

RESEARCH ARTICLE

iDREM: Interactive visualization of dynamic regulatory networks

Jun Ding¹, James S. Hagood², Namasivayam Ambalavanan³, Naftali Kaminski⁴, Ziv Bar-Joseph^{1*}

1 Computational Biology Department, School of Computer Science, Carnegie Mellon University, Pittsburgh, Pennsylvania, United States of America, **2** Division of Respiratory Medicine, Department of Pediatrics, University of California, San Diego and Rady Children's Hospital, La Jolla, California, United States of America, **3** Division of Neonatology, Department of Pediatrics, University of Alabama at Birmingham, Birmingham, Alabama, United States of America, **4** Section of Pulmonary, Critical Care and Sleep Medicine, Yale School of Medicine, New Haven, Connecticut, United States of America

* zivbj@cs.cmu.edu



OPEN ACCESS

Citation: Ding J, Hagood JS, Ambalavanan N, Kaminski N, Bar-Joseph Z (2018) iDREM: Interactive visualization of dynamic regulatory networks. *PLoS Comput Biol* 14(3): e1006019. <https://doi.org/10.1371/journal.pcbi.1006019>

Editor: Dina Schneidman, Hebrew University of Jerusalem, ISRAEL

Received: November 15, 2017

Accepted: February 4, 2018

Published: March 14, 2018

Copyright: © 2018 Ding et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The software can be accessed at the Github repository: <https://github.com/phoenixding/idrem> including a detailed instruction and an example.

Funding: This work is supported by: (1) National Institutes of Health, U01HL122626, https://projectreporter.nih.gov/project_info_description.cfm?aid=9268658&icde=36922853; (2) National Institutes of Health, 1R01GM122096, https://projectreporter.nih.gov/project_info_description.cfm?aid=9350447&map=y; (3) National Science, DBI-1356505, Foundation <https://www.nsf.gov/>

Abstract

The Dynamic Regulatory Events Miner (DREM) software reconstructs dynamic regulatory networks by integrating static protein-DNA interaction data with time series gene expression data. In recent years, several additional types of high-throughput time series data have been profiled when studying biological processes including time series miRNA expression, proteomics, epigenomics and single cell RNA-Seq. Combining all available time series and static datasets in a unified model remains an important challenge and goal. To address this challenge we have developed a new version of DREM termed interactive DREM (iDREM). iDREM provides support for all data types mentioned above and combines them with existing interaction data to reconstruct networks that can lead to novel hypotheses on the function and timing of regulators. Users can interactively visualize and query the resulting model. We showcase the functionality of the new tool by applying it to microglia developmental data from multiple labs.

This is a PLoS Computational Biology Software paper.

Introduction

The analysis and modeling of dynamic regulatory networks remains a major goal of systems biology. Several methods for the analysis of such networks using a wide range of high throughput biological datasets have been developed over the last 15 years. Initial methods have mainly focused on using time series microarray data [1–3], though over the years these methods were extended by utilizing several other types of high throughput temporal and static data. Examples include methods that combine time series RNA-Seq and ChIP-Seq data [4, 5], methods for the analysis of epigenetic dynamics [6], microRNA regulation over time [7–9], time series proteomics [10–12] and, most recently, single cell RNA-Seq (scRNA-Seq) data [13, 14].

[awardsearch/showAward?AWD_ID=1356505](#); and (4) the Pennsylvania Department of Health, Grant 4100070287. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

While each of the above data types has been studied and modeled on its own, relatively few methods have been developed to integrate multiple time series data types and we are not aware of any current method that can integrate all of them in a comprehensive analysis and visualization framework. In 2007, we presented the Dynamic Regulatory Events Miner (DREM) that was developed to integrate time series gene expression and static protein-DNA interaction data [15]. DREM learns an Input Output Hidden Markov Model (IOHMM) which attempts to identify bifurcation points—time points in which a set of genes that are co-expressed up to that point start to diverge. These points are then annotated by the transcription factors (TFs) that are predicted to regulate these genes allowing the method to assign dynamics to the (often static) protein-DNA interaction data. Over the years we have extended DREM so that it can utilize time series miRNA data [16], static ChIP-Seq data [17] and static protein-interaction data [18]. DREM has been widely used, by us and others, to model regulatory networks in a wide range of conditions and species [19–21].

