Integrative Biology

Cite this: *Integr. Biol.*, 2012, **4**, 1470–1477

www.rsc.org/ibiology

**PAPER**

# Early gene regulation of osteogenesis in embryonic stem cells†

**Glen R. Kirkham,‡*[a] Anna Lovrics,‡[bc] Helen M. Byrne,[bd] Oliver E. Jensen,[b] John R. King,[b] Kevin M. Shakesheff[a] and Lee D. K. Buttery[a]**

***Received 29th June 2012, Accepted 15th August 2012***
**DOI: 10.1039/c2ib20164j**

The early gene regulatory networks (GRNs) that mediate stem cell differentiation are complex, and the underlying regulatory associations can be difficult to map accurately. In this study, the expression profiles of the genes Dlx5, Msx2 and Runx2 in mouse embryonic stem cells were monitored over a 48 hour period after exposure to the growth factors BMP2 and TGFβ1. Candidate GRNs of early osteogenesis were constructed based on published experimental findings and simulation results of Boolean and ordinary differential equation models were compared with our experimental data in order to test the validity of these models. Three gene regulatory networks were found to be consistent with the data, one of these networks exhibited sustained oscillation, a behaviour which is consistent with the general view of embryonic stem cell plasticity. The work cycle presented in this paper illustrates how mathematical modelling can be used to elucidate from gene expression profiles GRNs that are consistent with experimental data.

## 1 Introduction

The differentiation of mouse embryonic stem cells (mES) into numerous cell types has been widely demonstrated *in vitro*.[1–6] However, the signalling pathways and associated changes in gene expression that regulate this process are not well characterized. The differentiation of these pluripotent cells into bone is no exception, and a clearer understanding of the regulatory mechanisms that govern osteogenesis is vital for any future therapeutic applications, including those relating to tissue engineering.

Growth factors are potent regulators of differentiation and have been used in numerous studies to induce the osteogenic differentiation of mES cells.[6–20] Bone morphogenetic protein 2 (BMP2) and transforming growth factor β1 (TGFβ1) are examples of such growth factors.[12,14,15,6–10,17,19] The Runt-related transcription factor Runx2 is an essential regulatory gene within osteoblasts[21–26] and has been shown to mediate the downstream expression of numerous osteogenic genes including osteocalcin, alkaline phosphatase, bone sialoprotein and osteopontin.[22,27–30] The gene is known to be regulated by exposure to both BMP2 and TGFβ1;[10,14,23] however the upstream elements that control this process are not fully understood.

The homeobox proteins Dlx5 and Msx2 have both been shown to regulate the expression of osteocalcin and Runx2 in numerous cell types,[31–33] after stimulation by both BMP2 and TGFβ1.[14,15,34,35] Increases in Dlx5 expression have been induced in myogenic, osteoblast and mesenchymal cells after stimulation with BMP2.[36–39] Runx2 expression is known to be downstream of these genes, with over-expression of Dlx5

[a] *Centre for Biomolecular Sciences, University Park, Nottingham, NG7 2RD, UK. E-mail: glen.kirkham@nottingham.ac.uk; Tel: +44 (0)115 8232003*
[b] *School of Mathematical Sciences, University of Nottingham, University Park, Nottingham, NG7 2RD, UK*
[c] *Biotalentum Ltd., Szent-Györgyi Albert u 4, Gödöllö, H-2100, Hungary*
[d] *Oxford Centre for Collaborative Applied Mathematics, Mathematical Institute, University of Oxford, Oxford, OX1 3LB, UK*
† Electronic supplementary information (ESI) available: Supplementary information on mathematical modeling. See DOI: 10.1039/c2ib20164j
‡ Equal contributors.

### Insight, innovation, integration

Gene regulation is a complex system consisting of multiple elements (genes) that co-regulate one another, forming intricate networks that determine cellular activity. How these gene regulatory networks are affected by external inputs is a wide area of investigation but the level of complexity within these systems often makes accurate determinations difficult. We present an integrated approach that utilises mathematical models to map the interactions of a set of genes known to be involved in early bone cell formation. The formation of these models was then used to inform novel experimental characterisations of these genes in response to biological signals, demonstrating how small changes in biological inputs can affect larger changes in gene networks.

linked to increases in Runx2 expression,[10,15] and interactions between Dlx5 and the Runx2 promoter being observed in myoblast cell lines.[10] Msx2 is thought be a negative regulator of Runx2 expression[37,40,41] as studies have demonstrated that the protein binds to the Runx2 promoter without inducing its expression.[41]

In contrast to BMP2, studies using TGFβ1 have indicated that the growth factor is a negative regulator of osteogenesis.[12,15] High levels of TGFβ1 have been detected in both bone and cartilage tissue and may play key roles in their regulation.[7,42] In one study with myogenic cells, BMP2-mediated activation of Dlx5 was suppressed by exposure to TGFβ1.[15] Other studies have shown that Msx2 expression is also regulated after TGFβ1 exposure.[43,44] However, there is a large body of evidence that contradicts these studies, indicating that TGFβ1 could be a pro-osteogenic factor.[17,38,45,46]

In addition, *in vivo* studies have demonstrated that BMP2 induces ectopic bone formation,[47] whereas TGFβ1 can initiate *de novo* bone formation only when injected next to pre-existing bone tissue. Few *in vitro* studies have been carried out using the two growth factors in combination; however, Lee *et al.* exposed primary rat calvaria cells to both BMP7 and TGFβ1. They demonstrated that bone nodule formation was abrogated but the cells continued to express osteopontin. Taken together, these data suggest that the role of TGFβ1 in the regulation of bone formation could depend on cell type and the local environment.

The signalling and gene expression networks that regulate crosstalk between BMP2 and TGFβ1 regulation are not known. Candidates for such interactions are Dlx5 and Msx2 as both genes have been shown to compete for the same binding site on the Runx2 promoter.[34] Furthermore, in one study BMP-mediated Runx2 expression was further increased in myoblast cells after exposure to TGFβ1.[38] Msx2 regulation of Runx2 is somewhat contradictory, with some studies showing a suppression of Runx2 expression[40,41] and others indicating that the gene plays no role in the regulation of Runx2, and may in fact be pro-osteogenic *via* a Runx2-independent mechanism.[36]

Mathematical models have been developed in previous studies to characterise gene regulatory networks (GRNs)[48] and, in some cases, to make novel predictions.[49] Where information about the GRN is incomplete, mathematical modelling can be used to investigate alternative scenarios and to compare model predictions with available experimental data. Various modelling techniques have been used to achieve these goals, including ordinary differential equations (ODEs), Boolean networks, Petri nets, Bayesian or graphical Gaussian models, as well as Stochastic and Process Calculi.[48–66]

The idea to model GRNs with Boolean networks was originally proposed by Kauffman.[58] With increasing numbers of experimental studies performed to unravel genetic interactions, the Boolean approach has been used to model GRNs associated with a variety of processes in many different organisms including flower specification of *Arabidopsis thaliana*,[67,68] the cell cycle of budding yeast, fission yeast and mammals[69–71] and the embryonic segmentation of *Drosophilia melanogaster*.[72] The Boolean simulations in these studies were able to capture the qualitative dynamics observed in the corresponding experiments.

