**Title: Intricate genetic programs controlling dormancy in *Mycobacterium tuberculosis***

**Authors:**

Abrar A. Abidi1†

Eliza J. R. Peterson1†

Mario L. Arrieta-Ortiz1†

Amardeep Kaur1

Min Pan1

Vivek Srinivas1

Nitin S. Baliga1\*

†Co-first authors

\*Corresponding author

**Affiliations:**

1Institute for Systems Biology, 401 Terry Avenue North, Seattle, Washington 98109, USA

**Abstract:** *Mycobacterium tuberculosis* (MTB), responsible for the deadliest infectious disease worldwide, displays the remarkable ability to transition into and out of dormancy, a hallmark of the pathogen’s capacity to evade the immune system and opportunistically exploit immunocompromised individuals. Uncovering the gene regulatory programs that underlie the dramatic phenotypic shifts in MTB during disease latency and reactivation has posed an extraordinary challenge. We developed a novel experimental system to exquisitely control dissolved oxygen levels in MTB cultures in order to capture the chain of transcriptional events that unfold as MTB transitions into and out of hypoxia-induced dormancy. Using a comprehensive, genome-wide transcription factor binding location map, we identified regulatory circuits that deterministically drive sequential transitions across six transcriptionally and functionally distinct states encompassing more than three-fifths of the MTB genome. The architecture of the genetic programs explains the transcriptional dynamics underlying synchronous entry of cells into a dormant state that is primed to infect the host upon encountering favorable conditions.

**One Sentence Summary:** High-resolution transcriptional time-course reveals six-state genetic program that enables MTB to enter and exit hypoxia-induced dormancy.

**Main Text:** *Mycobacterium tuberculosis* (MTB) kills more people than any other infectious agent, causing ~10 million new cases of active tuberculosis (TB) disease and 1.7 million deaths each year1. TB remains a major human public health burden, in large part due to the sizeable reservoir of latently infected individuals, who may relapse into active disease decades after acquiring the infection. MTB can persist in a stable, non-replicative (termed dormant) state within the host for months or years without symptoms, and then revive to initiate the production of lesions and active TB. Moreover, dormant cells may be responsible for the slow treatment response of active TB patients. Elucidation of the factors that affect treatment outcome, latency and activation requires a better characterization of functional states adopted by the pathogen during progression of the disease, as well as a mechanistic understanding of the genetic program that orchestrates transitions between these states. Hypoxia, an environmental stress encountered by MTB within granulomas2, is sufficient to shift the pathogen into a defined non-growing survival form, which can be reversed upon aeration of the culture3. Therefore, hypoxia has been leveraged as an *in vitro* approximation to study MTB dormancy and the underlying genetic programs. However, previous transcriptional analyses under *in vitro* hypoxic conditions (via Wayne model in which MTB cultures are sealed and gradually depleted of oxgyen4,5 or the defined hypoxia model in which nitrogen gas is flowed into the headspace to rapidly deplete oxygen6,7) were limited to either static snapshots, or low-resolution time-course studies8-10. Moreover, deletion of previously identified transcriptional regulators thought crucial to hypoxia-induced dormancy (i.e. Δ*dosR*, Δ*sigE*Δ*sigH*), conferred only mild growth defects under hypoxic conditions8,11,12, suggesting a genetic circuit architecture that has evolved to withstand environmental and genetic perturbations. Here, we developed a novel experimental platform to characterize MTB’s response to changing oxygen (O2) levels in considerable more depth. We reveal detailed transcriptional dynamics and coordinated regulatory circuits that enable the pathogen’s transition into and out of hypoxia-induced dormancy.

To detail the genetic programs underlying hypoxia-induced dormancy in MTB, we needed to obtain accurate dynamic measurements of genome-wide expression, over an O2 gradient. Previous experimental models to study O2 tension and growth arrest in MTB were not suitable for the accuracy and resolution of measurements needed. In particular, the Wayne model has issues with reproducibility (ref?) and the defined hypoxia model depletes O2 very quickly, thereby hindering high-resolution sampling during critical transition periods. Moreover, neither model has been performed with real-time monitoring of O2 levels to accurately relate the transcriptional state of MTB with a precise O2 measurement. Therefore, we designed a new programmable multiplexed reactor system, the controlled O2 model, to precisely manipulate and remotely monitor O2 levels within the growth medium–even during sampling (**Fig. 1**). The exquisite control engineered into the system enabled high-resolution sampling across a time-course and O2 gradient, with minimal disturbance to the bacteria and high reproducibility across culture replicates and experiments (**Fig. S1**). Briefly, air and nitrogen (N2) gas lines were connected to separate mass flow controllers, which allowed for programmable gradients of gas mixtures to be streamed into the headspace of spinner flasks containing MTB in media. Moreover, we used O2 sensor spots and fiber optic technology to non-invasively measure the dissolved O2 content of the cultures. With the controlled O2 model, we performed a time-course experiment, which involved a steady depletion of dissolved oxygen (DO) over 2 days from full aeration (~80% DO) to hypoxia (0% DO). This steady depletion was achieved by programming the mass flow controllers to achieve the desired mixture of air and N2. The cultures were maintained in hypoxia for 2 days by streaming only N2, then reaerated over 1 day by a programmed increase in air flow (**Fig 1**). Over the time-course, we harvested samples in triplicates, one each from three independent reactors, via sampling ports that prevented aeration of the culture. We sampled at high frequency during the period when cultures transitioned from 10% to 0% DO, as well as from 0% to 10% DO and at lower but regular frequency across the remaining 120-hour experiment. Samples were flash frozen in liquid N2 and later processed for gene expression profiling by RNA-sequencing.

Over the course of the experiment, nearly 64% of all genes in the MTB genome were significantly differentially expressed (2,582 genes with adjusted *p*-value < 0.05 and estimated absolute log2 fold-change >1). The number of differentially expressed genes from our controlled O2 system is an order of magnitude greater than the differentially expressed genes from earlier microarray studies using the Wayne9 (411 genes) or defined hypoxia8 (279 genes) models. Nevertheless, there is significant overlap across gene sets between the three models of hypoxia-induced dormancy (**Table. SX**). Interestingly, the controlled O2 model also recapitulated differential expression observed from intracellular MTB (enrichment test *p*-value=4.29 x 10-30) (ref), whereas the defined model hypoxia series (ref) did not (*p*-value=0.1) (**Table SX**). These findings highlight the capability of our model to comprehensively capture MTB’s transcriptional program for entry into dormancy, even within host cells.

