Published in final edited form as:
Nature. 2013 July 11; 499(7457): 178–183. doi:10.1038/nature12337.

The *Mycobacterium tuberculosis* regulatory network and hypoxia

James E. Galagan[1,2,3,4], **Kyle Minch**[5,*], **Matthew Peterson**[1,*], **Anna Lyubetskaya**[3,*], **Elham Azizi**[3,*], **Linsday Sweet**[6,*], **Antonio Gomes**[3,*], **Tige Rustad**[5], **Gregory Dolganov**[7], **Irina Glotova**[3], **Thomas Abeel**[4,8], **Chris Mahwinney**[1], **Adam D. Kennedy**[9], **René Allard**[10], **William Brabant**[5], **Andrew Krueger**[1], **Suma Jaini**[1], **Brent Honda**[1], **Wen-Han Yu**[1], **Mark J. Hickey**[5], **Jeremy Zucker**[4], **Christopher Garay**[1], **Brian Weiner**[4], **Peter Sisk**[4], **Christian Stolte**[4], **Jessica K. Winkler**[5], **Yves Van de Peer**[8], **Paul Iazzetti**[1], **Diogo Camacho**[1], **Jonathan Dreyfuss**[1], **Yang Liu**[7], **Anca Dorhoi**[11], **Hans-Joachim Mollenkopf**[12], **Paul Drogaris**[10], **Julie Lamontagne**[10],

Correspondence and requests for materials should be addressed to J.E.G. (jgalag@bu.edu).
*These authors contributed equally to this work.

Supplementary Information is available in the online version of the paper.

Author Contributions J.E.G. led the project with G.K.S., oversaw ChIP-Seq, wrote the paper and produced figures, discussed results and implications, oversaw data integration, and performed analyses. K.M. co-designed and performed ChIP and transcriptomic experiments, discussed results and implications, and commented on the manuscript. M.P. developed the analysis pipeline for ChIP-Seq data, performed all ChIP-Seq data analysis, and contributed multiple figures and text. A.L. performed all analysis of the integration of TF induction transcriptomics with ChIP-Seq data, contributed to analysis of ChIP-Seq binding data, and contributed multiple figures and text. E.A. developed the predictive models of gene expression, and contributed all corresponding figures and text. L.S. performed lipidomics experiments and data analysis, discussed the results and implications, and contributed figure and text to the paper. A.G. developed the improved blind deconvolution algorithm for ChIP-Seq, contributed to analysis of all ChIP-Seq data, and contributed corresponding figures. T.R. designed and performed hypoxic time course and transcriptomic experiments, discussed results and implications and commented on the manuscript. G.D. performed all RT–PCR transcriptomics experiments and contributed analyses to the paper. I.G. performed the DREM analysis and provided corresponding the figure. T.A. analysed ChIP-Seq data, developed the interfaces for data sharing and public release, and provided text. C.M. performed all library preparation and sequencing for ChIP-Seq. A.D.K. performed the metabolomics measurements, data analysis and their interpretation, discussed the results and implications and commented on the manuscript. R.A. was responsible for overview of bioinformatics and statistical data analysis. W.B. performed hypoxic time course, ChIP and transcriptomic experiments, and discussed results and implications. A.K. performed the experimental analysis of KstR de-repression and provided the corresponding figure. S.J. performed the experimental analysis of KstR de-repression, and provided the corresponding figure. M.J.H. produced individual MTB strains for ChIP-Seq experiments, and discussed results and implications. J.Z. developed and curated the MTB metabolic model. C.G. contributed to analysis of profiling data. J.K.W. performed ChIP and transcriptomic experiments, and discussed results and implications. Y.V.P. provided support and advice. P.I. contributed to the analysis of KstR expression and the validation of KstR binding sites. B.W. contributed to the ChIP-Seq analysis pipeline. P.S. and C.S. developed the interfaces for data sharing and public release. D.C. contributed to initial network analysis. J.D. contributed to analysis of profiling data. Y.L. contributed expression data for TB under different lipids. P.D. was responsible for experimental design and mass spectrometry analysis. J.L. was responsible for coordinating sample analysis, data generation, annotation and results reporting Y.Z. was responsible for proteomics statistical data analysis. J.P. was responsible for analysis of LC-MS and LC-MS/MS data analysis, protein identification and maintenance of annotation databases. A.D. and H.-J.M. discussed the results and implications and commented on the manuscript. B.H. and W.-H.Y. developed the ChIP protocol; S.T.P. developed the ChIP protocol, performed the KstR RT–PCR experiments, and performed the MTB KstR native promoter ChIP-Seq experiments. S.R. developed the ChIP protocol, oversaw experimental work on KstR and commented on the manuscript. S.H.E.K. discussed the results and implications and commented on the manuscript. R.P.M. performed the metabolomics measurements, data analysis, and their interpretation; discussed the results and implications and commented on the manuscript. D.C. was responsible for overall scientific direction of the proteomic core. D.B.M. oversaw lipidomics experiments, contributed to integration of methods across mass spectral platforms, discussed the results and implications and commented on the manuscript. D.R.S. oversaw the hypoxic culture, ChIP and transcriptomic experiments, discussed results and implications, provided text and commented extensively on the manuscript. G.K.S. led the project with J.E.G., oversaw RT–PCR experiments, discussed results and implications, provided text and commented extensively on the manuscript. G.K.S. and D.R.S. are co-last authors.

Expression data were deposited at GEO (accession number GSE43466). The proteomics data have been deposited in the Proteome Xchange with the identifier PXD000045.

The authors declare no competing financial interests.

Yiyong Zhou[10], **Julie Piquenot**[10], **Sang Tae Park**[2], **Sahadevan Raman**[2], **Stefan H. E. Kaufmann**[11], **Robert P. Mohney**[9], **Daniel Chelsky**[10], **D. Branch Moody**[6], **David R. Sherman**[5,13], and **Gary K. Schoolnik**[7,14]

[1]Department of Biomedical Engineering, Boston University, Boston, Massachusetts 02215, USA

[2]Department of Microbiology, Boston University, Boston, Massachusetts 02215, USA

[3]Bioinformatics Program, Boston University, Boston, Massachusetts 02215, USA

[4]The Eli and Edythe L. Broad Institute of Harvard and MIT, Cambridge, Massachusetts 02142, USA

[5]Seattle Biomedical Research Institute, Seattle, Washington 98109, USA

[6]Division of Rheumatology, Immunology and Allergy, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts 02115, USA

[7]Departments of Medicine and of Microbiology and Immunology, Stanford Medical School, Stanford, California 94305, USA

[8]Department of Plant Biotechnology and Bioinformatics, Ghent University, 9052 Gent, Belgium

[9]Metabolon Inc., Durham, North Carolina 27713, USA

[10]Caprion Proteomics, Inc., Montreal, Quebec H4S 2C8, Canada

[11]Department of Immunology, Max Planck Institute for Infection Biology, 10117 Berlin, Germany

