

SOFTWARE

Open Access

DREM 2.0: Improved reconstruction of dynamic regulatory networks from time-series expression data

Marcel H Schulz^{1*}, William E Devanny², Anthony Gitter³, Shan Zhong¹, Jason Ernst⁴ and Ziv Bar-Joseph^{1,2*}

Abstract

Background: Modeling dynamic regulatory networks is a major challenge since much of the protein-DNA interaction data available is static. The Dynamic Regulatory Events Miner (DREM) uses a Hidden Markov Model-based approach to integrate this static interaction data with time series gene expression leading to models that can determine when transcription factors (TFs) activate genes and what genes they regulate. DREM has been used successfully in diverse areas of biological research. However, several issues were not addressed by the original version.

Results: DREM 2.0 is a comprehensive software for reconstructing dynamic regulatory networks that supports interactive graphical or batch mode. With version 2.0 a set of new features that are unique in comparison with other softwares are introduced. First, we provide static interaction data for additional species. Second, DREM 2.0 now accepts continuous binding values and we added a new method to utilize TF expression levels when searching for dynamic models. Third, we added support for discriminative motif discovery, which is particularly powerful for species with limited experimental interaction data. Finally, we improved the visualization to support the new features. Combined, these changes improve the ability of DREM 2.0 to accurately recover dynamic regulatory networks and make it much easier to use it for analyzing such networks in several species with varying degrees of interaction information.

Conclusions: DREM 2.0 provides a unique framework for constructing and visualizing dynamic regulatory networks. DREM 2.0 can be downloaded from: www.sb.cs.cmu.edu/drem.

Keywords: Systems biology, Gene regulatory networks, Times series expression data, Dynamic networks, ChIP-chip, ChIP-Seq

Background

Modeling gene regulatory networks (GRNs) is a key challenge when studying development and disease progression. These networks are dynamic with different (overlapping) sets of transcription factors activating genes at different points in time or developmental stages. Reconstructing the dynamics of these networks is a non-trivial task that requires the integration of datasets from different types of genome-wide assays.

Several methods were proposed for reconstructing GRNs (see the following reviews for a general overview: [1-3]). These methods often combine expression and protein-DNA interaction data to recover the underlying networks. However, most methods to date focused on reconstructing *static* networks and the resulting models did not provide any temporal information. In this paper we focus on the reconstruction of dynamic GRNs using time-series expression data. Such data is prevalent for several species, mostly from microarray studies [4,5] and more recently using RNA-Seq methods [6-8].

While several studies measure time series expression data, the available protein-DNA interaction data is almost always static (either from sequence motifs or from ChIP-chip or ChIP-Seq experiments). This creates a major

*Correspondence: maschulz@cs.cmu.edu; zivbj@cs.cmu.edu

¹Ray and Stephanie Lane Center for Computational Biology, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213 USA

²Machine Learning Department, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213 USA

Full list of author information is available at the end of the article

computational challenge when attempting to integrate these dynamic and static datasets.

Several methods were suggested for clustering time series expression data [9-11], or for constructing dynamic networks with regression-based techniques that rely on only the temporal expression data [12]. While these approaches led to some success, as we show in Results, methods that can utilize both the temporal expression data and the static interaction data can improve upon the expression-only methods.

A number of methods have been suggested for addressing these issues, though most of them were targeted at specific input datasets and did not offer any software to support their general use. For example, Luscombe et al. [13] created a dynamic network by overlaying TFs regulating differentially expressed genes for different time points. Lu et al. [14] created a 2D visualization for different dynamic measurements, including time series expression, histone modification, and Pol2-occupancy data using the GATE software [15] although no combined model is presented. Bromberg et al. measure TF activation as a time series and derive pathways that explain activated TFs by integrating subnetworks from PPI networks [16]. Baugh et al. relies on the expression data of transcription factors to identify representatives regulating early development of *C. elegans* embryos [17].

A different way of formulating the problem is to decompose the gene expression data into TF activity and TF affinity values for each expressed gene as suggested by Network Component Analysis [18]. From the matrix of TF affinity values one can construct a dynamic network with connections for each time point [19]. There have been many extensions to this idea with different underlying mathematical models, including ordinary differential equations [20] and Factor analysis [21]. Note however that such regression-based methods do not really take time into account. If one randomly reorders the temporal columns (exchanging, for example the second time point with the fourth etc.) these models will still result in the same network.

One of the first approaches to construct networks that change over time while still incorporating the ordering of time series data was suggested by Friedman [22] using dynamic Bayesian networks (DBNs). A DBN is a set of directed networks, one for each time point. Although general learning of DBNs is NP-hard there exist conditions where these networks can be learned optimally [23,24]. However, these methods do not scale to hundreds of regulators.

To provide a general method that can be widely applied to reconstructing dynamic regulatory networks, [25] presented DREM, a method that integrates times series and static data using an Input-Output Hidden Markov Model (IOHMM). DREM learns a dynamic GRN by identifying

bifurcation points, places in the time series where a group of co-expressed genes begins to diverge. These points are annotated with the TFs controlling the split leading to a combined dynamic model. Since its release 5 years ago the DREM software has been used for modeling a wide range of GRNs for example stress response in yeast [25] and *E. coli* [26], development in fly by the modENCODE consortium [8], stem cell differentiation in mice [27] and disease progression in human [28].

While DREM has been successfully used for multiple species, so far each group using it had to obtain its own protein-DNA interaction data. Since such data is often dispersed among several databases, websites and publications, this step was a major hurdle to using DREM. Other features not supported in the original DREM version included: the integration of motif discovery, the ability to utilize dynamic ChIP binding data [29,30] and TF expression data, and visualization of these new data types. In this paper we discuss a new version of DREM, termed DREM 2.0, that addresses all these limitations. As we show, by addressing these issues DREM 2.0 improves upon both methods that do not integrate static information in the analysis of dynamic data and the previous version of DREM which lacked the above features.

Implementation

DREM 2.0 is implemented entirely in Java and will work with any operating system supporting Java 1.5 or later. Portions of the interface of DREM 2.0 are implemented using third party libraries, the Java Piccolo toolkit from the University of Maryland [31] and the Batik toolkit for svg export of network images [32]. DREM 2.0 also supports batch mode for automated execution. DREM 2.0 makes use of external Gene Ontology (GO) and gene annotation files. DREM 2.0 downloads these files directly from the GO website [33].

