



OPEN

Dynamical gene regulatory networks are tuned by transcriptional autoregulation with microRNA feedback

Thomas G. Minchington¹, Sam Griffiths-Jones²✉ & Nancy Papalopulu¹✉

Concepts from dynamical systems theory, including multi-stability, oscillations, robustness and stochasticity, are critical for understanding gene regulation during cell fate decisions, inflammation and stem cell heterogeneity. However, the prevalence of the structures within gene networks that drive these dynamical behaviours, such as autoregulation or feedback by microRNAs, is unknown. We integrate transcription factor binding site (TFBS) and microRNA target data to generate a gene interaction network across 28 human tissues. This network was analysed for motifs capable of driving dynamical gene expression, including oscillations. Identified autoregulatory motifs involve 56% of transcription factors (TFs) studied. TFs that autoregulate have more interactions with microRNAs than non-autoregulatory genes and 89% of autoregulatory TFs were found in dual feedback motifs with a microRNA. Both autoregulatory and dual feedback motifs were enriched in the network. TFs that autoregulate were highly conserved between tissues. Dual feedback motifs with microRNAs were also conserved between tissues, but less so, and TFs regulate different combinations of microRNAs in a tissue-dependent manner. The study of these motifs highlights ever more genes that have complex regulatory dynamics. These data provide a resource for the identification of TFs which regulate the dynamical properties of human gene expression.

Cell fate changes are a key feature of development, regeneration and cancer, and are often thought of as a “landscape” that cells move through^{1,2}. Cell fate changes are driven by changes in gene expression: turning genes on or off, or changing their levels above or below a threshold where a cell fate change occurs. “Omic” technologies have been successful in cataloguing changes in gene expression during cell fate transitions. Many computational tools have been developed for the ordering of gene expression changes in pseudotime, delineating cell fate bifurcation points and linking genes into networks^{3–5}. However, while we have a good understanding of the fates/states that cells transition through and their order in time/space, the mechanisms that allow cells to move through the fate/state landscape are not well understood.

Gene regulatory networks are maps of interactions between different transcription factors (TFs), cofactors, and the genes or transcripts they target⁶. Networks are commonly represented diagrammatically as graphs of the connecting components, such as TFs and their targets. Network motifs are small repeating patterns found within larger networks⁶. Modelling of networks in this manner allows us to develop an understanding of how components interact and what behaviours they may generate^{6–9}.

Although it is clear that gene interactions are dynamic and change over time, current approaches in many biological studies focus on the qualitative analysis of genes or simple interactions between gene pairs: in short, how the perturbation of one gene affects the expression of another. However, gene expression is more nuanced than these traditional binary approaches can reveal. Biological networks are dynamical systems, that is systems that not only transform over time but resolve differentially depending on their parameter values, initial and boundary conditions, time delays, noise and the non-linearity of reactions. Autoregulation and cross-regulation

¹School of Medical Sciences, Faculty of Biology Medicine and Health, The University of Manchester, Oxford Road, Manchester M13 9PT, UK. ²School of Biological Sciences, Faculty of Biology Medicine and Health, The University of Manchester, Oxford Road, Manchester M13 9PT, UK. ✉email: sam.griffiths-jones@manchester.ac.uk; nancy.papalopulu@manchester.ac.uk

of components are the heart of a dynamical network's structure¹⁰. Dynamical models are better suited than binary systems to explain biological systems because they can account for phenomena such as robustness and plasticity¹¹.

Oscillations can emerge as a 'hallmark' of a dynamical system with multiple attractors. The fundamental properties that generate oscillations, such as the non-linearity of reactions, instability of components and time delays, are also the very properties that endow systems, including gene expression, with the ability to transition between different dynamic regimes. Oscillatory gene expression is now a well-recognised feature of several key TFs and signalling molecules. For example, p53 is expressed in an oscillatory manner following DNA damage, leading to the arrest of the cell cycle and DNA-repair, although, sustained expression of p53 leads to cell senescence (reviewed in Hafner et al¹²). p53 dynamics can also be altered in response to different stimuli¹³. Oscillatory vs sustained expression of p53 may differentially regulate target genes, leading to various outcomes including cell cycle arrest, or growth and recovery¹⁴.

Another example is the oscillatory dynamics of the Hes genes, which play essential roles in many different developmental processes. Notably, oscillations in Hes7 govern the timing of the somite segmentation during embryogenesis¹⁵, whereas oscillations of Hes1 have been shown to regulate the direction and timing of cell fate decisions in the developing neural tissue^{16,17}. Oscillations may be decoded by the phase, amplitude or duration of the oscillatory phase^{18–20}.

While gene expression oscillations are important, they may be viewed as only one of the possible states that a gene network can assume. Dynamical systems can have non-intuitive behaviours and it is necessary to employ mathematical modelling and quantitative approaches to understand their behaviour. Knowing which networks may show these properties is the first step. Coupling this approach to appropriate functional experimentation and mathematical modelling will allow a mechanistic understanding of cell fate/state transitions. Thus, we argue that discovering network motifs in biological networks that can drive a range of dynamics will facilitate a new conceptual and experimental approach to cell fate or cell state transitions, applicable to development, regeneration and disease, including cancer.

Results from synthetic biology show that the most straightforward motif that can produce oscillations is an autoregulatory negative feedback loop^{21,22}. Autoregulation is a critical component of other oscillatory motifs, including the amplified feedback^{23,24} and dual feedback loops²². A common element of the oscillatory motifs outlined here is negative feedback coupled with instability and delay. The repressilator, a synthetic gene interaction network of three repressors, forming a circuit of repression, utilised these principles to produce oscillatory expression of a fluorescent protein in *Escherichia coli*²⁵.

In addition to the network structure, gene expression oscillations have other principles in common, such as time delays and instability of the components (mRNA and protein). Recent evidence suggests that mRNA stability is key to the generation of oscillatory gene expression²⁶, and this may be regulated by microRNAs. MicroRNAs are a class small non-coding RNAs around 22nts in length and are critical regulators of gene expression, modifying mRNA stability and translation²⁷. MicroRNAs are transcribed from either intergenic microRNA genes or intragenic loci producing primary transcripts containing hairpin loops. These primary transcripts are spliced into shorter pre-microRNA which are exported from the nucleus where they are subsequently processed into a double-stranded microRNA duplex^{27–29}. One of these two transcripts, either the 5' or 3' arm, is selected to be the mature microRNA and enter into a repressor complex with Argonaute (AGO). MicroRNAs guide RISC (RNA-induced silencing complex) members to the 3' UTRs of target messenger RNAs (mRNAs)^{27,29}. RISC protein AGO can then repress translation and increase the rate of mRNA deadenylation of target transcripts³⁰. Increased deadenylation leads to decreased stability and therefore increased degradation of the target mRNA²⁷.

MicroRNAs are transcribed by Pol II (and in some cases Pol III) in the same way as mRNAs^{31–33}. As such, they are subject to the same regulatory input and are under the regulatory influence of transcriptional regulators such as TFs. MicroRNAs therefore are often incorporated into biological gene regulation circuitry, such as in the case of the Hes1/miR-9 oscillator and p53 oscillatory modulation by miR-16^{34,35}.

We have a good understanding of the structures that may allow dynamical behaviour in biological networks, such as bistability and oscillatory expression. However, the prevalence of such gene regulatory network structures is unknown. Are they common, or are they restricted to just a few cases of key TFs?

To address this knowledge gap, we constructed a gene interaction network by integrating data from well characterised human TF and microRNA databases. We then interrogated it for the presence of network motifs that have the potential of dynamical gene expression, in particular oscillations. Specifically, we have used the previously identified network structures of known synthetic and natural periodically expressed genes to interrogate a set of human TFs for the presence of simple motifs that incorporate TF autoregulation and reciprocal interaction of microRNAs and TFs. We used well-annotated databases and drew on information for TF binding from ChIP-seq data incorporated from ReMap database³⁶, as well as microRNA target predictions from miRTarBase³⁷. TF and microRNA target predictions from these data were used to generate a network, consisting of these genes and their targets.