[12]Microarray Core Facility, Max Planck Institute for Infection Biology, 10117 Berlin, Germany

[13]Interdisciplinary Program of Pathobiology, Department of Global Health, University of Washington, Seattle, Washington 98195, USA

[14]Division of Infectious Diseases and Geographic Medicine, Department of Medicine, Stanford Medical School, Stanford, California 94305, USA

Abstract

We have taken the first steps towards a complete reconstruction of the *Mycobacterium tuberculosis* regulatory network based on ChIP-Seq and combined this reconstruction with system-wide profiling of messenger RNAs, proteins, metabolites and lipids during hypoxia and re-aeration. Adaptations to hypoxia are thought to have a prominent role in *M. tuberculosis* pathogenesis. Using ChIP-Seq combined with expression data from the induction of the same factors, we have reconstructed a draft regulatory network based on 50 transcription factors. This network model revealed a direct interconnection between the hypoxic response, lipid catabolism, lipid anabolism and the production of cell wall lipids. As a validation of this model, in response to oxygen availability we observe substantial alterations in lipid content and changes in gene expression and metabolites in corresponding metabolic pathways. The regulatory network reveals transcription factors underlying these changes, allows us to computationally predict expression changes, and indicates that Rv0081 is a regulatory hub.

Mycobacterium tuberculosis (MTB) has been associated with human disease for thousands of years and its success is due in part to the ability to survive within the host for months to

decades in an asymptomatic state. The mechanisms underlying this persistence in the host are poorly understood, although adaptations to hypoxia are thought to have a prominent role[1,2]. Hypoxia produces widespread changes in the bacterium and induces a non-replicating state characterized by phenotypic drug tolerance. Within the host, MTB also shifts to lipids, including cholesterol, as a primary nutrient[3–6]. Lipid catabolism is, in turn, linked to the biosynthesis of lipids that serve as energy stores, factors associated with virulence and immunomodulation, and components of the unique and complex cell wall of MTB[7–9].

The regulatory mechanisms underlying these and other adaptations are largely unknown, as functions for only a small fraction of the 180+ MTB transcription factors (TFs) are known, direct DNA binding data exist for only a handful of sites, and the interactions between TFs necessary for complex behaviour have not been studied. We also lack a comprehensive understanding of the cellular changes underlying pathogenesis, with existing studies typically focused on specific molecular components that can be difficult to integrate with results from other studies. To address these challenges, we have performed a systems analysis of the MTB regulatory and metabolic networks, with an emphasis on hypoxic conditions thought to contribute to MTB persistence in the host.

Mapping and functional validation of TF binding sites

To systematically map TF binding sites, we performed chromatin immunoprecipitation followed by sequencing (ChIP-Seq)[10–12] using Flag-tagged transcription factors episomally expressed under control of a mycobacterial tetracycline-inducible promoter[13–15] (Supplementary Fig. 1). The inducible promoter system allows us to study all MTB TFs in a standard and reproducible reference state without a priori knowledge of the conditions that normally induce their expression. Using a custom pipeline (Supplementary Fig. 2 and Supplementary Table 1) we identified binding sites in regions of enrichment with high spatial resolution. Using this method, we mapped 50 TFs. We compared the results with previous reports for two well-studied regulators for which strong evidence for direct binding exists: the activator DosR (Rv3133c) and the repressor KstR (Rv3574).

Our method shows high sensitivity and reproducibility. We identified all known direct binding regions for DosR (Supplementary Fig. 3) and KstR (Fig. 1a) and recovered the known motifs for these factors (Supplementary Material). Coverage for enriched sites is highly correlated between replicates (Fig. 1b and Supplementary Fig. 4). There is also high reproducibility in binding location, with distances between replicate binding sites less than the length of predicted binding site motifs for the vast majority of sites (Fig. 1b). Moreover, for 11 different TFs we also see substantial concordance between binding observed in normoxia and binding observed in hypoxia (Supplementary Fig. 5).

ChIP enrichment is a function of the number of cells in which a site is bound[16] which in turn is governed by the affinity of the site and the concentration of the factor. Thus, increasing TF induction was predicted to increase the occupancy of strong sites up to a saturation limit while occupying weaker affinity sites. This is confirmed by comparing ChIP-Seq

experiments after inducing three different factors to different expression abundances (Fig. 1c, Supplementary Fig. 6 and Supplementary Fig. 7).

Consistent with this observation, at the highest levels of TF induction we identify more binding sites than previously reported for DosR and KstR (Fig. 1a); most, but not all, of these newly-identified sites have lower ChIP-Seq coverage than the majority of previously identified sites. Abundant binding of transcription factors, particularly to low affinity sites, has been reported in yeast, worm, fly and mammalian cells[16–18] but, to our knowledge, these data represent the first large-scale observation in a prokaryote. We have confirmed that many novel sites can be bound at physiological levels of these TFs, and that sites show sequence specificity for each TF. In addition, for DosR, nearly all novel sites are also found when performing ChIP using anti-DosR antibodies in a wild-type background (Supplementary Material Section 2.4).

To assess the degree to which binding is associated with transcriptional regulation, we performed transcriptomic analysis from the same cultures in which regulators were induced for ChIP-Seq. Using these data we developed a procedure for determining the possible regulatory roles of identified binding sites (Supplementary Fig. 11). This method identified a regulatory effect for 92% and 80% of previously identified DosR and KstR sites, respectively, and associated regulation with 43% and 36% of new DosR and KstR binding sites revealed using ChIP-Seq (false discovery rate (FDR) = 0.15). Many, but not all, newly identified sites show weaker ChIP-Seq enrichment, indicating evidence for regulatory effects of weak binding even for well-studied regulators[19–21]. This was corroborated by knockout expression data for these TFs (Supplementary Fig. 12).

Applying our method to all peaks from all 50 TFs, we could assign a potential regulatory role to 25% of peaks within 1,000 base pairs (bp) on either side of the site (FDR = 0.15; 18% of sites were significant with q value = 0) (Fig. 1e). Stronger binding sites are more often associated with regulation than weaker sites, independent of window size, suggesting a possible correlation between binding strength and regulatory impact (Supplementary Fig. 13). Such a correlation could explain why the stronger sites have been reported, as they would be more easily detected. The use of a 1-kilobase (kb) window ensures that predictions are not a priori biased to proximal promoter regions. However, even with 4-kb windows, the distance between binding sites and associated target genes is consistent with expectation: binding sites are typically located within 500 bp of the start codon of the predicted regulated gene (Fig. 1f), with 24% located in the upstream intergenic region. By contrast, 76% of sites fall into annotated coding regions and a significant proportion are associated with regulation. Extensive genic binding has been reported[17,18] and there remains no consensus on its functional significance. Prokaryotic binding sites have been largely mapped with lower resolution ChIP-Chip that frequently show broad binding overlapping both genic and intergenic regions[22]. Our method detects binding at high spatial resolution and indicates that some genic binding may reflect the extension of promoter regions into upstream genes, alternative promoter regions within genes, or errors in the current annotation of genic regions. As with previous reports[17], we cannot assign regulatory roles to all detected binding sites (Supplementary Fig. 13). We discuss potential issues with false positives and negatives in Supplementary Material.