To further characterize the MTB transcriptional states over the time course and oxygen gradient, we applied dimensionality reduction techniques that allowed us to define tightly clustered samples (*see methods*). The six identified clusters are shown in a two-dimension t-distributed stochastic neighbor embedding (tSNE) plot (**Fig. 1**). Each cluster represents a distinct transcriptional state and was classified with sets of non-overlapping differentially expressed genes (see methods): normoxia (81 genes), transition (446 genes), stage Ia hypoxia (328 genes), stage Ib hypoxia (320 genes), stage II hypoxia (978 genes), and resuscitation (429 genes) (**Table SX**). The average expression profiles for the gene sets reveal that the states transition from one to another and that transitions are oxygen (e.g. stage II into resuscitation occurred upon re-introducing air into the culture) and time dependent (e.g. stage Ia/b into stage II occured ~40 h after the culture reached 0% DO) (**Fig. X** and **Fig. SX**). As such, the six states were also defined with oxygen and time intervals (**Table SS**), with the exception of XX-XX h, where there was oscillation between stage Ia and stage II states as the culture went below ~3% DO (**Fig SX**). While this intriguing “flicker” behavior could be experimental noise, these anomalous time points (measured roughly one hour apart) clearly cluster with stage Ia and stage II states. Such oscillatory mechanisms could involve complex regulatory dynamics comprising feedforward loop motifs, which we describe later on.

Genes associated to the transition state (DO between 43% and 4%) were enriched for growth-related functions including amino acid metabolism, oxidative phosphorylation and translation (**Fig. X**). In stage Ia hypoxia, ATP synthase and genes involved in electron transport chain and lipid metabolism were highly enriched and expressed, even more so than in normoxia. Furthermore, these metabolic genes were then significantly down regulated during stage II hypoxia (need **Fig SX**). This indicates that stage Ia is a transient but active metabolic state that may exist for MTB to metabolically prepare itself for what could be a more long-term and metabolically quiescent stage II hypoxia. Stage Ib genes were enriched in stress response genes, indicating the bacteria is sensing and adapting to the anaerobic environment. In stage II hypoxia, genes essential for MTB to infiltrate host cells were induced (i.e. genes associated with mammalian cell entry category). Furthermore, genes for 32 proteins that belong to the proline-glutamic acid (PE) and proline-proline-glutamic acid (PPE) family, whose functions remain largely unknown13 were up-regulated. These PPE family proteins have been proposed to modulate the host’s immune response14, generate antigenic variation15 and were shown to be secreted by MTB’s Type VII secretion system16. Interestingly, genes encoding the components of the Type VII secretion system were rapidly activated as soon as MTB shifted from stage II hypoxia into resuscitation, minutes after air was introduced back into the culture. In addition, proteases, transposases and insertion sequences were enriched and activated in the resuscitation state. It is possible that stage II hypoxia not only engenders quiescence in MTB, but also sequesters a collection of PPE proteins in anticipation of resuscitation. Upon resuscitation, MTB may strategically avert immune recognition through antigenic heterogeneity (via Type VII secretion of PPE proteins), and simultaneously reorganize its genome to increase its chances for survival and transmission to a new host

Using the full expression dataset, we also analyzed changes in MTB metabolic pathways, as reconstructed by Kavvas *et al*17, along the hypoxia time series. The expression of many metabolic pathways reiterated the state transitions described above (**Fig XA-D**). For example, significant down-regulation of genes involved in oxidative phosphorylation during stage II (**Fig. XA**), confirming the loss of energy related pathways during late hypoxia. Furthermore, MTB’s dependency on alternative carbon sources was also observed in stage II, with significant up-regulation of genes related to cholesterol degradation and a simultaneous down-regulation of mycolic acid biosynthesis genes (**Fig XB-C**). Re-aeration of the MTB culture and entry into the resuscitation stage reversed these expression trends of late hypoxia. Additionally, we observed genes involved in mycobactin biosynthesis were significantly anti-correlated with decreasing % DO (R and P-vlaue?) (**Fig XD**). The increase in mycobactin, an iron chelator, is essential for MTB to access iron, particularly when it is competing with the host for the limiting metal18. Overall, MTB adaptation to hypoxia involves rewiring of several metabolic pathways, including those associated with intracellular MTB, suggesting an evolved response to stresses that typically co-exist within the host environment (e.g. hypoxia, starvation, iron limitation).

Further analysis of the six-state model across the time course and O2 gradient revealed distinct patterns of expression suggestive of intriguing and coordinated regulatory circuits. We explored two examples in greater detail. The first observation being that stage I and stage II hypoxia exhibit bistability: when one state is active, the other is largely inactive (**Fig. X**). Moreover, the switch from stage I to stage II hypoxia takes place ~40-hours after entering hypoxia. Such temporal expression dynamics is indicative of a delay element that is commonly associated with regulatory circuitry patterns known as feedforward loops (FFLs)19. To unbiasedly search for circuitry patterns, in particular certain classes of FFLs, we analyzed experimentally determined transcription factor (TF)-target interactions from ChIP-seq analysis of 154 of the ~217 TFs in MTB20 using FANMOD21 in the MotifNet Webserver (Yeger-lotem et al 2018). Indeed, MTB ChIP-seq network is significantly enriched with FFLs. We ran extensive permutation tests to confirm the likelihood (*p*-value < 0.05) of 7098 FFL instances emerging in a random network with the same number of nodes and edges. Interestingly, Rv0081 is the most frequent regulator at the top of the FFLs (22.5% of all detected instances). Rv0081 has been previously linked to MTB’s response to hypoxia22,23 and is itself a target gene of the well-characterized regulator of dormancy survival, DosR10,11,24,25. To evaluate the potential involvement of Rv0081-centered FFLs in the transcriptional changes observed during hypoxia, we evaluated the overlap of the set of differentially expressed genes in the Rv0081 KO MTB strain in hypoxia (Sun et al. 2018) and the genes associated to each of the six transcriptional states of the controlled O2 model. Surprisingly, the set of DEGs in the Rv0081 deletion strain is significantly enriched with stage Ia, stage Ib and resuscitation genes (**Table SX**). Moreover, most of the shared genes are down-regulated in the KO, suggesting that Rv0081 directly or indirectly activate those genes during hypoxia. Two important hypotheses are derived from these findings: i) Rv0081 plays a pivotal and early role in the adaptation of MTB to hypoxia. Thus, Rv0081 is at the top of the circuit controlling the cascade of events leading MTB adaptation to hypoxia . ii) Rv0081 behaves differently in hypoxia and aeration. *In hypoxic conditions, Rv0081 mainly exerts positive regulation on stage I genes (based on KO data and our time series). However, Rv0081 acts as repressor of some of the stage I genes in normoxia according to transcriptomics data of Rv0081over-expressing MTB strain36. In addition, we propose that Rv0081 exerts positive regulation on resuscitation genes during the brief period following resuscitation (before it falls toward its basal expression leve). Transcriptional data taken when Rv0081 is over-expressed in normal growth conditions endorses the idea that an oxygen-dependent conformational switch dictates Rv0081’s active targets in either condition (fig. S5) (18).*

Multiple instances of these FFLs interact between them to generate interlocking FFLs that together modulate a large number of downstream genes (331 genes) corresponding to stage II hypoxia (**Fig. S5**). Interlocking FFLs, as found in the identified circuit, predict pulses of TF expression upon activation. Clear pulses of TF expression were demonstrated in the time course expression data upon entry in hypoxia (**Fig. 3B**). Moreover, these TF pulses begin with Rv0081 expression and include all the TFs in our identified circuit. While the sequence of the TF cascade was not predicted, we inferred a temporal genetic program from a genome-wide TF binding location map that corresponds to MTB’s response to hypoxia.