Time-specific binding of regulators

The underlying Input-Output Hidden Markov Model learning can now accommodate dynamic input data for each time point in the following way. The transition probabilities for the IOHMM are derived from a logistic regression classifier that uses the protein-DNA interaction data as supervised input and utilizes them to classify genes into diverging paths at a split node in the model. In the new version the nodes in the input layer can be dynamic and thus the function can depend on input from the specific time point it is associated with. See Figure 1 for an illustration.

Results

Using DREM 2.0

Users input their time series expression data by using the graphical user interface (GUI) (see Figure 2). DREM 2.0

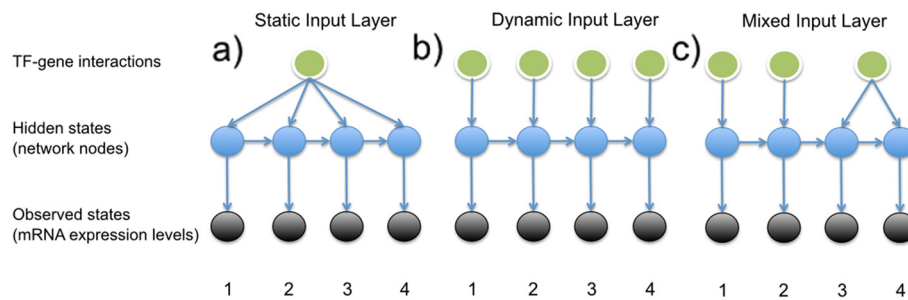


Figure 1 Possible IOHMM topologies in DREM 2.0. The basic topology for a DREM 2.0 IOHMM is shown. The hidden states represent the network nodes (in blue) that we are interested in. The observations (black nodes) are the gene expression ratios which are given to the model, these are dynamic and dependent on the time point. The protein-DNA interaction data (green nodes) are used as supervised *input* data to construct the network. (a) In the original DREM formulation only one static input node is connected to all hidden nodes. In DREM 2.0 the nodes in the input layer can be dynamic and dependent on the time point with a topology either fully dynamic (b) or a mix of static and dynamic input (c).

can transform the data and combine time point repeats. Next, users select a protein-DNA interaction data set for the species they are working with. DREM 2.0 includes protein-DNA interaction data for several species (see Table 1 for a full list). After selecting the species and interactions the user can set various learning parameters or use the default settings (see Additional file 1). Once the data is entered the user selects the ‘execute’ button which runs DREM 2.0 on the input data and results in the dynamic network learned by DREM 2.0 (for example, the one displayed in Figure 3). DREM 2.0 supports downstream analysis using external databases (for example GO as shown in Figure 4) and software (for example, DECOD and STAMP, as shown in Figure 5, see also below).

DREM 2.0 Analysis of asbestos induction

As a running example to illustrate the new features, we used the human protein-DNA data now available with DREM 2.0 to analyze an expression experiment studying the effects of asbestos on human lung adenocarcinoma cells (A549) [39] (Figure 3). Preprocessing and parameters for the analysis are described below. DREM 2.0 successfully predicts enrichment of TFs known to be relevant in asbestos exposure, *e.g.*, TFs from the FOS family [39], that are shown to be up-regulated at the 6 hour time point (blue IDs Figure 3).

Parameters and datasets for the asbestos analysis

The time series data for asbestos treatment of human lung cancer cells [39] was downloaded from GEO (record: GSE6013). The dataset contains gene expression data measured with Affymetrix human gene expression arrays 1, 6, 24, 48 hours, and 7 days after asbestos exposure and a control time series without exposure. The array data was normalized with quantile normalization using RMAExpress (version 1.0.5) with default parameters [40].

\log_2 ratios of exposed versus control were computed as input to DREM 2.0. The human binding predictions

(top 100 threshold, see Additional file 2) were used as the regulatory dataset for DREM 2.0. For the DREM 2.0 analysis the following options were not set to default values: (i) genes in the time course were discarded if “Minimum Absolute Expression Change” was smaller than 0.5, (ii) “incorporate expression in regulator data” was activated for transcription factors with “Expression scaling weight” set to 1. For the annotation of split nodes (Figure 3) the “Path significance conditional on Split” enrichment *p*-value in the GUI was set to be $\leq 5 \cdot 10^{-5}$.

For the motif analysis DECOD [41] version 1.01 was downloaded and connected with DREM 2.0 using the GUI interface. 8512 human promoter sequences (-499,+100 bp relative to transcription start site) were downloaded from the EPD promoter database (from the website: Last update 11 Nov. 2009) [42]. DECOD was run to search for motifs of length 7 with the exact mode and STAMP [43] motif similarity search was conducted against TRANSFAC (version 11.3) using default parameters [44]. The reported motif (below) is the 3rd motif found by DECOD with a similarity *E*-value of 3.93e-12 returned by STAMP.

Supporting additional species

DREM 2.0 utilizes time series expression data (from a specific condition, for example the asbestos data used in this paper) and static interaction data which is often condition-independent (for example, DNA binding motifs). The original version of DREM [25] only provided such static data for *S. cerevisiae*, which meant that users studying other species had to collect their own static data as well as the condition-specific time series data. Over the years we have included protein-DNA interaction data for *E. coli* and human, but several other species were still not supported, limiting DREM’s usage. We have now collected static data for a number of additional species (*M. musculus*, *D. melanogaster*, *A. thaliana*) and have added additional high throughput protein-DNA interaction datasets for human as well. With these additions DREM 2.0 now