We also tested the degree to which observed binding could be used to develop models predictive of gene expression. We developed computational models relating the expression of target genes to the expression of TFs predicted to bind the target (Supplementary Fig. 14). The relationship between TFs and target genes was parameterized based on subsets of the overexpression data and tested on the remaining using cross-validation. We could generate models that predict more accurately than random TF assignments for 28% of genes with binding (positive false discovery rate (pFDR) < 0.15; Supplementary Table 4). More importantly, as described below, we confirmed the ability of these models to predict expression for genes in an independent data set.

An MTB regulatory network model

Using the combination of binding site mapping and functional validation via expression profiling, we analysed the regulatory interactions of 50 TFs (26% of predicted MTB TFs). Our TF selection was weighted towards those that respond to hypoxia or are associated with lipid metabolism. By linking TFs with genes based on binding proximity (Supplementary Text) and potential regulation, we constructed the regulatory network model shown in Supplementary Fig. 15 (also Supplementary Fig. 16). The TB regulatory network model has topological features seen for other organisms (Supplementary Text), including the presence of ‘hubs’ or TFs that interact with many genes. Surprisingly, Rv0081 forms the largest hub identified among the TFs reported, and interacts with another hub, Lsr2, an MTB analogue of the H-NS nucleoid binding protein[23,24] (Supplementary Text).

The network also begins to reveal interactions between transcription factors mediating responses of MTB to its environment (Supplementary Material). Of particular interest is a subnet work involving responses to altered oxygen status and lipid availability (Fig. 2). These responses, among the most extensively studied in MTB, have been viewed largely as separate phenomena. DosR and Rv0081 mediate the initial response to hypoxia, whereas a larger stimulon termed the enduring hypoxic response (EHR) is induced later in hypoxia[25]. KstR controls a large regulon mediating cholesterol degradation and lipid and energy metabolism[26,27]. KstR was identified as part of the EHR, but the biology linking these responses was unclear.

We identified two potential regulators for KstR. Rv0081 is predicted to repress both Rv0324 and KstR, whereas Rv0324 is predicted to activate KstR. Rv0081 is the only regulator in the initial hypoxic response apart from DosR, and our network identifies an interaction underlying the known induction of Rv0081 by DosR. Rv0324 is a regulator associated with the EHR[25].

We also identify several potential regulators of DosR: Rv2034, Rv0767c and PhoP (Rv0757). Rv2034 is an EHR regulator predicted to activate DosR, thus providing possible positive feedback from the enduring to the initial hypoxic response (during revision, this link between Rv2034 and DosR was confirmed[28]). PhoP mediates a range of responses, including upregulating DosR[29–31], although direct regulation of DosR by PhoP had not been previously demonstrated. PhoP binding to DosR is the strongest among 50 TFs, providing a mechanism for this regulatory link and supporting the conclusion that regulation of hypoxia

a daptation by PhoP is indirect through this connection with DosR[29]. PhoP also mediates pH adaptation and our data confirm direct binding between PhoP and the *aprABC* locus required for this[32]. PhoP is known to modulate the production of virulence lipids and we predict PhoP to bind upstream of and directly regulate WhiB3 (Rv3416), which codes for a redox-sensitive protein that directly regulates the production of these lipids[33]. In addition to PhoP, both Rv0081 and Lsr2 also display binding to whiB3, with activation predicted by Rv0081. Taken together, the data reveal an interconnected subnet work linking hypoxic adaptation, lipid and cholesterol degradation, and lipid biosynthesis (Supplementary Text).

Profiling and prediction during hypoxia and re-aeration

To broadly assess the changes associated with altered O_2 availability, and assess the explanatory power of the regulatory network in these responses, we performed systems level lipidomic, proteomic, meta-bolomic and transcriptomics profiling of MTB during a time course of hypoxia and subsequent re-aeration (Supplementary Fig. 17 and Methods). We cultured MTB in a medium without detergent or exogenous lipids. All measurements were normalized to baseline levels before hypoxia, and integrated with a manually curated model of MTB metabolism (Supplementary Fig. 18). We summarize key results here and provide additional details and results in Supplementary Text.

Changes in oxygen availability result in expression changes to nearly one-third of all MTB genes (Supplementary Fig. 19A). To identify temporal trends and associate them with possible regulators, we clustered expression data into paths using DREM[34] (Supplementary Text). We identified Rv0081 as a candidate high-level regulator broadly predictive of the overall expression of sets of genes during hypoxia and re-aeration (Supplementary Fig. 19b).Abroad regulatory role for Rv0081 is thus supported by three independent sources of evidence: Rv0081 overexpression in normoxia alters the expression of numerous genes, Rv0081 ChIP-Seq reveals a large number of binding sites which are also detected during hypoxia (Supplementary Fig. 20), and the expression and predicted regulatory role of Rv0081 correlates with the expression of the genes it binds during hypoxia.

We next sought to assess the degree to which the regulatory network could be used to predict changes in the expression of individual genes during hypoxia and re-aeration. We used the regression models described above—parameterized by independent ChIP-Seq and TF overexpression transcriptomics data (Supplementary Material)—and generated predictions that are significantly better than random for 66% of genes with significant changes. Examples are shown in Fig. 3 and Supplementary Fig. 21. In particular, we correctly predict the pattern of expression of KstR, confirming an implication of the network topology. Importantly, these data also indicate that the regulatory network, built from a normoxic baseline, can generalize to hypoxia.

Alterations in lipid metabolism

Consistent with predictions of the regulatory network during hypoxia, we found strong induction of genes associated with lipid catabolism and cholesterol degradation, including the regulator kstR (Fig. 3, Supplementary Fig. 18 and Supplementary Fig. 22). KstR induction by hypoxia is predicted by the core regulatory network. However, kstR is a

repressor[26] and kstR-repressed cholesterol degradation genes are among those induced. KstR de-repression occurs during growth on cholesterol[27].However, no cholesterol or other exogenous lipids are present in our medium. Follow-up studies suggest that de-repression of kstR may be due to fatty acids endogenous to MTB or their metabolites (Supplementary Text).

The accumulation of triacylglycerides (TAGs) during hypoxia and in TB patient sputum samples, and their utilization upon re-aeration, has been reported[7,8,35]. We also observe TAG accumulation during hypoxia and rapid depletion during re-aeration (Fig. 4).A detailed systems view associated with these changes (Supplementary Text) suggests a scenario in which metabolites upstream of DAG decrease in production, and TAG accumulation results from conversion of existing DAGs to TAGs via triacylglyceride synthase. We also observe changes potentially related to TAG utilization. The regulatory network identifies several regulatory links potentially relevant to these changes (Supplementary Fig. 18). Induction of *tgs1* by DosR is well established[7,36,37], and we identify this link. The network also identifies oxygen-responsive regulators of *tgs2* (Rv0081, Rv0324) and *tgs4* (DosR, Rv0324) and our models predict positive regulation of these genes in hypoxia by these TFs (Fig. 3). Further, three of four lipase genes (*Rv3176, Rv1169c* and *Rv3097c*) induced during hypoxia are influenced by regulators in the core network, and in these three cases we are able to predict their expression profiles using our gene expression models (Fig. 3).