#####NEXT TWO PARAGRAPHS WILL CHANGE ONCE WE HAVE DONE THE EGRIN ,DREM ANALYSES, ADDITIONAL MODELLING, ETC

The expression data-driven circuitry is refined in **Fig. 3C** and shows a time delay element, consisting of the TF cascade, embedded within an incoherent FFL (I-FFL). Only when Rv0047c is repressed by Rv1033c (at the end of the TF cascade) is the I-FFL de-repressed and Rv0081 induces the expression of stage II genes. We find that this motif determines a delay following signal activation, ultimately toggling the cells from stage I into stage II hypoxia. The time-scale of the delay generated by the regulatory circuit, about 40-hours, is consistent with delayed translation observed in slow-growing MTB in response to nitric oxide26. Thus, the time-delay could allow for protein changes that are necessary for the switch to stage II hypoxia. Alternatively, the time delay could ensure that stage II hypoxia is only activated once the proper conditions are sensed for an extended and sustained period of time.

The second expression pattern we explored in detail was observed when DO levels dropped below ~3%. Upon entering these low O2 conditions, the MTB transcriptome oscillated between two states. As seen in the tSNE plot (**Fig. 2B**) and expression profiles (**Fig. 2C**), stage II genes were expressed (T8 and T9), then stage I genes (T10), then back to stage II (T11), before ultimately committing to stage I hypoxia (T12-T18). These fluctuations between stage I and stage II, measured roughly one hour apart, could emerge from oscillatory mechanisms involving negative feedback of the proposed I-FFL described above (**Fig. 3C**)27-29. Interestingly, the peak height between the first and second wave of stage II genes has roughly equal relative expression (**Fig. 4A)**, suggesting a mechanism for detection of fold change based on the I-FFL, as described by Goentoro and colleagues30. We fit our time-course expression data to their theoretical model of the circuit and discovered that the parameters fall within a regime where the I-FFL behaves as a fold-change detector (**Fig. 4B**). We hypothesize that nearing hypoxia, the absolute levels of TFs may vary from cell to cell, but by sensing relative changes in gene expression, this circuit can provide robustness and synchronize the O2 response across all cells of the population.

#####

While we have described in detail the circuit characteristics underlying two major expression patterns in this four state model, our data also suggest that the I-FFL does not act in isolation and is interlocked with other regulatory circuits to coordinate and distribute the appropriate responses to distinct O2 levels. In addition to the I-FFL, we also see evidence of mutual inhibition of Rv0081 regulatory targets (**Fig. S6**), single-input-modules generating just-in-time expression patterns during stage I hypoxia (**Fig. S7**), and parallel coherent FFLs that control the rapid activation and de-activation of stage II hypoxia genes (**Fig. S8**). Together, these intricate genetic circuits control the temporal gene expression programs and four-state model described in this paper.

Understanding the detailed circuit architecture that coordinates MTB’s response to stimuli is key to formulate rational strategies to intervene and improve the outcome of TB disease. The high-resolution gene expression profiling proved critical to reveal the detailed properties of the genetic program that temporally coordinates the expression of nearly three quarter of genes in the MTB genome to drive the pathogen through four distinct states during the entry and exit from hypoxia-induced dormancy. Importantly, two of the four states were transient (stage I hypoxia and resuscitation) and were undetectable in the two previous hypoxia models (summarized in **Fig X**). Interestingly, one of these transient states (stage I hypoxia), which existed for ~40 hours, is metabolically active and appears preparative for transition into stage II hypoxia. Consistent with the predicted active metabolism of stage I hypoxia, we observed MTB cultures continued to grow for 40 hours after DO dropped to 0% (**Fig S1**). Given this, we suspect that isoniazid and other drugs active against growing cells would be effective in killing MTB in stage I hypoxia, but not in stage II hypoxia Therefore, blocking the TF cascade that comprises the time-delay element in the I-FFL could potentially push (or keep) MTB in stage I hypoxia and susceptible to TB treatment. Inhibiting TFs in this cascade could potentiate the activity of isoniazid by blocking the circuit that expresses stage II hypoxia genes. However, several studies have shown that TF perturbation can still result in the expression of target genes via cooperative TFs that recognize similar DNA binding sites31-34. In contrast, some transcriptional programs lack redundancy in TF activity and are therefore ‘fragile’ to perturbation. We measured the fragility score of all TF-gene interactions from the ChIP-seq network using an approach devised by Dai *et al*35. All the TFs in the I-FFL and time-delay TF cascade are among those with top 20 fragility scores [**Fig. S9**]. This suggests the I-FFL regulatory circuit could be highly susceptible to TF perturbation and such perturbation could potentiate anti-tubercular drugs. Altogether, this study demonstrates how apprehending the intricate genetic programs underlying dormancy could pave the way for predictive and rational strategies to improve outcome of TB treatment.

**REFERENCES**

1 Murray, C. J. *et al.* Global, regional, and national incidence and mortality for HIV, tuberculosis, and malaria during 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* **384**, 1005-1070, doi:10.1016/S0140-6736(14)60844-8 (2014).

2 Tsai, M. C. *et al.* Characterization of the tuberculous granuloma in murine and human lungs: cellular composition and relative tissue oxygen tension. *Cell Microbiol* **8**, 218-232, doi:10.1111/j.1462-5822.2005.00612.x (2006).

3 Chao, M. C. & Rubin, E. J. Letting sleeping dos lie: does dormancy play a role in tuberculosis? *Annu Rev Microbiol* **64**, 293-311, doi:10.1146/annurev.micro.112408.134043 (2010).

4 Wayne, L. G. & Sohaskey, C. D. Nonreplicating persistence of mycobacterium tuberculosis. *Annu Rev Microbiol* **55**, 139-163, doi:10.1146/annurev.micro.55.1.139 (2001).

5 Wayne, L. G. & Hayes, L. G. An in vitro model for sequential study of shiftdown of Mycobacterium tuberculosis through two stages of nonreplicating persistence. *Infect Immun* **64**, 2062-2069 (1996).

6 Kempner, W. Oxygen tension and the tubercle bacillus. *Am Rev Tubercul* **40**, 157-168 (1939).

7 Yuan, Y. *et al.* The 16-kDa alpha-crystallin (Acr) protein of Mycobacterium tuberculosis is required for growth in macrophages. *Proc Natl Acad Sci U S A* **95**, 9578-9583 (1998).

8 Rustad, T. R., Harrell, M. I., Liao, R. & Sherman, D. R. The enduring hypoxic response of Mycobacterium tuberculosis. *PLoS One* **3**, e1502, doi:10.1371/journal.pone.0001502 (2008).

