

Inference of Causal Networks from Time-Varying Transcriptome Data via Sparse Coding

Kai Zhang, Ju Han, Torsten Groesser, Gerald Fontenay, Bahram Parvin*

Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California, United States of America

Abstract

Temporal analysis of genome-wide data can provide insights into the underlying mechanism of the biological processes in two ways. First, grouping the temporal data provides a richer, more robust representation of the underlying processes that are co-regulated. The net result is a significant dimensional reduction of the genome-wide array data into a smaller set of vocabularies for bioinformatics analysis. Second, the computed set of time-course vocabularies can be interrogated for a potential causal network that can shed light on the underlying interactions. The method is coupled with an experiment for investigating responses to high doses of ionizing radiation with and without a small priming dose. From a computational perspective, inference of a causal network can rapidly become computationally intractable with the increasing number of variables. Additionally, from a bioinformatics perspective, larger networks always hinder interpretation. Therefore, our method focuses on inferring the simplest network that is computationally tractable and interpretable. The method first reduces the number of temporal variables through consensus clustering to reveal a small set of temporal templates. It then enforces simplicity in the network configuration through the sparsity constraint, which is further regularized by requiring continuity between consecutive time points. We present intermediate results for each computational step, and apply our method to a time-course transcriptome dataset for a cell line receiving a challenge dose of ionizing radiation with and without a prior priming dose. Our analyses indicate that (i) the priming dose increases the diversity of the computed templates (e.g., diversity of transcriptome signatures); thus, increasing the network complexity; (ii) as a result of the priming dose, there are a number of unique templates with delayed and oscillatory profiles; and (iii) radiation-induced stress responses are enriched through pathway and subnetwork studies.

Citation: Zhang K, Han J, Groesser T, Fontenay G, Parvin B (2012) Inference of Causal Networks from Time-Varying Transcriptome Data via Sparse Coding. *PLoS ONE* 7(8): e42306. doi:10.1371/journal.pone.0042306

Editor: Huaijun Zhou, University of California, Davis, United States of America

Received: December 6, 2011; **Accepted:** July 6, 2012; **Published:** August 20, 2012

Copyright: © 2012 Zhang et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by the Director, Office of Science, Office of Biological and Environmental Research, Radiochemistry and Imaging Instrumentation, Low Dose Radiation program with the U.S. Department of Energy under Contract Nos. DE-AC02-05CH11231, and a grant from the National Cancer Institute (Grant R01CA140663). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: B_Parvin@lbl.gov

Introduction

Biological systems often operate as networks of interacting components that are highly regulated [1]. These networks enable a cell to integrate external stimuli and biochemical reactions that can potentially lead to the activation of transcription factors (TFs). In turn, these TFs recognize a specific regulatory region for manipulating gene expressions. Characterization of network biology has been further advanced through mathematical analysis of genome-wide array data for hypothesis generation. In the context of mathematical modeling, logical (e.g., Boolean [2], stochastic [3,4], petri net [5]) and continuous (e.g., ordinary differential equations [6], flux balance analysis [7]) techniques have been proposed. Recent reviews of these techniques can be found in [8,9]. Each of these techniques has its own pros and cons with distinct application domains. In this paper, we introduce a method to hypothesize a causal network that is derived from the analysis of the time-varying genome-wide array data, where causality is interpreted in a weak sense to show a potential relationship between groups of transcripts at two consecutive time-points. Given the complexities of a biological network and inherently high dimensionality of an array-based data coupled with a low sample size, we aim at deriving the simplest network for

hypothesizing causality. We suggest that causality can be inferred through either perturbation studies or time-course data. The latter has the potential to enrich the genome-wide array data by grouping time-course profiles; thereby, leading to a lower dimensional representation. Subsequently, such a low dimensional representation can then be modeled as a layered signaling network, where each output at a given time layer is expressed as a function of inputs from a previous time point. The net result is a causal network (e.g., a wiring diagram) that fits the time-varying data according to a cost function. The concept of “simplicity” is enforced by requiring that (i) not all input variables from a given time point contribute to an output at the next time point, (ii) an output is a linear combination of input variables, and (iii) there is a notion of continuity in the signaling network. Collectively, these constraints lead to a highly regularized sparse linear model. The method is validated against different configurations of synthetic data and then applied to an experimental dataset to examine the effects of a higher dose of ionizing radiation with and without a priming low dose of radiation, which is known as an adaptive response. The proposed computational protocol is applied to a unique experiment in radiation biology, where a cell line has been treated in one of two different ways: (a) with a challenge dose of 200 cGy or (b) with a priming dose of 10 cGy followed by the

challenge dose of 200 cGy. The latter is referred to as an “adaptive response” [10–12], since adaptation is attributed to reduced damages as a result of adding the priming dose. Consequently, it is our goal to characterize and differentiate induced perturbations in terms of the (a) shape and number of computed templates, (b) architecture of the wiring diagrams, and (c) biological interpretation through enrichment analysis.

Results

We provide an analysis of clustered temporal profiles, followed by an interpretation of the causal networks.

Analysis of temporal profiles

The initial sets of gene expression data for the treatment groups with and without the priming dose are reduced to 682 and 527 genes, respectively, in accordance with the policy outlined in the Method section. These genes have highly variable expression values across different time points. Consensus clustering of filtered transcript data indicates there are 8 clusters that correspond to samples receiving the priming dose (e.g., adaptive response) versus the 5 clusters that do not receive the priming dose (e.g., challenge), as shown in Figures 1A and 1B. Each cluster from each of the treatment groups corresponds to a unique temporal profile as shown in Figures 2C and 2D. The details for selecting the number of clusters are summarized in Figures S1, S2, S3, S4, and Text S1. We label each template, in its own cluster, as $T^i_{adaptive}$ and $T^i_{challenge}$, where i represents the template number (e.g., 1–8 for those with the priming dose), *adaptive* represents those samples receiving the priming dose prior to the challenge dose, and *challenge* representing those that only receive the challenge dose.

A comparison of Figures 1C and 1D indicate both similarities and dissimilarities in the radiation response with and without the priming dose. These differences can then be used to probe for bioinformatics analysis. We used *Pathway Studio* to analyze the computed clusters through pathway and subnetwork enrichment analysis for identical and differential responses with the results shown in Tables 1 and 2.