We report that motifs with the potential to generate oscillatory gene expression (hereafter termed "oscillatory motifs") are widespread in TF networks. In particular, the autoregulatory motif, which is a core structure of all oscillatory motifs examined, is prevalent in our human TF dataset. Oscillations can occur with negative autoregulation^{21,34}, or positive autoregulation coupled with indirect negative feedback^{22,23}. We also demonstrate, through the use of tissue-specific sub-networks, that autoregulation with microRNA feedback is well conserved between tissues in terms of network topology. Surprisingly TFs were found to utilise different microRNAs for preserving the network structure in different cellular environments; possibly by utilising the expression of tissue-specific microRNAs.

Results

Most TFBS cluster in cis-regulatory modules. To understand the prevalence of network motifs within transcriptional and post-transcriptional networks, we collected datasets of experimentally-validated TF and microRNA interactions to construct a gene regulatory network. For the TFBS data, we took advantage of the ReMap project, which acquired human ChIP-seq experiments from GEO³⁸ (Gene Expression Omnibus). ReMap (v1.2) reprocessed the ChIP-seq data to identify 2,829 high-quality data sets^{36,39} and combined with ENCODE release V3⁴⁰ TFBS data to annotate a total of 80,129,424 TFBS positions for 485 TFs.

The ReMap database defined a set of cis-regulatory modules, which are loci where more than one TF binds. The vast majority of binding sites fall within these defined cis-regulatory modules and have at least one overlapping peak from another TF (98.6%; Fig. 1). We reasoned that TFBSs outside of these cis-regulatory modules were more likely to represent non-specific binding, and therefore excluded those sites, leaving 79,037,581 peaks, an average of 165,005 binding sites per TF.

TFBSs were assigned to the closest annotated Transcriptional Start Site (TSS) in Ensembl within a range of 50 kb upstream and 10 kb downstream of the annotated TSS. This range is expected to contain the proximal promoter region of the gene, but may also include more proximal enhancers⁴¹. Of the 79,037,581 TFBS remaining after cis-regulatory module (CRM) filtering, 65,222,825 TFBS (82.5%) were successfully assigned to TSSs of potential target genes. Binding-sites outside this range are likely to represent either non-specific binding or more distal enhancers. To investigate we compared the loci of the TFBS filtered from the data with candidate regulatory elements from SCREEN (Search Candidate cis-Regulatory Elements by ENCODE^[75,76]). Of the TFBS excluded for falling outside CRMs 0.02% were associated with promoter-like elements (PLS), 0.28% proximal enhancer-like elements (pELS) and 2.9% with distal enhancer-like elements (dELS). For TFBS outside our defined promoter range 0.02% intersected PLS, 0.28% pELS and 47% dELS. These results are consistent with the CRMs selecting for predicted regulatory elements and sites outside our promoter regions representing more distal enhancers. This relatively simple approach to the assignment of TFBS to genes thus provides a good approximation of gene regulatory input.

The set of annotated TSSs from Ensembl to which we have mapped our TFBS data include the 5' ends of protein-coding transcripts, but also a number of classes of non-coding RNA, including lincRNAs and microRNAs. 68% of human microRNAs in our datasets overlap with longer transcript annotations; the majority of these microRNAs are found in introns of protein-coding genes. Intragenic microRNAs may be co-transcribed with their host genes or independently regulated^{42,43}. For example, Marsico et al.⁴⁴ identified that around 60% of intragenic microRNAs share the transcriptional regulation of their host gene. To ensure that regulatory inputs for microRNAs are as complete as possible in our network, we combined the regulatory binding sites for the host gene with that of any TSS associated with the microRNA itself. Assignment of host gene regulation to microRNA results in 72,592,550 TFBS to gene interactions in the network. On average across all intragenic microRNAs, 13% of the regulatory inputs in our network are associated with the microRNA alone, and 87% come from the host gene.

Integrating microRNA target information. Next, we sought to integrate post-transcriptional regulation by microRNAs into our network. Several options are available for annotation of microRNA binding sites. For example, tools are available to predict microRNA target sites; most rely on seed matching⁴⁵ that is, comparing the seed sequence of the microRNA with sequences within the 3' UTRs of target genes. The signal-to-noise ratio for predictions using such short sequence matches is known to be relatively low, leading to high proportions of both false positives and false negatives⁴⁶. Instead, we collected microRNA target information from the miRTarBase (v7.0) database and integrated into the network. miRTarBase is an online database of validated microRNA targets, utilising experimental data from the literature, including luciferase reporter assays and CLIP-seq data³⁷.

TFBS filtering selects for higher confidence interactions. Different TFs have substantially different numbers of targets in the raw network. For example, ZNF335 and MDM2 have one protein-coding target each while MYC and CTCF have over 18,500 targets each, targeting over 93% of all protein-coding genes each. In the raw network, the median protein-coding target number for TFs as assigned is 10,651 (> 53% of protein-coding genes).

More binding sites for a given TF at a given TSS has previously been shown to be a good indicator of regulation⁴⁷. We therefore developed further filters to select for higher confidence interactions, based on the distribution of number of binding sites for each TF across all TSSs. Essentially, we simply remove interactions between a TF and a TSS region where the number of binding sites for that TF is fewer than the mean number of binding sites per gene for that TF (see “Methods”). In this way, we focus on TF/TSS region interactions where the number of TFBSs is above average. After this mean filtering is applied, the median number of protein-coding targets per TF is 4,303 (22% of protein-coding genes) a reduction of 60%. Some TFs are identified as targeting up to 9,369 protein-coding targets. Filtering TF-target interactions in this manner does not appear to bias or select for any particular target type, with the distribution of targets remaining proportional between different classes of gene targets, such as microRNA or protein-coding genes (see Fig. 2).

The TF/microRNA regulatory network is highly connected. The complete network comprises 479 TFs (ReMap) and 2,599 microRNA (miRTarBase v7.0). The interactions in a network are described as edges that link one component (node) to another. These edges can be divided into two groups: in-edges, which are receiving signals, such as a TF gene promoter receiving input from another TF; and out-edges, where a TF or microRNA regulates a promoter or mRNA. All components of a gene are linked together in our network. In the case of protein-coding genes, this means that the interactions at the TSS, mRNA and protein level all link to one

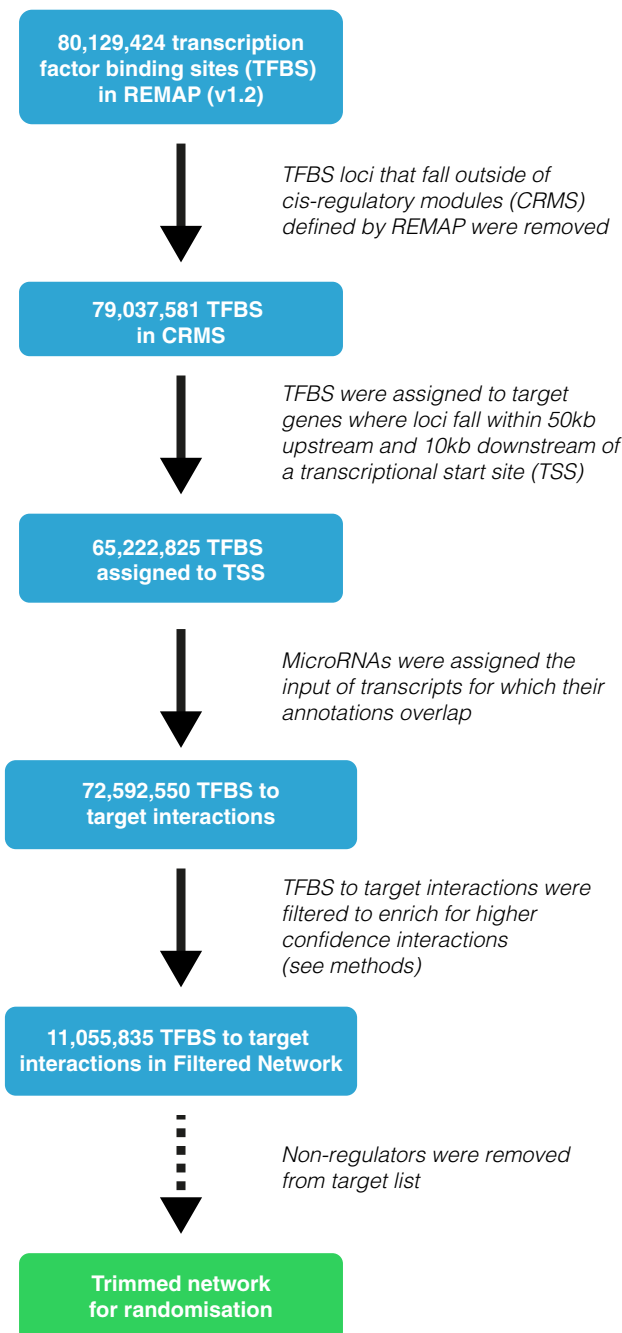


Figure 1. Refinement of TFBS data to final network. The original REMAP (v1.2) data contain 80,129,424 TFBS. These data were first filtered to keep only those found in CRMs (regions where TFBS were found to overlap). These peaks were then assigned to TSS of all genes as annotated in Ensembl 89 in a window of 50 kb upstream of the TSS and 10 kb downstream of the TSS (see “Methods”). To increase the confidence in the interactions between TFs and target genes, TSS-TF interactions were filtered based on the mean binding profiles of each TF. Interactions \geq mean binding were maintained. This filtering produces the final network. As all interactions in which we are interested contain feedback, we produced a trimmed network that only contains genes which are both targets and regulators.

