Dissecting ensemble networks in ES cell populations reveals micro-heterogeneity underlying pluripotency†

Jamie Trott, a Katsuhiko Hayashi, b Azim Surani, b M. Madan Babu c and Alfonso Martinez-Arias a

Received 26th September 2011, Accepted 16th November 2011
DOI: 10.1039/c1mb05398a

Analysis of transcription at the level of single cells in prokaryotes and eukaryotes has revealed the existence of heterogeneities in the expression of individual genes within genetically homogeneous populations. This variation is an emerging hallmark of populations of Embryonic Stem (ES) cells and has been ascribed to the stochasticity associated with the biochemical events that mediate gene expression. It has been suggested that these heterogeneities play a role in the maintenance of pluripotency. However, for the most part, studies have focused on individual genes in large cell populations. Here we use an existing dataset on the expression of eight genes involved in pluripotency in eighty-three ES cells to create Gene Regulatory Networks (GRNs) at the single cell level. We observe widespread heterogeneities in the expression of the eight genes, but analysis of correlations within individual cells reveals three distinct classes centered on the expression of Nanog, a marker of pluripotency, and Fgf5, a gene associated with differentiation: high levels of Nanog and low levels of Fgf5, low levels of Nanog and high levels of Fgf5, and low levels of both. Each of these classes is associated with a collection of active sub-networks, with differing degrees of connectivity between their elements, which define a cellular state: self-renewal, primed for differentiation or transition between the two. Though every cell should be governed by the same network, the active sub-networks may emerge due to considerations such as variation in (i) the expression level of active transcription factors (e.g. through post-translational modification or ligand/co-factor availability) or (ii) access to the target gene locus (e.g. via changes in chromatin status or epigenetic modifications). We conclude that heterogeneities in gene expression should not be interpreted as representing different states of a single unique network, but as a reflection of the activity of different sub-networks in sub-populations of cells.

Introduction

Regulatory interactions between transcription factors and genes involved in particular processes are represented by Gene Regulatory Networks (GRNs). In such a representation, nodes denote either transcription factors or target genes and a directed link denotes the binding of a transcription factor to the promoter of a target gene. The binding of a transcription factor to the promoter regions of target genes is typically inferred through experimental measurements performed on a population of genetically homogeneous cells. In some cases it is possible to synchronize the state of the cells but, for practical reasons, variations in the cell state (e.g., expression levels of transcription factors) are generally ignored in the analysis. Such approaches assume homogeneity in expression levels of genes and, more generally, in the cellular states of members of a cell population, and hence should be applicable only in situations where these assumptions are justified (Fig. 1A).

While several systems conform to this premise, there is increasing evidence of heterogeneity in gene expression levels across individuals of genetically homogeneous cell populations. In some instances this heterogeneity has a dynamic component and this will impact on the regulatory networks that are inferred from measurements in these cell populations. Thus, reconstructing a single gene regulatory network based on average population measurements may not provide the true underlying regulatory network but only an ensemble average network (Fig. 1B). These considerations raise a fundamental question: are all individuals from a cell population that display non-genetic heterogeneities in gene expression levels governed by a single, unique gene regulatory network? In other words, does the non-genetic heterogeneity arise because distinct sub-populations of cells exhibit different states of the same gene.
regulatory network? Here we use mouse embryonic stem (ES) cells as a system to analyze this question and suggest that different sub-populations of cells are governed by distinct active sub-networks (Fig. 1B). We propose that the heterogeneities in gene expression in a cell population should not be interpreted as the variable output of a single “averaged” network (i.e., representing different states of a multi-stable system that is governed by a single unique network), but as a reflection of the operation of different active sub-networks in sub-populations of cells. These sub-networks may arise due to considerations such as altered access to the target gene locus (e.g. epigenetic barriers) or due to variation in the expression level of the active state of transcription factors (e.g. fluctuations in levels of co-factor availability).

