# Early Gene Regulation of Osteogenesis in Embryonic Stem Cells - Supplementary Information on Mathematical Modeling

**Glen R Kirkham,[*a#] Anna Lovrics,[b,c#] Helen M Byrne,[b,d] Oliver E Jensen,[b] John R King,[b] Kevin M Shakesheff,[a] and Lee DK Buttery[a]**

## S-1 Experimental data

Expression levels of Dlx5, Msx2 and Runx2 mRNA normalized to GAPDH expression are shown in table S-1. Experimental conditions are cells with no growth factors (CO); cells with 300ng/ml BMP2; with 10ng/ml TGF$\beta$; or 300ng/ml BMP2 and 10ng/ml TGF$\beta$1.

| **Dlx5** | | | | | |
|---|---|---|---|---|---|
| | 0h | 8h | 16h | 24h | 48h |
| CO | 1.00 | $1.23 \pm 0.05$ | $1.44 \pm 0.14$ | $1.18 \pm 0.07$ | $1.44 \pm 0.07$ |
| BMP2 | 1.00 | $1.68 \pm 0.07$ | $1.65 \pm 0.08$ | $2.95 \pm 0.44$ | $1.48 \pm 0.12$ |
| TGF$\beta$1 | 1.00 | $1.32 \pm 0.02$ | $0.87 \pm 0.07$ | $1.10 \pm 0.44$ | $0.57 \pm 0.07$ |
| BMP2 & TGF$\beta$1 | 1.00 | $1.49 \pm 0.16$ | $1.73 \pm 0.04$ | $4.28 \pm 0.33$ | $1.41 \pm 0.20$ |

| **Msx2** | | | | | |
|---|---|---|---|---|---|
| CO | 1.00 | $0.92 \pm 0.16$ | $1.53 \pm 0.05$ | $1.58 \pm 0.10$ | $1.33 \pm 0.08$ |
| BMP2 | 1.00 | $1.68 \pm 0.07$ | $1.65 \pm 0.08$ | $2.74 \pm 0.16$ | $2.36 \pm 0.04$ |
| TGF$\beta$1 | 1.00 | $1.39 \pm 0.16$ | $0.61 \pm 0.03$ | $1.06 \pm 0.04$ | $0.98 \pm 0.18$ |
| BMP2 & TGF*beta*1 | 1.00 | $1.41 \pm 0.11$ | $1.41 \pm 0.18$ | $1.91 \pm 0.08$ | $1.08 \pm 0.02$ |

| **Runx2** | | | | | |
|---|---|---|---|---|---|
| CO | 1.00 | $1.02 \pm 0.02$ | $1.08 \pm 0.03$ | $1.68 \pm 0.26$ | $1.49 \pm 0.26$ |
| BMP2 | 1.00 | $1.23 \pm 0.04$ | $1.04 \pm 0.04$ | $1.52 \pm 0.05$ | $1.43 \pm 0.01$ |
| TGF$\beta$1 | 1.00 | $1.23 \pm 0.02$ | $1.09 \pm 0.05$ | $2.20 \pm 0.22$ | $1.32 \pm 0.22$ |
| BMP2 & TGF*beta*1 | 1.00 | $1.09 \pm 0.09$ | $1.12 \pm 0.03$ | $0.92 \pm 0.13$ | $1.78 \pm 0.08$ |

**Table S-1** Expression levels of Dlx5, Msx2 and Runx2 mRNAs.

## S-2 Logic rules generated from GRNs

Here we explain how Boolean logic rules can be generated from GRNs. A Boolean logic rule $l$ is a set of binary values that relates the state of a target gene to the expression levels of its regulators. Thus the expression level of the Boolean variable $B$ at simulation step $(n+1)$ is given by

$$B(n+1) = l(R(n)), \tag{S-1}$$

where the vector $R(n)$ contains the Boolean values of the regulators of $B$ at simulation step $n$ and $l(.)$ is the Boolean logic rule.