component (node). The resulting network is highly connected: TFs on average have 120 in-edges and 170 out-edges, and microRNAs have 120 in-edges and 145 out-edges. Since we are primarily interested in motifs where the targets of transcriptional and microRNA regulation are themselves regulators, we also produced a trimmed network using only edges that connect two regulators (TFs and microRNAs). In this trimmed network, the average TF has 200 in-edges and 830 out-edges, and microRNAs have an average of 120 in-edges and 7 out-edges.

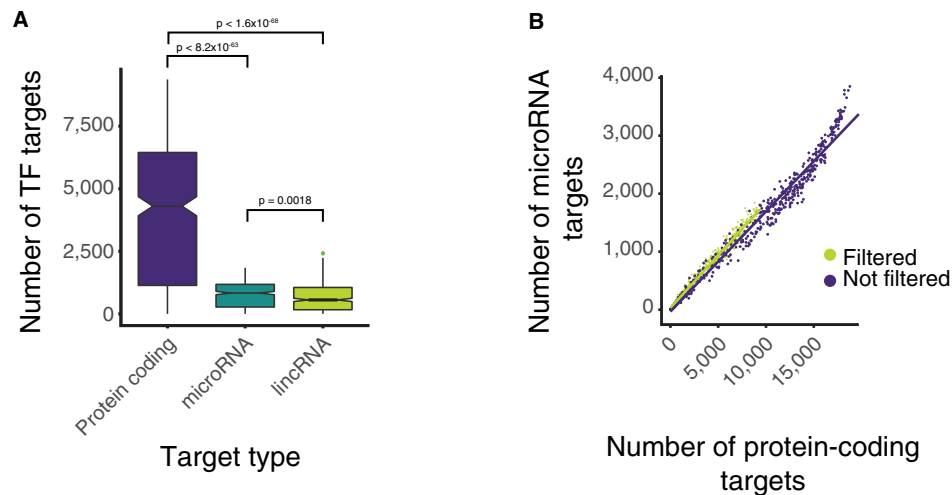


Figure 2. Network properties. (A) A boxplot showing the distribution of TF targets for protein coding, microRNA and lincRNA target types (p-values are calculated by Wilcoxon test). (B) Shows the relationship between the number of protein-coding and microRNA targets per TF before and after mean filtering. Due to the non-linearity of the non-filtered data at high values and the non-normal distribution of protein-coding targets and microRNA targets before and after filtering correlations were calculated Spearman's correlation (filtered $\rho = 0.989$, not filtered $\rho = 0.990$). Both remain highly correlated and no significant difference was identified between these correlations when Fisher Z transform was applied ($z = -0.79$, $p = 0.43$).

The difference here is due to the quantity of each type of network components—there are far fewer TFs in the network than microRNAs.

Transcriptional autoregulation with feedback by TFs and microRNAs is common. The network was interrogated to identify network motifs with structures that match those of existing oscillatory motifs. In all cases, the microRNA activity is assumed to be repressive, but the TF may be positively or negatively regulating. Nonetheless, we expect that a subset of the instances of the motifs identified will be connected in a way which could produce oscillations. The motifs selected were M1 (TF autoregulation), M2 (autoregulation with microRNA feedback), M3 (autoregulation with TF feedback), and M4 (two autoregulators with co-regulation) (Fig. 3A). TF autoregulation (M1) is a key feature of all our chosen network motifs and was observed for 266 of 479 (56%) TFs in the network (Fig. 3B). 237 of the autoregulating TFs (89%) are subject to microRNA feedback (M2). In total, the M2 motif was found 3,809 times within the network (Fig. 3B), comprising combinations of 237 (49%) TFs and 1,254 (48%) microRNAs (Fig. 3C). The M3 motif was observed 24,381 times and M4 19,622 times (Fig. 3B). The TFs in M3 and M4 motifs have more targets than in the M2 motifs and are more highly interconnected (Fig. 3D).

A randomisation experiment was conducted to determine that motif structure and prevalence is a product of underlying biology, not the inherent connectivity of the network. The trimmed network was randomised by rewiring, maintaining the number of inputs and outputs of each node. All network motifs investigated include an autoregulatory loop. If the autoregulatory loops (M1) are present more often than expected by chance, this could lead to all motifs appearing to be enriched. We therefore used two randomisation models: (1) “unlocked” randomisation, where all connections in the network are randomised, and (2) “locked” randomisation, where autoregulatory edges were fixed. TF-target and microRNA-target interactions were randomised independently as they represent fundamentally different levels of control (transcriptional and post-transcriptional regulation, respectively). Separating the interaction types also acts to ensure the random networks retain structural comparability to the real network.

The randomisation experiment showed that all four studied motifs are found more frequently than would be expected by chance. M1 motifs were investigated using the unlocked randomisation, whereas all other motifs used the locked method due to the presence of autoregulation in all motifs. M1, M2 and M4 motifs were never observed in the 1,000 random networks at numbers greater than or equal to their frequency within the real data ($p < 0.001$; M1 $z = 40.4$; M2 $z = 6.10$; M4 $z = 4.42$). M3 network motifs were less significantly enriched than other motifs within the network, though they are still observed more frequently than would be expected by chance ($p = 0.01$; $z = 2.42$). We therefore conclude that all four of the transcriptional autoregulation motifs, with and without feedback from TFs and microRNAs, are over-represented in the gene regulatory network. This over-representation is likely to reflect their importance in an underlying biological process.

Autoregulatory genes are regulatory nodes in the network. It has been hypothesised that genes which autoregulate do so due to the need for more precise control over the expression of these genes⁴⁸. Autoregulation, when negative, may also help buffer the response of the gene to multiple inputs to produce more robust expression behaviour. Connections into and out of autoregulatory genes were measured in comparison to non-