Similar profiles such as: (i) $T^7_{adaptive}$ and $T^2_{challenge}$ appear to have a down-regulated profile from 1 to 8 hours. Here, the pathway and subnetwork enrichment analysis reveals a significant amount of overlap through the cell cycle regulation and inflammatory responses. (ii) $T^8_{adaptive}$ and $T^1_{challenge}$ are initially flat and then down-regulated with similar pathway (e.g., cell cycle and Hedgehog) and subnetwork (e.g., E2F, RB1, TP53, PDGF) enrichment analyses. The role of E2F and RB1 in regulating cell cycle, cell fate, DNA damage repair and apoptosis has been well established [13,14]. It has also been shown that the cellular response to DNA damage utilizes the RB/E2F cell cycle pathway [15]. Their time profiles are shown in Figure 2, where RB1 is significantly down-regulated in samples receiving the priming dose. The E2F family shows similar profile for its transcripts; a representative, E2F8, is also shown here. (iii) Conversely, $T^1_{adaptive}$ and $T^5_{challenge}$ share the same temporal signature, but in the opposite direction of the templates in (ii). Specifically, they are initially flat, but then upregulated. These templates share a significant amount of overlap in terms of pathway (e.g., apoptosis) and subnetwork (e.g., TP53, SP1, IL1B, and EGFR) enrichment analyses. (iv) Although $T^4_{adaptive}$ and $T^3_{challenge}$ also has similar temporal profiles, the ensemble of genes representing $T^3_{challenge}$ are poorly enriched, which is largely due to the lack of related annotation data. On the other hand, $T^4_{adaptive}$ is highly enriched for

FOXO3A and NF-kB. (v) Lastly, $T^2_{adaptive}$ and $T^4_{challenge}$ are initially upregulated and then plateau after the 2 hour time point with the enrichment of CD43 (for regulating immune function) and TP53 pathways in adaptive and challenged response treatment groups, respectively. These analyses suggest that every template in the challenge group has a profile similar to those in the treatment group with the priming dose, and in one case, pathway enrichment has been limited only to the immune function activation.

Dissimilar profiles are $T^5_{adaptive}$ and $T^6_{adaptive}$ (upregulated and downregulated, respectively, at the 4 hour time point) as well as the oscillatory profile of $T^3_{adaptive}$. The first two templates indicate a delayed response as a result of the priming dose, and enrichment analysis indicates a number of components that overlap with the existing templates. In $T^5_{adaptive}$ and $T^6_{adaptive}$, STAT signaling is an enriched dominant pathway that integrates cellular stresses such as ultraviolet radiation, inflammation, and infection [16], as reflected in the subnetwork enrichment analysis and reported earlier as a result of low dose radiation on inbred strains of mice [17]. Further examination indicates that $T^6_{adaptive}$ shares some of the components of $T^1_{challenge}$, but with a delayed response, as shown in the supplementary section. These shared transcripts, for example, HIST1H4B, are nuclear-bound and are involved in chromatin remodeling. However, the subnetwork enrichment analysis of the oscillatory profile of $T^3_{adaptive}$ indicates enrichment of NF-kB, TP53, and TGFBI, shown in Figure 3. Figure 4 shows the temporal profiles for a subset of transcripts presented in Figure 3, which are also known to be sensitive to ionizing radiation. Although the oscillatory signature is more dominant in the adaptive response where it forms a distinct cluster (e.g., template), some of the transcripts in the challenge groups also show similar profiles. Finally, TGFBI has been identified as a common regulator of radiation response in Figure 4, though its transcriptome profile is not significantly altered as a result of ionizing radiation.

Analysis of causal network

Causal networks, shown in Figures 5 and 6, indicate interactions between computed temporal profiles. These networks are constrained to infer the simplest wiring diagram, in terms of directed graphs, for interpretation. The wiring diagram provides the means for interpretation of each computed template and its role in the dynamical system. Below, we analyze both the initial and final active templates.

The initial active templates are $T^1_{adaptive}$, $T^2_{adaptive}$, $T^3_{adaptive}$, $T^7_{adaptive}$, $T^2_{challenge}$ and $T^4_{challenge}$, and bioinformatic analyses suggest that the network is enriched by cell cycle, inflammatory, and apoptosis regulation processes. Although for the treatment group receiving the priming dose, enrichment is highly biased with inflammatory and immune responses. It is also clear that the oscillatory template $T^3_{adaptive}$ and delayed activation $T^6_{adaptive}$ play a role in the adaptive response (e.g., the treatment group receiving the priming dose) at an intermediate stage. Had there been no connections, then these templates would have been non-essential.

The final active templates (e.g., the interval between 4–8 hour time point) are largely enriched by the identical temporal signatures in $T^1_{adaptive}$, $T^4_{adaptive}$, $T^6_{adaptive}$, and $T^3_{challenge}$. While FOXO3A and NFkB pathways are highly enriched through multiple receptors with the group receiving the priming dose, enrichment in the challenge group is limited to EGRI. Enrichment of FOXO3A is potentially due to PDPK1, which is differentially expressed between the two treatment groups, as shown in the top row of Figure 7. However, there is no