A GRN can be used to formulate a logic rule in the following way. If only activators are expressed at step $n$ then the target gene will be 'on' at simulation step $(n+1)$. If no regulators, or only inhibitors, are expressed at step $n$ than the target gene will be 'off' at simulation step $(n+1)$. If activators and inhibitors are coexpressed then the target gene may be 'on' or 'off' at simulation

[#]Equal contributors
[a] *Centre for Biomolecular Sciences, University Park, Nottingham, NG7 2RD, UK. Tel: +44(0)115 8232003; E-mail: glen.kirkham@nottingham.ac.uk*
[b] *School of Mathematical Sciences, University of Nottingham, University Park , Nottingham, NG7 2RD, UK.*
[c] *Biotalentum Ltd., Szent-Györgyi Albert u 4, Gödöllő, H-2100, Hungary*
[d] *Oxford Centre for Collaborative Applied Mathematics, Mathematical Institute, University of Oxford, Oxford, OX1 3LB, UK.*

| regulators at step $n$ | $\begin{bmatrix} A_1 \\ A_2 \\ I \end{bmatrix}$ | $\begin{smallmatrix}0\\0\\0\end{smallmatrix}$ | $\begin{smallmatrix}0\\0\\1\end{smallmatrix}$ | $\begin{smallmatrix}0\\1\\0\end{smallmatrix}$ | $\begin{smallmatrix}0\\1\\1\end{smallmatrix}$ | $\begin{smallmatrix}1\\0\\0\end{smallmatrix}$ | $\begin{smallmatrix}1\\0\\1\end{smallmatrix}$ | $\begin{smallmatrix}1\\1\\0\end{smallmatrix}$ | $\begin{smallmatrix}1\\1\\1\end{smallmatrix}$ |
|---|---|---|---|---|---|---|---|---|---|
| output | $l_1(A_1,A_2,I)$ | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 |
| at step | $l_2(A_1,A_2,I)$ | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 |
| $(n+1)$: | $l_3(A_1,A_2,I)$ | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 1 |
| T$(n+1)$= | $l_4(A_1,A_2,I)$ | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 |
| $l_i(A_1,A_2,I)$ | $l_5(A_1,A_2,I)$ | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 |
| inconsistent | $k_1(A_1,A_2,I)$ | 0 | 0 | 1 | **1** | 1 | 0 | 1 | **0** |
| logic | $k_2(A_1,A_2,I)$ | 0 | 0 | 1 | 0 | 1 | **1** | 1 | **0** |
| rules | $k_3(A_1,A_2,I)$ | 0 | 0 | 1 | 1 | 1 | **1** | 1 | **0** |

**Table S-2** The results of possible Boolean logic functions ($l_1$-$l_5$) corresponding to the GRN shown in figure S-1. Row 1: expression levels of regulators at simulation step $n$. Rows 2-6: expression level of the output gene at simulation step $(n+1)$ in the case of a consistent Boolean logic $l_i$ (row $i$+1, $i$=1,..,5). Rows 7-9: expression levels of the output gene at simulation step $(n+1)$ for the inconsistent Boolean logic rules $k_j$ (row $j$+6, $j=1,2,3$). Bold entries in the same row indicate that the output levels and, hence, the corresponding logic rule $k$ are inconsistent and omitted. (For definition of inconsistency see text.)

step $(n+1)$, depending on the relative strengths of the regulators. In such cases we propose alternative logic rules to account for all possibilities.

Consider for example the regulation of target gene $T$ by activators $A_1$ and $A_2$ and inhibitor $I$ (see figure S-1). Table S-2 shows the eight possible states of $(A_1,A_2,I)$ at simulation step $n$ and the output state of $T$ at simulation step $(n+1)$. The output is ambiguous when both activators and inhibitors are upregulated (see columns IV, VI and VII). In each case, the target gene $T$ will be either 'on' or 'off' at simulation step $(n+1)$, so at most eight logic rules can be constructed. However some of the logic rules are inconsistent, *i.e.* the effects of certain regulators are contradictory. For example, under rule $k_1$, when $A_1$ and I are 'on' and $A_2$ is 'off', the target gene (T) expression will be 1 (column IV), whereas when $A_1$, $A_2$ and I are all upregulated the target gene expression will be 0 (column VIII). This logic rule is inconsistent with the activatory role of $A_2$. Two other logic rules ($k_2$ and $k_3$) are also inconsistent and therefore omitted from our simulations. When there are two activators and one inhibitor, there are five consistent logic rules, $l_1$, $l_2$, ..., $l_5$.

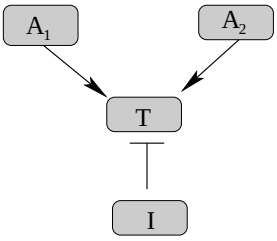



**Fig. S-1** Example regulatory connection: the target gene (T) is regulated by two activators ($A_1$ and $A_2$) and an inhibitor (I).

In summary, alternative logic rules are needed if regulators of different types act on a particular node. For each such logic rule, consistency is required as described above. The logic rules for a GRN consist of all combinations of the consistent logic rules for each target gene.

