



Published in final edited form as:

Nat Biotechnol. 2009 September ; 27(9): 829–839. doi:10.1038/nbt.1563.

Genome-wide Identification of Post-translational Modulators of Transcription Factor Activity in Human B-Cells

Kai Wang^{1,2,*†}, **Masumichi Saito**^{3,*°}, **Brygida C. Bisikirska**², **Mariano J. Alvarez**², **Wei Keat Lim**^{1,2,§}, **Presha Rajbhandari**², **Qiong Shen**³, **Ilya Nemenman**^{2,#}, **Katia Basso**³, **Adam A. Margolin**^{1,2,◆}, **Ulf Klein**³, **Riccardo Dalla-Favera**^{3,4}, and **Andrea Califano**^{1,2,3,‡}

¹ Department of Biomedical Informatics, Columbia University, 1130 St. Nicholas Ave., New York, New York, 10032, USA

² Joint Centers for Systems Biology, Columbia University, 1130 St. Nicholas Ave., New York, New York, 10032, USA

³ Institute of Cancer Genetics and Herbert Irving Comprehensive Cancer Center, Columbia University, 1130 St. Nicholas Ave., New York, New York, 10032, USA

⁴ Department of Pathology and Genetics & Development, Columbia University, 1130 St. Nicholas Ave., New York, New York, 10032, USA

Abstract

The ability of a transcription factor 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 transcription factor activity are still largely unavailable. Here we introduce a systems biology framework, MINDy (Modulator Inference by Network Dynamics), for the genome-wide identification of post-translational modulators of transcription factor 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 turn-over, transcriptional complex formation, and selective enzyme recruitment. MINDy is generally applicable to study the post-translational modulation of mammalian transcription factors in any cellular context. As such it provides a useful resource to dissect context-specific signaling pathways and combinatorial transcriptional regulation.

Users may view, print, copy, download and text and data- mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: http://www.nature.com/authors/editorial_policies/license.html#terms

[‡]To whom correspondence should be addressed. califano@c2b2.columbia.edu, Telephone: 212-851-5183, Fax: 212-851-5149.

^{*}These authors contributed equally to this work.

[†]Present affiliation: Pfizer, Inc., San Diego, CA 92121

[°]Present affiliation: Division of Human Immunology, Kumamoto University, Kumamoto City, Japan

[§]Present affiliation: Therasis Inc., New York, NY 10016

[#]Present affiliation: Computer, Computational and Statistical Sciences Division and Center For Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, NM 87545, USA

[◆]Present affiliation: Cancer Biology Program, The Broad Institute, Cambridge, MA 02142

INTRODUCTION

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, such as phosphorylation, acetylation and ubiquitination enzymes⁷, as well as of those participating in transcription complexes (co-factors), thus making cellular networks highly context dependent.

Although the large-scale reprogramming of the cell's transcriptional logic was studied in yeast^{8, 9}, identification of a repertoire of genes that effect these events remains elusive. Indeed, compared to tools such as ChIP-on-Chip 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 context¹¹, which inferred substrates of 73 kinases. Here we propose and experimentally validate MINDy (Modulator Inference by Network Dynamics), a gene expression profile based method for the systematic identification of genes that modulate a transcription factor's (TF) transcriptional program at the post-translational level, i.e. those encoding proteins that affect the TF's activity without changing its mRNA abundance. These proteins may post-translationally modify the TF (e.g., 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 Information (SI) Section 1.1, this can be efficiently accomplished by computing an information theoretic measure known as the *Conditional Mutual Information* (CMI), $I[T F; t | M]$ ¹², between the expression profile of a transcription factor (TF) and one of its putative targets (t), given the expression of a modulator gene (M). Accurate estimation of the CMI requires exceedingly large datasets. Thus MINDy infers candidate modulators using a related, yet simpler estimator, which we denote as I (see Methods). Briefly, the estimator assesses the statistical significance of the difference in Mutual Information (MI) between the TF and a target in two subsets, each including 35% of samples, in which the modulator is least and most expressed, respectively. The 35% parameter was determined empirically as the one optimizing the identification of proteins in the B-cell receptor signaling pathway as modulators of the MYC TF (see 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, \dots) and a list of potential TF targets (t_1, t_2, \dots). The I 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 Methods. Candidate modulators may include all genes satisfying these constraints (unbiased analysis) or may be filtered by additional criteria, e.g., their molecular functions. Each possible triplet (TF, M, t) is then independently tested using the I estimator. False positives are controlled using appropriate statistical thresholds (see Methods).

A positive or negative *Mode of Action* is determined, depending on whether the TF-target MI increases or decreases as a function of the modulator abundance (Figure 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 MI may decrease as a function of the modulator. Details on the CMI analysis and on how to assess both *Mode of Action* and *Biological Activity* of a modulator are provided in the Methods section.

For illustrative purposes, we show a simple synthetic network (Figure 1B, see Methods), which explicitly models two post-translational modulation events (activation by phosphorylation and co-factor 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 (Figure 1C).

MINDy-based identification of MYC modulators

We applied MINDy to the genome-wide identification of modulators of the MYC proto-oncogene, 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 (see Methods). MYC encodes 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 cancers¹⁷ and can either activate or repress a large number of targets, depending on the cellular context (for a review see [18]).

To study MINDy's robustness and generality, we tested its performance using different MYC-target selections. First, we used 340 literature-validated targets from the MYC database¹⁹ (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 ARACNe4 (AR-targets). Finally, see SI section 1.9, 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.

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 (Table S1) at a False Discovery Rate (*FDR*) of 4.5×10^{-3} (see 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 ($p = 0.002$), transcription factor 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} (see Methods and Table S2 for details on the GO enrichment analysis).

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 (Table S3), using the Ingenuity software (Ingenuity® Systems, www.ingenuity.com). The assumption is that physical interactors are likely to affect MYC protein function, although there may be false positives. While not exhaustive, this provides an independently-established, literature-based dataset to assess algorithm performance (see SI Section 1.8 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 since Ingenuity modulators may not be B cell specific) with recall comparable to high-throughput assays for protein-protein interactions, which on average detect 20% of known interactions²³. 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. Indeed, almost twice as many modulators (18 vs. 11) were found by MINDy among proteins not known to interact directly with MYC.

