INFECTION AND IMMUNITY, Feb. 2006, p. 1233–1242 Vol. 74, No. 2
0019-9567/06/$08.00+0 doi:10.1128/IAI.74.2.1233–1242.2006

Unique Transcriptome Signature of *Mycobacterium tuberculosis* in Pulmonary Tuberculosis†

Helmy Rachman,[1] Michael Strong,[2] Timo Ulrichs,[1] Leander Grode,[1] Johannes Schuchhardt,[3] Hans Mollenkopf,[1] George A. Kosmiadi,[4] David Eisenberg,[2] and Stefan H. E. Kaufmann[1]*

Max Planck Institute for Infection Biology, Department of Immunology, Berlin, Germany[1]; Howard Hughes Medical Institute, UCLA-DOE Institute for Genomics and Proteomics, Molecular Biology Institute, University of California, Los Angeles, California[2]; MicroDiscovery GmbH, Berlin, Germany[3]; and Central Tuberculosis Research Institute, Department of Immunology 2, Moscow, Russian Federation[4]

Received 28 April 2005/Returned for modification 15 June 2005/Accepted 8 November 2005

Although tuberculosis remains a substantial global threat, the mechanisms that enable mycobacterial persistence and replication within the human host are ill defined. This study represents the first genome-wide expression analysis of *Mycobacterium tuberculosis* from clinical lung samples, which has enabled the identification of *M. tuberculosis* genes actively expressed during pulmonary tuberculosis. To obtain optimal information from our DNA array analyses, we analyzed the differentially expressed genes within the context of computationally inferred protein networks. Protein networks were constructed using functional linkages established by the Rosetta stone, phylogenetic profile, conserved gene neighbor, and operon computational methods. This combined approach revealed that during pulmonary tuberculosis, *M. tuberculosis* actively transcribes a number of genes involved in active fortification and evasion from host defense systems. These genes may provide targets for novel intervention strategies.

Mycobacterium tuberculosis results in approximately 8 million new cases of active tuberculosis and over 2 million deaths annually. Although many cases can be treated by chemotherapy, a dramatic increase in the emergence of multidrug-resistant (MDR) strains has been reported in numerous parts of the world (13, 18). To date, it is estimated that more than 50 million individuals are infected with MDR strains of *M. tuberculosis*. Coinfections with human immunodeficiency virus, which presently concern more than 50 million individuals, have also dramatically exacerbated disease development, increasing the annual risk of developing active tuberculosis to 1 in 10 (13, 18).

Currently, approximately 2 billion individuals (one-third of the total world population) are infected with *M. tuberculosis*. Most of these individuals harbor the pathogen in a dormant stage and will not develop active disease (18). The outbreak of the disease is typically accompanied by the liquefaction and caseation of granulomatous lesions. These lesions can contain large numbers of bacteria ($>10^9$ organisms), and rupture of these lesions can result in the dissemination of bacteria to other organs through the circulatory system and contagious spreading to the environment through the alveolar system.

Previous efforts to elucidate the mechanisms of *M. tuberculosis* survival within the host have employed animal models or murine bone-marrow-derived macrophages. Global expression profiling of *M. tuberculosis* within murine bone-marrow-derived macrophages revealed that the local environment within macrophages is likely fatty acid rich, DNA and cell wall damaging, and iron deficient (29). A recent examination of *M. tuberculosis* gene expression using reverse transcription (RT)-PCR, however, revealed differences between the gene expression levels of *M. tuberculosis* isolates obtained from mice and humans (37). Although only a subset of *M. tuberculosis* genes were examined, the differences in gene expression levels revealed distinct differences between the pulmonary tissue environments of humans and mice. In order to gain a more insightful picture of the transcriptome of *M. tuberculosis* during human disease, it is necessary therefore to examine the genome-wide transcription profiles of *M. tuberculosis* isolates obtained from pulmonary tuberculosis patients.

The increasing incidence of MDR strains of *M. tuberculosis* in several parts of Russia has rendered surgical lung resection an unavoidable measure of tuberculosis treatment. We took advantage of surgical resection of this highly affected pulmonary tissue from MDR tuberculosis patients to determine the gene expression profiles of *M. tuberculosis* at three different sites of pulmonary infection.

MATERIALS AND METHODS

Lung tissue. Tuberculosis patients suffering from extensive tuberculosis lung disease (often MDR) underwent surgery at the Central Tuberculosis Research Institute in Moscow, Russia. The use of the resected tissue for further immunological and genetic analyses has been approved by ethics committees in both Moscow and Berlin. Tissue was used for this study only after informed consent was given by the patient. Following surgery, the tissue was separated into three different types: granuloma, pericavital tissue (pericavity), and macroscopically normal lung tissue from distant parts of the removed tissue (distant lung) (Fig. 1). In general, not all three types of tissue could be obtained from each patient. Therefore, we decided to take each sample from different patients for DNA array analysis. This method also guarantees diversified statistical data. The samples in one group were considered replicates. Ten tuberculosis patients who had

* Corresponding author. Mailing address: Max Planck Institute for Infection, Immunology, Schumannstrasse 21-22, Berlin 10117, Germany. Phone: 49-30-28460500. Fax: 49-30-28460501. E-mail: kaufmann@mpiib-berlin.mpg.de.

† Supplemental material for this article may be found at http://iai.asm.org/.

