Supporting Information

Vallat et al. 10.1073/pnas.1211130110

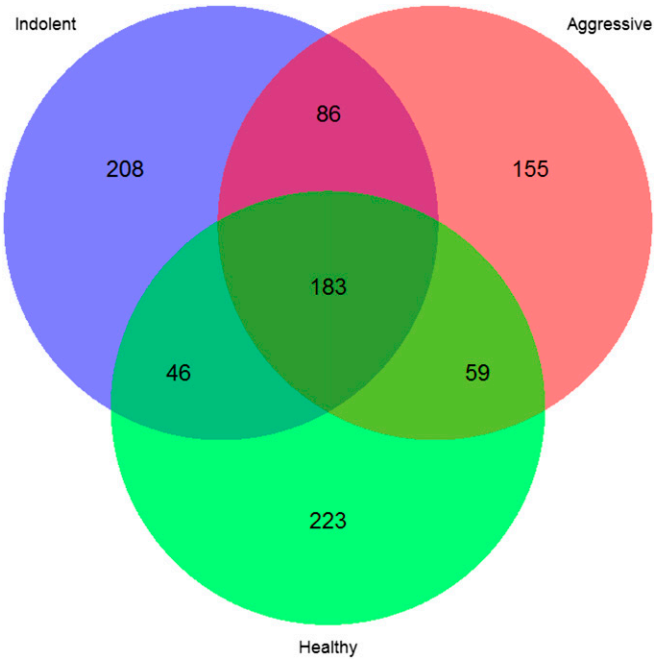

Fig. S1. Venn diagram: distribution of the 960 probe sets between the three cell groups. A total of 960 probe sets was retained for all of the subjects across the three different cell groups. A core of 183 probe sets is shared by the three groups.

Fig. S2. Each graphic represents genes in a specific categorical time label (1–4, from left to right) and their connections, showing how the signal is spreading through the aggressive network.

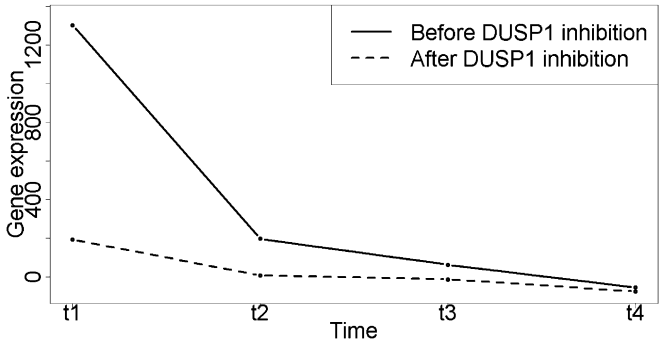

Fig. S3. *DUSP1* is the targeted gene for the knock-down experiment. We show its expression before and after the inhibition experiment.

