Bioinformatics, 32(6), 2016, 884–892
doi: 10.1093/bioinformatics/btv673
Advance Access Publication Date: 14 November 2015
Original Paper

OXFORD

Systems biology

CHRONOS: a time-varying method for microRNA-mediated subpathway enrichment analysis

Aristidis G. Vrahatis[1,2], Konstantina Dimitrakopoulou[3], Panos Balomenos[1,2], Athanasios K. Tsakalidis[1] and Anastasios Bezerianos[2,4,*]

[1]Department of Computer Engineering and Informatics, [2]Department of Medical Physics, School of Medicine, University of Patras, Patras 26500, Greece, [3]Centre for Cancer Biomarkers CCBIO and Computational Biology Unit, Department of Informatics, University of Bergen, Bergen, Norway and [4]SINAPSE Institute, Center of Life Sciences, National University of Singapore, Singapore 117456

*To whom correspondence should be addressed.

Associate Editor: Igor Jurisica

Received on July 24, 2015; revised on October 19, 2015; accepted on November 9, 2015

Abstract

Motivation: In the era of network medicine and the rapid growth of paired time series mRNA/microRNA expression experiments, there is an urgent need for pathway enrichment analysis methods able to capture the time- and condition-specific 'active parts' of the biological circuitry as well as the microRNA impact. Current methods ignore the multiple dynamical 'themes'—in the form of enriched biologically relevant microRNA-mediated subpathways—that determine the functionality of signaling networks across time.

Results: To address these challenges, we developed time-vaRying enriCHment integrOmics Subpathway aNalysis tOol (CHRONOS) by integrating time series mRNA/microRNA expression data with KEGG pathway maps and microRNA-target interactions. Specifically, microRNA-mediated subpathway topologies are extracted and evaluated based on the temporal transition and the fold change activity of the linked genes/microRNAs. Further, we provide measures that capture the structural and functional features of subpathways in relation to the complete organism pathway atlas. Our application to synthetic and real data shows that CHRONOS outperforms current subpathway-based methods into unraveling the inherent dynamic properties of pathways.

Availability and implementation: CHRONOS is freely available at http://biosignal.med.upatras.gr/chronos/.

Contact: tassos.bezerianos@nus.edu.sg.

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

In the last decade, pathway-based approaches have become the first choice in complex disease analysis for gaining more delicate biological insights and have enabled the comprehension of the higher-order functions of the biological systems (Khatri *et al.*, 2012). In parallel, the increasing number of time course experiments has opened new avenues for pathway analysis to unravel the dynamic features of the interdependencies among a cell's molecular components leading to deeper characterization of the functional, molecular and causal relationships among apparently distinct phenotypes (Barabási *et al.*, 2011). Toward this direction, few approaches have considered the dynamic aspects of pathways (Khatri *et al.*, 2012) and, to our knowledge, no method has been developed to deal with their context-dependent and systematic temporal variations.

© The Author 2015. Published by Oxford University Press. All rights reserved. For Permissions, please e-mail: journals.permissions@oup.com

Downloaded from http://bioinformatics.oxfordjournals.org/ at University of California, Los Angeles on August 30, 2016

The first generation of pathway-based approaches related expression data to pathways by examining the entire pathway and well-established methods in this category are GSEA (Subramanian *et al.*, 2005) and SPIA (Tarca *et al.*, 2009). The next generation shifted the focus toward subpathways (local areas of the entire biological pathway), which represent the underlying biological phenomena more accurately, and have emerged as even more targeted and context-specific molecular candidate communities for the treatment of complex diseases (Chen *et al.*, 2011; Haynes *et al.*, 2013; Jacob *et al.*, 2012; Li *et al.*, 2012a, 2015; Martini *et al.*, 2013; Nam *et al.*, 2014; Sebastian-Leon *et al.*, 2014). A representative example is the Subpathway-GM method which identifies key metabolic subpathways based on information from genes and metabolites by searching for similarities of signature nodes within the pathway structure (Li *et al.*, 2013). Another relevant method is the topology enrichment analysis framework (TEAK) (Judeh *et al.*, 2013) which extracts linear and non-linear subpathways and scores them using the Bayes Net Toolbox to fit a context-specific Gaussian Bayesian network for each subpathway. However, these approaches do not take the temporal aspects of pathways into consideration and ignore possible perturbations on intermediate time points. Moving forward, timeClip combines dimension reduction techniques and graph decomposition theory to explore and identify the portion of pathways that is mostly time-dependent (Martini *et al.*, 2014). Although the temporal variation of the pathway is explored so as to identify the most time-dependent subpathway, information about the regulation between nodes is not included and the exact time period a subpathway is active is not provided.

In parallel, nowadays genome-wide expression studies of the endogenous non-coding small RNAs (microRNAs) have given a strong impulse to the comprehension of cell regulatory mechanisms (Calura *et al.*, 2014; Dimitrakopoulou *et al.*, 2015). Recent studies have examined the microRNA (miRNA) regulation upon pathways by showing that the perturbations on pathways does not necessarily occur at the overall level but more likely at specific areas of the topology (subpathways) (Li *et al.*, 2012b, 2014; Ooi *et al.*, 2011; Wu *et al.*, 2013; Zhang *et al.*, 2014).

Toward addressing these challenges, we introduce the time-vaRying enriCHment integrOmics Subpathway aNalysis tOol (CHRONOS) which detects significantly enriched miRNA-mediated subpathways per time point from paired miRNA/mRNA time series expression profiles (Fig. 1). Our key assumption is that during the time course, different subareas of the pathway topology are active and the transitions of subpathway activity at time t depend solely on time $t-1$ (Jethava *et al.*, 2011). We illustrate the enhanced performance of CHRONOS over other state-of-the-art subpathway-based methods based on synthetic time series expression data. Finally, we evaluate our method on a publicly available time series microarray dataset (Nazarov *et al.*, 2013) which investigates the transcriptional changes of miRNA and mRNA expression levels over time after activation of the Janus kinase/Signal transducer and activator of transcription (Jak/STAT) pathway by interferon-gamma stimulation (IFN-g) of melanoma cells. Our results show how time-evolving subpathway analysis generates new robust and dynamically context-rich biological hypotheses that complement those generated by other classical bioinformatics pipelines.