MTB uses methylmalonyl-CoA as a precursor to synthesize a complex set of surface-exposed methyl-branched lipids including acylated trehaloses (PAT/DAT), sulphoglycolipids (SGL) and phthiocerol dimycocerosates (PDIM), the latter two associated with virulence in murine models[38–42]. During hypoxia, the expression of biosynthetic genes for SGL, PAT/DAT, PDIMs and methylmalonyl are generally downregulated (Supplementary Fig. 18). Correspondingly, during hypoxia mass spectral signals corresponding to diacylated sulphoglycolipid (AC_2SGL) (a precursor to SL-1, the major SGL in MTB) and DATs seemed unaltered, whereas ions corresponding to PDIMs showed a modest decline (Fig. 4, DATs not shown). Conversely, during re-aeration, we observed induction of genes encoding enzymes in the methylmalonyl pathway. The activation of the methylcitrate cycle and accumulation of methylcitrate suggests the availability of precursors for methylmalonate. Consistent with this hypothesis, we see statistically significant increases in AC_2SGL (Fig. 4).

The regulation of the methylmalonyl pathway is partially explained by the regulatory network. All three subunits of the propionyl-CoA carboxylase (PCC) complex (*AccA3*, *AccD5* and *AccE5*) are regulated by hypoxia regulators (Fig. 3). Both MutA and MutB also display regulation by KstR and Lsr2. Regulation associated with methyl-branched lipid biosynthesis, in contrast, is complex. Whib3 is regulated by PhoP in the model, and both are known to modulate the production of PAT/DAT (via *pks3*) and SL (via *pks2*)[29,33]. Our network predicts a PhoP/WhiB3 FFL underlying this phenomenon, with PhoP regulating *whiB3* and both regulating *pks2/pks3* (Supplementary Fig. 25). Similar regulatory complexity is seen for DIM, although regulation of key steps in DIM synthesis by Rv0081, PhoP, DosR and KstR is predicted.

Mycolyl glycolipids are important immunomodulatory components of the mycobacterial cell wall. As seen in other systems[43–45], we observe increases in free mycolic acids during hypoxia that are reversed during re-aeration (Fig. 4). Conversely, we observe the opposite effects on trehalose monomycolates (TMMs) (Fig. 4) and trehalose dimycolates (TDMs) (not shown). Similar effects have recently been reported for TDMs in *Mycobacterium smegmatis* during biofilm formation[45] and TMMs in MTB during the transition into a dormant "non-culturable" state induced by a potassium-free medium[43].The rapid, reversible and nearly complete mobilization of glycosylated to free mycolates during hypoxic dormancy is also compatible with decreased need to deliver mycolic acids to non-dividing cells.

Concluding remarks

This report presents an initial step in the reconstruction of the MTB regulatory network, based on 50 TFs, and its integration with system-wide profiling of MTB during a time-course of hypoxia and re-aeration. Although necessarily incomplete, the regulatory network confirms previously known physical interactions, provides possible mechanisms for known regulatory interactions, provides a framework for reinterpreting existing data, and identifies network motifs thought to underlie dynamic behaviour. The predictive models take a first step towards systems modelling, and integration of the network model with profiling data provides new insight about the physiological consequences of regulatory programs induced by changes in oxygen availability—a perturbation relevant to host adaptation. The results provide a foundation for ongoing efforts to map the complete transcriptional regulatory network, and to extend it to include signalling and non-coding RNAs[46]. The results presented here identify compelling questions for further investigation (Supplementary Text). Studies now focus on determining how the *in vitro* network connections and physiological changes identified here relate to adaptations of the microbe in the intracellular environment of the macrophage.

METHODS SUMMARY

MTB H37Rv was used for all experiments with the single exception of one experiment performed in *M. smegmatis* (Supplementary Fig. 21). This MTB strain was fully sequenced by the Broad Institute (GI:397671778). For Chip-Seq, cells were cultured in Middle brook 7H9 with ADC (Difco), 0.05% Tween 80, and 50 μg ml^{-1} hygromycin B at 37 °C with constant agitation and induced with 100 ng ml^{-1} anhydrotetracycline (ATc) during mid-log-phase growth, and ChIP was performed using a protocol optimized for mycobacteria and related Actinomycetes. For the hypoxia and re-aeration time-course, bacilli were cultured in bacteriostatic oxygen-limited conditions (1% aerobic O_2 tension) for seven days, followed by re-aeration. Bacteria were cultured in Sauton's medium without detergent or exogenous lipid source. Profiling samples were collected as described in the Supplementary Text. All data available at http://TBDB.org. Expression data also available at GEO (accession number GSE43466).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This project has been funded in whole or in part with Federal funds from the National Institute of Allergy and Infectious Diseases National Institute of Health, Department of Health and Human Services, under contract no. HHSN272200800059C and U19 AI 076217, R01 AI 071155, the Paul G. Allen Family Foundation (to DRS), the National Science Foundation Pre-doctoral Fellowship Program (to K.M.), and the Burroughs Wellcome Fund Award for Translational Research. We acknowledge D. C. Young for lipidomics mass spectrometry services and advice. We would also like to thank L. Carvalho for his advice on the statistical analysis of the gene expression modelling. We are grateful for the administrative assistance of S. Shiviah and S. Tucker and for the support and advice of V. Di Francesco, K. Lacourciere, P. Dudley and M. Polanski.