While useful, DREM and its extensions are still unable to utilize several recent high throughput time series data types. These include epigenetic data (methylation, histone modification etc.), time series proteomics datasets and time series scRNA-Seq data. While past studies have usually profiled only one of these data types, more recent work often profiles multiple data types over time [22] which necessitates methods that can combine all of these in a single analysis and visualization framework. In addition, the current DREM output is a dynamic network figure (Fig 1) which does not allow for interactive analysis of the resulting model. To address these issues we developed the interactive DREM (iDREM) tool that provides support for more data types and greatly improves the visualization allowing users to interactively query the reconstructed network. We also allow users to project scRNA-Seq data on the resulting model helping highlight the relationships between different cell types and the trajectories observed in bulk expression analysis (S3 Fig).

Design and implementation

In previous DREM versions ([15, 17, 18]), we discussed the integration of time-series mRNA expression, time-series miRNA expression and static TF-gene and protein-protein interaction data. Here we focus on the new capabilities of iDREM including the ability to utilize time-series proteomics, epigenomics, and scRNA-Seq data and the interactive visualization options.

Incorporating time series proteomics data

We use the proteomics data to improve our ability to detect the *time* of TF activation. In previous versions of DREM we used a static, prior regulatory interaction matrix (inferred from previous experiments not necessarily related to the condition being studied). To obtain a dynamic version of such matrix we do the following. First, if a TF protein is highly expressed at a specific time point we increase the prior on its activity for that time point. Second, to account for post-translational modifications which are not always reflected by the protein levels we also use protein interaction information. Specifically, for each TF we look at the average expression of its known interaction partners at each time point. If the levels of proteins that interact with the TF are increased (decreased) we increase (decrease) the prior on that TF for that time point by adjusting the values in the prior regulation matrix for that TF. See S1 Text for complete details. The interactive visualization (S1 and S2 Figs) further supports exploration of the proteomics data and its impact. Users can view the protein levels of the specific genes and TFs. To determine the impact of the proteomics data, users can run iDREM with and without this data and directly compare the resulting models.