ES cells are a genetically homogeneous cell population derived from early mammalian embryos which can be induced to differentiate into all cell types of an organism i.e. they are pluripotent.20,21,22 Studies with mouse ES cells have shown that a growing population exhibits heterogeneities in protein and gene expression levels that are linked to a dynamic equilibrium between states of self-renewal and differentiation.13,15,23–26 Genetic analysis of pluripotency has identified a small group of transcription factors, centered around Sox2, Oct4 and Nanog (SON), which are essential for the establishment and maintenance of the pluripotent state in ES cells and early embryos.27–31 In particular Nanog, a homeobox containing transcription factor, plays a central role in the transcriptional network governing pluripotency. In the absence of Nanog, the pluripotent state is not established and, after establishment, at least in ES cells, the degree of pluripotency correlates with the dosage of Nanog: in the absence of Nanog the equilibrium between self-renewal and differentiation is biased towards differentiation, whereas increasing the levels of Nanog favors self-renewal.13,32–35 There are other genes associated with pluripotency, e.g. Rex1, Stella and Gbx2, but their respective roles are unclear.25,26,36–39 In addition, a balance of growth factor mediated signaling is also required to maintain a population of ES cells in the pluripotent state.29,40,41

Results and discussion

Data on single cell measurements of key stem cell factors

Attempts to uncover the GRNs involved in pluripotency have produced a collection of very similar networks, centered on Sox2, Oct4 and Nanog, with varying degrees of involvement of other factors.27,28,42–53 These networks, which are typically derived from ChIP-Seq or ChIP-chip experiments on a population of cells, will be referred to as “structural” networks. They highlight “average” regulatory interactions, and do not reveal quantitative relationships between genes that operate in any one cell from a cell population. Given the degree of heterogeneity in the expression of many of these genes (see the bimodal distribution of expression levels of the pluripotency genes in Fig. S1, ESI†), the network of physical interactions between the transcription factor and promoter of the target genes, and the strength of interaction should be interpreted with caution as they represent ensemble averages of different sub-networks operating in individual cells54 (Fig. 1B). It would be of fundamental importance to access these individual networks and compare them to the average.
Recent advances in single-cell transcriptomics allow quantitative measurements of the levels of particular mRNAs in single cells.\textsuperscript{55-58} Such measurements could be useful in trying to relate ensemble behaviour to that of the individual elements of a population of cells. Here we explore these issues by using the expression levels of eight genes involved in pluripotency in 83 individual mouse ES cells\textsuperscript{25} (Fig. S1, ESI†). The genes studied comprise elements of the pluripotency network (Sox2, Nanog, Oct4, Rex1, Gbx2, Stella, Pecam1)\textsuperscript{25,26,36–39,59–64} and a marker of differentiation (Fgf5).\textsuperscript{65–67} Transcript levels were obtained for each gene in each cell as described previously.\textsuperscript{25,55,56} Transcript numbers for each gene were normalized by dividing the individual transcript levels by the levels of Gapdh (glyceraldehyde 3-phosphate dehydrogenase) transcripts from the respective cells to account for technical variations.

**Defining sub-population of cells**

When the dataset is averaged over the whole population or over large groups of cells, the expression levels of pluripotency genes correlate well with each other: high, medium and low levels of Nanog correspond to the same levels of, for example, Sox2 or Oct4. However, as the bins used to carry out the averaging become narrower, the correlations between many of the genes become weaker (Fig. S2, ESI†). Ultimately, at the level of single cells, a degree of variation arises in the overall correlations, but individual relationships emerge. We ordered the 83 cells individually from higher to lower levels of Nanog expression and observe some distinct groups of cells in terms of expression of other pluripotency markers (Fig. 2; Fig. S2, ESI†).

We first defined a Nanog threshold level as the minimum level of Nanog expression, above which cells were never observed to express Fgf5 (Fig. 2). At high levels of Nanog expression (HN: High Nanog cells) there is little variation in the expression of many of the other pluripotency genes. However, below the threshold level of Nanog expression (LN: Low Nanog cells—see Methods summary and Fig. 2), simple correlations are lost and one can observe large cell-to-cell variability in the levels of the different genes.

This is particularly clear in the cases of Fgf5 and Rex1. Fgf5 is, by definition, only expressed below the threshold level of Nanog, but further reductions in the level of Nanog do not affect the probability of a cell expressing Fgf5. Interestingly, immediately below the same threshold level of Nanog expression, cells can dispense with Rex1 expression and also (but less frequently) Sox2, in an apparently stochastic manner (Fig. 2). Furthermore, although there is no statistically significant correlation between Fgf5 and Nanog levels, expression of Fgf5 is often associated with low levels of Rex1, and less frequently Sox2, in cells in the LN region (Fig. 2). Indeed, whilst Fgf5 expression was found to anti-correlate, at least weakly, with all of the other genes we looked at, by far the strongest predictors of Fgf5 expression are low levels of Rex1 (Spearman’s coefficient $\rho = -0.47$, $P = 6.3 \times 10^{-3}$) (see Fig. S3A, ESI†).