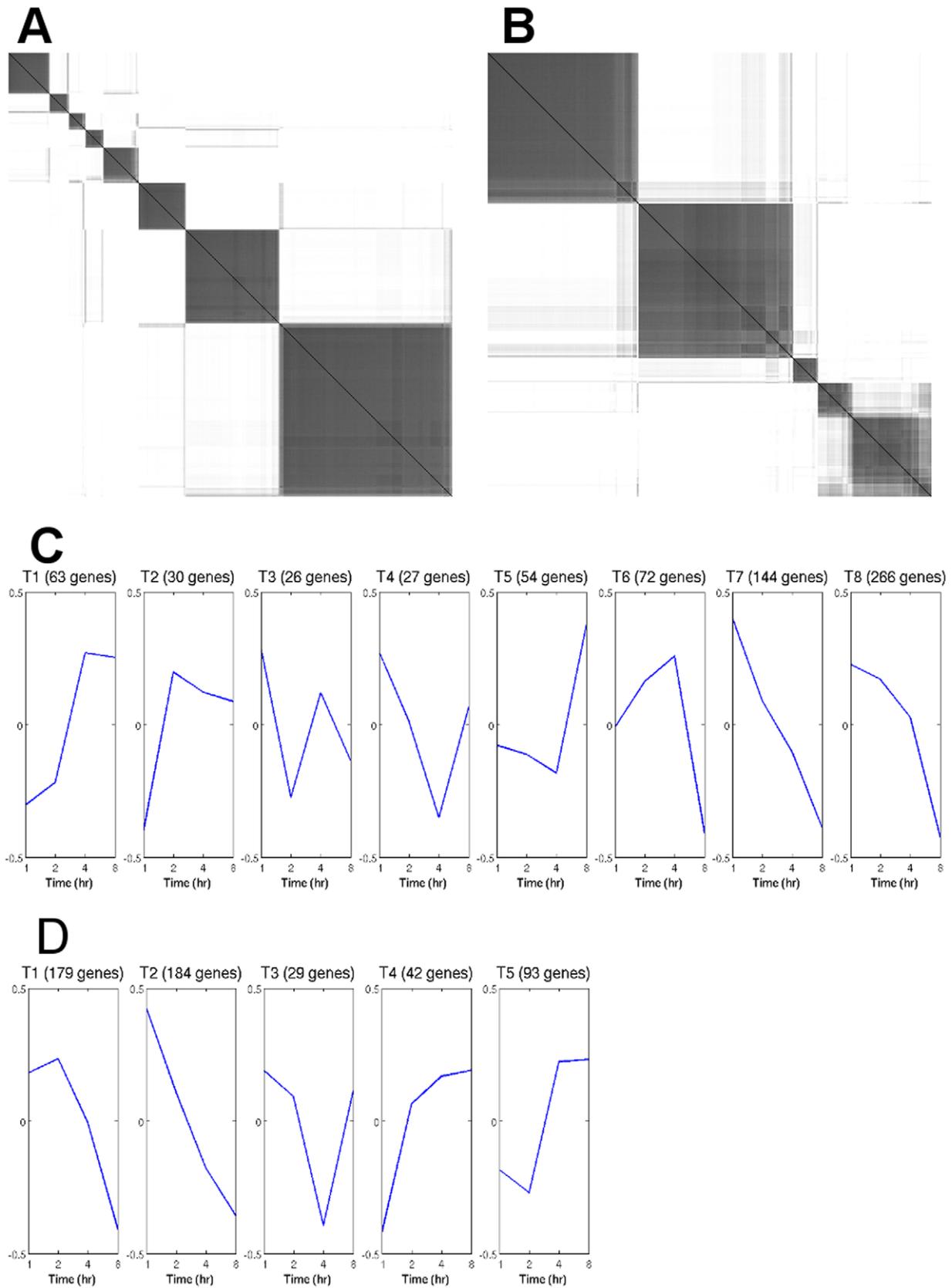


Figure 1. Analysis of temporal profile. Consensus clustering indicates A) 8 clusters for transcript data with the priming dose and B) 5 clusters in the control group. Each cluster, in A) and B), corresponds to a unique temporal profile as shown in C) and D), respectively.
doi:10.1371/journal.pone.0042306.g001

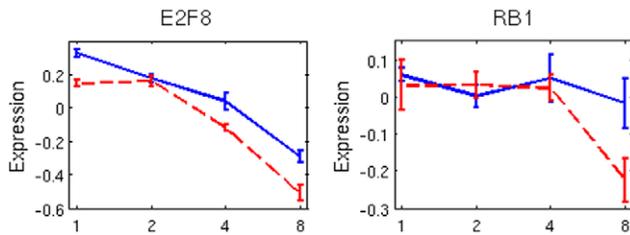


Figure 2. Transcriptome profile of E2F family and RB1 indicates down-regulation for the treatment group with the priming dose. RB1 remains unchanged in the absence of the priming dose. Dash red-lines correspond to adaptive response.
doi:10.1371/journal.pone.0042306.g002

differentially expressed signal in FOXO3A transcription factor even though EGR1 and FOXO3A are shown to be associated with the exposure to the ionizing radiation [18,19]. EGR1 functions as a hub for many signaling cascades that is essential for growth and proliferation [20], and is down-regulated as a result of ionizing radiation in both groups. Within the FOXO group, FOXO4 has a slight upregulation in both treatment groups. FOXO transcription factors are highly conserved genetic pathways, at the intersection of aging and cancer, that are phosphorylated in response to insulin and growth factors [21]. Upregulation of FOXO can cause apoptosis through an independent p53 pathway [19], whereas loss of Forkhead FOXO transcription factors in a cancer cell may decrease cell cycle arrest or apoptosis as a result of DNA damage or genomic instability [22,23]. Enrichment of NFkB is potentially through the TNF family. More specifically, TNFRSF10B is an activator of NFkB while TNFRSF11B is an inhibitor of NFkB. Their profiles are shown again in Figure 7 (lower row).

Discussion

From the perspective of a strict gene expression, the fold changes are generally low and appear to be stochastic as a result of ionizing radiation. This observation is consistent with previous literature [24,25]. Nevertheless, the temporal patterns from the gene expression provide more candidates and are more informative than a single time point observation, i.e., any transcript with a small value, at a given time point, can be eliminated using standard filtering techniques. The richness of the temporal gene

expressions is crucial in grouping and hypothesizing causal relationships from high dimensional transcriptome data. Typically, inference of the causal relationships can be ambiguous; there is significant literature in support of it [26] and against it [27], but most researchers agree that through carefully designed experimental data, ambiguities in the inference of causation can be reduced or eliminated. Such an experimental design may include a specific set of perturbations (e.g., siRNA) that may also include the time-course data. The time-course enables identification of a set of similar profiles that will (i) reduce complexities in the causal network, (ii) provide pseudo replicates for sampling and cross validation, and (iii) constrain the network structure (and the solution) by enforcing temporal continuity. In short, the proposed computational protocol enables interpretation of a complex dataset at multiple steps. However, the main theme is inference of the simplest network that is computationally tractable, and at the same time, interpretable. The method is initiated by identifying temporally co-regulated transcripts into a distinct set of templates or groups. This step not only reduces the dimensionality of the data, but also reduces the number of variables that need to be estimated for building the causal network, i.e., transition matrices. The network construction assumes a model for which every node, at a given time point, is a sparse linear combination of nodes in the previous time point. The concept of sparseness also enforces the notion of network simplicity. Finally, the solution is regularized by eliciting continuity of the transition matrices between consecutive time points. It should be noted the method has been applied to transcriptome data, but it is also extensible to other time-course data, i.e., identifying aberrant signal transduction pathways.

The method has been validated on synthetic data and then applied to transcriptome data that has been collected from a cell strain, which was exposed to 2 Gy ionizing radiation (e.g., the challenge dose) with and without the priming dose of 10 cGy applied 4 hours prior to the higher dose of radiation. Bioinformatics analyses revealed that computed templates (e.g., clusters) without the priming dose (e.g., in the challenge group) are a subset of those that received the priming dose (e.g., the adaptive response group). Furthermore, the adaptive response group included templates with delayed activations and oscillatory behavior. It is clear that the priming dose has induced a significant amount of diversity in how the networks are modulated. In both treatment groups, the initial active templates of the causal networks are highly enriched by the down-regulation of the cell cycle

Table 1. Ariadne signaling pathway and subnetwork enrichment with priming dose with the default p-value of 0.05.

Templates with priming dose	Signaling pathway enrichment	Subnetwork enrichment
T1	Apoptosis regulation, TNFR->(AP1, NF-kB, CREB), IL1R->STAT3, FOXO3A with p-value<0.02	TP53, SP1, GF, TNF, IL1B, TGFB1, IFNG, Cytokines with p-value<0.004
T2	TP53 signaling with p-value<0.008	TP53, TP73, STAT1, E2F1, interferon, ESR1, CREB1, NF-kB, IFNG, SP1
T3	(TNFR,TLR3, CSF2R, DDR1)->NF-kB, Mast cell activation, DDIT3, HSF1, AP-1 with p-value<0.03	NF-kB, TGFB1, TNF
T4	(HGFR, PDGFR, VEGFR)->FOXO3A, multiple receptors->NF-kB signaling	Proteasome endopeptidase complex with p-value<0.03
T5	STAT signaling	POU2F1, CEBPA, RCA1
T6	Hedgehog, Cell cycle, Notch,STAT3, STAT, STAT1	NOG, PDGF, EDN1, IL1A
T7	Cell cycle regulation with p-value<0.0002	IGF1, VEGFA, Ras
T8	Cell Cycle regulation, hedgehog	E2F, RB1, TP53, MYC, PDGF

doi:10.1371/journal.pone.0042306.t001

Table 2. Pathway and subnetwork enrichment without the priming dose with the default p-value of 0.05.