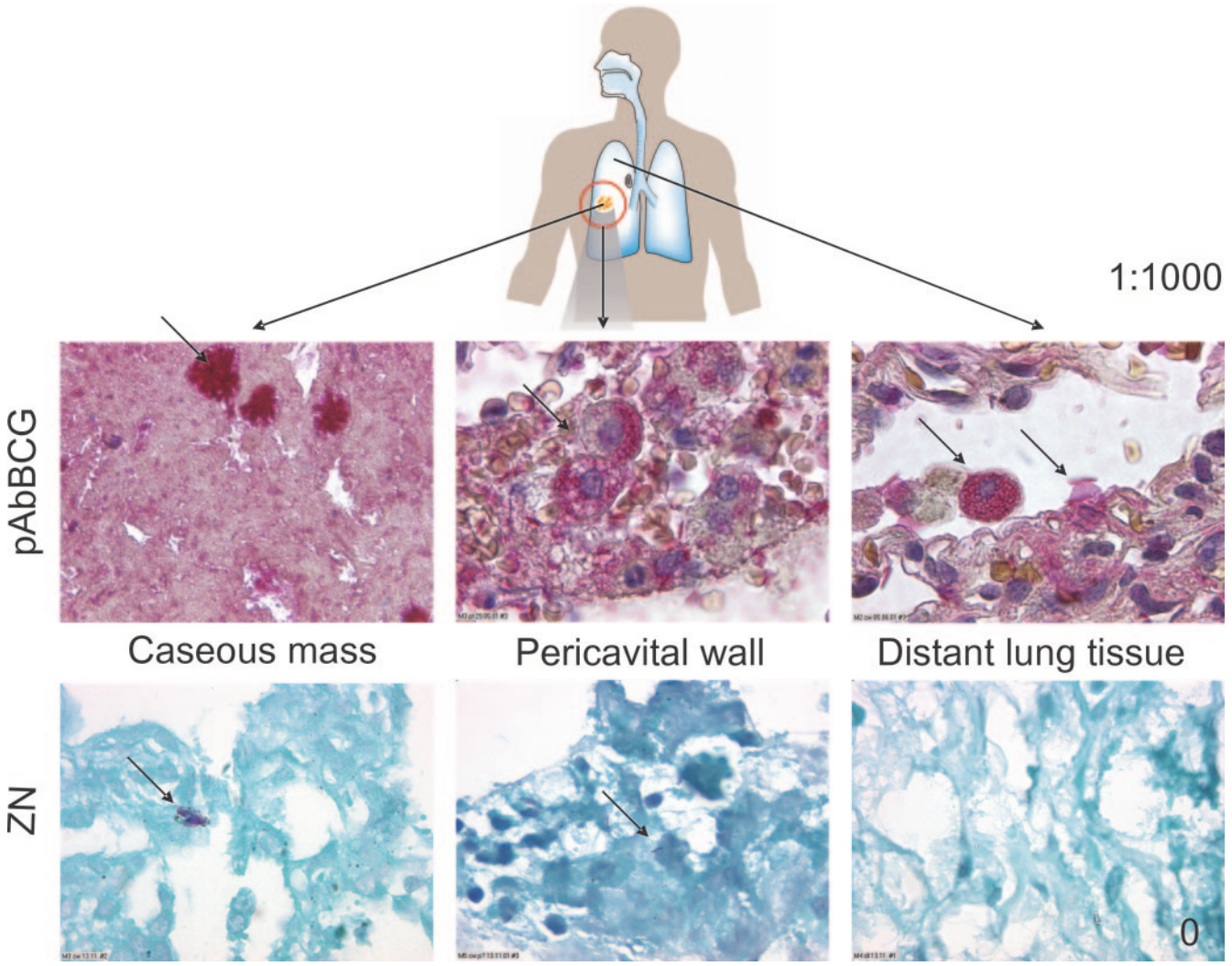


FIG. 1. Tissue samples from different regions of the human lung. The tissue samples shown were stained for *M. tuberculosis* by the Ziehl-Neelsen (ZN) staining method or with polyclonal serum raised against *M. bovis* BCG (pAbBCG). The three regions of the lung examined were the caseous mass of the granuloma, the pericavital tissue, and tissue of the distant lung. Only in the caseous mass of the granuloma, and to a lesser extent the pericavital tissue, were mycobacteria ZN positive, although all three regions harbored abundant mycobacteria.

received chemotherapy were included in this study, and one patient suffering from lung cancer was included as a tissue control. The patients had a mean age of 39 ± 15.3 years and a male/female ratio of 2:1. CFU analyses of the removed tissue revealed a mean density of 6.3×10^8/g (minimum, 5.5×10^3/g; maximum, 3.7×10^9/g) of *M. tuberculosis* bacteria. All tissue specimens were analyzed histologically for the presence of *M. tuberculosis* bacteria by Ziehl-Neelsen staining and by immunohistological staining using a polyclonal serum raised against *Mycobacterium bovis* BCG (31).

In vitro culture of bacteria and culture conditions. Immediately after removal from patient lungs, the samples of the human lung granuloma, pericavity, and distant lung were homogenized. An aliquot of the granuloma homogenate was inoculated into Middlebrook 7H9 medium supplemented with 10% (vol/vol) albumin-dextrose-catalase enrichment (Difco Laboratories) and 0.05% (vol/vol) Tween 80 (Sigma). The culture was grown to mid-log phase at 37°C with shaking in screw-cap bottles. Only the granuloma samples, in which the tubercle bacilli were able to be cultivated, were chosen for DNA array analysis. The other aliquot was immersed in TRIzol (Invitrogen) and immediately processed further for RNA extraction. Genotyping analysis revealed that all *M. tuberculosis* strains used in this study showed the characteristics of the Beijing strain.

RNA extraction. The cell pellet from 50 ml of each mid-log-phase culture was resuspended in 5 ml phenol and 5 ml chloroform-methanol (3:1) and vortexed for 1 min or until the formation of an interface occurred. RNA was extracted with 4 ml RLT buffer from the RNeasy Kit (QIAGEN) containing 0.5% sarcosyl and 1% β-mercaptoethanol (added prior to the use of buffer). The suspension was centrifuged to separate the aqueous phase from the organic phase. The aqueous layer was precipitated in ethanol, and RNA was redissolved in 400 μl RLT buffer and further purified using RNeasy columns (QIAGEN) according to the manufacturer's instructions. The size distribution and the quantity of the isolated total RNA samples were determined using an Agilent 2100 bioanalyzer (Agilent) high-resolution electrophoresis system. The RNA samples were first diluted with injection buffer according to the manufacturer's instructions and then analyzed in parallel with an external RNA 6000 size ladder (Ambion).

cDNA synthesis and labeling. Total RNA (500 ng) from in vitro culture or 8 μg total RNA from patient lung samples was mixed with 5 pmol mycobacterial-genome-directed primers (35) and heated to 70°C for 10 min, followed by immediate cooling to 4°C (on ice). For negative controls, total RNA isolated from the tissue of a lung cancer patient and commercial total human RNA (Ambion) were applied in equal amounts. RNA was then reverse transcribed using a RevertAid first-strand cDNA synthesis kit (Fermentas) in a 20-μl reaction mixture containing 200 U RevertAid Moloney murine leukemia virus reverse transcriptase; 20 U RNase inhibitor; 0.5 mM dATP, dGTP, and dTTP; 0.5 μM dCTP; and 50 μCi of [α-^{33}P]dCTP (Amersham Biosciences) in 1× reverse transcription buffer. The reaction was carried out at 25°C for 10 min and subsequently at 42°C for 2 h. The unincorporated radioactive nucleotide was re-

moved from the labeled cDNA by use of a G-50 column (Amersham Biosciences).

Hybridization of DNA arrays. DNA arrays were prehybridized at 50°C for at least 1 h in 10 ml ULTRArray hybridization buffer (Ambion). The radioactively labeled cDNA probe was denatured at 95°C for 10 min, cooled immediately on ice for 2 min, and then transferred to the hybridization buffer. The hybridization was performed at 50°C overnight. The DNA arrays were washed three times in washing buffer (0.5% sodium dodecyl sulfate, 0.1× SSC [1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate]) at 55°C for 30 min each. After washing, the DNA arrays were wrapped with clear wrap film and exposed to a phosphorimaging screen for 4 days. The array images were scanned using a Fuji BAS 2500 phosphorimaging instrument at 50-μm-pixel resolutions.

