mu_t^-) > 0 \\ \text{antagonist} & \text{if } \rho\,(\mu_t^+ - \mu_t^-) < 0 \\ \text{undetermined} & \text{if } \rho\,(\mu_t^+ - \mu_t^-) \approx 0 \end{cases}$$

where ρ is the Pearson correlation between the TF and the candidate target t, and $\mu_t^\pm$ is the mean expression of t in $L_m^\pm$. We assess these differences using a two-sample Student's t-test with 10% type-I error rate (two sided). For modulators affecting more than one MYC target, the biological mode is labeled as undetermined if the undetermined triplets are the majority (>50%) or if neither mode dominates the other by more than 30%. Otherwise it is assigned the dominant mode (**Supplementary Note 13** and **Supplementary Table 12**).

GO enrichment analysis. GO molecular function categories with fewer than 20 and more than 500 genes were excluded. The enrichment of each category was computed using the Fisher's exact test and corrected for multiple testing using the method of Storey and Tibshirani[43].

Promoter analysis. For TFs with an appropriate DNA binding profile in the TRANSFAC professional database (version 6.0)[24], we determined the binding-site enrichment in the promoter regions (defined as 2 kb upstream and 2 kb downstream of the transcription initiation site, masked for repetitive elements) of the targets they modulate. Sequences were retrieved from the UCSC Golden Path database (build 35, May 2004)[44]. The binding profile match threshold was calibrated for a FDR of ≤5% per 1 Kb (in both directions). The binding-site enrichment, compared to a 13,000 random human promoter[5] background, was computed by Fisher's exact test.

Cell lines and cell culture. The human embryonic kidney 293T cells were maintained in DMEM supplemented with 10% FBS and antibiotics. The Burkitt lymphoma cell lines, Ramos, P3HR1, ST486 and ODH-III, were maintained in IMDM supplemented with 10% FBS and antibiotics.

Plasmids. The mammalian expression vectors encoding MYC and TERT-800Luc have been previously described[28]. The mammalian expression vectors encoding BHLHB2/Stra13-FLAG, HDAC1-FLAG and HA-MEF2B were kindly provided by R. Taneja (Mount Sinai School of Medicine, New York), S.L. Schreiber (Harvard University, Cambridge, Massachusetts) and R. Prywes (Columbia University, New York), respectively. HA-STK38 (pcDNA3.1-NDR1-wt) expression vector was provided by B. Hemmings (FMI, Basel, Switzerland). MYC-HA and FLAG-MYC plasmids were constructed by subcloning the corresponding human cDNA amplified by PCR into the pcDNA3 (Invitrogen) and pCMV-Tag2A (Stratagene) vectors, respectively.

Transient transfection and reporter assays. 293T cells were transiently transfected by using the calcium phosphate precipitation method, and luciferase reporter assays were performed as previously described[45,46]. Each transfection was done in duplicate, and luciferase activities were measured 48 h after transfection using the dual-luciferase reporter assay kit (Promega) according to the manufacturer's protocol.

Co-immunoprecipitation assay and western blot analysis. Nuclear cell extracts and whole cell lysates were prepared as previously described[47]. Proteins were analyzed by SDS-PAGE and subsequently by western blot using the following antibodies: anti-MYC (C33 and N262), anti-HDAC1 (N-19) and anti-NDR1 (STK38) (G15) (Santa Cruz Biotechnology); anti-BHLHB2/DEC1 (BL2928) (Bethyl); anti-MEF2B (ab33540) (Abcam); anti-STK38 (2G8-1F3) (Novus Biologicals); Flag M2 and anti-HA beads (Sigma); and hemagglutinin (Roche).

Quantitative chromatin immunoprecipitation (qChIP). ChIP assays were performed as previously described[48,49]. Antibodies used were anti-MYC (N-262, Santa Cruz), anti-BHLHB2/DEC1 (BL2928, Bethyl) and anti-HDAC1 (AB7028, Abcam). The immunoprecipitated chromatin fragments from two independent experiments were pooled together, and the amounts of sample immunoprecipitated

by individual antibody were assessed by quantitative real-time PCR. The oligonucleotide pairs are listed in **Supplementary Table 13**.

In vitro **de-acetylation assay.** FLAG-MYC and CBP-HA expression vectors were transiently transfected into 293T cells. Acetylated-FLAG-MYC was purified with FLAG-beads. FLAG-HDAC1 was purified by FLAG-beads from 293T cells transiently transfected with FLAG-HDAC1 expression vector. Acetylated-FLAG-MYC and FLAG-HDAC1 were incubated at 30 °C for 2 h in a buffer containing 50 mM Tris-HCl, 50 mM NaCl, 4 mM $MgCl_2$, 0.5 mM DTT, 0.2 mM PMSF, 0.02% NP-40 and 5% glycerol with or without 2 µM trichostatin A (TSA) for inhibition of HDAC1 activity.