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 (see Methods). Among these, MINDy identified 91 signaling proteins (Table S4) and 99 TFs (Table S5), respectively, as MYC modulators (FDR = 0.0053). For each modulator, virtually 100% of the *I* tests inferred the same Mode of Action (see columns “T+” and “T–” in Table S4 and S5) 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 (i.e. precision), we performed manual literature curation. Given the labor-intensive nature of this step, only 29 signaling proteins and 35 TFs affecting more than 30 MYC targets were considered. 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 co-factors or antagonists (precision = 17.1%). For TFs with informative binding profiles in TRANSFAC²⁴, we tested whether their binding sites were enriched in promoters of the modulated targets (see Methods). 14 of 35 TFs had appropriate binding profiles. Of these, 11 were highly enriched. Overall, 17 of 35 TFs (precision = 48.6%) were either literature-validated or had enriched binding sites (see Table 1 and Table S6). This suggests that MINDy’s precision may approach experimental assays²³ 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, 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 (See SI Section 1.9 for more details). This suggests that the method is highly robust 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 de-acetylating enzyme (HDAC1). These genes were selected based on the availability of reagents allowing the thorough validation of their activity, and the diversity of the possible mechanisms by which they may modulate MYC activity. Since 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—As a first candidate, we validated STK38, a Serine/Threonine kinase²⁵ inferred by MINDy as a strong positive modulator of MYC activity, affecting 60 targets (Figure 2A, Table S4). We silenced STK38 by lentiviral vector-mediated shRNA expression in ST486 cells. While qRT-PCR analysis showed that MYC mRNA concentration was unchanged following STK38 silencing, MYC protein levels were significantly affected (Figure 2B), suggesting 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 up-regulated following STK38 silencing (Figure 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 down-regulated following STK38 silencing, while the third one, p21^{CIP1}, which is known to be repressed by MYC²⁹, was up-regulated (Figure 2C). These results confirm the observed target-independent down-regulation 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 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 (Figure 2D). Finally, co-immunoprecipitation (co-IP) of epitope-tagged STK38 (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 (Figure 2E). Immunoprecipitation of endogenous MYC using specific antibodies in Ramos cells confirmed that the two proteins can interact physiologically in native cells (Figure 2F). These results suggest that STK38 modulates MYC activity by directly affecting MYC protein stability.

BHLHB2—MINDy infers this TF as a negative modulator of MYC activity, affecting the regulation of 80 targets (Figure 3A, Table S5). 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

repressor³⁰, but not 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 target²⁸. Transient co-transfection of a reporter gene driven by a segment of the human TERT promoter region, carrying two E-boxes [TERT-Luc800²⁸], and vectors encoding MYC or BHLHB2 in 293T cells showed that BHLHB2 represses MYC-mediated transcriptional activation on TERT in a dose-dependent manner. This effect is MYC-dependent since the basal transcriptional activity of the reporter gene is actually mildly increased by BHLHB2 (1.2 fold, Figure 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 (Figure 3C). In order 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 may be highly target specific and thus only a subset of MYC-target may 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 (Figure S5). 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 targets³¹. Thus, without prior knowledge of MYC 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 (see Methods). The analysis confirmed a highly significant enrichment of canonical MYC targets within differentially expressed genes ($p < 0.001$) (Figure 3D). Among the 80 MINDy inferred BHLHB2-modulated MYC-targets, 30 were among the most differentially expressed genes. This constitutes approximately a 2-fold increase over their enrichment in non-differentially expressed genes (FET $p = 8 \times 10^{-5}$). MYC mRNA and protein levels were not affected (data not shown and Figure S5) indicating a post-translational effect. These results validate BHLHB2 as an antagonist of MYC activity.

MEF2B: this TF was inferred as a positive modulator of MYC activity, affecting 14 MYC targets (Figure 4A, Table S5). 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 promoters³³. Details on the validation assays are provided in SI Section 1.11. Briefly, similar to BHLHB2, we showed (a) that MEF2B physically interacts with MYC both exogenously in 293T cells (Figure 4B), and endogenously in P3HR1 and Ramos cells (Figure 4C), (b) that MYC and MEF2B can synergistically activate a TERT reporter gene (Figure 4D), and (c) that genes differentially

expressed following shRNA-mediated silencing of MEF2B were highly enriched in MYC targets by GSEA ($p < 0.001$, Figure 4E), while MYC expression was not affected (Figure S6B).

HDAC1—Finally, MINDy identified the histone deacetylase and well-known transcriptional co-repressor HDAC134, 35 as a modulator of MYC transcriptional activity on 8 MYC targets (Figure 5A, Table S4). Experiments (see SI Section 1.12 for more details) confirmed (a) that HDAC1 and MYC can interact *in vivo*, both exogenously in 293T cells (Figure 5B) as also reported in [36], and endogenously in Ramos and P3HR1 cells (Figure 5C), (b) that genes differentially expressed following shRNA-mediated silencing of HDAC1 were highly enriched in MYC targets by GSEA ($p < 0.001$, Figure 5D), while MYC expression was not affected (Figure S7A), (c) that HDAC1 can deacetylate MYC *in vitro*, following its CBP-mediated acetylation (Figure 5E) which has been shown to increase MYC's activity as a transcriptional activator, and (d) that, based on 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 (Figure 5F). These results suggested that MYC could recruit HDAC1 to repress transcription on selected target genes. Taken together, we have demonstrated that HDAC1 may modulate MYC activity both by de-acetylation of the MYC protein and by transcriptional repression of selected targets.

Extension to other Transcription Factors: 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 which we had already collected gene expression profile data following shRNA-mediated silencing. Specifically, we tested whether their MINDy-inferred targets were enriched in differentially expressed genes (by GSEA), following lentivirus mediated shRNA silencing of the corresponding modulators. As shown in Table S13, 75% of the TFs inferred by MINDy as modulated by any of the four modulators with support greater than 100 targets (33 out of 44, $p = 5.1 \times 10^{-34}$) could be experimentally validated by the analysis. 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 TF 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. (See SI Section 1.13 for further details).

DISCUSSION

We have introduced MINDy, a novel 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, both of MYC and of other TF.

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 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 vs. AR-targets). Additionally, MINDy's formulation is relatively simple, requiring only the availability of a large gene expression profile dataset ($n = 200$ 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, include mostly either genes that are not expressed or genes whose availability is so tightly regulated (e.g. 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, less 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. If desired, the more general test may be used without relying on the assumption $I[TF; M] = 0$ (see SI Section 1.1). 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 MI may not change substantially. In this case the multi-information could be used instead of the conditional.

Comparison with existing methods reveals that MINDy is unique in its ability to discover large numbers of modulators of the same TF. Finding the optimal Bayesian Network structure, assuming arbitrary interactions among genes' parents, for instance, is hyper-exponential in the number of parents. As a result, dissecting network topologies with large numbers (tens to hundreds) of upstream modulators, as is the case for MYC, may be difficult. Other methods³⁹, while promising in a yeast context, have not yet been extended to mammalian networks. Finally, comparison with a recently introduced algorithm, NetworKIN11, shows that the latter is restricted to substrates of only 73 kinases, from 20 families, while MINDy can be used to dissect post-translational interactions of a much wider nature, including phosphorylation, acetylation, chromatin modification, formation of transcriptional complexes, and binding site antagonism, as shown for STK38, HDAC1, MEF2B and BHLHB2, just to mention those experimentally validated in this manuscript.

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 genes⁴⁰. This could be valuable because TFs are generally considered difficult pharmacological target.

Although preliminarily applied to the identification of modulators of MYC for experimental validation purposes, MINDy has already provided novel biological insights of general significance. First, the results indicate that not all modulators can influence the program of a TF in a global fashion, but 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 TF they control in different cellular contexts or depending on

different signals. Second, MINDy provided novel information about 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 NCI and by the National Center for Biomedical Computing at Columbia University (<http://wiki.c2b2.columbia.edu/workbench/index.php/Home>).