Booelan networks model gene expression *via* nodes that represent either genes or input factors (for example a growth factor) and edges that represent interactions between the nodes, with the associations being characterized as either activating or inhibiting. The expression level of regulators at simulation step n determines whether the target gene is expressed (on) or not (off) in simulation step $n + 1$.[58] The network is updated at each (discrete) time-step through Boolean functions, so transitions between states in the network are deterministic and synchronous. Stable steady states of the Boolean model correspond to cell phenotypes that can be compared to experimental results. Discrete Boolean logic rules can also be converted to continuous ODEs.[66]

ODE models permit more detailed and quantitative characterization of GRNs. Concentrations of inputs and parameters associated with reactions involving promoters and genes can relatively easily be modelled within an ODE based framework,[53] with stable steady states in the resulting systems of ODEs again representing cell phenotypes.

In this study we monitored the expression of Dlx5, Msx2 and Runx2 over 48 hours in mES cells. Based on publised results, candidate GRNs were incorporated into Boolean and equivalent ODE models. Simulations of Boolean models were conducted to select networks that are consistent with the observed 0 and 24 hour expression profiles and then ODEs corresponding to these GRNs were simulated to identify model parameters that best fit experimental data.

## 2 Results and discussion

### 2.1 Gene expression in mES

Embryonic stem cells were exposed to the growth factors BMP2 and TGFβ1 for 48 hours and expression levels of the genes Dlx5, Msx2 and Runx2 were monitored at 0, 8, 16, 24 and 48 hours. All gene expression levels were normalised against control conditions (cells without growth factors) so that relative changes due to the growth factors could be determined. Non-normalised expression levels are summarized in Table S1 (ESI†).

Expression of Dlx5 in mES cells (Fig. 1(a)) did not significantly rise above control levels in the 16 hours following exposure to the growth factors. After 24 hours both the BMP2 and BMP2/TGFβ1 groups showed a significant up-regulation ($p \leq 0.05$) which was higher than the TGFβ1 group ($p \leq 0.001$) and demonstrated no significant changes in the expression of Dlx5. In addition, the BMP2/TGFβ1 group showed significantly higher levels of expression at 24 hours than those treated with either BMP2 ($p \leq 0.05$) or TGFβ1 ($p \leq 0.001$) alone. All groups returned to initial expression levels after 48 hours stimulation except for the TGFβ1 group which showed a small decrease in expression compared to the other groups ($p \leq 0.01$) and its initial Dlx5 expression level ($p \leq 0.001$).

Similarly mES cells exposed to TGFβ1 alone, showed significantly reduced expression of Msx2 (Fig. 1(b)) over the 48 hour time-course ($p \leq 0.001$), with reduced levels being observed from the 16 hours time point (Fig. 1(b)). Expression levels were also lower than those of the other two experimental groups between 16 and 24 hours ($p \leq 0.001$). The highest levels of expression of Msx2 were associated with the BMP2 group at 24 hours, these being significantly higher than those observed within the TGFβ1 ($p \leq 0.001$) and BMP2/TGFβ1 ($p \leq 0.05$) groups. Levels of expression returned to initial

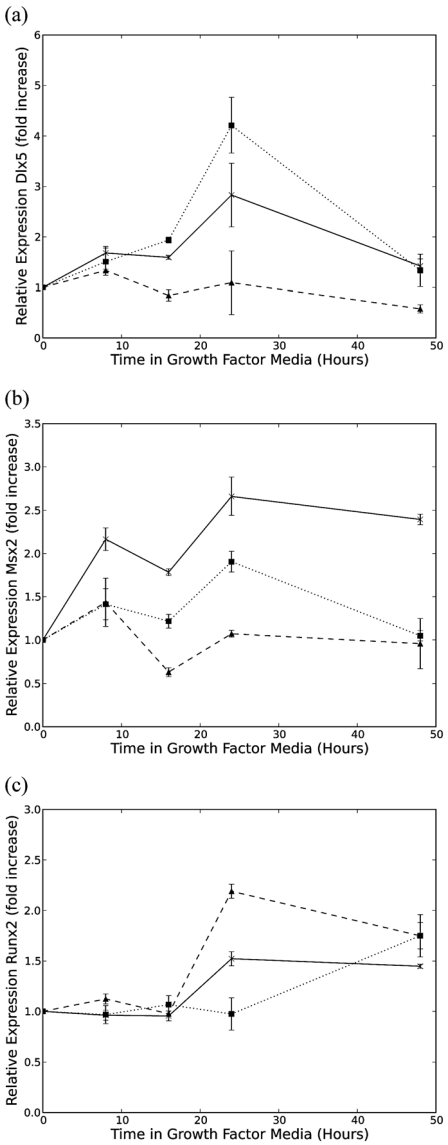


**Fig. 1** Quantitative expression of (a) Dlx5, (b) Msx2 and (c) Runx2 in mouse embryonic stem cells, after exposure to BMP2 (unbroken line) TGFβ1 (broken line) and BMP2/TGFβ1 (dotted line). Gene expression was normalized against housekeeping gene and then normalized against control expression (cells without growth factors). The zero time point was used as a calibrator for relative expression.

values in both TGFβ1 and BMP2/TGFβ1 groups at the 48 hours time point. In contrast, the levels of expression of Msx2 observed within the BMP2 group remained elevated 48 hours after stimulation ($p \leq 0.01$).

Increases in the levels of Runx2 expression (Fig. 1(c)) were observed at 24 hours in both the BMP2 and TGFβ1 groups compared to both the initial expression levels ($p \leq 0.001$ and $\leq 0.05$ respectively) and the levels within the BMP2/TGFβ1 group ($p \leq 0.001$ and $\leq 0.05$).

### 2.2 Mathematical modelling of mES cells

Potential GRNs of transcription factors (TFs) Dlx5, Msx2, Runx2 and input growth factors BMP2 and TGFβ1 were created based on previous experimental results (see Fig. 2). Fig. 2 represents 243 different GRNs: all of them were incorporated into Boolean models.

