Molecular Systems Biology 4; Article number 225; doi:10.1038/msb.2008.63
**Citation:** ***Molecular Systems Biology*** **4:225**

www.molecularsystemsbiology.com

molecular systems biology

**REPORT**

# The temporal response of the *Mycobacterium tuberculosis* gene regulatory network during growth arrest

**Gábor Balázsi[1,*], Allison P Heath[2], Lanbo Shi[3] and Maria L Gennaro[3]**

[1] Department of Systems Biology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA, [2] Department of Computer Science, Rice University, Houston, TX, USA and [3] Public Health Research Institute, New Jersey Medical School, Newark, NJ, USA
* Corresponding author. Department of Systems Biology, U950, University of Texas MD Anderson Cancer Center, 7435 Fannin Street, Houston, TX 77054, USA. Tel.: +1 713 834 6157; Fax: +1 713 563 4235; E-mail: gbalazsi@mdanderson.org

Received 8.4.08; accepted 19.9.08

**The virulence of *Mycobacterium tuberculosis* depends on the ability of the bacilli to switch between replicative (growth) and non-replicative (dormancy) states in response to host immunity. However, the gene regulatory events associated with transition to dormancy are largely unknown. To address this question, we have assembled the largest *M. tuberculosis* transcriptional-regulatory network to date, and characterized the temporal response of this network during adaptation to stationary phase and hypoxia, using published microarray data. Distinct sets of transcriptional subnetworks (origons) were responsive at various stages of adaptation, showing a gradual progression of network response under both conditions. Most of the responsive origons were in common between the two conditions and may help define a general transcriptional signature of *M. tuberculosis* growth arrest. These results open the door for a systems-level understanding of transition to non-replicative persistence, a phenotypic state that prevents sterilization of infection by the host immune response and promotes the establishment of latent *M. tuberculosis* infection, a condition found in two billion people worldwide.**
*Molecular Systems Biology* 4 November 2008; doi:10.1038/msb.2008.63
*Subject Categories:* metabolic and regulatory networks; computational methods; microbiology and pathogens
*Keywords:* tuberculosis; network; origon; growth arrest; hypoxia



## Introduction

A hallmark of *Mycobacterium tuberculosis* infection is the switching of tubercle bacilli between replicative (growth) and non-replicative (dormancy) states in response to environmental cues generated by the host immune response (Wayne and Sohaskey, 2001; Warner and Mizrahi, 2007). When infection has progressed enough to induce adaptive immune responses, the bacilli survive by slowing down their growth and eventually entering a phenotypic state called dormancy, which enables *M. tuberculosis* to persist in the immunocompetent host for many years, causing asymptomatic (latent) infection. When host immunity falters, tubercle bacilli can resume growth and reactivate disease (Wayne and Sohaskey, 2001; Warner and Mizrahi, 2007).

Little is known about the dormant state of tubercle bacilli in human infection. A tractable surrogate for dormancy is the arrest (or drastic slowdown) of bacterial growth in particular stress conditions *in vitro*, including gradual $O_2$ depletion, treatment with nitric oxide (NO), and nutrient starvation. Microarray studies of *in vitro* cultures have defined transcriptional changes during hypoxia (Sherman *et al*, 2001; Voskuil *et al*, 2004), NO treatment (Voskuil *et al*, 2003), nutrient starvation (Betts *et al*, 2002; Hampshire *et al*, 2004), altered pH (Fisher *et al*, 2002), and treatment with detergents such as SDS (Manganelli *et al*, 2001). Robust markers of dormancy have emerged, such as the upregulation of the *dosR* regulon (Park *et al*, 2003). The induction of *dosR*-regulated genes in various dormancy models was further underscored by a recent meta-analysis of published microarray data (Murphy and Brown,

2007). Nevertheless, most of these studies have focused only on changes in the expression of individual genes in *M. tuberculosis* dormancy. Little, if any, attention has been given to the dynamic series of events that occur at the level of the gene regulatory network (Albert, 2005). To understand these aspects of gene regulation during transition to dormancy, time course microarray data should be overlaid with the large-scale transcriptional-regulatory (TR) network of *M. tuberculosis*, as done earlier for *Escherichia coli* (Balázsi *et al*, 2005; Ernst *et al*, 2008) and *Saccharomyces cerevisiae* (Ihmels *et al*, 2004; Farkas *et al*, 2006). However, the current database of *M. tuberculosis* gene regulation (Jacques *et al*, 2005) contains far fewer interactions than the TR network of *S. cerevisiae* (Harbison *et al*, 2004; Balaji *et al*, 2006) and *E. coli* (Salgado *et al*, 2006).

To address this problem, we assembled a large *M. tuberculosis* TR network and used previously published microarray data (Voskuil *et al*, 2004) to analyze the network-level response of *M. tuberculosis* to hypoxia and transition into stationary phase. Although the goal of most microarray data analysis methods is to identify individual genes that are significantly up- or downregulated, we aimed to identify significantly responsive subnetworks. This is motivated by the modular structure of biological networks (Wagner *et al*, 2007), where various sets of modules respond specifically to various types of environmental change. We found a distinct set of transcriptional subnetworks (origons) affected early and late during adaptation to hypoxia and stationary phase, indicating a progressive shift of modular network response to growth arrest. Most of the origons were affected in both conditions, suggesting the existence of a general, condition-independent repertoire of transcriptional modules utilized in *M. tuberculosis* growth arrest.