At the single cell level, no other qualitative relationships could be found between the levels of any of the other genes studied (Fig. S3, ESI†). These observations are consistent with the proposal that low levels of Nanog provide a substrate for differentiation and reveal an element of stochasticity in the decision of a cell to express or not to express Fgf5. They also suggest that Rex1 may play a role in regulating differentiation. Thus, we qualitatively observe three sub-populations within an ES cell population: High Nanog, Fgf5 negative (HN: Fgf5$^-$); Low Nanog, Fgf5 negative (LN: Fgf5$^-$) and Low Nanog, Fgf5 positive (LN: Fgf5$^+$), each corresponding to a cell state with low Rex1 being a key element in the LN: Fgf5$^+$ state.

**Principal component analysis (PCA) of transcript levels**

Principal component analysis (PCA) of transcript levels of all 8 genes in individual cells allows us to observe how similar the expression profile of any cell is to any other cell and reveals...
different degrees of heterogeneity between the sub-populations (Fig. 3). PCA suggests that the High Nanog Fgf5− and Low Nanog Fgf5+ cells are largely separate populations, whilst the Low Nanog Fgf5− cells appear to be in a state that overlaps those two distinct states. Removing Nanog from the PCA does not affect the results significantly (data not shown) and each of the genes contributes approximately equally to the separation of the different cells along each axis. This indicated that the High Nanog Fgf5− and Low Nanog Fgf5+ are truly discrete sub-populations. To determine whether particular patterns of gene expression underlie these trends we tested the degree to which the transcriptional activity of individual cells conforms to an ensemble network derived from the correlations observed between different genes across the entire dataset (Fig. 4).

Deciphering active sub-networks in sub-population of cells

The quantitative and single-cell nature of the expression data allows us to elucidate “co-expression networks” in which genes (nodes) with correlated activity are joined by edges that contain quantitative information about the correlation of expression across the individual cells. While correlation does not imply causation, causal links between genes will be correlated and hence one can potentially infer regulatory relationships from such a network. This representation contrasts with the commonly used “structural networks”, which is a reconstruction of regulatory interactions from a mixture of results of protein-DNA interactions and gene expression using whole populations of cells. Often, the structural networks are highly interconnected making their interpretation difficult as they do not contain information about the dynamics of the participating nodes in individual cells.

Here we first generated an average co-expression network for the genes under study (Fig. 4A). For comparison, we reconstructed the corresponding structural network from published data68 (Fig. 4B). The co-expression network contains some of the edges in the corresponding structural network (representing causal relationships), as well as some additional edges (denoting non-causal correlations). Each type of the “average” ensemble network has unique information and both reveal a hierarchical structure which might have biological significance. By mapping the expression levels of the genes from individual cells onto the “average” structural network, we inferred the active structural sub-network that is likely to operate in each cell (see Methods). When this procedure is applied to different sub-populations, it is possible to see that they are governed by distinct sub-networks, with unique distinguishing features (Fig. 4B and C).

Analysis of the networks in the sub-populations reveals that the heterogeneities are probably a result of different network connectivities: every sub-population of cells is characterized by a distinct active structural sub-network, which is reflected in the corresponding co-expression network (Fig. 4B, C, and 5). Furthermore this analysis reveals some dominant patterns, which allow rationalization of the structural network and provide some intuition for its activity (Fig. 5). In particular, a tetrad of Sox2, Oct4, Nanog and Rex1 emerges as a common...
motif in that it is interconnected and active in all cells with different levels of gene activity in different sub-populations. The activity of this tetrad exhibits striking differences between cells that express Fgf5 and those that do not, with Rex1 emerging as a connector of the SON triad to Fgf5: for Fgf5 to be on, Rex1 has to be off (Fig. 5).

Cells expressing Fgf5 are likely to be in an epistem cell (EpiSC) state, where cells have been shown to express low levels of pluripotency markers\(^{65,67}\) and this is the case here. Our analysis also suggests that the LN Fgf5\(^+\) population is a mixture of cells that sample distinct parts of the same structural network which span the other two states. We also noticed...
that when cells express high levels of Nanog, the connectivity between different genes in the structural network is at its highest (Fig. 4B) and, as a consequence, the state of the network—the relative levels of the genes within it—is at its most stable. On the other hand, cells that express low levels of Nanog, but which have not yet differentiated, express lower levels of the core transcription factors and have a higher probability of losing expression of one or more of the pluripotency genes, as might be expected. This perspective contrasts with a deterministic view of the network, whereby a gradual and consistent reduction in the expression levels of all pluripotency network elements relieves repression of differentiation genes and differentiation is associated with a single network state.