References

1. Manabe YC, Bishai WR. Latent *Mycobacterium tuberculosis*-persistence, patience, and winning by waiting. Nature Med. 2000; 6:1327–1329. [PubMed: 11100115]
2. Flynn JL, Chan J. Tuberculosis: latency and reactivation. Infect. Immun. 2001; 69:4195–4201. [PubMed: 11401954]
3. Schnappinger D, et al. Transcriptional adaptation of *Mycobacterium tuberculosis* within macrophages: insights into the phagosomal environment. J. Exp. Med. 2003; 198:693–704. [PubMed: 12953091]
4. Yang X, Nesbitt NM, Dubnau E, Smith I, Sampson NS. Cholesterol metabolism increases the metabolic pool of propionate in *Mycobacterium tuberculosis*. Biochemistry. 2009; 48:3819–3821. [PubMed: 19364125]
5. Miner MD, Chang JC, Pandey AK, Sassetti CM, Sherman DR. Role of cholesterol in *Mycobacterium tuberculosis* infection. Indian. J. Exp. Biol. 2009; 47:407–411. [PubMed: 19634704]
6. Chang JC, et al. *igr* genes and *Mycobacterium tuberculosis* cholesterol metabolism. J. Bacteriol. 2009; 191:5232–5239. [PubMed: 19542286]
7. Daniel J, Maamar H, Deb C, Sirakova TD, Kolattukudy PE. *Mycobacterium tuberculosis* uses host triacylglycerol to accumulate lipid droplets and acquires a dormancy-like phenotype in lipid-loaded macrophages. PLoS Pathog. 2011; 7:e1002093. [PubMed: 21731490]
8. Low KL, et al. Triacylglycerol utilization is required for regrowth of *in vitro* hypoxic nonreplicating *Mycobacterium bovis* bacillus Calmette-Guerin. J. Bacteriol. 2009; 191:5037–5043. [PubMed: 19525349]
9. Russell DG, Mwandumba HC, Rhoades EE. *Mycobacterium* and the coat of many lipids. J. Cell Biol. 2002; 158:421–426. [PubMed: 12147678]
10. Robertson G, et al. Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. Nature Methods. 2007; 4:651–657. [PubMed: 17558387]
11. Mikkelsen TS, et al. Genome-wide maps of chromatin state in pluri potent and lineage-committed cells. Nature. 2007; 448:553–560. [PubMed: 17603471]
12. Johnson DS, Mortazavi A, Myers RM, Wold B. Genome-wide mapping of *in vivo* protein-DNA interactions. Science. 2007; 316:1497–1502. [PubMed: 17540862]
13. Ehrt S, et al. Controlling gene expression in mycobacteria with anhydrotetracycline and Tet repressor. Nucleic Acids Res. 2005; 33:e21. [PubMed: 15687379]
14. Ehrt S, Schnappinger D. Controlling gene expression in mycobacteria. Future Microbiol. 2006; 1:177–184. [PubMed: 17661663]
15. Klotzsche M, Ehrt S, Schnappinger D. Improved tetracycline repressors for gene silencing in mycobacteria. Nucleic Acids Res. 2009; 37:1778–1788. [PubMed: 19174563]

16. Farnham PJ. Insights from genomic profiling of transcription factors. Nature Rev. Genet. 2009; 10:605–616. [PubMed: 19668247]

17. MacQuarrie KL, Fong AP, Morse RH, Tapscott SJ. Genome-wide transcription factor binding: beyond direct target regulation. Trends Genet. 2011; 27:141–148. [PubMed: 21295369]

18. Galagan J, Lyubetskaya A, Gomes A. ChIP-Seq and the complexity of bacterial transcriptional regulation. Curr. Top. Microbiol. Immunol. 2013; 363:43–68. [PubMed: 22983621]

19. Chauhan S, Sharma D, Singh A, Surolia A, Tyagi JS. Comprehensive insights into *Mycobacterium tuberculosis* DevR (DosR) regulon activation switch. Nucleic Acids Res. 2011; 39:7400–7414. [PubMed: 21653552]

20. Gautam US, Chauhan S, Tyagi JS. Determinants outside the DevR C-terminal domain are essential for cooperativity and robust activation of dormancy genes in *Mycobacterium tuberculosis*. PLoS ONE. 2011; 6:e16500. [PubMed: 21304599]

21. Vasudeva-Rao HM, McDonough KA. Expression of the *Mycobacterium tuberculosis* acr-coregulated genes from the DevR (DosR) regulon is controlled by multiple levels of regulation. Infect. Immun. 2008; 76:2478–2489. [PubMed: 18391009]

22. Cho BK, Federowicz S, Park YS, Zengler K, Palsson BO. Deciphering the transcriptional regulatory logic of amino acid metabolism. Nature Chem. Biol. 2012; 8:65–71. [PubMed: 22082910]

23. Colangeli R, et al. The multifunctional histone-like protein Lsr2 protects mycobacteria against reactive oxygen intermediates. Proc. Natl Acad. Sci. USA. 2009; 106:4414–4418. [PubMed: 19237572]

24. Gordon BR, et al. Lsr2 is a nucleoid-associated protein that targets AT-rich sequences and virulence genes in *Mycobacterium tuberculosis*. Proc. Natl Acad. Sci. USA. 2010; 107:5154–5159. [PubMed: 20133735]

25. Rustad TR, Harrell MI, Liao R, Sherman DR. The enduring hypoxic response of *Mycobacterium tuberculosis*. PLoS ONE. 2008; 3:e1502. [PubMed: 18231589]

26. Kendall SL, et al. A highly conserved transcriptional repressor controls a large regulon involved in lipid degradation in *Mycobacterium smegmatis* and *Mycobacterium tuberculosis*. Mol. Microbiol. 2007; 65:684–699. [PubMed: 17635188]

27. Nesbitt NM, et al. A thiolase of *Mycobacterium tuberculosis* is required for virulence and production of androstenedione and androstadienedione from cholesterol. Infect. Immun. 2010; 78:275–282. [PubMed: 19822655]

28. Gao CH, Yang M, He ZG. Characterization of a novel ArsR-like regulator encoded by Rv2034 in *Mycobacterium tuberculosis*. PLoS ONE. 2012; 7:e36255. [PubMed: 22558408]

29. Gonzalo-Asensio J, et al. PhoP: a missing piece in the intricate puzzle of *Mycobacterium tuberculosis* virulence. PLoS ONE. 2008; 3:e3496. [PubMed: 18946503]

30. Gonzalo Asensio J, et al. The virulence-associated two-component PhoP-PhoR system controls the biosynthesis of polyketide-derived lipids in *Mycobacterium tuberculosis*. J. Biol. Chem. 2006; 281:1313–1316. [PubMed: 16326699]

31. Ryndak M, Wang S, Smith I. PhoP, a key player in *Mycobacterium tuberculosis* virulence. Trends Microbiol. 2008; 16:528–534. [PubMed: 18835713]

32. Abramovitch RB, Rohde KH, Hsu FF, Russell DG. aprABC: a *Mycobacterium tuberculosis* complex-specific locus that modulates pH-driven adaptation to the macrophage phagosome. Mol. Microbiol. 2011; 80:678–694. [PubMed: 21401735]

33. Singh A, et al. *Mycobacterium tuberculosis* WhiB3 maintains redox homeostasis by regulating virulence lipid anabolism to modulate macrophage response. PLoS Pathog. 2009; 5:e1000545. [PubMed: 19680450]

34. Ernst J, Vainas O, Harbison CT, Simon I, Bar-Joseph Z. Reconstructing dynamic regulatory maps. Mol. Syst. Biol. 2007; 3:74. [PubMed: 17224918]

35. Garton NJ, et al. Cytological and transcript analyses reveal fat and lazy persister-like bacilli in tuberculous sputum. PLoS Med. 2008; 5:e75. [PubMed: 18384229]

36. Park HD, et al. Rv3133c/dosR is a transcription factor that mediates the hypoxic response of *Mycobacterium tuberculosis*. Mol. Microbiol. 2003; 48:833–843. [PubMed: 12694625]

