MIDREG: A METHOD OF MINING DEVELOPMENTALLY REGULATED GENES USING BOOLEAN IMPLICATIONS

D. Sahoo[*], J. Seita, D. Bhattacharya, M.A. Inlay, I.L. Weissman

Institute of Stem Cell Biology and Regenerative Medicine, Stanford University, 1050 Arastradero Rd (A148)
Stanford, CA 94305,USA
[*]Email: sahoo@stanford.edu

S.K. Plevritis

Department of Radiology, Stanford University, The Lucas Center for MR Spectroscopy and Imaging - P268
Stanford, CA 94305, USA
Email: sylvia.plevritis@stanford.edu

D.L. Dill[+]

Department of Computer Science, Stanford University, 353 Serra Mall (Gates 344)
Stanford, CA 94305, USA
Email: dill@stanford.edu

We present a new method termed Mining Developmentally Regulated Genes (MiDReG) to predict genes whose expression is either activated or repressed as precursor cells differentiate. The MiDReG algorithm bases its predictions on the gene expression patterns between the initial and terminal stages of the differentiation pathway, coupled with "if-then" rules (Boolean implications) mined from large-scale microarray databases. MiDReG uses two gene expression-based seed conditions that mark the initial and the terminal stages of a given differentiation pathway, and combine the statistically inferred Boolean implications from these seed conditions in a unique way to identify the relevant genes. On B cell development the algorithm predicted 62 genes that are expressed after the KIT+ progenitor cell stage and remain expressed through CD19+ and AICDA+ germinal center B cells. Review of the published literature of knockout mice revealed that of the predicted genes, 63.4% have defects in B cell differentiation and function, and 22% have a role in B cell according to other experiments. Therefore, our method identified novel gene candidates for future examination of their role in B cell develo pment. In a companion study we tested the functional significance of expression of one of the identified genes, Ly6d: Ly6d expression marks the earliest B cell progenitor yet defined in mice, and splits the common lymphoid progenitors (CLP) into Ly6d+ B cell committed and Ly6d- CLP. These data demonstrate the power of MiDReG in predicting functionally important genes in a given developmental pathway that is defined by a mutually exclusive gene expression pattern.

Reference/Funding source: National Institutes of Health, award 5U56CA112973 (to S.K.P) and 5R01CA086065 (to I.L.W.), National Institutes of Health fellowship 5K01DK078318 (to D.B), the California Institute of Regenerative Medicine fellowship (to J.S.).

1. Sahoo D, Seita J, Bhattacharya D, Inlay MA, Weissman IL, Plevritis SK, Dill DL. MiDReG: A Method of Mining Developmentally Regulated Genes Using Boolean Implications. *PNAS*, 2010 Mar 30;**107**(13):5732-7. Epub 2010 Mar 15.
2. Inlay MA, Bhattacharya D, Sahoo D, Serwold T, Seita J, Karsunky H, Plevritis SK, Dill DL, Weissman IL. Ly6d marks the earliest stage of B-cell specification and identifies the branchpoint between B-cell and T-cell development. *Genes Dev*. 2009 Oct 15;**23**(20):2376-81.

[*] Corresponding author.