STUDY DESIGNS

Studying and modelling dynamic biological processes using time-series gene expression data

Ziv Bar-Joseph[1], Anthony Gitter[2] and Itamar Simon[3]

Abstract | Biological processes are often dynamic, thus researchers must monitor their activity at multiple time points. The most abundant source of information regarding such dynamic activity is time-series gene expression data. These data are used to identify the complete set of activated genes in a biological process, to infer their rates of change, their order and their causal effects and to model dynamic systems in the cell. In this Review we discuss the basic patterns that have been observed in time-series experiments, how these patterns are combined to form expression programs, and the computational analysis, visualization and integration of these data to infer models of dynamic biological systems.

[1]*Lane Center for Computational Biology and Machine Learning Department, School of Computer Science, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA.*
[2]*Computer Science Department, School of Computer Science, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA.*
[3]*Department of Microbiology and Molecular Genetics, IMRIC, Hebrew University — Hadassah Medical School, Jerusalem, Israel.*
e-mails: zivbj@cs.cmu.edu; agitter@cs.cmu.edu; itamarsi@ekmd.huji.ac.il
doi:10.1038/nrg3244

As most biological processes are dynamic, time-series experiments are key to our ability to understand and model these processes. Although several types of genomics data can be measured over time, one of the most abundant and available types of such data is time-series gene expression data. Such data can be used to gain a wide range of insights, such as characterizing the functions of specific genes, the relationships among these genes, their regulation and coordination and the clinical implications of differential dynamics. Not surprisingly, generating time-series expression data has become one of the most fundamental methods for querying biological processes that range from various responses during development to cyclic biological systems. Recent improvements in methods for measuring gene expression — such as high-throughput RNA sequencing (RNA-seq) — and the increased focus on clinical applications of genomics make time-series expression studies more feasible and relevant. Indeed, the number of time-series data sets deposited in major expression databases has grown exponentially over the past few years[1]. In addition, the vast increase in sequencing capacity, which so far has been used primarily to generate static data sets, makes time-series expression data especially attractive as a form of complementary data for understanding dynamic systems.

However, despite its many benefits, time-series data also raise several experimental and computational challenges. In this Review we discuss these challenges, arguing that by adequately addressing them we could make better use of time-series data, and we highlight the insights that have been gained from using such data. We present the basic experimental considerations and computational methods that have been developed for representing, clustering and classifying time-series gene expression data. We then discuss the different types of temporal response patterns that are observed when studying biological processes and the set of genes that participates in these responses. We mention various other types of genomics data that have been measured as a time series and the insights that have been gained from these studies. Finally, we discuss advanced computational methods for integrating time-series data with other types of dynamic and static genomics data, and how the resulting models can provide temporal resolution for static genomic data sets. We note that although this Review is largely focused on gene expression data derived from microarrays, most of the points discussed also apply to sequencing-based gene expression data.

Experimental design and data analysis

Is a time-series experiment the right approach? The first question that should be asked when designing a time-course experiment is: what are the added benefits of performing multiple, rather than single (static), measurements? First, time-series gene expression experiments allow researchers to capture information about genes with transient expression changes. This is true for all types of biological processes; for example, in developmental or cycling processes different sets of genes are

Box 1 | **Sampling rates**

Sampling rates are tightly coupled to the goals of the experiment. When measuring a cyclic process (for example, the cell cycle[14–18] or circadian rhythms[19–21]) the sampling should be uniform and should cover multiple cycles in order to capture consistent changes. However, for developmental and response processes there is no simple practical approach for predicting the ideal sampling times. For developmental studies, two approaches are commonly used. The first approach relies on morphological markers of the embryo as a proxy for the important stages of transcriptional regulation (as was done, for example, in zebrafish[11] and in *Caenorhabditis elegans*[12]). The second approach is to change the sampling rate during the life cycle according to the expected rates of changes in gene expression. Thus, initial studies of *Drosophila melanogaster* sampled at overlapping one-hour intervals during embryonic stages and multi-day intervals during adulthood[10]. More recently, a combination of these two approaches was used in the *D. melanogaster* modENCODE project[13], in which embryos were sampled in hour-scale intervals, pupae and adults were sampled in day-scale intervals and larvae were sampled at distinct developmental stages.

In contrast to developmental and cyclic systems, in perturbation–response experiments most attention should be given to early time points because cumulative experience has shown that most of the transcriptional response occurs at the early time points, and sampling late time points does not usually add much information. For example, a study of the immune response to the yellow fever vaccine 17D (YF17D) that monitored gene expression of 15 people on days 0, 3, 7, 10, 14, 28 and 60 after vaccination[96] revealed that almost the entire response occurred in the first three time points. Similar results were obtained in yeast[2], *Escherichia coli*[86] and mice[5]. Thus, a better use of funds when carrying out response studies is denser sampling of the early time points. Nevertheless, for capturing gene expression patterns, a late time point is essential in order to differentiate between transient and sustained responses.

Once the general timescale is worked out there is still the issue of choosing the sampling density, for which the decision depends on the experimental goal. In general, given a fixed budget an important consideration is whether to invest in more replicates for each time point or in more time points (each with fewer replicates). If the goal is only to identify the genes that are significantly differentially expressed at certain points in the time series, then it may be better to invest more heavily in replicates. By contrast, if the goal is to determine the complete set of genes that are changing (including genes for which expression levels spike for short time intervals) or to investigate the pattern and kinetics of the response, a denser sampling rate (with fewer replicates) is an advantage. In that regard, because of the temporal autocorrelations of expression levels, denser sampling may also help to overcome noise in individual time points even without repicates. If a continuous representation framework (for example, cubic splines[32]) is used, the values that are estimated for a specific time point are dependent on the values of the preceding and following time points. Denser sampling results in more measurements being used to estimate each time point and thus less noise.

In practice, selecting the appropriate sampling rate is difficult because our prior knowledge of the system is limited. One useful practical solution is to monitor the expression of a few genes using dense sampling over a long time period and to try to deduce from their behaviours the preferred experimental design. Such an approach was recently used by Amit *et al.*[8] to study an immune response.

activated and repressed at diverse time points. Another example is perturbation–response experiments, in which different sets of genes respond with different kinetics, and therefore the entire response can be captured only by sampling the process at multiple time points[2]. Second, such experiments provide a view of the sequence of events that take place. This is important both in order to understand the type of processes that are activated at each stage and also for inferring causality. Finally, a time-course experiment allows study of the kinetics and the temporal pattern of a response. This is important for understanding the dynamic use of transcriptional networks[3], for defining early responses versus later responses[4] and for studying the differences in the kinetics of the response of the same genes to different types of perturbations[5].

General considerations for time-series experiments. Most genomic methods require a large quantity of starting material and therefore are often applied only to populations of cells rather than single cells. Technological advances already allow the measurement of the transcriptome of single cells[6], but sampling a cell at multiple time points is still a problem (an exception being imaging methods, although currently they can only follow a few genes simultaneously[7]). Hence, time-course measurements have so far been applied to cases in which fairly uniform cell populations exist. Fortunately, several processes produce such populations, including: responses to external signals (such as stress[2] or pathogens[8,9]); developmental processes that have a clear starting point (for example, embryonic development[10–13]); and cyclic internal processes in which the entire cell population can be synchronized (for example, the cell cycle[14–18] and circadian clock[19–21]). The next step is to determine the duration and sampling rates for the process being studied. BOX 1 discusses how this question can be addressed for the different types of biological processes mentioned above.

In static experiments and in time-course experiments, thousands of genes are measured for each sample, and both types of experiment often use several microarrays in a single experiment for both cases and controls. It is thus not surprising that early work using time-series gene expression data relied heavily on methods that were originally designed for static data. Although such methods were successful (see below), by ignoring the dynamics these methods may inadequately analyse and represent the results and thus may miss important conclusions.

Although most time-series gene expression data sets have been generated using microarrays and microarray analysis tools are currently more mature, over the past few years numerous time-series RNA-seq studies have been carried out[22–24]. There are various advantages to using RNA-seq for this purpose. First, several noise issues that are associated with microarrays — including background correction and cross-hybridization — are eliminated by using RNA-seq. In addition, sequencing-based studies make it easier to determine the expression of alternatively spliced genes, and such studies open the door to expression experiments for species that do not yet have an assembled genome[25]. Finally, there are indications that sequencing-based methods are more replicable and lead to more-accurate results compared with microarray-based methods[26]. Although researchers still face several problems when using RNA-seq (including higher costs and sequencing read errors[27]), we expect that over the next few years most time-series expression experiments will migrate to the RNA-seq technology. However, the issues discussed in this Review — including experimental design, representation and clustering — span both technologies and remain relevant regardless of the actual method that is used to quantify RNA levels.