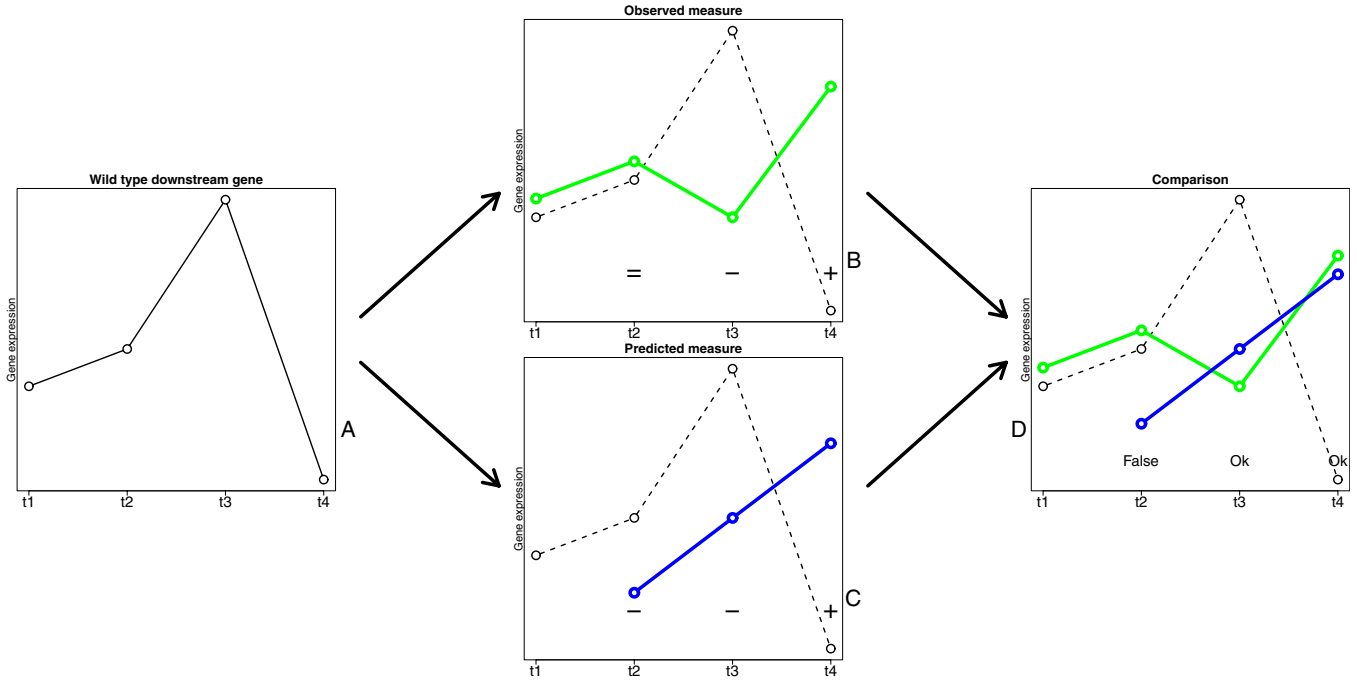

Fig. S4. Principle of the validation experiment. Graphic *A* represents a gene expression before inhibition of a targeted gene. Graphic *B* shows how this gene expression evolves after silencing this targeted gene, and graphic *C* shows the predicted gene expression. For these two last graphics, for time t_2–t_4 we assigned a plus (+), minus (−), or equal (=) sign label if gene expression after silencing is greater, smaller, or equal, respectively, to gene expression before silencing. For this gene, graphic *D* shows that we made two good predictions of three in this example.

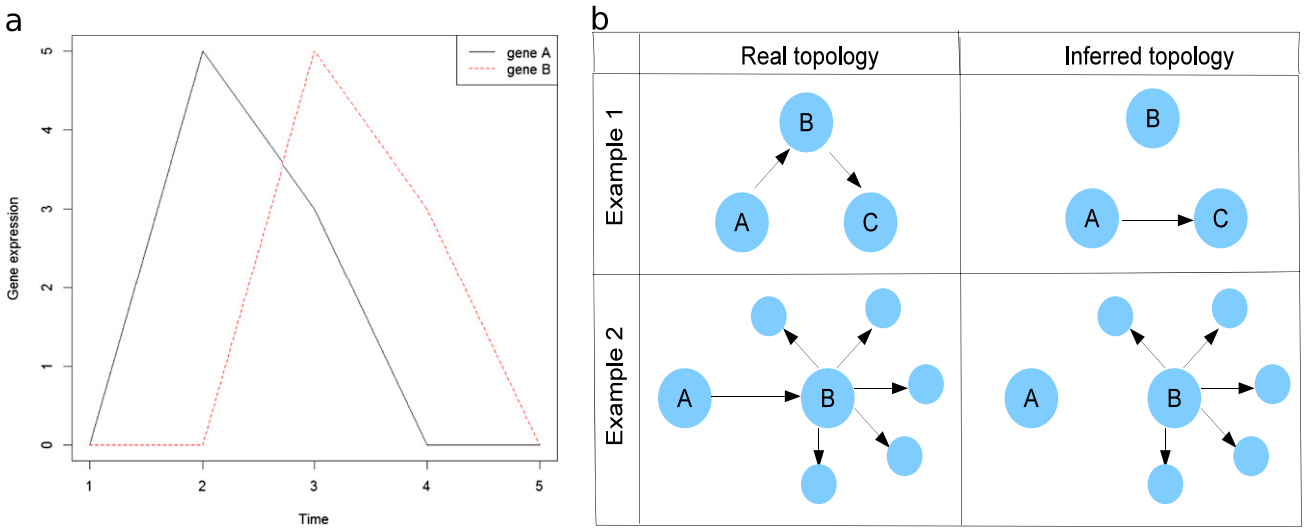

Fig. S5. Schematic representation of specific constraints related to prediction abilities in model inference. This ability to predict the transcriptional effect of a modulation in the network is crucial to predict a gene expression level modification after a knockdown experiment. For instance, given a situation where a gene A regulates the expression of a gene B (with a time lag between activation of gene A and gene B, as schematically shown in *A*, which in turn regulates gene C, we want to predict the absence of a link between B and C if gene A is knocked down. Importantly, this predictive capacity requires much more complexity than inference alone. More than inferring a network topology, a predictive method should be able to learn how the biological signal spreads in this network. To go further, the best algorithms for reverse-engineering are not necessarily the best methods for predicting purposes, as explained in *B* with two simple examples. In the first example, a real network is composed of a gene A that activates a gene B, which in turn activates gene C (*Upper Left*). An inference method could infer a statistical link between A and C, leading to two false-negative links (two existing links are not present) and one false-positive link (*Upper Right*). However, to predict gene C's expression, given the expression of gene A, this inference method will probably give adequate results. In the second example (*Lower Right*), a better inference method could give six true-positive inferred links and only one false negative, omitting the link between A and B. However, in this case, we have a dramatic situation for prediction purposes because gene A can no longer activate gene B.