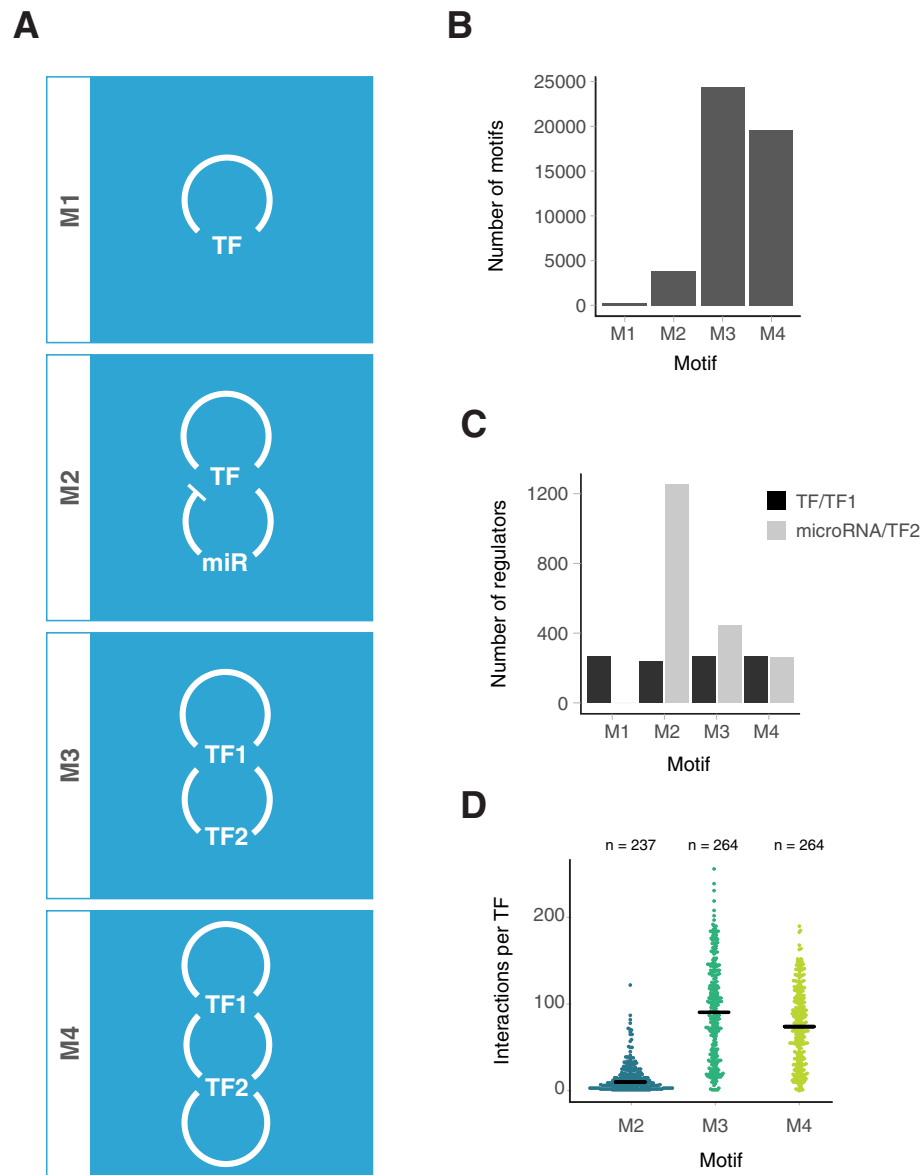


Figure 3. Known oscillatory motifs are abundant in the gene interaction network. **(A)** A key to the motifs studied here. The M1 motif is based on the classic Goodwin oscillator²¹. The M2 motif is based on the HES1 oscillator²⁴. The M3 motif is based on the amplified feedback motif outlined in Atkinson et al.²³. The M4 motif is based on the dual feedback motif from Stricker et al.²². **(B)** Bar graph showing the total number of motifs of each type within the network. **(C)** Bar graph showing the number of TFs and microRNAs which contribute to each of the motifs. M2 motifs have the highest number of components due to the large number of microRNAs in the network. **(D)** Beeswarm plot showing the number of other regulators each TF interacts within each motif. Black lines show medians.

autoregulatory genes to assess the connectivity of two groups in the network. Autoregulatory genes were found to have significantly more target genes than non-autoregulatory genes, an observation that holds across multiple target types (Fig. 4A). The microRNAs within the network were also found to have significantly more autoregulatory targets than non-autoregulatory targets (Fig. 4B). This observation did not appear to be specific to the group of microRNAs that show a preference for autoregulatory genes. Rather, most microRNAs (62%) regulate more autoregulatory genes than non-autoregulatory genes. Autoregulators on average have more inputs and outputs than non-auto-regulators independent of regulators or target type. These results taken together may indicate that auto-regulators form small local hubs in transcriptional and post-transcriptional networks. Gene Ontology enrichment analysis reveals that 70% (185) of the autoregulatory genes are associated with terms related to developmental processes (Supplementary Table 1). No enrichment was found for the non-autoregulatory group. Control of these genes, provided by autoregulation, may be necessary due to the high number of targets and their roles in development.

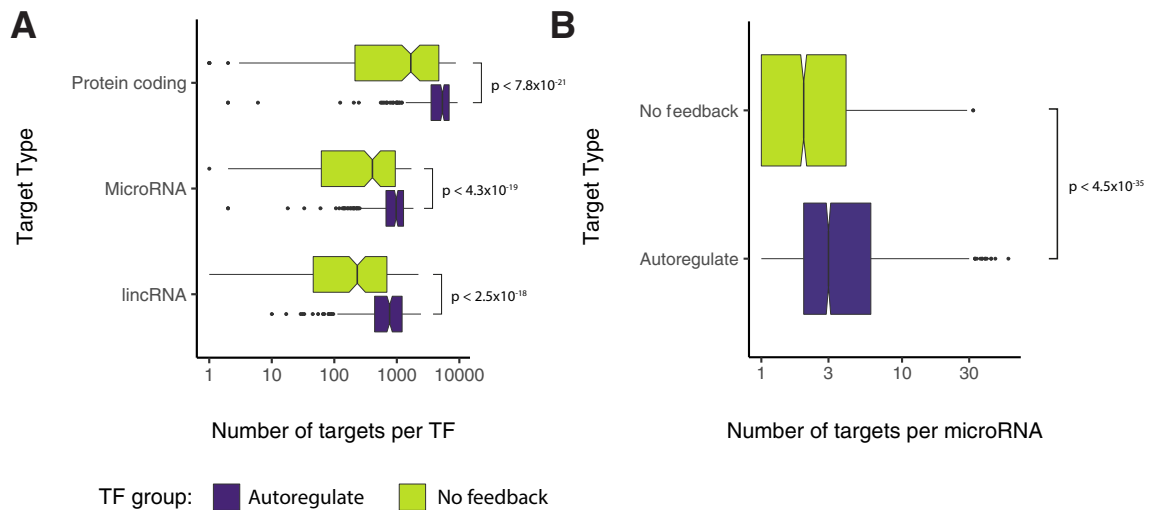


Figure 4. Autoregulatory genes are more connected than non-autoregulatory genes. **(A)** The number, and gene type, of down-stream targets for both autoregulatory genes and non-autoregulatory genes. Autoregulatory genes have more targets than non-autoregulatory genes. **(B)** The number of targets of microRNAs in the network which are autoregulatory or non-autoregulatory. MicroRNAs have more autoregulatory targets than non-autoregulatory targets. The p-values for both A and B were calculated using the Wilcoxon test using R.

Tissue networks reveal conserved motifs between tissues. The ReMap data contains information on the cell types from which all TFBS data were obtained. We used this cell type information to group TFBS datasets into 28 tissue types (Supplementary Table 2). This information was exploited to identify patterns of interactions between TFs and microRNAs that were maintained between different tissue types. To this end, we built sets of tissue-specific networks involving TFs and microRNAs. Each tissue-specific network necessarily contains a subset of the nodes and edges present in the entire network. The tissue networks are quite different in terms of TFs with 91% of TFs found in fewer than 20% of the networks and each TF on average contributing to 2.4 (mean) networks. The highest Jaccard index between the TF/target interactions of two networks is 31.3% for embryonic stem cells and lung (Supplementary Fig. 1A) with 83% of pairwise comparisons between networks resulting in a Jaccard index below 10% (Supplementary Fig. 1B). The difference between tissue-specific networks is largely a function of the studies that have been conducted. Of TFs that autoregulate, 43% had binding site data in only one tissue type, and were therefore not considered further here. 19% of autoregulating TFs were found in 2 tissues, 13% were observed in three tissues, and the remaining 24% were detected in four or more tissues (Fig. 5A).

Where TFs have binding site data in more than one tissue, we asked whether their connections were conserved between those tissues. To this end, we investigated the conservation of M1 and M2 type motifs between tissues. M1 autoregulatory motifs are well conserved between different tissue types: the autoregulatory ability of over half of all TFs was conserved in 100% of tissues where data exist (Fig. 5B). This suggests that their biological function may be linked to their ability to autoregulate. The presence of M2 motifs was only slightly less well conserved between tissues (Fig. 5C). When M2 motifs are not conserved between tissues, the most significant factor is loss of autoregulation—19% of all M2 motifs lose their autoregulatory signal between tissues, whereas only 1.3% of motifs are not conserved because they are missing TF regulation of the microRNA. However, even when the presence of the M2 motif is conserved, the identity of the components of the motifs may change. 18% of autoregulating TFs in M2 motifs, while retaining a core set of microRNAs, target different miRNAs in different tissues. The data suggest that the specific combinations of microRNAs and TFs vary between tissues, but the M2 motif structures themselves are maintained, sometimes by co-option of different microRNAs into those regulatory processes (Fig. 5D).