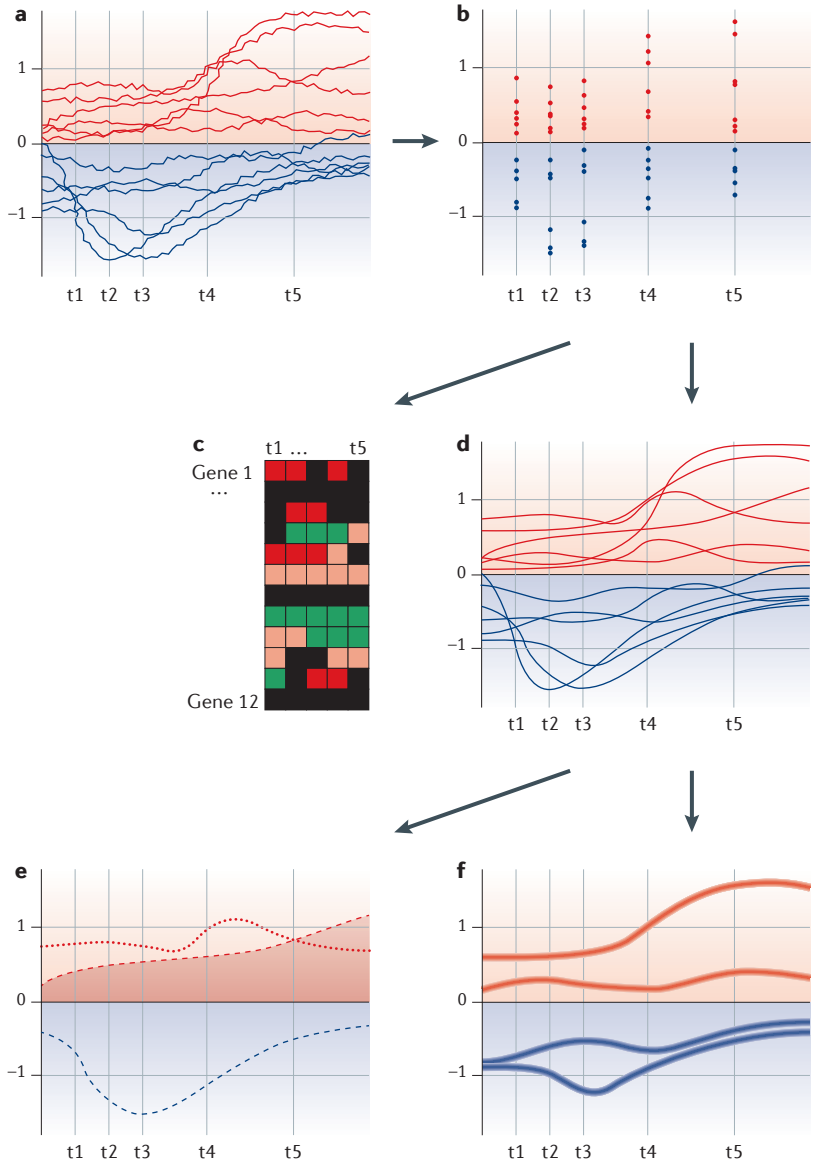


Figure 1 | **Computational workflow for time-series gene expression data.** **a** | The actual trajectories of gene expression levels over time are complex and noisy. The x-axis represents time (arbitrary units) and the y-axis represents $\log_2$ fold change of gene expression. Vertical lines show the time points at which samples were collected, which may be non-uniform. **b** | Actual values measured in a time-series experiment. The sampling rate defines the resolution of the resulting set of data points. **c** | Heat maps are a common way to display gene expression levels, but they do not provide information about duration. **d** | Before further downstream analysis, computational methods can be used to provide a more accurate representation of the expression profile of each gene. For example, splines provide a smooth, continuous representation of gene expression[32]. **e** | One ad-hoc approach for identifying differentially expressed genes requires that a gene exceeds a predetermined fold change threshold (shown as horizontal lines at 1 and -1) at two or more consecutive time points. This can help to discriminate truly differentially expressed genes (blue dashed curve) from genes with noisy expression (red dotted curve) that exceed the threshold at a single time point. Specialized computational techniques (for example, Significance Analysis of Microarrays (SAM)) can examine the area under the curve (red shaded area under the red dashed curve) or use temporal information along the entire curve (for example, Extraction of Differential Gene Expression (EDGE) and Bayesian Estimation of Temporal Regulation (BETR)) to call significant genes. **f** | Clustering assigns genes with similar expression trajectories to the same group. Curves that summarize the expression of all group members can reveal the common patterns of a response.

In the following subsections we focus on methods that are specifically designed for time-series experiments and we highlight how the temporal nature of the experiment can lead to improved downstream analysis and results. FIGURE 1 summarizes the computational workflow and TABLE 1 lists several software packages for each step in this workflow.

Normalization and data processing. Because normalization is primarily focused on recalibrating the data values of individual microarrays, normalization methods are usually the same for time-series and static experiments. However, in some cases researchers should be more careful when dealing with time-series data. One such example are experiments that measure RNA decay rates over time, which often involve blocking the transcription of new mRNAs (transcription shutdown)[28]. Such experiments violate one of the primary assumptions on which most normalization methods are based: that the total quantity of mRNA is the same at different time points[29]. The optimal normalization strategy for such cases is to rely on spike controls, if such controls are available. If not, some normalization methods, such as dChip[30], that do not rely on total RNA quantities should be preferred. Such methods rely on rank-invariant genes, which probably exist even after transcription shutdown[31].

Another issue related to data processing is synchronization. In several time-series experiments (for example, those studying cyclic processes) cells need to be synchronized over the entire time series. Achieving such synchronization relies on specific treatments and often requires both experimental and computational follow-up to correct for synchronization loss, as we discuss in BOX 2.

Representation. The next step is to plot graphically the expression levels over time, so that an overall view of the results can be obtained. This step is sometimes coupled with clustering (see below). The most popular representation for time-series data is still a heat map (FIG. 1c). However, heat maps are plotted with equal width for each time point, and cannot show whether the samples were taken at uneven time intervals. An alternative and very popular approach is to use piecewise linear curves, in which every two consecutive time points are connected by a line[32]. Extensions of these display methods include the representation of genes using other types of continuous curves (usually approximating splines) which can reduce noise through the use of temporal correlations[33]. FIGURE 1d provides a sketch of the use of cubic splines, which can be applied to time-series data sets that have more than four time points. Splines can also be assigned using a mixed effects model that uses expression data from co-expressed genes when reconstructing the curve representation for a single gene. This approach was shown to improve the accuracy of the continuous representation of expression data from multiple species[33,34]. Various recent methods attempt to further highlight the dynamics of expression changes. For example, Calvano *et al.*[35] presented the progression of gene expression during inflammation as a set of projections

Table 1 | **Software for the analysis of time-series gene expression data**

Task	Software	Description	Link	Refs
Identifying differentially expressed genes	Linear Models for Microarray Data (LIMMA)	Uses linear models to analyse gene expression and is part of the popular Bioconductor project	http://www.bioconductor.org/packages/release/bioc/html/limma.html	107
	Significance Analysis of Microarrays (SAM)	Permutation-based analysis of gene expression	http://www-stat.stanford.edu/~tibs/SAM	39
	Extraction of Differential Gene Expression (EDGE)	Statistical analysis that specifically leverages the time structure in the expression data	http://www.genomine.org/edge	108, 109
	Bayesian Estimation of Temporal Regulation (BETR)	Bayesian technique that exploits time-dependent structure in the expression data and is available with the MultiExperiment Viewer (MeV) application and Bioconductor	http://www.tm4.org/mev	41
Clustering	Short Time-series Expression Miner (STEM)	Maps genes to representative expression profiles with an emphasis on short time-series experiments; also implements k-means	http://www.sb.cs.cmu.edu/stem	110
	Graphical Query Language (GQL)	Hidden Markov model (HMM)-based clustering	http://ghmm.org/gql	111
	Cluster Analysis of Gene Expression Dynamics (CAGED)	Models gene expression using autoregressive equations	http://dcommon.bu.edu/xmlui/handle/2144/1290	43
	TimeClust	Implements hierarchical clustering, self-organizing maps, and two novel time-series clustering algorithms	http://aimed11.unipv.it/TimeClust	112
	Dynamic modelling and clustering (DynaMiteC)	Simultaneously clusters genes and fits groups of similar genes to impulse models	http://www.compbio.cs.huji.ac.il/DynaMiteC/Site/DynaMiteC.html	113
	Platform for Processing Expression of Short Time Series (PESTS)	Summarizes expression profiles with various features and can also identify significant genes	http://www.mailman.columbia.edu/academic-departments/biostatistics/research-service/software-development	114
Classification	GQL	Extensions of GQL enable it to classify clinical responses on the basis of gene expression	http://ghmm.org/gql	49
	Treatment-Response Alignment Models (TRAM)	Discriminative HMM-based classification	http://www.cs.cmu.edu/~thlin/tram	48
	MVQueries	Uses HMMs to model expression response as piecewise constant functions	http://bioinformatics.rutgers.edu/Software/MVQueries	103
Dynamic regulatory networks	Inferelator	Ordinary differential equations are used to model transcriptional changes in terms of environmental and transcription factor influence	http://err.bio.nyu.edu/inferelator	115
	Network Component Analysis	Decomposes a dynamic gene expression matrix to learn transcription factor activities over time	http://www.seas.ucla.edu/~liaoj/download.htm	71,83, 116
	Dynamic Regulatory Events Miner (DREM)	HMM-based algorithm for identifying transcription factors that control divergence points in gene expression profiles	http://www.sb.cs.cmu.edu/drem	85
	Time-Series Network Identification (TSNI)	Constructs a local regulatory network of genes that are affected by an external perturbation	http://dibernardo.tigem.it/wiki/index.php/Time_Series_Network_Identification_TSNI	84
Simulation	GeneNetWeaver	Generates realistic regulatory networks and dynamic gene expression data	http://gnw.sourceforge.net	117

on an interaction network, and Lu *et al.*[36] created movies to track the progression of gene expression during stem cell differentiation by arranging genes in a (fixed) two-dimensional grid.

Differentially expressed genes. Tightly linked to representation is the question of how to identify differentially expressed genes. A heuristic, ad-hoc solution that is commonly used[37,38] is to require that at least two consecutive points are above a chosen fold change in order to declare a gene as differentially expressed (FIG. 1e). However, such heuristics use arbitrary thresholds that may not be appropriate for all genes (depending on the baseline expression levels, a twofold increase can be either very significant or just noise[39]). To address this issue numerous statistical methods have been developed, or extended, to identify differentially expressed genes in time-series data (TABLE 1). Unlike the ad-hoc methods, these often rely on analysing a more continuous version of the experimental results for each gene and thus use more time points to identify differentially expressed genes. Comparing these methods to methods that were developed for the analysis of static gene expression data (for example, t-tests on individual time-point

Spike controls
Special control transcripts that are mixed into a biological sample before microarray hybridization. Because the quantity added is known, the control probe signal can be used for accurate normalization.