## S-3 ODE representation of Boolean models

When formulating an ordinary differential equation (ODE) model of a GRN, the evolution of each variable is often described by an equation of the following form:

$$\frac{dx_B}{dt} = f(\{x_R\}, \boldsymbol{p}) - \gamma x_B, \tag{S-2}$$

where $x_B$ is the protein (or mRNA or gene) of interest, $\{x_R\}$ is the set of its regulators, $\boldsymbol{p}$ is the vector of input growth factors, $f(\{x_R\},\boldsymbol{p})$ is the rate of production of the protein and $\gamma$ is a decay rate and $t$ denotes time.

By scaling time with the inverse decay rate ($t=\tau/\gamma$) and scaling the protein concentration with the maximal protein level ($x_B(t)=x_{B,\max}\hat{x}_B(\tau)$) equation (S-2) transforms to

$$\frac{d\hat{x}_B}{d\tau}=\hat{f}(\{x_R\},\mathbf{p})-\hat{x}_B, \tag{S-3}$$

where $\hat{f}(\{x_R\},\mathbf{p})=f(\{x_R\},\mathbf{p})/(\gamma x_{B,\max})$.

The condition for a steady state of the Boolean model (S-1) is

$$B(n)=B(n+1). \tag{S-4}$$

The corresponding steady state condition for the ODE model (S-3) is

$$\hat{x}_B(\tau)=\hat{f}(\{x_R(\tau)\},\boldsymbol{p}). \tag{S-5}$$

Comparing equations (S-4) and (S-5) we note that if the continuous function $\hat{f}$ agrees with the Boolean function $l$ when the regulators take the values 0 or 1, then each steady state of the Boolean model will also be a steady state of the continuous model (as shown in Reference [1]). We note that the converse is not true, *i.e.* the continuous system may possess steady states not present in the Boolean model. Guided by this result, our general approach is to extend a discrete Boolean model to a continuous model by requiring that the function $\hat{f}$ in the continuous model (S-3) gives the same result as the Boolean function $l$ when the regulators take the values 0 or 1.

Following Reference [2] we use the normalized HillCube method to transform the Boolean models to continuous ones, because it is a general method which can be applied to any Boolean logic rule. A polynomial $\hat{g}$ which fulfils the requirement stated above can be derived from a general Boolean logic rule $l$ (see equation (S-1)) as follows:

$$\hat{g}(x_{R,1},x_{R,2},...,x_{R,N}):= \tag{S-6}$$
$$\sum_{x_{R,1}=0}^{1}\sum_{x_{R,2}=0}^{1}...\sum_{x_{R,N}=0}^{1}\left[l(R_1,R_2,...,R_N)\prod_{j=1}^{N}\{R_jx_{R,j}+(1-R_j)(1-x_{R,j})\}\right],$$

where the continuous function $\hat{g}(.)$ replaces the Boolean logic rule $l$, $R_j$ ($j=1,..,N$) are the discrete Boolean regulators and $x_{R,j}$ ($j=1,..,N$) are the corresponding continuous regulators.

The structure of (S-6) guarantees that $\hat{g}(x_{R,1},x_{R,2},...,x_{R,N})$ depends linearly on each of the regulators $x_{R,i}$ ($i=1,..,N$). However transcriptional regulation is usually switch-like, *i.e.* as the value of a regulator crosses a threshold, the transcription rate changes rapidly. This switch-like behaviour is achieved in the ODE representation by defining the function $g(.)$ as follows:

$$g(x_{R,1},x_{R,2},...,x_{R,N}):=\hat{g}(H^+(x_{R,1},\Theta_{R,1}),H^+(x_{R,2},\Theta_{R,2}),...,H^+(x_{R,N},\Theta_{R,N})), \tag{S-7}$$

where the function $H^+(Y,\Theta_Y)$ is a Hill-function of the form:

$$H^+(Y,\Theta_Y)=\frac{Y^m}{Y^m+\Theta_Y^m}. \tag{S-8}$$

In (S-8), $Y$ is the concentration of an activator protein, $\Theta_Y$ is the threshold concentration of the activator above which transcription occurs and $m$ is the Hill coefficient. In the simulations, we fix $m=15$ to ensure switch-like behaviour.