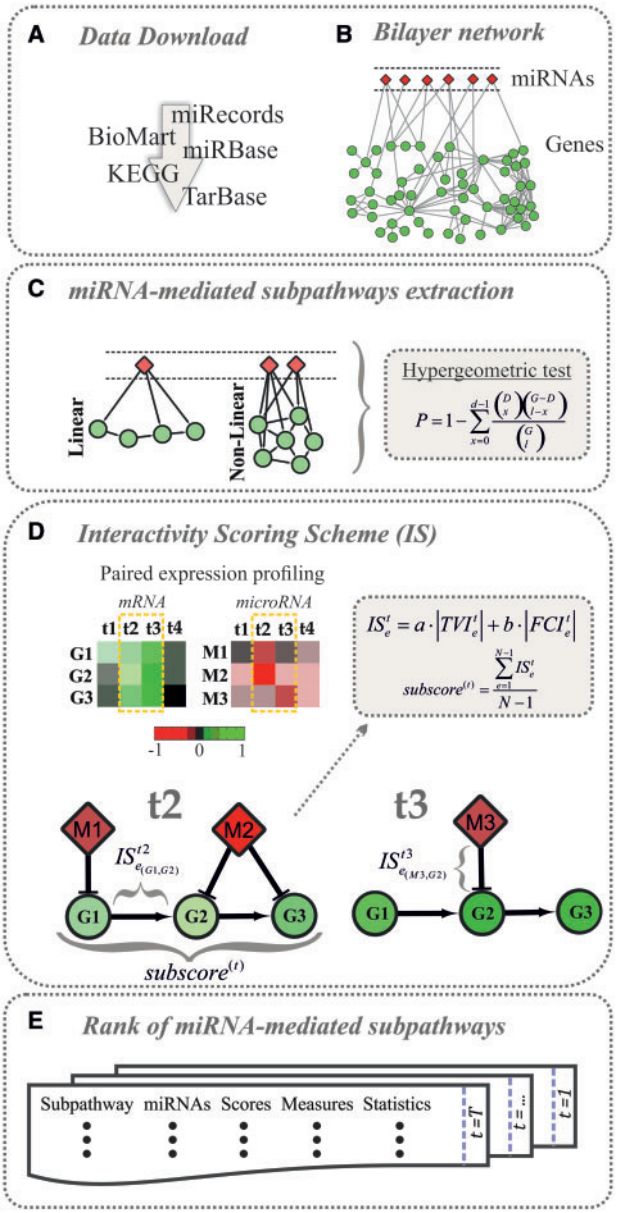

Fig. 1. CHRONOS overview. (**A**) Data retrieval of pathway maps from KEGG database and miRNA–mRNA interactions from miRecords and TarBase. (**B**) Converting pathway interactions and miRNA target interactions into a bilayer gene–gene/miRNA-gene network. (**C**) Extraction of linear and non-linear subpathways from each pathway; candidate miRNA regulators for each subpathway are defined based on cumulative hypergeometric distribution. (**D**) Based on user's data input, i.e. paired time series mRNA and miRNA expression data, the miRNA-mediated subpathways are evaluated based on two scoring schemes which manage to encapsulate the fold changes in expression and the temporal aspects at the level of interactions as well as the flow of information imposed by the pathway topology. (**E**) The top-scoring subpathways can be further examined for enrichment in any user-defined set of 'interesting genes' based on cumulative hypergeometric distribution. Information regarding structural and functional aspects of subpathways is provided via three subpathway-adapted graph-based measures: the subpathway degree (*subDEG*), the betweenness centrality (*subBC*) and the *subPathness*

2 Materials and methods

2.1 CHRONOS framework

2.1.1 User input

Time series expression data from paired mRNA/miRNA experiments are required as input; of note, CHRONOS is applicable in case only the mRNA expression data are available. Twelve common gene label systems are supported, such as EntrezGene ID, UniGene ID, Refseq mRNA ID, HGNC ID, Ensembl Gene ID, based on the web-based data mining tool BioMart (Durinck *et al.*, 2005). Also,

Downloaded from http://bioinformatics.oxfordjournals.org/ at University of California, Los Angeles on August 30, 2016

mature miRNA Ids as reference are obtained by miRBase database (http://www.mirbase.org/) (Supplementary text 1). We denote that the terms 'gene', 'mRNA' and their encoded 'protein' are used interchangeably in this article.

2.1.2 Data retrieval

All available organism-specific pathway maps, both metabolic and non-metabolic, are downloaded from KEGG pathway database (Kanehisa *et al.*, 2002). miRNA–mRNA interactions (micronome) are downloaded from miRecords (Xiao *et al.*, 2009) which integrates the predicted targets of eleven established miRNA target prediction tools (Supplementary text 2). Validated interactions are downloaded from miRecords and DIANA-TarBase (Vlachos *et al.*, 2015).

2.1.3 Converting pathways to gene–gene networks

All KEGG metabolic and non-metabolic pathways are converted to directed gene–gene networks based on information in KGML files (Supplementary text 3 and Fig. S1), without compromising the structural topology of the pathway. In parallel, the biological assumptions set prior to conversion can be adapted according to the study under investigation; similar tools lack such flexibility.

KEGG pathway entry elements representing genes (type = 'gene') may correspond to single or multiple gene products containing most likely gene families or genes with similar biochemical function. Similarly, entries named 'group' usually include multiple gene products representing protein complexes. To tackle this, these entries are expanded into separate nodes by rewiring the incoming and outgoing links of the entry (Supplementary Fig. S2) (Sales *et al.*, 2012). Optionally, these entries can be considered (user-defined) as single nodes by unifying the corresponding gene products for downstream analysis (Judeh *et al.*, 2013). The chemical compounds (type = 'compound') are omitted from the graph and the pathway structure is preserved by connecting the nodes in cases a compound (or hash of compounds) acts as a bridge in-between (Supplementary Fig. S3) (Li *et al.*, 2013).

Another major issue is the categorization of the multiple KEGG supported relations among entries with regard to their biological interpretation. We define three kinds of relations; '*activation*' is translated as enhancement of gene regulation, '*inhibition*' as suppression and '*unknown*' with ambiguous definition. Of note, the user can define a fourth kind of relation named '*no interaction*' so as to exclude specific relation types while focus on those of interest. In addition, CHRONOS allows the user to define the single/multiple relation types and directionality that characterizes an interaction (Supplementary Table S1). Also, with regard to metabolic pathways, gene–gene relations are constructed either by accounting all enzyme–enzyme (type = 'ECrel') relations (Judeh *et al.*, 2013) or/and by extracting all metabolites participating in reactions as substrate or product (Li *et al.*, 2013) (Supplementary text 3 and Fig. S4). An indicative example about pathways conversion to gene–gene network is provided in Supplementary Figure S5.

2.1.4 miRNA-mediated subpathway extraction

Linear and non-linear subpathways are extracted from the gene–gene pathway network. The former are comprised of genes in linear cascades with an in-house computationally fast approach (Supplementary text 4), while the latter contain highly connected gene communities. Notably, our extraction process is highly advantageous over other similar methods in terms of capturing into subpathways the maximum number of nodes present in the KEGG pathway maps (called hereafter 'node coverage') (Fig. 2A and Supplementary Fig. S21).