## Results and discussion

### Assembly of a large-scale *M. tuberculosis* TR network

We compiled a large-scale *M. tuberculosis* TR network using three main sources. The core of the TR network consists of 381 gene regulatory interactions documented in the literature, 222 of which have been collected in MtbRegList (Jacques *et al*, 2005), whereas 159 links were added in this study (see Materials and methods). We enlarged this core network including 223 *M. tuberculosis* gene pairs that have orthologs with confirmed TR relationship in *E. coli* (Babu *et al*, 2006). Finally, we augmented the network based on the full list of *M. tuberculosis* operons (Roback *et al*, 2007), assuming that transcription factor (TF) binding to the promoter region affects the expression of all genes within an operon. This is a reasonable assumption, as TF-promoter binding dictates the rate at which genes in a typical operon are co-transcribed into polycistronic mRNA (Jacob *et al*, 1960), although still allowing for post-transcriptional modulation of individual

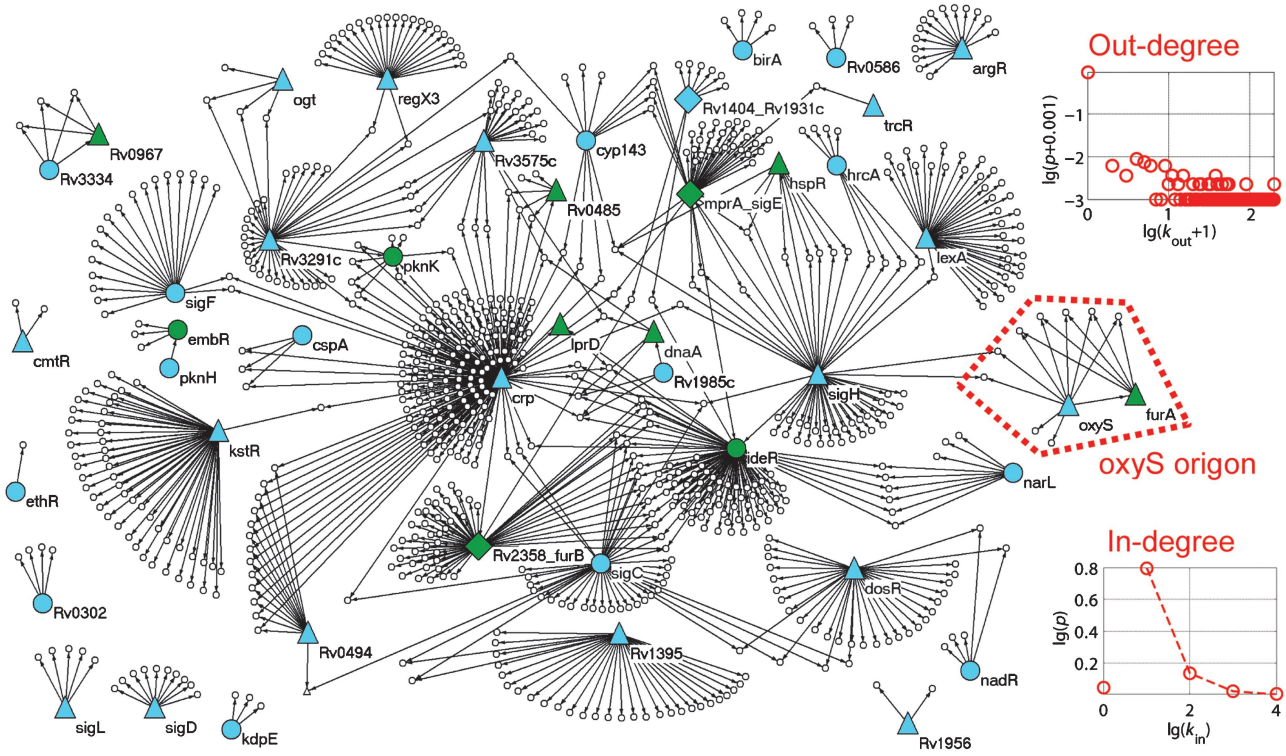


**Figure 1** The *M. tuberculosis* TR network assembled from publicly available sources. Input nodes (genes with no known transcriptional regulators) are shown in blue, whereas transit nodes (TFs with known transcriptional regulators) are shown in green. The white nodes represent output nodes (genes encoding proteins with no TF activity). Triangles mark nodes that autoregulate their own expression, whereas diamonds represent nodes that are part of two-gene feedback loops. As an example, The *oxyS* origon is indicated by the dashed red line. The insets show the distributions of out-degree (number of target genes a TF can regulate, on the top) and in-degree (number of regulators a gene can have, on the bottom). The dashed line indicates the exponential fit $f(x)=0.8e^{-1.78(x-1)}$.

**Table I** List of transcription factors in the large-scale TR network of *M. tuberculosis*

| Transcription factor | TF type |
|---|---|
| *Rv0117* (oxyS), *Rv0212c* (nadR), *Rv0302* (TetR/AcrR family), *Rv0491* (regX3), *Rv0494* (GntR family), *Rv0586* (GntR family), *Rv0735* (sigL), *Rv0844c* (narL), *Rv1027c* (kdpE), *Rv1033c* (trcR), *Rv1266c* (pknH), *Rv1316c* (ogt), *Rv1657* (argR), *Rv1785c* (FruR-like/cyp143), *Rv1931c* (AraC/xylS family), *Rv1956* (HTH TF), *Rv1985c* (LysR family), *Rv1994c* (cmtR), *Rv2069* (sigC), *Rv2374c* (hrcA), *Rv2720* (lexA), *Rv3133c* (dosR), *Rv3223c* (sigH), *Rv3279c* (birA), *Rv3286c* (sigF), *Rv3291c* (Lrp/AsnC family), *Rv3334* (MerR family), *Rv3414c* (sigD), *Rv3574* (TetR family), *Rv3575c* (LacI family), *Rv3648c* (cspA), *Rv3676* (Crp/Fnr family), *Rv3855* (ethR), *Rv1395* (AraC/xylS family) | Input |
| *Rv0001* (dnaA), *Rv0353* (hspR), *Rv0485* (NagC/XylR family), *Rv0967* (YvgZ-like), *Rv1221* (sigE), *Rv1267c* (embR), *Rv1343c* (LprD), *Rv1909c* (furA), *Rv2359* (furB), *Rv2711* (ideR), *Rv3080c* (pknK) | Transit |

gene expression (Nudler and Gottesman, 2002; Li and Altman, 2004; Isaacs *et al*, 2006; Pfleger *et al*, 2006).