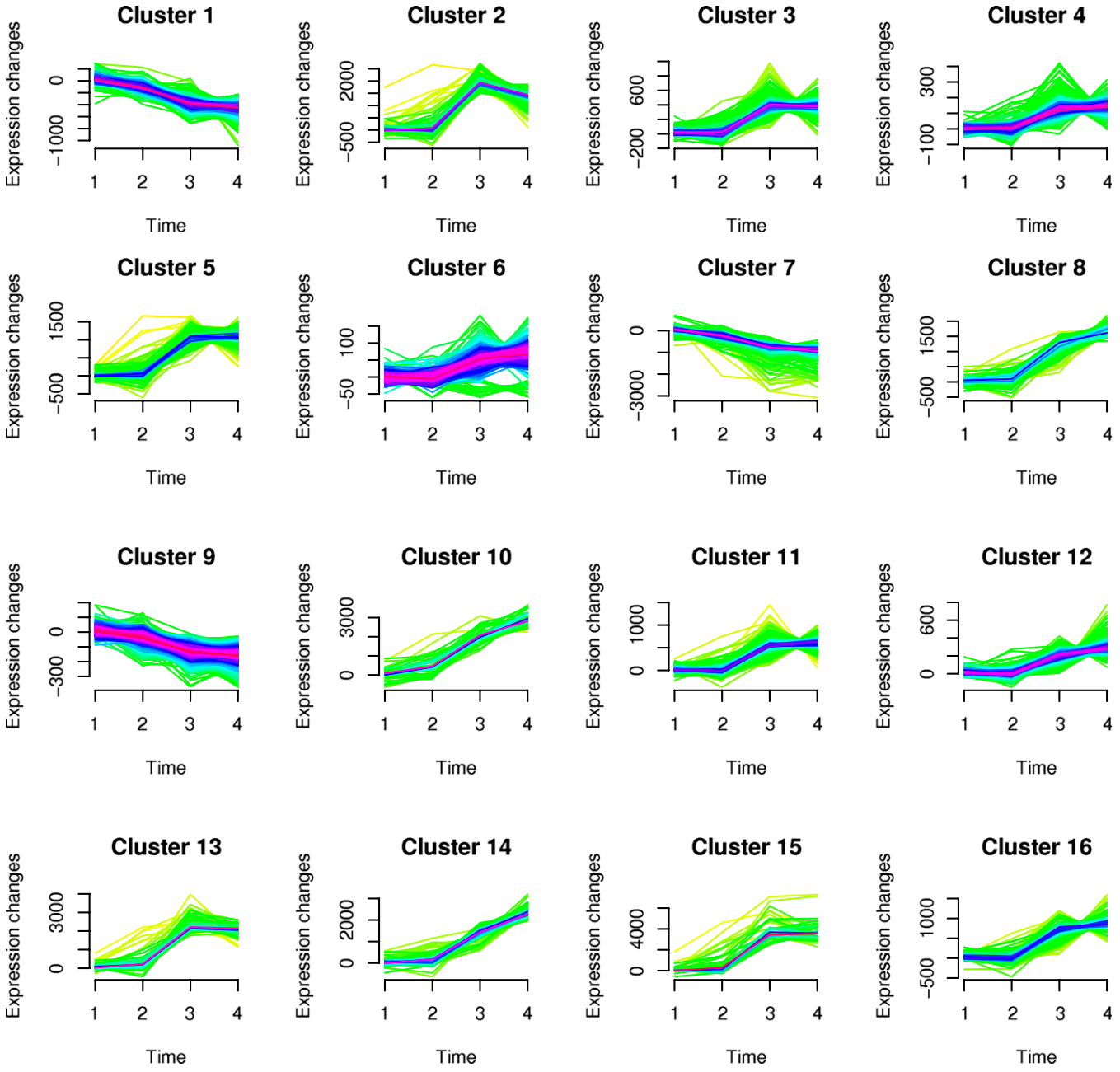

Fig. S6. Significance of selected patterns in the clustering step. To evaluate the relevance of our selected patterns used for enrichment, we compared these patterns with various temporal gene clusters obtained with a gold standard unsupervised clustering method. One of the most widely used clustering methods is fuzzy c-means (1). The preponderant aspect of this algorithm relies on the fuzzy parameter that allows taking into account the inherent noise of transcriptional data (when this parameter increases, more genes are randomly assigned into clusters). For comparison purposes, we focused on the biological data set of patients with more-aggressive CLL, and we first select relevant genes with Limma (2), using a *P* value of 0.01. An unsupervised temporal clustering of the 8,113 genes retained with Limma (2) is then performed showing 16 distinct clusters. Importantly, these clusters emphasize the existence of genes with transient expressions (peaks) at t_1 (cluster 1, 7, 9), t_2 (within cluster 2, 4), t_3 (cluster 2, 3, 4, 11, 13, 15), and t_4 (cluster 5, 6, 8, 10, 12, 14, 16), as shown by our method. The fact that through this unsupervised clustering method we reach patterns similar to those produced by our method confirms the pertinence of our own gene selection process.

1. Cannon RL, Dave JV, Bezdek JC (1986) Efficient implementation of the fuzzy c-means clustering algorithms. *IEEE Trans Pattern Anal Mach Intell* 8(2):248–255.
2. Smyth GK, Michaud J, Scott HS (2005) Use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinformatics* 21(9):2067–2075.

Table S1. Short description of selected methods used for inference methods comparison

Method	Possible prediction	Designed for large networks	Short description
TD-ARACNE (1)	No	No	Time-delay regression combined to ARACNE method (2). ARACNE is an information theory model based on mutual information
GeneNet (3)	No	Yes	Graphical Gaussian model using partial correlation
GeneReg (4)	Yes	No	Regression with spline interpolation to increase the number of time points
Morrissey et al. (5)	Yes	No	Dynamic Bayesian network

1. Zoppoli P, Morganella S, Ceccarelli M (2010) TimeDelay-ARACNE: Reverse engineering of gene networks from time-course data by an information theoretic approach. *BMC Bioinformatics* 11:154.
2. Margolin AA, et al. (2006) ARACNE: An algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* 7(Suppl 1):S7.
