Inferring Boolean Networks with perturbation from sparse gene expression data: a general model applied to the Interferon regulatory network.

Le Yu,*[a] Steven Watterson*[bc], Stephen Marshall[a] and Peter Ghazal[bc]

[a] *Department of Electronic and Electrical Engineering, University of Strathclyde, Royal College Building, 204 George Street, Glasgow, G1 1XW, Scotland. Fax: +44 141 552 2487; Tel: +44 141 548 2350; Email: l.yu@eee.strath.ac.uk.*

[b] *Division of Pathway Medicine, University of Edinburgh Medical School, Chancellor's Building, 49 Little France Crescent, Edinburgh, EH16 4SB, Scotland. Fax: +44 131 242 6244; Tel: +44 131 242 6242; Email: s.watterson@ed.ac.uk.*

[c] *Centre for Systems Biology at Edinburgh, Darwin Building, The King's Buildings, West Mains Road, Edinburgh, EH3 9JY, Scotland. Fax: +44 131 651 9068; Tel: +44 651 9065.*

Supplementary Material

Section 1 – The limit of equation (4)

Equation (2) describes a transition of of the system as

$$
\begin{pmatrix} P_1(t+\Delta t) \\ P_2(t+\Delta t) \\ P_3(t+\Delta t) \end{pmatrix} = \begin{pmatrix} 1-p+pq_{11} & pq_{12} & pq_{13} \\ pq_{21} & 1-p+pq_{22} & pq_{23} \\ pq_{31} & pq_{32} & 1-p+pq_{33} \end{pmatrix} \begin{pmatrix} P_1(t) \\ P_2(t) \\ P_3(t) \end{pmatrix}
$$

As t gets large, we expect this to converge on a specific distribution of $P_i(t)$. Thus, in the limit t→∞, we have equation (4)

$$
\begin{pmatrix} P_1(t) \\ P_2(t) \\ P_3(t) \end{pmatrix} \underset{t \to \infty}{\lim} = \begin{pmatrix} 1-p+pq_{11} & pq_{12} & pq_{13} \\ pq_{21} & 1-p+pq_{22} & pq_{23} \\ pq_{31} & pq_{32} & 1-p+pq_{33} \end{pmatrix} \begin{pmatrix} P_1(t) \\ P_2(t) \\ P_3(t) \end{pmatrix}
$$

We can see how we would obtain this limit, by considering the constraints it places on the derivatives of the $P_i(t)$. By expanding equation (2) in a Taylor series, we have

$$
\begin{pmatrix} P_1(t) \\ P_2(t) \\ P_3(t) \end{pmatrix} + \Delta t \begin{pmatrix} P_1'(t) \\ P_2'(t) \\ P_3'(t) \end{pmatrix} + \frac{1}{2}\Delta t^2 \begin{pmatrix} P_1''(t) \\ P_2''(t) \\ P_3''(t) \end{pmatrix} + \ldots = \begin{pmatrix} 1-p+pq_{11} & pq_{12} & pq_{13} \\ pq_{21} & 1-p+pq_{22} & pq_{23} \\ pq_{31} & pq_{32} & 1-p+pq_{33} \end{pmatrix} \begin{pmatrix} P_1(t) \\ P_2(t) \\ P_3(t) \end{pmatrix}
$$

If this is to be true for arbitrary step size Δt, we require that all the derivatives of $P_i(t)$ must go to zero. Thus the solution to equation (4) is defined in the region where $d^n P_i(t)/dt^n \to 0$, for n≥1 and for all values of i.

Section 2 – Experimental details and results

We applied the algorithm to expresion level data taken from experiments in which bone marrow derived macrophages were exposed to Murine Cytomegalovirus. The macrophages were exposed to cytomegalovirus at time zero and the resulting expression levels were measured at 30 minute intervals, for a period of 12 hours, using Agilent technology's whole mouse genome oligo microarrays (forerunner to Agilent part number G4122F).

The expression data was annotated using the mgug4121a package on Bioconductor software. Alongside this time course, a control time course was implemented on the same microarray platform.

Portions of each sample from the control, taken across all time points were pooled and used as a common control RNA. The control RNA and each experimental sample were labelled with separate dyes (Cyanines 3 and 5) and the relative levels of the dyes were used to provide a measure of the expression level of the RNA corresponding to each probe for each time point. Expression levels were obtained in the form of log2 scale signal intensity ratios between the sample and the control RNA. Samples from different time points were normalised using subset median normalisation, in which the difference between the mean median expresion value of the control across the time course and the median expresion at each time point of the control, is used to correct the experimental values. (Y. Yang, S. Dudoit, P. Luu, D. Lin, V. Peng, J. Ngai, and T. Speed (2002). Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Research* **30**(4), pp15.)

Table 1. The time course of expression levels for Irf1, Irf2, Irf3 and Irf4 together with their binarised values.

Time	Irf1		Irf2		Irf3		Irf4	
0	0.51	0	-0.16	0	0.01	1	0.37	1
30	0.90	0	-0.04	0	-0.24	1	0.57	1
60	1.26	0	-0.42	0	-0.21	1	0.41	1
90	1.97	1	-0.95	0	-0.54	0	0.53	1
120	2.15	1	-1.07	0	-0.40	0	0.36	1
150	2.26	1	-0.95	0	-0.54	0	0.37	1
180	4.33	1	1.26	1	-0.55	0	0.41	1
210	3.47	1	1.44	1	-0.51	0	0.52	1
240	3.08	1	1.16	1	-0.32	0	0.66	1
270	2.89	1	1.53	1	-0.35	0	0.18	0
300	3.10	1	1.82	1	-0.38	0	0.22	1
330	3.05	1	1.49	1	-0.48	0	0.49	1
360	3.33	1	1.66	1	-0.27	1	0.26	1
390	3.08	1	1.59	1	-0.30	0	0.25	1
420	3.30	1	1.32	1	-0.23	1	0.33	1
450	2.54	1	1.04	1	-0.28	0	0.24	1
480	2.83	1	1.13	1	-0.15	1	0.33	1
510	2.43	1	1.20	1	-0.16	1	0.45	1
540	3.49	1	1.04	1	-0.18	1	0.46	1
570	2.32	1	1.11	1	-0.11	1	0.45	1
600	2.21	1	0.93	1	-0.10	1	0.41	1
630	2.24	1	1.17	1	-0.09	1	0.24	1
660	2.46	1	0.78	1	-0.26	1	-0.11	0
690	2.14	1	1.11	1	-0.18	1	-0.03	0
720	2.22	1	0.76	1	-0.05	1	-0.27	0