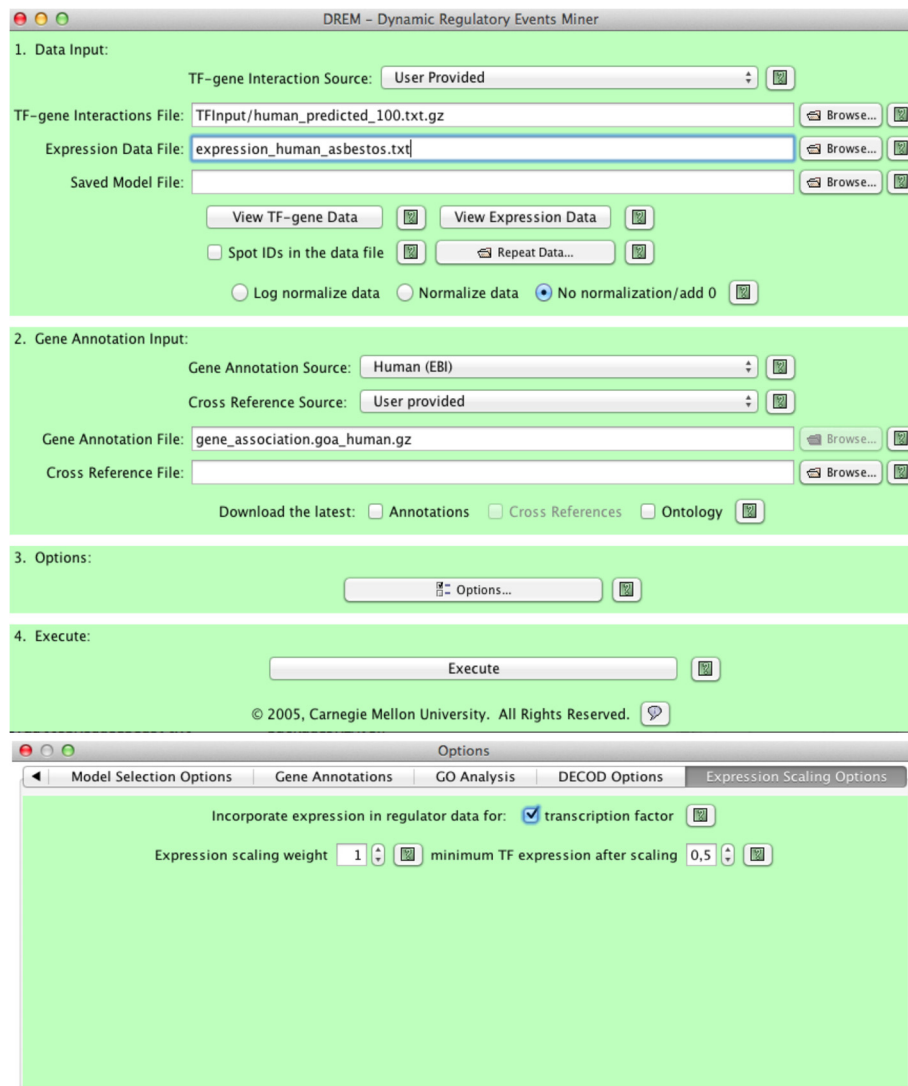
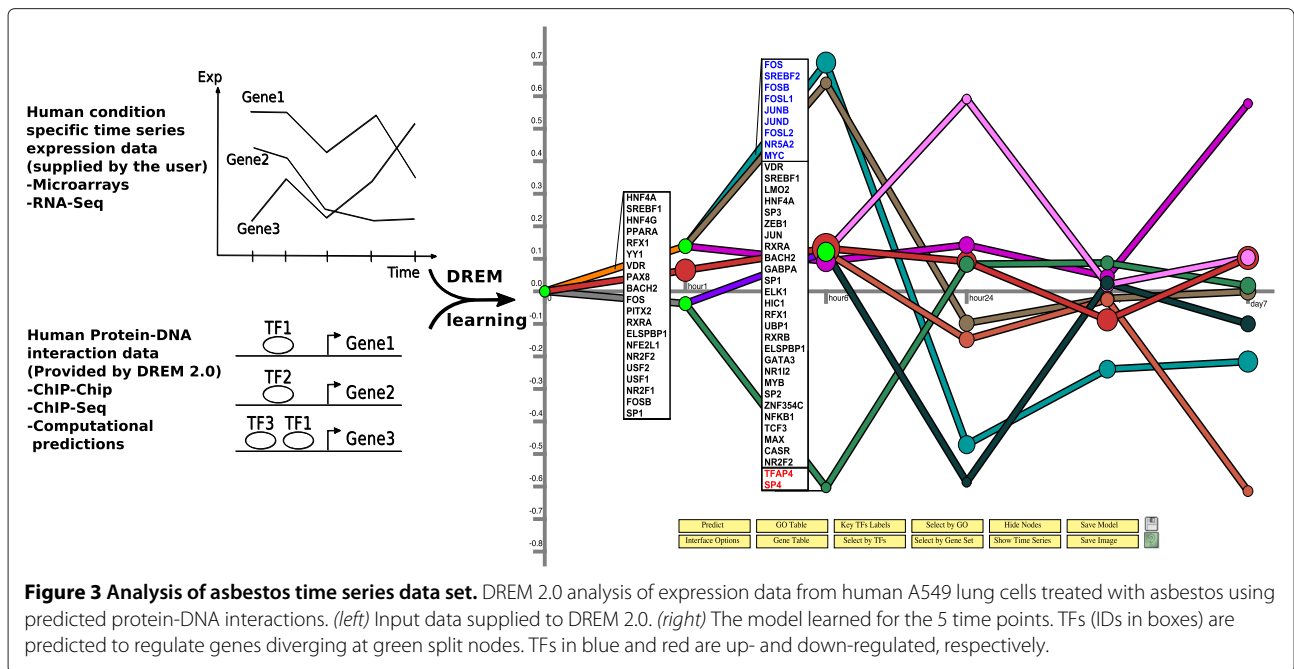


Figure 2 DREM 2.0 input dialog. Top: Input dialog for the DREM 2.0 software. Bottom: Selected tab for the Options window, shows the dialog for the activation of TF expression level scaling, see text for details.