Box 2 | **Synchronization of time-series gene expression experiments**

Microarray-based and high-throughput RNA sequencing (RNA-seq)-based experiments currently require a population of cells. Therefore, the synchronization of these cells — so that all are at the same phase of the system or response at the beginning of the time series and at any subsequent time point — is an important issue.

The issue of within-series synchronization was highlighted when studying the cell cycle. As mentioned in the main text, one of the first applications of a time-series analysis was the study of cell-cycle-regulated genes[14]. Such experiments require a synchronized population of cells at multiple time points during the cell cycle. The most commonly used synchronization method in these experiments is arresting cells at a specific point in the cell cycle and then releasing them. This approach was initially used in budding yeast[14] and was later extended to other species[16,17,97].

However, although methods involving arrest were effective for characterizing genes in lower organisms that show cyclic expression levels, their application to wild-type mammalian cells, in which the cell cycle is much longer, resulted in cells quickly losing their synchronization and thus produced inaccurate results[72]. Even for yeast cells such arrest methods did not lead to complete synchronization, which affected downstream analysis[98].

Various methods have been introduced for re-synchronizing cells in a cyclic experiment. These rely on matching the profiles for the first and second cycle for each gene[98]. Although these work well for cases in which at least the first cycle is fairly synchronized (including the cell cycle in yeast and circadian rhythms), they do not generalize to the mammalian cell cycle or to other responses in which the activity is not cyclic.

An alternative set of approaches is to synchronize cells *in silico*. One approach along these lines relies on additional measurements to characterize the population of cells at each time point (for example, flow-cytometry-based analysis of DNA quantity or other markers). Using these markers, a model for the mixture of cells at each time point is constructed and is then used to deconvolve the time-series expression data[18,99]. In these papers it was shown that such a model can recover human cell cycle expression data in primary cells. Additional methods that assume other types of mixing models were successfully applied to study the cell cycle in *Caulobacter crescentus* bacteria[100] and meiosis in yeast[101].

Another type of synchronization is required when combining or comparing time-series experiments from multiple studies, or from different patients. In such cases, even though each individual series may be synchronized, response rates may differ between these series, thus making it difficult to compare results on the basis of the actual time reported (for example, a certain individual may respond to a drug more slowly than another, depending on their age and other background factors). Several approaches have been suggested to address these problems; most either rely on the alignment of expression profiles between the experiments using a time-wrapping method[32,34,102] or use hidden Markov models for the alignment process[48,103].

values) indicates that in at least some cases these statistical methods can improve the identification of differentially expressed genes[40]. Some of these methods require replicates at each time point (for example, Bayesian Estimation of Temporal Regulation (BETR)[41]), whereas others do not (for example, new versions of Significance Analysis of Microarrays (SAM)[39]).

Clustering. Although static-based clustering methods (including hierarchical clustering[42] and k-means) have been used widely for time-series data, several methods have been developed specifically for time-series data. These include: methods that use regression analysis to group genes on the basis of their trajectories (for example, Cluster Analysis of Gene Expression Dynamics (CAGED)[43]); methods based on graphical models (specifically, hidden Markov models (HMMs)) that group genes on the basis of their transcriptional trends, regardless of the specific values (for example, Graphical Query Language (GQL)[44]); and methods that attempt to assign genes to one of several previously defined temporal trajectories, thus allowing users to determine significance levels for the different clusters (for example, Short Time-series Expression Miner (STEM)[45]). By using the temporal information, these methods are often an improvement on the static-based methods, when judged by the biological coherence of the groupings obtained.

Classification. In disease and clinical studies it has been observed that dynamic differences in gene expression profiles may provide insights into the severity of the disease and the responsiveness of the patients to treatment[9,46] (see also BOX 3). Numerous methods have been proposed recently to exploit such data by trying to classify outcomes on the basis of the dynamics of expression changes[47–49]. In many cases it was shown that by using the temporal information, these methods are an improvement on methods that only use the baseline (time-point zero) values[48].

Causality. A key advantage of time-series data is the ability to infer causality without perturbing the system, using causal modelling. By observing the cascade of expression changes, their specific profiles and their temporal autocorrelations, researchers can derive several hypotheses regarding causal relationships between genes. Early work in this direction used an alignment approach to match similar or opposite subsections of expression patterns that were temporally separated (one preceding the other)[50]. These alignments were used to identify potential activators and repressors[50]. Several additional methods, many using various types of regression analysis, were suggested for identifying such causal relationships[51]. In these methods researchers attempt to model the expression profile of a specific gene as a function of the expression of another gene (that is expressed earlier). Again, methods that use continuous representation seem to be more appropriate for this type of analysis[52]. For example, regression methods were used in a zebrafish study in which it was shown that during somitogenesis the expression of some transcription factors precedes that of their target genes[11]. Although most work in this area has focused on model organisms, recently dynamic Bayesian networks — which rely on the expression of a regulator at one time point to explain the expression of a target at the next — were successfully applied to identify causal candidates for the temporal changes in a human blood transcriptional network [53]. Note, however, that owing to the high dimensionality of the data, false positives remain a major concern when carrying out such causality analysis[52]. In addition, because many transcription factors are only post-transcriptionally regulated, such an analysis may miss key regulators. Thus, a better approach is to integrate additional types of data when carrying out causal modelling, as discussed below.

Patterns of temporal gene expression

Principles of dynamic gene expression programs. As noted above, time-course experiments fall into three

dChip
A software for viewing and normalizing probe-level microarray data. It relies on the assumption that some genes are rank-invariant between different samples.

Approximating splines
A smooth piecewise polynomial function that can be fitted to a temporal gene expression profile.

Cubic splines
Splines composed of third-order polynomial functions.

Box 3 | **Clinical applications**

Time-series gene expression data are being increasingly used to monitor patient responses in clinical studies that are focused on human responses to injury and disease[9,35,104], as well as to treatments and preventive measures[46,96,105]. Patient heterogeneity can render the analysis of absolute expression levels meaningless[104], and ethnicity can affect the transcriptional responses to therapy[105]. Therefore, dynamic measurements that allow the assessment of within-patient expression changes are especially beneficial. Such studies often face a unique set of challenges. Ethical considerations may preclude certain types of sample collection that would be most relevant to the scientific hypotheses. For example, although the analysis of hepatocytes would have been most relevant in a study of patients chronically infected with hepatitis C virus, harvesting hepatocytes requires a liver biopsy, which cannot be carried out repeatedly on humans[105]. Instead, many such studies examine peripheral blood mononuclear cells[46,96,105] or other peripheral blood cells[9] as an approximation, owing to their accessibility. It is difficult to assess how accurately this approximation reflects the actual gene expression dynamics in specific conditions, but an analysis of which genes are expressed in peripheral blood versus nine other tissues indicated that it may be an effective surrogate tissue[106]. For infections by milder viruses (for example, certain influenza strains[9]) it is possible to expose volunteers to the pathogen and to measure phenotypic changes directly, but studies of more lethal viruses cannot do so (although studies of responses to vaccines, for example yellow fever vaccination[96], have been carried out).

Determining the correct sampling rate can also be difficult. In many cases the most discriminating transcriptional changes occur within days[96,104,105], but it may take years to assess whether a patient has responded successfully to treatment [46], suggesting that a long duration of sampling would be more appropriate in those studies. Studies that are carried out outside of a controlled laboratory environment may generate time-course data from variable numbers of samples and collection times across patients[104]. This problem can be overcome by regressing the log gene expression levels on time to quantify the fold change per hour and using this linear slope for subsequent analyses[104]. Such slope-based analysis, and other techniques that use within-patient expression dynamics, are widely applicable and they also help to mitigate the aforementioned effects of heterogeneity among patients or laboratories.

The primary goal of many clinical studies is often classification to predict patient outcome. For example, in a study in which healthy volunteers were exposed to influenza, researchers applied logistic regression classifiers to identify genes that can discriminate between pairs of phenotypic classes (pre-inoculation, asymptomatic, symptomatic before onset, and symptomatic after onset)[9]. Similarly, correlation with dynamic clinical phenotypes was also instrumental in a study of patients that had suffered blunt-force traumas[104]. An adjusted Spearman rank-based correlation test detected significant associations between gene expression fold change per hour and a temporal multiple organ failure score. In a study of patients with multiple sclerosis that were undergoing treatment with recombinant human interferon-β, Baranzini *et al.*[46] analysed the dynamics of gene expression and also identified gene triplets for which expression levels at time point 0 (before treatment began) were predictive of the therapeutic outcome. Follow-up studies of the same multiple sclerosis data set have shown that even better predictive results can be obtained by using the complete temporal responses for selected genes[48,49]. Clinical studies often suffer from a lack of reproducibility, which inhibits their applicability to medical practice. Thus, it is crucial to guard against overfitting the gene expression data by: concentrating on small subsets of relevant genes on the basis of prior knowledge[46]; using cross-validation strategies[46,104]; or validating results using independent patients and/or experimental methods[35,96].

large categories: responses to external signals, developmental processes and cyclic processes. Each type of experiment has a characteristic expression outcome (FIG. 2A). However, these distinct outcomes are achieved by two types of basic expression patterns — short impulse and long sustained — which are common to all responses. Short impulses are defined as upregulation or downregulation of transcript levels for a short period of time (FIG. 2Ba) and then a change to a new steady state, which is often a return to the original levels. Long sustained responses are defined as changes in the transcript level of a gene that persist for a long time (FIG. 2Bb). Combinations of these basic expression patterns give rise to the more complex outcomes. In the following subsections, we start by describing common combinations of these basic patterns and then we summarize the major results of numerous measurements, grouped into the three main types of time-course experiments.