Data analysis of DNA arrays. The feature extraction from the resulting image files and data analysis were carried out using the customized software "Genespotter" (MicroDiscovery, Berlin, Germany) (30). Hierarchical clustering of DNA array data was performed using hierarchical clustering methods described previously (12). The correspondence analysis of DNA array data was performed as described previously (14). Since the hybridization of DNA arrays with the negative control sample resulted in signals that were only about 8% detectable (signals with intensities higher than two times those of the background signals), it was impossible to extract the features from the DNA array image properly. For this reason, the effects of host total RNA in the lung samples of the tuberculosis patients used in this study were ignored. The DNA array data were normalized as follows. The raw signal intensities from the image extraction of each array were logarithmized. After the exclusion of negative controls, the average value of the data set from each array was calculated and used to normalize the data set. Since each open reading frame was represented by two spots on each array, and there were at least three RNA samples for each experimental condition, at least six normalized values were obtained from each experimental condition for each open reading frame. The average value of this data set was calculated and used for further data analysis. In addition, the data sets were analyzed using a program for significance analysis of microarrays (38). A false discovery rate of $<1\%$ was applied in this analysis. The genes were considered upregulated and used for protein linkage analysis when the upregulation was >3-fold. For the comparison of pericavity and distant-lung data, the cutoff was set to twofold upregulation. The overlapping results from both analysis methods are available at http://www.doe-mbi.ucla.edu/strong/kaufmann and in supplemental tables S8 to S14.

Rosetta stone method. Proteins were functionally linked by the Rosetta stone method when individual proteins were found to be present as a single fusion protein in another organism (22, 23). When individual *M. tuberculosis* proteins have significant homologies to distinct regions of a single fusion protein in another organism, they are indicated as functionally linked by this method.

Phylogenetic profile method. Phylogenetic profiles were used to identify proteins that occurred in a correlated manner in multiple genomes (25). A phylogenetic profile for each *M. tuberculosis* protein was created in the form of a bit vector by searching for the presence or absence of homologs in each of the available fully sequenced genomes. The presence of an identifiable homolog in a particular genome was indicated by the integer 1 in the bit vector at the position corresponding to that genome, while the absence of a homolog was indicated by the integer 0. Phylogenetic profiles were then clustered based on the similarity of the profiles.

Conserved gene neighbor method. Functional links were established by the conserved gene neighbor method in cases where genes appeared as chromosomal neighbors in multiple genomes (9, 24). For all possible pairs of *M. tuberculosis* genes, the nucleotide distances between the homologs of these genes in all available sequenced genomes were calculated. Genes that were in close proximity in multiple genomes were indicated as functionally linked by this method.

Operon method. A series of genes are considered functionally linked by the operon method if the nucleotide distance between genes in the same orientation is less than or equal to a specified distance threshold. Multiple genes are linked if a series of genes in the same orientations all have intergenic distances less than or equal to the defined distance threshold (33, 34).

Validation of computationally assigned functional linkages. To evaluate the quality of functional linkages inferred by two or more computational methods, we compared the Swiss-Prot keywords of all linked pairs of annotated genes. The percentage of gene pairs that had at least some function in common was calculated as the number of annotated gene pairs that had at least one Swiss-Prot keyword in common divided by the total number of annotated gene pairs. An annotated gene pair is defined as a pair of genes that have functions assigned in the form of Swiss-Prot keywords to each of the genes of the pair. Comparison of the Swiss-Prot keywords revealed that 80% of the annotated gene pairs have at least one keyword in common (ignoring the keywords "hypothetical protein," "3D structure," "transmembrane," and "complete proteome"), and therefore 80% of the annotated gene pairs have some function in common. Although 20% of annotated gene pairs do not have any Swiss-Prot keywords in common, this may arise from a number of factors, including incomplete annotation or incomplete functional characterization. We infer from these data that at least 80% of the gene pairs linked by two or more computational methods have some function in common.

Protein networks. Protein networks were constructed using functional linkages between proteins that are inferred by two or more computational methods. Networks were constructed using the NetPlot web utility (http://www.doe-mbi.ucla.edu/~morgan/NETPLOT/). Upregulated genes are indicated as red nodes within the networks. Keyword comparisons between Swiss-Prot annotated proteins have been made previously to assess the reliability of computationally inferred linkages and to assess biochemical experiments such as yeast two-hybrid experiments (23). Linkages inferred by two or more methods have been shown to be of a quality similar to that of experimental interaction data.

RESULTS AND DISCUSSION

Transcriptome comparisons of *M. tuberculosis* isolates from different tissue sites. Due to the scarcity of available RNA from *M. tuberculosis* isolated from tuberculosis patients, it was difficult to repeat the hybridization experiments for a given sample. We overcame this obstacle by performing hybridization experiments using RNA samples from the same tissue sites of at least three different tuberculosis patients. We analyzed the DNA array data sets using hierarchical clustering algorithms (12) and correspondence analysis (14). Figure 2A depicts the hierarchical clustering of 13 DNA array data sets obtained from this study. Three distinct clusters can be observed. Although the in vitro growth data sets are quite distinct from the granuloma, pericavity, and distant-lung data sets, there is considerable overlap among the pericavity and distant-lung data sets. This result indicates that the environments in the pericavity and the distant lung are similar. This correlates well with the histological data of *M. tuberculosis* bacteria at the investigated sites (39). Whereas most tubercle bacilli are extracellular in the center of the granuloma, they reside within macrophages in the pericavity and the distant lung. The functional categories of upregulated genes are shown in Fig. 3 and in Tables S1 through S7 in the supplemental material.

Figure 2B shows planar embedding of the 13 DNA array data sets based on correspondence analysis (14). The result of this method confirms that of the hierarchical clustering. Based on this assessment, we conclude that gene expression profiles of *M. tuberculosis* from the same sites of infection but from different patients are highly similar and can thus be reliably treated as experimental duplicates. The lists of genes that are upregulated during human infection are available at http://www.doe-mbi.ucla.edu/~strong/kaufmann and in Tables S8 through S14 in the supplemental material.