Table 1 Statistics for protein-DNA datasets supplied with DREM 2.0

Species	#TFs	#genes	Protein-DNA interactions	Type	Reference
<i>S. cerevisiae</i>	205	6,230	22,167	ChIP-Chip, conservation	[34,35]
<i>E.coli</i>	124	1,763	3,520	curated + computational	[26]
<i>D. melanogaster</i>	77	12,504	158,558	ChIP-Chip, ChIP-Seq	[8]
<i>M. musculus</i>	336	16,641	468,319	computational prediction, supplement	[36]
<i>H. sapiens</i>	127	19,755	954,377	ChIP-Seq	[37]
<i>H. sapiens</i>	349	17,848	514,925	computational prediction	[36]
<i>A. thaliana</i>	68	8,132	11,354	diverse experimental evidence	[38]

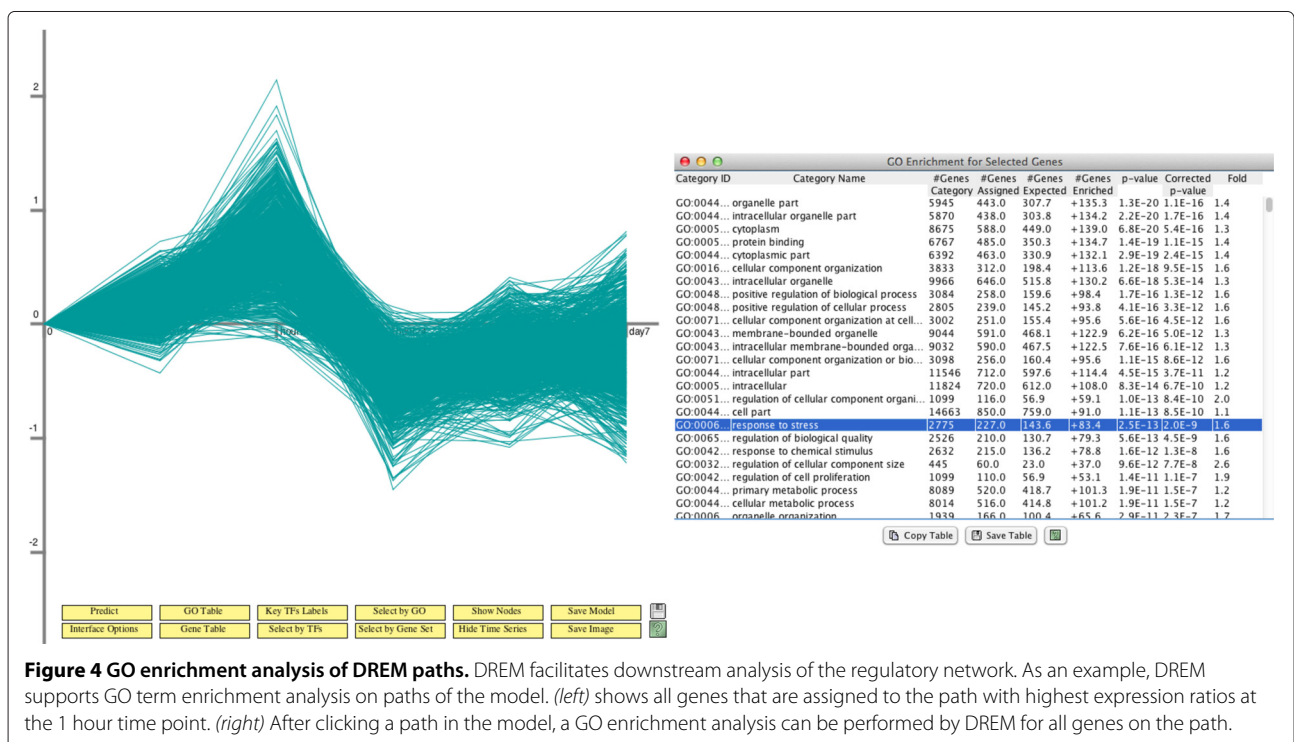
Number of protein-DNA interactions for TFs and target genes for the six supplied species *H. sapiens*, *A. thaliana*, and *M. musculus*, *S. cerevisiae*, and *E.coli* since DREM 2.0. Higher-confidence subsets of these interactions are also provided for some species. More details can be found in Additional file 2.

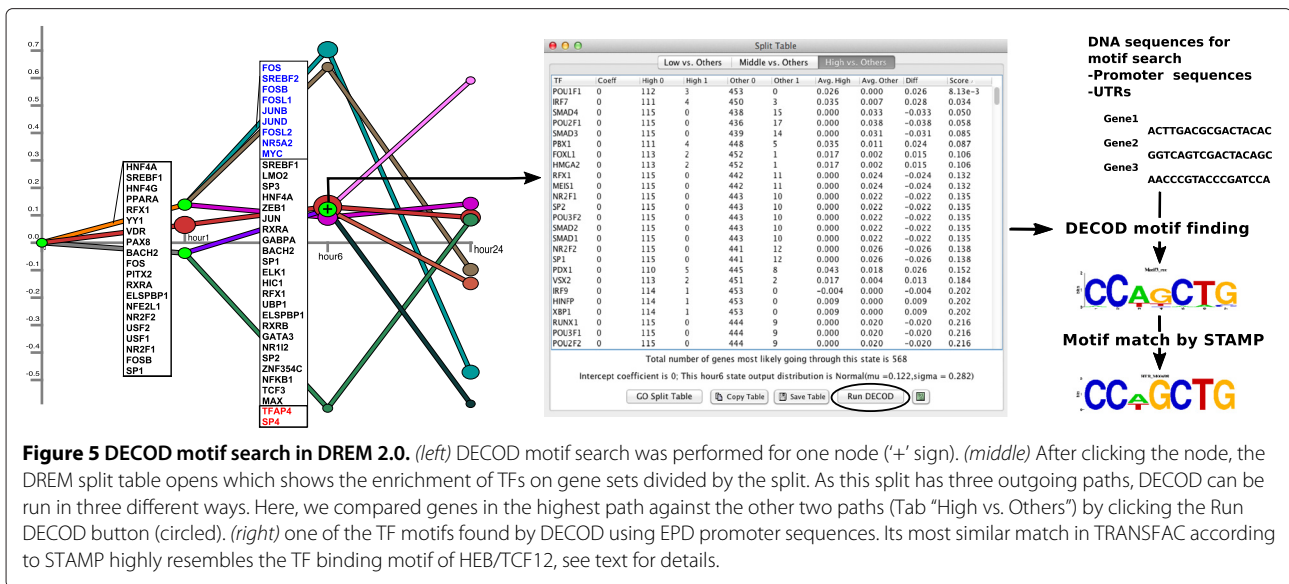


supports most of the well-studied organisms facilitating much wider use of the method. Table 1 lists the current species supported, the number of interactions we have for each species and where these interactions were obtained. More details regarding these datasets can be found in Additional file 2.