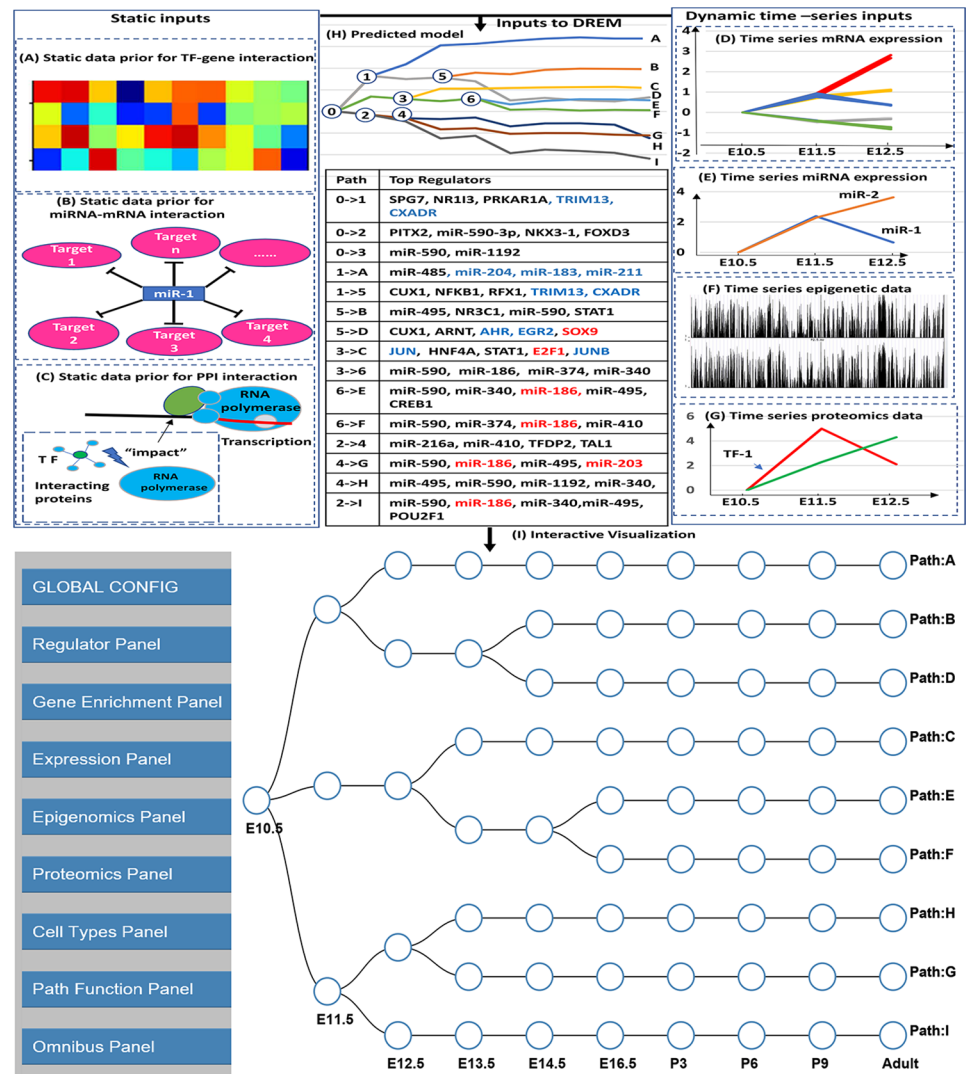


Fig 1. DREM and iDREM flowchart. **Top:** Data types integrated to learn the DREM model include general, static interaction data (A) Transcription factor (TF)-gene interaction; (B) miRNA-mRNA interaction; (C) protein-protein interaction (PPI) and condition specific time series data (right): (D) mRNA expression; (E) miRNA expression; (F) Epigenetic data; (G) Proteomics data. The resulting model (H) provides a summary of different gene groups in the experiment, their expression level, their temporal profiles and the regulators (TFs and miRNAs) that control different bifurcation events. **Bottom I:** The iDREM representation of the learned DREM model above. Note that this representation removes the actual levels and only provide a schematic view for the paths and splits in the model. The actual expression levels and several other aspects of the model and the data can be interactively viewed by using the various panels available (left).

<https://doi.org/10.1371/journal.pcbi.1006019.g001>

Utilizing time series epigenomics data

iDREM adds support for dynamic epigenetic data. Here we discuss time series histone methylation (H3K4me2) data, though iDREM supports other types of epigenetic data as well (S1 Text). Epigenetic data is used to further improve our ability to assign temporal activity to TFs. Specifically, depending on the type of time series data that the user provides, iDREM either increases or decreases the prior on the likelihood of binding of a specific TF to each of its targets. For example, H3K4me2 methylation is associated with “activation” [23], and thus we use it to increase the likelihood of binding in cases where a TF binding site is methylated for a

specific target at a specific time point. See [S1 Text](#) for details on how the epigenetic data is used and integrated into the IOHMM learning process. Additionally, iDREM provides a number of options for visualizing epigenetic data and its relationship with other data types. For genes, users can plot the temporal profiles of their promoters and explore the overall impact of the epigenetic data on targets of specific TFs/ miRNAs. Users can also explore the difference in epigenetic scores between two time points and can view the data directly on the UCSC genome browser [24] ([Fig 2\(G\)](#)).

scRNA-Seq and sorted cell data

A new and exciting type of high-throughput time series data is available from experiments that profile the expression in single cells (e.g. scRNA-Seq) [13]. Other studies have profiled different types of homogeneous cells over time [25, 26] (often termed sorted cells). To enable the integration of single and sorted cell data with bulk studies iDREM allows users to superimpose cell type studies on the reconstructed models. This is performed using the “Cell Types” panel which allows users to upload single cell data (for specific time points) and then intersects the top differentially expressed (DE) genes in these datasets with genes assigned to nodes that represent the same time points in the iDREM model. This enables users to determine the cell type composition of the different nodes and paths and to infer whether specific changes observed are related to activation of TFs in existing cells or the formation of new cell types.

Interactive visualization of the reconstructed model

In addition to visualizing the new data types discussed above, several additional panels are provided for users to explore the reconstructed model, trajectories and interactions of specific TFs, genes and miRNAs. The panels are shown in [S1 Fig](#). They include the “Global Config panel” which provides general functions for the appearance of the schematic network. The “Expression panel” allows users to interactively look at the expression of specific genes, sets of genes and miRNAs ([Fig 2\(A\)](#)) and determine the path they were assigned to. The “Regulator panel” allows users to determine regulators for specific splits ([Fig 1](#)) and paths. It can also be used to determine all paths controlled by a specific TF or miRNA. Users can change the setting to only select those paths for which the regulator is one of the top X regulators (where X is user defined) or based on the assigned p -value. See [S1 Text](#) for complete details on all panels.

Implementation

iDREM is implemented using a combination of Java and Javascript. The regulatory model prediction part is implemented in Java and the interactive visualization is implemented in Javascript with D3.js and Google charts. Users only need to run the main java program `idrem.jar` to get all results including the interactive visualization. Please refer to iDREM readme for more details (<https://github.com/phoenixding/idrem>).