Signatures inferred from DNA array results. Figure 3 summarizes the functional categories of *M. tuberculosis* genes that are upregulated under four different growth conditions. The percentages were calculated from previously assigned functional annotations (7). A high proportion of differentially expressed genes under all growth conditions investigated was associated with the biosynthesis of the components of the cell envelope (Fig. 3). Although *M. tuberculosis* induces similar percentages of genes involved in cell envelope metabolism in vivo (10% for the granuloma and 9.6% for the pericavity/distant lung) and in vitro (9.7%), different genes were induced in vivo and in vitro, as identified by our DNA array experi-

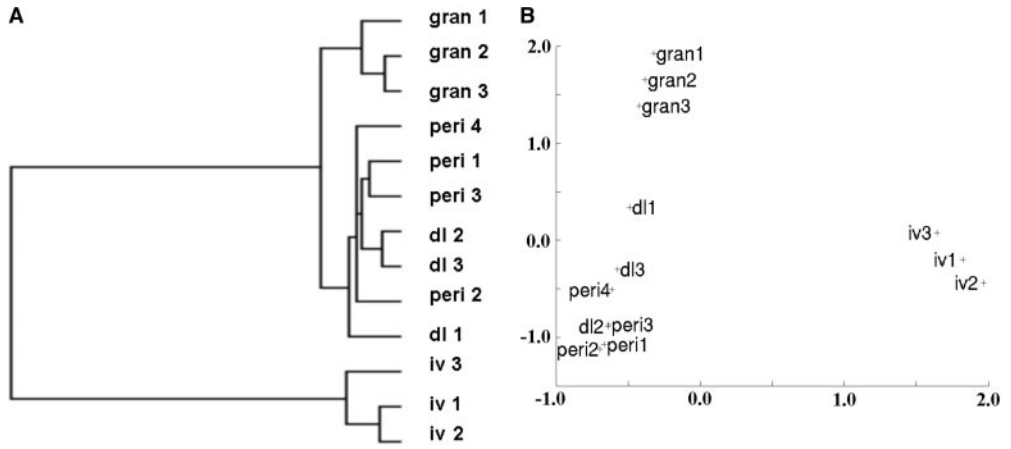


FIG. 2. Hierarchical clustering and correspondence analysis of DNA array data. (A) Dendrogram obtained from the hierarchical clustering of all genes of *M. tuberculosis* gene expression data generated from 13 data sets. Four different growth environments were assayed: in vitro culture (iv), granuloma (gran), distant lung (dl), and pericavity (peri). The numbers indicate different patients or in vitro cultures from which the *M. tuberculosis* RNA samples were isolated. The average linkage hierarchical clustering algorithm was used to construct the dendrogram. (B) Result of correspondence analysis of the 13 data sets. Data sets from each growth environment appear similar and were therefore treated as experimental duplicates. Moreover, the data sets from the pericavity and distant-lung sites are highly correlated, indicating similar growth environments at the two infection sites.

ments (see Tables S8 through S14 in the supplemental material). We infer that *M. tuberculosis* undergoes immense changes in cell envelope composition upon infection of human lungs. Recent observations have revealed that *M. tuberculosis* bacteria isolated from macrophages have a unique morphology, existing as small ovoid and coccoid forms with intact permeability barriers (3). The isolated bacteria also exhibited alterations in normal cell surface characteristics, such as alterations in adsorption of bacteriophage, adherence, and the ability to take up stains (3). We observed that twice the proportion of lipid biosynthesis genes was upregulated in vivo than in vitro (Fig. 3). In addition, we observed that a series of fatty and mycolic acid modification genes, including two of the three desaturase genes (Rv3229c and Rv0824c), were induced in vivo (Fig. 4A). Since some lipids and some cell wall components of *M. tuberculosis* can modulate immune responses (1, 2), these observations suggest that the pathogen may mobilize mechanisms aimed at evasion from host immune responses by modifying lipid and cell wall components. Moreover, Rv3229c (*desA3*) has been reported to be the target of the drug thiourea isoxyl, which is known to be active against some MDR strains of *M. tuberculosis* (26).

A similar observation was noticed for the genes of the PE and PPE families (Fig. 3). Some of the PE and PPE members are known to be highly immunogenic. We observed the induction of one of the best-studied PE-PGRS proteins, Rv1818c, at all sites of pulmonary tuberculosis. This observation corresponds well with the results of the previous study, in which a significant immune response to the PGRS domain of Rv1818c in infected mice was observed (10). While the exact function of these proteins is still a matter of debate, the PE-PGRS proteins have been found to localize at the mycobacterial cell surface and hence may participate in cell surface interactions.

The proportions of genes involved in detoxification and encoding chaperones were also higher in vivo than in vitro (Fig. 3 and 4A), indicating highly toxic and stressful conditions at the sites of pulmonary tuberculosis. In addition to their function as stress response proteins, some chaperones have been proposed as potential virulence determinants and may play an important role in inducing the host inflammatory response (19). Noticeable was the induction of two genes encoding putative glyoxylase II enzymes (Rv0634c and Rv2581c) (Fig. 4A), which suggests that the environments at the sites of pulmonary tuberculosis may be rich in alpha-ketoaldehydes, especially methylglyoxal, and that mycobacteria may be susceptible to this group of small molecules.

A number of genes that are associated with DNA repair and modification, such as *dinX* (Rv1537), *dinF* (Rv2836c), *gyrA* (Rv0004), *gyrB* (Rv0005), and a series of insertion elements, were also upregulated during pulmonary tuberculosis (Fig. 4A). A significant proportion of insertion elements and transposases was recently reported to be highly induced in *M. tuberculosis* during growth in environments promoting DNA damage (5). This may indicate that the environment of *M. tuberculosis* in vivo is rich in DNA-damaging agents produced by host cells due to an effort to eradicate the bacilli. A similar environment was also observed during *M. tuberculosis* infection of macrophages and mice (24).

Some genes involved in the transport of amino acids were upregulated in vivo (Fig. 4A), indicating a condition of nutrient starvation in vivo. Nutrient deprivation may contribute to the host defense by depriving the tubercle bacilli of necessary nutrients. The concentration of L-tryptophan has been suggested to influence the intracellular survival of *Bordetella pertussis*, and the gamma interferon-mediated induction of tryptophan-degrading enzymes has been hypothesized to serve as a host defense mechanism against *B. pertussis* (21).