Utilizing the expression levels of TFs

The original version of DREM did not use any information regarding the expression levels of the TFs predicted to regulate split nodes. The underlying reason for this was the fact that many TFs are post-transcriptionally regulated and relying on their expression to determine activity





may lead to missing important TFs. In the new version, we still maintain the ability to identify TFs that are only post-transcriptionally regulated. However, we have added a new computational module that allows the method to utilize expression information for those TFs that are transcriptionally regulated. For each TE, its binding prior is elevated based on the TF's expression level using a logistic function. Thus, active TFs have a stronger prior of being selected as regulators by DREM 2.0 (see Additional file 2). We have also changed the visualization in DREM 2.0 to highlight such factors. In Figure 3, which is a screenshot from DREM 2.0, active TFs are highlighted in blue and repressed TFs in red.

Finding DNA motifs at split nodes with DECOD

During learning DREM assigns genes to paths in the network model and uses split nodes (light green nodes in Figure 3) to represent sets of genes that change their expression between consecutive time points. TFs are assigned to split nodes allowing DREM to infer their time of activation. When the protein-DNA interaction data is unable to explain some of the split nodes (i.e. no TF is assigned to that split), it could mean that the interaction data is incomplete. To still allow the identification of such TFs, we integrated with DREM 2.0 the discriminative motif finder DECOD [41]. The user can search for discriminative DNA motifs between DNA, e.g. promoter, sequences of genes assigned to diverging paths emerging out of any split node. The method uses two sets (genes going up and down from the split) to discriminatively search for motifs. The predicted DNA motifs can be matched against known motif databases using STAMP [43]. To highlight the utility of this new feature in DREM 2.0 we used it on the asbestos data described above. As can

be seen, not all split nodes had been assigned in Figure 3. We have thus used the new DECOD feature to identify TFs for one of these splits ('+' sign in Figure 5). A database motif search with STAMP reveals a motif with significant similarity to HEB/TCF12. TCF12 was indeed missing among significant TFs in the split table (Figure 5, middle), perhaps because of incomplete data. However, a DNA inversion close to the *TCF12* gene was recently found in lung cancer patients [45] indicating that this protein may be playing a role in regulating gene response in the lung.

In order to test the ability of DECOD to recover TF binding motifs at DREM split nodes for the case where no TF-gene interaction data is available, we have conducted the following analysis. A DREM model using the asbestos expression data was built without using the TF-gene interaction data. Then, EPD promoter sequences for genes at the 6 hour split node where used for motif search with DECOD. We searched for motifs of length 6-8 and selected all those with significant matches in TRANSFAC (using the STAMP motif comparison tool). After grouping TFs from the same family, 10 of the 24 TFs identified in the original run of DREM for this split were found in the DECOD derived set (see Additional file 2 for details).

Supporting continuous and dynamic binding data

The original version of DREM only supported three binding states (activator/ repressor/ no regulation) interaction data. DREM 2.0 now supports continuous binding values. These can be derived from *p*-values of ChIP-Seq calling procedures or from computational affinity predictions [46]. Thus, in the new version the same regulator may have a different binding value for each gene. The classifier weighs a target with a large binding value higher than targets with a lower binding value. A plausible way to turn

ChIP binding p -values into DREM 2.0 binding values is to set $b = -\log p$ -value. These continuous binding values can then be passed to DREM 2.0.

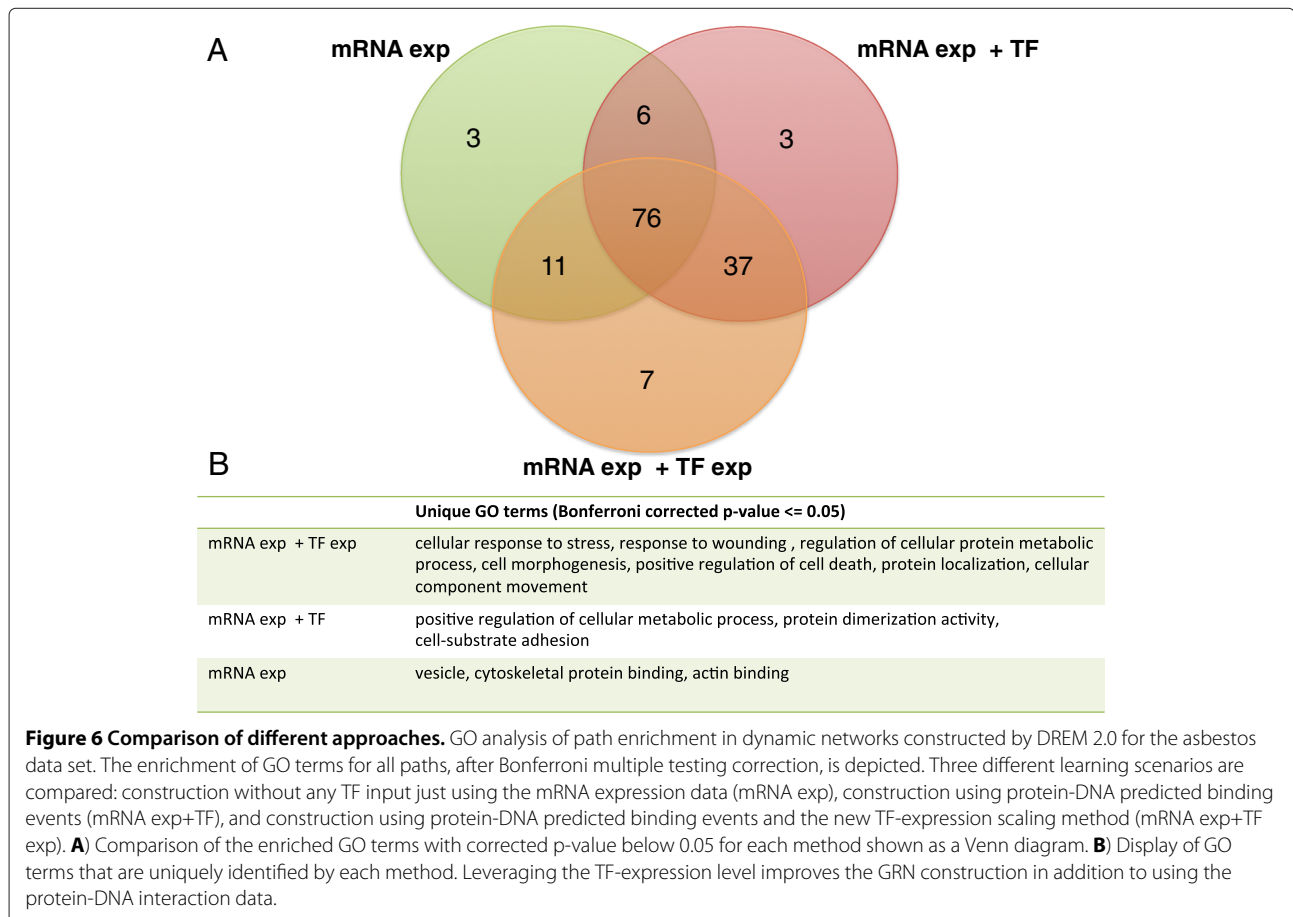
In addition, DREM 2.0 also supports temporal binding data. While most interaction data is still static, dynamic binding data is becoming available. Recent studies have shown that TFs may alter their binding behavior depending on the time point [29,30] necessitating methods that can utilize such information when available. In its original implementation DREM could only use static protein-DNA interaction data when learning logistic regression classifiers for the transition probabilities in the IOHMM. We have now revised this allowing the learning algorithm to support dynamically changing protein-DNA interaction data (see Implementation). For each time point an independent data set can be passed to the logistic regression classifier. Since dynamic binding data is often only available for a (small) subset of TFs, DREM 2.0 supports a joint static-dynamic input format for protein-DNA interactions.