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 standard deviation $\sigma < 0.3\mu$, were considered uninformative and were thus excluded, leaving 7907 probe sets for the analysis. Table S9 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 (Figure 1B). It includes the transcription factor (TF), an activating protein kinase (K), a co-factor (cTF) forming a transcriptionally active complex with the TF, and three downstream TF targets. The full kinetic model is described in Table S10. 250 synthetic expression profiles were generated from this model by: (a) randomly sampling the TF, K, and cTF mRNA concentration from independent Normal distributions ($\mu = 4$ and $\sigma = 1$), (b) simulating network dynamics until a steady state was reached, and (c) 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 SI Section 1.3. The statistical significance test for the *independence constraint* is based on the Mutual Information, as described in [59] (see also SI Sections 1.5 and 1.6).

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 $t \rightarrow TF$ and $t \leftarrow M$, MINDy assesses whether the CMI, $I[TF; t|M]$, is constant as a function of M . Assuming that the CMI is a monotonic function of M (see SI Section 1.10), this can be efficiently tested by measuring

$\Delta I = I[TF; t|M \in L_m^+] - I[TF; t|M \in L_m^-] \neq 0$, 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 activity⁴²) on regulating canonical MYC target genes (see SI Section 1.2). MI was computed using the Gaussian Kernel estimator of [59] (see also SI Section 1.5). The p -value corresponding to a specific I is obtained by permutation tests (see SI Section 1.1), 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 $I[TF; x] > I[TF; t]$ and $I[t; x] > I[t; TF]$ in both L_m^\pm , showing an indirect relationship between TF and t mediated by x , as suggested by the Data Processing Inequality⁵⁹ (see SI Section 1.7 and Figure S8 for details).

Modulator minimum support

Once MINDy inference is made on individual (TF, M, t) , modulators are selected based on its support, i.e. the number of TF targets it modulates. 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, based on which we can compute the modulator support under the null. The minimum support is then selected as the support that gives the smallest FDR (i.e. the ratio between the numbers of selected modulators in the permuted run vs. real run). For MINDy analysis based on DB-targets, fifteen 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 $I > 0$ ($I < 0$). 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 (i.e. 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 μ_t^\pm is the mean expression of t in L_m^\pm . We assess these differences using a two-sample Student 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 (See SI Section 1.4 and Table S11 for more description).

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 Tibshirani⁶⁰.

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 2Kb upstream and 2Kb downstream of the transcription initiation site, masked for repetitive elements) of the targets they modulate. Sequences were retrieved from the UCSC Golden Path database (build35, May 2004)⁶¹. The binding profile match threshold was calibrated for a FDR = 5% per 1 Kb (in both directions). The binding-site enrichment, compared to a 13,000 random human promoter⁵ 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 described²⁸. The mammalian expression vectors encoding BHLHB2/Stra13-FLAG, HDAC1-FLAG, and HA-MEF2B were kindly provided by Dr R. Taneja (Mount Sinai School of Medicine, New York, NY), Dr S.L. Schreiber (Harvard University, Cambridge, MA), and Dr R. Prywes (Columbia University, New York, NY), respectively. HA-STK38 (pcDNA3.1-NDR1-wt) expression vector was provided by Dr. 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^{62, 63}. Each transfection was done in duplicate and luciferase activities were measured 48h post-transfection using 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 described⁶⁴. 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^{65, 66}. 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 Table S12.

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 2hrs in a buffer containing 50mM Tris-HCl, 50mM NaCl, 4mM MgCl₂, 0.5mM DTT, 0.2mM PMSF, 0.02% NP-40 and 5% Glycerol with or without 2μM 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^{67, 68}. For infection, ODH-III, P3HR1 and ST486 cells (2×10⁶ cells/ml) were mixed with viral supernatants, supplemented with 8μg/ml polybrene and centrifuged for 120 min at 450g. The ODH-III and ST486 cells were collected for analysis 96h and 60h post-infection, respectively. The lentiviral infected P3HR1 cells were selected with puromycin (0.5 μg/ml) for 5 days 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 Table S12.

Gene-expression profiling after lentiviral mediated silencing of the modulator gene

For each modulator of interest, 5 samples infected with the modulator-specific shRNA and 5 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 μ g of total RNA and following the one-cycle cDNA synthesis protocol (Affymetrix, 701025 Rev.6). 15 μ g of fragmented cRNA were hybridized to HG-U95Av2 microarrays (Affymetrix).

Expression profile data analysis

We determined the gene expression values by 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 [32] for details on 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 vs. the background list, such that the curve starts and ends at $y = 0$. The statistical significance of the GSEA statistics (i.e., the maximum height of the curve, called Enrichment Score, ES) was determined by permutation test where the ranks of the probe sets were randomly shuffled 1000 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 (i.e. the increasing phase of GSEA profile) were determined to be significantly differentially expressed, and the enrichment was calculated using the Fisher's exact test.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the National Cancer Institute (R01CA109755), the National Institute of Allergy and Infectious Diseases (R01AI066116), and the National Centers for Biomedical Computing NIH Roadmap Initiative (U54CA121852). I.N. was supported in part by the National Institute for General Medical Sciences (R21GM080216). A.A.M. is supported by an IBM Ph.D. fellowship.

REFERENCES

1. Faith JJ, et al. Large-Scale Mapping and Validation of Escherichia coli Transcriptional Regulation from a Compendium of Expression Profiles. *PLoS Biol.* 2007; 5:e8. [PubMed: 17214507]
2. Friedman N. Inferring cellular networks using probabilistic graphical models. *Science.* 2004; 303:799–805. [PubMed: 14764868]