We observed the in vivo upregulation of a series of genes that are involved in anaerobic respiration, including nitrate reductase genes. Nitrate reduction has been proposed as a marker for the hypoxic transition of *M. tuberculosis* (41). Figure 4B shows that more genes involved in anaerobic respira-

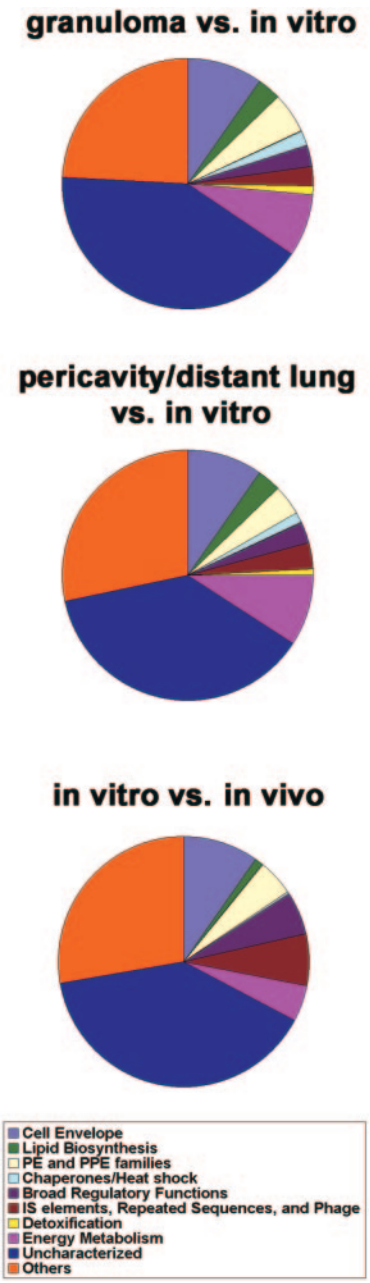


FIG. 3. Functional categories of *M. tuberculosis* genes that are upregulated during human infection. Pie graphs represent *M. tuberculosis* genes that are upregulated in (top) granuloma versus in vitro, (middle) the pericavity and distant lung versus in vitro, and (bottom) in vitro versus in vivo (in the granuloma, pericavity, and distant lung).

tion were induced in the pericavity and the distant lung and that some genes were induced more strongly at these sites than in the granuloma. *narX* and *frdA* were recently reported to be induced in tubercle bacilli residing in activated macrophages (29). Our DNA array results showed that *narX* was induced in the granuloma and that *narX*, *narG*, and *frdA* were upregulated in the pericavity and the distant lung. We also observed the upregulation of a panel of genes involved in aerobic respiration (Fig. 4B). Some of these genes were more strongly induced in the granuloma than in the pericavity and the distant lung. Tubercle bacilli in our study were isolated from patients with active tuberculosis, in whom the granuloma and surrounding tissue eroded through the bronchial wall. In such cases, the liquefied material is usually discharged into an airway, and a cavity is formed in the lung. Oxygen and carbon dioxide are then able to freely enter the cavity. Due to the reactive and small-molecular nature of oxygen, a proportion of this gas may diffuse from the center of the granulomas to the pericavity and the distant lung. The induction of both aerobic and anaerobic respiration systems supports the idea of a microaerophilic environment in vivo during active tuberculosis. In addition, the expression profiles of tubercle bacilli in macrophages of the pericavity and the distant lung have features which may be more consistent with a response to hypoxia than those of the granuloma.

Functional linkages and protein networks. Figure 3 shows that in all categories, the largest percentages of upregulated genes are annotated as either "hypothetical proteins" or "hypothetical conserved proteins" and are thus listed as "uncharacterized." To aid the identification of potential functions for these proteins, we constructed protein networks based on computationally inferred functional linkages. For each of the genes that we identified as differentially expressed, we examined functional linkages established by the Rosetta stone (22, 23), phylogenetic profile (25), conserved gene neighbor (9, 24), and operon (33, 34) computational methods. The Rosetta stone method identifies genes that occur as a single fusion gene in another organism, the phylogenetic profile method identifies genes that have correlated occurrences across numerous genomes, the conserved gene neighbor method identifies genes that occur in close chromosomal proximity in numerous genomes, and the operon method identifies genes likely to belong to common operons based on the distances between genes in the same genomic orientation. These linkages are available at http://www.doe-mbi.ucla.edu/~strong/kaufmann.

Protein networks containing both annotated and nonannotated genes that are upregulated during pulmonary tuberculosis were constructed. Protein networks provide a framework for upregulated genes within the context of other genes to which they are functionally linked and were used to infer the function of previously uncharacterized *M. tuberculosis* genes.

The protein networks depicted in Fig. 5 represent functional linkages among *M. tuberculosis* proteins inferred by two or more of the computational methods mentioned above. Each protein is represented by a single node within the illustration, and each line represents a functional linkage between a pair of protein nodes. Although the use of linkages inferred by two or more methods decreases the coverage of possible functional linkages, the accuracy increases (33, 34). *M. tuberculosis* genes that are upregulated during human infection are also indicated in the networks in Fig. 5. Proteins that are functionally linked may represent members of a common macromolecular complex, a biochemical pathway, or proteins of related functions.

In Fig. 5A, the uncharacterized gene Rv0459 is linked to a network containing eight other genes involved in aldehyde dehydrogenase activity. Although the upregulated gene Rv0459 does not have any sequence homology to any genes of known function, we infer that Rv0459 has a functional relationship to the proteins of this network, specifically with the Rv0458-encoded aldehyde dehydrogenase. Formaldehyde de-