The ability to incorporate temporal binding data allows DREM to reduce false positive assignments by only assigning TFs that are active at that time point (based

on the time points binding data). This in turn can both help identify co-regulators for which only computational predictions exists and also lead to the identification of different waves of transcriptional regulation, where the same TFs activate different sets of genes at different time points.

Comparison to previous methods

We used the asbestos data to compare some of the new features in DREM 2.0 to other methods and to the previous version of DREM. First, to compare DREM 2.0 to methods that only use one type of data (clustering the expression data) we ran DREM 2.0 without using the static protein-DNA interaction information. This is similar to several clustering methods that have been suggested for time series data [9,10]. To compare to the original version of DREM we also reran the asbestos data using TF-DNA interaction data but without using the TF expression information. As a performance metric we used the number of enriched GO terms, a common comparison strategy [11,47]. In Figure 6 the significant GO terms after multiple testing correction are compared for the three methods. Leveraging the TF-expression leads to the highest number of significant GO terms (Figure 6A) and the identification



of additional relevant functions that are not identified by the other two variants, including the GO terms *cellular response to stress* and *positive regulation of cell death* (Figure 6B).

Discussion and conclusions

While several methods can be used to reconstruct GRNs using time series expression data, most such methods either rely only on the expression data itself or result in static networks that do not consider the ordering of the time points. DREM provides not only an alternative to these methods but also a rich GUI and as such, has been used by several groups in multiple species.

Although here we used both treatment and control time series, DREM can also be used with only the treatment time series by taking the log fold change w.r.t. time point 0, see [25] for an example.

The new version eases the application to several species by directly supplying protein-DNA interaction data and incorporating de-novo discriminative motif discovery. In addition we have made other improvements including the ability to utilize and view the expression levels of the TFs and to use dynamic protein-DNA interaction data. Combined, we believe that these improvements will make DREM 2.0 a more widely used software package for the reconstruction of dynamic GRNs.

Availability and requirements

- **Project name:** DREM
- **Project homepage:** www.sb.cs.cmu.edu/drem
- **Operating system(s):** Platform independent
- **Other requirements:** Java 1.5 or higher
- **License:** Free to academics/non-profit
- **Any restrictions to use by non-academics:** License needed

Additional files

Additional file 1: DREM 2.0 Manual. The Manual for using the DREM 2.0 software with details of all parameters and the different dialogs in the GUI.

Additional file 2: Supplementary Methods. Additional description for DREM 2.0 for TF expression level scaling, data collection for the protein-DNA binding data sets and the analysis with DECOD on an unannotated split node.

Abbreviations

DREM, Dynamic Regulatory Events Miner; TF, Transcription factor; GRN, Gene regulatory network; DBN, Dynamic Bayesian network; ChIP, Chromatin immunoprecipitation; IOHMM, Input-output hidden Markov model; GUI, Graphical user interface; GO, Gene Ontology; MGD, Mouse Genome Database; HGNC, HUGO Gene Nomenclature Committee; RNA-Seq, Next generation sequencing of messenger RNAs.

Competing interests

The authors declare that they have no competing interests.

Author's contributions

MHS, WED, AG, SZ designed and implemented the new version. MHS, AG, SZ, JE performed the data collection and analysis. ZBJ supervised the work. MHS and ZBJ wrote the manuscript. All authors read and approved the final manuscript.

Authors' information

Marcel H. Schulz and William E. Devanny joint first authors.

Funding

Work supported in part by NIH grant 1RO1 GM085022.

Acknowledgements

We would like to acknowledge all groups that have contributed and made available the human ChIP-Seq predictions for human as part of the ENCODE project.

Author details

¹Ray and Stephanie Lane Center for Computational Biology, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213 USA. ²Machine Learning Department, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213 USA. ³Computer Science Department, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213 USA. ⁴Department of Biological Chemistry, University of California Los Angeles, Los Angeles, CA, 90095 USA.