9 Muttucumaru, D. G., Roberts, G., Hinds, J., Stabler, R. A. & Parish, T. Gene expression profile of Mycobacterium tuberculosis in a non-replicating state. *Tuberculosis (Edinb)* **84**, 239-246, doi:10.1016/j.tube.2003.12.006 (2004).

10 Sherman, D. R. *et al.* Regulation of the Mycobacterium tuberculosis hypoxic response gene encoding alpha -crystallin. *Proc Natl Acad Sci U S A* **98**, 7534-7539, doi:10.1073/pnas.121172498 (2001).

11 Boon, C. & Dick, T. Mycobacterium bovis BCG response regulator essential for hypoxic dormancy. *J Bacteriol* **184**, 6760-6767 (2002).

12 Rustad, T. R., Sherrid, A. M., Minch, K. J. & Sherman, D. R. Hypoxia: a window into Mycobacterium tuberculosis latency. *Cell Microbiol* **11**, 1151-1159, doi:10.1111/j.1462-5822.2009.01325.x (2009).

13 Bottai, D. & Brosch, R. Mycobacterial PE, PPE and ESX clusters: novel insights into the secretion of these most unusual protein families. *Mol Microbiol* **73**, 325-328, doi:10.1111/j.1365-2958.2009.06784.x (2009).

14 Tiwari, B. M., Kannan, N., Vemu, L. & Raghunand, T. R. The Mycobacterium tuberculosis PE proteins Rv0285 and Rv1386 modulate innate immunity and mediate bacillary survival in macrophages. *PLoS One* **7**, e51686, doi:10.1371/journal.pone.0051686 (2012).

15 Cole, S. T. *et al.* Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. *Nature* **393**, 537-544, doi:10.1038/31159 (1998).

16 Abdallah, A. M. *et al.* The ESX-5 secretion system of Mycobacterium marinum modulates the macrophage response. *J Immunol* **181**, 7166-7175 (2008).

17 Kavvas, E. S. *et al.* Updated and standardized genome-scale reconstruction of Mycobacterium tuberculosis H37Rv, iEK1011, simulates flux states indicative of physiological conditions. *BMC Syst Biol* **12**, 25, doi:10.1186/s12918-018-0557-y (2018).

18 McMahon, M. D., Rush, J. S. & Thomas, M. G. Analyses of MbtB, MbtE, and MbtF suggest revisions to the mycobactin biosynthesis pathway in Mycobacterium tuberculosis. *J Bacteriol* **194**, 2809-2818, doi:10.1128/JB.00088-12 (2012).

19 Mangan, S., Zaslaver, A. & Alon, U. The coherent feedforward loop serves as a sign-sensitive delay element in transcription networks. *J Mol Biol* **334**, 197-204 (2003).

20 Minch, K. J. *et al.* The DNA-binding network of Mycobacterium tuberculosis. *Nat Commun* **6**, 5829, doi:10.1038/ncomms6829 (2015).

21 Wernicke, S. & Rasche, F. FANMOD: a tool for fast network motif detection. *Bioinformatics* **22**, 1152-1153, doi:10.1093/bioinformatics/btl038 (2006).

22 Prosser, G. *et al.* The bacillary and macrophage response to hypoxia in tuberculosis and the consequences for T cell antigen recognition. *Microbes Infect* **19**, 177-192, doi:10.1016/j.micinf.2016.10.001 (2017).

23 Galagan, J. E. *et al.* The Mycobacterium tuberculosis regulatory network and hypoxia. *Nature* **499**, 178-183, doi:10.1038/nature12337 (2013).

24 Park, H. D. *et al.* Rv3133c/dosR is a transcription factor that mediates the hypoxic response of Mycobacterium tuberculosis. *Mol Microbiol* **48**, 833-843 (2003).

25 Vasudeva-Rao, H. M. & McDonough, K. A. Expression of the Mycobacterium tuberculosis acr-coregulated genes from the DevR (DosR) regulon is controlled by multiple levels of regulation. *Infect Immun* **76**, 2478-2489, doi:10.1128/IAI.01443-07 (2008).

26 Cortes, T. *et al.* Delayed effects of transcriptional responses in Mycobacterium tuberculosis exposed to nitric oxide suggest other mechanisms involved in survival. *Sci Rep* **7**, 8208, doi:10.1038/s41598-017-08306-1 (2017).

27 Geva-Zatorsky, N. *et al.* Oscillations and variability in the p53 system. *Mol Syst Biol* **2**, 2006 0033, doi:10.1038/msb4100068 (2006).

28 Kholodenko, B. N. Negative feedback and ultrasensitivity can bring about oscillations in the mitogen-activated protein kinase cascades. *Eur J Biochem* **267**, 1583-1588 (2000).

29 Novak, B. & Tyson, J. J. Design principles of biochemical oscillators. *Nat Rev Mol Cell Biol* **9**, 981-991, doi:10.1038/nrm2530 (2008).

30 Goentoro, L., Shoval, O., Kirschner, M. W. & Alon, U. The incoherent feedforward loop can provide fold-change detection in gene regulation. *Mol Cell* **36**, 894-899, doi:10.1016/j.molcel.2009.11.018 (2009).

31 Horak, C. E. *et al.* Complex transcriptional circuitry at the G1/S transition in Saccharomyces cerevisiae. *Genes Dev* **16**, 3017-3033, doi:10.1101/gad.1039602 (2002).

32 Kwon, Y. S. *et al.* Sensitive ChIP-DSL technology reveals an extensive estrogen receptor alpha-binding program on human gene promoters. *Proc Natl Acad Sci U S A* **104**, 4852-4857, doi:10.1073/pnas.0700715104 (2007).

33 Phuc Le, P. *et al.* Glucocorticoid receptor-dependent gene regulatory networks. *PLoS Genet* **1**, e16, doi:10.1371/journal.pgen.0010016 (2005).

34 Yang, A. *et al.* Relationships between p63 binding, DNA sequence, transcription activity, and biological function in human cells. *Mol Cell* **24**, 593-602, doi:10.1016/j.molcel.2006.10.018 (2006).

35 Dai, Z., Dai, X., Xiang, Q. & Feng, J. Robustness of transcriptional regulatory program influences gene expression variability. *BMC Genomics* **10**, 573, doi:10.1186/1471-2164-10-573 (2009).

36 Vignali, M. *et al.* NSR-seq transcriptional profiling enables identification of a gene signature of Plasmodium falciparum parasites infecting children. *J Clin Invest* **121**, 1119-1129, doi:10.1172/JCI43457 (2011).

37 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359, doi:10.1038/nmeth.1923 (2012).