Overall, our data show that autoregulation of TFs is prevalent and that microRNAs engage more frequently with autoregulated TFs. Indirect feedback, both with microRNAs and between TFs is also widespread within in the network. The variety and quantity of feedback observed here, coupled with the conservation for these network structures between tissues, may indicate an important role for these motifs in conferring dynamic behaviour to the TFs involved.

Discussion

We generated a human gene interaction network including both transcriptional (TF) and posttranscriptional (microRNA) regulation. Previous work has indicated the high prevalence of autoregulatory loops in the human gene regulation network⁴⁹. Kielbasa and Vingron⁴⁹ used computational TF target predictions based on binding DNA motifs to look for autoregulatory loops. Here we build upon this work by utilising experimentally predicted gene targets, rather than relying on DNA binding motifs, while also increasing the number of genes investigated from 292 TFs to 479. Our data also include the addition of post transcriptional feedback with the inclusion of microRNA target data. The motifs investigated in this study are capable of producing a range of different dynamic

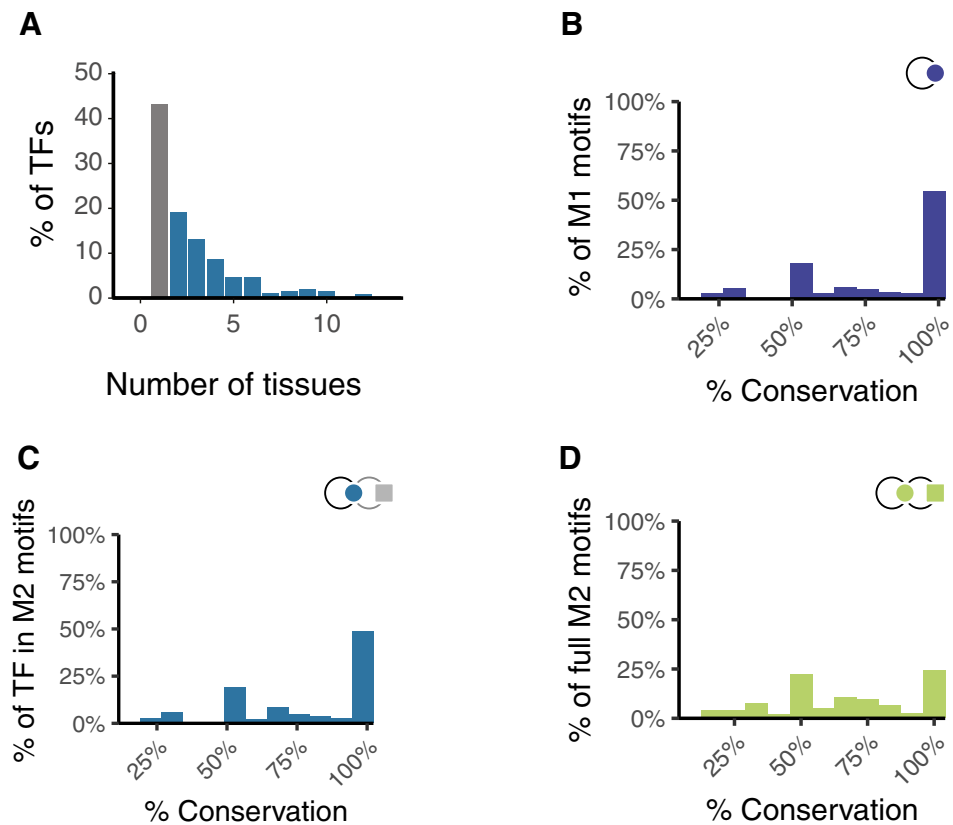


Figure 5. Conservation of TF in motifs between tissues. **(A)** Percentage of TFs found in multiple tissues. 45% of TFs have data in only one tissue (grey bar). The remaining TFs have TFBS data in 2 or more tissues. **(B)** Histogram showing the % conservation of M1 motifs between tissues. **(C)** Histogram showing the % conservation of TFs (circle) in M2 motifs between tissues, showing which transcription factors are in M2 motifs between tissues. **(D)** Histogram showing the % conservation of full M2 motifs between tissues involving specific combinations of TFs (circle) and microRNA (square) are conserved between tissues.

behaviours in their component genes, including noise modulation, bistability and oscillatory expression. Analysing the prevalence and behaviour of these motifs is therefore essential to our understanding of fundamental gene regulatory processes and their resulting molecular and cellular phenotypes.

Autoregulating TFs are highly enriched within our network, consistent with previous observations^{49,50}. These autoregulatory genes are also highly connected within the network, possessing more inputs and outputs to protein and microRNA genes than non-autoregulatory genes. The high number of connections of autoregulatory genes indicates that they play central roles in processing both inputs and outputs in the network. The wide range of gene expression dynamics which can arise from autoregulation may explain the prevalence of the motif in biological networks.

Autoregulation (M1, Fig. 3A) can produce many different behaviours depending on the properties of the components involved, and the type of inputs into the system. Negative autoregulation has previously been shown to decrease variability in the expression of a gene, buffering the noise of transcription and translation⁵¹. Negative autoregulation can also decrease the variability in the levels of a gene in a population of cells. Conversely, positive autoregulation can increase variability of protein levels between cells in a population leading to bistability⁵². Negative autoregulation may also decrease the response time of a gene when compared with simple regulation of one gene targeting another⁵³. The response time of a gene is the time required to achieve 50% the concentration of steady-state⁶. Steady-state is achieved when the degradation and production rates of the gene products are balanced. Negative autoregulation adjusts the balance between degradation and production by decreasing the transcription rate as the level of protein increases. Therefore negative autoregulation leads to faster equilibrium, as the degradation rate more quickly matches the production rate⁵³. Positive autoregulation has the opposite effect: the rate of transcription increases as the levels of protein increase, and the system therefore takes longer to reach an equilibrium state^{54,55}.

MicroRNA feedback has also been previously predicted to be an enriched motif in human and mouse gene networks⁵⁶. While microRNAs themselves are often suggested to act as buffers of gene expression, the interaction between a repressing TF and a microRNA in an M2 motif under certain conditions may act as a bistable switch⁵⁷. For example, ZEB and miR-200 form a negative feedback loop where miR-200 inhibits mZEB and ZEB inhibits the transcription of miR-200. This loop is thought to be involved in mesenchymal to epithelial-mesenchymal

transition⁵⁸. Modelling of miR-200 and ZEB has shown the system to switch states depending on input: high levels of miR-200 and low ZEB levels favour the epithelial phenotype, while low miR-200 and high ZEB produce a mesenchymal phenotype^{59,60}. The M2 motif involving ZEB1 and miR-200b-3p is present in our network analysis (Supplementary File 2).

One of the goals of this work was to identify motifs with similar structures to existing oscillators, and thereby predict new oscillators. Many motifs investigated in this study are capable of producing oscillatory patterns of gene expression, provided the regulatory feedback is negative, whether direct (M1 and M2), indirect (M3) or a combination of the two (M4) and other dynamical properties such as component instability and non-linearity of reactions are satisfied⁶¹. Oscillatory gene expression is an emerging area of research across many fields of biology including inflammation, stem cell heterogeneity and neurogenesis^{62–65}. The core conditions required to produce oscillations are negative feedback, instability and delay^{61,66,67}. Negative autoregulation is the most straightforward oscillatory motif and was the first synthetic oscillator to be described²¹. Oscillations can occur in just this simple system if a gene's mRNA and protein half-lives are shorter than the delay between gene activation and autorepression⁶¹. Negative autoregulation was shown to produce noisy oscillations *in vivo* when implemented through a synthetic autoregulatory module²².

The addition of negative feedback by a microRNA modifies the M1 motif to the M2 motif. This motif matches the network structure of the known ultradian oscillator HES1, a highly conserved oscillatory gene which drives cell fate decisions during neurogenesis^{17,68}. HES1 partners miR-9 in an M2 motif, and the levels of miR-9 are able to drive neuronal differentiation through the modulation of HES1 dynamics^{34,69}. It has also been hypothesised that the microRNA may also act as a self-limiting timer accumulating gradually through rounds of oscillations to a level where it destabilises the oscillations³⁴.