It has been suggested that the heterogeneous expression of pluripotency related genes, characteristic of ES cells, reflects the dynamics of the underlying gene regulatory networks and might be the essence of the pluripotent state.\(^{14}\) In this view, the observed variations in the levels of gene and protein expression reflect fluctuations in the dynamics of the network. Our analysis suggests that these fluctuations are likely to occur but that they might reflect not just changes in the kinetics of the network, but the existence of different strengths of connectivity between the nodes, which in different cells (or sub-population of cells) lead to different structural sub-networks in operation.\(^{54}\) As long as the output is attained, whether HN or expression of Fgf5, the system will allow for the variability to be present and tolerated. This also implies that a particular cell state is not the result of a single GRN, but of an active sub-network, which can produce the output and operate in individual cells of a sub-population. While every cell should be governed by the same network, the active sub-networks may emerge due to considerations such as variation in (i) the expression level of active transcription factors (e.g., through post-translational modifications or ligand/co-factor availability) or (ii) access to the target gene locus (e.g., via changes in chromatin status or epigenetic modifications). The structure of the LN:Fgf5\(^{-}\) population is interesting in that it seems to be intermediate and might represent individual cells which do have the option to return to the HN level. Taken together, these observations suggest that distinct parts (with varying strengths for an interaction) of the same structural network are being sampled by the different sub-populations and that this may correspond to the distinct states that one observes in the sub-population of cells.

An important consequence of our observations is that the heterogeneities in the expression of the previously described genes might not only be a simple consequence of the performance of the ensemble network, but might also reflect different “connectivity” of the elements of the structural network in

---

**Fig. 5** Network activity in single mES cells. (A) Simplification of the co-expression network. The co-expression network was progressively simplified so that patterns in gene expression across single cells become apparent. The expression levels of Nanog, Oct4 and Sox2 (NOS), and Gbx2, Pecam1 and Stella (GPS) in each cell were averaged (middle) as our analysis suggested that genes within these two sets might have related expression patterns. Rex1 was not included in the averaging as our analysis suggests that the Rex1 expression level is the best predictor of the cell state. Fgf5 was not included as it is the sole marker of epiblast differentiation. Fgf5 was omitted from the final network structure (bottom and 5B) to simplify pattern recognition. The expression levels of Rex1 and Fgf5 and the average expression levels of the two groups (NOS and GPS) are color-coded according to the scale (red and blue signify high and low levels of expression, respectively). (B) Network activity in single cells. Cells were ordered according to decreasing Rex1 expression from left to right and top to bottom. Cells that express Fgf5 are highlighted in green. Note that the level of NOS falls consistently with the level of Rex1, whilst the level of GPS remains fairly stable down to a certain level of Rex1 before dropping off in an apparently stochastic manner. Though the overall level of the GPS module, defined as the mean level of the three genes within each cell, remains stable, the individual genes that make up the GPS module show considerable variability even at high levels of Rex1.
individual cells, i.e. what varies between the cells and determines the heterogeneity is the connectivity and its strength.

Identification of Rex1 as a key player that connects to the Sox2-Oct4-Nanog network

We also observe the existence of patterns, which can be treated as elements in larger networks. The Sox2, Oct4 and Nanog (SON) triad, which has been identified as a central element in pluripotency, is certainly a key motif in pluripotency, which we would like to suggest be called the pluripotency motif. Rex1, a zinc finger containing protein, is a connector between the SON triad and its strength of interaction with other genes might be the determinant of the cell’s state. The structural and co-expression sub-networks (reconstructed from single-cell measurements) that operate in distinct sub-populations of cells thus provide high-resolution information about the state of a cell when compared to the ensemble networks obtained from interrogating cell populations.

The identification of Rex1 as a connector between the SON triad and differentiation, as represented by the Fgf5 expression level, highlight the importance of the single cell analysis. This connection also provides an explanation for a paradox in the literature. Rex1 expression fluctuates in an ES cell population and Rex1 negative cells have been shown to contribute less to chimeras whereas Rex1 positive cells do; surprisingly, Rex1 mutant cells do contribute to chimeras. Our results suggest that in the wild-type the Rex1 negative cells are in an epiblast state and epiblast cells do not contribute efficiently to chimeras. The ability of Rex1 mutant cells to contribute to chimeras probably reflects that Rex1 is not the main determinant of the state of the cells and a population of Rex1 mutant cells clearly contains the same micro-heterogeneities in gene expression of certain as yet uncharacterised factors as in the wild type cells. Thus a population of Rex1 mutant cells contains cells in the ES cell state, which can therefore contribute to chimeras.