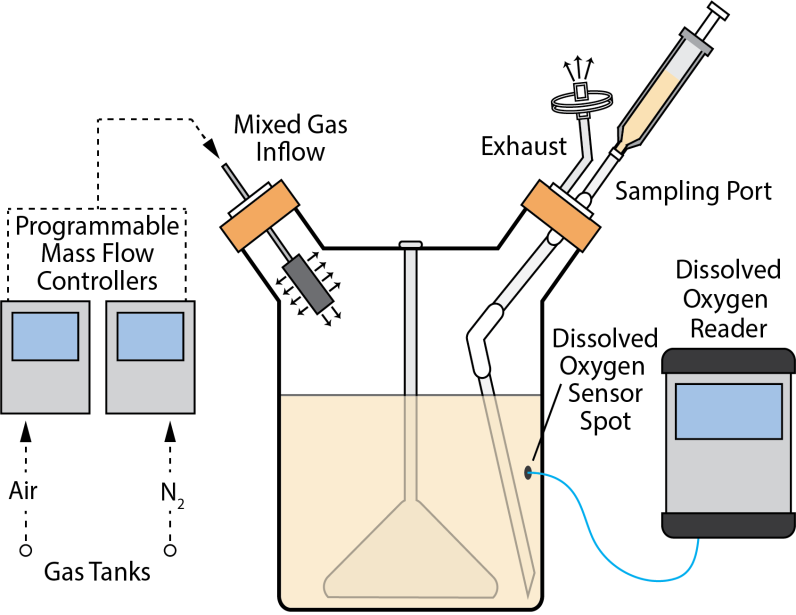
38 Rustad, T. R. *et al.* Mapping and manipulating the Mycobacterium tuberculosis transcriptome using a transcription factor overexpression-derived regulatory network. *Genome Biol* **15**, 502, doi:10.1186/PREACCEPT-1701638048134699 (2014).

**Acknowledgements**: We thank members of the Baliga (in particular Christopher Plaisier now at Arizona State University) and Shmulevich labs for critical discussions and Lee Hiner for help with mass flow controllers. Funding was provided by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health: supplementary grant awarded to Christopher Plaisier and E.J.R.P through U19 AI135976; [R01 AI128215]; and [U19 AI10676].

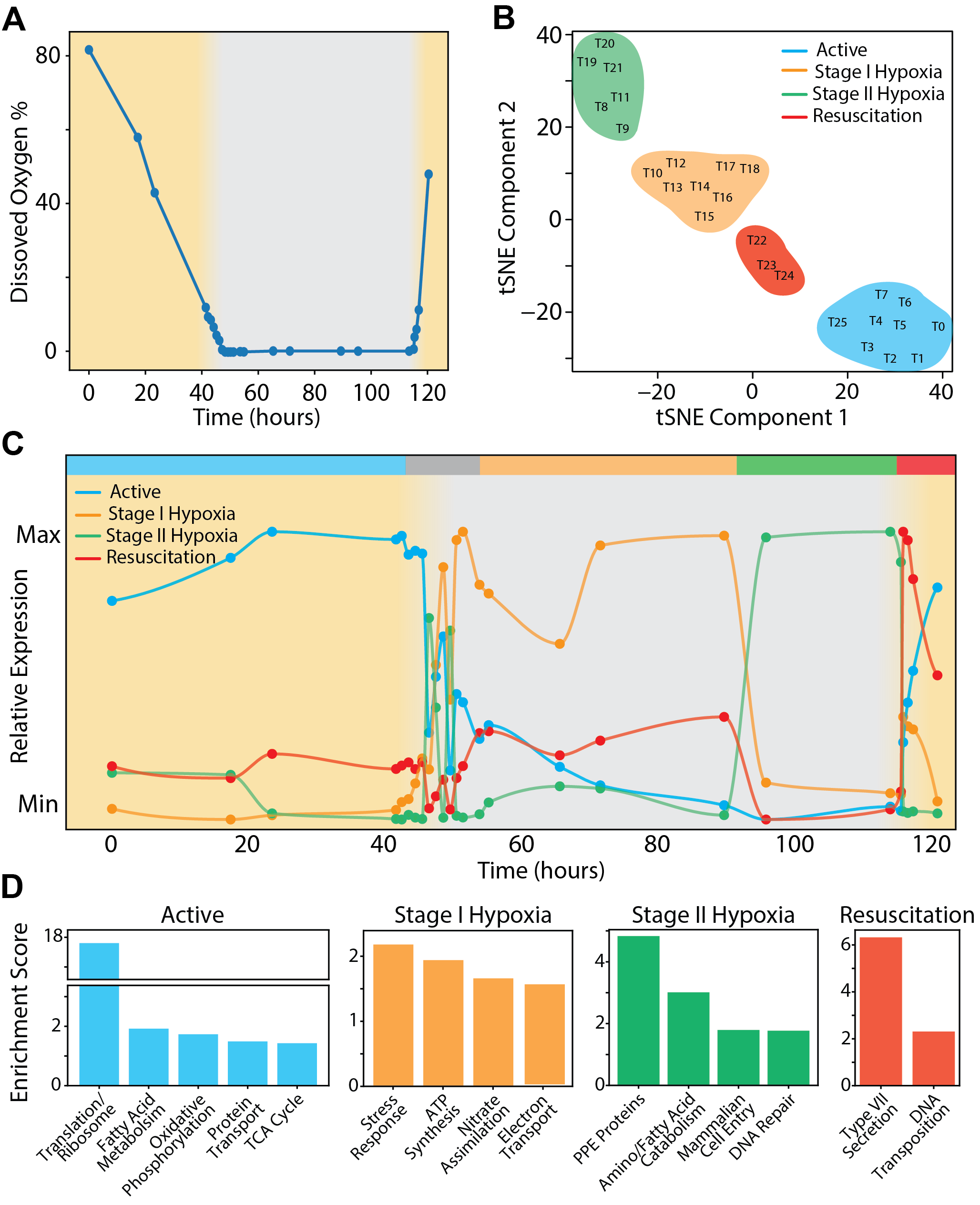
**Author contributions:** A.A., E.J.R.P., B.A., I.S., and N.S.B. designed research; A.A., E.J.R.P, M.P., A.K., and V.S. performed research; A.A., E.J.R.P., M.A-O., and B.A. analyzed data and performed computational analyses; A.A., E.J.R.P, M.A-O and N.S.B. wrote the paper.

**Competing interests:** The authors declare no competing financial interests.

**Materials & Correspondence:** Correspondence and requests for materials should be addressed to N.S.B. (nitin.baliga@systemsbiology.org).



**Figure 1. Schematic of the controlled O2 system.** Programmablemass flow controllers were used to modulate the ratio of air and nitrogen (N2) in a gas mixture that was flowed into the headspace of spinner flasks containing cultures of MTB. Dissolved oxygen sensor spots and fiber optic technology non-invasively provided real time and remote readout of the dissolved oxygen (DO) levels within the culture media. Samples were drawn from a sampling port attached to one of the side arms of the spinner flask. Four reactors were multiplexed and individually monitored for DO levels to obtain biological replicates.

****

**Figure. 2. The controlled O2 model captures distinct cell states over time course and O2****gradient. (A)** DO levels across the 120-hour time course. Points are the average of three biological replicates and the yellow shading indicate the periods of controlled O2 depletion and reaeration, whereas the white background indicates a sustained 2 day immersion in hypoxia. **(B)** t-SNE analysis of all samples across the time course and hypoxia gradient. **(C)** Average expression profiles for state-specific gene sets across the time course and hypoxia gradient. Error analysis for all genes in the four states is shown in **Fig. S4**. **(D)** Gene ontology enrichment analysis reveals the representative gene sets of the four clusters are functionally distinct.