The M3 motif (Fig. 3A) is similar in structure to the M2 motif, with the microRNA substituted for a TF. In cases where the autoregulator in the M3 motif is a positive regulator of its own transcription, and the second TF negatively regulates the first, the M3 motif has been shown in synthetic biology to generate oscillations *in vivo*²³. The M4 motif is structurally similar to the M3 motif; however, both TFs are autoregulatory. Work by Stricker *et al.*²² has shown that the M4 motif can produce robust oscillations in situations where one TF is a positive regulator of transcription, and the other is negative.

Instances of autoregulation (M1) and autoregulation with microRNA feedback (M2) are highly conserved between tissues. The level of conservation indicates that these network motifs are necessary for core functional roles in different tissues. The component TFs were found to be conserved in M2 motifs in multiple tissues, but in many cases regulating different microRNAs between different tissues. Previous studies have shown that microRNAs have highly specific patterns of tissue expression⁷⁰. The switching of microRNAs in M2 motifs between tissues may therefore be due to differential incorporation of microRNAs into the network based on their availability in those tissues.

Our data show that autoregulation is prevalent amongst TFs and that genes which autoregulate are highly integrated into transcriptional networks. Further, autoregulatory genes preferentially engage with microRNAs, targeting and being targeted by more microRNA than non-autoregulatory genes. All of the oscillatory type motifs were identified more than would be expected by chance within our data. The wide range of behaviours that TFs can exhibit when contained within these motifs suggest that these network structures of TFs, microRNAs and their target genes allow dynamic expression to control core cellular processes. Of course, a complete understanding of the detail of how these motifs drive complex behaviours such as oscillations will require extensive and in-depth study of individual instances. However, the types of dynamical behaviours that these motifs predict are difficult to detect experimentally without real-time single-cell analysis. Our data therefore provides an invaluable resource in the search for dynamically expressed genes.

Methods

All scripts can be found at https://github.com/TMinchington/network_codes. All graphs were generated using R⁷¹ and the ggplot2 library⁷².

ReMap data filtering. TFBS data were obtained from the ReMap2018 (v1.2) database (hg38, All peaks). The cis-regulatory module data were also downloaded from ReMap2018 for the matching version. ReMap2018 (v1.2) contains TFBS data for 485 TFs. TF datasets generated using non-specific antibodies were removed from the data. For example, the binding data for RUNX contains bindings which are non-specific and may relate to any member of the RUNX family. We removed tracks where non-specificity was evident from the ReMap annotation. After this filtering was applied, 479 TFs remained. Transcriptional regulator binding sites were filtered to remove sites which were not found within cis-regulatory modules as described in the ReMap paper³⁶. The filtering was performed by running script `chip_in_cmfs.py` (arguments: CRM_file, chip_file). This custom python script uses the coordinates of the cis-regulatory modules from the ReMap database to cycle through the ReMap TFBS file and outputs TFBS which have loci contained within the cis-regulatory modules based on the genome coordinates.

Assigning TFBSs to target genes. Human gene annotation data were downloaded from Ensembl 89 BioMart (human genes; GRCh38.p10). TFBS binding site coordinates for 485 TFs were utilised from the ReMap database (V1.2)³⁶. TFBS (ReMap v1.2)³⁶ were assigned to the TSS (Ensembl 89) which was closest within a 60 kb region (10 kb downstream and 50 kb upstream of TSSs) using the custom python scripts `peak_MULTI.py` and `peakME_functions.py`.

Assigning microRNAs to host transcripts. The coordinates of human genes and exons were downloaded from Ensembl⁷³ (89, GRCh38.p10). The genomic positions of microRNAs were downloaded from miRBase⁷⁴ FTP server (release 22.0). MicroRNAs were then assigned to all transcripts with which the annotations overlap. Overlap with transcripts and positions within the transcripts were performed using microME.py. MicroRNAs within the peak_MULTI.py output files then inherited a copy of the TSS regulation of the genes in which they were contained while also maintaining the TSS as listed in the Ensembl 89 data.

Generating the transcriptional and translational network. MicroRNA target sites were downloaded from MiRTarBase³⁷ (Release 7.0) and converted into a compatible input format using mirTarbase_convert.py. Output files from microME_plus2018.py were combined with mirTarbase_convert.py output files to produce an edge list of regulators to target interactions. Duplicate interactions were collapsed using collapse_maps.py and the number of interactions recorded as an edge weight.

ENCODE cRE comparison. To compare the number of peaks filtered from the network against promoter and enhancer data 3 sets of Candidate cis-Regulatory Elements from SCREEN (Search candidate cis-regulatory elements by ENCODE)^{75,76}. Promoter-like, Proximal enhancer-like and distal enhancer-like elements were downloaded in BED format for human GRCH38. The peaks filtered from the data were recorded in BED format and compared against the promoter and enhancer files using Bedtools²⁷⁷ version v2.29.2-35-g07124422 intersect function using the unique option. The filter data were the peaks removed for being outside CRMs and those outside the 60 kb region investigated around the promoter.

Mean filtering of TFBS to TSS interactions. For each dataset the number of binding sites for each TF was recorded at each TSS. Following the removal of outliers (> 2 SD of the mean), the mean number of binding sites for each TF is calculated. If the average number of binding sites at TSSs for a given TF is n , then any TSS where fewer than n sites are observed are removed from the network. Means for interactions are calculated after removing extreme outliers. TSS, where $\geq n$ sites are found, would be maintained. Filtering is performed by edge_weight_filter.py and run on the collapsed network generated in the previous step.

Correlation coefficients were calculated using cor.test in R with the method 'spearman' for both the filtered and unfiltered data, due to the non-parametric distribution of both protein coding and microRNA coding targets and the non-linearity of the non-filtered data at higher values. The z and p values were obtained by using paired.r from the psych library, as a two-tailed test⁷⁸.

Network numbers. The number and type of edges in the networks were counted using the count_edge_type.py script. This cycles over the network and quantifies the type of interactions seen in the network. e.g. protein_coding-microRNA, microRNA-protein_coding and protein_coding-protein_coding.

Motif identification. Network motifs were identified by looking for patterns within the transcriptional-translational network edge list utilising the custom python script get_motifs_quicker.py. Autoregulatory loops are calculated first as other motifs are dependent on these loops. TFs in autoregulatory loops were then used to shortlist the search for further motifs. Motif discovery is based on Boolean logic looking for distinct patterns on interaction. For example, for autoregulation instances were identified where gene-A targets gene-A. Motifs were identified by looking at all patterns of interaction required to generate the motif.

Randomisation of networks. Gene networks were randomised 1,000 times utilising python script rewire_full_csf_array.py and rewire_locked_csf_array.py by randomly swapping network edges. Both scripts take two position arguments, the network to be randomised supplied as an edge list and the repeat number. This allows multiple instances to be run in parallel in a distributed computing environment. Network rewiring is a principle previously employed in the randomisation of networks⁷⁹. The swapping of edges was constrained such that microRNA to target interaction and TF to microRNA interactions were randomised separately. This prevents microRNAs being able to target the TSS of genes for example. Motifs were identified in each random network and compared to the real data. The number of edges exchanged was $1.1 \times$ the number of edges per group (protein_coding, microRNA). Any edge swaps which resulted in the same edge or duplicated an edge already in the network were discarded. Locked randomisation was used for testing the enrichment of all motifs except autoregulation which was investigated using the full randomisation method. Networks statistics were calculated as in Eqs. (1) and (2)⁸⁰.

Equation 1: Number of observations in random data greater than the observed data

$$r = n \geq \text{observed data.} \quad (1)$$

Equation 2: p -value

$$p = \frac{(r + 1)}{(n + 1)}. \quad (2)$$

Gene Ontology enrichment. Gene ontology enrichment analysis was performed using GOnet⁸¹. The set of TFs tested (autoregulators/non-autoregulators) for enrichment were supplied as a CSV file. Enrichment was

performed for the GO namespace “biological_process”. The complete list of TFs was supplied as a background. The q-value threshold was left at the default value of ≤ 0.05 .