As linear subpathways, we consider all ordered linear sequences of gene interactions from biologically meaningful start-nodes to end-nodes. In this case, the start-nodes have no incoming interactions and the end-nodes have no outgoing interactions. For a given gene–gene pathway network, the subpathways are obtained by searching all possible paths between start-nodes and end-nodes (Supplementary text 4 and Figs. S6–S8).

Regarding non-linear subpathways, the biological assumption is that the functional similarity between two genes increases as their distance in pathways decreases (Guo *et al.*, 2006). Under this notion, we aim to find subpathways in which all genes have a neighbor with a common functional role. For this, we detect non-linear subpathways at each pathway map by adopting the *k*-clique concept used in social network analysis (Li *et al.*, 2009). A *k*-clique is a subgraph where the distance between any two nodes is no greater than *k*. In our default settings, every 2-clique is a non-linear subpathway (Supplementary text 5 and Figs. S9–S11).

The micronome layer above each subpathway is constructed based on miRecords and TarBase by searching for miRNAs targeting the respective gene members. A miRNA is considered as a potential regulator of a subpathway if its targets are significantly enriched in the relevant subpathway members. Enrichment analysis is performed based on cumulative hypergeometric distribution:

$$P = 1 - \sum_{x=0}^{d} \frac{\binom{D}{x}\binom{G-D}{l-x}}{\binom{G}{l}}, \tag{1}$$

where we suppose that the user input list has G genes, and l genes are included in the subpathway. D is the number of the miRNA target genes and d those included in the subpathway (Li *et al.*, 2012b). miRNAs with $P < 0.05$ are regarded as candidate regulators. The final miRNA regulators of a subpathway are defined later during scoring process (see later Section 2.1.6).

2.1.5 Interactivity scoring schemes

The core component of CHRONOS is two scoring schemes which manage, first on interaction and later on pathway level, to encapsulate the fold changes in expression and the temporal aspects at the level of interactions as well as the flow of information imposed by the pathway topology.

The change in the expression level of a gene is adequately controlled by measuring its fold change activity (relative to a control condition). Under the perspective of pathway networks, it is crucial to examine fold change in expression at the level of edges when referring to interacting genes. To accomplish this, CHRONOS uses the Fold Change Interactivity (*FCI*) score [Equation (2)] based on two multivariate logistic functions, which capture pairs of nodes with high absolute fold change values and with high positive or negative correlation (Kim *et al.*, 2011). The fold activity score of the corresponding interaction e at time t for two connected nodes i,j is:

$$FCI_e^t = sgn(f_i^t, f_j^t)\left(\left(1 + c\sum_{k=i,j\in e} e^{(-K(|f_k^t|-T))}\right)^{-1} - \left(1 + c\sum_{k=i,j\in e} e^{(-K(-|f_k^t|-T))}\right)^{-1}\right), \tag{2}$$

where f_i^t and f_j^t are the $\log_2$-fold change values of nodes i and j, respectively. C, K denote the parameters controlling the shape of the multivariate logistic distribution, and T is a shifting parameter. An indicative example of *FCI* score is provided in Supplementary Figure S12. Over the course of the activation of a pathway there exist

Downloaded from http://bioinformatics.oxfordjournals.org/ at University of California, Los Angeles on August 30, 2016