3. Gardner TS, di Bernardo D, Lorenz D, Collins JJ. Inferring genetic networks and identifying compound mode of action via expression profiling. *Science*. 2003; 301:102–105. [PubMed: 12843395]
4. Basso K, et al. Reverse engineering of regulatory networks in human B cells. *Nat Genet*. 2005; 37:382–390. [PubMed: 15778709]
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*. 2003; 13:773–780. [PubMed: 12727897]
6. Stuart JM, Segal E, Koller D, Kim SK. A Gene-Coexpression Network for Global Discovery of Conserved Genetic Modules. *Science*. 2003; 302:249–255. [PubMed: 12934013]
7. Zeitlinger J, et al. Program-Specific Distribution of a Transcription Factor Dependent on Partner Transcription Factor and MAPK Signaling. *Cell*. 2003; 113:395. [PubMed: 12732146]
8. Luscombe NM, et al. Genomic analysis of regulatory network dynamics reveals large topological changes. *Nature*. 2004; 431:308–312. [PubMed: 15372033]
9. Segal E, et al. Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nat Genet*. 2003; 34:166–176. [PubMed: 12740579]
10. Ren B, et al. Genome-wide location and function of DNA binding proteins. *Science*. 2000; 290:2306–2309. [PubMed: 11125145]
11. Linding R, et al. Systematic discovery of in vivo phosphorylation networks. *Cell*. 2007; 129:1415–1426. [PubMed: 17570479]
12. Cover, TM.; Thomas, J. *Elements of Information Theory*. 2. Wiley Interscience; 2006.
13. Mani KM, et al. A systems biology approach to prediction of oncogenes and molecular perturbation targets in B-cell lymphomas. *Mol Syst Biol*. 2008; 4:169. [PubMed: 18277385]
14. Basso K, et al. Reverse engineering of regulatory networks in human B cells. *Nat Genet*. 2005; 37:382–390. [PubMed: 15778709]
15. Dang CV, et al. The c-Myc target gene network. *Seminars in Cancer Biology*. 2006; 16:253–264. [PubMed: 16904903]
16. Dominguez-Sola D, et al. Non-transcriptional control of DNA replication by c-Myc. *Nature*. 2007; 448:445–451. [PubMed: 17597761]
17. Pelengaris S, Khan M, Evan G. c-MYC: more than just a matter of life and death. *Nat Rev Cancer*. 2002; 2:764–776. [PubMed: 12360279]
18. Zeller KI, Jegga AG, Aronow BJ, O'Donnell KA, Dang CV. An integrated database of genes responsive to the Myc oncogenic transcription factor: identification of direct genomic targets. *Genome Biol*. 2003; 4:R69. [PubMed: 14519204]
19. Zeller K, 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 Biology*. 2003; 4:R69. [PubMed: 14519204]
20. Levens DL, Reconstructing MYC. *Genes Dev*. 2003; 17:1071–1077. [PubMed: 12730130]
21. Patel JH, et al. The c-MYC Oncoprotein Is a Substrate of the Acetyltransferases hGCN5/PCAF and TIP60. *Mol Cell Biol*. 2004; 24:10826–10834. [PubMed: 15572685]
22. Sears R, et al. Multiple Ras-dependent phosphorylation pathways regulate Myc protein stability. *Genes Dev*. 2000; 14:2501–2514. [PubMed: 11018017]
23. Yu H, et al. High-quality binary protein interaction map of the yeast interactome network. *Science*. 2008; 322:104–110. [PubMed: 18719252]
24. Matys V, et al. TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic acids research*. 2003; 31:374–378. [PubMed: 12520026]
25. Tamaskovic R, Bichsel SJ, Hemmings BA. NDR family of AGC kinases--essential regulators of the cell cycle and morphogenesis. *FEBS Lett*. 2003; 546:73–80. [PubMed: 12829239]
26. Patel JH, McMahon SB. BCL2 is a downstream effector of MIZ-1 essential for blocking c-MYC-induced apoptosis. *J Biol Chem*. 2007; 282:5–13. [PubMed: 17082179]
27. Schuhmacher M, et al. The transcriptional program of a human B cell line in response to Myc. *Nucl Acids Res*. 2001; 29:397–406. [PubMed: 11139609]

28. Wu KJ, et al. Direct activation of TERT transcription by c-MYC. *Nature genetics*. 1999; 21:220–224. [PubMed: 9988278]
29. Seoane J, Le HV, Massague J. Myc suppression of the p21(Cip1) Cdk inhibitor influences the outcome of the p53 response to DNA damage. *Nature*. 2002; 419:729–734. [PubMed: 12384701]
30. St-Pierre B, Flock G, Zacksenhaus E, Egan SE. Stra13 homodimers repress transcription through class B E-box elements. *J Biol Chem*. 2002; 277:46544–46551. [PubMed: 12297495]
31. Wanzel M, Herold S, Eilers M. Transcriptional repression by Myc. *Trends Cell Biol*. 2003; 13:146–150. [PubMed: 12628347]
32. Subramanian A, et al. From the Cover: Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *PNAS*. 2005; 102:15545–15550. [PubMed: 16199517]
33. Molkenin JD, et al. MEF2B is a potent transactivator expressed in early myogenic lineages. *Mol Cell Biol*. 1996; 16:3814–3824. [PubMed: 8668199]
34. Wade PA. Transcriptional control at regulatory checkpoints by histone deacetylases: molecular connections between cancer and chromatin. *Hum Mol Genet*. 2001; 10:693–698. [PubMed: 11257101]
35. Satou A, Taira T, Iguchi-Arigo SMM, 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*. 2001; 276:46562–46567. [PubMed: 11585818]
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*. 2008; 99:309–315. [PubMed: 18271930]
37. Margolin AA, et al. ChIP-on-chip significance analysis reveals large-scale binding and regulation by human transcription factor oncogenes. *Proc Natl Acad Sci U S A*. 2009; 106:244–249. [PubMed: 19118200]
38. Robertson G, et al. Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat Methods*. 2007; 4:651–657. [PubMed: 17558387]
39. Steffen M, Petti A, Aach J, D'haeseleer P, Church G. Automated modelling of signal transduction networks. *BMC Bioinformatics*. 2002; 3
40. Hopkins AL, Groom CR. The druggable genome. *Nat Rev Drug Discov*. 2002; 1:727–730. [PubMed: 12209152]
41. Luscher B, Kuenzel EA, Krebs EG, Eisenman RN. Myc oncoproteins are phosphorylated by casein kinase II. *EMBO J*. 1989; 8:1111–1119. [PubMed: 2663470]
42. Niiro H, Clark EA. Regulation of B-cell fate by antigen-receptor signals. *Nature Reviews Immunology*. 2002; 2:945.
43. Feng XH, Liang YY, Liang M, Zhai W, Lin X. Direct interaction of c-Myc with Smad2 and Smad3 to inhibit TGF-beta-mediated induction of the CDK inhibitor p15(Ink4B). *Mol Cell*. 2002; 9:133–143. [PubMed: 11804592]
44. Li ZH, et al. Cross-linking of surface immunoglobulin activates src-related tyrosine kinases in WEHI 231 cells. *Biochem Biophys Res Commun*. 1992; 187:1536–1544. [PubMed: 1384474]
45. Swanson HI, Yang JH. Specificity of DNA binding of the c-Myc/Max and ARNT/ARNT dimers at the CACGTG recognition site. *Nucl Acids Res*. 1999; 27:3205–3212. [PubMed: 1093/nar/27.15.3205]
46. Machida N, et al. Mitogen-activated protein kinase kinase kinase 4 as a putative effector of Rap2 to activate the c-Jun N-terminal kinase. *J Biol Chem*. 2004; 279:15711–15714. [PubMed: 14966141]
47. Noguchi K, et al. Regulation of c-Myc through Phosphorylation at Ser-62 and Ser-71 by c-Jun N-Terminal Kinase. *J Biol Chem*. 1999; 274:32580–32587. [PubMed: 10551811]
48. Jeffrey KL, et al. Positive regulation of immune cell function and inflammatory responses by phosphatase PAC-1. *Nat Immunol*. 2006; 7:274–283. [PubMed: 16474395]
49. Lang R, Hammer M, Mages J. DUSP meet immunology: dual specificity MAPK phosphatases in control of the inflammatory response. *J Immunol*. 2006; 177:7497–7504. [PubMed: 17114416]