Results

Applying iDREM to study mouse microglia development

We illustrate the functionality of iDREM by applying it to reconstruct mouse microglia developmental regulatory networks from a diverse set of high throughput biological data types ([S1 Table](#)). Microglia are a type of small macrophage-like glial cell and these cells comprise up to 15% of all cells in the brain. Most of the data we used for this analysis, including mRNA expression data, histone methylation data and single cell RNA-Seq data is from a study of

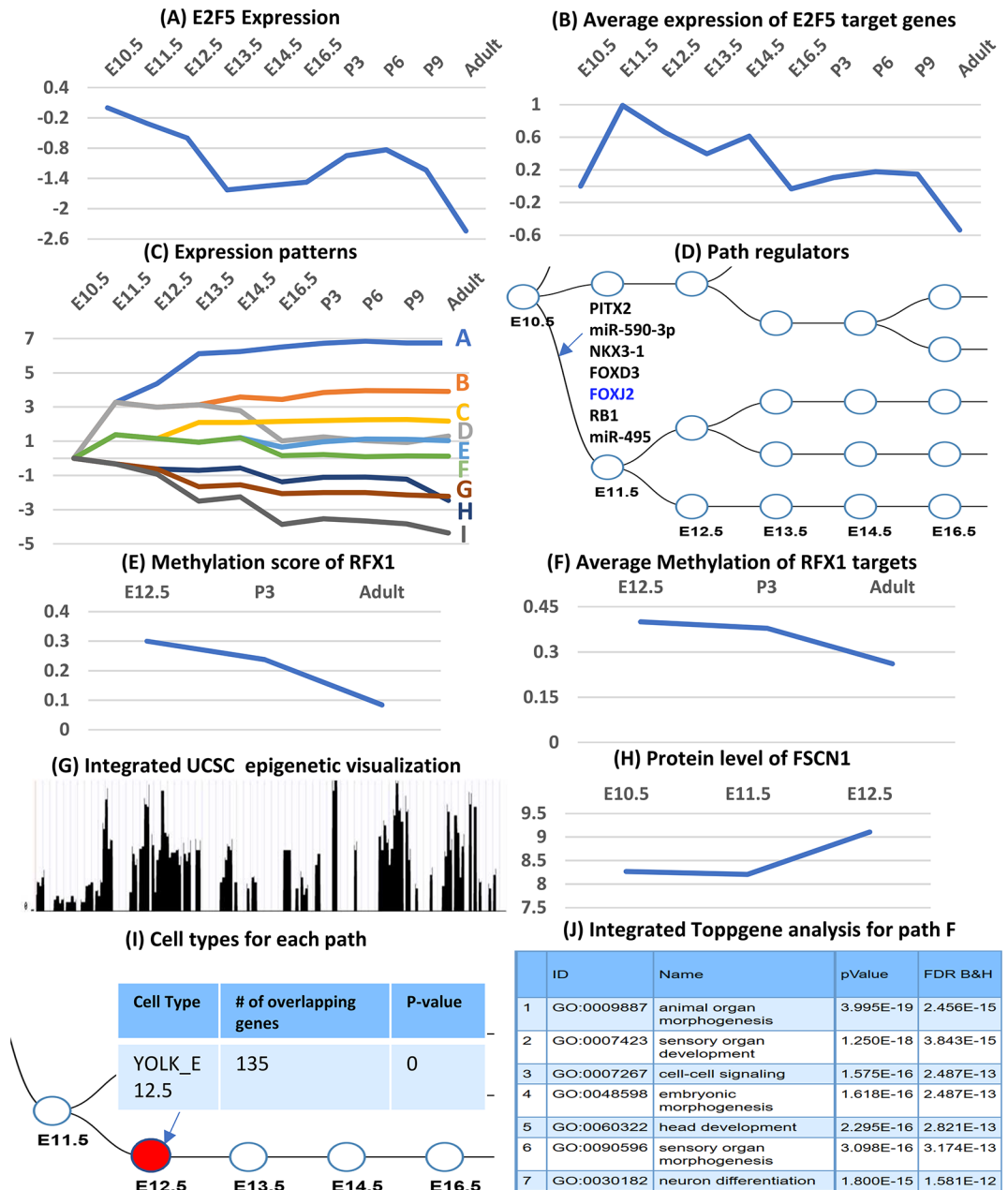


Fig 2. iDREM visualization functions. Top: Expression of a regulator (E2F5) (A) and its targets (B). 2nd row: Expression patterns (similar to the original DREM result, can be viewed from the tool as well) (C) and the regulators for each of these splits (D). 3rd row: Methylation of a regulator (E) and its targets (F). 4th row: Integration with additional browsers for viewing epigenetic data for specific TFs / genes (G) and protein level for specific TFs/proteins (H). 5th row: Intersection of path genes with single cell data (I) and integrated GO functional analysis (J).