The 783 nodes in the TR network (see Figure 1 and Supplementary information) correspond to *M. tuberculosis* genes and their protein products, whereas the 937 links correspond to 45 TFs (Table I) directly regulating the expression of target genes. Remarkably, 29 of these 45 TFs regulate their own expression, demonstrating the importance of autoregulation in prokaryotic gene networks (Thieffry *et al*, 1998). In addition, the gene pairs *Rv2358-furB*, *Rv1404-Rv1931c*, and *mprA-sigE* participate in two-gene feedback loops (Figure 1).

We consider 381 (41%) of the 937 interactions in this network relatively reliable, because they are based on experimental studies of 26 TFs binding and regulating the expression of 355 target genes. The operon-based extension of the literature-derived network has 581 links (62%) among 518 genes, which are somewhat less reliable. Finally, the 223 regulatory interactions among 201 genes inferred from orthology with *E. coli* TF–target gene pairs (Babu *et al*, 2006) might have the lowest confidence because orthologous TFs in bacteria can have different functions and can regulate different genes (Price *et al*, 2007). In fact, only 4 of the 223 orthology-based links are in common with the literature-based links. Still, the operon-based expansions of these two networks (581 and 410 links, respectively) share 54 links, supporting the inclusion of orthology-based links into the network.

To the best of our knowledge, this is the largest TR network of *M. tuberculosis* that has been assembled to date, comprising ~20% of its genome. In comparison, the current version of the *E. coli* TR network (excluding sigma factors) contains 1364 genes (~35%) of the *E. coli* genome (Salgado *et al*, 2006). We expect that this large-scale TR network will be a valuable resource for the *M. tuberculosis* research community, complementing existing efforts of genome-scale data integration (see, for example http://www.tbdb.org).

## Topological properties of the *M. tuberculosis* TR network

To quantitatively characterize the topology of the newly assembled *M. tuberculosis* TR network, we analyzed and compared its connectivity distribution with that of other existing TR networks. The out-degree distribution (Albert, 2005) did not follow a power law (Khanin and Wit, 2006), but had a heavy tail, indicating that a small number of TF hubs regulate a very large number of targets, whereas most TFs regulate few or no targets. On the other hand, the in-degree distribution had a near-exponential tail to the right of a peak for genes with one regulator, indicating that most genes have only one known transcriptional regulator (Figure 1). Such differences between in- and out-degree distributions have been observed for other TR networks (Thieffry *et al*, 1998; Guelzim *et al*, 2002), suggesting a general property of TR network topology (see the Supplementary information for a detailed analysis and comparison with the TR networks of *E. coli* and *S. cerevisiae*).

The 783 genes in the *M. tuberculosis* TR network can be arranged hierarchically (Balázsi *et al*, 2005) into four layers, which reflect the flow of information from the 34 *input nodes* (representing 15 TFs that are transcriptionally unregulated and 19 TFs that are regulated only by feedback loops) to the 735 *output nodes* (representing genes that do not directly regulate the expression of other genes). The 11 nodes that are neither input nor output nodes are *transit nodes* (Table I). Input and transit nodes mediate information entry into the TR network because their TR activity is affected by various intra- or extracellular changes (Martinez-Antonio *et al*, 2006). Most (34/45) TFs are input nodes, similar to *E. coli*, but unlike *S. cerevisiae* (see the Supplementary information). This may reflect the simplicity of bacterial TR networks as compared with eukaryotes, indicating that bacteria are equipped with a specialized sensing apparatus for diverse environmental stimuli that undergo relatively simple processing before a response is developed.

Because of the directionality and sparseness of links, TFs control the expression of only a limited number of genes in the current version of the TR network. The set of genes regulated directly or indirectly by a given TF forms an *origon* (an example is shown in Figure 1). This is a generalization of the earlier concept of regulatory subnetworks originating only at the input layer (Balázsi *et al*, 2005). By contrast, here we allow origons to originate at either input or transit TFs, because any TF can be affected by intra- or extracellular signal(s) (Martinez-Antonio *et al*, 2006) and relay the perturbation to target genes directly or indirectly. Thus, the number of origons is equal to the number of TFs in the network, and we will refer to the resulting 45 origons by the name of the TF at which they originate (see the Materials and methods).

## Origons significantly affected by growth arrest

Having assembled a large-scale TR network of *M. tuberculosis*, we set out to identify transcriptional subnetworks affected by various conditions. The reason for shifting focus from individual genes to subnetworks is that particular TFs can mediate the up- or downregulation of downstream target