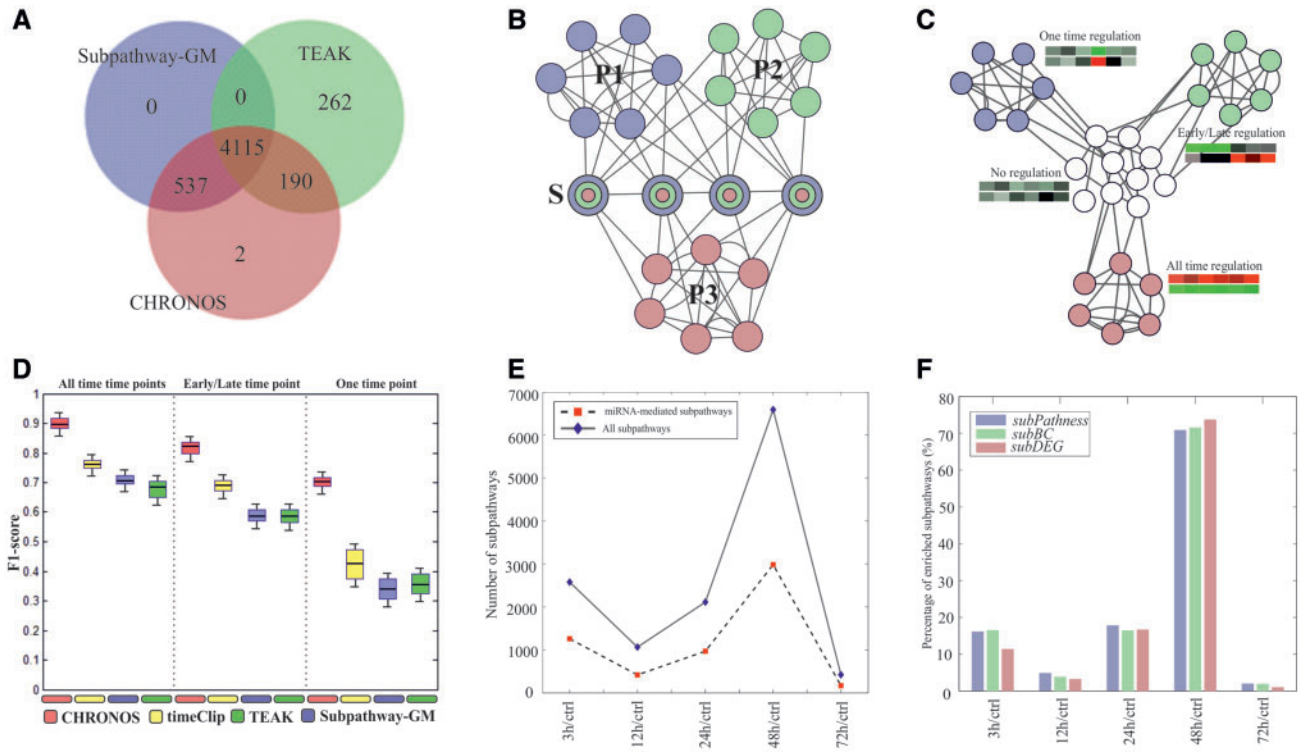

Fig. 2. (**A**) Venn diagrams showing the node coverage of TEAK (4567), Subpathway-GM (4652) and CHRONOS (4844) during subpathway extraction process of human (hsa) KEGG pathway maps. (**B**) Illustration of *subPathness* functional measure. P1, P2 and P3 are three hypothetical pathways with each node denoting a gene and *S* is a subpathway with four nodes all shared among the three pathways. *S* is considered as a subpathway with high *subPathness* value. (**C**) (Simulation analysis) Four expression patterns, based on different types of regulation, were assigned to the nodes of the gene–gene network descending from the human KEGG metabolic pathway maps. (**D**) Boxplot depicting the F1-score results of CHRONOS, TEAK, timeClip, Subpathway-GM and DEAP based on synthetic data. All tools were applied on 100 independent synthetic network models for each duration period of regulation (in all time points, in early/late time points, in one time point). CHRONOS outperforms the other tools in all cases with statistical significance (two-sided Wilcoxon signed rank test, *P-value* < *0.01*). (**E**) (Application to real expression data) The number of enriched non-metabolic subpathways with subscore > 0.4 in each time point and *q*-value < 0.05 is shown. With rhomb, we refer to all enriched subpathways and with square to the subset that is regulated by at least one miRNA. The maximum number of (miRNA-mediated) subpathways was observed at time 48 h. (**F**) (Application to real expression data) Bar plots visualizing per time point the percentage of enriched non-metabolic subpathways that belong to the top scoring 25% (of the total pool of subpathways) with regard to the structural and functional measures (*subDEG, subBC, subPathness*). As shown, the majority of enriched subpathways displayed at 48 h the highest value in all three metrics

multiple time-varying topological 'themes', i.e. time-/condition-specific enriched biologically relevant subgraphs, that determine the output of the topology (Martini *et al.*, 2014; Song *et al.*, 2009). Under this notion, we employed a model rooted in Markov dynamics for analyzing the interactivity dynamics among nodes based on the dynamics of their expression changes (Jethava *et al.*, 2011). To accomplish this, we introduce the Time-Varying Interactivity (*TVI*) score arising from a probabilistic generative model.

Initially, we define the pathway interaction graph with V nodes (genes) and E edges (interactions) as $G = (V, E)$. x_v^t is the fold change expression level for gene $v \in V$ at time t and w_e^t is the interaction strength of $e \in E$ at time t. Weight w can take values in the range $w = \{-2, -1, 0, 1, 2\}$ and represents the interaction strength and higher values correspond to increasing degrees of positive correlation of the interacting genes. The probability of the fold change expression levels for genes i and j at time t , conditioned on the interaction strength w_e^t , is calculated as:

$$P(\vec{X} = \vec{x}^t | \vec{W} = \vec{w}^t) = \frac{1}{Z(\vec{w}^t)} exp\left(\sum_{e=(i,j)\in E} w_e^t x_i^t x_j^t\right) \quad , \quad (3)$$

where $Z(\vec{w}^t)$ is the normalization constant. It is assumed that the weights evolve according to the Markov property, thus the probability of the transition from state $\vec{w}^{t-1}$ at time $t-1$ to state $\vec{w}^t$ at time t is:

$$Q\left(\vec{w}^{(t-1)}, \vec{w}^t\right) = \frac{1}{Z(\vec{w}^t)} exp\left(\sum_{e=(i,j)\in E} w_e^t x_i^t x_j^t\right) \; P(\vec{X}^{t-1} = \vec{x}^{t-1} | \vec{W}^{t-1} = \vec{w}^{t-1}) \quad (4)$$

From the earlier equations, we obtain the TVI_e^t score for each interaction $e \in E$ at each time t which is $TVI_e^t = \max_w Q(\vec{w}^{t-1}, \vec{w}^t)$ (Supplementary text 6). Interactions with $TVI_e^t = \{-2, 0, 2\}$ indicate that the fold changes of the corresponding genes of interaction e at time $t-1$ and t, are strongly negatively correlated, uncorrelated and strongly positively correlated, respectively. Similarly, interactions with $TVI_e^t = \{-1, 1\}$ indicate weak negative and weak positive correlation, respectively.

The two aforementioned scoring schemes are unified into the *Interactivity Score* (*IS*) after accounting for the KEGG pathway interaction trends. More specifically, we defined two consecutive criteria before unifying FCI and TVI scores. First, we check if the two scores agree in terms of sign. Second, if the first criterion is fulfilled, we check if the *interaction trend* sign [(+) for 'activation' and (−) for 'inhibition'] agrees with the sign of the scores. In case of agreement, the two scores are added multiplied by two weighting

Downloaded from http://bioinformatics.oxfordjournals.org/ at University of California, Los Angeles on August 30, 2016

factors [e.g. IF $FCI > 0$ AND $TVI > 0$ AND *interaction trend* > 0 THEN *IS* is calculated based on Equation (5)]; in case the first and/or second criterion is not fulfilled the *IS* is set to zero (e.g. IF $FCI < 0$ AND $TVI > 0$ THEN $IS = 0$, IF $FCI < 0$ AND $TVI < 0$ AND *interaction trend* > 0 THEN $IS = 0$). Notably, in the case of 'unknown' interactions, the second criterion is not checked since the interaction trend cannot be unambiguously determined.

For interaction e at time t, *IS* is calculated as:

$$IS_e^t = a \cdot |TVI_e^t| + b \cdot |FCI_e^t| \quad , \tag{5}$$

where a, b are weighting parameters indicating the percentage of contribution of *TVI* and *FCI* score, respectively. With these criteria, we promote, first on interaction and later on pathway level, the interactions among nodes that simultaneously change in terms of expression in each time point and these changes accord well with the constraints imposed by the topology.

miRNA-gene interactions are also examined in similar manner [Equation (5)] as previous to identify miRNAs correlated to their target genes (Ooi *et al.*, 2011) during time. Notably, for these interactions the type of relation is not well-defined since, according to literature evidence, miRNAs mainly exhibit suppression effect but in many cases can also exhibit activation effect (Place *et al.*, 2008). For this reason, we selected to ignore the type of relation to avoid biasing the scoring process in favor of the suppressor type of relation.

2.1.6 Time-varying enriched miRNA-mediated subpathways

Moving forward from interaction to pathway level, subpathways are evaluated based on the *IS* since we search for simultaneously context- and time-specific subpathways, whose expression changes in specific time points accord well with the flow of information as provided by the subpathway structure. Hence, subpathway score (*subscore*) is calculated as the summand of all involved *ISs* divided with the number of subpathway interactions, so that the scores among subpathways are comparable. For a subpathway with N gene members and $N-1$ interactions at time t, *subscore* is:

$$subscore^{(t)} = \frac{\sum_{e=1}^{N-1} IS_e^t}{N-1}, \tag{6}$$

where IS_e^t is the *Interactivity Score* of interaction e at time t. Further, via cumulative hypergeometric distribution the enrichment of a subpathway in any user-defined set of 'interesting genes' (e.g. list of differentially expressed genes (DEGs), list of disease-related genes, etc.) can be examined. In detail, the statistical significance is calculated based on Equation (1) where G is the user-defined gene set, D the number of user-defined set of interesting genes, l all the genes involved in the subpathway and d the interesting genes included in the subpathway. Also, the false positive (FP) discovery rate of P results is reduced by providing false discovery rate (FDR) corrected P-values (q-values). Subpathways with $subscore > 0.4$ in at least one time point and $q\text{-}value < 0.05$ (in case a set of 'interesting genes' is provided) are considered enriched. The *subscore* threshold was determined by setting a FDR of 1% cutoff on the results derived from 100 permutations; in each permutation a small fraction of subpathway interaction scores changed and subpathway scores were recomputed (Allantaz *et al.*, 2012).