<https://doi.org/10.1371/journal.pcbi.1006019.g002>

microglia development [22]. We have also included whole brain time series proteomics data [27] and miRNA expression data [28]. While the whole brain data may only partially overlap with the microglia profiles, since the focus here is on the methods and visualization, we have added that data to fully showcase the ability of iDREM to integrate and interactively visualize diverse types of time series data.

The datasets overlapped in some of the time points used (S1 Table) though the overlap was only partial. This highlights another advantage of iDREM, the ability to utilize some data types in only a subset of time points which can improve the ability of researchers to integrate their data with other, publicly available, data. In addition to the condition-specific, time series data sets iDREM also uses general static TF-DNA interactions data similar to DREM 2.0 [17], static miRNA-mRNA interactions data [16] and protein-protein interactions data which are used for the time series proteomic data analysis and were downloaded from STRING(V10.5) [29].

Fig 1 provides an overview of the data used by iDREM to reconstruct the networks, the resulting DREM model and a screenshot from the interactive visualization tool (S2 Fig). The model determines the different paths and splits, the genes assigned to them and the TFs and miRNAs that control each of the paths and splits. The model reconstructed for the microglia development data (Fig 1) includes 9 different paths, which have each been assigned a set of regulating TFs and miRNAs. Several of the paths are correctly enriched for GO functions related to immune defense and development of the central nervous system, which have been reported as the primary function of microglia cells [30]. S2 Table presents the top GO terms associated with each path.

Several of the regulators identified for the paths are known to regulate microglia development (S3 Table). Specifically, the reconstructed network includes 5 of the 7 TFs identified manually in the original microglia study [22], all of which are determined to be very significant. In addition, the method identified a number of additional microglia relevant TFs including CD40 which is known to be a microglia marker [31], SMAD1 which is an immune system factor [32], TRAF4 which is reported to be involved in multiple immune functions [33] and more. Fig 1H presents many of the top TFs and miRNAs identified by iDREM as controlling the various paths in the model.

Fig 2 displays some of the visualization capabilities of iDREM. It also shows how the new functionality improves the accuracy of the reconstructed model. For example, regulatory factor X1 (RFX1) is an immune response factor [34], consistent with the function of microglia cell. However, without the time series methylation data RFX1 cannot be identified as a regulator. The large increase in the activation prior for RFX1 (Fig 2(E)) leads to much higher probability that RFX1 is regulating path B resulting in its inclusion in the reconstructed model. Note, The TF binding prior is smaller for genes with larger methylation score in iDREM model (might need a pre-processing for methylation associated with increased TF binding activities such as H3K4me2 methylation, please refer the iDREM manual for details). Similarly, the elevated protein expression levels of fascin actin-bundling protein 1 (FSCN1), an immune system regulator [35], enabled iDREM to correctly identify it as controlling the path from E12.5 to E13.5 (Fig 2(H)).

In this study, we provided some anecdotal evidence for the impact of these newly introduced features such as proteomics and epigenetics data (in Fig 2). We also performed additional analysis in which we removed one data type at a time and analyzed the differences in the resulting networks, significant GO functions associated with different paths and the set of regulators identified by the models. Specifically, we compared the 4 iDREM models: I) Does not use any of the new datasets (only uses miRNA, mRNA expression and the static interaction data); II) the data used by I + the time series proteomics data; III) the data used by I + the time series methylation data; IV) The model presented in the paper that uses all data types. We see an improvement when using more data types and the best results are obtained by model IV indicating that including all data types can lead to more accurate models. Please refer to S1 Text, S4, S5, S6, S7 and S8 Figs and S4 Table for the complete details.

Availability and future directions

The iDREM code and software, with an example input dataset and detailed instructions are available from GitHub (<https://github.com/phoenixding/idrem>). All the data, code and results are also available at the supporting website (<http://www.cs.cmu.edu/~jund/idrem/>). Future work of iDREM will focus on better integration of new data (e.g. time series Single-cell ATAC-Seq).

Supporting information

S1 Text. Supporting methods and results. This file provides the detailed method description and also the supporting results.