Implications

Measurements of gene expression in embryonic stem (ES) cells have revealed transcriptional heterogeneities at the level of single cells. Our analysis suggests that these heterogeneities arise from the activity of different gene regulatory networks in different cells. We suggest that the transcriptional state of embryonic stem cells is not a deterministic property, homogeneous across a cell population, but that it needs to be considered as an ensemble property of the population, with individual cells differing in the connectivity of specific networks. In this view different connectivities can lead to the same phenotype. This situation is reminiscent of the relationship established in statistical physics between a macrostate and an ensemble of microstates: a macrostate refers to the value of a state function (e.g. temperature) that can be attained through, or is compatible with, a number of different configurations of the component elements of a system (atoms, molecules, etc.), and each of these configurations would correspond to a microstate.69 In our analysis, the macrostates would be the phenotypes, e.g. High Nanog, correlated with pluripotency, Fgf5+, correlated with differentiation, and LN:Fgf5−, which corresponds to the transition state between them respectively.

Each of these states (the equivalent of a macrostate) is represented by a collection of related but differentially connected GRNs (the microstates compatible with each macrostate). It is interesting that in LN:Fgf5− cells, there is a large variation, as if in this state the cells explore possible network connections that could lead them into one or another state.

Taken together, these observations highlight the importance of single cell analysis in regulatory network inference and the need to take into consideration statistical properties of single cell ensembles when investigating developmental processes.

Materials and methods

Methods summary

We obtained transcript numbers for 9 genes in 83 individual mESCs from Hayashi et al.25 where the authors performed RT-PCR on single cells.55,56 Briefly, cells were grown under conditions that promote pluripotency (DMEM/F12, Invitrogen; supplemented with 15% fetal calf serum, 1 mM glutamine, 100 u ml⁻¹ penicillin/streptomycin, 0.1 mM β-mercaptoethanol, and 1000 u ml⁻¹ LIF on mitomycin C-treated MEFs) and the expression level distributions of those genes in this cell population were assessed. When cells were ordered according to Nanog expression, it became apparent that below a certain level of Nanog expression, certain genes exhibit different behavior. Specifically, cells are able to express the Epi-SC marker Fgf5 and the levels of certain genes such as Rex1 and Sox2 start to show much more variation. This Nanog threshold level was defined as the minimum level of Nanog expression, above (HN; High Nanog) which cells were never observed to express Fgf5. Cells expressing Nanog below this level (LN; Low Nanog) did, however, express Fgf5 stochastically. These observations allowed us to assign each cell to one of three groups based on its expression of Nanog and Fgf5 (LN Fgf5−; HN Fgf5+; LN Fgf5−). We then used PCA and hierarchical clustering to determine whether the similarities within, and differences between groups went beyond expression of Nanog and Fgf5. Finally, we superimposed the average expression levels of the different genes onto the average structural networks (representing protein–DNA interactions) and the co-expression networks (representing correlations between the expression levels of the different genes) to infer sub-networks that operate in these sub-populations of cells. The sub-networks served to illustrate how network connectivity varied at the single cell level, whilst preserving the overall relationship between genes observed at the population level.

Single-cell expression dataset and normalization procedure

The data consist of transcript numbers for 9 different genes in 87 mESCs and were obtained from Hayashi et al.25 where the authors performed RT-PCR on single cells.55,56 They randomly selected 87 single mESCs and Q-PCR was used to determine the level of expression of different transcripts in each cell. Known quantities of spike RNAs from B. subtilis were added to each sample prior to reverse transcription and amplified alongside the endogenous RNAs. Transcript levels were obtained for each gene in each cell by comparing experimental C; values to those obtained for known numbers
of spike RNAs. To prepare the data for analysis, we first removed the two cells that had missing values for one of the genes, and then added the minimum observed value (0.0217) to every other value to account for the sensitivity of the method. Next we divided the transcript number for each gene in each cell by the number of transcripts of Gapdh in that cell. Cells 3 and 25 were excluded from the analysis as they contain exceptionally high levels of virtually all genes when normalized to Gapdh, thereby leaving 83 cells. Finally, throughout the analysis, we used the natural logarithms of each putative transcript number. The expression level distributions were obtained using the R statistical package. The same expression level scale, split into 30 units, was used for each gene.