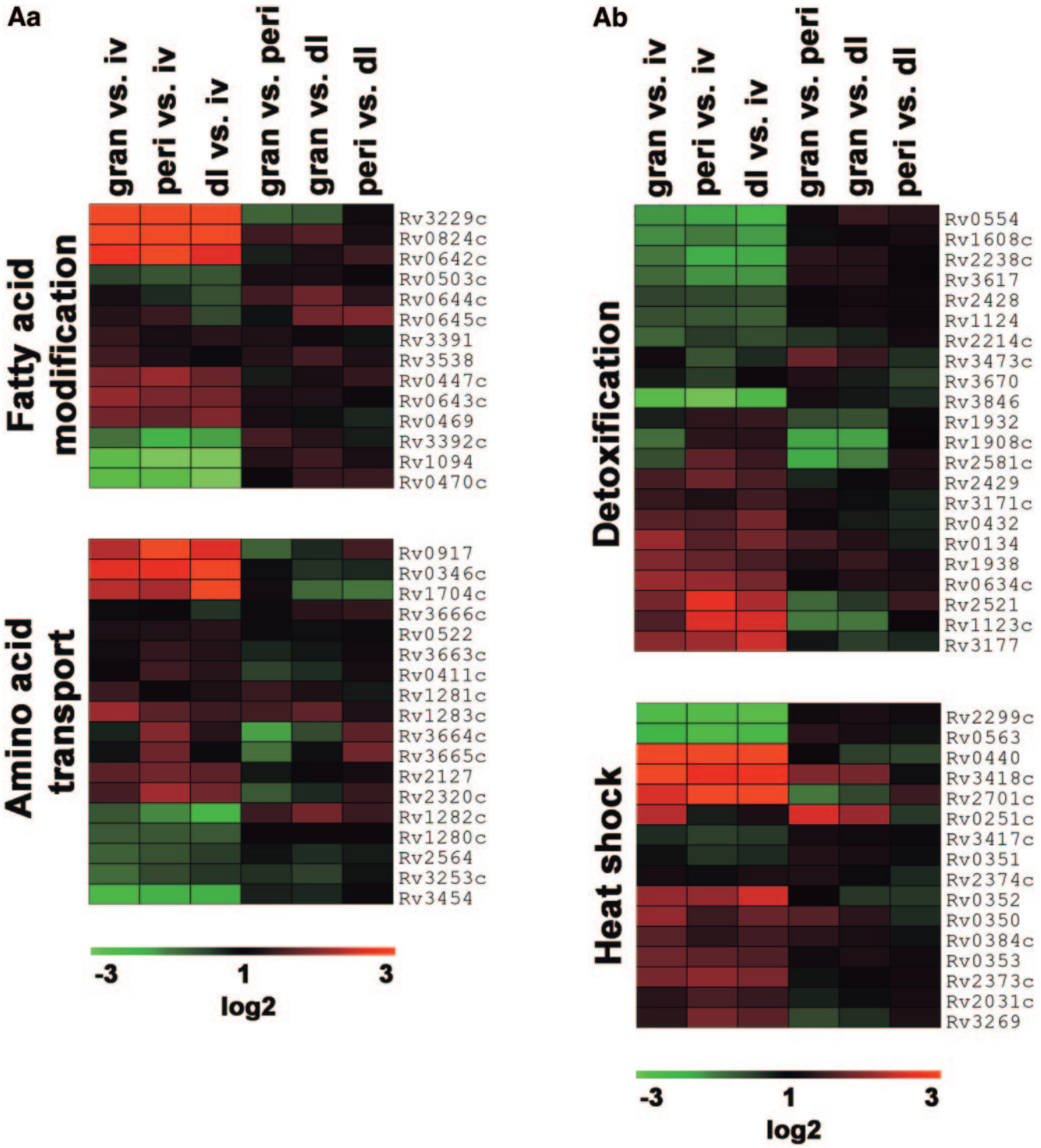
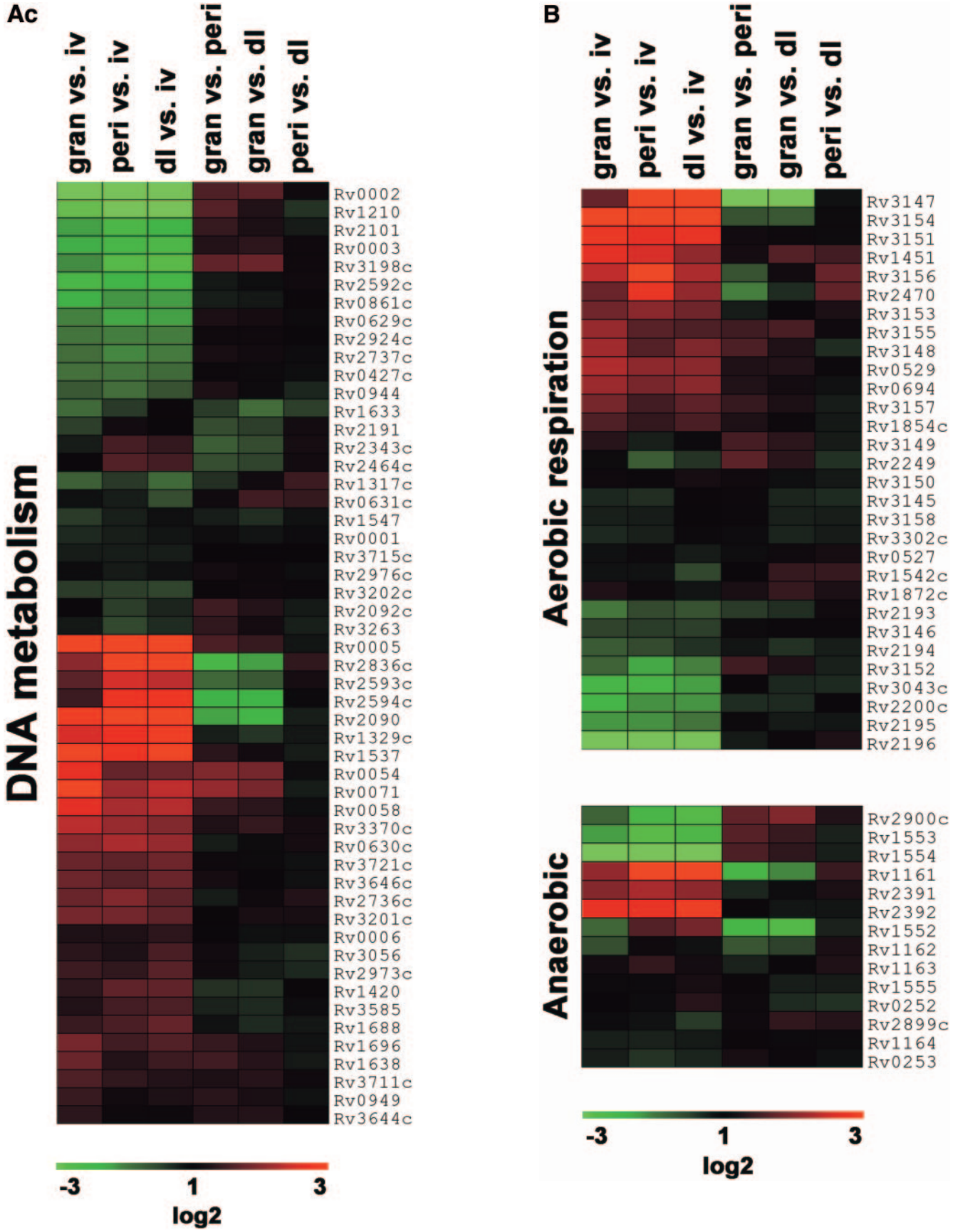


FIG. 4. Red-green display summarizing the regulation of selected genes at the sites of pulmonary tuberculosis. Genes were selected based on the categories described by Cole et al. (7). Log_2 transformations of the averaged cDNA ratios of the experimental conditions (in vitro [iv], granuloma [gran], pericavity [peri], distant lung [dl]) indicated at the top of each panel is displayed according to the color scale at the bottom of the panel. Genes were selected according to their functional categories as described previously and ordered based on the average linkage hierarchical clustering. (Aa) Modification of fatty and mycolic acids; (Ab) chaperones/heat shock; (Ac) DNA replication, repair, recombination, and restriction/modification; (B) aerobic and anaerobic respiration.