Throughout the simulations a two-step approach was applied: Boolean models were used to identify networks that are consistent with the observed expression profiles and corresponding ODE models were fit to experimental data. First, in order to compare Boolean model results to our experimental data, $t = 0$ hour and $t = 24$ hours mRNA expression levels associated with distinct cell phenotypes were categorized as *on* or *off* (see Table 1). The binary categorization of expression levels was supplemented by biological knowledge: mRNA levels are low in mES cells, so the initial mRNA expressions were considered to be *off*. The mRNA level for each TF at 24 hours was considered to be *on* if there was significant difference compared to the mRNA level at $t = 0$ hours and *off* otherwise.

On the second step, each matching Boolean logic rule was transformed into an equivalent ODE model using the HillCube method (see ESI†, S2).[66] Expression values at time $t = 0$ were used as initial conditions and the model results were compared to experimental data at times $t = 8$, 16 and 24 hours. The system reaches steady state by $t = 24$ hours, so fitting to data at $t = 48$ hours as well did not significantly change the results. In order to fit the model results to the experimental data, parameter optimization was conducted with the EcsPy[73] module in Python using a genetic algorithm.[74]

One or more Boolean models were constructed (see detailed explanation is ESI†, S2) for each of the $3^5 = 243$ GRNs shown in Fig. 2. During the Boolean simulations the mRNA levels at $t = 0$ hour (see Table 1) were considered as input and steady states of the models were compared to the binary representation of the experimental data at $t = 24$ hours (see Table 1). Using this method three different GRNs were found to reproduce the experimental data, as shown in Fig. 3 and Table S3 (ESI†). In these GRNs, the mechanism by which Runx2 regulates Dlx5 can not be resolved: in the different models the regulation is either absent, positive or negative.

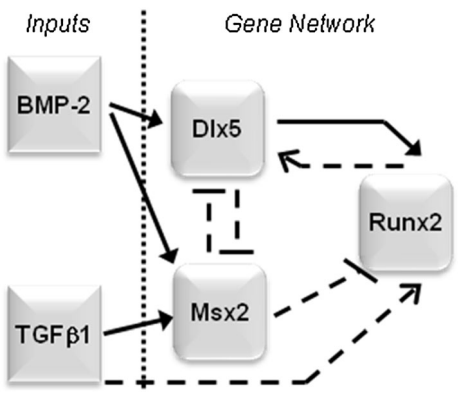


**Fig. 2** Possible gene regulatory networks of early osteogenesis (based on previous studies). Arrowheads denote activation and blunted ends inhibition. Associations depicted by continuous lines were demonstrated in at least two studies; associations based on single studies are depicted by a broken line. During simulations the nature of connections depicted by broken lines was allowed to be activating, inhibitory or inactive; hence the figure represents $3^5 = 243$ different GRNs.

 

**Table 1** Experimental results of mRNA expression in mES cells quantified as *on* or *off* at 0 h and 24 h after the addition of growth factors scaled to expression level in control medium

| Time (hours) | Media | Gene expression | | |
|---|---|---|---|---|
| | | Dlx5 | Msx2 | Runx2 |
| 0 | | Off | Off | Off |
| 24 | BMP2 | On | On | On |
| | TGFβ1 | Off | Off | On |
| | BMP2/TGFβ1 | On | On | Off |

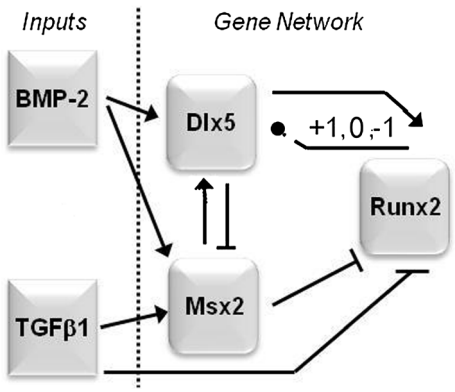


**Fig. 3** Summary of Boolean modelling results for mES cells. Graphical representation of logic rules that matched the experimental data (*cf.* ESI†, S4). Arrowheads denote activation and blunted ends inhibition, rounded arrowheads indicate that regulation is different for different matching logic rules: the possible regulations are denoted by numbers on the arrows: +1 refers to positive, −1 to negative regulation; 0 denotes no regulation.

In order to guide further experiments the matching logic rules were used to predict the effect of under- and over-expression of Dlx5, Msx2 or Runx2. Under-expression was simulated by holding the value of the relevant variable fixed at zero, over-expression was simulated by holding the value of the relevant variable fixed at one.

Steady state levels of Dlx5 and Msx2 were predicted for the different GRNs summarized in Fig. 3. If Runx2 positively regulates Dlx5, then upregulation is predicted both in control and TGFβ1 medium, and downregulation of Msx2 is predicted in TGFβ1 medium. In all other cases (when Runx2 negatively regulates or does not regulate Runx2), downregulation of Dlx5 is predicted both in control and TGFβ1 medium, whereas upregulation of Msx2 is predicted in TGFβ1 medium. By comparing our theoretical predictions with experimental results in which Runx2 is over-expressed it should be possible to reduce the number of feasible GRNs for osteogenesis in mES cells.

Simulation of each set of matching logic rules using Boolean models generated oscillatory results in TGFβ1 medium, as shown in Fig. S2 (ESI†). However the corresponing ODE models generated sustained oscillations only for the GRN in which Runx2 negatively regulates Dlx5. This is in accordance with the network topology: oscillation only occured when each TF (Dlx5, Msx2 and Runx2) could be both up- and down-regulated (see Fig. 3). A parameter fit was conducted for the system of ODEs corresponing to the GRN in which Runx2 negatively regulates Dlx5 (see ESI†, eqn (S10)–(S12)). Simulation results are shown in Fig. 4(a)–(c).

In accordance with the experimental data (see Fig. 1), at most time-points the expression levels of Dlx5 generated by the ODE simulations for both the BMP2 and BMP2/TGFβ1 groups are higher than the TGFβ1 group (Fig. 4(a)). Similarly, the simulated expression levels of Msx2 (Fig. 4(b)) for the BMP2 and BMP2/TGFβ1 groups are higher than the TGFβ1 group. In contrast to the Dlx5 and Msx2 data, Runx2 expression in the BMP2 and TGFβ1 groups is higher than in the BMP2/TGFβ1 group and this is also reflected by the simulation results (Fig. 4(c)).