Common combinations of the basic expression patterns. One recurring pattern is an immediate-early impulse of key regulators that is followed by sustained changes in transcription (FIG. 2Ca). For example, such a response was observed during phorbol myristate acetate (PMA)-induced differentiation of THP-1 human myelomonocytic leukaemia cells[54]. Typically, impulses of key regulators are followed by the sustained repression of cell cycle genes and the activation of genes that characterize the differentiated phenotype (for example, immune-response genes).

Another common pattern is a succession of impulses of expression[4,8,38,55]. This pattern has been found in various types of cellular processes. For example, the response of dendritic cells to pathogens involves several waves of induction[8]. Some developmental processes are also regulated by such cascades[56]. For example, in sea urchin embryos, skeletogenic cell development is facilitated by a regulatory cascade in which transcription factors that are active during one phase activate genes for which the products are needed in the next phase[57]. Cyclic processes are also characterized by sequences of pulses of gene expression[14]. Indeed, for the yeast cell cycle it was shown that the cyclic behaviour is controlled by a cascade of transcription factors; the expression level of each factor peaks at a different stage of the cell cycle and this regulates the expression of transcription factors of the next stage[58]. Investigation of the temporal orders of gene activation in a pathway revealed several recurring temporal patterns[3]. One of the major temporal patterns is the 'just-in-time' pattern, which is characterized by a match between the temporal activation of a set of genes and their order of activity in the process[3,59,60] and is a consequence of a cascade of impulse responses (FIG. 2Cb). This temporal pattern is commonly observed among genes encoding metabolic and biosynthetic enzymes in microorganisms[3]. For example, following deprivation of amino acids in *Escherichia coli* the expression of amino acid metabolic genes is induced in the same order that their encoded enzymes are required in the relevant amino acid biosynthetic pathway[61]. This just-in-time pattern has also been observed in other bacterial processes, such as flagellar biogenesis[62].

Responses to external signals. The most common type of time-course gene expression experiment is carried out to measure the responses of cells to external signals. A comprehensive review of the observed cellular responses to such perturbations was recently published[63], so here

Hierarchical clustering
A greedy clustering approach in which pairs of genes or clusters are sequentially connected until they form a tree-like structure.

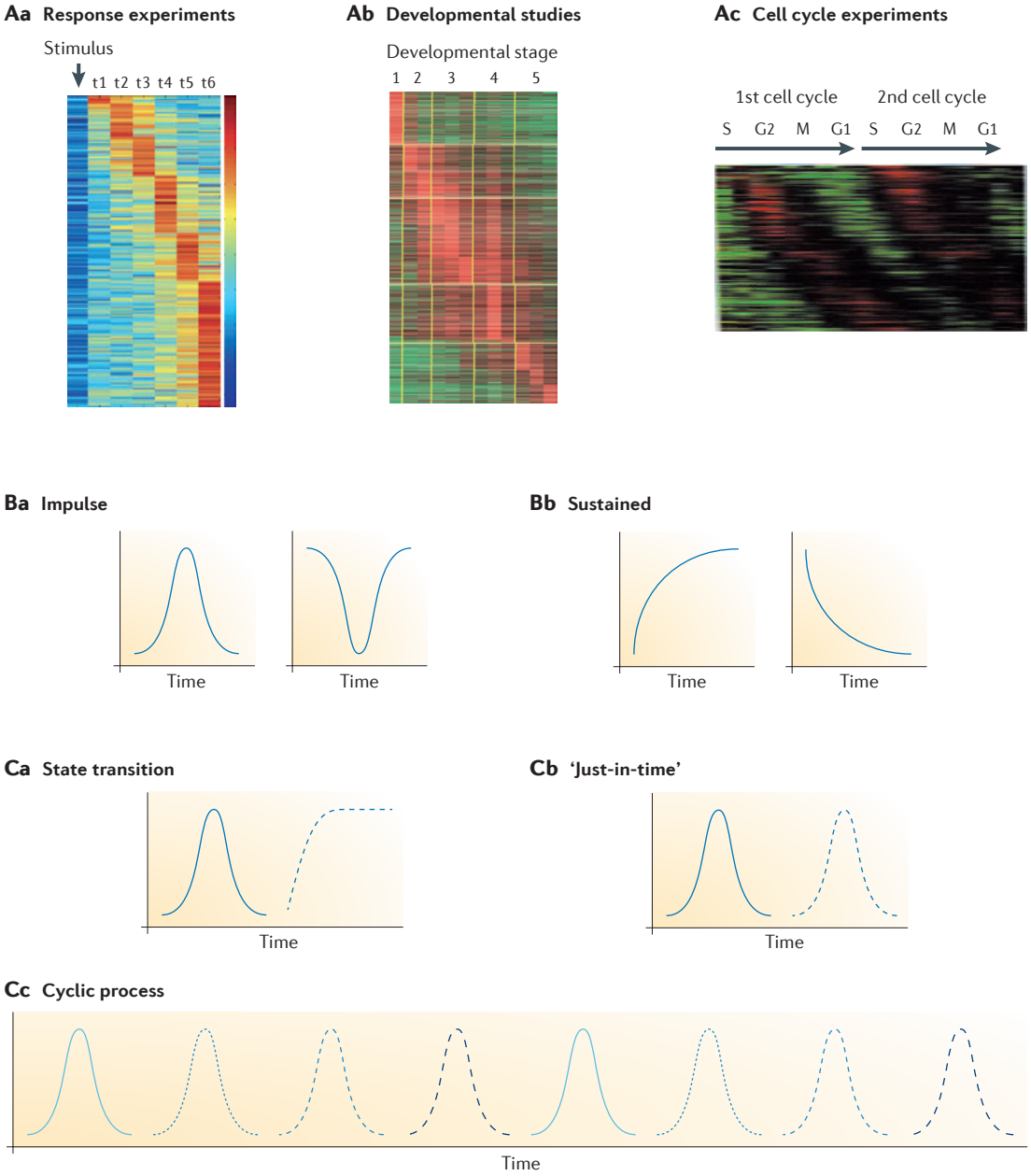


Figure 2 | **Patterns of gene expression.** **A** | Three major types of biological scenarios lead to different observed outcomes: **Aa** | Response experiments are characterized by transient responses. **Ab** | Developmental studies display a shift from one steady state to another. **Ac** | Cell cycle experiments show a repeated pattern of gene expression for each of the cycles measured. **B** | Basic temporal patterns. (For the graphs in **B** and **C**, time is shown on the x-axis and gene expression levels are shown on the y-axis (both arbitrary units)). **Ba** | Impulse patterns (upregulation or downregulation) followed by a return to basal levels. **Bb** | Sustained patterns in which the gene remains overexpressed or underexpressed. **C** | Expression programs often combine several of the basic patterns. Different genes are depicted in different line styles. **Ca** | A cascade in which an impulse response leads another gene to a sustained response. **Cb** | 'Just-in-time' activation as part of a response program. **Cc** | Cyclic process using recurrent just-in-time activation of specific genes. Part **Aa** is modified from REF. 4 © (2007) Macmillan Publishers Ltd. All rights reserved. Part **Ab** is modified from REF. 11. Part **Ac** is modified, with permission, from REF. 14 © (1998) The American Society for Cell Biology.

K-means
A clustering approach that searches for a specific number of clusters (k) maximizing a global target function. Clusters are defined by their centre. Iteratively, genes are assigned to the best-matching cluster and then the clusters' centre values are updated.

Causal modelling
A causal model asserts that a gene controls its target genes and changes their expression levels. This is in contrast to a model that merely identifies genes for which expression is correlated over time.

Phorbol myristate acetate
(PMA). Phorbol 12-myristate 13-acetate is a diester of phorbol that is frequently used to activate the signal transduction enzyme protein kinase C (PKC).

Just-in-time
A temporal expression pattern that is characterized by a match between the temporal activation of a set of genes and the time in which their products are required.

we briefly summarize the main insights. The major cellular transcriptional response to perturbations of the surrounding environment is in an impulse-like pattern of expression[3,64] (FIG. 2Aa). Impulse patterns in responses to environmental changes are common in all organisms, from bacteria to mammals[2,65–68]. Nevertheless some environmental signals can cause a sustained response. For example, starvation in yeast causes the cells to initiate sustained programs, such as quiescence, filamentation or sporulation[67].

Developmental processes. Gene expression has been monitored during the development of various organisms for which there is an access to embryonic material, including zebrafish[11], *Drosophila melanogaster*[10,13] and *Caenorhabditis elegans*[12]. In general, the outcomes of developmental processes are shifts from one steady state to another (FIG. 2Ab). Such changes are not restricted to developmental processes[57,69] but are common also in pathogenesis[70] and in immune responses[8,55]. This type of outcome is not surprising because in both developmental and differentiation time courses we are actually monitoring different cell types. Genes expressed at consecutive time points are often from similar cells, whereas distant time points may represent totally different cell types. Thus, a developmental time course is analogous to monitoring multiple cell types, some of which are more closely related than others, and thus a partial overlap between consecutive time points is frequently observed (FIG. 2Ab). As mentioned above, such sustained changes are frequently the outcome of a combination of short impulses followed by a sustained response (FIG. 2Ca).

The extent of stage-specific expression varies among organisms. In zebrafish only 22% of genes are found to be differentially expressed during development; some genes show short time spans of activity, whereas other genes have extended periods of expression[11]. In *C. elegans*, approximately half of the genes are found to be differentially expressed during development, and only a few genes show stage-specific expression (approximately 100 genes per stage)[12]. In *D. melanogaster*, most genes (86%) show significant changes in their expression levels during development, with approximately one-third of these genes showing a single major peak of expression and two-thirds being expressed at multiple stages. Interestingly, the re-use of genes at multiple developmental stages in *D. melanogaster* has a distinct pattern. Many of the genes that are expressed in the embryonic stage are re-expressed in the pupae, and many of the larval genes are re-expressed in adults[10].