37. Baek SH, Li AH, Sassetti CM. Metabolic regulation of mycobacterial growth and antibiotic sensitivity. PLoS Biol. 2011; 9:e1001065. [PubMed: 21629732]

38. Cox JS, Chen B, McNeil M, Jacobs WR Jr. Complex lipid determines tissue-specific replication of *Mycobacterium tuberculosis* in mice. Nature. 1999; 402:79–83. [PubMed: 10573420]

39. Camacho LR, Ensergueix D, Perez E, Gicquel B, Guilhot C. Identification of a virulence gene cluster of *Mycobacterium tuberculosis* by signature-tagged transposon mutagenesis. Mol. Microbiol. 1999; 34:257–267. [PubMed: 10564470]

40. Converse SE, et al. MmpL8 is required for sulfolipid-1 biosynthesis and *Mycobacterium tuberculosis* virulence. Proc. Natl Acad. Sci. USA. 2003; 100:6121–6126. [PubMed: 12724526]

41. Domenech P, et al. The role of MmpL8 in sulfatide biogenesis and virulence of *Mycobacterium tuberculosis*. J. Biol. Chem. 2004; 279:21257–21265. [PubMed: 15001577]

42. Rousseau C, et al. Production of phthiocerol dimycocerosates protects *Mycobacterium tuberculosis* from the cidal activity of reactive nitrogen intermediates produced by macrophages and modulates the early immune response to infection. Cell. Microbiol. 2004; 6:277–287. [PubMed: 14764111]

43. Nazarova EV, et al. Role of lipid components in formation and reactivation of *Mycobacterium smegmatis* “nonculturable” cells. Biochemistry. 2011; 76:636–644. [PubMed: 21639843]

44. Ojha AK, et al. Growth of *Mycobacterium tuberculos* is biofilms containing free mycolic acids and harbouring drug-tolerant bacteria. Mol. Microbiol. 2008; 69:164–174. [PubMed: 18466296]

45. Ojha AK, Trivelli X, Guerardel Y, Kremer L, Hatfull GF. Enzymatic hydrolysis of trehalose dimycolate releases free mycolic acids during mycobacterial growth in biofilms. J. Biol. Chem. 2010; 285:17380–17389. [PubMed: 20375425]

46. Arnvig K, Young D. Non-coding RNA and its potential role in *Mycobacterium tuberculosis* pathogenesis. RNA Biol. 2012; 9:427–436. [PubMed: 22546938]

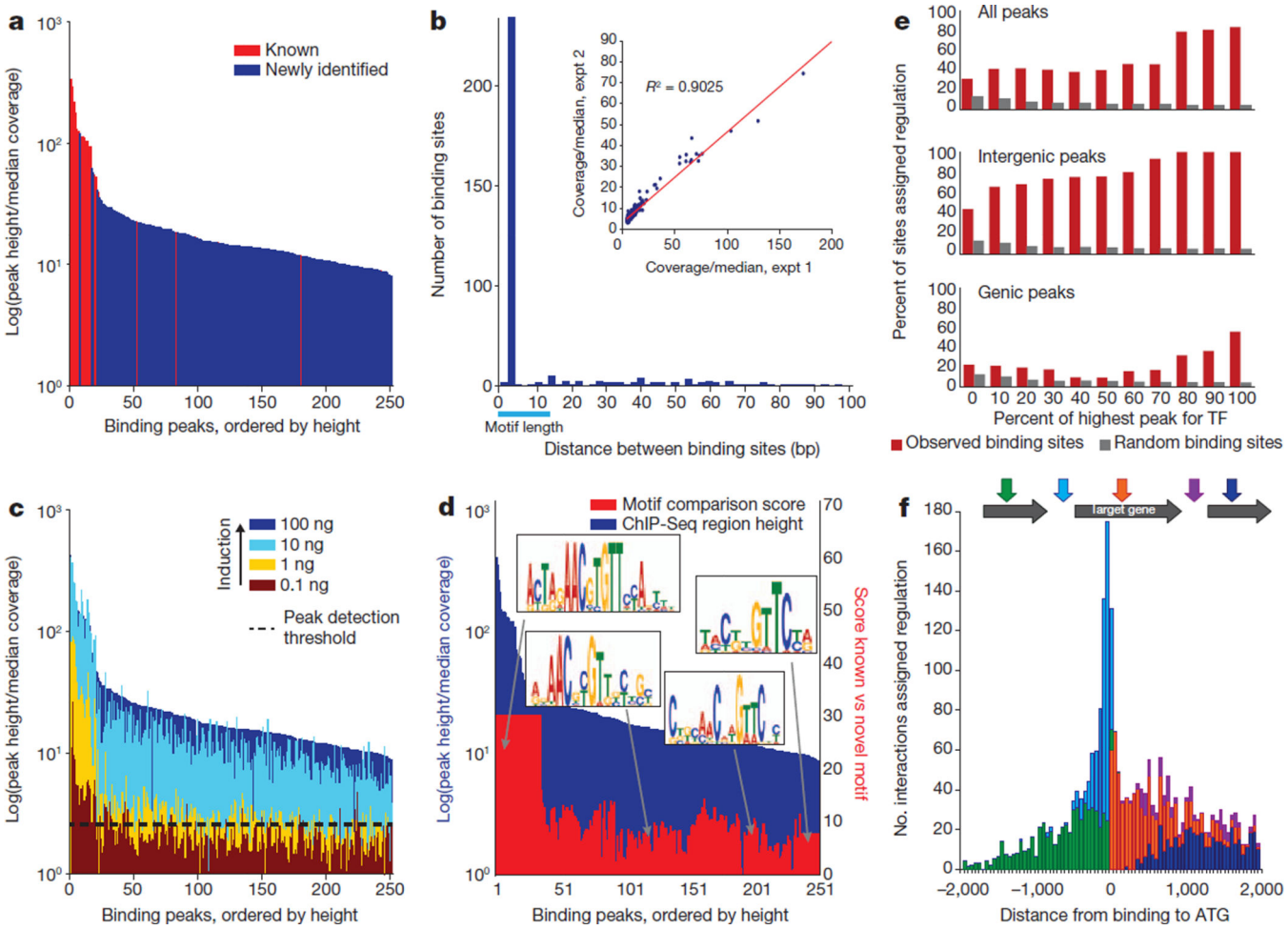

Figure 1. ChIP-Seq binding shows high sensitivity, reproducibility and sequence specificity

a, We identify all known binding sites (red bars) for KstR and DosR (Supplementary Fig. 3). Binding site heights plotted as bars and ordered by peak height. **b**, Binding site identification is highly reproducible. Bar plot shows the distance between corresponding sites in two KstR replicates. The majority of replicates fall within the motif (cyan line). Inset shows correlation of heights of corresponding peaks in two replicates ($R^2 > 0.83$ for all TFs). **c**, Increasing TF expression increases peak height. Shown are plots of peaks identified at different levels of KstR induction. Corresponding peaks are plotted at the same position on the horizontal axis. **d**, KstR binding peak height correlated with motif structure. The canonical palindromic motif is identified in all strong binding sites. At weaker sites, however, we detect degraded motifs. **e**, Fraction of peaks assigned regulation as a function of relative peak height. **f**, Stacked histogram of the number of peaks assigned regulation as a function of the distance to the start codon of the predicted target gene and coloured by genomic location relative to the target gene and genic or intergenic context.