**Fig. 3. Hypoxia transcriptional states are controlled by an incoherent feedforward loop with time delay element. (A)** Stage I (orange border) and stage II (green border) dormancy genes are regulated in a bistable manner, such that genes from only one of the two states are active at a time. **(B)** Heatmap showing the components of the I-FFL and the time delay element show hierarchical ordering of the circuit components following entry and exit from hypoxia. **(C)** Embedded within the I-FFL, a time delay element, which consists of a long TF cascade, determines the 40 hour interval of stage I hypoxia.

**Fig. 4. Fold-change detection determines expression of stage II hypoxia genes. (A)** Oscillation of stage II hypoxia genes as cells adjust to hypoxia. **(B)** Predicted parameters for I-FFL fall within regime where circuit behaves as fold-change detector.

**Materials and Methods**

The approaches used in this study include both computational and biological methods. Plots were generated using Python and R, and images prepared using Adobe Illustrator CS6.

Culturing conditions

Experiments were performed using H37Rv grown at 37°C in Middlebrook 7H9 supplemented with ADC and 0.05% Tween in spinner flasks. Working stocks were expanded from frozen aliquots shortly before experiments began. For hypoxia time-course experiment, a 50 mL culture was grown to mid-log phase, and diluted in 700 mL 7H9 media within each bioreactor to a starting A600 of 0.01. Cultures were stirred over a range of speeds throughout the experiment.

Controlled O2 model design and operation

An Oxygen Sensor Spot (PreSens, Regensburg, Germany) was adhered within a 1L disposable spinner flask with two side arms (Corning, Corning, NY). A velcro belt with a screw-on port for the fiber optic cable was wrapped around the flask. A gas line input was fastened on one arm of the flask, and a luer-lock/filter sampling port was connected to the other arm. Air and N2 gas lines were run into the Biological safety laboratory and connected to gas-specific mass flow controllers (Alicat, Tucson, AZ), whose outputs were connected downstream through a Y-connector that led into an incubator. Three separate flasks, all prepared as described above, were placed onto a stir plate inside an incubator at 37° C. The mixed gas line was split via additional Y-connecters, streamed through 0.2 um filters, and attached to the gas line inputs of each flask. Media was incubated overnight and checked for contamination before innoculated with MTB.

The mass flow controllers and oxygen sensor were linked to a computer and controlled through special retail software, which could be remotely accessed and monitored in real-time. After inoculation, we programmed the mass flow controllers to achieve a changing gas mixture gradient, which allowed us creating a steady two-day depletion, followed by two-days of sustained hypoxia, and re-aeration by flowing pure air into the headspace of the vessels and increasing the speed of the stir bars in each vessel.

RNA isolation

Samples were collected by attaching a luer-lock syringe to the sampling port. Samples were centrifuged at high speed for 5 min, supernatant discarded and cell pellet was immediately flash frozen in liquid nitrogen. Cell pellets were stored at -80° C until all samples collected and then resuspended in 600 μL of fresh lysozyme solution in TE pH 8.0 (5 mg/mL). The resuspended cells were transferred to a tube containing Lysing Matrix B (QBiogene, Inc.) and incubated at 37° C for 30 min. Following incubation, 60 μL (1/10th volume of lysate volume) of 10% SDS was added and then tubes were vigorously shaken at max speed for 30 s in a FastPrep 120 homogenizer (QBiogene) three times. Tubes were centrifuged for 1 min (max speed), then 66 μL of 3 M sodium acetate pH 5.2 added and mixed well. Acid phenol (pH 4.2) was added at 726 μL and tubes were inverted to mix well (~60 times). Samples were incubated at 65° C for 5 min, inverting tubes to mix samples every 30 s. Then, centrifuged at 14000 rpm for 5 min and upper aqueous phase was transferred to a new tube. 3M sodium acetate (pH 5.2) was added at 1/10th volume along with 3x volumes of 100% ethanol. The addition of 1 μL is optional for visualization of the pellet. Sample was mixed well and incubated at -20° C for 1 hr or overnight. Following incubation, samples were centrifuged at 14000 rpm for 30 min at 4° C, ethanol was discarded and 500 μL of 70% ethanol was added. Samples were centrifuged again at 14000 rpm for 10 min at 4° C, supernatant discarded, and any residual ethanol removed using pipet. Pellet was allowed to air dry, resuspended in 30-40 μL of RNase free water and quantified by Nanodrop (Thermo Scientific). This was followed by in solution genomic DNA digestion using RQ1 Dnase (Promega) following manufacturer’s recommendation. RNA quality was analyzed in a 2100 Bioanalyzer system (Agilent Technologies). Total RNA samples were depleted of ribosomal RNA using the Ribo-Zero Bacteria rRNA Removal Kit (Illumina).

Processing and analysis of RNA-seq data

Sample collection and RNA-extraction was performed as described above. Quality and purity of mRNA samples was determined with 2100 Bioanalyzer (Agilent, Santa Clara, CA). Samples were prepared with TrueSeq Stranded mRNA HT library preparation kit (Illumina, San Diego, CA) and multiplexed into a single run. All samples were sequenced on the NextSeq sequencing instrument in a high output 150 v2 flow cell. Paired-end 75 bp reads were checked for technical artifacts using Illumina default quality filtering steps. Raw FASTQ read data were processed using the R package DuffyNGS as described previously36. Briefly, raw reads pass through a 3-stage alignment pipeline: (i) a prealignment stage to filter out unwanted transcripts, such as rRNA, mitochondrial RNA, albumin, and globin; (ii) a main genomic alignment stage against the genome(s) of interest; and (iii) a splice junction alignment stage against an index of standard and alternative exon splice junctions. Reads were aligned to *M. tuberculosis H37Rv* (ASM19595v2) with Bowtie237, using the command line option “very-sensitive.” BAM files from stages 2 and 3 are combined into read depth wiggle tracks that record both uniquely mapped and multiply mapped reads to each of the forward and reverse strands of the genome(s) at single-nucleotide resolution. Gene transcript abundance is then measured by summing total reads landing inside annotated gene boundaries, expressed as both RPKM and raw read counts. Two stringencies of gene abundance are provided using all aligned reads and by just counting uniquely aligned reads.

Differential expression

Analysis of ChIP-seq data

To uncover the core gene regulatory circuits that govern these cell-state transitions, we combined our transcriptional data with a previously published transcription factor binding map20. This dataset was assembled by performing ChIP-seq on overexpression strains for all known regulators in MTB. While this caused the cells to exist in a non-physiological state within a different context from the one we were testing (i.e., normal growth conditions rather than hypoxia), TF binding from an overexpression strain for a hypoxic-response regulator (Rv0081) in normal growth conditions was shown to recapture the majority of sites found when that regulator bound DNA in a wild-type strain in hypoxia. Therefore, though our binding data is necessarily incomplete, we use the comprehensive TF-binding network to approximate the most densely-connected nodes of a more complete hypoxia-specific network. Indeed, to properly assay TF-binding for the genetic program outlined in this paper, one would need to perform ChIP-seq on all the relevant TFs for each separate time-point in the hypoxia time-course—a highly laborious project that goes beyond the scope of this paper. It remains for a future study to thus exhaustively “fill in the gaps” of the network we devise in this paper.