ODE models can illustrate the order of appearance of TFs: simulation results indicate that Runx2 expression starts later than Dlx5 or Msx2. This result is consistent with the view that

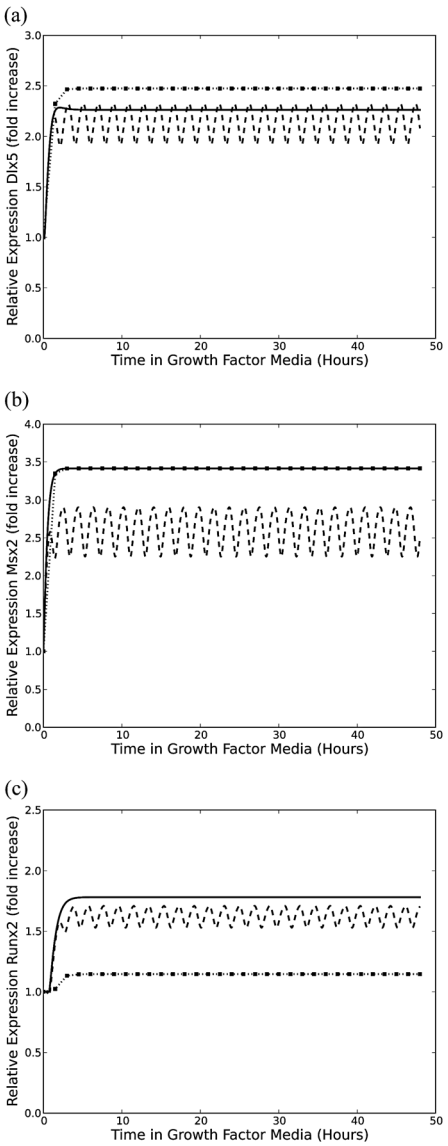


**Fig. 4** Levels of (a) Dlx5, (b) Msx2 and (c) Runx2 obtained from simulation of eqn (S10)–(S12) (see ESI†, S4) using initial conditions and parameter values summarized in Table S5 (ESI†): modelling exposure to BMP2 (unbroken line), exposure to TGFβ1 (broken line) and exposure to both BMP2 and TGFβ1 (dotted line and square symbol). Simulation results were normalized against simulation results modelling control medium.

Dlx5 and Msx2 are upstream regulators of Runx2,[10,44] however it is not readily observable from the experimental data.

In summary, based on a literature search, 243 possible GRNs of osteogenesis were proposed. By comparing Boolean models to experimental data, the number of possible networks were reduced to three. Simulations of the corresponding systems of ODEs show sustained oscillation in one of the three networks and indicate that Runx2 expression starts later than Dlx5 or Msx2 expression.

### 2.3 Discussion

While stem cells have the capacity to differentiate into different cell types, the mechanisms that regulate their transformation are poorly understood. In this study three early regulators of osteogenesis were monitored over a 48 hour time period in mES cells. Previous experimental studies were used to formulate network models of the early regulation of the central osteogenic gene Runx2 (Fig. 2).[10,15,43,44,34–41,75,76] The gene expression profiles of each cell type were then monitored experimentally and the data used to determine likely network interactions.

In mES cells, the transcription factors Dlx5 and Msx2 (Fig. 1(a) and (b)) were upregulated when stimulatied with BMP2. Cultures exposed to both TGFβ1 and BMP2 did not demonstrate significant differences in expression of Dlx5 and Msx2 compared to the cultures with BMP2 only. In contrast, Runx2 expression in mES cells showed an increase at 24 hours in response to both BMP2 and TGFβ1 but not BMP2/TGFβ1 (Fig. 1(c)). The increases in expression of Dlx5 and Msx2 in mES cells in response to BMP2 is consistent with previous studies, with Dlx5 serving as a pro-osteogenic factor[10,15,36–39] and Msx2 as a negative regulator of osteogenesis.[37,40,41] There are however, some studies that suggest a pro-osteogenic role for TGFβ1.[17,38,45,46,77–79]

Significant differences between the experimental groups across the four time points can be difficult to rationalize. As a result it can be difficult to infer interactions between the genes of interest (Dlx5, Msx2 and Runx2). For these reasons, mathematical models were developed to identify those gene regulatory networks that are consistent with the experimental data.

Analysis of the mES data (Fig. 3) indicates that Runx2 upregulation in response to BMP2 is mediated by changes in Dlx5 expression caused by direct stimulation with BMP2 or a BMP2-Msx2 mediated mechanism. The observed increase in Runx2 expression in response to TGFβ1 stimulation suggests that this process is mediated by the Msx2 stimulation of Dlx5 (Fig. 3). In addition, Dlx5 acts as a negative regulator of Msx2, a gene known to be involved in the regulation of proliferation,[34,76] and could therefore mediate a reduction in cell proliferation.

While the link between Dlx5 and increases in Runx2 expression is well established,[10,15,44] the response to TGFβ1 and the possible mediation of this response through Msx2 expression suggested by the model have not previously been demonstrated.

In accordance with the view that oscillation of TFs is a possible source of heterogeneous differentiation in stem cells,[80] our model predicts sustained oscillation of TFs in response to TGFβ1 if Runx2 negatively regulates Dlx5. In fact, Runx2 oscillation has been observed in osteoblastic cells.[81] We concede though that based on our experimental data, regulation of Dlx5 by Runx2 could not be classified (see Fig. 3) and previously Dlx5 was shown to be upregulated by Runx2.[75] In our model oscillation is predicted only in response to TGFβ1 stimulation and not BMP2 stimulation. The osteoinductive role of BMP2 is well established[36–39] and TGFβ1 is considered pro-ostogenic as well.[17,38,45,46] However there are studies that have indicated that TGFβ1 is a negative regulator of osteogenesis.[12,15]

In this paper a general methodology was used to identify candidate GRNs (Fig. 3) of early osteogenesis using experimental data from mES cells (Fig. 1). Expression of mRNAs in mouse primary bone cells were also measured (see Fig. 5) and using the same methodology the GRNs shown in Fig. 6 were obtained. GRNs corresponding to mES cells were different from GRNs corresponding to mouse primary bone cells (see Fig. 3 and 6). Whereas in mES cells Dlx5 inhibits Msx2, in the predicted GRN for primary bone cells Dlx5 activates Msx2. Similarly, Msx2 inhibits Runx2 in mES cells, but activates it in primary bone cells. In addition oscillatory behaviour was only possible in GRNs corresponding to mES cells. Our results indicate that the interaction between TFs changes as cells differentiate into a mature phenotype.