Matching Boolean logic rules for mES cells (see section S-4) were transformed to ODEs using the general equation

$$\frac{\mathrm{d}x_i}{\mathrm{d}t}=b_i+g(x_{R,1},x_{R,2},...,x_{R,N})-x_i, \tag{S-9}$$

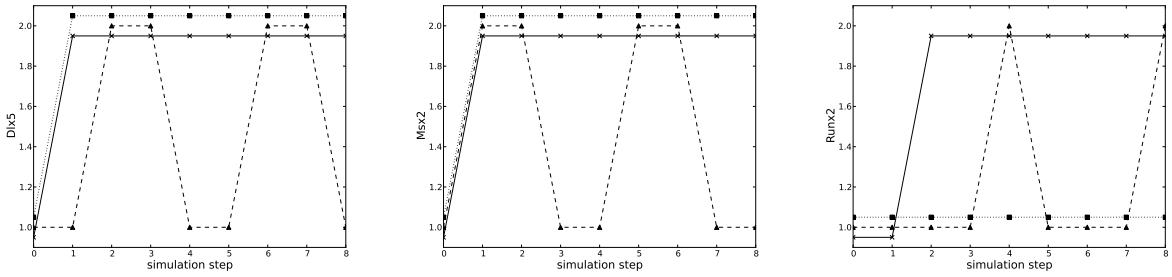
where $x_i$ correspond to variables Dlx5, Msx2 or Runx2, $b_i$ represent a basal transcription rate and $g(.)$ is the function defined in equation (S-7).

| | | B M R | 0 0 0 | 0 0 1 | 1 0 0 | 1 1 0 | 0 0 1 | 0 1 1 | 1 0 0 | 1 1 1 | B T D | 0 0 0 | 0 0 1 | 0 1 0 | 0 1 1 | 1 0 0 | 1 0 1 | 1 1 0 | 1 1 1 | T D M | 0 0 0 | 0 0 1 | 0 1 0 | 0 1 1 | 1 0 0 | 1 0 1 | 1 1 0 | 1 1 1 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| RøD | $l_1$ | D1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | M1 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | R1 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 0 |
| R→D | $l_2$ | D2 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | M2 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | R2 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 0 |
| R–\|D | $\boldsymbol{l_3}$ | **D3** | **0** | **0** | **1** | **0** | **1** | **0** | **1** | **1** | **M3** | **0** | **0** | **1** | **0** | **1** | **1** | **1** | **1** | **R4** | **0** | **0** | **1** | **1** | **0** | **0** | **1** | **0** |
| R–\|D | $l_4$ | D4 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | M4 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | R4 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 0 |
| R–\|D | $l_5$ | D5 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | M5 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | R5 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 0 |

**Table S-3** Logic rules that match the experimental data from the mES cells. These five logic rules correspond to three different GRNs (see figure 3): $l_1$ and $l_2$ represent GRNs where Runx2 does not regulate or positively regulate Dlx5 respectively and logic rules $l_3$-$l_5$ represent negative regulation of Dlx5 by Runx2. The header shows the expression state of the regulators for a particular TF and each row shows one possible matching logic rule $l_i$. The ODE representation of logic rule $l_3$ (denoted by bold) is shown in equations (S-10)-(S-12).

## S-4 Matching logic rules

The matching logic rules for the mES cells are summarized in table S-3. The header shows the possible states of the regulators for Dlx5 (D), Msx2 (M) and Runx2 (R) respectively and each row represents a matching logic rule where the expression states of the TFs are shown as a response to the expression state of the regulators. Each logic rule corresponds to one of the GRNs shown in figure 5 as indicated in the first column: Runx2 either activates (R→D) or inhibits (R–|D) Runx2 or it does not regulate Runx2 (RøD).




(a) (b) (c)

**Fig. S-2** Simulation results from Boolean model associated with logic rule $l_3$ (see table S-3). Level of Dlx5 (a), Msx2 (b) and Runx2 (c) modelling exposure to BMP2 (unbroken line with symbol 'x'), exposure to TGF$\beta$1 (broken line with symbol '▲') and exposure to both BMP2 and TGF$\beta$1 (dotted line with symbol '■'). Simulation results were normalized against simulation results modelling control medium. Expression levels are 1 or 2, these values representing either an 'off' or 'on' state of a gene. The binary values are shifted slightly so that all results can be presented on the same graph.