Using the TF-binding map, we assigned targets to a particular TF when a statistically significant central peak for that TF was found within a -150 and + 70 window around the TSS for a given gene. When a TF was found to bind a gene within an operon, we also assigned the genes rightward (i.e. toward the 5’ end) of the direct target that belonged to the operon that the TF targeted.

Computational modeling of network motifs

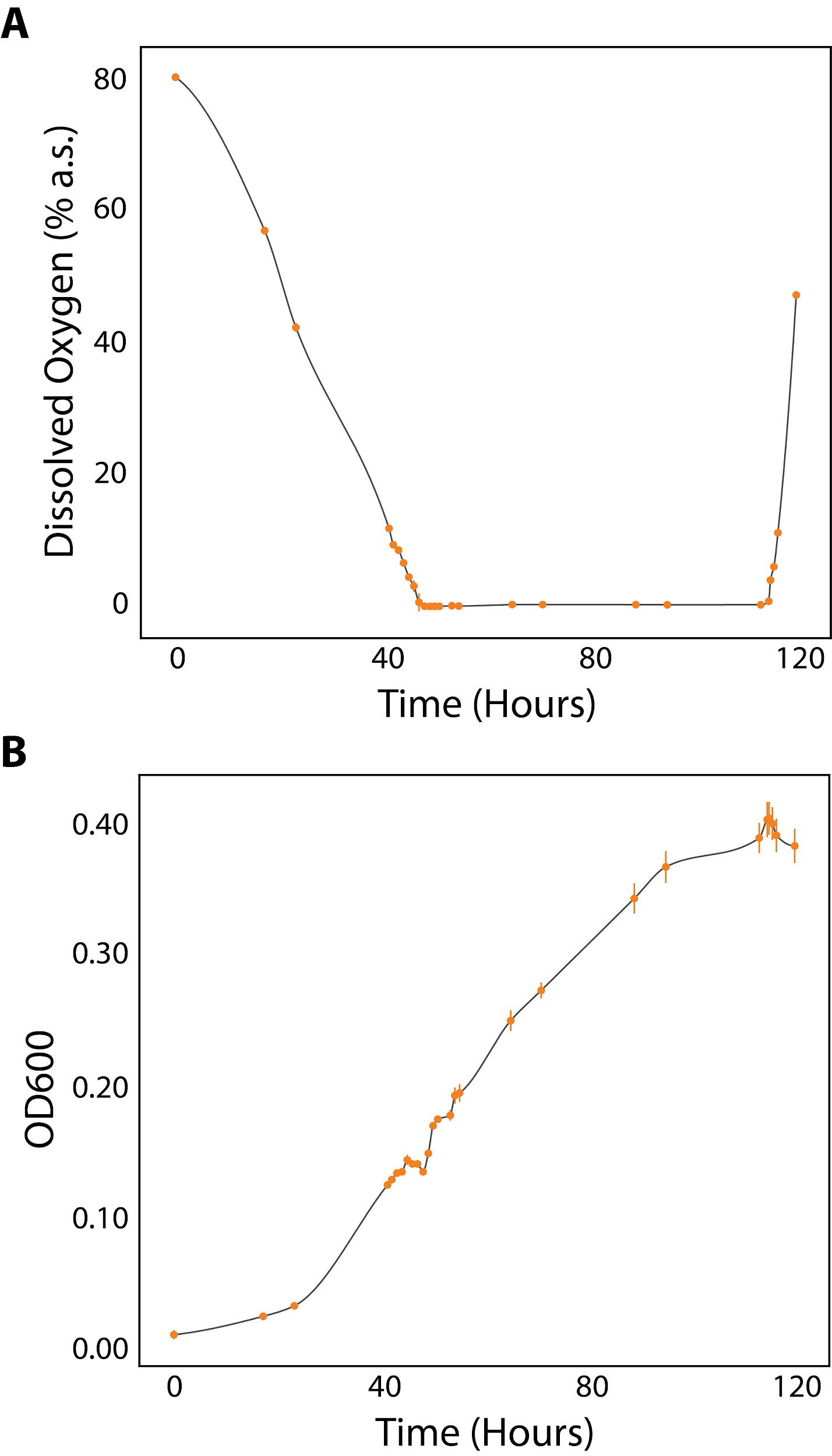
For both C-FFLs and I-FFLs, we required that a primary TF was found to bind a secondary TF, and that both those TFs were found to bind a shared target. We designated a FFL coherent when the target gene was only induced while both its regulators were also induced. We also required that the median inter-TF binding distance of both the primary and secondary TFs fell within a +/- 5 interval of 10 base-pairs. We tested the statistical significance of these median values by comparing them against the median inter-TF binding distance of all TFs that were shown to bind a common gene but lacked the inter-TF edge required for a FFL. We designated a FFL incoherent when the target gene satisfied two conditions: 1) the target was induced while its primary TF was induced but its secondary TF was repressed, and 2) the target was not induced when both its primary and secondary TFs were induced.