Received: 28 February 2012 Accepted: 18 July 2012

Published: 16 August 2012

References

1. Friedman N: **Inferring cellular networks using probabilistic graphical models.** *Science (New York, N.Y.)* 2004, **303**(5659):799–805.
2. Markowitz F, Spang R: **Inferring cellular networks—a review.** *BMC Bioinf* 2007, **8**(Suppl 6):S5.
3. Lee WP, Tzou WS: **Computational methods for discovering gene networks from expression data.** *Briefings Bioinf* 2009, **10**(4):408–423.
4. Schena M, Shalon D, Davis RW, Brown PO: **Quantitative monitoring of gene expression patterns with a complementary DNA microarray.** *Science (New York, N.Y.)* 1995, **270**(5235):467–470.
5. Gasch AP, Spellman PT, Kao CM, Carmel-Harel O, Eisen MB, Storz G, Botstein D, Brown PO: **Genomic expression programs in the response of yeast cells to environmental changes.** *Mol Biol cell* 2000, **11**(12):4241–4257.
6. Sultan M, Schulz MH, Richard H, Magen A, Klingenhoff A, Scherf M, Seifert M, Borodina T, Soldatov A, Parkhomchuk D, Schmidt D, O'Keefe S, Haas S, Vingron M, Lehrach H, Yaspo ML: **A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome.** *Science (New York, N.Y.)* 2008, **321**(5891):956–960.
7. Wang Z, Gerstein M, Snyder M: **RNA-Seq: a revolutionary tool for transcriptomics.** *Nature Rev. Genet* 2009, **10**:57–63.
8. modENCODE Consortium, Roy S, Ernst J, Kharchenko PV, Kheradpour P, Negre N, Eaton ML, Landolin JM, Bristow CA, Ma L, Lin MF, Washietl S, Arshinoff BI, Ay F, Meyer PE, Robine N, Washington NL, Di Stefano L, Berezhikov E, Brown CD, Candeias R, Carlson JW, Carr A, Jungreis I, Marbach D, Sealfon R, Tolstorukov MY, Will S, Alekseyenko AA, Artieri C, et al.: **Identification of functional elements and regulatory circuits by *Drosophila* modENCODE.** *Science (New York, N.Y.)* 2010, **330**(6012):1787–1797.
9. Ernst J, Nau GJ, Bar-Joseph Z: **Clustering short time series gene expression data.** *Bioinformatics (Oxford, England)* 2005, **21**(Suppl 1):i159–68.
10. Schliep A, Costa IG, Steinhoff C, Schönhuth A: **Analyzing gene expression time-courses.** *IEEE/ACM Trans Comput Biol Bioinf / IEEE, ACM* 2005, **2**(3):179–193.
11. Costa IG, Roepcke S, Hafemeister C, Schliep A: **Inferring differentiation pathways from gene expression.** *Bioinformatics (Oxford, England)* 2008, **24**(13):i156–i164.
12. Song L, Kolar M, Xing EP: **KELLER: estimating time-varying interactions between genes.** *Bioinformatics (Oxford, England)* 2009, **25**(12):i128–36.
13. Luscombe NM, Babu MM, Yu H, Snyder M, Teichmann SA, Gerstein M: **Genomic analysis of regulatory network dynamics reveals large topological changes.** *Nature* 2004, **431**(7006):308–312.