Ac
gran vs. iv
peri vs. iv
dl vs. iv
gran vs. peri
gran vs. dl
peri vs. dl
DNA metabolism
Rv0002
Rv1210
Rv2101
Rv0003
Rv3198c
Rv2592c
Rv0861c
Rv0629c
Rv2924c
Rv2737c
Rv0427c
Rv0944
Rv1633
Rv2191
Rv2343c
Rv2464c
Rv1317c
Rv0631c
Rv1547
Rv0001
Rv3715c
Rv2976c
Rv3202c
Rv2092c
Rv3263
Rv0005
Rv2836c
Rv2593c
Rv2594c
Rv2090
Rv1329c
Rv1537
Rv0054
Rv0071
Rv0058
Rv3370c
Rv0630c
Rv3721c
Rv3646c
Rv2736c
Rv3201c
Rv0006
Rv3056
Rv2973c
Rv1420
Rv3585
Rv1688
Rv1696
Rv1638
Rv3711c
Rv0949
Rv3644c
-3
1
3
log2
B
gran vs. iv
peri vs. iv
dl vs. iv
gran vs. peri
gran vs. dl
peri vs. dl
Aerobic respiration
Rv3147
Rv3154
Rv3151
Rv1451
Rv3156
Rv2470
Rv3153
Rv3155
Rv3148
Rv0529
Rv0694
Rv3157
Rv1854c
Rv3149
Rv2249
Rv3150
Rv3145
Rv3158
Rv3302c
Rv0527
Rv1542c
Rv1872c
Rv2193
Rv3146
Rv2194
Rv3152
Rv3043c
Rv2200c
Rv2195
Rv2196
Anaerobic
Rv2900c
Rv1553
Rv1554
Rv1161
Rv2391
Rv2392
Rv1552
Rv1162
Rv1163
Rv1555
Rv0252
Rv2899c
Rv1164
Rv0253
-3
1
3
log2

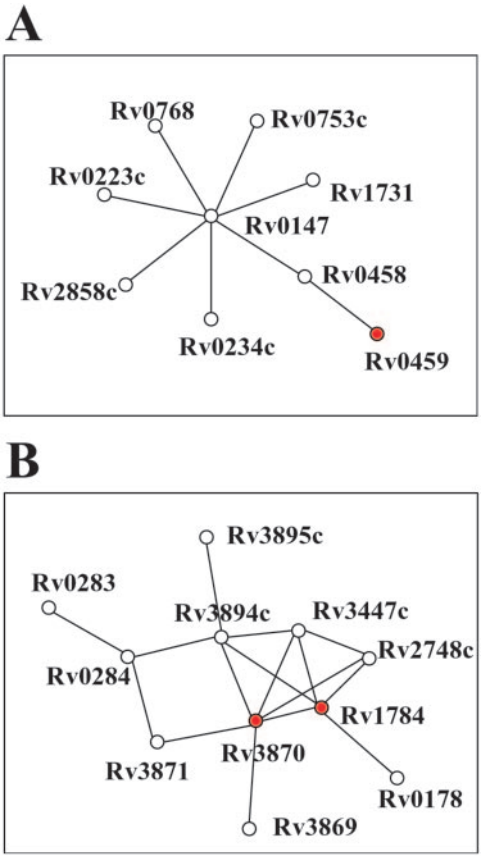


FIG. 5. *M. tuberculosis* protein networks containing genes that are upregulated during pulmonary tuberculosis. Protein networks were constructed using computationally assigned functional linkages inferred by the Rosetta stone, phylogenetic profile, operon, and conserved gene neighbor methods. Functional linkages inferred by two or more methods are indicated. Each node (circle) in the network represents a particular *M. tuberculosis* gene/protein, with in vivo-upregulated genes indicated in red. (A) Protein network containing genes encoding aldehyde dehydrogenases; (B) protein network containing a large proportion of genes with unknown functions.

hydrogenases have been proposed as selective drug targets against pathogenic mycobacteria (11). Hence, Rv0459 may represent a new potential drug target for *M. tuberculosis*.

Figure 5B illustrates another example of inference of protein function based on protein linkages. Rv1784 is induced in the pericavity and distant lung versus in vitro, whereas Rv3870 is upregulated in the granuloma but not in the pericavity or the distant lung. Both Rv1784 and Rv3870 have yet to be fully characterized at the molecular level, although Rv3870 has been hypothesized to be part of the RD1 specialized secretion machinery of *M. tuberculosis* (8, 16, 27, 32). Rv3870 is directly linked to five genes in the network, including Rv1784, Rv2748c, Rv3869, Rv3871, and Rv3894. Interestingly, three of the five genes that Rv3870 is linked to are located in or around the RD1 genomic region. The Rv3870 product contains sequence motifs that are homologous to the conserved FtsK domain and the FtsK_SpoIIIE domain. The FtsK_SpoIIIE domain is a putative ATP-binding P-loop found in known FtsK cell division proteins and SpoIIIE sporulation proteins. Divergent members of the FtsK_SpoIIIE family have been hypothesized to be involved in functions ranging from the extrusion and partitioning of DNA during cell division to the secretion of peptides through membrane pores (6, 17). The Rv1784 product also contains two conserved FtsK_SpoIIIE domains. Based on this information, we hypothesize that the genes of this network may be involved in processes related to the extrusion of biomolecules, either nucleic acids or proteins, specifically involving the FtsK_SpoIIIE domain-containing proteins.

The networks depicted in Fig. 5 are two examples of the nearly 70 protein networks available at our website (http://www.doe-mbi.ucla.edu/~strong/kaufmann) and were constructed using functional linkages inferred by a combination of the Rosetta stone, phylogenetic profile, conserved gene neighbor, and operon computational methods. Specifically, these networks represent proteins that are linked by two or more computational methods. A complete list of linked proteins, including the methods linking all proteins, is included in a searchable format at our website. Although these networks depict proteins that are functionally linked and that may therefore serve related cellular functions, the proteins may in fact have different expression profiles since they may perform similar cellular functions at different stages of the cell cycle or may be activated under different conditions. For example, some cell division proteins, such as FtsK, participate in dual roles, such as cell division and chromosome localization (20), or have expression profiles that differ from typical cell division proteins, such as FtsZ (28).

In addition to aiding in the inference of protein function, the computational methods we have applied in this study may suggest additional pathways that are important for the survival and persistence of *M. tuberculosis* within the human host. Additional protein networks are available at http://www.doe-mbi.ucla.edu/~strong/kaufmann.

To verify the results obtained from DNA array experiments, we performed real-time RT-PCR on some randomly chosen genes, including 4 of the 10 genes in Table S12 in the supplemental material (see Table S16 in the supplemental material). The results of real time RT-PCR confirmed those of the DNA array experiments. Especially for the genes listed in Table S12 in the supplemental material, these results indicate that these genes do not represent background noise arising during the DNA array experiments.