Tissue networks. Tissue networks were generated by using the cell type and experiment data from the ReMap database and manually curating it into 28 tissue groups. Grouping was performed manually by identifying the tissue associated to each cell type. The network edges were then divided into network specific files based on the which tissue the data was grouped into.

Motifs were recalculated as above for each tissue network. The presence of each motif was then compared between each tissue network. The number of networks each TF was located in was calculated. If a TF was only found in one tissue network, then the data were discounted. For each motif a TF was located in was then compared between the tissue networks for which data for a given TF was found. Conservation was therefore the number of tissues where the motif was found as a percentage of the tissues where data for the TF exists.

To investigate the similarity of the tissue networks we calculated the Jaccard index between all combinations of pairs of tissues. For this analysis, we only considered TF/target interactions, as we do not have tissue-specific data with which to filter microRNA interactions, and the microRNA edges are therefore identical in each tissue network. The Jaccard index between a pair of networks was calculated by dividing the number of edges found in both networks by the number of edges unique to each network.

Data availability

TFBS and CRM data was obtained from ReMap2018 database (<https://pedagogix-tagc.univ-mrs.fr/remap/>). MicroRNA loci were obtained from miRBase (r22.0, <https://www.mirbase.org/>). MicroRNA target data was obtained from miRTarBase (<https://mirtarbase.mbc.nctu.edu.tw/php/index.php>). Genome coordinates and TSS were downloaded from Ensembl (89) (<https://www.ensembl.org/index.html>). The networks are available at https://github.com/TMinchington/network_codes.

Received: 5 March 2020; Accepted: 6 July 2020

Published online: 31 July 2020

References

- Waddington, C. H. The strategy of the genes. A discussion of some aspects of theoretical biology. With an appendix by H. Kacser. *Strateg. genes. A Discuss. some Asp. Theor. Biol. With an Append. by H. Kacser.* (1957).
- Huang, S., Ernberg, I. & Kauffman, S. Cancer attractors: A systems view of tumors from a gene network dynamics and developmental perspective. *Semin. Cell Dev. Biol.* **20**, 869–876 (2009).
- Boukouvalas, A., Hensman, J. & Rattray, M. BGP: Identifying gene-specific branching dynamics from single-cell data with a branching Gaussian process. *Genome Biol.* **19**, 65 (2018).
- Guo, L. *et al.* Resolving cell fate decisions during somatic cell reprogramming by single-cell RNA-Seq. *Mol. Cell* **73**, 815–829.e7 (2019).
- Aibar, S. *et al.* SCENIC: Single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017).
- Alon, U. Network motifs: Theory and experimental approaches. *Nat. Rev. Genet.* **8**, 450–461 (2007).
- Gennarino, V. A. *et al.* Identification of microRNA-regulated gene networks by expression analysis of target genes. *Genome Res.* **22**, 1163–1172 (2012).
- Neems, D. & Kosak, S. T. Turning down the volume on transcriptional noise. *Nat. Cell Biol.* **12**, 929–931 (2010).
- Osella, M., Bosia, C., Corá, D. & Caselle, M. The role of incoherent microRNA-mediated feedforward loops in noise buffering. *PLoS Comput. Biol.* **7**, e1001101 (2011).
- Miller, P. Dynamical systems, attractors, and neural circuits [version 1; referees: 3 approved]. *F1000Research* **5**(F1000 Faculty Rev):992, (2016).
- DiFrisco, J. & Jaeger, J. Beyond networks: Mechanism and process in evo-devo. *Biol. Philos.* **34**, 1–24 (2019).
- Hafner, A., Bulyk, M. L., Jambhekar, A. & Lahav, G. The multiple mechanisms that regulate p53 activity and cell fate. *Nat. Rev. Mol. Cell Biol.* **20**, 199–210 (2019).
- Batchelor, E., Loewer, A., Mock, C. & Lahav, G. Stimulus-dependent dynamics of p53 in single cells. *Mol. Syst. Biol.* **7**, 488 (2011).
- Purvis, J. E. *et al.* p53 Dynamics Control Cell Fate. *Science* (80-). **336**, 1440LP–1444 (2012).
- Bessho, Y., Hirata, H., Masamizu, Y. & Kageyama, R. Periodic repression by the bHLH factor Hes7 is an essential mechanism for the somite segmentation clock. *Genes Dev.* **17**(12), 1451–1456. (2003).
- Hatakeyama, J. *et al.* Hes genes regulate size, shape and histogenesis of the nervous system by control of the timing of neural stem cell differentiation. *Development* **131**, 5539–5550 (2004).
- Ohtsuka, T., Sakamoto, M., Guillemot, F. & Kageyama, R. Roles of the basic helix-loop-helix genes Hes1 and Hes5 in expansion of neural stem cells of the developing brain. *J. Biol. Chem.* **276**, 30467–30474 (2001).
- Sonnen, K. F. *et al.* Modulation of phase shift between Wnt and notch signaling oscillations controls mesoderm segmentation. *Cell* **172**, 1079–1090.e12 (2018).
- Levine, J. H., Lin, Y. & Elowitz, M. B. Functional roles of pulsing in genetic circuits. *Science* (80-). **342**, 1193–1200 (2013).
- Batchelor, E., Mock, C. S., Bhan, I., Loewer, A. & Lahav, G. Recurrent initiation: A mechanism for triggering p53 pulses in response to DNA damage. *Mol. Cell* **30**, 277–289 (2008).
- Goodwin, B. C. Oscillatory behavior in enzymatic control processes. *Adv. Enzym. Regul.* **3**, 425–438 (1965).
- Stricker, J. *et al.* A fast, robust and tunable synthetic gene oscillator. *Nature* **456**, 516–519 (2008).
- Atkinson, M. R., Savageau, M. A., Myers, J. T. & Ninfa, A. J. Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in *Escherichia coli*. *Cell* **113**, 597–607 (2003).
- Barkai, N. & Leibler, S. Biological rhythms: Circadian clocks limited by noise. *Nature* **403**, 267–268 (2000).
- Elowitz, M. B. & Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335–338 (2000).
- Moss Bendsen, K., Jensen, M. H., Krishna, S. & Semsey, S. The role of mRNA and protein stability in the function of coupled positive and negative feedback systems in eukaryotic cells. *Sci. Rep.* **5**, 1–10 (2015).
- Kim, V. N. MicroRNA biogenesis: Coordinated cropping and dicing. *Nat. Rev. Mol. Cell Biol.* **6**, 376–385 (2005).
- Ambros, V. *et al.* A uniform system for microRNA annotation. *RNA* **9**, 277–279 (2003).
- Ha, M. & Kim, V. N. Regulation of microRNA biogenesis. *Nat. Rev. Mol. Cell Biol.* **15**, 509–524 (2014).
- Chen, C.-Y.A. & Shyu, A.-B. Mechanisms of deadenylation-dependent decay. *Wiley Interdiscip. Rev. RNA* **2**, 167–183 (2011).