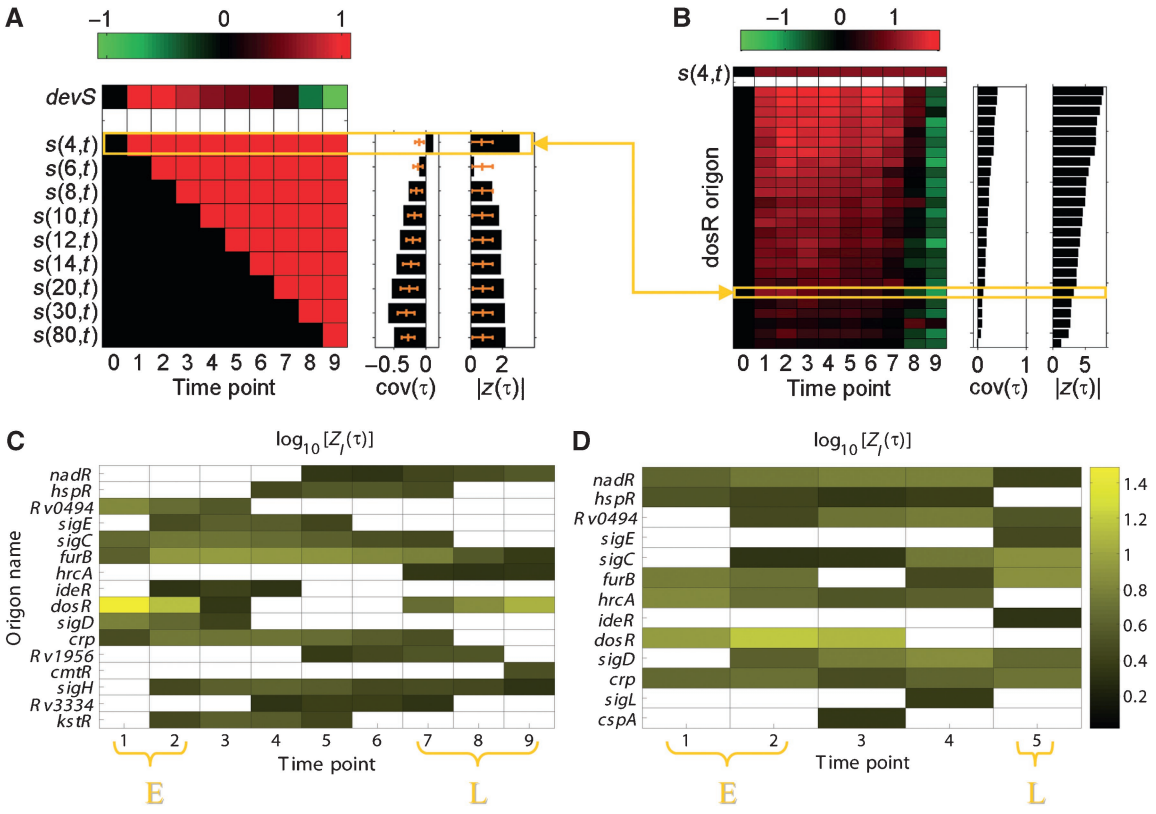


**Figure 2** Responsiveness of genes and origons. (**A**) The gene expression profile of the gene *devS* (top row, left panel) combined with each of nine time-shifted step functions (bottom rows, left panel) give the normalized cross-covariance (middle panel), and then the responsiveness $|z(\tau)|$ (right panel) of *devS* at each of the nine hypoxia time points starting with day 4. The orange error bars indicate averages and standard deviations over all *M. tuberculosis* genes. (**B**) Similar to (A), except the cross-covariance and responsiveness are calculated by combining a single step function $s(4, t)$ with the expression profile of each gene in the *dosR* origon. The yellow rectangles indicate identical values of $\text{cov}(\tau)$ and $|z(\tau)|$. (**C**) $Z_I(\tau)$ scores of significantly responsive origons during growth arrest in hypoxia (time points correspond to 4, 6, 8, 10, 12, 14, 20, 30, and 80 days). (**D**) $Z_I(\tau)$ scores of significantly responsive origons during aerated growth (time points correspond to days 6, 8, 14, 24, and 60). Eleven origons (*nadR*, *hspR*, *Rv0494*, *sigE*, *sigC*, *furB*, *hrcA*, *ideR*, *dosR*, *sigD*, and *crp*) responded significantly in both time courses. E and L denote the time points of peak response for early and late origons, respectively. Since a step function can only jump at time point 1 or later, time point 0 (day 0) is excluded from panels (C) and (D).

genes by post-translational modification while maintaining relatively constant mRNA expression levels. Traditional approaches focusing only on individual genes with significantly altered mRNA expression could miss such TFs (Ideker *et al*, 2002).

We developed a new method, *NetReSFun* (Network Response to Step Functions), which takes a network and time course data as inputs, and generates a list of significantly affected subnetworks for each time point as output. *NetReSFun* is the extension of an earlier approach (Balázsi *et al*, 2005), with a new scope and modified methodology (see the Materials and methods). We have tested *NetResFun* on random data, and showed that it can reliably detect the time when a major expression change occurs in a group of genes, such as an origon (see the Supplementary information).

We identified significantly affected *M. tuberculosis* origons during hypoxia-induced growth arrest by feeding the newly assembled TR network and the recently published time course microarray data GSE8786 (Voskuil *et al*, 2004) into *NetReSFun*. Briefly, the program calculates scaled cross-covariances $\text{cov}_i(\tau)$ between the expression profile $x_i(t)$ of each gene $i$ and a set of step functions $s(\tau, t)$ that jump at subsequent time points $\tau$ of microarray data collection, e.g., $\tau \in \{4, 6, 8, 10, 12, 14, 20, 30, 80 \text{ days}\}$ in hypoxia (Figure 2A). Next, the responsiveness $|z_i(\tau)|$ of each gene at time point $\tau$ is determined as the $z$-score of $\text{cov}_i(\tau)$ when compared with $\text{cov}_j(\tau)$ for all other genes (Figure 2A). Finally, the program calculates the responsiveness $Z_I(\tau)$ of each origon $I$ as the $z$-score of the average $<|z_i(\tau)|>_I$ over all genes in the origon (Figure 2B), when compared to the average $<|z_i(\tau)|>_R$ of the same number of genes chosen randomly from the network (see the Materials and methods). The output of *NetReSFun* consists of origons with $Z_I(\tau) > 2$, considered 'significantly responsive' at time point $\tau$. Importantly, the responsiveness $Z_I(\tau)$ of origon $I$ peaks at times when many genes within the origon have a large expression change (Figure 2C and D). Therefore, the times $\tau$ when $Z_I(\tau)$ peaks occur can be used to classify origons as early or late responders.

We classified significantly responding origons as 'early', 'intermediate' or 'late' based on the peak in their responsiveness $Z_I(\tau)$ over the time course (Figure 2C). For example, the *dosR* origon was most responsive at day 4, as nearly all *dosR*-controlled genes changed their expression at this time point (Figure 2B and 3A). *Rv0494* and *sigD* were also early origons, with a $Z_I(\tau)$ peak on or before day 6. Most of the significantly responsive origons peaked between days 8 and 14. These intermediate origons included *furB/zur*, *crp*, *sigH*, *kstR*, and *sigE-mprA*. Finally, late origons such as *nadR*, *Rv1956*, and