Similarly, for the micronome layer, the involved candidate interactions are evaluated based on Equation (5) and the miRNAs whose relevant interactions achieve $IS > 0.4$ are considered as final regulators. This threshold (0.4) was determined by setting a FDR cutoff of 1% on the results derived from 100 permutations; in each permutation a small fraction of miRNA–mRNA interaction scores within the boundaries of a subpathway changed and average scores were recomputed.

2.1.7 Subpathway structural and functional measures

CHRONOS accounts for the interpathway dependencies and as such provides information about the position of the subpathways in the complete pathway atlas. This zoom-out overview of subpathways can assist significantly when the goal is the monitoring of the signal propagation. In our tool, subpathways are considered as single nodes, called meta-nodes, upon which known established topological metrics are adapted. In detail, we calculate the subpathway degree (*subDEG*) and betweenness centrality (*subBC*) so as to capture the local and global structural aspects of each subpathway, respectively (Supplementary Fig. S13).

Moving forward, it is crucial to examine the functional aspects of pathway topologies. When dealing with the organism map, it is preferable to view pathways as modules defined not only through structural measures but also through biological criteria that discriminate the mechanisms among pathways. Under this notion, we introduce *subPathness* (Fig. 2B) which measures the degree to which a subpathway serves as bridge (Dimitrakopoulou *et al.*, 2014; Kovács *et al.*, 2010) among the different pathways of an organism (Supplementary text 7).

2.1.8 CHRONOS output

CHRONOS is a computationally fast approach (Supplementary Fig. S15) and outputs a ranked summary of the final results. The (miRNA-mediated) subpathways are exported in serial alignment based on their *subscore* [Equation (5)] and FDR corrected P-values (q-values). The (miRNA-mediated) subpathways per time point are exported separately for each type of pathways (metabolic and non-metabolic) and for each type of extraction (linear and non-linear) accompanied by the graph-based metric values (*subDEG, subBC, subPathness*). KEGG pathway map links and circular plots (Gu *et al.*, 2014) are also available for each subpathway per each time point (Supplementary Figs. S16–S19).

2.2 Generating synthetic networks and synthetic expression data

Synthetic networks and synthetic expression data were generated to run a simulation performance analysis of CHRONOS and other similar tools (Section 3.1). Random graphs were generated as biological pathway networks by using two closely related model graphs, the Erdős–Rényi and Edgar Gilbert. These graphs were built with $N = 500$ nodes with rewiring probability $P = 0.9$.

Synthetic gene expression values per time point were generated either with time-dependent up/down regulation or with no change over time relevant to a control state ($\log_2$-fold change difference 1, −1 and 0, respectively). In addition, the up/down regulated expression profiles were categorized according to the duration of their regulation. These categories correspond to the up/down regulated gene expressions in (i) all time points, in (ii) half time points (early or late) and in (iii) one time point. For each category, a synthetic network model was constructed with 60% of nodes set as up/down regulated and the rest as non-regulated. Nodes were randomly drawn from a normal distribution and Gaussian noise with zero mean and standard deviation (SD = 0.2) was added.

2.3 mRNA/miRNA time series expression data of IFN-g-treated melanoma cells

We applied CHRONOS on two real time series mRNA and miRNA datasets after stimulation of human A375 melanoma cell line with

Downloaded from http://bioinformatics.oxfordjournals.org/ at University of California, Los Angeles on August 30, 2016

IFN-g (Nazarov *et al.*, 2013). The datasets contain samples either left untreated (control) or stimulated with human IFN-g at specific time points. Datasets are available at ArrayExpress (www.ebi.ac.uk/arrayexpress) under accession numbers E-MEXP-3544 and E-MEXP-3720, respectively. The datasets were normalized with the Robust Multi-array Average (RMA) method in MATLAB environment and probes referring to the same Entrez Id were averaged. Also, 6,777 Entrez Ids were mapped onto KEGG human pathways (accessed May 2015) and 4048 onto our extracted subpathways. For downstream analysis, only the matching samples among mRNA and miRNA were kept, i.e. 0, 3, 12, 24, 48 and 72 h (the Janus kinase inhibitor I second control samples were excluded). Also, limma R package (Smyth, 2005) was used to identify significantly (with FDR < 0.05) DEGs and miRNAs (DEMs) by comparing in each time point the expression of IFN-g–treated samples versus untreated controls; the DEMs common with those identified in (Nazarov *et al.*, 2013), were used as input to CHRONOS. Of note, CHRONOS was run including all genes mapped to KEGG pathways and to alleviate a more comprehensive evaluation of our results relative to the results of the original study (Nazarov *et al.*, 2013), the CHRONOS output subpathways were examined for enrichment in DEGs (used as 'interesting genes' for the calculation of *q*-values) so as to obtain those subpathways that are most relevant.

3 Results and discussion

3.1 Application of CHRONOS to synthetic dataset

The performance of CHRONOS in capturing the time-varying 'themes' underlying the pathway topology was evaluated relative to other similar methods (TEAK, timeClip, Subpathway-GM and DEAP) based on a synthetic dataset. Figure 2C illustrates four expression patterns assigned to the nodes of a gene–gene network descending from the human KEGG metabolic pathway maps (non-metabolic pathways and miRNAs were not included since their analysis is not applicable in all methods). All tools were applied on 100 independent synthetic network models and the 100 top ranking subpathways were considered as reliable for downstream analysis. To overcome the different subpathway topologies identified by each method, we assessed the performance (precision, recall, F1-score) based solely on the unique nodes included in the 100 top ranking subpathways. The up-/downregulated and non-regulated genes included in final results were defined as true positives and FPs, respectively. As false negatives were defined the up-/downregulated genes belonging to the corresponding perturbed pathways which were not included in the final subpathways. The boxplots of Figure 2D (see also Supplementary Fig. S20) illustrate that CHRONOS outperformed TEAK, timeClip, Subpathway-GM and DEAP in all types of regulation with statistical significance (two-sided Wilcoxon signed rank test, $P\text{-value} < 0.01$) with the highest efficiency achieved in the one time point regulation case. These observations point out that CHRONOS succeeds into realizing the main goal of a time series experiment, that is to zoom into the time-specific phenomena of a pathway (Song *et al.*, 2009) that are not attainable using current subpathway-based approaches.