(PDF)

S1 Fig. iDREM visualization configuration panels. (A) Global config, which can be used to customize the visualizations (e.g. background color, node color, visualization size). (B) Regulator Panel, which can be used to visualize the gene/miRNA expression. (C) Enrichment panel, which can be used to find the enriched paths/nodes in iDREM model for any given inputs. (D) Expression panel, which can be used to visualize gene or miRNA expression. (E) Epigenomics Panel, which can be used to explore and visualize the epigenomics data used in the study. (F) Proteomics Panel, which can be used to visualize the protein levels. (G) Cell Types Panel, which can be used to explore/visualize the Single-cell or Sorted-Cell data. (H) Path Function Panel, which can be used to visualize the associated GO functions and regulators for each path. (I) Omnibus Panel, which can be used to explore and visualize the TF/gene in all possible panels. For a more detailed description, please refer to iDREM manual.

(PDF)

S2 Fig. iDREM interactive visualization. This figure shows the interactive visualization for the microglia data used in the study.

(PDF)

S3 Fig. An example of using single-cell RNA-seq data in iDREM. (A) The single-cell RNA-seq data. (B) Cluster the cells into different sub-types based on the expression profile. (C) Identify the signature genes (marker genes) for each cell type. (D) Intersect the marker genes (of specific cell-type) with the predicted paths/nodes in iDREM model to identify enriched paths/nodes. (E) This enables users to determine the cell type composition of the different nodes and paths and to infer whether specific changes observed are related to activation of TFs in existing cells or the formation of new cell types.

(PDF)

S4 Fig. Predicted paths for models I, II, III, IV. I: only use miRNA and mRNA expression data; II: data used by I + time series proteomics data; III: the data used by I + the time series methylation data; IV: using all data presented in the study.

(PDF)

S5 Fig. Sankey Diagram for model I. The Sankey Diagram shows the GO functions and regulators associated with each of the predicted paths.

(PDF)

S6 Fig. Sankey Diagram for model II. The Sankey Diagram shows the GO functions and regulators associated with each of the predicted paths.

(PDF)

S7 Fig. Sankey Diagram for model III. The Sankey Diagram shows the GO functions and regulators associated with each of the predicted paths.

(PDF)

S8 Fig. Sankey Diagram for model IV. The Sankey Diagram shows the GO functions and regulators associated with each of the predicted paths.

(PDF)

S1 Table. Mouse microglia development time points used in this paper.

(PDF)

S2 Table. Top Go Terms associated with each path.

(PDF)

S3 Table. Supported regulating factors predicted by iDREM.

(PDF)

S4 Table. Regulator comparison for models using different sets of input data.

(PDF)

Author Contributions

Conceptualization: Jun Ding, James S. Hagood, Namasivayam Ambalavanan, Naftali Kaminski, Ziv Bar-Joseph.

Data curation: Jun Ding.

Funding acquisition: James S. Hagood, Namasivayam Ambalavanan, Naftali Kaminski, Ziv Bar-Joseph.

Investigation: Ziv Bar-Joseph.

Methodology: Jun Ding, Ziv Bar-Joseph.

Software: Jun Ding, Ziv Bar-Joseph.

Supervision: Ziv Bar-Joseph.

Validation: Jun Ding, Ziv Bar-Joseph.

Visualization: Jun Ding.

Writing – original draft: Jun Ding, Ziv Bar-Joseph.

References

1. Kim SY, Imoto S, Miyano S. Inferring gene networks from time series microarray data using dynamic Bayesian networks. *Briefings in bioinformatics*. 2003; 4(3):228–235. <https://doi.org/10.1093/bib/4.3.228> PMID: 14582517
2. Raychaudhuri S, Stuart JM, Altman RB. Principal components analysis to summarize microarray experiments: application to sporulation time series. In: *Pacific Symposium on Biocomputing*. Pacific Symposium on Biocomputing. NIH Public Access; 2000. p. 455.
3. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. *Proceedings of the National Academy of Sciences*. 2001; 98(9):5116–5121. <https://doi.org/10.1073/pnas.091062498>
4. Lachmann A, Xu H, Krishnan J, Berger SI, Mazloom AR, Ma'ayan A. ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X experiments. *Bioinformatics*. 2010; 26(19):2438–2444. <https://doi.org/10.1093/bioinformatics/btq466> PMID: 20709693
5. Jones CJ, Newsom D, Kelly B, Irie Y, Jennings LK, Xu B, et al. ChIP-Seq and RNA-Seq reveal an AmrZ-mediated mechanism for cyclic di-GMP synthesis and biofilm development by *Pseudomonas*