3. Schäfer J, Strimmer K (2005) An empirical Bayes approach to inferring large-scale gene association networks. *Bioinformatics* 21(6):754–764.
4. Huang T, et al. (2010) Using GeneReg to construct time delay gene regulatory networks. *BMC Res Notes* 3(1):142.
5. Morrissey ER, Juarez MA, Denby KJ, Burroughs NJ (2011) Inferring the time-invariant topology of a nonlinear sparse gene regulatory network using fully Bayesian spline autoregression. *Biostatistics* 12(4):682–694.

Table S2. Settings of selected methods used for inference methods comparisons

Method	User-fixed parameter	Note
TD-ARACNE (1)	The number *N* of bin in the discretization process	The best-performing value of *N* was retained after sequential tests (*N* value varying from 6 to 20, by 0.5)
GeneNet (2)	None	None
GeneReg (3)	Number of interpolated time points	The ratio (number of interpolated time points)/(number of initial time points) used by the authors has been conserved
Morrissey et al. (4)	None	None

1. Zoppoli P, Morganella S, Ceccarelli M (2010) TimeDelay-ARACNE: Reverse engineering of gene networks from time-course data by an information theoretic approach. *BMC Bioinformatics* 11:154.
2. Schäfer J, Strimmer K (2005) An empirical Bayes approach to inferring large-scale gene association networks. *Bioinformatics* 21(6):754–764.
3. Huang T, et al. (2010) Using GeneReg to construct time delay gene regulatory networks. *BMC Res Notes* 3(1):142.
4. Morrissey ER, Juarez MA, Denby KJ, Burroughs NJ (2011) Inferring the time-invariant topology of a nonlinear sparse gene regulatory network using fully Bayesian spline autoregression. *Biostatistics* 12(4):682–694.

Table S3. Comparison of our method with three other benchmark methods applied to the more-aggressive CLL data set

	Our method	TD-ARACNE	GeneNet	GeneReg
Our method	1,528	28	39	7
TD-ARACNE (1)	28	5,236	87	66
GeneNet (2)	39	87	2,241	86
GeneReg (3)	7	66	86	1,567

Total number of inferred links for each selected method and intersection between the methods, in total common inferred links. The biological data set of patients with more-aggressive CLL type is used.

1. Zoppoli P, Morganella S, Ceccarelli M (2010) TimeDelay-ARACNE: Reverse engineering of gene networks from time-course data by an information theoretic approach. *BMC Bioinformatics* 11:154.
2. Schäfer J, Strimmer K (2005) An empirical Bayes approach to inferring large-scale gene association networks. *Bioinformatics* 21(6):754–764.
3. Huang T, et al. (2010) Using GeneReg to construct time delay gene regulatory networks. *BMC Res Notes* 3(1):142.