An important feature in subpathway analysis is the ability to exploit the maximum number of nodes present in pathway maps ('node coverage'). We chose to test CHRONOS subpathway extraction efficiency only against TEAK and Subpathway-GM since these two methods provide both an approach for converting KEGG KGML pathway files into biologically concise gene–gene networks and an approach for subpathway extraction. The node coverage in the extracted subpathways was examined in both real and synthetic networks based on the human KEGG pathway maps and Erdős-Renyi and Edgar Gilbert model graphs, respectively. CHRONOS outperformed the other two approaches both in real (Fig. 2A) as well as in synthetic networks (Supplementary Fig. S21) with statistical significance (two-sided Wilcoxon signed rank test, $P\text{-value} < 0.01$).

3.2 Application of CHRONOS to real data

IFN-g production is highly related with tumor progression and prognosis and IFN-g is the most important endogenous mediator of immunity and inflammation and plays a key role in macrophage activation, inflammation, host defense against intracellular pathogens and T helper 1 cell responses (Hu and Ivashkiv, 2009). IFN-g, apart from activating the Jak-STAT pathway, has been shown to cross-regulate multiple signaling pathways driven by endogenous or exogenous factors. Moreover, several miRNAs were found to be dynamically regulated after stimulation with IFN-g (Schmitt *et al.*, 2012). In general, our time-varying analysis showed significant reranking of subpathways during time evolution and in Supplementary files S2–S5 we provide the full list of the enriched (miRNA-mediated)-subpathways ($subscore > 0.4$ and $q\text{-value} < 0.05$) per time point. In Figure 2E, the timeline of (miRNA-mediated) non-metabolic subpathway frequency distribution points out 48 h as the peak of activity (see also Supplementary Fig. S24); this observation agrees with the findings of the original study where no significant alteration of the mRNA levels was observed after 48 h although this analysis differs conceptually from our approach; it is focused mainly on differentially expressed molecules and evaluates them through functional, pathway, upstream regulator analysis and regulatory loop identification (Nazarov *et al.*, 2013).

Moreover, when evaluating the subpathways in terms of the structural and functional measures (*subDEG*, *subBC*, *subPathness*), we observed that at 48 h (Fig. 2F) the majority of enriched subpathways acquired high scores indicating so that at this time point several hub- and bridges-subpathways take over that in turn propagate and scatter quickly the signal to multiple signaling pathways (see also Supplementary Figs. S14, S22 and S25). Regarding Jak-STAT signaling, we observed that the related subpathways (in all time points) represent the 10.5% of all enriched non-metabolic subpathways with the majority being active at 3 and 12 h, in accordance with the observations of the original study, with most subpathways regulated at 3 h by miR-105 (other regulators: miR-18a, miR-19b, miR-149, miR-29a). In parallel, we run TEAK, timeClip, Subpathway-GM and DEAP on the gene expression data and compared Jak-STAT and other indicative pathways whose timeline is discussed in (Nazarov *et al.*, 2013) and verified that CHRONOS is efficient not only in capturing more information about the active pathways but also in specifying the exact time period of activity (supplementary Table S3).

To illustrate the performance of CHRONOS into capturing time-specific enriched topological 'themes', which in other case would be neglected or implicated in the whole time period by other current similar methods, we selected few indicative non-metabolic exemplar pathways that underwent severe transformations in response to IFN-g during time course. Toll-like receptor signaling related subpathways were not active at 72 h; it has been shown that IFN-g enhances TLR-activated signaling, promotes NF-kB activation and induces transcription factors necessary for the expression of TLR-responsive genes (Hu and Ivashkiv, 2009). Also, most of the TNF-signaling related subpathways were enriched at 24 h and no active subpathway was found at 72 h; evidence has showed that IFN-g and TNF-α synergistically induce mesenchymal stem cell impairment

Downloaded from http://bioinformatics.oxfordjournals.org/ at University of California, Los Angeles on August 30, 2016

and tumorigenesis via NFκB signaling (Wang *et al.*, 2013). RIG-I and Ras signaling related subpathways vanished at 24 h; induction of RIG-I in response to IFN-g was found in epithelial cell lines (Imaizumi *et al.*, 2005) while the signaling initiated by activated *k*-ras affects the IFN/STAT signaling pathway and controls the responsiveness of cancer cells to interferons (Klampfer *et al.*, 2003). Finally, TGF-beta signaling related subpathways appeared only at 48 h; evidence has shown that IFN-g induces expression of *SMAD7* which in turn inhibits the TGF-b-induced activation of the activating Smad3 and of TGF-b-responsive genes (Hu and Ivashkiv, 2009). Similarly, we observed significant temporal transformations in the micronome layer with many miRNAs appearing or disappearing as regulators of the active subpathways. Based on the circular plots of Figure 3, we argue that the miRNA effect is present from the early time point (3 h) compared with the observations of the initial study where miRNAs are mostly involved in later time points (Nazarov *et al.*, 2013); however, it is evident that the regulatory effect is concentrated on specific pathways at the early time points whereas at 48 h is scattered upon multiple pathways—probably through the hub-/bridge-subpathways (for the remaining time points see also Supplementary Fig. S23).