14. Lu R, Markowitz F, Unwin RD, Leek JT, Airoidi EM, MacArthur BD, Lachmann A, Rozov R, Ma'ayan A, Boyer LA, Troyanskaya OG, Whetton AD, Lemischka IR: **Systems-level dynamic analyses of fate change in murine embryonic stem cells.** *Nature* 2009, **462**(7271):358–362.
15. MacArthur BD, Lachmann A, Lemischka IR, Ma'ayan A: **GATE: software for the analysis and visualization of high-dimensional time series expression data.** *Bioinformatics (Oxford, England)* 2010, **26**:143–144.
16. Bromberg KD, Ma'ayan A, Neves SR, Iyengar R: **Design logic of a cannabinoid receptor signaling network that triggers neurite outgrowth.** *Science (New York, N.Y.)* 2008, **320**(5878):903–909.
17. Baugh LR, Hill AA, Claggett JM, Hill-Harfe K, Wen JC, Slonim DK, Brown EL, Hunter CP: **The homeodomain protein PAL-1 specifies a lineage-specific regulatory network in the *C. elegans* embryo.** *Development (Cambridge, England)* 2005, **132**(8):1843–1854.
18. Liao JC, Boscolo R, Yang YL, Tran LM, Sabatti C, Roychowdhury VP: **Network component analysis: reconstruction of regulatory signals in biological systems.** *Proc Natl Acad Sci USA* 2003, **100**(26):15522–15527.
19. Seok J, Xiao W, Moldawer LL, Davis RW, Covert MW: **A dynamic network of transcription in LPS-treated human subjects.** *BMC Syst Biol* 2009, **3**:78.
20. Bansal M, Della Gatta G, di Bernardo D: **Inference of gene regulatory networks and compound mode of action from time course gene expression profiles.** *Bioinformatics (Oxford, England)* 2006, **22**(7):815–822.
21. Pournara I, Wernisch L: **Factor analysis for gene regulatory networks and transcription factor activity profiles.** *BMC Bioinf* 2007, **8**:61.
22. Friedman N, Murphy K: **Learning the structure of dynamic probabilistic networks.** In *UAI'98 Proceedings of the Fourteenth conference on Uncertainty in Artificial Intelligence*. San Francisco: Morgan Kaufmann Publishers Inc.; 1998:139–147.
23. Wilczyński B, Dojer N: **BNFinder: exact and efficient method for learning Bayesian networks.** *Bioinformatics (Oxford, England)* 2009, **25**(2):286–287.
24. Vinh N, Chetty M, Coppel R: **GlobalMIT: learning globally optimal dynamic bayesian network with the mutual information test criterion.** *Bioinformatics (Oxford, England)* 2011, **27**:2765–2766.
25. Ernst J, Vainas O, Harbison CT, Simon I, Bar-Joseph Z: **Reconstructing dynamic regulatory maps.** *Mol Syst Biol* 2007, **3**:74.
26. Ernst J, Beg QK, Kay KA, Balázsi G, Oltvai ZN, Bar-Joseph Z: **A semi-supervised method for predicting transcription factor-gene interactions in *Escherichia coli*.** *PLoS Comput Biol* 2008, **4**(3):e1000044.
27. Mendoza-Parra MA, Walia M, Sankar M, Gronemeyer H: **Dissecting the retinoid-induced differentiation of F9 embryonic stem cells by integrative genomics.** *Molecular Systems Biol* 2011, **7**:538.
28. Gu F, Hsu PY, Wu J, Ma Y, Parvin J, Huang THM, Jin VX: **Inference of hierarchical regulatory network of estrogen-dependent breast cancer through CHIP-based data.** *BMC Syst Biol* 2010, **4**:170.
29. Ni L, Bruce C, Hart C, Leigh-Bell J, Gelperin D, Umansky L, Gerstein MB, Snyder M: **Dynamic and complex transcription factor binding during an inducible response in yeast.** *Genes & Dev* 2009, **23**(11):1351–1363.
30. Wilczyński B, Furlong EEM: **Dynamic CRM occupancy reflects a temporal map of developmental progression.** *Mol Syst Biol* 2010, **6**:383.
31. Bederson B, Grosjean J, Meyer J: **Toolkit design for interactive structured graphics.** *Software Eng, IEEE Trans*, **30**(8):535–546.
32. The Apache XML Graphics Project: **Batik SVG Toolkit.** [http://xmlgraphics.apache.org/batik/]
33. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G: **Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.** *Nat Gen* 2000, **25**:25–29.
34. Harbison CT, Gordon DB, Lee TI, Rinaldi NJ, Macisaac KD, Danford TW, Hannett NM, Tagne JB, Reynolds DB, Yoo J, Jennings EG, Zeitlinger J, Pokholok DK, Kellis M, Rolfe PA, Takusagawa KT, Lander ES, Gifford DK, Fraenkel E, Young RA: **Transcriptional regulatory code of a eukaryotic genome.** *Nature* 2004, **431**(7004):99–104.
35. Macisaac KD, Wang T, Gordon DB, Gifford DK, Stormo GD, Fraenkel E: **An improved map of conserved regulatory sites for *Saccharomyces cerevisiae*.** *BMC Bioinf* 2006, **7**:113.
36. Ernst J, Plasterer HL, Simon I, Bar-Joseph Z: **Integrating multiple evidence sources to predict transcription factor binding in the human genome.** *Genome Res* 2010, **20**(4):526–536.
37. ENCODE Project Consortium, Birney E, Stamatoyannopoulos JA, Dutta A, Guigó R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET, Thurman RE, Kuehn MS, Taylor CM, Neph S, Koch CM, Asthana S, Malhotra A, Adzhubei I, Greenbaum JA, Andrews RM, Flicek P, Boyle PJ, Cao H, Carter NP, Clelland GK, Davis S, Day N, Dharmi P, Dillon SC, Dorschner MO, et al.: **Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project.** *Nature* 2007, **447**(7146):799–816.
38. Palaniswamy SK, James S, Sun H, Lamb RS, Davuluri RV, Grotewold E: **AGRIS and AtRegNet. a platform to link cis-regulatory elements and transcription factors into regulatory networks.** *Plant Physiol* 2006, **140**(3):818–829.
39. Nymark P, Lindholm PM, Korpela MV, Lahti L, Ruosaari S, Kaski S, Hollmén J, Anttila S, Kinnula VL, Knuutila S: **Gene expression profiles in asbestos-exposed epithelial and mesothelial lung cell lines.** *BMC Genomics* 2007, **8**:62.
40. Bolstad BM, Irizarry RA, Astrand M, Speed TP: **A comparison of normalization methods for high density oligonucleotide array data based on variance and bias.** *Bioinformatics (Oxford, England)* 2003, **19**(2):185–193.
41. Huggins P, Zhong S, Shiff I, Beckerman R, Laptenko O, Prives C, Schulz MH, Simon I, Bar-Joseph Z: **DECOD: fast and accurate discriminative DNA motif finding.** *Bioinformatics (Oxford, England)* 2011, **27**(17):2361–2367.
42. Schmid CD, Perier R, Praz V, Bucher P: **EPD in its twentieth year: towards complete promoter coverage of selected model organisms.** *Nucleic Acids Res* 2006, **34**(Database issue):D82–D85.
43. Mahony S, Benos PV: **STAMP: a web tool for exploring DNA-binding motif similarities.** *Nucleic Acids Res* 2007, **35**(Web Server issue):W253–W258.
44. Matys V, Kel-Margoulis OV, Fricke E, Liebich I, Land S, Barre-Dirrie A, Reuter I, Chekmenev D, Krull M, Hornischer K, Voss N, Stegmaier P, Lewicki-Potapov B, Saxel H, Kel AE, Wingender E: **TRANSFAC and its module TRANSCOMP: transcriptional gene regulation in eukaryotes.** *Nucleic Acids Res* 2006, **34**(Database issue):D108–D110.
45. Lee W, Jiang Z, Liu J, Haverty PM, Guan Y, Stinson J, Yue P, Zhang Y, Pant KP, Bhatt D, Ha C, Johnson S, Kennemer MI, Mohan S, Nazarenko I, Watanabe C, Sparks AB, Shames DS, Gentleman R, de Sauvage FJ, Stern H, Pandita A, Ballinger DG, Drmanac R, Modrusan Z, Seshagiri S, Zhang Z: **The mutation spectrum revealed by paired genome sequences from a lung cancer patient.** *Nature* 2010, **465**(7297):473–477.
46. Roider HG, Kanhere A, Manke T, Vingron M: **Predicting transcription factor affinities to DNA from a biophysical model.** *Bioinformatics (Oxford, England)* 2007, **23**(2):134–141.
47. Kuo D, Tan K, Zinman G, Ravasi T, Bar-Joseph Z, Ideker T: **Evolutionary divergence in the fungal response to fluconazole revealed by soft clustering.** *Genome Biol* 2010, **11**(7):R77.

doi:10.1186/1752-0509-6-104

Cite this article as: Schulz et al.: DREM 2.0: Improved reconstruction of dynamic regulatory networks from time-series expression data. *BMC Systems Biology* 2012 **6**:104.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