Templates without the priming dose	Signaling pathway enrichment	Subnetwork enrichment
T1	Hedgehog, Cell cycle regulation	E2F, TP53, RB1, MYC, SP1, PDGF
T2	Cell cycle, Notch->(MEF, TCF3, NF-kB)	Proteasome endopeptidase complex, TP53, cyclines, NF-kB, IL1B, TGFB1, TNF with p-value<0.01
T3	none	EGR1, GH1
T4	TNFR->(AP1, TP53, NF-kB, CREB), CD43, TNFRSF6->(HSF1, FOXO3A, RB1, E2F)	TP53, AKT1, MAP, SP1, NF-kB, IFANG, TNF with p-value<0.02
T5	Cell cycle regulation, Apoptosis	TP53, SP1, IL1B, EGF, TNF

doi:10.1371/journal.pone.0042306.t002

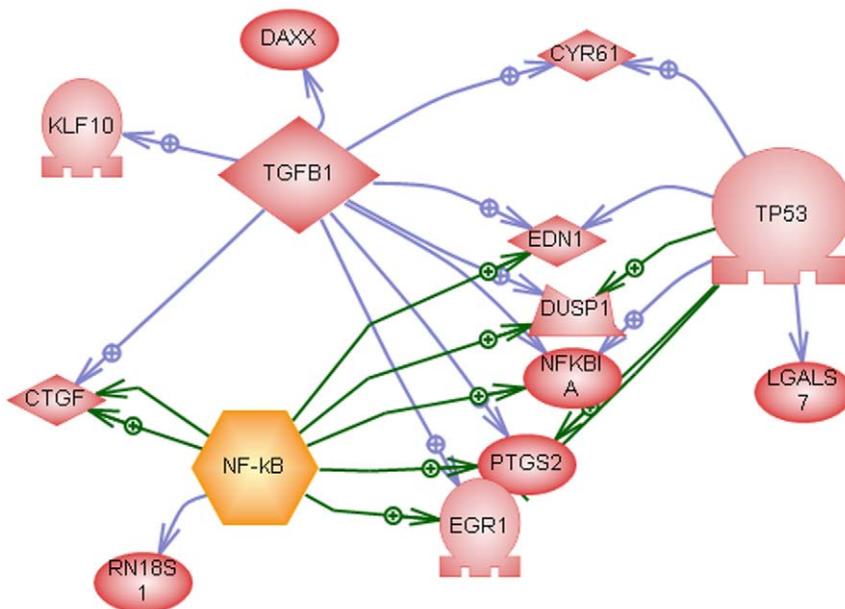
machinery. However, in the case of the adaptive causal network, the network is also modulated by the up-regulation of the inflammatory processes. On the other hand, with the exception of EGR-1, the network is poorly enriched at the late stages for the treatment group that does not receive a priming dose. It has been suggested that both EGR-1 and p53 are essential for mediating radiation-induced apoptosis [28]. The effect of priming at a late stage indicates a network modulation through pro-inflammatory responses and proteasomes, which is also consistent with the literature on low dose exposure [29,30].

Another way to examine experimental data is through enrichment of cellular processes. Initially, both treatment groups are enriched by DNA double strand repair, apoptosis, and cell cycle processes. However, the group receiving the priming dose is also enriched with single strand base excision and mismatch DNA repair. Within the group receiving the priming dose, these processes are modulated with a chromatin remodeling (e.g., $T_{adaptive}^6$) at 8 hour time point. On the other hand, the group receiving no priming dose appears to be poorly enriched in the final stages. The bioinformatics analysis suggests that the priming dose changes the network architecture by delaying the effects of the chromatin remodeling.

Materials and Methods

Experimental design

The experiment is designed around the human diploid, embryonic lung, fibroblast cell strain WI-38 (TP53 proficient) [31,32], which is publicly available from the Coriell Institute. These cells were grown as a monolayer (2D) under a physiologically relevant oxygen concentration (3%) with 10% CO₂, instead of at high oxygen levels (about 20%). Additionally, the cells were asynchronously growing when exposed to 2 Gy (e.g., the challenge dose) of ionizing radiation (160 kV X-rays), with or without a priming dose of 10 cGy (e.g., the adaptive dose), 4 hours prior to the challenge dose. Three biological replicates for each treatment group (e.g., with and without the priming dose) were collected 1, 2, 4 and 8 hours after the challenge dose. The time-course was selected on the basis of our prior research on early responses to ionizing radiation [33,34]. Purified total cellular RNA was extracted using an RNeasy Mini Kit (Qiagen) and quantified for Affymetrix microarray analysis using the Human Gene 1.0 ST Array. A robust multi-array analysis (RMA) was performed to the normalize data collected from the different samples. Samples were then examined for quality control using the NUSE protocol, which