Each matching logic rule was used to generate an ODE model using the methods outlined in section S-3. A representative example is shown: the system of ODEs obtained from logic rule $l_3$ (see table S-3) is presented in equations (S-10)-(S-12). In order to be able to modify the expression level and the time-scale, parameters $p_i$, i=*D*, *M*, *R* (transcription rates of Dlx5, Msx2 and Runx2 respectively) and $\gamma_i$, i=*D*, *M*, *R* (decay rates) were added to obtain equations

$$\frac{\mathrm{dD}}{\mathrm{d}t} = b_D + p_D\left\{2H^+(g_{\mathrm{BMP}},\Theta_{BD})H^+(\mathrm{M},\Theta_{MD})H^+(\mathrm{R},\Theta_{RD}) - H^+(g_{\mathrm{BMP}},\Theta_{BD})H^+(\mathrm{M},\Theta_{MD}) - H^+(g_{\mathrm{BMP}},\Theta_{BD})H^+(\mathrm{R},\Theta_{RD}) - H^+(\mathrm{M},\Theta_{MD})H^+(\mathrm{R},\Theta_{RD}) + H^+(g_{\mathrm{BMP}},\Theta_{BD}) + H^+(\mathrm{M},\Theta_{MD})\right\} - \gamma_D\mathrm{D}, \quad \text{(S-10)}$$

$$\frac{\mathrm{dM}}{\mathrm{d}t} = b_M + p_M\left\{H^+(g_{\mathrm{BMP}},\Theta_{MM})H^+(g_{\mathrm{TGF}},\Theta_{TM})H^+(\mathrm{D},\Theta_{DM}) - H^+(g_{\mathrm{BMP}},\Theta_{MM})H^+(g_{\mathrm{TGF}},\Theta_{TM}) - H^+(g_{\mathrm{TGF}},\Theta_{TM})H^+(\mathrm{D},\Theta_{DM}) + H^+(g_{\mathrm{TGF}},\Theta_{TM}) + H^+(g_{\mathrm{BMP}},\Theta_{MM})\right\} - \gamma_M\mathrm{M}, \quad \text{(S-11)}$$

$$\frac{\mathrm{dR}}{\mathrm{d}t} = b_R + p_R\left\{H^+(\mathrm{D},\Theta_{DR}) - H^+(g_{\mathrm{TGF}},\Theta_{TR})H^+(\mathrm{D},\Theta_{DR})H^+(\mathrm{M},\Theta_{MR})\right\} - \gamma_R\mathrm{R}, \quad \text{(S-12)}$$

where $b_i$ (i=$D$, $M$, $R$) represent the basal transcription rate of Dlx5, Msx2 and Runx2 respectively, $\Theta_{ij}$ (i=$B$, $T$, $D$, $M$, $R$, j=$D$, $M$, $R$) is a threshold concentration above which the regulator $i$ has an effect of target $j$. Parameters $g_{\mathrm{BMP}}$ and $g_{\mathrm{TGF}}$ are control parameters representing the presence (1) or absence (0) of the corresponding growth factor as shown in table S-4.

| | Parameter Values | |
|---|---|---|
| Media | $g_{\mathrm{BMP}}$ | $g_{\mathrm{TGF}}$ |
| Control | 0 | 0 |
| BMP2 | 1 | 0 |
| TGF$\beta$1 | 0 | 1 |
| BMP2/TGF$\beta$1 | 1 | 1 |

**Table S-4** Control parameter values used in Boolean and ODE simulations to represent the different combinations of external growth factors

Parameter values of equations (S-10)-(S-12) were optimized to fit the experimental data. The simulation results shown in figures 6-8 were obtained with initial conditions $D(0) = M(0) = R(0) = 1$ and parmeters listed in table S-5.

| name | value | name | value |
|---|---|---|---|
| $b_D$ | 1.8 | $\Theta_{BD}$ | 0.1 |
| $b_M$ | 1.7 | $\Theta_{MD}$ | 0.2 |
| $b_R$ | 1.4 | $\Theta_{RD}$ | 0.9 |
| $p_D$ | 2.7 | $\Theta_{BM}$ | 0.1 |
| $p_M$ | 4 | $\Theta_{TM}$ | 0.9 |
| $p_R$ | 1.1 | $\Theta_{DM}$ | 0.9 |
| $\gamma_D$ | 2.0 | $\Theta_{TR}$ | 0.9 |
| $\gamma_M$ | 2.7 | $\Theta_{DR}$ | 0.1 |
| $\gamma_R$ | 1.6 | $\Theta_{MR}$ | 0.8 |

**Table S-5** Parameter values of equations (S-10)-(S-12) that were used to obtain simulation results shown in figures 6-8.

## References

1 L. Glass and S. Kauffman, *J. Theor. Biol.*, 1973, **39**, 103–129.
2 D. Wittmann, J. Krumsiek, J. Saez-Rodriguez, D. Lauffenburger, S. Klamt and F. Theis, *BMC Systems Biology*, 2009, **3**, 98.