## 3 Experimental

### 3.1 Cell culture

The mouse ES cell line denoted columnar epiblast epithelium (CEE) (52, 53). Cells were cultivated on feeder layers consisting of a mitomycin-C inactivated (Calbiochem) mouse embryonic cell line (SNL). Medium consisted of Dulbeccos modified Eagles medium (DMEM) (Gibco) supplemented with 10% (v/v) foetal bovine serum (Invitrogen), 2mM L-glutamine (Invitrogen), 0.1M 2-mercaptoethanol (Sigma-Aldrich), 50 g $ml^{-1}$ penicillin, 50 g $ml^{-1}$ streptomycin (Invitrogen) and 5000 U $ml^{-1}$ leukaemia inhibitory factor (Calbiochem).

Primary mouse calvaria cells (bone) were extracted from CD1 neonatal mice. An incision was made through the skin of the scalp, the calvaria were then located and excised from the skull, ensuring that no soft tissue was attached. The individual calvaria were then washed in phosphate buffered saline (PBS) with 2% Antibiotic/Antimicrobial (AA) (Sigma-Aldrich), and digested at 37 °C for 45 minutes in a 1.4 mg $ml^{-1}$ Collagenase I-A and 0.5 mg $ml^{-1}$ Trypsin II-S made up in AlphaMEM (Gibco). The resulting digest was passed through a 0.7 m cell strainer, centrifuged at 250*g* for 10 minutes, re-suspended in complete media and seeded into T75 culture flaks (3 flasks per litter). The cells were further expanded and cyropreserved in FBS containing 10% dimethylsulphoxide. After resuscitation cells were cultured until they reached 80–90% confluence at which point they were used for experimentation.

Each cell type was seeded into 10 $cm^2$ cell culture plastic well plates at a seeding density of 10 000 cells per $cm^{-2}$. Cells were left overnight in medium consisting of MEM (DMEM for mES and alpha MEM for bone cells) with 10% (v/v) foetal bovine serum (Invitrogen), 2 mM L-glutamine (Invitrogen), 0.1 M 2-mercaptoethanol (Sigma-Aldrich), 50 g $ml^{-1}$ penicillin, 50 g $ml^{-1}$ streptomycin (Invitrogen). Four experimental

 

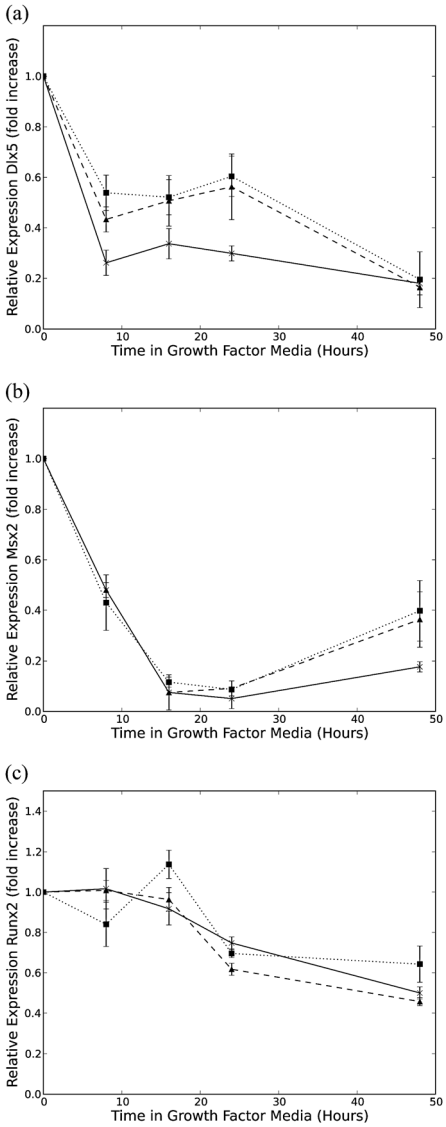


**Fig. 5** Quantitative expression of Dlx5 (a), Msx2 (b) and Runx2 (c) in mouse primary bone cells, after exposure to BMP2 (unbroken line) TGFβ1 (broken line) and BMP2/TGFβ1 (dotted line). Expression levels are normalized as in Fig. 1.

conditions were established, each consisting of five replicates: cells alone, with no growth factors (CO); cells with 300 ng $ml^{-1}$ BMP2; with 10 ng $ml^{-1}$ TGFβ; or 300 ng $ml^{-1}$ BMP2 and 10 ng $ml^{-1}$ TGFβ1. Cultures were lysed for RNA analysis at 0, 8, 16, 24 and 48 hours after exposure to growth factors.

### 3.2 Real-Time PCR

All cell lysis and RNA extractions were carried out using a Qiagen RNeasy Mini kit as per manufacturers instructions, with an on column DNAse digestion step (Qiagen). RNA was quantified using an ND100 spectrophotometer (Nanodrop) integrity assessed with an Agilent 2100 Bioanalyzer (Agilent Technologies).

Synthesis of cDNA was achieved by use of Superscript III reverse transcriptase as per manufacturers instructions (Invitrogen). Quantitative Real-Time PCR was performed for Msx2 and Dlx5 genes by use of SYBR Green fluorescent

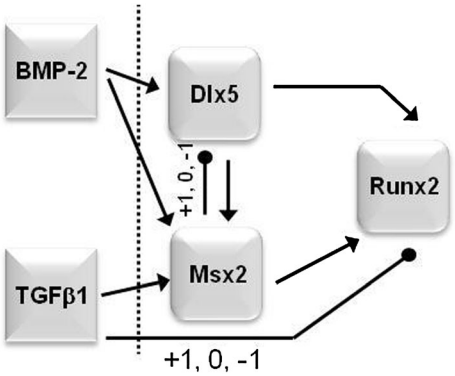


**Fig. 6** Summary of modelling results for mouse primary bone cells. Graphical representation of logic rules that matched the experimental data. Arrowheads denote activation and blunted ends inhibition, rounded arrowheads indicate that regulation is different for different matching logic rules: the possible regulations are denoted by numbers on the arrows: +1 refers to positive, −1 to negative regulation; 0 denotes no regulation.

dye (Applied Biosystems 7500). Primers were designed using Primer Express software (Applied BIosystems) and the primer sequences used were as follows: Dlx5 5′-ATGACAGGAGTGTTTGACA-GAAGAGT-3′ and 3′-GGGAACGGAGCTTGGAAGTC-5′, Msx2 5′-ACGCGGCGCAGAAAGTC-3′ and 3′-CCTCCTCAT-CCGACGAAAAC-5′. The housekeeping gene used in each case was GAPDH with primer sequences 5′-CATGGCCTTCCGT-GTTCCTA-3′ and 3′-GCGGCACGTCAGATCCA-5′. Runx2 gene expression was quantified using a TaqMan probe with commercially purchased primers for both Runx2 and GAPDH (Applied Biosystems). All results were first normalized against the housekeeping gene using δδCt analysis with a 0 time point used as a calibrator for relative expression analysis. All experimental readings (BMP2, TGFβ1 and BMP2/TGFβ1) were further normalized against control expression levels (CO) giving the fold increase (or decrease) over (or under) control expression.