shRNA and lentiviral infections. Lentiviral vectors for BHLHB2 shRNA (TRCN0000013249), MEF2B shRNA (TRCN0000015739), HDAC1 shRNA (TRCN0000004818), STK38 shRNA (TRCN0000010216) and non-target control shRNA (SHC002) were purchased from Sigma. Lentiviral supernatants were produced by transiently co-transfecting the lentiviral vectors, the packaging vector delta 8.9 and the VSV-G envelope glycoprotein vector as previously described[50,51]. For infection, ODH-III, P3HR1 and ST486 cells (2×10^6 cells/ml) were mixed with viral supernatants, supplemented with 8 µg/ml polybrene and centrifuged for 120 min at 450g. The ODH-III and ST486 cells were collected for analysis 96 h and 60 h after infection, respectively. The lentiviral-infected P3HR1 cells were selected with puromycin (0.5 µg/ml) for 5 d and collected for analysis.

qRT-PCR analysis. Polymerase chain reaction with reverse transcription (RT-PCR) analysis was performed using FastLane Cell cDNA Kit (Qiagen) according to the manufacturer's instructions. qRT-PCR was performed with Quanti Tect SYBR Green PCR kit (Qiagen) using the 7300 Real Time PCR systems (Applied Biosystems) according to manufacturer's instructions. The oligonucleotide primers are described in **Supplementary Table 13**.

Gene-expression profiling after lentiviral mediated silencing of the modulator gene. For each modulator of interest, five samples infected with the modulator-specific shRNA and five infected with non-target control shRNA were obtained. Total RNA was extracted with Trizol reagent (Invitrogen) and purified using the RNeasy kit (Qiagen). Biotinylated cRNA was produced according to the manufacturer instructions, starting with 6 µg of total RNA and following the one-cycle cDNA synthesis protocol (Affymetrix, 701025 Rev.6). Fifteen micrograms of fragmented cRNA were hybridized to HG-U95Av2 microarrays (Affymetrix).

Expression profile data analysis. We determined the gene expression values by the MAS5.0 normalization method provided in Affymetrix's GeneChip Operating Software (GCOS). Differential expression between modulator shRNA and control

shRNA samples was analyzed using two-sample t-test. For the GSEA analysis probe sets were sorted in decreasing order by the $-\log_{10}$ of their t-test P values. Readers are referred to ref. 32 for details of the GSEA algorithm. In the GSEA plot, specific targets (MYC- or MINDy-signatures) are shown as vertical bars against the background of all B cell–expressed genes in the expression profiles, sorted from the most to the least differentially expressed following silencing of the candidate modulator. The curve represents a random walk where the value on the y axis is increased proportionally each time the gene on the x axis is one of the selected targets and decreased if it is part of the background. Weights are chosen proportionally to the statistical significance of the differential expression and to the relative number of targets in the signatures versus the background list, such that the curve starts and ends at $y = 0$. The statistical significance of the GSEA statistics (that is, the maximum height of the curve, called the enrichment score, ES) was determined by permutation test where the ranks of the probe sets were randomly shuffled 1,000 times. To determine the enrichment of MINDy-predicted modulator-specific targets of MYC among the differentially expressed genes, probe sets that rank before the GSEA leading edge (the increasing phase of GSEA profile) were determined to be significantly differentially expressed, and the enrichment was calculated using the Fisher's exact test.

41. Margolin, A. *et al.* ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* **7**, S7 (2006).
42. Niiro, H. & Clark, E.A. Regulation of B-cell fate by antigen-receptor signals. *Nat. Rev. Immunol.* **2**, 945–956 (2002).
43. Storey, J.D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. USA* **100**, 9440–9445 (2003).
44. Karolchik, D. *et al.* The UCSC Genome Browser Database. *Nucleic Acids Res.* **31**, 51–54 (2003).
45. Bereshchenko, O.R., Gu, W. & Dalla-Favera, R. Acetylation inactivates the transcriptional repressor BCL6. *Nat. Genet.* **32**, 606–613 (2002).
46. Chang, C.C., Ye, B.H., Chaganti, R.S. & Dalla-Favera, R. BCL-6, a POZ/zinc-finger protein, is a sequence-specific transcriptional repressor. *Proc. Natl. Acad. Sci. USA* **93**, 6947–6952 (1996).
47. Phan, R.T., Saito, M., Basso, K., Niu, H. & Dalla-Favera, R. BCL6 interacts with the transcription factor Miz-1 to suppress the cyclin-dependent kinase inhibitor p21 and cell cycle arrest in germinal center B cells. *Nat. Immunol.* **6**, 1054–1060 (2005).
48. Niu, H., Cattoretti, G. & Dalla-Favera, R. BCL6 controls the expression of the B7–1/CD80 costimulatory receptor in germinal center B cells. *J. Exp. Med.* **198**, 211–221 (2003).
49. Pasqualucci, L. *et al.* Mutations of the BCL6 proto-oncogene disrupt its negative autoregulation in diffuse large B-cell lymphoma. *Blood* **101**, 2914–2923 (2003).
50. Lois, C., Hong, E.J., Pease, S., Brown, E.J. & Baltimore, D. Germline transmission and tissue-specific expression of transgenes delivered by lentiviral vectors. *Science* **295**, 868–872 (2002).
51. Naldini, L. *et al. In vivo* gene delivery and stable transduction of nondividing cells by a lentiviral vector. *Science* **272**, 263–267 (1996).