- aeruginosa. *PLoS pathogens*. 2014; 10(3):e1003984. <https://doi.org/10.1371/journal.ppat.1003984> PMID: 24603766
6. Xia J, Mandal R, Sinelnikov IV, Broadhurst D, Wishart DS. MetaboAnalyst 2.0—a comprehensive server for metabolomic data analysis. *Nucleic acids research*. 2012; 40(W1):W127–W133. <https://doi.org/10.1093/nar/gks374> PMID: 22553367
 7. Huang GT, Athanassiou C, Benos PV. mirConnX: condition-specific mRNA-microRNA network integrator. *Nucleic acids research*. 2011; 39(suppl_2):W416–W423. <https://doi.org/10.1093/nar/gkr276> PMID: 21558324
 8. Setty M, Helmy K, Khan AA, Silber J, Arvey A, Neezen F, et al. Inferring transcriptional and microRNA-mediated regulatory programs in glioblastoma. *Molecular systems biology*. 2012; 8(1):605. <https://doi.org/10.1038/msb.2012.37> PMID: 22929615
 9. Huang JC, Babak T, Corson TW, Chua G, Khan S, Gallie BL, et al. Using expression profiling data to identify human microRNA targets. *Nature methods*. 2007; 4(12):1045–1049. <https://doi.org/10.1038/nmeth1130> PMID: 18026111
 10. Ideker T, Thorsson V, Ranish JA, Christmas R, Buhler J, Eng JK, et al. Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science*. 2001; 292(5518):929–934. <https://doi.org/10.1126/science.292.5518.929> PMID: 11340206
 11. Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, et al. The Perseus computational platform for comprehensive analysis of (prote) omics data. *Nature methods*. 2016; 13(9):731–740. <https://doi.org/10.1038/nmeth.3901> PMID: 27348712
 12. Borirak O, Rolfe MD, de Koning LJ, Hoefsloot HC, Bekker M, Dekker HL, et al. Time-series analysis of the transcriptome and proteome of *Escherichia coli* upon glucose repression. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*. 2015; 1854(10):1269–1279. <https://doi.org/10.1016/j.bbapap.2015.05.017>
 13. Treutlein B, Brownfield DG, Wu AR, Neff NF, Mantalas GL, Espinoza FH, et al. Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature*. 2014; 509(7500):371–375. <https://doi.org/10.1038/nature13173> PMID: 24739965
 14. Deng Q, Ramsköld D, Reinius B, Sandberg R. Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Science*. 2014; 343(6167):193–196. <https://doi.org/10.1126/science.1245316> PMID: 24408435
 15. Ernst J, Vainas O, Harbison CT, Simon I, Bar-Joseph Z. Reconstructing dynamic regulatory maps. *Molecular systems biology*. 2007; 3(1):74. <https://doi.org/10.1038/msb4100115> PMID: 17224918
 16. Schulz MH, Pandit KV, Cardenas CLL, Ambalavanan N, Kaminski N, Bar-Joseph Z. Reconstructing dynamic microRNA-regulated interaction networks. *Proceedings of the National Academy of Sciences*. 2013; 110(39):15686–15691. <https://doi.org/10.1073/pnas.1303236110>
 17. Schulz MH, Devanny WE, Gitter A, Zhong S, Ernst J, Bar-Joseph Z. DREM 2.0: Improved reconstruction of dynamic regulatory networks from time-series expression data. *BMC systems biology*. 2012; 6(1):104. <https://doi.org/10.1186/1752-0509-6-104> PMID: 22897824
 18. Gitter A, Carmi M, Barkai N, Bar-Joseph Z. Linking the signaling cascades and dynamic regulatory networks controlling stress responses. *Genome research*. 2013; 23(2):365–376. <https://doi.org/10.1101/gr.138628.112> PMID: 23064748
 19. Ciofani M, Madar A, Galan C, Sellars M, Mace K, Pauli F, et al. A validated regulatory network for Th17 cell specification. *Cell*. 2012; 151(2):289–303. <https://doi.org/10.1016/j.cell.2012.09.016> PMID: 23021777
 20. Gudbjartsson DF, Helgason H, Gudjonsson SA, Zink F, Oddson A, Gylfason A, et al. Large-scale whole-genome sequencing of the Icelandic population. *Nature genetics*. 2015; 47(5):435. <https://doi.org/10.1038/ng.3247> PMID: 25807286
 21. Song L, Huang SsC, Wise A, Castanon R, Nery JR, Chen H, et al. A transcription factor hierarchy defines an environmental stress response network. *Science*. 2016; 354(6312):aag1550. <https://doi.org/10.1126/science.aag1550> PMID: 27811239
 22. Matcovitch-Natan O, Winter DR, Giladi A, Aguilar SV, Spinrad A, Sarrazin S, et al. Microglia development follows a stepwise program to regulate brain homeostasis. *Science*. 2016; 353(6301):aad8670. <https://doi.org/10.1126/science.aad8670> PMID: 27338705
 23. Wang Y, Li X, Hu H. H3K4me2 reliably defines transcription factor binding regions in different cells. *Genomics*. 2014; 103(2):222–228. <https://doi.org/10.1016/j.ygeno.2014.02.002> PMID: 24530516
 24. Karolchik D, Baertsch R, Diekhans M, Furey TS, Hinrichs A, Lu Y, et al. The UCSC genome browser database. *Nucleic acids research*. 2003; 31(1):51–54. <https://doi.org/10.1093/nar/gkg129> PMID: 12519945