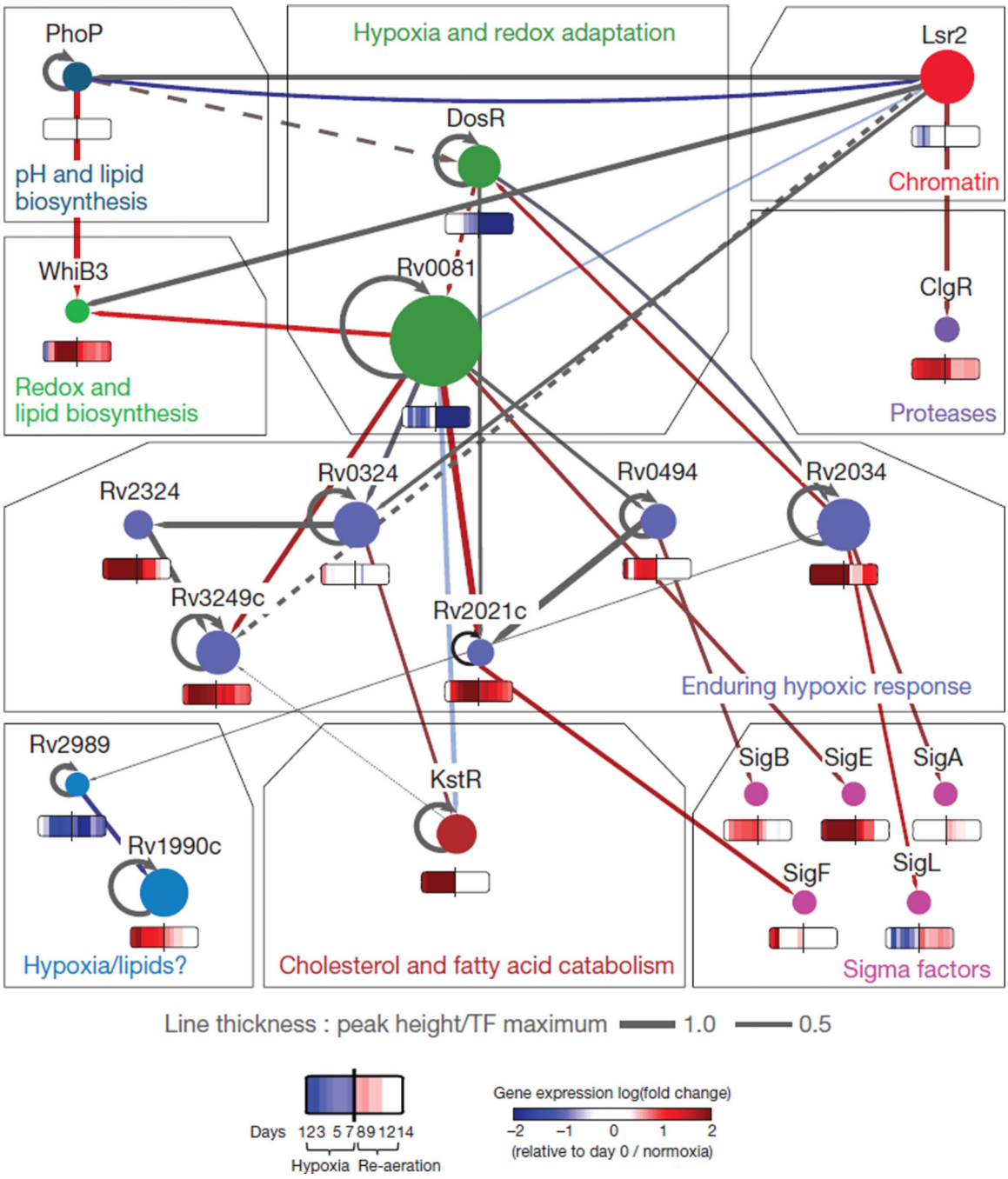

Figure 2. TF regulatory interaction subnet work linking hypoxia, lipid metabolism and protein degradation

The figure shows a subset of the regulatory network model for selected transcription factors. Edges are coloured by *z*-score (see text) with red edges indicating positive *z*-scores and activation, and blue indicating negative *z*-scores and repression. Grey edges indicate links without significant *z*-scores, TFs without induction expression data, or autobinding. The width of edges indicates the height of the corresponding binding site relative to the maximum binding site for the corresponding TF. Selected TFs are colour-coded by

functional association and heat maps show expression data during hypoxia and re-aeration as shown in legend.