31. Borchert, G. M., Lanier, W. & Davidson, B. L. RNA polymerase III transcribes human microRNAs. *Nat. Struct. Mol. Biol.* **3**, 1097–1101 (2006).
32. Lee, Y. *et al.* MicroRNA genes are transcribed by RNA polymerase II. *EMBO J.* **23**, 4051–4060 (2004).
33. Monteys, A. M. *et al.* Structure and activity of putative intronic miRNA promoters. *RNA* **16**, 495–505 (2010).
34. Bonev, B., Stanley, P. & Papalopulu, N. MicroRNA-9 modulates hes1 ultradian oscillations by forming a double-negative feedback loop. *Cell Rep.* **2**, 10–18 (2012).
35. Issler, M. V. C. & Mombach, J. C. M. MicroRNA-16 feedback loop with p53 and Wip1 can regulate cell fate determination between apoptosis and senescence in DNA damage response. *PLoS ONE* **12**, e0185794 (2017).
36. Chèneby, J., Gheorghe, M., Artufel, M., Mathelier, A. & Ballester, B. ReMap 2018: An updated atlas of regulatory regions from an integrative analysis of DNA-binding ChIP-seq experiments. *Nucleic Acids Res.* **46**, D267–D275 (2018).
37. Chou, C. H. *et al.* MiRTarBase update 2018: A resource for experimentally validated microRNA-target interactions. *Nucleic Acids Res.* **46**, D296–D302 (2018).
38. Edgar, R., Domrachev, M. & Lash, A. Gene expression omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **30**, 207–210 (2002).
39. Griffon, A. *et al.* Integrative analysis of public ChIP-seq experiments reveals a complex multi-cell regulatory landscape. *Nucleic Acids Res.* **43**, e27–e27 (2015).
40. Wang, J. *et al.* Factorbook.org: A Wiki-based database for transcription factor-binding data generated by the ENCODE consortium. *Nucleic Acids Res.* **41**, D171–D176 (2013).
41. Maston, G. A., Evans, S. K. & Green, M. R. Transcriptional regulatory elements in the human genome. *Annu. Rev. Genom. Hum. Genet.* **7**, 29–59 (2006).
42. Baskerville, S. & Bartel, D. P. Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. *RNA* **11**, 241–247 (2005).
43. Morlando, M. *et al.* Primary microRNA transcripts are processed co-transcriptionally. *Nat. Struct. Mol. Biol.* **15**, 902–909 (2008).
44. Marsico, A. *et al.* PROMiRNA: A new miRNA promoter recognition method uncovers the complex regulation of intronic miRNAs. *Genome Biol.* **14**, R84 (2013).
45. Bartel, D. P. MicroRNA target recognition and regulatory functions. *Cell* **136**, 215–233 (2009).
46. Oliveira, A. C. *et al.* Combining results from distinct microRNA target prediction tools enhances the performance of analyses. *Front. Genet.* **8**, 59 (2017).
47. Sikora-Wohlfeld, W., Ackermann, M., Christodoulou, E. G., Singaravelu, K. & Beyer, A. Assessing computational methods for transcription factor target gene identification based on ChIP-seq data. *PLoS Comput. Biol.* **9**(11), e1003342 (2013).
48. Crews, S. T. & Pearson, J. C. Transcriptional autoregulation in development. *Curr. Biol.* **19**, R241–R246 (2009).
49. Kielbasa, S. M. & Vingron, M. Transcriptional autoregulatory loops are highly conserved in vertebrate evolution. *PLoS ONE* **3**, e3210 (2008).
50. Thieffry, D., Huerta, A. M., Pérez-Rueda, E. & Collado-Vides, J. From specific gene regulation to genomic networks: A global analysis of transcriptional regulation in *Escherichia coli*. *BioEssays* **20**, 433–440 (1998).
51. Becskel, A. & Serrano, L. Engineering stability in gene networks by autoregulation. *Nature* **405**, 590–593 (2000).
52. Becskei, A., Séraphin, B. & Serrano, L. Positive feedback in eukaryotic gene networks: Cell differentiation by graded to binary response conversion. *EMBO J.* **20**, 2528–2535 (2001).
53. Rosenfeld, N., Elowitz, M. B. & Alon, U. Negative autoregulation speeds the response times of transcription networks. *J. Mol. Biol.* **323**, 785–793 (2002).
54. Maeda, Y. T. & Sano, M. Regulatory dynamics of synthetic gene networks with positive feedback. *J. Mol. Biol.* **359**, 1107–1124 (2006).
55. Mitrophanov, A. Y., Hadley, T. J. & Groisman, E. A. Positive autoregulation shapes response timing and intensity in two-component signal transduction systems. *J. Mol. Biol.* **401**, 671–680 (2010).
56. Tsang, J., Zhu, J. & van Oudenaarden, A. MicroRNA-mediated feedback and feedforward loops are recurrent network motifs in mammals. *Mol. Cell* **26**, 753–767 (2007).
57. Goodfellow, M., Phillips, N. E., Manning, C., Galla, T. & Papalopulu, N. MicroRNA input into a neural ultradian oscillator controls emergence and timing of alternative cell states. *Nat. Commun.* **5**, 3399 (2014).
58. Lamouille, S., Subramanyam, D., Belloch, R. & Derynck, R. Regulation of epithelial–mesenchymal and mesenchymal–epithelial transitions by microRNAs. *Curr. Opin. Cell Biol.* **25**, 200–207 (2013).
59. Tian, X. J., Zhang, H. & Xing, J. Coupled reversible and irreversible bistable switches underlying TGFβ-induced epithelial to mesenchymal transition. *Biophys. J.* **105**, 1079–1089 (2013).
60. Lai, X., Wolkenhauer, O. & Vera, J. Understanding microRNA-mediated gene regulatory networks through mathematical modeling. *Nucleic Acids Res.* **44**, 6019–6035 (2016).
61. Novák, B. & Tyson, J. J. Design principles of biochemical oscillators. *Nat. Rev. Mol. Cell Biol.* **9**, 981–991 (2008).
62. Graf, T. & Stadtfeld, M. Heterogeneity of embryonic and adult stem cells. *Cell Stem Cell* **3**, 480–483 (2008).
63. Kobayashi, T. & Kageyama, R. Hes1 oscillation: Making variable choices for stem cell differentiation. *Cell Cycle* **9**, 207–208 (2010).
64. Nguyen, K. D. *et al.* Circadian gene Bmal1 regulates diurnal oscillations of Ly6C(hi) inflammatory monocytes. *Science (80-)*. **341**, 1483–1488 (2013).
65. Shimojo, H. & Kageyama, R. Oscillatory control of Delta-like1 in somitogenesis and neurogenesis: A unified model for different oscillatory dynamics. *Semin. Cell Dev. Biol.* **49**, 76–82 (2016).
66. Bratsun, D., Volfson, D., Tsimring, L. S. & Hasty, J. Delay-induced stochastic oscillations in gene regulation. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 14593–14598 (2005).
67. Monk, N. A. M. Oscillatory expression of Hes1, p53, and NF-κB driven by transcriptional time delays. *Curr. Biol.* **13**, 1409–1413 (2003).
68. Hirata, H. *et al.* Oscillatory expression of the BHLH factor Hes1 regulated by a negative feedback loop. *Science (80-)*. **298**, 840–843 (2002).
69. Tan, S. L., Ohtsuka, T., González, A. & Kageyama, R. MicroRNA9 regulates neural stem cell differentiation by controlling Hes1 expression dynamics in the developing brain. *Genes Cells* **17**, 952–961 (2012).
70. Ludwig, N. *et al.* Distribution of miRNA expression across human tissues. *Nucleic Acids Res.* **44**, 3865–3877 (2016).
71. R Core Team. R: A Language and Environment for Statistical Computing. (2019).
72. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer, New York, 2016).
73. Yates, A. *et al.* Ensembl 2016. *Nucleic Acids Res.* **44**, D710–D716 (2016).
74. Kozomara, A. & Griffiths-Jones, S. miRBase: Integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res.* **39**, D152–D157 (2011).
75. Dunham, I. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
76. Myers, R. M. *et al.* A user's guide to the Encyclopedia of DNA elements (ENCODE). *PLoS Biol.* **9**, e1001046 (2011).
77. Quinlan, A. R. & Hall, I. M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
78. Revelle, W. psych: Procedures for psychological, psychometric, and personality research. Northwestern University, Evanston, Illinois. R package version 1.9.12 (2019).

79. Milo, R. *et al.* Network motifs: Simple building blocks of complex networks. *Science* (80-). **298**, 824–827 (2002).
80. North, B. V., Curtis, D. & Sham, P. C. A note on the calculation of empirical P values from Monte Carlo procedures. *Am. J. Hum. Genet.* **72**, 498–499 (2003).
81. Pomaznoy, M., Ha, B. & Peters, B. GOnet: A tool for interactive Gene Ontology analysis. *BMC Bioinform.* **19**, 470 (2018).

Acknowledgements

The authors would like to acknowledge the assistance given by Research IT and the use of the Computational Shared Facility at The University of Manchester. The authors would like to thank Cerys Manning for proof-reading the manuscript. This work was supported by a Wellcome Trust PhD Studentship to T.G.M. (110566/Z/15/Z), a Wellcome Trust Senior Research Fellowship to N.P. (090868/Z/09/Z), and a BBSRC Grant to S.G.J. (BB/M011275/1). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contributions

S.G.J. and N.P. supervised this study. T.G.M., S.G.J. and N.P. conceived and designed this study. All computational and bioinformatics experiments and subsequent analysis were conducted by T.G.M.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-69791-5>.

Correspondence and requests for materials should be addressed to S.G.-J. or N.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020