50. Kitaura H, et al. Reciprocal regulation via protein-protein interaction between c-Myc and p21(cip1/waf1/sdi1) in DNA replication and transcription. *J Biol Chem.* 2000; 275:10477–10483. [PubMed: 10744738]
51. Sutherland CL, Heath AW, Pelech SL, Young PR, Gold MR. Differential activation of the ERK, JNK, and p38 mitogen-activated protein kinases by CD40 and the B cell antigen receptor. *J Immunol.* 1996; 157:3381–3390. [PubMed: 8871635]
52. Shi Y, Sharma A, Wu H, Lichtenstein A, Gera J. Cyclin D1 and c-myc internal ribosome entry site (IRES)-dependent translation is regulated by AKT activity and enhanced by rapamycin through a p38 MAPK- and ERK-dependent pathway. *J Biol Chem.* 2005; 280:10964–10973. [PubMed: 15634685]
53. Derijard B, et al. Independent human MAP-kinase signal transduction pathways defined by MEK and MKK isoforms. *Science.* 1995; 267:682–685. [PubMed: 7839144]
54. Burkhardt AL, Brunswick M, Bolen JB, Mond JJ. Anti-immunoglobulin stimulation of B lymphocytes activates src-related protein-tyrosine kinases. *Proceedings of the National Academy of Sciences of the United States of America.* 1991; 88:7410–7414. [PubMed: 1714601]
55. Chen CR, Kang Y, Siegel PM, Massague J. E2F4/5 and p107 as Smad cofactors linking the TGFbeta receptor to c-myc repression. *Cell.* 2002; 110:19–32. [PubMed: 12150994]
56. Chuang CF, Ng SY. Functional divergence of the MAP kinase pathway. ERK1 and ERK2 activate specific transcription factors. *FEBS Lett.* 1994; 346:229–234. [PubMed: 8013639]
57. Liou J, et al. HPK1 is activated by lymphocyte antigen receptors and negatively regulates AP-1. *Immunity.* 2000; 12:399–408. [PubMed: 10795738]
58. Cheng SWG, et al. c-MYC interacts with INI1/hSNF5 and requires the SWI/SNF complex for transactivation function. *Nature genetics.* 1999; 22:102–105. [PubMed: 10319872]
59. Margolin A, et al. ARACNE: An Algorithm for the Reconstruction of Gene Regulatory Networks in a Mammalian Cellular Context. *BMC Bioinformatics.* 2006; 7:S7. [PubMed: 16723010]
60. Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences of the United States of America.* 2003; 100:9440–9445. [PubMed: 12883005]
61. Karolchik D, et al. The UCSC Genome Browser Database. *Nucl Acids Res.* 2003; 31:51–54. [PubMed: 12519945]
62. Bereshchenko OR, Gu W, Dalla-Favera R. Acetylation inactivates the transcriptional repressor BCL6. *Nature genetics.* 2002; 32:606–613. [PubMed: 12402037]
63. Chang CC, Ye BH, Chaganti RS, Dalla-Favera R. BCL-6, a POZ/zinc-finger protein, is a sequence-specific transcriptional repressor. *Proceedings of the National Academy of Sciences of the United States of America.* 1996; 93:6947–6952. [PubMed: 8692924]
64. Phan RT, 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.* 2005; 6:1054–1060. [PubMed: 16142238]
65. 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.* 2003; 198:211–221. [PubMed: 12860928]
66. Pasqualucci L, et al. Mutations of the BCL6 proto-oncogene disrupt its negative autoregulation in diffuse large B-cell lymphoma. *Blood.* 2003; 101:2914–2923. [PubMed: 12515714]
67. Lois C, Hong EJ, Pease S, Brown EJ, Baltimore D. Germline transmission and tissue-specific expression of transgenes delivered by lentiviral vectors. *Science.* 2002; 295:868–872. [PubMed: 11786607]
68. Naldini L, et al. In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. *Science.* 1996; 272:263–267. [PubMed: 8602510]

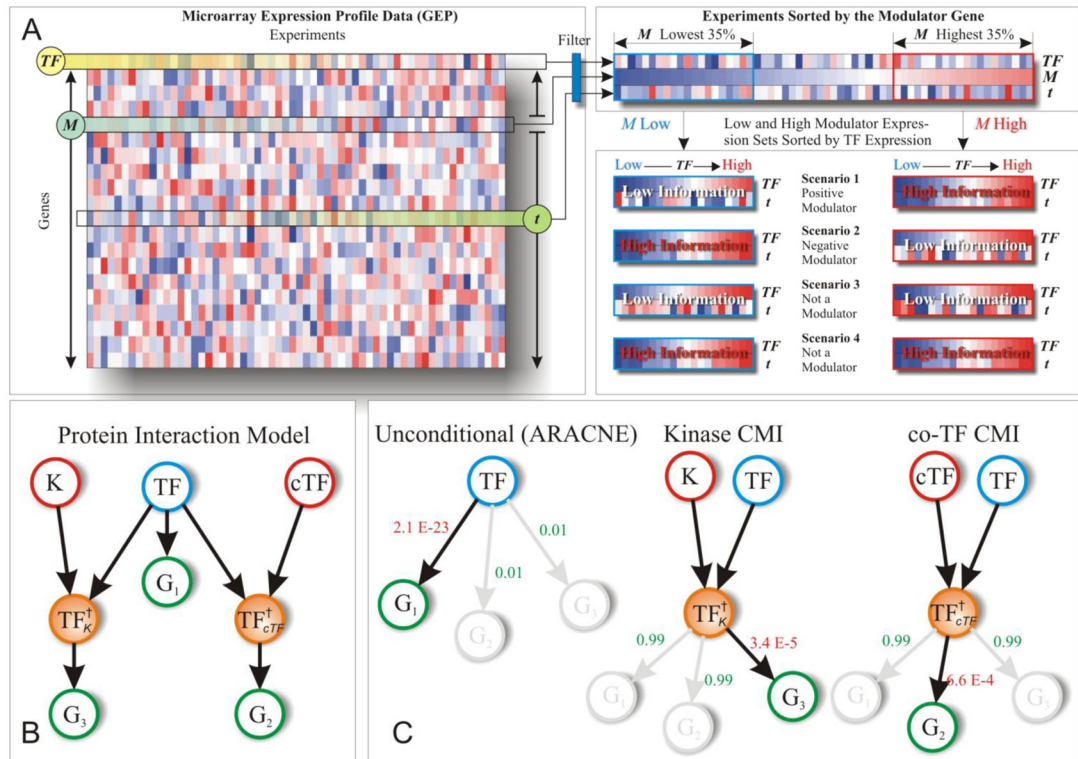


Figure 1. MINDy algorithm

(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 and a target to be tested by MINDy (M and t respectively) 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 (e.g., 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 vs. M -low. **Scenario 1 (Positive Inferred Modulator):** Significant increase in MI (i.e. more correlation in the M -high set than in the M -low set); **Scenario 2 (Negative Inferred Modulator):** Significant decrease in MI (i.e. less correlation in the M -high set than in the M -low set); **Scenario 3&4 (Not a Modulator):** No significant MI 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_K^\dagger** : activated TF 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. More details on the synthetic network are provided in the Methods.

(C) Networks reconstructed by ARACNe and MINDy. ARACNe (left) is based on pairwise MI, thus captures only the unconditional interaction between TF and G_1 . MINDy, on the

other hand, infers three-way interactions using the CMI. 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.