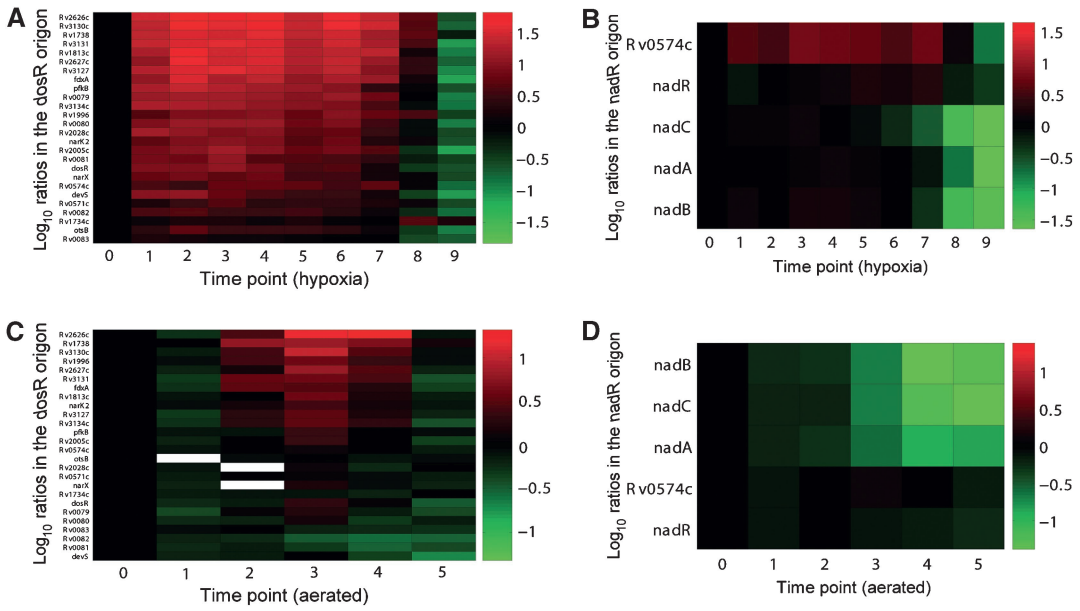


**Figure 3** Gene expression profiles in two *M. tuberculosis* origons affected early and late during hypoxia and stationary phase. The $\log_{10}$ ratios of all genes are shown for (**A**) the *dosR* origon during transition to dormancy in hypoxia, (**B**) the *nadR* origon during transition to dormancy in hypoxia, (**C**) the *dosR* origon during transition to stationary phase following aerated growth (white boxes indicate missing data) and (**D**) the *nadR* origon during transition to stationary phase following aerated growth.

*hrcA* were most responsive on or after day 20 (Figure 2C). Interestingly, the *dosR* origon had a second prominent $Z_I(\tau)$ peak at day 80, corresponding to a gene expression change opposite to day 4 (Figure 2B).

We performed a similar analysis for the time course microarray data collected by the same authors at days 0, 6, 8, 14, 24, and 60 in aerated cultures (Voskuil *et al*, 2004). Surprisingly, 11 of the origons responsive in hypoxia were also significantly responsive during transition to stationary phase (Figure 2D). We found that *dosR* was again the most prominently responding early origon, but it remained significant longer than in hypoxia (until day 14), presumbly because aerated cultures reach stationary phase later (day 20) than hypoxic cultures stop growing (day 10) (Voskuil *et al*, 2004), and prompted us to classify origons with a $Z_I(\tau)$ peak on or before day 8 as 'early' in the aerated time course. In addition to *dosR*, other early origons during aerated growth were *hrcA* and *hspR*. The origons *sigD*, *nadR*, and *Rv0494* were most prominently responsive at intermediate time points (days 14 and 24) (Figure 2D), whereas the origons *sigC* and *furB* had a $Z_I(\tau)$ peak on day 60. This indicates that, although the two types of growth arrest elicit response from the same origons, the temporal sequence of these responses is not always identical.

The most consistent early responder is the *dosR* origon (Figure 3A and C), which seems to be upregulated immediately before the bacteria stop growing in both time courses. By contrast, the origons *sigD*, *hrcA*, and *Rv0494* respond early in only one of the time courses, raising the possibility that they are condition-dependent initiators of growth arrest along with *dosR*. Finally, the origons *nadR* (Figure 3B and D), *sigE*, *sigC*, and *furB* peak consistently after *dosR* in both time courses (Figure 2C and D), suggesting that they orchestrate the maintenance (rather than the initiation) of dormancy. It will be important to experimentally test how inhibiting early versus late TFs affects the transition to dormancy.

In particular, the condition-dependent activation of other, alternate early origons in addition to *dosR* might explain the controversy between the early upregulation of the *dosR* regulon during growth arrest *in vitro* (Park *et al*, 2003) and *in vivo* (Shi *et al*, 2003) with the ability of a *dosR* deletion mutant to stop growing in hypoxic cultures and in mice weeks into the time course (Rustad *et al*, 2008). Considering the likelihood of alternate origon partners joining *dosR* to initiate dormancy in a condition-dependent manner, the ability of appropriate multiple deletion mutants (including *dosR*, *sigD*, *hrcA*, and *Rv0494*) to prevent growth arrest should be tested experimentally. Also, the apparent contradiction between the early hyper-virulence and fast growth of the *dosR* deletion mutant (Parish *et al*, 2003) and its unaltered dormancy after weeks of culture (Rustad *et al*, 2008) could be resolved by late origons governing growth arrest regardless of *dosR* status.

We performed additional control analyses to test the sensitivity of these results to random network rewiring and node removal. Specifically, we used *NetReSFun* to detect significantly responding regulons instead of origons, and we performed the same analysis on a higher confidence (literature-based) network. All these tests supported the robustness of our findings (see the Supplementary information).

## Conclusions

In summary, we have assembled the largest *M. tuberculosis* TR network available to date, and analyzed its topology, comparing it with two other large-scale TR networks. We have developed a novel method to unravel the temporal network response to a cellular program (growth arrest), and identified early, intermediate, and late origons based on their peak responsiveness during the time course. We found that the sets of TFs governing temporal network response to growth arrest

in two different conditions (hypoxia and stationary phase) were highly similar. As growth arrest is key to *M. tuberculosis* virulence, these regulators can be regarded as potential drug targets.