Another type of developmental experiment involves monitoring changes in gene expression as a response to a signal that induces differentiation. For example, this includes the differentiation of mouse embryonic stem (ES) cells in response to the downregulation of NANOG[36], or the differentiation of human THP-1 cells into mature monocytes following PMA treatment[54]. These types of experiments resemble classical developmental work because the different time points capture diverse differentiation stages of the system rather than temporal differences in the response of a single cell type to the differentiation signal.

Cyclic processes. Various important biological process — most notably the cell division process, circadian rhythms and other reparative systems — operate by expressing, in a just-in-time manner, the set of required genes during each cycle[19,71] (FIG. 2Cc). As can be seen in FIG. 2Ac, for such processes a similar set of genes is activated (or repressed) at every iteration of the cyclic process. This cyclic behaviour is the basis for the identification of the genes that are involved in the process as well as the temporal and functional relationships between these gene sets. One of the first, and most influential, time-series gene expression data sets was obtained from the analysis of the cell cycle in budding yeast. This study identified 800 genes as cell-cycle-regulated (~15% of all yeast genes)[14]. Follow-up work studied the cell cycle in a range of organisms, including fission yeast[17], plants[15], mice and in human cancer cells and normal cells[16,18]. Conclusions regarding the large set of cell-cycle-regulated genes in each species, their just-in-time activation and their consistency in multiple cycles were the same[16]. Similar results were observed for other cyclic systems that were studied using time-series data[20]. As these studies relied on populations of cells, issues related to cell cycle arrest and synchronization have a key role in the success of these experiments[72,73] (BOX 2). Interestingly, even though the cell cycle is one of the most conserved biological systems, a comparison of expression experiments between budding and fission yeast did not identify a conserved set of cycling genes[17]. Issues related to differences in the experimental setup and the data analysis — including the methods for cell cycle arrest and release, and the choice of cutoff values for determining cyclic behaviour — may have affected this observation[74]. Another possible explanation, suggested by Jensen *et al.*[75], is that the just-in-time activation may apply to different members of the same protein complex in different species. Thus, even though orthologous genes have differential cycling patterns, complexes that contain the proteins encoded by these genes might still be activated in the same phases, thus leading to similar outcomes across a wide range of species.

Additional types of genomic time-series data

Several other temporal high-throughput data sets are now becoming available. These provide additional information that can be integrated with the expression data to model biological systems more accurately.

Transcription factor occupancy and epigenetic data. When studying gene regulation, additional levels of information are frequently used besides expression profiling. Data from experiments using chromatin immunoprecipitation followed by microarray (ChIP–chip) or by high-throughput sequencing (ChIP–seq) provide information about the genome-wide occupancy of DNA binding proteins, and such information is frequently incorporated into attempts to infer gene regulatory networks. Chromatin structure also has a central role in gene regulation and can now be measured by using ChIP–chip or ChIP–seq for various histone modifications and by using various methods to map DNA methylation across the genome. However, these chromatin features have almost always been monitored in a static way, owing to the difficulty in collecting a sufficient number of cells at multiple time points. Occasionally, temporal data have been obtained; for example, from ChIP–chip time courses for transcription factors during the *Schizosaccharomyces pombe* cell cycle[76] and during *D. melanogaster* development[77–79].

NANOG
A transcription factor that is crucial for the self-renewal of undifferentiated embryonic stem cells.

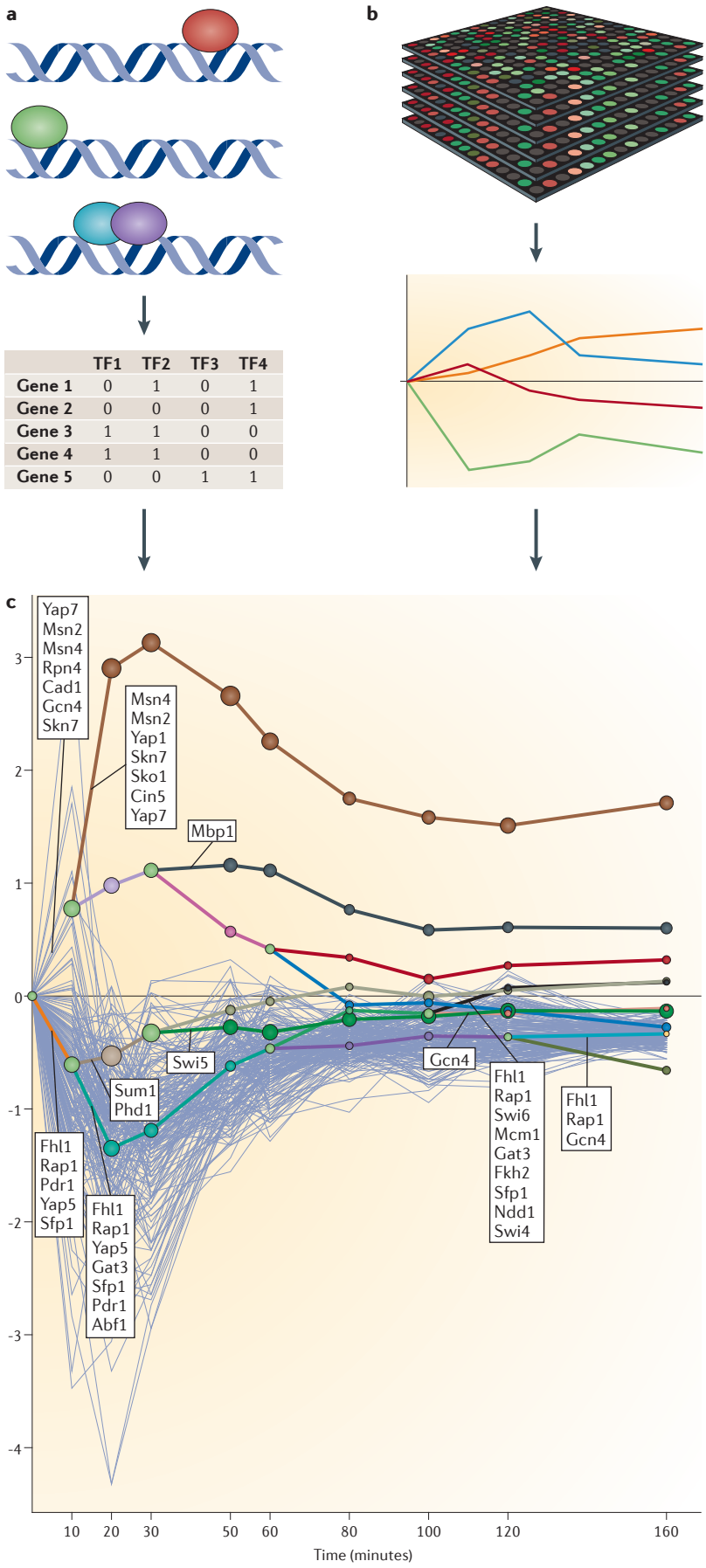
More comprehensive data sets were collected as part of the modENCODE project, in which a few transcription factors, RNA polymerase II (Pol II) and several histone marks were monitored at multiple time points over multiple developmental stages in *D. melanogaster*[13] and to a lesser extent also in *C. elegans*[12]. In mammals, the abundance of Pol II and of acetylation of histone H3 lysine 9 (H3K9) and H3K14 was monitored at multiple time points following NANOG depletion from mouse ES cells[36]. Finally, changes in DNA methylation, histone H3 trimethylated on lysine 27 (H3K27me3) levels and Pol II occupancy during mouse development were monitored in three cell lines (ES cells, neuronal precursor cells and terminally differentiated neurons) that represent consecutive differentiation states[80]. Future advancement both in sequencing technologies and in ChIP protocols will most probably change this situation; the development of additional computational tools will then be required to take advantage of such dynamic data.

Integrating time-series expression data with other 'omics' data. Time-series expression data provide a unique viewpoint regarding the dynamics of cellular activity. However, this is by no means a comprehensive picture of the activity in the cell. Modelling the regulatory and signalling networks that lead to the observed gene expression changes requires the integration of several additional types of biological data. Indeed, numerous recent studies have combined time-series data for mRNA expression, protein levels, protein–DNA interactions and chromatin features for modelling such networks. An early example is a study of *D. melanogaster* that combined a time-series expression analysis of mutant strains with ChIP–chip and *in situ* hybridization data to identify direct and indirect targets of the myocyte-specific enhancer factor 2 (MEF2) transcription factor during *D. melanogaster* development[77]. Using this integrated approach the researchers identified more than 200 direct targets of MEF2, which indicated a much broader role for MEF2 in myogenesis than was previously thought. Another example, a study of stem cell differentiation in mice[36], showed that although there was usually good temporal agreement between Pol II occupancy and mRNA expression levels, approximately half of the proteins did not exhibit good correlations between protein and mRNA levels over time; the authors attributed this to translational and post-translational regulation. A study using a similar approach[12] concluded that in *C. elegans* the initial activation of a gene in a developmental process occurs through Pol II recruitment, whereas later changes in its expression occur without changes in Pol II occupancy. Another type of temporal omics data, microRNA (miRNA) expression data, and mRNA expression data from samples taken at the same time points were combined to identify regulatory interactions[81]. The authors analysed miRNAs and their predicted target mRNAs and identified those pairs that displayed significant time-lagged correlation. Such mRNAs were thus classified as temporally regulated in that experiment.