**Figure 3.** Subnetwork enrichment analysis of template 3 with priming dose.

doi:10.1371/journal.pone.0042306.g003

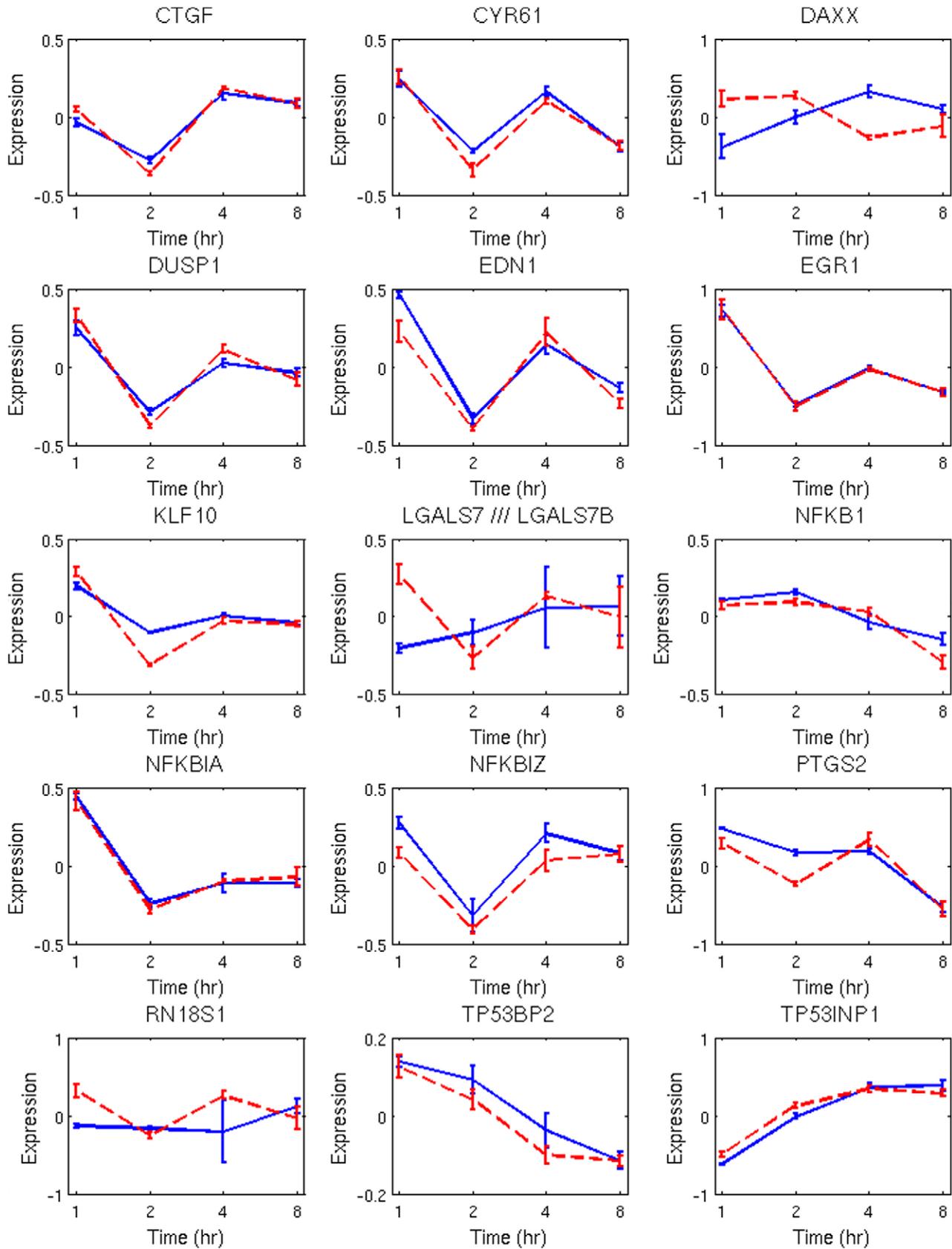


Figure 4. Temporal profiles of endpoints in Figure 4 with a significant fold change. Dash red-lines correspond to adaptive response. doi:10.1371/journal.pone.0042306.g004

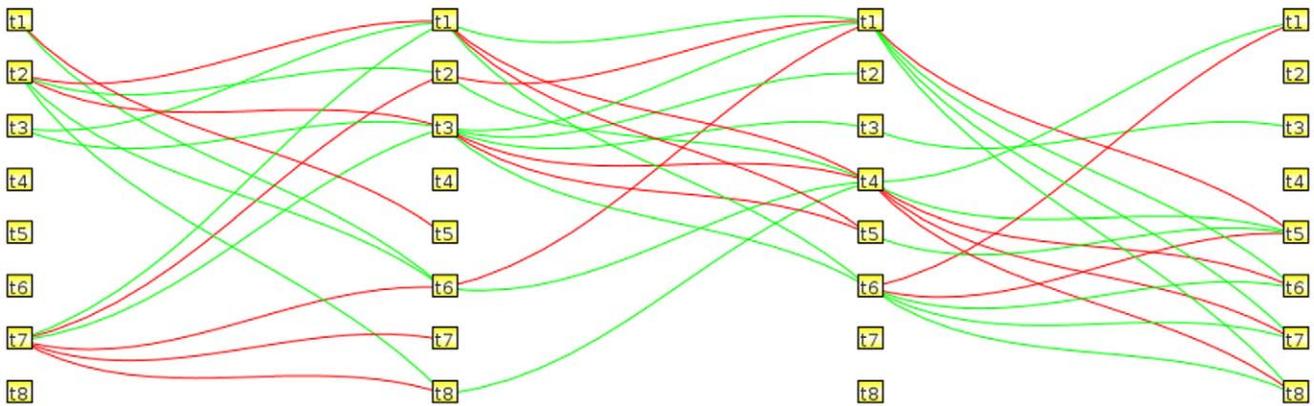


Figure 5. Causal networks for the 8 templates with priming dose in Figure 2C.
doi:10.1371/journal.pone.0042306.g005

is shown in Figure S5 and Text S2. The array data is publicly available with accession number of GSE37688.

Computing temporal templates from gene expression data

The first step of our protocol is to eliminate transcripts with little variation, which are maximum folds of change less than 0.5. The net result is a significant reduction in the number of candidate transcripts, with those having similar temporal profiles being grouped together. The basic assumption is that co-regulated transcripts have a similar biological basis and is a step towards significant dimensionality reduction through clustering and categorization. Currently, there is an abundance of literature available on the clustering of time-varying expression data that includes predefined templates [35], autoregressive models [36], curve-based clustering [37], and mixture models [38]. Nonetheless, this is not the main theme of our research. Our approach relies on constructing template profiles through consensus clustering [39], widely used for class discovery, and then leveraging higher level enrichment analysis for evaluation. Consensus uses a voting strategy on the resampled data, from different runs, with a clustering algorithm (e.g., k -means), and facilitates visualizing of computed clusters for quality control. In our implementation, the clustering algorithm is based on k -means, where the distance measured is one minus the sample correlation. The clustering procedure is repeated for 1000 runs, and each run is performed on the randomly sampled genes with a sampling rate of 0.8. The optimal number of clusters is determined by examining the clustering stability and similarity matrix.

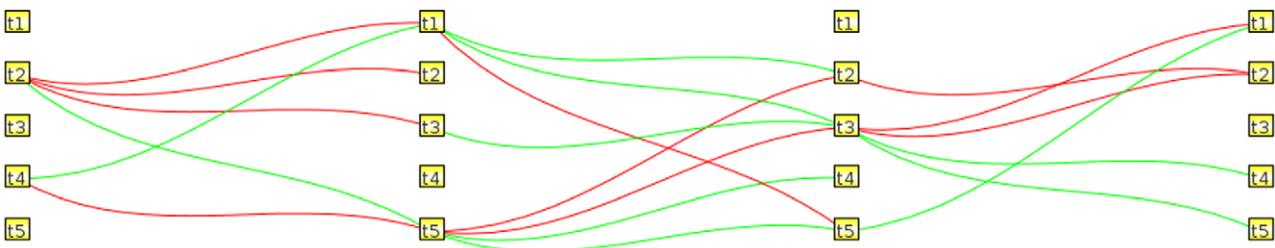


Figure 6. Causal networks for the 5 templates without priming dose in Figure 2D.
doi:10.1371/journal.pone.0042306.g006

Inference of causal networks

Suppose we have clustered genes into K groups, and each group contains n_k genes, i.e., $\mathbf{G}^K = \{g_1^K, g_2^K, \dots, g_{n_k}^K\}$. All the genes in the same group are assumed to have similar temporal patterns, which can be thought of as being denoted by a representative pattern, $v^K \in \mathbb{R}^{1 \times T}$, where T is the number of time frames. By concatenating the representative pattern for the K groups, we can obtain a K -by- T data matrix

$$P = \begin{bmatrix} v^1 \\ \dots \\ v^K \end{bmatrix}. \quad (1)$$

Our objective is to analyze the temporal causal relationships among the gene expression profiles for the K groups, i.e., whether the expression level of group- i at time t will have an impact on the expression level of group- j . More systematically, we use the following matrix equation to encode the causal relationships,

$$P^{t+1} = W^t P^t. \quad (2)$$

Here, $P^t \in \mathbb{R}^{K \times 1}$ denotes the t -th column of P , and $W^t \in \mathbb{R}^{K \times K}$ is the time-varying coefficient matrix, where $W^t(i, j)$ depicts how the expression level for group- j at time t affects the expression level of group- i at time $t+1$. A positive $W^t(i, j)$ indicates a positive correlation, and a negative one means a depression in the expression level.