### 3.3 Experimental statistics

All data sets analyzed in this study had five replicates. Gaussian distribution was assessed by use of Kolmogorov and Smirnov test. Differences data sets that had Gaussian distribution were assessed for statistical significance using a one-way analysis of variance (ANOVA) with Tukey multiple comparison test. The Bartlett method was used to assess equal standard deviations. Comparisons between non-parametric data sets were achieved by use of the Kruskal–Wallis test with the Dunn multiple comparison test.

### 3.4 Mathematical modelling – initial GRNs

The TFs Dlx5, Msx2 and Runx2 are the dependent variables in our models and the externally supplied growth factors BMP2 and TGFβ1 are the control parameters. Previous experimental studies were used to construct the GRNs shown in Fig. 2. Direct regulatory connections between the control parameters and model variables that have been reported by at least two different studies are depicted by continuous lines; additional regulatory connections identified in only one study are depicted by dashed lines.[10,15,43,44,34–41,75,76]

These regulatory connections were treated as incomplete and needing to be tested with simulations. During simulations the additional connections were allowed to be activating, inhibitory or inactive. As shown in Fig. 2, our compound GRN includes four 'fixed' and five 'uncertain' regulatory connections thus representing $3^5 = 243$ candidate GRNs.

### 3.5 Mathematical modelling – Boolean modelling

For each GRN shown in Fig. 2 a Boolean logic rule (a set of binary values relating the state of the target gene to the expression levels of its regulators) was constructed. Parameters $g_{BMP}$ and $g_{TGF}$ were control parameters with a constant binary value (0 or 1) assigned to each, depending on whether or not the corresponding growth factor was present in the culture medium (see Table S4, ESI†). Levels of Dlx5, Msx2 and Runx2 were the nodes of the network. See ESI†, S2 for detailed descriptions of how the Boolean logic rules are generated from each GRN.

Simulations were performed to identify those logic rules that are consistent with the experimental data from the mES and primary bone cells. Binary representation of experimentally obtained mRNA levels at time point 0 were used as initial conditions and on each simulation step the state of each variable was updated according to a particular logic rule. The simulations were continued until a steady state or a limit cycle was reached. A particular GRN was said to match the data if the large-time limit of the simulation matched the 24 hour experimental values.

If a simulation yielded a steady state for each choice of medium, then the result was directly compared to the binary representation of the experimental results. If a simulation yielded a limit cycle, then the values of each variable (Dlx5, Msx2 and Runx2) were averaged, and the average values for each variable were ordered according to their magnitudes. A logic rule that yielded a limit cycle was considered a 'matching result' if, for all variables, the average values associated with a medium corresponding to an experimental *on* state were higher than those corresponding to *off* state.

### 3.6 Mathematical modelling – continuous representation of Boolean logic rules

Each matching Boolean logic rule was transformed into an equivalent ODE model using the HillCube method (see ESI†, S3).[66] The system of ODEs capable of producing sustained oscillations was further analysed to obtain a parameter fit to the experimental data. Optimization was conducted with the EcsPy[73] modul in Python using a genetic algorithm.[74]

## 4 Conclusions

We have presented a general methodology to identify GRNs based on literature information and *in vitro* expression data. This approach was illustrated by identifying GRNs of early osteogenesis.

Guided by the literature involving mES cells, 243 GRNs of early osteogenesis were proposed. Using Boolean modelling three of 243 GRNs were found to be consistent with our experimental data. Experimentally untested transcriptional connections were either ruled out or suggested by our modelling results, one of the hypotheses being that Runx2 stimulation by TGFβ1 is mediated *via* Msx2. In addition, simulation of system of ODEs corresponding to the three GRNs consistent with our experimental data showed that some of these models exhibit sustained oscillation, a possible source of heterogeneous differentiation capacity in stem cells. Oscillation was only observed when addition of TGFβ1 was modelled, a growth factor that was identified as either a negative[12,15] or positive[17,38,45,46] regulator of osteogenesis.

Using this approach, GRNs were identified which were consistent with expression data of primary bone cells stimulated by BMP2 and/or TGFβ1. The obtained GRNs differ from predicted GRNs of early osteogenesis, indicating that TF interactions depend on the cells' differentiation state.

The mathematical model and the experimental data neglect numerous signalling processes (*e.g.* Smads, MAPK, Wnt signalling) and gene interactions that regulate osteogenesis. However, the accuracy of the model will increase as more experimental data are incorporated and the approach we have adopted could be adapted for any gene regulatory system.

## Acknowledgements

The BBSRC/EPSRC for funding project number BB/D008522/1. This publication was based on work supported in part by Award No. KUK-C1-013-03, made by King Abdullah University of Science and Technology (KAUST).