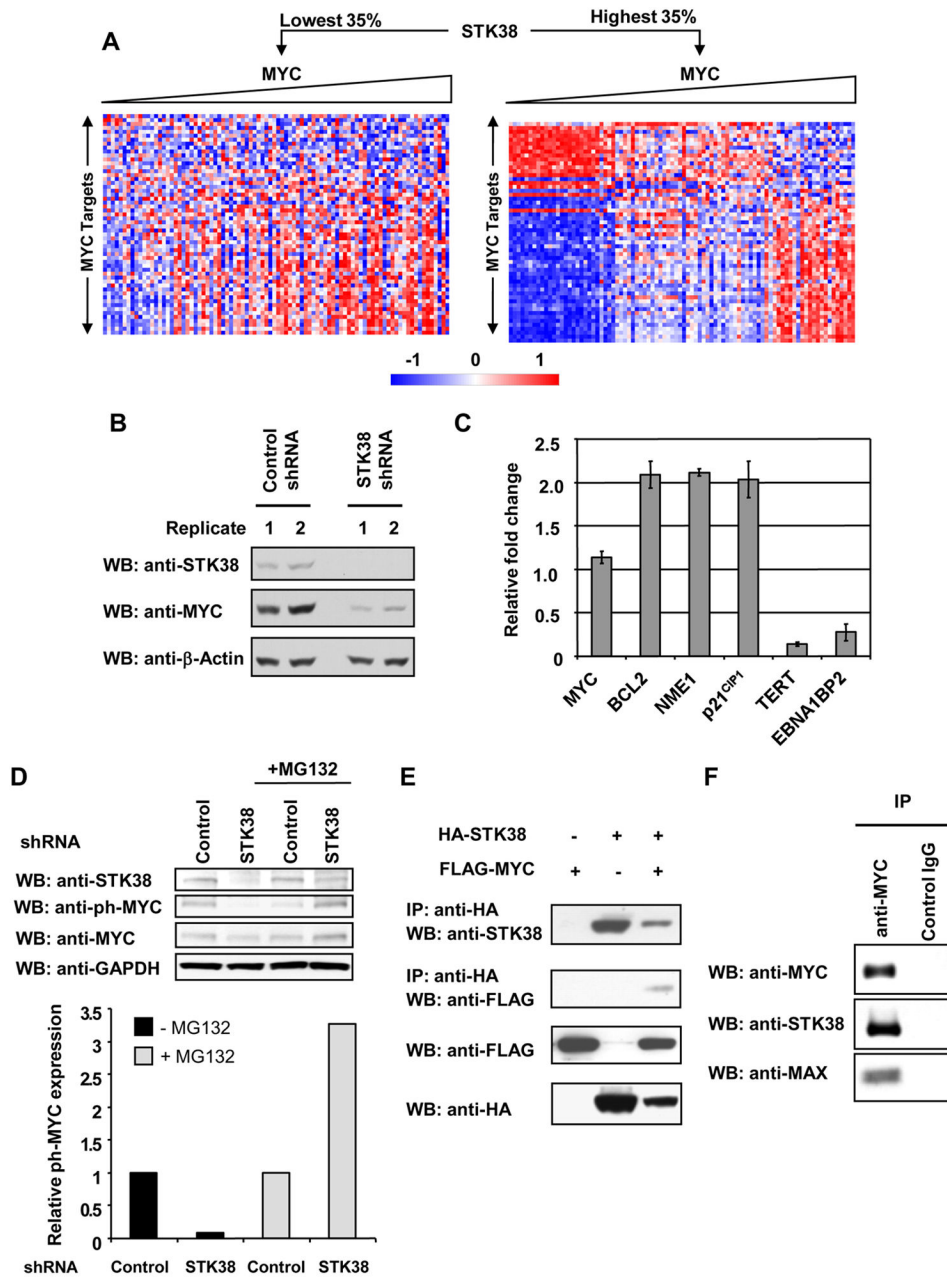


Figure 2. STK38 regulates the stability of MYC protein

(A) Visualization of MINDy output. Gene expression profiles were 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 hours post-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 genes 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 housekeeping gene. Bars represent mean \pm standard error (SE) 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 hours post-infection). ST486 cells were infected with lentivirus expressing STK38 shRNA or control shRNA. The cells were treated with or without MG132 for 4 hours 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 (i.e. ratio of ph-MYC to total MYC) are normalized to that of cells transduced with control shRNA.

(E) HA-STK38 expression vector was transiently transfected into 293T cells with FLAG-MYC expression vector. At 24-hour post-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.