Figure 3 | **Identifying transcription factor dynamics with DREM.** Joint analysis of static transcription factor (TF) binding interactions and dynamic gene expression data enables Dynamic Regulatory Events Miner (DREM)[85] to reconstruct the dynamic regulatory network underlying the response of yeast to hydrogen peroxide. **a** | Protein–DNA binding interactions, from static experiments, are used to define the set of genes that each transcription factor may potentially control. These interactions can be obtained from chromatin immunoprecipitation followed by microarray (ChIP–chip), ChIP followed by high-throughput sequencing (ChIP–seq), or from computational predictions based on DNA sequence analysis. **b** | The time-series gene expression input captures the dynamic transcriptional changes (upper panel). Individual genes are grouped by DREM on the basis of the similarity of their expression trajectories and the transcription factors that regulate them (lower panel) (note that the colours of the hypothetical genes in panel **b** do not correspond to the colours in panel **c**). **c** | The DREM model for the response of yeast to hydrogen peroxide (data from REF. 2). The thick, coloured, lines depict average expression profiles for each cluster of genes. DREM associates transcription factors with points at which these clusters diverge (light green circles). These transcription factors are predicted to control the split (by activating or repressing a subset of the genes). Transcription factor names are listed on the branches out of the splits, showing the time points at which the transcription factors are predicted to be active. We selected one path (the path that is initially the most strongly repressed and later recovers to end with the light blue segment that is second from the bottom) in the network whose member gene expression profiles are shown as thin blue lines. This group of genes is repressed for 20 minutes and then enters a recovery phase, which is a common pattern of stress responses.

However, to date, most of the available high-throughput data have been measured at a single time point. Thus, a major computational challenge has been to integrate these static data sets with temporal data sets to reconstruct dynamic networks. Early work along these lines[82] identified differentially expressed genes at different time points and used these to assign, for each time point, a set of transcription factors that were bound to the differentially expressed genes, even though the transcription factor binding was only determined in a static experiment. Several studies have relied on regression analysis to identify potential regulators of differentially expressed genes. These methods attempt to explain the temporal response of genes as a combination of the temporal activity profile of transcription factors that are known to regulate them and of sequence motifs in their upstream regions. By solving a large-scale optimization problem (in which several genes are assumed to be regulated by a small number of transcription factors) researchers aim to identify this (hidden) activity profile and to determine the set of interactions at the various time points[71,83,84]. Such an approach was used to study the regulation of monocyte differentiation[54]. In that study, a Motif Activity Response Analysis (MARA) framework was developed to associate promoter activity with DNA binding motifs

Motif Activity Response Analysis
(MARA). A method for inferring DNA-binding-motif activity and for linking motifs to promoters. MARA models promoter expression as a linear function of motif activity and the number of functional binding sites.



over time. The term 'activity motifs'[3] has been used to describe the activity of a gene on the basis of its temporal mRNA or protein expression levels. These activity motifs are then used to identify, in static signalling and metabolism networks, cascades of proteins that are also temporally ordered on the basis of their expression values. By also analysing the binding affinity to specific transcription factors, these motifs can partially explain mechanisms that are used for ordered transcription.

A different approach, which integrates static protein–DNA interaction or motif data with time-series expression data in a unified model, is the Dynamic Regulatory Events Miner (DREM)[85]. DREM searches the time-series data for bifurcation events, which are time points in the time series at which a set of genes transitions from being co-expressed to having divergent expression (FIG. 3). These events are annotated with transcription factors that are predicted to regulate the divergently expressed sets of genes at the time of the split, on the basis of the set of genes that they bind in the static experiment or on their activity motif. DREM uses an extension of HMMs to search for these split events and to identify the most likely transcription factors that are responsible for each split. Using the dynamic network that was inferred by DREM, researchers were able to identify master regulators and secondary response regulators and to determine differences in their regulation and binding strategies in various stress responses in yeast and *E. coli*[85]. DREM was also used to identify mammalian transcription factors that are differentially targeted during various infections[5]. For developmental studies, the ability of DREM to assign transcription factors to specific developmental stages in flies helped to determine their regulatory roles and modes of action (as repressors or activators)[13] and was also used as the basis for follow-up studies of stem cell differentiation[86].

Conclusions and future directions

Time-series expression experiments provide a wealth of information regarding the complete set of genes that are activated, their dynamics and interactions, the role that different genes have in the process and the differences in processes within and between species. By integrating such data with other temporal or static omics data sets researchers can obtain a global view of the dynamic networks that are activated in cells. Given the importance of dynamic biological processes, the insights and models that are derived from current high-throughput dynamic data and our increased ability to study various molecular mechanisms over time, temporal data will play an even bigger part in future studies.

We expect that both experimental and computational advances will lead to better modelling and understanding of dynamic systems in the cell. As has been stated above, a current limitation of high-throughput expression analyses is that they must be carried out on a population of cells. Another technical limitation is the sampling rate. Owing to the high cost of genomic experiments it is not practical to sample the system very densely. These two limitations can be overcome by using time-lapse microscopy for measuring the dynamic

Phosphoproteomic
The use of experimental technology such as mass spectrometry to identify and quantify protein or peptide phosphorylation.

changes in gene expression. Combining the use of fluorescent protein reporters, time-lapse microscopy and quantitative image analysis has enabled the direct observation of multiple cellular components over time in individual cells[7]. This approach was used initially in bacteria for studying flagella formation[87] and metabolic pathways[61], and more recently was also applied to mammalian cells[88]. Currently, such experiments can only monitor the products of a few genes simultaneously in a cell. However, improvements in this approach and in single-cell sequencing techniques may enable high-throughput temporal studies at a single-cell level.

Regarding computational future directions, recovering temporal transcription factor activity is only a first step in understanding the complex dynamic processes that drive changes in gene expression. For example, when responding to an external signal, the transcription factors themselves are activated by signalling pathways that initially detect the stimulus. A major computational challenge will be to infer these signalling cascades and connect them to the downstream transcription factors and transcriptional dynamics. Existing approaches for integrating signalling and regulatory networks rely on gene knockouts[89], genetic screen hits[90], phosphoproteomic data[91], or known plasma membrane-localized receptors[92] to define a set of source nodes in a protein–protein interaction network and then search for explanatory paths from these sources to the transcription factors of interest. Despite their successful applications in diverse biological conditions, these methods are limited because they are inherently static. As discussed in recent reviews[93,94], considering the dynamic behaviour of signalling networks is vital for understanding their biological properties (such as the specificity of different responses), and improvements in experimental technology have enabled the quantification of signalling network dynamics. Because signalling and transcriptional regulatory networks generally operate on different timescales, temporal gene expression data alone are insufficient for inferring signalling dynamics. Learning the temporal behaviour of signalling pathways will probably require combining experimental measurements of protein dynamics (such as the dynamic phosphorylation data[95] that were previously used to validate predicted signalling pathways[92]) with time-series gene expression data. We expect that developing unified computational models of signalling and transcriptional dynamics will yield more complete representations and additional biological insights.