## References

1 L. Buttery, J. X. S. Bourne, H. Wood, F. Hughes, S. Hughes, V. Episkopou and J. Polak, *Tissue Eng.*, 2001, **7**(1), 89–99.
2 T. Doetschman, H. Eistetter, M. Katz, W. Schmidt and R. Kemler, *J. Embryol. Exp. Morphol.*, 1985, **87**, 27–45.
3 A. Smith, *Curr. Biol.*, 1998, **8**, R802–R804.
4 J. Thomson, J. Itskovitz-Eldor, S. Shapiro, M. Waknitz, J. Swiergiel, V. Marshall and J. Jones, *Science*, 1998, **282**, 1145–1147.
5 M. Weiss and S. Orkin, *J. Clin. Invest.*, 1996, **97**, 591–595.
6 N. zur Nieden, G. Kempka, D. Rancourt and H. Ahr, *BMC Dev. Biol.*, 2005, **5**, 1.
7 L. Bonewald and G. Mundy, *Clin. Orthop. Relat. Res.*, 1990, **250**, 261–276.
8 S. Cheifetz, I. Li, C. McCulloch, K. Sampath and J. Sodek, *Connect. Tissue Res.*, 1996, **35**, 71–78.
9 R. Harland, *Proc. Natl. Acad. Sci. U. S. A.*, 1994, **91**, 10243–10246.
10 M. Hassan, R. Tare, S. Lee, M. Mandeville, M. Morasso, A. v. W. A. Javed, J. Stein, G. Stein and J. Lian, *J. Biol. Chem.*, 2006, **281**(52), 40515–40526.
11 S. Hassel, S. Schmitt, A. Hartung, M. Roth, A. Nohe, N. Petersen, M. Ehrlich, Y. Henis, W. Sebald and P. Knaus, *J. Bone Jt. Surg., Am. Vol.*, 2003, **85–A**(Suppl 3), 44–51.
12 T. Katagiri, A. Yamaguchi, M. Komaki, E. Abe, N. Takahashi, T. Ikeda, V. Rosen, J. Wozney, A. Fujisawa-Sehara and T. Suda, *J. Cell Biol.*, 1994, **127**, 1755–1766.
13 M. Kawabata, T. Imamura and K. Miyazono, *Cytokine Growth Factor Rev.*, 1998, **9**, 49–61.
14 M. Lee, Y. Kim, H. Kim, H. Park, A. Kang, H. Kyung, J. W. J. H. Sung, H. Kim and H. Ryoo, *J. Biol. Chem.*, 2003, **278**(36), 34387–34394.
15 M. Lee, T. Kwon, H. Park, J. Wozney and H. Ryoo, *Biochem. Biophys. Res. Commun.*, 2003, **309**, 689–694.
16 I. Li, S. Cheifetz, C. McCulloch, K. Sampath and J. Sodek, *J. Cell. Physiol.*, 1996, **169**, 115–125.
17 A. Roberts, K. Flanders, P. Kondaiah, N. Thompson, E. Van Obberghen-Schilling, L. Wakefield, P. Rossi, B. De Crombrugghe, U. Heine and M. Sporn, *Recent Prog. Horm. Res.*, 1988, **44**, 157–197.
18 L. Shum, X. Wang, A. Kane and G. Nuckolls, *Int. J. Dev. Biol.*, 2003, **47**, 423–432.

 