However, in practice, we are confronted with the big challenge of processing limited data while estimating the large parameter

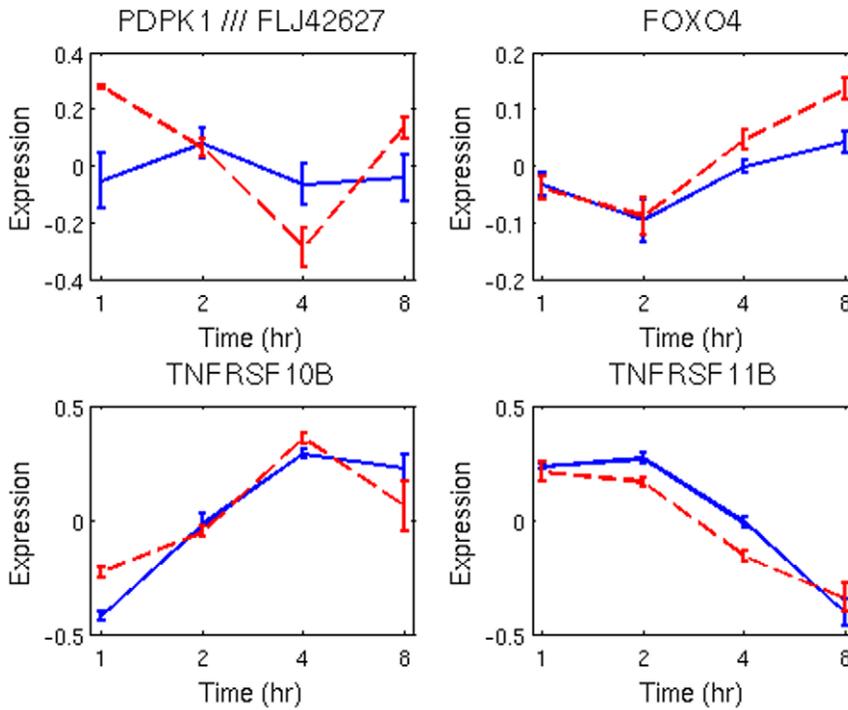


Figure 7. Temporal profiles of PDPK1, FOXO4, and TNF family involved in activation and inhibition of NFkB. Dash red-lines correspond to adaptive response.
doi:10.1371/journal.pone.0042306.g007

space of W^t . Additionally, more complications are introduced as a result of the time-varying nature of W^t which increases the number of parameters by a factor of T . To alleviate the lack of samples in the inference problem, we propose the following extensions:

(i) A temporal regression framework for the time course data.

Instead of using only one column in Equation (2), we will apply a few columns together, corresponding to a small sliding window (multiple adjacent time frames) in the forward direction:

$$[f_{t+1} P^{t+1}; f_{t+2} P^{t+2}; \dots; f_{t+\Delta} P^{t+\Delta}] = W^t [f_t P^t; f_{t+1} P^{t+1}; \dots; f_{t+\Delta} P^{t+\Delta}], \quad (3)$$

Where $f_t = e^{-(t-t_0)/h^2}$ is a weighting function that assigns smoothly decaying weight to the distance between t and the current centering frame t_0 . We simplify the notation of the linear equation group (3) by

$$P_{\Delta}^{(t+1)} = W^t P_{\Delta}^{(t)}, \quad (4)$$

where $P_{\Delta}^{(t)} = [f_t P^t; f_{t+1} P^{t+1}; \dots; f_{t+\Delta} P^{t+\Delta}] \in R^{K \times (\Delta+1)}$. This is illustrated in Figure 8, where we use a sliding window of $\Delta = 2$. As one can see, for current centering frame $t_0 = 1$, the equality $P^2 = W^1 P^1$ is assigned the largest weight, $P^3 = W^1 P^2$ has a smaller weight, and $P^4 = W^1 P^3$ has a diminishing weight. This means that when computing W^1 , most of the emphasis should be placed on the constraint closest to the current time frame, i.e., the relationship between P^1 and P^2 . By doing this, we equivalently increase the number of samples used to compute each W^t , while at the same time assigning smaller weights to samples that are temporally too far away from the current centering frame. Mathematically, we use the regression framework for computing

the coefficient matrices,

$$\min \sum_{t=1}^{T-1} \|P_{\Delta}^{(t+1)} - W^t P_{\Delta}^{(t)}\|_F^2. \quad (5)$$

(ii) The coefficient matrices should vary smoothly with time.

Temporal coherence of W^t is an important constraint that can be utilized to enforce reliable estimations. We assume that the mechanism for gene-gene interactions is unlikely to change drastically over small time intervals. This is then translated into a smoothness term for W^t 's. More explicitly, if $t1$ and $t2$ are in close proximity to each other (according to a predefined range), then W^{t1} and W^{t2} should also be close to each other. To enforce similarity between adjacent W^t 's, we also include the following term

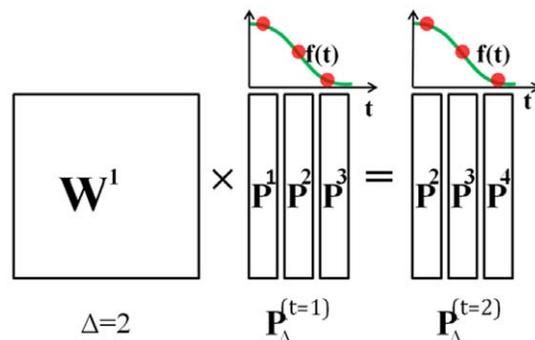


Figure 8. Illustration of the sliding window regression.
doi:10.1371/journal.pone.0042306.g008

$$\min \sum_{i,j=1}^{T-1} c_{ij} \|W^i - W^j\|_F^2.$$

Here, c_{ij} is an indicating factor that determines whether i and j are close to each other: if so, c_{ij} should be a positive number, enforcing the closeness between W^i and W^j in the above minimization term; otherwise c_{ij} will be zero, indicating no constraints on W^i and W^j if i and j are far away. In practice, one can have either a hard indicator,

$$c_{ij} = \begin{cases} 1, & |i-j| < decay \\ 0, & |i-j| > decay \end{cases},$$

or a soft indicator,

$$c_{ij} = e^{-\frac{(i-j)^2}{h^2}}.$$

(iii) The coefficient matrices are sparse.

We note that the matrix W^t uniquely specifies a directed graph, where each node is a gene group and the edge weights indicate the interaction between the groups. However, only a subset of genes is active at any time point, which means that the network structure is sparse, i.e., only a small portion of the node pairs have interactions with each other. The notion of sparsity can be enforced via penalizing the l_1 -norm of the coefficient matrix, as

$$\min \sum_{t=1}^{T-1} \|W^t\|_1.$$

This is typically known as sparse constraint. In recent years, it has been applied extensively in signal processing, image reconstruction, and model selection [40,41].

By combining the three terms, we have the following optimization problem

$$\min_{\substack{W^t \in \mathbb{R}^{K \times K} \\ t=1,2,\dots,T-1}} \sum_{t=1}^{T-1} \|P_{\Delta}^{(t+1)} - W^t P_{\Delta}^{(t)}\|_F^2 + \lambda_1 \sum_{i,j=1}^{T-1} c_{ij} \|W^i - W^j\|_F^2 + \lambda_2 \sum_{t=1}^{T-1} \|W^t\|_1$$

Here, λ_1 and λ_2 are positive regularization parameters that control the balance between the loss term, the temporal smoothness term, and the sparsity of solutions.