CHRONOS has the explanatory power to offer focused serial temporal snapshots of the mechanisms perturbed under specific conditions with resolution up to every single time point and offers a holistic view of the dynamics of molecules participating in multiple subpathways and in many cases multiple pathways. We comment on the signaling neighborhood around indicative interesting genes involved in biological functions highly related to the mechanisms investigated in this experiment (cell adhesion, apoptosis, immune response and cell cycle) which were described by transcriptional regulatory network motifs consisting of feed forward loops (FFLs) involving transcription factors, miRNAs and joint targets (Nazarov *et al.*, 2013). Intercellular adhesion Molecule 1 (*ICAM1*) was found in subpathways belonging to 'leukocyte transendothelial migration' and 'cell adhesion molecules' pathways with the majority being active at 12 and 24 h and few of them under miRNA control at 12 and 48 h (time point 24 h was observed in the initial study); IFN-g has been shown to facilitate lymphocyte adhesion to endothelial junctional regions enabling so the transendothelial migration (Jaczewska *et al.*, 2014). *RelA/NF-kB* complex acts as transcription factor and it was found in subpathways descending from 16 distinct pathway maps with most activity observed at 3, 12 and 48 h and miRNA regulation at 24 h (time point 24 h was observed in the initial study). *TP53* (tumor protein 53) acts as transcription factor and it was found in the subpathways descending from eight pathway maps with most activity and miRNA regulation observed at 48 h (time point 48 h was also observed in the initial study). One indicative exemplar pathway involving this gene is 'MAPK signaling'; IFN-g controls tissue damage by deflating MAPK pathways that enhance the transcriptional activity of AP-1 proteins essential for the expression of matrix metalloproteinases whose inhibition limits tissue destruction (Hu and Ivashkiv, 2009). *BCL2L1* (Bcl-2-like protein 1) was found in the subpathways descending from four pathway maps (mostly in Jak-STAT signaling) with most activity and miRNA regulation observed at 3 and 12 h (time points 48 and 72 h were observed in the initial study). *CCND1* (cyclin D1) was found in the subpathways descending from 13 pathway maps, mainly in prolactin, Jak-STAT and focal adhesion signaling, with most activity and miRNA regulation observed at 48 h (time point 72 h was observed in the initial study). The discrepancies between methods are explained by the limitation of KEGG topology in terms of the number of pathway annotated genes and by the fact that only differentially expressed molecules are used in (Nazarov *et al.*, 2013) and the inferred transcriptional regulatory network motifs are related to signaling pathways but the latter refer mostly to the targets of FFLs. As such, our method is complementary and can serve as a push forward step for unraveling more about the dynamics of 'crossroad' molecules involved in many cases in multiple downstream signal propagation

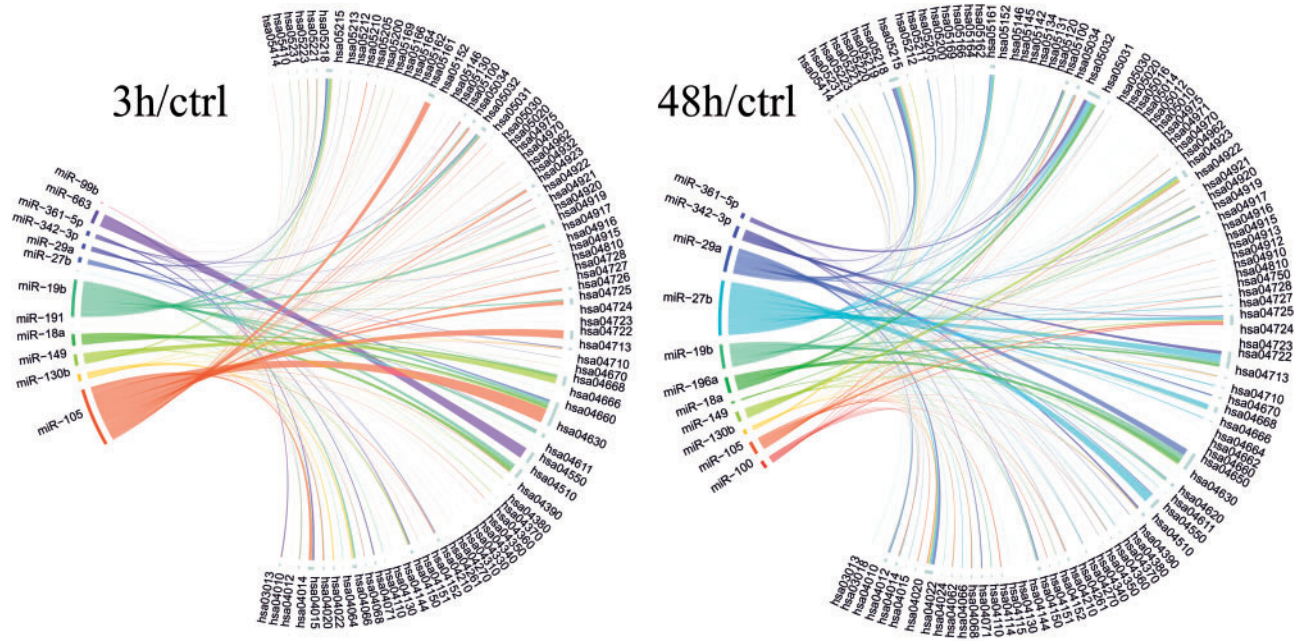

Fig. 3. Circular plots visualizing the rewiring of miRNA effect upon non-metabolic subpathways in two indicative time points (3 and 48 h). The width of categories is analogous to the number of human subpathways (denoted with KEGG pathway Ids) targeted by a miRNA at the specific time point. The miRNA-mediated subpathways (linear and non-linear) had *subscore* > *0.4* with *q-value* < *0.05.* For better readability non–miRNA-mediated subpathways are not included. The circular plots for the remaining time points are provided in Supplementary Figure S23

Downloaded from http://bioinformatics.oxfordjournals.org/ at University of California, Los Angeles on August 30, 2016

routes not only under the umbrella of gross signaling pathways—with unclear in many cases definitions and boundaries—but on the more finite context-rich scale of subpathways.

4 Conclusion

CHRONOS is a computationally fast and efficient approach for scanning the complete pathway topology to track the time-/context-specific biologically enriched subpathways along with their time-/context-specific miRNA regulators. The method and the supporting R package are flexible by allowing the user to intervene and adapt all discrete phases to the needs of the study under investigation. CHRONOS can assist significantly in complex disease analysis by enabling the experimentalists to acquire a more realistic time-varying dynamical view of the involved perturbed mechanisms.

Funding

This research has been cofinanced by the European Union (European Social Fund ± ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Program: Thalis. Investing in knowledge society through the European Social Fund (grant number: MIS: 377001).

Conflict of Interest: none declared.

References

Allantaz,F. *et al.* (2012) Expression profiling of human immune cell subsets identifies miRNA-mRNA regulatory relationships correlated with cell type specific expression. *PloS One*, **7**, e29979.

Barabási,A.L. *et al.* (2011) Network medicine: a network-based approach to human disease. *Nat. Rev. Genet.*, **12**, 56–68.

Calura,E. *et al.* (2014) Wiring miRNAs to pathways: a topological approach to integrate miRNA and mRNA expression profiles. *Nucleic Acids Res.*, **42**, e96–e96.

Chen,X. *et al.* (2011) A sub-pathway-based approach for identifying drug response principal network. *Bioinformatics*, **27**, 649–654.

Dimitrakopoulou,K. *et al.* (2014) Tamoxifen integromics and personalized medicine: dynamic modular transformations underpinning response to tamoxifen in breast cancer treatment. *OMICS*, **18**, 15–33.

Dimitrakopoulou,K. *et al.* (2015) Integromics network meta-analysis on cardiac aging offers robust multi-layer modular signatures and reveals micronome synergism. *BMC Genomics*, **16**, 147.

Durinck,S. *et al.* (2005) BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics*, **21**, 3439–3440.