19 E. Wang, D. Israel, S. Kelly and D. Luxenberg, *Growth Factors*, 1993, **9**, 57–71.
20 B. Yoon and K. Lyons, *J. Cell. Biochem.*, 2004, **93**, 93–103.
21 J. Choi, J. Pratap, A. Javed, S. Zaidi, L. Xing, E. Balint, S. Dalamangas, B. Boyce, A. Van Wijnen and J. Lian, *et al.*, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 8650–8655.
22 P. Ducy, R. Zhang, V. Geoffroy, A. Ridall and G. Karsenty, *Cell*, 1997, **89**, 747–754.
23 T. Komori, H. Yagi, S. Nomura, A. Yamaguchi, K. Sasak, K. Deguchi, Y. Shimizu, R. Bronson, Y. Gao, M. S. M. Inada, R. Okamoto, Y. Kitamura, S. Yoshiki and T. Kishimoto, *Cell*, 1997, **89**, 755–764.
24 B. Lee, K. Thirunavukkarasu, L. Zhou, L. Pastore, A. Baldini, J. Hecht, V. Geoffrey, P. Ducy and G. Karsenty, *Nat. Genet.*, 1997, **16**, 307–310.
25 S. Mundlos, F. Otto, C. Mundlos, J. Mulliken, A. Aylsworth, S. Albright, D. Lindhout, W. Cole, W. Henn and J. Knoll, *et al.*, *Cell*, 1997, **89**, 773–779.
26 F. Otto, A. Thornell, T. Crompton, A. Denzel, K. Gilmour, I. Rosewell, G. Stamp, R. Beddington, S. Mundlos, B. Olsen, P. Selby and M. Owen, *Cell*, 1997, **89**, 765–771.
27 B. Byers, G. Pavlath, T. Murphy, G. Karsenty and A. Garca, *J. Bone Miner. Res.*, 2002, **17**, 1931–1944.
28 V. Geoffroy, P. Ducy and G. Karsenty, *J. Biol. Chem.*, 1995, **270**, 30973–30979.
29 H. Harada, S. Tagashira, M. Fujiwara, S. Ogawa, T. Katsumata, A. Yamaguchi, T. Komori and M. Nakatsuka, *J. Biol. Chem.*, 1999, **274**, 6972–6978.
30 S. Yang, D. Wei, D. Wang, M. Phimphilai, P. Krebsbach and R. Franceschi, *J. Bone Miner. Res.*, 2003, **18**, 705–715.
31 H. Hoffmann, K. Catron, A. Van Wijnen, L. McCabe, J. Lian, G. Stein and J. Stein, *Proc. Natl. Acad. Sci. U. S. A.*, 1994, **91**, 12887–12891.
32 H. Ryoo, H. Hoffmann, T. Beumer, B. Frenkel, D. Towler, G. Stein, J. Stein, A. Van Wijnen and J. Lian, *Mol. Endocrinol.*, 1997, **11**, 1681–1694.
33 D. Towler, S. Rutledge and G. Rodan, *Mol. Endocrinol.*, 1994, **8**, 1484–1493.
34 M. Lee, Y. Kim, W. Yoon, J. Kim, B. Kim, Y. Hwang, J. Wozney, X. Chi, S. Bae, K. Choi, J. Cho, J. Choi and H. Ryoo, *J. Biol. Chem.*, 2005, **280**, 35579–35587.
35 T. Takahashi, S. Kato, N. Suzuki, N. Kawabata and M. Takagi, *J. Oral. Sci.*, 2005, **47**(4), 199–207.
36 S. Cheng, J. Shao, N. Charlton-Kachigian, A. Loewy and D. Towler, *J. Biol. Chem.*, 2003, **278**(48), 45969–45977.
37 F. Ichida, R. Nishimura, K. Hata, T. Matsubara, F. Ikeda, K. Hisada, H. Yatani, X. Cao, T. Komori, A. Yamaguchi and T. Yoneda, *J. Biol. Chem.*, 2004, **279**(32), 34015–34022.
38 K. Lee, H. Kim, Q. Li, X. Chi, C. Ueta, T. Komori, J. Wozney, E. Kim, J. Choi, H. Ryoo and S. Bae, *Mol. Cell. Biol.*, 2000, **20**(23), 8783–8792.
39 K. Miyama, G. Yamada, T. Yamamoto, C. Takagi, K. Miyado, M. Sakai, N. Ueno and H. Shibuya, *Dev. Biol.*, 1999, **208**, 123–133.
40 K. Shirakabe, K. Terasawa, K. Miyama, H. Shibuya and E. Nishida, *Genes Cells*, 2001, **6**, 851–856.
41 T. Yoshizawa, F. Takizawa, F. Iizawa, O. Ishibashi, H. Kawashima, A. Matsuda, N. Endo and H. Kawashima, *Mol. Cell. Biol.*, 2004, **24**(8), 3460–3472.
42 S. Seyedin, A. Thompson, H. Bentz, D. Rosen, J. McPherson, A. Conti, N. Siegel, G. Galluppi and K. Piez, *J. Biol. Chem.*, 1986, **261**, 5693–5695.
43 S. Brunelli, E. Tagliafico, F. De Angelis, R. Tonlorenzi, S. Baesso, S. Ferrari, M. Niinobe, K. Yoshikawa, R. Schwartz and I. Bozzoni, *et al.*, *Circ. Res.*, 2004, **94**, 1571.
44 M. Hassan, A. Javed, M. Morasso, J. Karlin, M. Montecino, A. van Wijnen, G. Stein, J. Stein and J. Lian, *Mol. Cell. Biol.*, 2004, **24**, 9248–9261.
45 M. Lee, A. Javed, H. Kim, H. Shin, S. Gutierrez, J. Choi, V. Rosen, J. Stein, A. van Wijnen, G. Stein, J. Lian and H. Ryoo, *J. Cell. Biochem.*, 1999, **73**, 114–125.
46 J. Massague, *Annu. Rev. Cell Biol.*, 1990, **6**, 597–641.
47 M. Urist, *Science*, 1965, **150**, 893–899.
48 T. Lee, N. Rinaldi, F. Robert, D. Odom, Z. Bar-Joseph, G. Gerber, N. Hannett, C. Harbison, C. Thompson, I. Simon, J. Zeitlinger, E. Jennings, H. Murray, D. Gordon, B. Ren, J. Wyrick, J. Tagne, T. Volkert, E. Fraenkel, D. Gifford and R. Young, *Science*, 2002, **298**, 799–804.
49 J. Vohradsky, *J. Biol. Chem.*, 2001, **276**, 36168–36173.
50 A. Arkin, J. Ross and H. McAdams, *Genetics*, 1998, **149**, 1633–1648.
51 W. Blake, M. Kærn, C. Cantor and J. Collins, *Nature*, 2003, **422**, 633–637.
52 J. Bower and H. Bolouri, *Computational modeling of genetic and biochemical networks*, The MIT Press, 2004.
53 D. Chu, N. Zabet and B. Mitavskiy, *J. Theor. Biol.*, 2009, **257**, 419–429.
54 M. Elowitz and M. Leibler, *Nature*, 2000, **403**, 335–338.
55 M. Elowitz, A. Levine, E. Siggia and P. Swain, *Science*, 2002, **297**, 1183–1186.
56 T. Gardner, C. Cantor and J. Collins, *Nature*, 2000, **403**(6767), 339–342.
57 N. Geard and J. Wiles, *Artificial Life*, 2005, **11**, 249–267.
58 S. Kauffman, *J. Theor. Biol.*, 1969, **22**, 437–467.
59 Y. Kaznessis, *BMC Syst. Biol.*, 2007, **1**, 47.
60 J. Knabe, C. Nehaniv, M. Schilstra and T. Quick, Proceedings of the Artificial Life X Conference, 2006, pp. 15–21.
61 J. Knabe, M. Schilstra and C. Nehaniv, Artificial Life XI: Proceedings of the Eleventh International Conference on the Simulation and Synthesis of Living Systems, 2008, pp. 321–328.
62 R. Leclerc, *Mol. syst. Biol.*, 2008, **4**, 213.
63 J. Raser and E. O'Shea, *Science*, 2005, **309**, 2010–2013.
64 A. Ribeiro, R. Zhu and S. Kauffman, *J. Comput. Biol.*, 2006, **13**, 1630–1639.
65 A. Ribeiro and J. Lloyd-Price, *Bioinformatics*, 2007, **23**, 777.
66 D. Wittmann, J. Krumsiek, J. Saez-Rodriguez, D. Lauffenburger, S. Klamt and F. Theis, *BMC Syst. Biol.*, 2009, **3**, 98.
67 L. Mendoza, D. Thieffry and E. Alvarez-Buylla, *Bioinformatics*, 1999, **15**, 593.
68 Y. Sánchez-Corrales, E. Álvarez Buylla and L. Mendoza, *J. Theor. Biol.*, 2010, **264**, 971–983.
69 F. Li, T. Long, Y. Lu, Q. Ouyang and C. Tang, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 4781.
70 M. Davidich and S. Bornholdt, *PLoS One*, 2008, **3**, 1672.
71 A. Fauré, A. Naldi, C. Chaouiya and D. Thieffry, *Bioinformatics*, 2006, **22**, e124.
72 R. Albert and H. Othmer, *J. Theor. Biol.*, 2003, **223**, 1–18.
73 T. Back, F. Hoffmeister and H.-P. Schwefel, A survey of evolution strategies, in *Proceedings of the 4th International Conference on Genetic Algorithms*, ed. R. K. Belew and L. B. Booker, Morgan Kaufman, 1991, pp. 2–9.
74 J. H. Holland, *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*, University of Michigan Press, Ann Arbor, 1975, viii, p. 183.
75 C. Gersbach, B. Byers, G. Pavlath and A. García, *Exp. Cell Res.*, 2004, **300**, 406–417.
76 N. Holleville, S. Matéos, M. Bontoux, K. Bollerot and A. Monsoro-Burq, *Dev. Biol.*, 2007, **304**(2), 860–874.
77 L. Bonewald and S. Dallas, *J. Cell. Biochem.*, 1994, **55**, 350–357.
78 B. Boyan, Z. Schwartz, S. Park-Snyder, D. Dean, F. Yang, D. Twardzik and L. Bonewald, *J. Biol. Chem.*, 1994, **269**, 28374–28381.
79 S. Harris, L. Bonewald, M. Harris, M. Sabatini, S. Dallas, J. Feng, N. Ghosh-Choudhury, J. Wozney and G. Mundy, *J. Bone Miner. Res.*, 1994, **9**, 855–863.
80 T. Graf and M. Stadtfeld, *Cell Stem Cell*, 2008, **3**, 480–483.
81 M. Galindo, J. Pratap, D. Young, H. Hovhannisyan, H. Im, J. Choi, J. Lian, J. Stein, G. Stein and A. Van Wijnen, *J. Biol. Chem.*, 2005, **280**(21), 20274–20285.