Robustness analysis

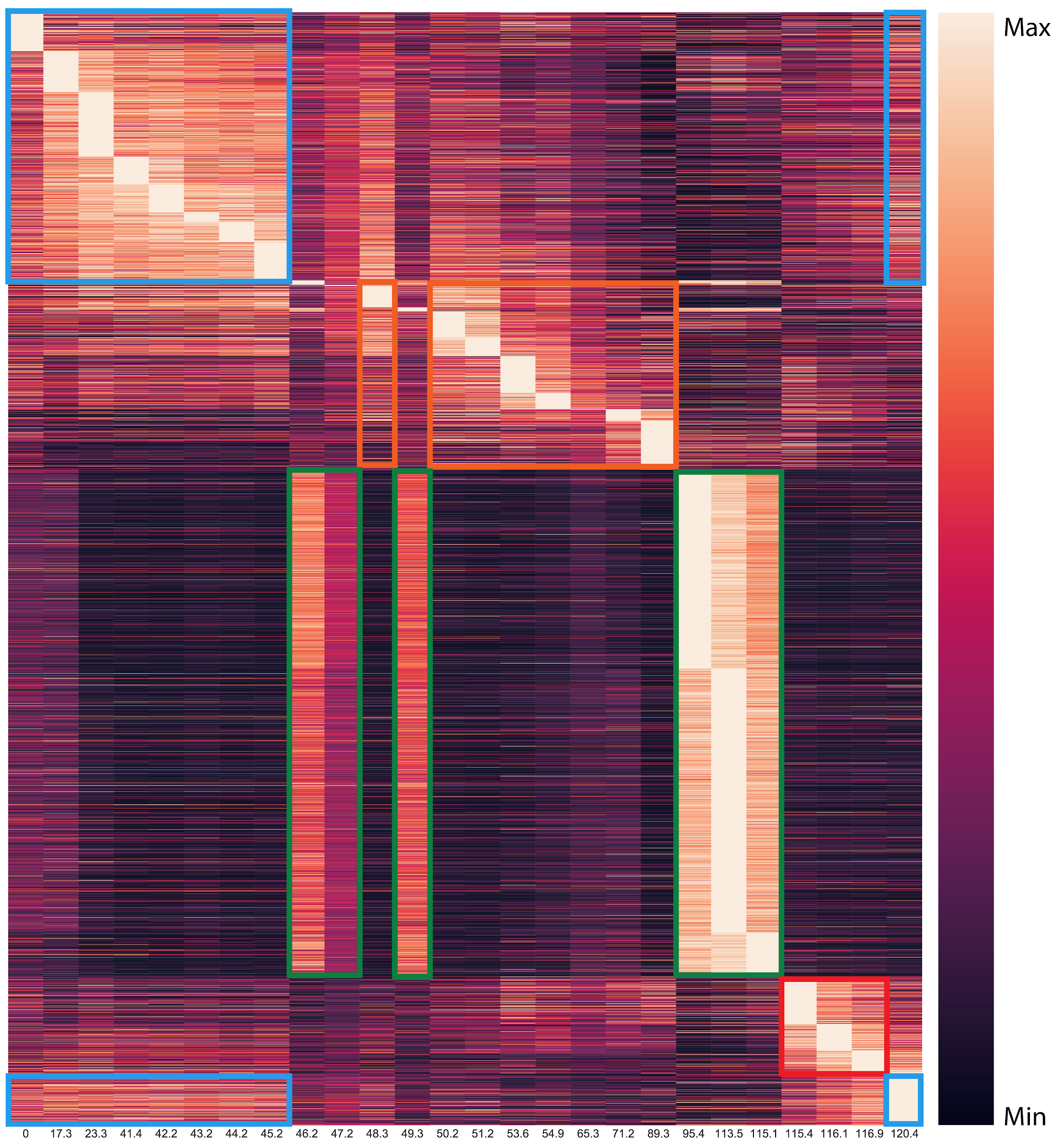
We adapted our procedure for quantifying the fragility of nodes in MTB’s gene regulatory network from a study done with yeast that employed expression data from a TF-knockout library and a database of ChIP-chip-validated TF binding sites35. Interestingly, this study found that target genes that showed significant expression changes when their regulator was knocked out also showed expression changes when that regulator was overexpressed. We exploited this fact and used our database of microarrays performed on overexpression strains to quantify expression changes in target genes, along with the comprehensive TF-binding map used previously in this paper38. We regarded a target gene’s expression change as significant when the overexpression of its regulator resulted in either a positive or negative fold-change amounting to square root 2. When the overexpression of at least five TFs resulted in a target gene’s experiencing a significant expression change, we labeled that gene as fragile. Finally, we assigned the fragility score of a given TF by the number of fragile target genes it was known to regulate.

Fold-change detector analysis

**Supplemental Materials**



**Fig. S1. Bioreactor system is highly reproducible across three biological replicates. (A)** Dissolved oxygen curve of growth medium across hypoxia and re-aeration time-course measured with non-invasive fiber-optic technology. **(B)** Optical density of samples drawn for RNA extraction. (Datapoints show median of three replicates, and error bars show Standard Error of measurements).

****

**Fig. S2. MTB transcriptomes from each time point cluster into one of six distinct cell states.** Heatmap of all2,723 significantly differentially expressed genes across the time course. Colored boxes indicate the gene sets that define the four cell states: active (786 genes outlined in blue); stage I hypoxia (439 genes in orange), stage II hypoxia (1,264 in green), and resuscitation (234 genes in red). Expression for each gene represents the median of three biological replicates. Gene sets and expression values are included in **Table S1**.

**Fig. S3. Dynamical behavior of four gene modules across three experimental replicates.** Relative expression profiles of representative genes for **(A)** active (blue), **(B)** stage I dormant (orange), **(C)** stage II dormant (green), **(D)** and resuscitation (red) cell states. Datapoints show the median of three replicates, and error bars show Standard Error of measurements between relative expression profiles for the three replicates. The shaded region represents the standard deviation for the entire set of genes in the gene module across all three replicates over time.

**Fig. S5.** Using ChIP-seq data and network motif mining tools, we inferred a temporal genetic program underlying MTB’s response to hypoxia. The sub-circuit is a cascade of interlocking FFLs, controlled by the master regulator *Rv0081* (blue), known to be involved in hypoxia 22,23. TFs (dark grey) that lie downstream of the Rv0081 are also implicated in another layer of FFLs (light grey TFs that are also included within target genes).

**Fig. S6.** “Mutual Inhibition of Rv0081”

Stage I and Stage II dormancy genes are expressed in a bi-stable manner, where only one or the other set of genes appears active at a time in the dormancy program. A simple, and common, circuit architecture that achieves this dynamic is the mutual inhibition circuit, where regulators in one stable state repress the genes in the other stable state, and vice versa. We hypothesize that in addition to activating genes in their own state, state-specific regulators also act to repress the genes in the other state.

**Fig S7**. “Just-in-time expression pattern”

Stage I dormancy genes activate in a non-uniform, time-dependent manner, resembling the so-called “just-in-time” expression patterns reported in metabolic pathways, where enzymes are synthesized in the order of their placement within a pathway hierarchy. Just-in-time expression is typically controlled by single regulators that bind to a set of targets with varying affinities, with stronger affinities indicating enzymes higher up in a pathway, and lower affinities indicating enzymes lower down in a pathway. We hypothesize that such a mechanism, driven by a distinct set of TFs activated near the start of stage I dormancy and terminated by a distinct set of TFs activated at the end of stage I dormancy, controls the transcriptional dynamics observed in stage I dormancy.

**Fig S8**. “parallel coherent FFL”

Genes in stage II dormancy activate in a near-uniform fashion, rapidly and coordinately inducing after 40 hours in hypoxia. A motif known as the dense overlapping regulon, composed of a backbone of parallel feed-forward loops as well as single-input regulators, integrates many signals to rapidly activate a broad swath of genes. Evidence for such a backbone from both motif topologies and very proximate binding between FFL regulators (as would underlie an AND gate) suggest that a motif like the DOR may form the basis for stage II dormancy.

**Fig. S9. More extensive ranking of TF perturbability.** TFs with the top 20 fragility scores show a broad distribution of scores. DosR, ranked as #11, belongs to stage I hypoxia and yielded only minor phenotypes when its deletion strain was forced to undergo hypoxia. The higher ranked TFs not only belong to stage II hypoxia but also, in many cases, appear to interact cooperatively with Rv0081. Thus we believe that double knockouts of Rv0081 and its auxiliary binding partners will deal more lethal damage to MTB’s ability to adapt to hypoxia.