Using multiple replicates for regression

As we previously indicated, a critical issue is the low sample size given the high dimensionality of the parameter space. Thus, to improve robustness and stability of the solution, we adopt the policy of using individual transcripts as replicates as they have similar signatures within a clustered group. In practice, one can compute a representative expression profile v^i for the i -th group of genes, as discussed. However, in some cases, genes in the same cluster still have a certain level of variation, and using their average pattern for regression might lead to loss of information. To solve this problem and be able to fully utilize available patterns, we randomly sample genes from each group as the representative v^i to form the data matrix (1); we then repeat this process and create an altogether \mathcal{N} data matrices; each data matrix will lead to one objective term, as specified in (4). We will then sum up the

objective terms associated with all the data matrices as the ultimate objective function. We can sample as many times as needed, i.e., injecting more constraints to the optimization problem, using certain heuristics for sampling. For example, given K groups and n_k genes for k -th group, then the total number of different data matrices can be $\prod_{k=1}^K n_k$.

Optimization protocol

There are two ways to obtain the solution: (i) concatenate the columns of each W^t as a K^2 -by-1 vector, and concatenate the T - I matrices to form a $K^2(T-1)$ -by-1 vector as the variable. The whole objective function can be transformed to a standard quadratic programming problem with a sparsity constraint, which can then be transformed into a l_1 -regularized least-square problem for which many advanced solvers are available; or (ii) the block coordinate descent method can be incorporated. Instead of computing all the variables at one time, we update one W^t matrix at a time while fixing all other W^i 's ($i \neq t$). The first approach requires larger memory due to the need to manipulate the $K^2(T-1)$ -by- $K^2(T-1)$ Hessian matrix. In practice, when K is very large, we apply the Nystrom low-rank approximation by sampling only a subset of the rows/columns of the Hessian matrix. This allows us to maintain enough energy in the eigen-spectrum in the reconstructed Hessian. The second approach is more memory efficient, but may require many cycling updates to converge.

Selection of regularization parameters

The evaluation of regularization parameters $\{\lambda_1, \lambda_2\}$ is based on Bayesian information criterion (BIC). Similar to Ahmed and Xing [42], we define the BIC as

$$BIC(\lambda_1, \lambda_2) = n \ln(\sigma^2) + d \ln(n)$$

Where $n = \sum_{t=1}^T N_t$, $\sigma^2 = \frac{1}{n-1} \sum_{t=1}^{T-1} \|P_{\Delta}^{(t+1)} - W^t P_{\Delta}^{(t)}\|_F^2$, and $d = \sum_{t=1}^{T-1} \sum_{i,j} I[W_{ij}^t \neq 0, W_{ij}^{t-1} = 0]$. We chose $\{\lambda_1, \lambda_2\}$, which gives the smallest BIC from the candidate models.

Validation with the synthetic data

We validated the network inference against a set of synthetic data. In the example, shown in Figure S6, the network consists of $k=9$ nodes with $T=4$ time-varying states. The transition matrices, W^t with $t=1,2,\dots,T$, is designed as sparse k -by- k matrices with roughly 10% non-zero entries in the range of $[-1,+1]$. Furthermore, adjacent matrices are designed to be similar with random perturbation of 10 entries in W^t and replacing them in W^{t+1} with $t=1,2,\dots,T$. This is the same policy, used in [42], by collecting samples from each state and adding random noise proportional to the 10% of the signal.

Comparison with related methods

With respect to network recovery, Ahmed and Xing [42] proposed to recover the network in social and biological studies by a temporally smoothed, l_1 -regularized logistic regression formalism. Promising results were reported on reverse engineering of the latent sequence of temporally rewiring political and academic social networks, and the evolving gene networks during the life cycle of *Drosophila melanogaster*. However, there are several differences between our method and theirs: (i) the loss function in our formulation integrates the error term between the actual and reconstructed gene expression value over multiple time frames in a sliding-window, while in [42] the loss term for each time t is the

log-likelihood of the model at time t ; (ii) [42] only considers the closeness between W^t 's of directly adjacent time frames, while we can control the clones of all W^t 's with a flexible decaying function to adjust the weight that is highly desired in cases of limited training data; (iii) we perform a pre-clustering step to reduce the number of parameters, which also allows us to generate a large number of "pseudo-samples" (data matrices) by sampling from each group. In summary, we have improved the stability of inference by incorporating the pre-clustering step. An alternative approach includes Bayesian techniques, where the notion of sparsity is often enforced by specifying the prior distribution for the graph that panelizes the number of edges. While Bayesian methods have been shown to infer biological networks [43,44], they (i) are generally greedy in terms of structure inference, and when coupled with advanced MCMC methods to reduce local trapping among possible structures [45], they are compute intensive; (ii) have limitations in terms of inference of feedback loops since network inference is a feed forward action (e.g., network is acyclic) [46]; and (iii) can infer edges based on probabilistic values, and as a result the notion of excitatory and/or inhibitory (e.g., positive and negative edges), they are lost. The main pitfalls are that the number of templates, during the pre-clustering stage, may alter the topology of the causal network, and network inference may be sensitive to a small sample size. The latter is a shortcoming that persists in most systems that infer structures from the data. A potential improvement includes further regularization by incorporating prior knowledge (e.g., KEGG, HPRD) to enforce locality and consistency with the published literature.

Supporting Information

Figure S1 Visualization of the consensus matrix of $N = 2, 3, \dots, 9$ clusters for the adaptive dose.

(TIF)