The present work has several limitations. First, the network-level analysis presented here would benefit from microarray data collected more frequently during the transition of *M. tuberculosis* into dormancy. The lower the number of samples, the higher the chance of observing high covariance values by pure chance. Second, time course data obtained under additional growth-arresting conditions, such as NO treatment and nutrient starvation, are needed to confirm that the observed repertoire of transcriptional modules generally governs growth arrest. However, no other time course data on *M. tuberculosis* growth arrest with sufficient time points is currently available. Third, a more complete TR network would improve our analysis significantly. The current version of the *M. tuberculosis* TR network contains only 45 of the 194 TFs listed in TubercuList. A systematic effort to identify the genes directly regulated by each of the 149 TFs is necessary to obtain an unbiased TR network. Future studies would also benefit from including non-TF regulators of gene expression in the network, such as signaling kinases, the alarmone (p)ppGpp, small peptides, and so on. Fourth, the majority of TFs implicated in network response are feedback-regulated, implying that their expression dynamics during growth arrest needs to be studied at the single cell level to better understand their role in adaptation and cell decision-making (Maamar *et al*, 2007; Sureka *et al*, 2008).

Despite the limitations mentioned above, our analysis defines an early, transient involvement of the *dosR* origon (Rustad *et al*, 2008), along with origons *sigD*, *hrcA*, and *Rv0494* in a condition-dependent manner during growth arrest. We also observed that the response of the origons *nadR*, *sigE*, *sigC*, and *furB* consistently replace *dosR* late in the time course, independent of the growth arrest conditions. This is in agreement with the proposition that the hypoxic response is maintained by genes that are not *dosR*-regulated (Rustad *et al*, 2008). However, our results also indicate that these 'later' origons are associated not specifically with hypoxia, but rather with the growth arrest *per se*, largely independent of the initiating stimulus.

Combining time course microarray data and large-scale gene regulatory networks might provide new means to dissect the cellular response to environmental changes at the network level. Such analyses should provide important novel insights into microbial biology and will likely suggest new drug targets.

# Materials and methods

## Assembly of the large-scale *M. tuberculosis* TR network

The TR network used in this study was assembled in several steps as follows. First, we created a gene regulatory network consisting of 222 links among 216 genes based on MtbRegList (Jacques *et al*, 2005), a database that lists the binding sites of 21 TFs and sigma factors. Next, we added to this network 159 links among 164 genes, based on recent studies on the transcriptional regulatory activity of *mprA*, *dosR*, *Rv1395*, *Rv2358*, *furB*, *Rv0967*, *kstR*, *pknH*, *embR*, *trcR*, and *crp* (Zahrt and Deretic, 2001; Park *et al*, 2003; Kendall *et al*, 2004, 2007; Bai *et al*, 2005; Canneva *et al*, 2005; Haydel and Clark-Curtiss, 2006; Sharma *et al*, 2006; Liu *et al*, 2007). We also downloaded and included an *M. tuberculosis* TR network (223 links among 201 genes) inferred from gene orthology with 29 *E. coli* TFs and their targets (Babu *et al*, 2006). Finally, we completed the network based on the list of *M. tuberculosis* operons (Roback *et al*, 2007), assuming that if a TF regulates a gene within an operon, it also regulates all other gene members of the operon. Following a similar procedure, we have also assembled a separate, purely literature-derived network, with 581 links among 518 genes that should have higher confidence than those in the full network.

The full *M. tuberculosis* TR network is available for download as Supplementary Table S1. This file contains regulator and target gene pairs identified by their GenBank IDs, their Rv numbers and traditional names whenever available. The last two columns provide information about the source of each regulatory interaction with respect to the literature and gene orthology, respectively. For example, the numbers 0, 1 and 2 in the last column indicate whether a link is not orthology-based (0), is from the original orthology-based network (1) or has been inferred by operon-based extension of the original orthology-based network (2). We used the software Pajek (Batagelj and Brandes, 2005) for network visualization.

## Naming of genes and origons

Throughout this paper, we used gene names obtained from FTGPRED (http://www.imtech.res.in/raghava/ftgpred/ANNOTATION/), TubercuList (http://genolist.pasteur.fr/TubercuList/) and the recent literature whenever possible. When the gene name was unknown, we used the Rv number instead.

We mapped origons as subtrees reachable from 45 of the 47 TFs. Two TFs (*Rv0144* and *Rv3744*) regulate no other genes except themselves, and therefore were not considered as origons. Origons were named based on the TF at which they originate. For the feedback loops involving more than one TF (*Rv2358-furB*, *Rv1404-Rv1931c*, and *mprA-sigE*), we chose the TF with more target genes to name the corresponding origons *furB*, *Rv1931c*, and *sigE*, respectively.

## Origons significantly affected during transition to non-replicative persistence

The GSE8786 microarray dataset that we used (Voskuil *et al*, 2004) consisted of two time series: aerated growth and growth arrest in hypoxia. For each gene, we used its expression at day 0 in aerated growth as a control intensity value. We determined the $\log_{10}$ ratios of expression, dividing the intensity on both the hypoxia and aerated growth arrays by this control intensity value.

The tool *NetReSFun* (available for download as Supplementary information) measures the effect of various stages of growth arrest on each gene's expression by the scaled covariance $\mathrm{cov}_i(\tau)$ between the expression ($\log_{10}$ ratio) profile $x_i(t)$ of gene $i$ and a step function $s(\tau, t)$ that jumps at time point $\tau$:

$$\mathrm{cov}_i(\tau) = \langle [x_i(t) - \bar{x}_i][s(\tau,t) - \bar{s}(\tau)] \rangle / \sigma[s(\tau,t)] \qquad (1)$$

where the brackets indicate averaging over genes, the horizontal bar indicates averaging over time, and the letter $\sigma$ denotes standard deviation. Thus, the covariance $\mathrm{cov}_i(\tau)$ is scaled by the standard deviation of the step function $s(\tau, t)$:

$$s(\tau,t) = \begin{cases} 0, t < \tau \\ 1, t \geqslant \tau \end{cases} \qquad (2)$$

ensuring that only the variance of gene expression contributes to gene responsiveness at time $\tau$, defined as the $z$-score

$$z_i(\tau) = \frac{\mathrm{cov}_i(\tau) - \langle \mathrm{cov}_i(\tau) \rangle}{\sigma[\mathrm{cov}_i(\tau)]} \qquad (3)$$

Similarly, the responsiveness of origon $I$ at time point $\tau$ was defined as the $z$-score of $z$-scores, or 'double $Z$-score':

$$Z_I(\tau) = \frac{\langle |z_i(\tau)| \rangle_I - \langle |z_i(\tau)| \rangle_R}{\sigma[|z_i(\tau)|]_R} \qquad (4)$$

The subscripts *I* and *R* indicate averaging over all genes in the origon, and over the same number of genes chosen randomly from the network.