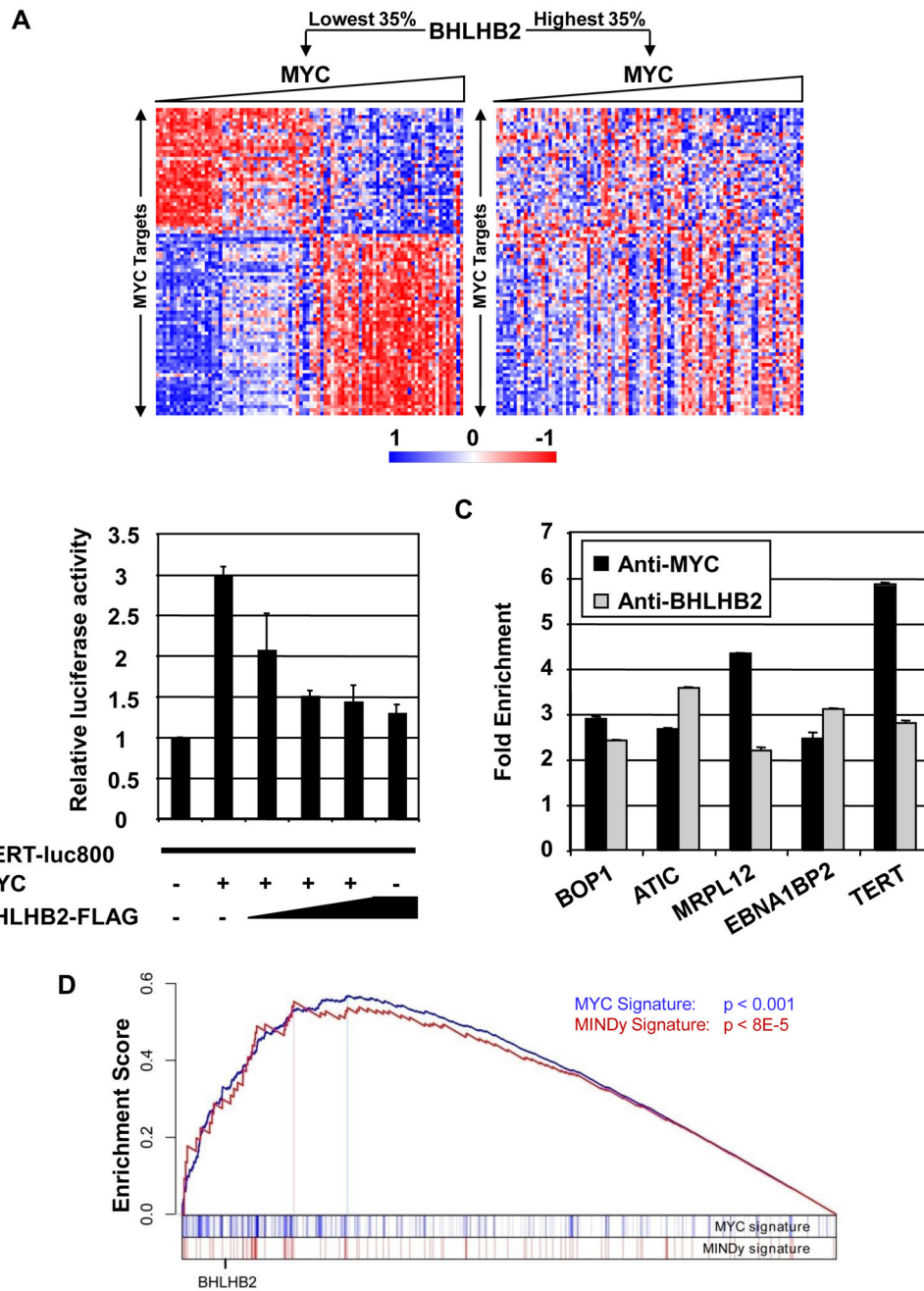


Figure 3. BHLHB2 antagonizes MYC activity in B-cells

(A) Visualization of MINdy output. See Figure 2A for interpretation of this graph.

(B) TERT promoter-driven luciferase reporter construct, TERT-800Luc, 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 hours post transfection and normalized against the Renilla activity. Each experiment was done in duplicate and data were shown as mean \pm SE 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 an irrelevant antibody (rabbit IgG) as a control. The precipitated DNA fragments were assessed by quantitative PCR and shown as mean \pm SE.

(D) GSEA enrichment analysis for BHLHB2 silencing. Gene expression profiles were generated for 5 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 vs. the control shRNA. Most differentially expressed genes are towards 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, while 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. (See Methods for details on GSEA test and statistical significance assessment.)

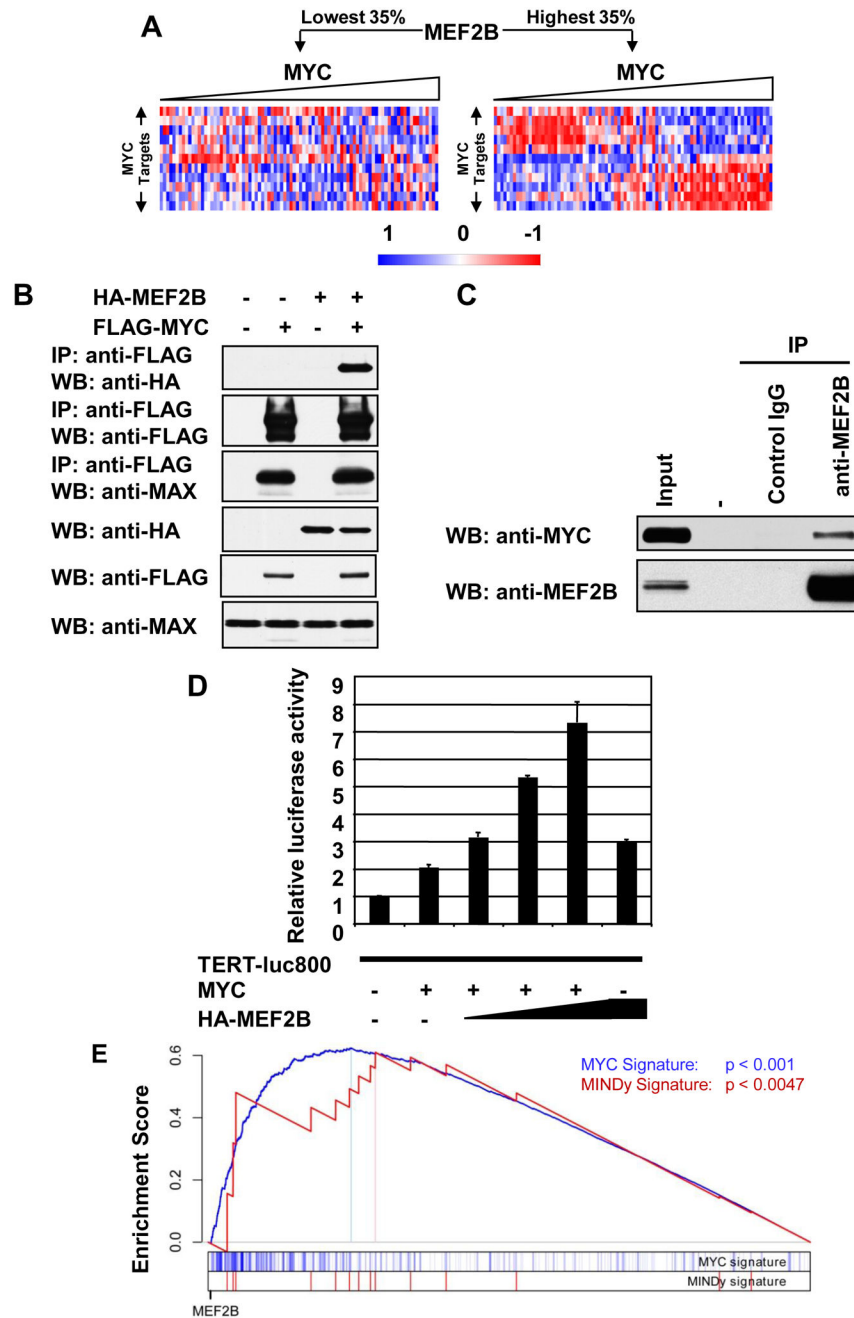


Figure 4. MEF2B enhances MYC activity via protein-protein interaction

(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-hour post-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 rabbit serum as a control. The precipitates were analyzed by Western blotting with anti-MEF2B and anti-MYC antibodies.

(D) 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 hours post-transfection and normalized against the Renilla activity. Each experiment was done in duplicate and data were shown as mean \pm SE of three independent experiments.

(E) GSEA enrichment analysis for MEF2B silencing. Gene expression profiles were generated for 5 samples of P3HR1 cells transduced with MEF2B-specific or control shRNA using a lentiviral vector. See Figure 3D for further interpretation of this graph.