1. Barrett, T. *et al.* NCBI GEO: archive for functional genomics data sets-10 years on. *Nucleic Acids Res.* **39**, D1005–D1010 (2011).
2. Gasch, A. P. *et al.* Genomic expression programs in the response of yeast cells to environmental changes. *Mol. Biol. Cell* **11**, 4241–4257 (2000).
An extensive study of the yeast response to various stresses, which revealed a large set of general environmental stress response genes.
3. Chechik, G. *et al.* Activity motifs reveal principles of timing in transcriptional control of the yeast metabolic network. *Nature Biotech.* **26**, 1251–1259 (2008).
4. Amit, I. *et al.* A module of negative feedback regulators defines growth factor signaling. *Nature Genet.* **39**, 503–512 (2007).
This study used a clever design strategy to sample and integrate temporal immune response data. It identified several cascades that are activated as part of mammalian immune response.
5. Zinman, G. *et al.* Large scale comparison of innate responses to viral and bacterial pathogens in mouse and macaque. *PLoS ONE* **6**, e22401 (2011).
6. Tang, F., Lao, K. & Surani, M. A. Development and applications of single-cell transcriptome analysis. *Nature Methods* **8**, S6–S11 (2011).
7. Locke, J. C. W. & Elowitz, M. B. Using movies to analyse gene circuit dynamics in single cells. *Nature Rev. Microbiol.* **7**, 383–392 (2009).
8. Amit, I. *et al.* Unbiased reconstruction of a mammalian transcriptional network mediating pathogen responses. *Science* **326**, 257–263 (2009).
9. Huang, Y. *et al.* Temporal dynamics of host molecular responses differentiate symptomatic and asymptomatic influenza A infection. *PLoS Genet.* **7**, e1002234 (2011).
An experimental and computational analysis of influenza infection in human volunteers that identified dynamic expression patterns that discriminate symptomatic and asymptomatic reactions.
10. Arbeitman, M. N. *et al.* Gene expression during the life cycle of *Drosophila melanogaster*. *Science* **297**, 2270–2275 (2002).
One of the first developmental studies using time-series expression data. It identified several cascades of genes that are expressed during fly development.
11. Mathavan, S. *et al.* Transcriptome analysis of zebrafish embryogenesis using microarrays. *PLoS Genet.* **1**, 260–276 (2005).
12. Gerstein, M. B. *et al.* Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE project. *Science* **330**, 1775–1787 (2010).
13. Roy, S. *et al.* Identification of functional elements and regulatory circuits by *Drosophila* modENCODE. *Science* **330**, 1787–1797 (2010).
A comprehensive study of fly development that collected and integrated several types of temporal and static omics data.
14. Spellman, P. T. *et al.* Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol. Biol. Cell* **9**, 3273–3297 (1998).
Among the first studies of cyclic responses using time-series experiments. It identified a large set of yeast cycling genes and it served as the basis for several computational methods that have been developed to study such data.
15. Menges, M., Hennig, L., Gruissem, W. & Murray, J. A. H. Cell cycle-regulated gene expression in *Arabidopsis*. *J. Biol. Chem.* **277**, 41987–42002 (2002).
16. Whitfield, M. L. *et al.* Identification of genes periodically expressed in the human cell cycle and their expression in tumors. *Mol. Biol. Cell* **13**, 1977–2000 (2002).
17. Rustici, G. *et al.* Periodic gene expression program of the fission yeast cell cycle. *Nature Genet.* **36**, 809–817 (2004).
18. Bar-Joseph, Z. *et al.* Genome-wide transcriptional analysis of the human cell cycle identifies genes differentially regulated in normal and cancer cells. *Proc. Natl Acad. Sci. USA* **105**, 955–960 (2008).
19. McDonald, M. J. & Rosbash, M. Microarray analysis and organization of circadian gene expression in *Drosophila*. *Cell* **107**, 567–578 (2001).
20. Schaffer, R. *et al.* Microarray analysis of diurnal and circadian-regulated genes in *Arabidopsis*. *Plant Cell* **13**, 113–123 (2001).
21. Storch, K.-F. *et al.* Extensive and divergent circadian gene expression in liver and heart. *Nature* **417**, 78–83 (2002).
22. Pauli, A. *et al.* Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. *Genome Res.* **22**, 577–591 (2012).
23. Otto, T. D. *et al.* New insights into the blood-stage transcriptome of *Plasmodium falciparum* using RNA-Seq. *Mol. Microbiol.* **76**, 12–24 (2010).
24. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotech.* **28**, 511–515 (2010).
25. Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotech.* **29**, 644–652 (2011).
26. Marioni, J. C., Mason, C. E., Mane, S. M., Stephens, M. & Gilad, Y. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res.* **18**, 1509–1517 (2008).
27. Kelley, D. R., Schatz, M. C. & Salzberg, S. L. Quake: quality-aware detection and correction of sequencing errors. *Genome Biol.* **11**, R116 (2010).
28. Shalem, O. *et al.* Transient transcriptional responses to stress are generated by opposing effects of mRNA production and degradation. *Mol. Syst. Biol.* **4**, 223 (2008).
29. Bolstad, B. M., Irizarry, R. A., Astrand, M. & Speed, T. P. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**, 185–193 (2003).
30. Li, C. & Hung Wong, W. Model-based analysis of oligonucleotide arrays: model validation, design issues and standard error application. *Genome Biol.* **2**, RESEARCH0032 (2001).
31. Ernst, J. *et al.* IL-3 and oncogenic Abl regulate the myeloblast transcriptome by altering mRNA stability. *PLoS ONE* **4**, e7469 (2009).
32. Bar-Joseph, Z., Gerber, G., Simon, I., Gifford, D. K. & Jaakkola, T. S. Comparing the continuous representation of time-series expression profiles to identify differentially expressed genes. *Proc. Natl Acad. Sci. USA* **100**, 10146–10151 (2003).
33. Bar-Joseph, Z., Gerber, G. K., Gifford, D. K., Jaakkola, T. S. & Simon, I. Continuous representations of time-series gene expression data. *J. Comput. Biol.* **10**, 341–356 (2003).
34. Kaminski, N. & Bar-Joseph, Z. A patient-gene model for temporal expression profiles in clinical studies. *J. Comput. Biol.* **14**, 324–338 (2007).
35. Calvano, S. E. *et al.* A network-based analysis of systemic inflammation in humans. *Nature* **437**, 1032–1037 (2005).
36. Lu, R. *et al.* Systems-level dynamic analyses of fate change in murine embryonic stem cells. *Nature* **462**, 358–362 (2009).
This study measured several types of temporal high-throughput data sets in stem cell differentiation. It determined that correlation between dynamic protein and mRNA expression levels is lower than was previously thought.
37. Nau, G. J. *et al.* Human macrophage activation programs induced by bacterial pathogens. *Proc. Natl Acad. Sci. USA* **99**, 1503–1508 (2002).