Using the scaled covariance to determine gene affectedness offers the advantage of simultaneously measuring the amplitude of gene expression changes as well as their similarity to a pre-defined signal. More widely used measures, such as the cross-correlation coefficient would only measure the similarity of the expression profile to the external signal, regardless of the amplitude of gene expression changes, ignoring an important characteristic of gene response.

### Supplementary information

Supplementary information is available at the *Molecular Systems Biology* website (www.nature.com/msb).

## Acknowledgements

We thank Bhaskar Dutta for discussions. This study was supported by MD Anderson start-up funds (GB) and by the NIH (MLG).

## Conflict of interest

The authors declare that they have no conflict of interest.

## References

Albert R (2005) Scale-free networks in cell biology. *J Cell Sci* **118:** 4947–4957

Babu M, Teichmann S, Aravind L (2006) Evolutionary dynamics of prokaryotic transcriptional regulatory networks. *J Mol Biol* **358:** 614–633

Bai G, McCue LA, McDonough KA (2005) Characterization of *Mycobacterium tuberculosis* Rv3676 (CRPMt), a cyclic AMP receptor protein-like DNA binding protein. *J Bacteriol* **187:** 7795–7804

Balaji S, Babu MM, Iyer LM, Luscombe NM, Aravind L (2006) Comprehensive analysis of combinatorial regulation using the transcriptional regulatory network of yeast. *J Mol Biol* **360:** 213–227

Balázsi G, Barabasi AL, Oltvai ZN (2005) Topological units of environmental signal processing in the transcriptional regulatory network of *Escherichia coli*. *Proc Natl Acad Sci USA* **102:** 7841–7846

Batagelj V, Brandes U (2005) Efficient generation of large random networks. *Phys Rev E Stat Nonlin Soft Matter Phys* **71:** 036113

Betts JC, Lukey PT, Robb LC, McAdam RA, Duncan K (2002) Evaluation of a nutrient starvation model of *Mycobacterium tuberculosis* persistence by gene and protein expression profiling. *Mol Microbiol* **43:** 717–731

Canneva F, Branzoni M, Riccardi G, Provvedi R, Milano A (2005) Rv2358 and FurB: two transcriptional regulators from *Mycobacterium tuberculosis* which respond to zinc. *J Bacteriol* **187:** 5837–5840

Ernst J, Beg QK, Kay KA, Balazsi G, Oltvai ZN, Bar-Joseph Z (2008) A semi-supervised method for predicting transcription factor–gene interactions in *Escherichia coli*. *PLoS Comput Biol* **4:** e1000044

Farkas IJ, Wu C, Chennubhotla C, Bahar I, Oltvai ZN (2006) Topological basis of signal integration in the transcriptional-regulatory network of the yeast, *Saccharomyces cerevisiae*. *BMC Bioinformatics* **7:** 478

Fisher MA, Plikaytis BB, Shinnick TM (2002) Microarray analysis of the *Mycobacterium tuberculosis* transcriptional response to the acidic conditions found in phagosomes. *J Bacteriol* **184:** 4025–4032

Guelzim N, Bottani S, Bourgine P, Kepes F (2002) Topological and causal structure of the yeast transcriptional regulatory network. *Nat Genet* **31:** 60–63

Hampshire T, Soneji S, Bacon J, James BW, Hinds J, Laing K, Stabler RA, Marsh PD, Butcher PD (2004) Stationary phase gene expression of *Mycobacterium tuberculosis* following a progressive nutrient depletion: a model for persistent organisms? *Tuberculosis (Edinb)* **84:** 228–238

Harbison CT, Gordon DB, Lee TI, Rinaldi NJ, Macisaac KD, Danford TW, Hannett NM, Tagne JB, Reynolds DB, Yoo J, Jennings EG, Zeitlinger J, Pokholok DK, Kellis M, Rolfe PA, Takusagawa KT, Lander ES, Gifford DK, Fraenkel E, Young RA (2004) Transcriptional regulatory code of a eukaryotic genome. *Nature* **431:** 99–104

Haydel SE, Clark-Curtiss JE (2006) The *Mycobacterium tuberculosis* TrcR response regulator represses transcription of the intracellularly expressed Rv1057 gene, encoding a seven-bladed beta-propeller. *J Bacteriol* **188:** 150–159

Ideker T, Ozier O, Schwikowski B, Siegel AF (2002) Discovering regulatory and signalling circuits in molecular interaction networks. *Bioinformatics* **18** (Suppl 1)**:** S233–S240

Ihmels J, Bergmann S, Barkai N (2004) Defining transcription modules using large-scale gene expression data. *Bioinformatics* **20:** 1993–2003

Isaacs FJ, Dwyer DJ, Collins JJ (2006) RNA synthetic biology. *Nat Biotechnol* **24:** 545–554

Jacob F, Perrin D, Sanchez C, Monod J (1960) [Operon: a group of genes with the expression coordinated by an operator.]. *C R Hebd Seances Acad Sci* **250:** 1727–1729

Jacques PE, Gervais AL, Cantin M, Lucier JF, Dallaire G, Drouin G, Gaudreau L, Goulet J, Brzezinski R (2005) MtbRegList, a database dedicated to the analysis of transcriptional regulation in *Mycobacterium tuberculosis*. *Bioinformatics* **21:** 2563–2565

Kendall SL, Movahedzadeh F, Rison SC, Wernisch L, Parish T, Duncan K, Betts JC, Stoker NG (2004) The *Mycobacterium tuberculosis* dosRS two-component system is induced by multiple stresses. *Tuberculosis (Edinb)* **84:** 247–255