Gu,Z. *et al.* (2014) circlize implements and enhances circular visualization in R. *Bioinformatics*, **30**, 2811–2812.

Guo,X. *et al.* (2006) Assessing semantic similarity measures for the characterization of human regulatory pathways. *Bioinformatics*, **22**, 967–973.

Haynes,W.A. *et al.* (2013) Differential expression analysis for pathways. *PLoS Comput. Biol.*, **9**, e1002967.

Hu,X. and Ivashkiv,L.B. (2009) Cross-regulation of signaling pathways by interferon-γ: implications for immune responses and autoimmune diseases. *Immunity*, **31**, 539–550.

Imaizumi,T. *et al.* (2005) Involvement of retinoic acid-inducible gene-I in the IFN-γ/STAT1 signalling pathway in BEAS-2B cells. *Eur. Respir. J.*, **25**, 1077–1083.

Jacob,L. *et al.* (2012). More power via graph-structured tests for differential expression of gene networks. *Ann. Appl. Stat.*, **6**, 561–600.

Jaczewska,J. *et al.* (2014) TNF-α and IFN-γ promote lymphocyte adhesion to endothelial junctional regions facilitating transendothelial migration. *J. Leukoc. Biol.*, **95**, 265–274.

Jethava,V. *et al.* (2011) Netgem: network embedded temporal generative model for gene expression data. *BMC Bioinformatics*, **12**, 327.

Judeh,T. *et al.* (2013) TEAK: topology enrichment analysis framework for detecting activated biological subpathways. *Nucleic Acids Res.*, **41**, 1425–37.

Kanehisa,M. *et al.* (2002) The KEGG databases at GenomeNet. *Nucleic Acids Res.*, **30**, 42–46.

Khatri,P. *et al.* (2012) Ten years of pathway analysis: current approaches and outstanding challenges. *PLoS Comput. Biol.*, **8**, e1002375.

Kim,Y. *et al.* (2011). Principal network analysis: identification of subnetworks representing major dynamics using gene expression data. *Bioinformatics*, **27**, 391–398.

Klampfer,L. *et al.* (2003) Oncogenic Ki-ras inhibits the expression of interferon-responsive genes through inhibition of STAT1 and STAT2 expression. *J. Biol. Chem.*, **278**, 46278–46287.

Kovács,I.A. *et al.* (2010) Community landscapes: an integrative approach to determine overlapping network module hierarchy, identify key nodes and predict network dynamics. *PloS One*, **5**, e12528.

Li,C. *et al.* (2009) SubpathwayMiner: a software package for flexible identification of pathways. *Nucleic Acids Res.*, **37**, e131.

Li,C. *et al.* (2012) Identifying disease related sub-pathways for analysis of genome-wide association studies. *Gene*, **503**, 101–109.

Li,C. *et al.* (2013) Subpathway-GM: identification of metabolic subpathways via joint power of interesting genes and metabolites and their topologies within pathways. *Nucleic Acids Res.*, **41**, e101.

Li,J. *et al.* (2014) The detection of risk pathways, regulated by miRNAs, via the integration of sample-matched miRNA-mRNA profiles and pathway structure. *J. Biomed. Inform.*, **49**, 187–197.

Li,X. *et al.* (2015). Subpathway analysis based on signaling-pathway impact analysis of signaling pathway. *PloS One*, **10**, e0132813.

Li,X. (2012) Dissection of human MiRNA regulatory influence to subpathway. *Brief. Bioinform.*, **13**, 175–186.

Martini,P. *et al.* (2013) Along signal paths: an empirical gene set approach exploiting pathway topology. *Nucleic Acids Res.*, **41**, e19–e19.

Martini,P. *et al.* (2014) timeClip: pathway analysis for time course data without replicates. *BMC Bioinformatics*, **15**, 1–10.

Nam,S. *et al.* (2014) PATHOME: an algorithm for accurately detecting differentially expressed subpathways. *Oncogene*, **33**, 4941–4951.

Nazarov,P.V. *et al.* (2013) Interplay of microRNAs, transcription factors and target genes: linking dynamic expression changes to function. *Nucleic Acids Res.*, **41**, 2817–2831.

Ooi,C.H. *et al.* (2011) A densely interconnected genome-wide network of microRNAs and oncogenic pathways revealed using gene expression signatures. *PLoS Genetics*, **7**, e1002415.

Place,R.F. *et al.* (2008) MicroRNA-373 induces expression of genes with complementary promoter sequences. *Proc. Natl. Acad. Sci. USA*, **105**, 1608–1613.

Sales,G. *et al.* (2012) graphite-a Bioconductor package to convert pathway topology to gene network. *BMC Bioinformatics*, **13**, 20.

Sebastian-Leon,P. *et al.* (2014) Understanding disease mechanisms with models of signaling pathway activities. *BMC Syst. Biol.*, **8**, 121.

Subramanian,A. *et al.* (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA*, **43**, 15545–15550.

Song,L. *et al.* (2009) Time-varying dynamic Bayesian networks. In *Advances in Neural Information Processing Systems*, pp. 1732–1740.

Smyth,G.K. (2005) Limma: linear models for microarray data. In: *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*, pp. 397–420.

Schmitt,M.J. *et al.* (2012) Interferon-gamma-induced activation of Signal transducer and activator of transcription 1 (STAT1) up-regulates the tumor suppressing microRNA-29 family in melanoma cells. *Cell Commun. Signal.*, **10**, 41.

Downloaded from http://bioinformatics.oxfordjournals.org/ at University of California, Los Angeles on August 30, 2016

Tarca,A.L. *et al*. (2009) A novel signaling pathway impact analysis. *Bioinformatics*, **25**, 75–82.

Vlachos,I.S. *et al*. (2015) DIANA-TarBase v7. 0: indexing more than half a million experimentally supported miRNA: mRNA interactions. *Nucleic Acids Res.*, **43**, D153–D159.

Wang,L. *et al*. (2013) IFN-γ and TNF-α synergistically induce mesenchymal stem cell impairment and tumorigenesis via NFκB signaling. *Stem Cells*, **31**, 1383–1395.

Wu,B. *et al*. (2013) Dissection of miRNA-miRNA interaction in esophageal squamous cell carcinoma. *PloS One*, **8**, e73191.

Xiao,F. *et al*. (2009) miRecords: an integrated resource for microRNA–target interactions. *Nucleic Acids Res.*, **37** (suppl. 1), D105–D110.

Zhang,C. *et al*. (2014) Identification of miRNA-mediated core gene module for glioma patient prediction by integrating high-throughput miRNA, mRNA expression and pathway structure. *PloS One*, **9**, e96908.

Downloaded from http://bioinformatics.oxfordjournals.org/ at University of California, Los Angeles on August 30, 2016