25. Eramo A, Lotti F, Sette G, Pilozzi E, Biffoni M, Di Virgilio A, et al. Identification and expansion of the tumorigenic lung cancer stem cell population. *Cell death and differentiation*. 2008; 15(3):504. <https://doi.org/10.1038/sj.cdd.4402283> PMID: 18049477
26. Du Y, Kitzmiller JA, Sridharan A, Perl AK, Bridges JP, Misra RS, et al. Lung Gene Expression Analysis (LGEA): an integrative web portal for comprehensive gene expression data analysis in lung development. *Thorax*. 2017; p. thoraxjnl–2016. <https://doi.org/10.1136/thoraxjnl-2016-209598> PMID: 28070014
27. Hartl D, Irmeler M, Römer I, Mader MT, Mao L, Zabel C, et al. Transcriptome and proteome analysis of early embryonic sorting lungyonic mouse brain development. *Proteomics*. 2008; 8(6):1257–1265.
28. Miska EA, Alvarez-Saavedra E, Townsend M, Yoshii A, Šestan N, Rakic P, et al. Microarray analysis of microRNA expression in the developing mammalian brain. *Genome biology*. 2004; 5(9):R68. <https://doi.org/10.1186/gb-2004-5-9-r68> PMID: 15345052
29. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: protein—protein interaction networks, integrated over the tree of life. *Nucleic acids research*. 2014; 43(D1): D447–D452. <https://doi.org/10.1093/nar/gku1003> PMID: 25352553
30. Filiano AJ, Gadani SP, Kipnis J. Interactions of innate and adaptive immunity in brain development and function. *Brain research*. 2015; 1617:18–27. <https://doi.org/10.1016/j.brainres.2014.07.050> PMID: 25110235
31. Ponomarev ED, Shriver LP, Dittel BN. CD40 expression by microglial cells is required for their completion of a two-step activation process during central nervous system autoimmune inflammation. *The Journal of Immunology*. 2006; 176(3):1402–1410. <https://doi.org/10.4049/jimmunol.176.3.1402> PMID: 16424167
32. Malhotra N, Kang J. SMAD regulatory networks construct a balanced immune system. *Immunology*. 2013; 139(1):1–10. <https://doi.org/10.1111/imm.12076> PMID: 23347175
33. Cherfils-Vicini J, Vingert B, Varin A, Tartour E, Fridman WH, Sautès-Fridman C, et al. Characterization of immune functions in TRAF4-deficient mice. *Immunology*. 2008; 124(4):562–574. <https://doi.org/10.1111/j.1365-2567.2008.02810.x> PMID: 18284467
34. Pugliatti L, Derre J, Berger R, Ucla C, Reith W, Mach B. The genes for MHC class II regulatory factors RFX1 and RFX2 are located on the short arm of chromosome 19. *Genomics*. 1992; 13(4):1307–1310. [https://doi.org/10.1016/0888-7543\(92\)90052-T](https://doi.org/10.1016/0888-7543(92)90052-T) PMID: 1505960
35. Abbas AR, Baldwin D, Ma Y, Ouyang W, Gurney A, Martin F, et al. Immune response in silico (IRIS): immune-specific genes identified from a compendium of microarray expression data. *Genes and immunity*. 2005; 6(4):319. <https://doi.org/10.1038/sj.gene.6364173> PMID: 15789058