References

- Barabasi A, Oltvai Z (2004) Network biology: understanding the cell's functional organization. *Nature Review Genetics* 5: 101–103.
- Fauer A, Naldi A, Chaouiya C, Thieffry D (2006) Dynamical analysis of generic Boolean model for control of mammalian cell cycle. *Bioinformatics* 22: 124–131.
- McAdams H, Arkin A (1997) Stochastic mechanisms in gene expression. *Proceedings of National Academy of Science* 94: 814–819.
- Cinquermani E, Millias-Argetitis A, Summers S, Lygeros J (2008) Stochastic dynamics of genetic networks. *Bioinformatics* 24: 2748–2754.
- Chaouiya C (2007) Petri Net modeling of biological networks Briefings in *Bioinformatics* 8: 210–219.
- Li S, Brazhnik P, Sobral B, Tyson J (2008) A quantitative study of the division of the cell cycle of *Caulobacter Crescentus* stalked cells. *PLoS Computational Biology* 4.
- Edwards J, Ibarra R, Palsson B (2001) In Silico predictions of *Escherichia Coli* Metabolic capabilities are consistent with experimental data. *Nature Biotechnology* 19: 125–130.
- Karlebach G, Shamir R (2008) Modeling and analysis of gene regulatory networks. *Nature Reviews in Molecular and Cell Biology* 9: 770–780.
- Tenazinha N, Vinga S (2011) A survey of methods for modeling and analyzing integrated biological networks. *IEEE Transactions on Computational Biology and Bioinformatics* 8: 943–958.
- Tanya K, Hooker A, Zeng G, Sykes P (2007) Low dose X-radiation adaptive response in spleen and prostate of Atm knockout heterozygous mice. *International Journal of Radiation Biology* 83: 523–534
- Prise K (2006) New advances in radiation biology. *Occupational Medicine* 56: 156–161.
- Tapio S, Jacob V (2007) Radioadaptive response revisited. *Radiation Environment Biophysics* 46: 1–12.
- Nevins J (1988) Toward an understanding of the functional complexity of the E2F and Retinoblastoma families. *Cell Growth and Differentiation* 9: 585–593.
- Worku D, Jouhra F, Jiang G, Patani N, Newbold R, et al. (2008) Evidence of a tumor suppressive function of E2F1 gene in human breast cancer. *Anticancer Research* 28: 2135–2140.
- Lin W, Lin F, Nevins J (2011) Selective induction of E2F1 in response to DNA damage, mediated by ATM-dependent phosphorylation. *Genes and Development* 15: 1833–1844.
- Barcellos-Hoff M (2005) How tissues respond to damage at the cellular level: Orchestration by transforming growth factor- β (TGF- β). *BJR Suppl* 27: 123–127.
- Voy B, Scharff J, Perkins A, Saxton A, Borate B, et al. (2006) Extracting gene networks for low-dose radiation using graph theoretical algorithms. *PLoS Computational Biology* 2(7): e89.
- Wichselbaum R, Hallhan D, Fuks Z, Kufe D (1994) Radiation induction of immediate early genes: effectors of the radiation-stress response. *Int J Radiation Oncology Biol Phys* 30: 229–234.
- Yang J, Xia W, HU M (2006) Ionizing radiation activates expression of FOXO3a, Fas Ligand, Bim, and induces cell apoptosis. *International Journal of Oncology* 29: 643–648.
- Thiel G, Cibelli G (2002) Regulation of life and death by the Zinc finger transcription factor Egr-1. *Cellular Physiology* 193: 287–292.
- Greer E, Brunet A (2008) FOXO transcription factors in ageing and cancer. *Acta Physiol* 192: 19–28.
- Accili D, Arden K (2004) FoxOs at the crossroads of cellular metabolism, differentiation, and transformation. *Cell* 117: 421–426.
- Burgering B, Kpos G (2002) Cell cycle and death control: long live Forkheads. *Trends Biochem Sci* 27: 352–360.
- Roy L, Gruel G, Vaurijoux A (2009) Cell responses to ionising radiation analyzed by gene expression patterns. *Ann Ist Super Sanita* 45: 272–277.
- Turtot A, Brown I, Oskamp D, Schneeweiss F (2008) Early gene expression in human lymphocytes after gamma irradiation—a genetic pattern with potential biodosimetry. *Int J Radiation Biology* 84: 375–387.
- Pearl J (2000) *Causality: Models, reasoning and inference*: Cambridge university Press.
- Robin JM, Wasserman L, editors (1999) *On the impossibility of inferring causation from association without background knowledge*. Menlo Park, Ca: AAAI Press/MIT Press. 305–321 p.
- Das A, Chendil D, Dey S, Mohiuddin M, Mohiuddin M, et al. (2001) Ionizing Radiation Down-regulates p53 Protein in Primary Egr-1 $^{-/-}$ Mouse Embryonic Fibroblast Cells Causing Enhanced Resistance to Apoptosis. *Biological Chemistry* 276: 3279–3286.
- McBride W, Pajonk F, Chiang C, Sun J (2002) NF-kappa B, cytokines, proteasomes, and low-dose radiation exposure. *Mil Med* 167: 66–67.
- Pervan M, Iwamoto K, McBride W (2005) Proteasome structure affected by ionizing radiation. *Mol Cancer Res* 3: 381–390.

Figure S2 Consensus CDF for the consensus matrix of $N = 2, 3, \dots, 9$ clusters for the adaptive dose as shown in Figure S1.

(TIF)

Figure S3 Visualization of the consensus matrix of $N = 2, 3, \dots, 9$ clusters for the adaptive dose.

(TIF)

Figure S4 Consensus CDF for the consensus matrix of $N = 2, 3, \dots, 9$ clusters for the challenge dose as shown in Figure S3.

(TIF)

Figure S5 NUSE (Normalized Unscaled Standard Error) plot of the microarray data.

(TIF)

Figure S6 An example of validation data in the top row with positive (red) and negative (green) values in the transition matrices. The bottom row shows inferred matrices through application of the computational method.

(TIF)

Text S1 Quality control for microarray data.

(DOCX)

Text S2 Selection of the number of clusters from consensus clustering.

(DOCX)

Author Contributions

Conceived and designed the experiments: BP. Performed the experiments: KZ JH TG. Analyzed the data: KZ JH GF BP. Wrote the paper: KZ JH BP.

31. Hayflick L (1965) The limited in vitro lifetime of human diploid cell strains. *Experimental cell research* 37: 614–636.
32. Hayflick L, Moorhead P (1961) The serial cultivation of human diploid cell strains. *Experimental cell research* 25: 585–621.
33. Groesser T, Chang H, Fontenay G, Chen J, Costes S, et al. (2011) Persistent of gamma-H2AX and 53BP1 foci in proliferating and non-proliferating human mammary epithelial cells after exposure to gamma-rays or iron ions. *International Journal of Radiation Research* 87: 696–710.
34. Han J, Chang H, Yang Q, Fontenay G, Groesser T, et al. (2010) Multiscale iterative voting for differential analysis of stress response for 2D and 3D cell culture models. *Microscopy* 241: 315–326.
35. Moller-Levet C, Chu K, Wolkenhauer O (2003) DNA microarray data clustering based on temporal variation: FcV with tcd preclustering. *Bioinformatics* 19: 834–841.
36. Ramoni M, Sebastiani P, Kohane P (2002) Cluster analysis of gene expression dynamics. *PNAS* 99: 9121–9126.
37. Luan Y, Li H (2003) Clustering of time-course gene expression data using mixed effects model with B-spline. *Bioinformatics* 19: 474–482.
38. Celeux G, Martin O, Lavergne C (2005) Mixture of linear models for clustering gene expression profiles from repeated microarray experiments. *Statistical modeling* 5: 243–267.
39. Monti S, Tamayo P, Mesirov J, Golub TR (2003) Consensus clustering: a resampling-based method for class discovery and visualization of gene expression microarray data. *Mach Learn* 52: 91–118.
40. Tibshirani R (1996) Regression shrinkage and selection via the lasso. *J Royal Statist Soc B* 58: 267–288.
41. Donoho D (2006) Compressed sensing. *IEEE Trans on Information Theory* 52: 1289–1306.
42. Ahmed A, Xing EP (2009) Recovering time-varying networks of dependencies in social and biological studies. *PNAS* 106: 11878–11883.
43. Ellis B, Wong WH (2008) Learning causal Bayesian network structures from experimental data. *American Statistical Association* 103: 778–789.
44. Hill SM, Neve RM, Bayani N, Kuo W-L, Ziyad S, et al. (2012) Integrated biological knowledge into variable selection: an empirical Bayes approach with an application in cancer biology. *BMC Bioinformatics* 13: 94.
45. Kou S, Zhou Q, Wong WH (2006) Equi-energy sampler: Applications in statistical inference and statistical mechanics. *The annals of Statistics* 34: 1581–1619.
46. Hena O, Winther O (2009) Bayesian sparse factor models and DAGs inference and comparison. pp. 736–744.