Kendall SL, Withers M, Soffair CN, Moreland NJ, Gurcha S, Sidders B, Frita R, Ten Bokum A, Besra GS, Lott JS, Stoker NG (2007) A highly conserved transcriptional repressor controls a large regulon involved in lipid degradation in *Mycobacterium smegmatis* and *Mycobacterium tuberculosis*. *Mol Microbiol* **65:** 684–699

Khanin R, Wit E (2006) How scale-free are biological networks. *J Comput Biol* **13:** 810–818

Li Y, Altman S (2004) Polarity effects in the lactose operon of *Escherichia coli*. *J Mol Biol* **339:** 31–39

Liu T, Ramesh A, Ma Z, Ward SK, Zhang L, George GN, Talaat AM, Sacchettini JC, Giedroc DP (2007) CsoR is a novel *Mycobacterium tuberculosis* copper-sensing transcriptional regulator. *Nat Chem Biol* **3:** 60–68

Maamar H, Raj A, Dubnau D (2007) Noise in gene expression determines cell fate in *Bacillus subtilis*. *Science* **317:** 526–529

Manganelli R, Voskuil MI, Schoolnik GK, Smith I (2001) The *Mycobacterium tuberculosis* ECF sigma factor sigmaE: role in global gene expression and survival in macrophages. *Mol Microbiol* **41:** 423–437

Martinez-Antonio A, Janga SC, Salgado H, Collado-Vides J (2006) Internal-sensing machinery directs the activity of the regulatory network in *Escherichia coli*. *Trends Microbiol* **14:** 22–27

Murphy DJ, Brown JR (2007) Identification of gene targets against dormant phase *Mycobacterium tuberculosis* infections. *BMC Infect Dis* **7:** 84

Nudler E, Gottesman ME (2002) Transcription termination and anti-termination in *E. coli*. *Genes Cells* **7:** 755–768

Parish T, Smith DA, Kendall S, Casali N, Bancroft GJ, Stoker NG (2003) Deletion of two-component regulatory systems increases the virulence of *Mycobacterium tuberculosis*. *Infect Immun* **71:** 1134–1140

Park HD, Guinn KM, Harrell MI, Liao R, Voskuil MI, Tompa M, Schoolnik GK, Sherman DR (2003) Rv3133c/dosR is a transcription factor that mediates the hypoxic response of *Mycobacterium tuberculosis*. *Mol Microbiol* **48:** 833–843

Pfleger BF, Pitera DJ, Smolke CD, Keasling JD (2006) Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. *Nat Biotechnol* **24:** 1027–1032

Price MN, Dehal PS, Arkin AP (2007) Orthologous transcription factors in bacteria have different functions and regulate different genes. *PLoS Comput Biol* **3:** 1739–1750

Roback P, Beard J, Baumann D, Gille C, Henry K, Krohn S, Wiste H, Voskuil MI, Rainville C, Rutherford R (2007) A predicted operon map for *Mycobacterium tuberculosis*. *Nucleic Acids Res* **35:** 5085–5095

Rustad TR, Harrell MI, Liao R, Sherman DR (2008) The enduring hypoxic response of *Mycobacterium tuberculosis*. *PLoS ONE* **3:** e1502

Salgado H, Santos-Zavaleta A, Gama-Castro S, Peralta-Gil M, Penaloza-Spinola MI, Martinez-Antonio A, Karp PD, Collado-Vides J (2006) The comprehensive updated regulatory network of *Escherichia coli* K-12. *BMC Bioinformatics* **7:** 5

Sharma K, Gupta M, Pathak M, Gupta N, Koul A, Sarangi S, Baweja R, Singh Y (2006) Transcriptional control of the mycobacterial embCAB operon by PknH through a regulatory protein, EmbR, *in vivo*. *J Bacteriol* **188:** 2936–2944

Sherman DR, Voskuil M, Schnappinger D, Liao R, Harrell MI, Schoolnik GK (2001) Regulation of the *Mycobacterium tuberculosis* hypoxic response gene encoding alpha-crystallin. *Proc Natl Acad Sci USA* **98:** 7534–7539

Shi L, Jung YJ, Tyagi S, Gennaro ML, North RJ (2003) Expression of Th1-mediated immunity in mouse lungs induces a *Mycobacterium tuberculosis* transcription pattern characteristic of nonreplicating persistence. *Proc Natl Acad Sci USA* **100:** 241–246

Sureka K, Ghosh B, Dasgupta A, Basu J, Kundu M, Bose I (2008) Positive feedback and noise activate the stringent response regulator rel in mycobacteria. *PLoS ONE* **3:** e1771

Thieffry D, Huerta AM, Perez-Rueda E, Collado-Vides J (1998) From specific gene regulation to genomic networks: a global analysis of transcriptional regulation in *Escherichia coli*. *Bioessays* **20:** 433–440

Voskuil MI, Schnappinger D, Visconti KC, Harrell MI, Dolganov GM, Sherman DR, Schoolnik GK (2003) Inhibition of respiration by nitric oxide induces a *Mycobacterium tuberculosis* dormancy program. *J Exp Med* **198:** 705–713

Voskuil MI, Visconti KC, Schoolnik GK (2004) *Mycobacterium tuberculosis* gene expression during adaptation to stationary phase and low-oxygen dormancy. *Tuberculosis (Edinb)* **84:** 218–227

Wagner GP, Pavlicev M, Cheverud JM (2007) The road to modularity. *Nat Rev Genet* **8:** 921–931

Warner DF, Mizrahi V (2007) The survival kit of *Mycobacterium tuberculosis*. *Nat Med* **13:** 282–284

Wayne LG, Sohaskey CD (2001) Nonreplicating persistence of *Mycobacterium tuberculosis*. *Annu Rev Microbiol* **55:** 139–163

Zahrt TC, Deretic V (2001) *Mycobacterium tuberculosis* signal transduction system required for persistent infections. *Proc Natl Acad Sci USA* **98:** 12706–12711



 