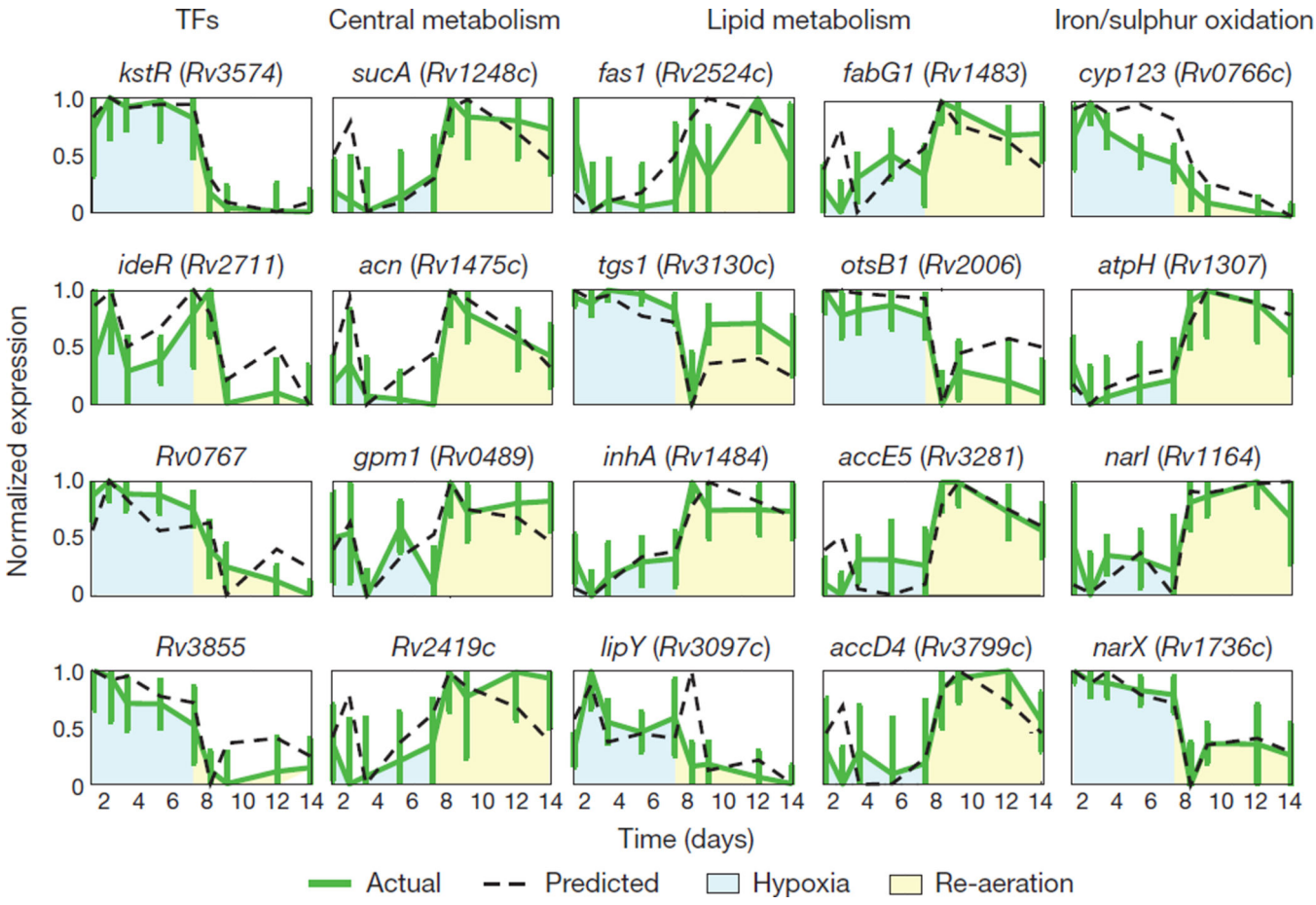

Figure 3. Predicting gene expression during hypoxia and re-aeration

Using the models described in text, we predict the expression pattern of 66% of genes (533) whose expression changes during hypoxia and re-aeration. Selected examples shown. Green lines, actual scaled expression with error bars from replicates; dashed black lines, model-predicted expression.

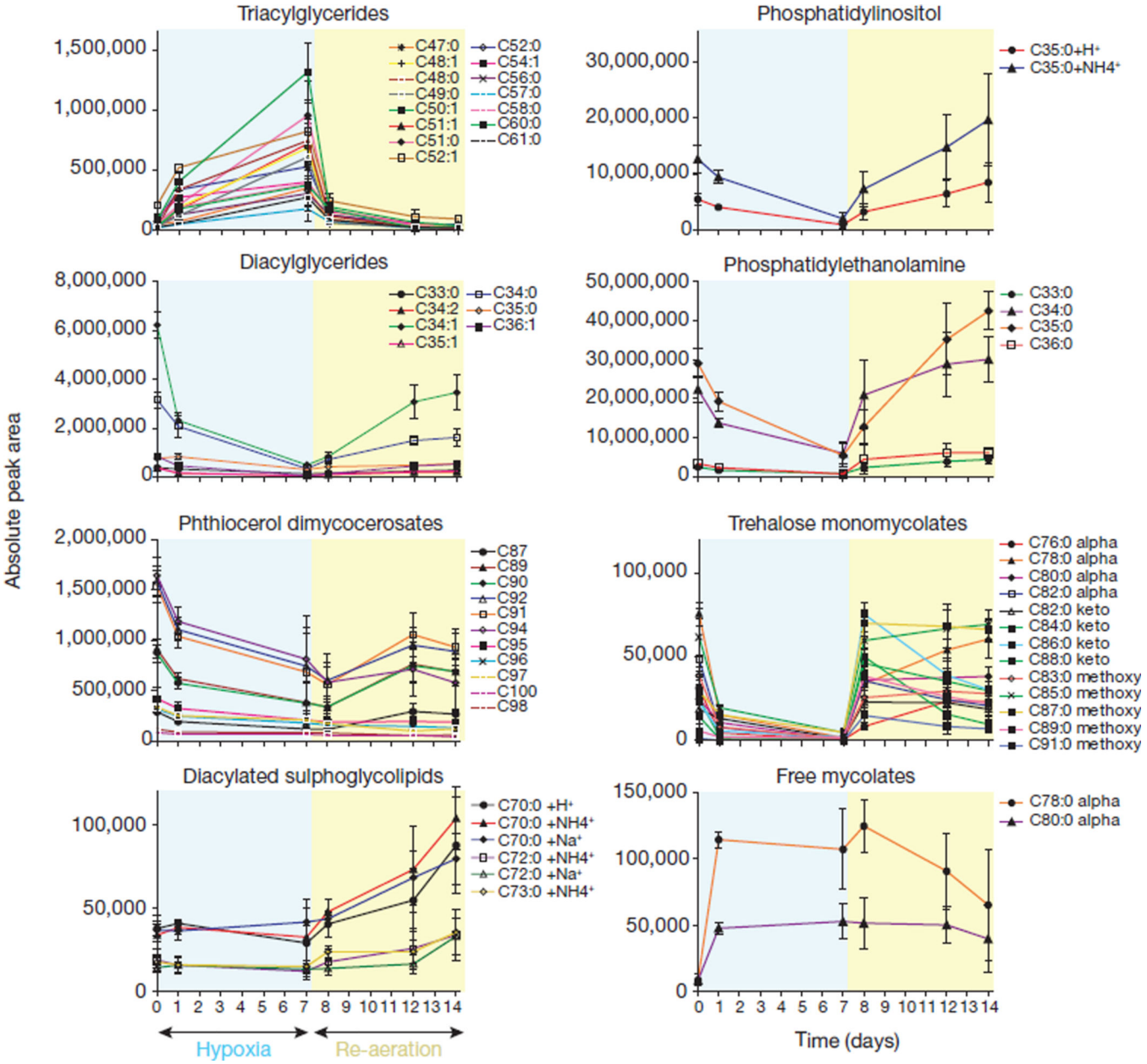

Figure 4. Lipid changes during hypoxia and re-aeration

HPLC-MS of total lipids from *M. tuberculosis* analysed in the positive-ion mode as ammoniated adducts unless otherwise indicated. Among more than 5,000 ions detected at each time point, *m/z* values for unnamed lipids were converted to named lipids when they matched the masses (< 10 p.p.m.) retention time (< 1 min) and collisional mass spectrometry patterns in MycoMass and MycoMap databases. Within each lipid class individual molecular species are reported by intensity and tracked by mass, converted to deduced empiric formulas and reported separately corresponding to the R group variants of mycolic acids (alpha, keto, methoxy) and as CX:Y, where X is the alkane chain length and Y is the

unsaturation in the combined fatty acyl, mycolyl, phthioceranyl, pthiocerol, mycocerosyl units of one molecule. Error bars are standard deviations from four replicates.