38. Shapira, S. D. *et al.* A physical and regulatory map of host-influenza interactions reveals pathways in H1N1 infection. *Cell* **139**, 1255–1267 (2009).
39. Tusher, V. G., Tibshirani, R. & Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl Acad. Sci. USA* **98**, 5116–5121 (2001).
40. Storey, J. D., Xiao, W., Leek, J. T., Tompkins, R. G. & Davis, R. W. Significance analysis of time course microarray experiments. *Proc. Natl Acad. Sci. USA* **102**, 12837–12842 (2005).
41. Aryee, M. J., Gutiérrez-Pabello, J. A., Kramnik, I., Maiti, T. & Quackenbush, J. An improved empirical bayes approach to estimating differential gene expression in microarray time-course data: BETR (Bayesian Estimation of Temporal Regulation). *BMC Bioinformatics* **10**, 409 (2009).
42. Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl Acad. Sci. USA* **95**, 14863–14868 (1998).
43. Ramoni, M. F., Sebastiani, P. & Kohane, I. S. Cluster analysis of gene expression dynamics. *Proc. Natl Acad. Sci. USA* **99**, 9121–9126 (2002).
44. Schliep, A., Schönhuth, A. & Steinhoff, C. Using hidden Markov models to analyze gene expression time course data. *Bioinformatics* **19**, i255–i263 (2003).
45. Ernst, J., Nau, G. J. & Bar-Joseph, Z. Clustering short time series gene expression data. *Bioinformatics* **21**, i159–i168 (2005).
46. Baranzini, S. E. *et al.* Transcription-based prediction of response to IFNβ using supervised computational methods. *PLoS Biol.* **3**, e2 (2005).
47. Borgwardt, K. M., Vishwanathan, S. V. N. & Kriegel, H.-P. Class prediction from time series gene expression profiles using dynamical systems kernels. *Pac. Symp. Biocomput.* **2006**, 547–558 (2006).
48. Lin, T., Kaminski, N. & Bar-Joseph, Z. Alignment and classification of time series gene expression in clinical studies. *Bioinformatics* **24**, i147–i155 (2008).
49. Costa, I. G., Schönhuth, A., Hafemeister, C. & Schliep, A. Constrained mixture estimation for analysis and robust classification of clinical time series. *Bioinformatics* **25**, i6–i14 (2009).
50. Qian, J., Dolled-Filhart, M., Lin, J., Yu, H. & Gerstein, M. Beyond synexpression relationships: local clustering of time-shifted and inverted gene expression profiles identifies new, biologically relevant interactions. *J. Mol. Biol.* **314**, 1053–1066 (2001).
51. Mukhopadhyay, N. D. & Chatterjee, S. Causality and pathway search in microarray time series experiment. *Bioinformatics* **23**, 442–449 (2007).
52. Shi, Y., Mitchell, T. & Bar-Joseph, Z. Inferring pairwise regulatory relationships from multiple time series datasets. *Bioinformatics* **23**, 755–763 (2007).
53. Zhu, J. *et al.* Characterizing dynamic changes in the human blood transcriptional network. *PLoS Comput. Biol.* **6**, e1000671 (2010).
54. The FANTOM Consortium & Riken Omics Science Center. The transcriptional network that controls growth arrest and differentiation in a human myeloid leukemia cell line. *Nature Genet.* **41**, 553–562 (2009).
55. Ramsey, S. A. *et al.* Uncovering a macrophage transcriptional program by integrating evidence from motif scanning and expression dynamics. *PLoS Comput. Biol.* **4**, e1000021 (2008).
56. Davidson, E. H. Emerging properties of animal gene regulatory networks. *Nature* **468**, 911–920 (2010).
57. Oliveri, P., Tu, Q. & Davidson, E. H. Global regulatory logic for specification of an embryonic cell lineage. *Proc. Natl Acad. Sci. USA* **105**, 5955–5962 (2008).
58. Simon, I. *et al.* Serial regulation of transcriptional regulators in the yeast cell cycle. *Cell* **106**, 697–708 (2001).
59. Ihmels, J., Levy, R. & Barkai, N. Principles of transcriptional control in the metabolic network of *Saccharomyces cerevisiae*. *Nature Biotech.* **22**, 86–92 (2004).
60. Alon, U. Network motifs: theory and experimental approaches. *Nature Rev. Genet.* **8**, 450–461 (2007).
61. Zaslaver, A. *et al.* Just-in-time transcription program in metabolic pathways. *Nature Genet.* **36**, 486–491 (2004).
62. Kalir, S., Mangan, S. & Alon, U. A coherent feed-forward loop with a SUM input function prolongs flagella expression in *Escherichia coli*. *Mol. Syst. Biol.* **1**, 2005.0006 (2005).
63. Yosef, N. & Regev, A. Impulse control: temporal dynamics in gene transcription. *Cell* **144**, 886–896 (2011).
A comprehensive review of impulse response in time-series gene expression data. It highlights both basic response patterns and the combination of such patterns in specific response programs.
64. Chechik, G. & Koller, D. Timing of gene expression responses to environmental changes. *J. Comput. Biol.* **16**, 279–290 (2009).
65. Braun, E. & Brenner, N. Transient responses and adaptation to steady state in a eukaryotic gene regulation system. *Phys. Biol.* **1**, 67–76 (2004).
66. Murray, J. I. *et al.* Diverse and specific gene expression responses to stresses in cultured human cells. *Mol. Biol. Cell* **15**, 2361–2374 (2004).
67. Lõpez-Maury, L., Marguerat, S. & Bähler, J. Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation. *Nature Rev. Genet.* **9**, 583–593 (2008).
68. Litvak, V. *et al.* Function of C/EBPδ in a regulatory circuit that discriminates between transient and persistent TLR4-induced signals. *Nature Immunol.* **10**, 437–443 (2009).
69. Basma, H. *et al.* Differentiation and transplantation of human embryonic stem cell-derived hepatocytes. *Gastroenterology* **136**, 990–999 (2009).
70. Iliopoulos, D., Hirsch, H. A. & Struhl, K. An epigenetic switch involving NF-κB, Lin28, Let-7 microRNA, and IL6 links inflammation to cell transformation. *Cell* **139**, 693–706 (2009).
71. Seok, J., Xiao, W., Moldawer, L. L., Davis, R. W. & Covert, M. W. A dynamic network of transcription in LPS-treated human subjects. *BMC Syst. Biol.* **3**, 78 (2009).
72. Shedden, K. & Cooper, S. Analysis of cell-cycle-specific gene expression in human cells as determined by microarrays and double-thymidine block synchronization. *Proc. Natl Acad. Sci. USA* **99**, 4379–4384 (2002).
73. Simon, I., Siegfried, Z., Ernst, J. & Bar-Joseph, Z. Combined static and dynamic analysis for determining the quality of time-series expression profiles. *Nature Biotech.* **23**, 1503–1508 (2005).
74. Lu, Y. *et al.* Combined analysis reveals a core set of cycling genes. *Genome Biol.* **8**, R146 (2007).
75. Jensen, L. J., Jensen, T. S., de Lichtenberg, U., Brunak, S. & Bork, P. Co-evolution of transcriptional and post-translational cell-cycle regulation. *Nature* **443**, 594–597 (2006).
A study of the conservation of cell cycle regulation across species. It determined that temporal complex activation, rather than specific gene expression, is conserved between species.
76. Eshaghi, M. *et al.* Deconvolution of chromatin immunoprecipitation-microarray (ChIP-chip) analysis of MBF occupancies reveals the temporal recruitment of Rep2 at the MBF target genes. *Eukaryot. Cell* **10**, 130–141 (2011).
77. Sandmann, T. *et al.* A temporal map of transcription factor activity: Mef2 directly regulates target genes at all stages of muscle development. *Dev. Cell* **10**, 797–807 (2006).
78. Jakobsen, J. S. *et al.* Temporal ChIP-on-chip reveals Biniou as a universal regulator of the visceral muscle transcriptional network. *Genes Dev.* **21**, 2448–2460 (2007).
79. Liu, Y.-H. *et al.* A systematic analysis of Tinman function reveals Eya and JAK-STAT signaling as essential regulators of muscle development. *Dev. Cell* **16**, 280–291 (2009).
80. Mohn, F. *et al.* Lineage-specific polycomb targets and *de novo* DNA methylation define restriction and potential of neuronal progenitors. *Mol. Cell* **30**, 755–766 (2008).
81. Jayaswal, V., Lutherborrow, M., Ma, D. D. F. & Hwa Yang, Y. Identification of microRNAs with regulatory potential using a matched microRNA-mRNA time-course data. *Nucleic Acids Res.* **37**, e60 (2009).
82. Luscombe, N. M. *et al.* Genomic analysis of regulatory network dynamics reveals large topological changes. *Nature* **431**, 308–312 (2004).
This was among the first studies to integrate time-series and static interaction data to determine temporal activity of transcription factors.
83. Liao, J. C. *et al.* Network component analysis: reconstruction of regulatory signals in biological systems. *Proc. Natl Acad. Sci. USA* **100**, 15522–15527 (2003).
84. Bansal, M., Della Gatta, G. & di Bernardo, D. Inference of gene regulatory networks and compound mode of action from time course gene expression profiles. *Bioinformatics* **22**, 815–822 (2006).
85. Ernst, J., Vainas, O., Harbison, C. T., Simon, I. & Bar-Joseph, Z. Reconstructing dynamic regulatory maps. *Mol. Syst. Biol.* **3**, 74 (2007).
This paper shows that integrating static protein–DNA binding interactions with temporal gene expression data allows the reconstruction of dynamic networks.
86. Mendoza-Parra, M. A., Walia, M., Sankar, M. & Gronemeyer, H. Dissecting the retinoid-induced differentiation of F9 embryonal stem cells by integrative genomics. *Mol. Syst. Biol.* **7**, 538 (2011).
87. Kalir, S. *et al.* Ordering genes in a flagella pathway by analysis of expression kinetics from living bacteria. *Science* **292**, 2080–2083 (2001).
88. Sigal, A. *et al.* Generation of a fluorescently labeled endogenous protein library in living human cells. *Nature Protoc.* **2**, 1515–1527 (2007).
89. Yeang, C.-H., Ideker, T. & Jaakkola, T. Physical network models. *J. Comput. Biol.* **11**, 243–262 (2004).
90. Yeger-Lotem, E. *et al.* Bridging high-throughput genetic and transcriptional data reveals cellular responses to α-synuclein toxicity. *Nature Genet.* **41**, 316–323 (2009).
91. Huang, S. C. & Fraenkel, E. Integration of proteomic, transcriptional, and interactome data reveals hidden signaling components. *Sci. Signal.* **2**, ra40 (2009).
92. Vinayagam, A. *et al.* A directed protein interaction network for investigating intracellular signal transduction. *Sci. Signal.* **4**, rs8 (2011).
93. White, F. M. Quantitative phosphoproteomic analysis of signaling network dynamics. *Curr. Opin. Biotechnol.* **19**, 404–409 (2008).
94. Kholodenko, B. N., Hancock, J. F. & Kolch, W. Signalling ballet in space and time. *Nature Rev. Mol. Cell Biol.* **11**, 414–426 (2010).
95. Olsen, J. V. *et al.* Global, *in vivo*, and site-specific phosphorylation dynamics in signaling networks. *Cell* **127**, 635–648 (2006).
96. Gaucher, D. *et al.* Yellow fever vaccine induces integrated multilineage and polyfunctional immune responses. *J. Exp. Med.* **205**, 3119–3131 (2008).
97. Cho, R. J. *et al.* Transcriptional regulation and function during the human cell cycle. *Nature Genet.* **27**, 48–54 (2001).
98. Lu, X., Zhang, W., Qin, Z. S., Kwast, K. E. & Liu, J. S. Statistical resynchronization and Bayesian detection of periodically expressed genes. *Nucleic Acids Res.* **32**, 447–455 (2004).
99. Bar-Joseph, Z., Farkash, S., Gifford, D. K., Simon, I. & Rosenfeld, R. Deconvolving cell cycle expression data with complementary information. *Bioinformatics* **20**, i23–i30 (2004).
100. Siegal-Gaskins, D., Ash, J. N. & Crosson, S. Model-based deconvolution of cell cycle time-series data reveals gene expression details at high resolution. *PLoS Comput. Biol.* **5**, e1000460 (2009).
101. Klutstein, M. *et al.* Combination of genomic approaches with functional genetic experiments reveals two modes of repression of yeast middle-phase meiosis genes. *BMC Genomics* **11**, 478 (2010).
102. Smith, A. A., Vollrath, A., Bradfield, C. A. & Craven, M. Similarity queries for temporal toxicogenomic expression profiles. *PLoS Comput. Biol.* **4**, e1000116 (2008).
103. Hafemeister, C., Costa, I. G., Schönhuth, A. & Schliep, A. Classifying short gene expression time-courses with Bayesian estimation of piecewise constant functions. *Bioinformatics* **27**, 946–952 (2011).
104. Desai, K. H. *et al.* Dissecting inflammatory complications in critically injured patients by within-patient gene expression changes: a longitudinal clinical genomics study. *PLoS Med.* **8**, e1001093 (2011).
105. Taylor, M. W. *et al.* Changes in gene expression during pegylated interferon and ribavirin therapy of chronic hepatitis C virus distinguish responders from nonresponders to antiviral therapy. *J. Virol.* **81**, 3391–3401 (2007).
106. Liew, C.-C., Ma, J., Tang, H.-C., Zheng, R. & Dempsey, A. A. The peripheral blood transcriptome dynamically reflects system wide biology: a potential diagnostic tool. *J. Lab. Clin. Med.* **147**, 126–132 (2006).
107. Smyth, G. in *Bioinformatics and Computational Biology Solutions Using R And Bioconductor* 397–420 (Springer, 2005).

108. Leek, J. T., Monsen, E., Dabney, A. R. & Storey, J. D. EDGE: extraction and analysis of differential gene expression. *Bioinformatics* **22**, 507–508 (2006).
109. Woo, S., Leek, J. T. & Storey, J. D. A computationally efficient modular optimal discovery procedure. *Bioinformatics* **27**, 509–515 (2011).
110. Ernst, J. & Bar-Joseph, Z. STEM: a tool for the analysis of short time series gene expression data. *BMC Bioinformatics* **7**, 191 (2006).
111. Schliep, A., Steinhoff, C. & Schönhuth, A. Robust inference of groups in gene expression time-courses using mixtures of HMMs. *Bioinformatics* **20**, i283–i289 (2004).
112. Magni, P., Ferrazzi, F., Sacchi, L. & Bellazzi, R. TimeClust: a clustering tool for gene expression time series. *Bioinformatics* **24**, 430–432 (2008).
113. Sivriver, J., Habib, N. & Friedman, N. An integrative clustering and modeling algorithm for dynamical gene expression data. *Bioinformatics* **27**, i392–i400 (2011).
114. Sinha, A. & Markatou, M. A. Platform for Processing Expression of Short Time Series (PESTS). *BMC Bioinformatics* **12**, 13 (2011).
115. Bonneau, R. *et al.* The Inferelator: an algorithm for learning parsimonious regulatory networks from systems-biology data sets *de novo*. *Genome Biol.* **7**, R36 (2006).
116. Galbraith, S. J., Tran, L. M. & Liao, J. C. Transcriptome network component analysis with limited microarray data. *Bioinformatics* **22**, 1886–1894 (2006).
117. Schaffter, T., Marbach, D. & Floreano, D. GeneNetWeaver: *in silico* benchmark generation and performance profiling of network inference methods. *Bioinformatics* **27**, 2263–2270 (2011).

Acknowledgements
This work was supported in part by the US National Institutes of Health (NIH) grant 1RO1 GM085022 to Z.B.-J. Research in the laboratory of I.S. is supported by the Israel Science Foundation (grant 567/10), the German–Israeli Foundation (grant 998/2008), the Weinkselbaum family medical research fund and the European Research Council Starting Grant (281306).

Competing interests statement
The authors declare no competing financial interests.

FURTHER INFORMATION
ModENCODE: http://www.modencode.org

ALL LINKS ARE ACTIVE IN THE ONLINE PDF