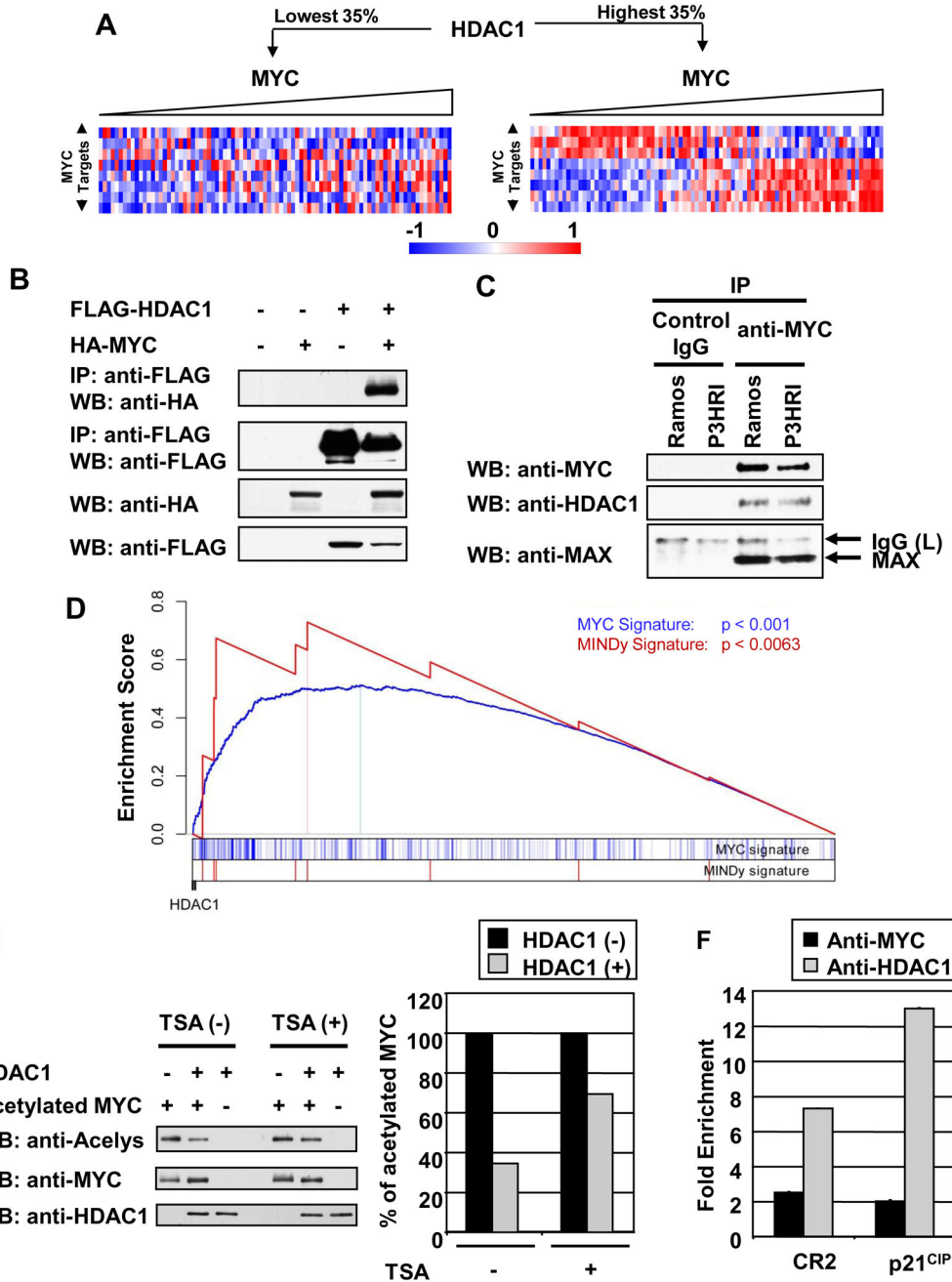


Figure 5. MYC selectively recruits HDAC1 to its targets as co-repressor

(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-hour post-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 5 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 transiently transfected 293T cells, was incubated with acetylated-MYC protein *in vitro*. The signal intensities of acetylated-MYC and total amount of MYC were measured by 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 \pm SE.

Table 1

Top modulators of MYC inferred by MINDy among signaling proteins and TFs.

Modulator	Signaling Protein				Transcription Factor				BSE <i>p</i> -value	
	#T	MoA	BA	Evidence	Modulator	#T	MoA	BA		Evidence
CSNK2A1	178	↑	+	Direct41, 42	SMAD3	159	↓	-	Direct43	-
HCK	111	↓	-	Pathway42, 44	AHR	156	↓	+	Direct45	-
PPAP2B	109	↓	-	-	CREM	125	↓	+	-	1×10^{-36}
SAT	99	↓	-	-	DDIT3	96	↓	-	-	0.04
MAP4K4	83	↓	+	Pathway46, 47	DRAP1	92	↑	-	-	-
DUSP2	82	↓	-	Pathway48, 49	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
CDKN1A	58	↓	-	Direct50	HOXB7	65	↑	-	-	-
MTMR6	52	↓	+	-	BACHI	64	↓	-	-	5×10^{-9}
PRKACB	50	↓	+	-	SOX5	64	↓	+	-	0.66
NEK9	47	↑	-	-	FOS	57	↓	-	-	2×10^{-3}
MYST1	45	↑	-	-	ARNT	55	↑	-	Direct45	-
MAPK13	41	↑	-	Pathway51, 52	IRF1	55	↓	-	-	6×10^{-8}
OXSRI	41	↓	+	-	ETV5	54	↓	+	-	-
DUSP4	40	↓	-	-	TCF12	51	↑	-	-	2×10^{-3}
MAP2K3	39	↓	-	Pathway53	SMAD2	50	↑	-	Direct43	-
FYN	38	↓	-	Pathway42, 54	NFATC4	46	↑	-	-	-
PPP4R1	35	↓	+	-	E2F5	45	↑	-	Direct55	-
MAPK1	35	↑	-	Direct22, 56	JUN	45	↓	-	-	1×10^{-7}
MAP4K1	34	↓	u	Pathway57	CUTL1	39	↓	+	-	-

Signaling Protein				Transcription Factor						
Modulator	#T	MoA	BA	Evidence	Modulator	#T	MoA	BA	Evidence	BSE <i>p</i> -value
CSNK1E	33	↑	-	-	ESR2	38	↓	+	-	-
NEK7	32	↑	-	-	ZNF354A	37	↑	-	-	-
CSNK2A2	31	↑	+	Direct4I	MAF	37	↓	+	-	-
TRIB2	30	↓	+	-	SMARCB1	37	↑	-	Direct58	-
HDAC1	8	↑	+	Validated	BRD8	34	↑	+	-	-
					DBP	34	↑	-	-	6×10^{-3}
					CBFA2T3	33	↑	-	-	-
					ESR1	31	↑	-	-	0.04
					RELB	31	↓	-	-	-
					TFEC	30	↓	u	-	0.92
					MEF2B	14	↑	+	Validated	-

Modulators on each side are sorted in decreasing order based on the number of MYC targets they modulate. #T: total number of MYC targets affected by the modulator; MoA - MINDy Mode of Action: MI increase (↑), MI decrease (↓); BA - Biological Activity: activator (+), antagonist (-), undetermined (u) (see Methods). Evidence: literature or experimental support: "Direct" (blue colored) denotes modulators physically interacting with MYC protein or MYC binding sites; "Pathway" (green colored) indicates literature validated role in pathway-mediated (e.g. B-cell receptor and mitogen-activated protein kinase pathways) modulation of MYC activity; "Validated" (red colored) indicates modulators experimentally validated in this study. BSE: Binding site enrichment analysis. *P*-values in red indicate statistical significance ($p < 0.05$) in BSE.