Identification of sub-populations of cells

The cells were ordered by Nanog expression then the levels of Nanog, Rex1, Oct4, Sox2 and Fgf5 were plotted as lines and bars (Fig. 2). The Nanog threshold level is defined as the minimum level of Nanog expression, above which cells were never observed to express Fgf5. It is interesting that below this level, expression of Sox2 and Rex1 becomes much more variable and can, in some cells, be lost entirely. Principal component analysis (PCA) is a form of multivariate analysis that allows one to capture the essence of the variability within a dataset, but using fewer dimensions to aid visualization. The mathematical procedure, carried out here using R, provides a set of principal component values for each gene, the first of which endeavors to explain as much of the data as possible and so on until the final principal component. The first two principal components are used to generate the PCA plot as these explain the most variability. To obtain the position of a cell along each axis, the expression level of each gene in the cell is multiplied by the corresponding principal component. The first two principal components are used to generate the PCA plot as these explain the most variability. To obtain the position of a cell along each axis, the expression level of each gene in the cell is multiplied by the corresponding principal component for that axis. Therefore, the greater the magnitude of a principal component, the more that gene contributes to the variability along that particular axis. Furthermore, the closer one cell is to another on this plot, the more similar their expression profiles will be. In the PCA plot, the cells are colored according to their expression levels of Nanog and Fgf5, where cells that express high levels of Nanog (as defined in Fig. 2; see above) are colored red and cells that express any Fgf5 are colored blue. Cells with low levels of Nanog, but no Fgf5, are colored green. Before calculating principal components, the transcript levels for each gene were scaled such that they had a mean of 0 and a variance of 1. To create the heat map, the cells were ordered according to decreasing Nanog levels and split into high and low Nanog populations according to the threshold as defined above; cells that expressed Fgf5 were then placed in a separate group, ordered according to decreasing Nanog expression. The heat map was created using the heatmap.2 function from the R package. The genes were arranged according to a dendrogram generated by hierarchical clustering using Euclidian distance as the distance metric. A discontinuous color scheme is used to represent expression levels.

Reconstruction of the ensemble and sub-networks for the structural and co-expression networks

The edges of the structural network represent promoter-binding events determined by ChIP-chip experiments performed on a population of cells (gene product at the arrow source binds the promoter of the gene at the arrow head). The structural network was reconstructed using data from the network of 264 genes known to play a role in stem cell self-renewal that can be found in the Integrated Stem Cell Molecular Interaction Database (iScMiD). Stella was omitted from the structural network as no data were available for this gene. This and all of the other network diagrams were generated using Cytoscape, with the nodes being arranged manually. The edges of the co-expression network represent strong correlations between genes, i.e. where the Pearson correlation coefficient for the expression levels of a pair of genes has $P < 0.01$. The $P$ values were calculated heuristically by randomizing the expression levels in the single cell dataset 1000 times; they represent the frequency with which a correlation coefficient equal to or greater in magnitude than that calculated was observed in the 1000 randomizations. In Fig. 4A, the widths of the edges are proportional to the magnitude of the Pearson correlation coefficient for the two genes. In each of the functional networks, the node color represents the expression level of that gene in a cell, or the average expression level across the group specified. Expression levels were mapped onto the networks using the Cytoscape Vizmapper function and the active structural sub-network in the sub-population of cells was inferred as in Luscombe et al. In the ensemble networks for ES cell subpopulations (Fig. 4B and C), when expression of a gene other than Fgf5 was lost, the connections between that gene and genes other than Fgf5 have been removed to reflect the loss of connectivity between the down-regulated gene and the rest of the network. In the case of Fgf5, a negative correlation is only shown when Fgf5 is expressed and the other gene is not expressed, or vice-versa. The co-expression network in Fig. 4 was progressively simplified in Fig. 5 to aid visualization of gene expression patterns for the single cells in our dataset.

Acknowledgements

The authors thank the Wellcome Trust (AMA, AS, KH and JT) and the ERC (AMA) for funding. MMB acknowledges the MRC, HFSP (RGY0073/2010), and EMBO YI program for funding. We thank L. Wernisch, J.G. Ojalvo, G. Chalancon, C. Ravarani, B. Lang and C. Marr for their comments on this work.

References