Since the patients from whom RNA samples were isolated received chemotherapy, we cannot exclude the possibility that the drugs influenced the gene expression profiles of tubercle bacilli in the patients' lungs. Recently, the gene expression profiles of *M. tuberculosis* bacteria from patients treated with various drugs in vitro have been reported (4, 40). One should exercise caution with the genes of *M. tuberculosis* which are similarly regulated in the lungs of tuberculosis patients and during drug treatment in vitro. Furthermore, it is possible that both conditions induce the expression of the same genes. Assuming that the antimycobacterial drugs exert similar influences at all sites of infection in the lung, differential gene expression profiles at different sites reflect the impact of the host environment on tubercle bacilli.

Recent studies on a select number of *M. tuberculosis* genes using RT-PCR and in situ hybridization revealed marked differences in the expression levels of mycobacteria isolated from mice and humans (15). Comparisons of our data with the

recent data from experiments with mice (29, 36) revealed similar signatures. Yet, relatively little overlap among genes overexpressed in mouse models and tuberculosis patients could be observed. This result is not entirely surprising since the environment of pulmonary tuberculosis in the human model is expected to be different from that of the mouse model, due in part to differences in the host immune response. This fact warrants the careful use of results obtained from animal models, for instance in the development of live vaccines and drug targets, since genes that may be important to virulence in one organism may not be crucial for virulence in another organism. Thus, we suggest that since the results of this study represent *M. tuberculosis* obtained directly from infected human lungs, they may provide an accurate representation of relevant gene expression profiles that can be used in the development of novel intervention strategies against this deadly microbial pathogen. Furthermore, the identification of gene expression signatures of *M. tuberculosis* indicative of distinct locations in the lung, which exert different stimuli on this pathogen, will facilitate the formulation of a set of biomarkers, even without knowledge of the functions of the gene products that contribute to this "biosignature."

ACKNOWLEDGMENTS

This work received financial support (to S.H.E.K.) from the German Ministry for Science and Technology (Competence Networks "Pathogenomics" and "Structural Genomics of *M. tuberculosis*"), the German Science Foundation (Priority Program "Novel Vaccination Strategies"), the EU FP6 Integrated Project TBVAC (LSHP-CT-2003-503367), and the Grand Challenge Program from the Bill and Melinda Gates Foundation. M.S. was supported by a USPHS National Research Service Award (GM07185).

H.R. thanks H. Witt and A. Saleh at AG Ruiz and H. Eickhoff, R. Reinhardt, and the whole automation crew at the Max Planck Institute for Molecular Genetics, Berlin, Germany. We thank M. Vingron for the correspondence analysis of our DNA array data. We thank H. Lehrach for his input and the company Chiron Behring for fruitful collaboration in the initial stage of this study. M.S. thanks M. Beeby, M. Pellegrini, and M. J. Thompson.

We declare that we have no competing financial interests.

REFERENCES

1. **Barry, C. E., III.** 2001. Interpreting cell wall 'virulence factors' of Mycobacterium tuberculosis. Trends Microbiol. **9:**237–241.
2. **Barry, C. E., III, R. E. Lee, K. Mdluli, A. E. Sampson, B. G. Schroeder, R. A. Slayden, and Y. Yuan.** 1998. Mycolic acids: structure, biosynthesis and physiological functions. Prog. Lipid Res. **37:**143–179.
3. **Biketov, S., G. V. Mukamolova, V. Potapov, E. Gilenkov, G. Vostroknutova, D. B. Kell, M. Young, and A. S. Kaprelyants.** 2000. Culturability of Mycobacterium tuberculosis cells isolated from murine macrophages: a bacterial growth factor promotes recovery. FEMS Immunol. Med. Microbiol. **29:**233–240.
4. **Boshoff, H. I., T. G. Myers, B. R. Copp, M. R. McNeil, M. A. Wilson, and C. E. Barry III.** 2004. The transcriptional responses of Mycobacterium tuberculosis to inhibitors of metabolism: novel insights into drug mechanisms of action. J. Biol. Chem. **279:**40174–40184.
5. **Boshoff, H. I., M. B. Reed, C. E. Barry III, and V. Mizrahi.** 2003. DnaE2 polymerase contributes to in vivo survival and the emergence of drug resistance in Mycobacterium tuberculosis. Cell **113:**183–193.
6. **Burts, M. L., W. A. Williams, K. DeBord, and D. M. Missiakas.** 2005. EsxA and EsxB are secreted by an ESAT-6-like system that is required for the pathogenesis of Staphylococcus aureus infections. Proc. Natl. Acad. Sci. USA **102:**1169–1174.
7. **Cole, S. T., R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S. V. Gordon, K. Eiglmeier, S. Gas, C. E. Barry III, F. Tekaia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, A. Krogh, J. McLean, S. Moule, L. Murphy, K. Oliver, J. Osborne, M. A. Quail, M. A. Rajandream, J. Rogers, S. Rutter, K. Seeger, J. Skelton, R. Squares, S. Squares, J. E. Sulston, K. Taylor, S. Whitehead, and B. G. Barrell.** 1998. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. Nature **393:**537–544.
8. **Converse, S. E., and J. S. Cox.** 2005. A protein secretion pathway critical for *Mycobacterium tuberculosis* virulence is conserved and functional in *Mycobacterium smegmatis*. J. Bacteriol. **187:**1238–1245.
9. **Dandekar, T., B. Snel, M. Huynen, and P. Bork.** 1998. Conservation of gene order: a fingerprint of proteins that physically interact. Trends Biochem. Sci. **23:**324–328.
10. **Delogu, G., and M. J. Brennan.** 2001. Comparative immune response to PE and PE_PGRS antigens of Mycobacterium tuberculosis. Infect. Immun. **69:**5606–5611.
11. **Duine, J. A.** 1999. Thiols in formaldehyde dissimilation and detoxification. Biofactors **10:**201–206.
12. **Eisen, M. B., P. T. Spellman, P. O. Brown, and D. Botstein.** 1998. Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. USA **95:**14863–14868.
13. **Espinal, M. A.** 2003. The global situation of MDR-TB. Tuberculosis (Edinburgh) **83:**44–51.
14. **Fellenberg, K., N. C. Hauser, B. Brors, A. Neutzner, J. D. Hoheisel, and M. Vingron.** 2001. Correspondence analysis applied to microarray data. Proc. Natl. Acad. Sci. USA **98:**10781–10786.
15. **Fenhalls, G., L. Stevens, L. Moses, J. Bezuidenhout, J. C. Betts, P. van Helden, P. T. Lukey, and K. Duncan.** 2002. In situ detection of *Mycobacterium tuberculosis* transcripts in human lung granulomas reveals differential gene expression in necrotic lesions. Infect. Immun. **70:**6330–6338.
16. **Hsu, T., S. M. Hingley-Wilson, B. Chen, M. Chen, A. Z. Dai, P. M. Morin, C. B. Marks, J. Padiyar, C. Goulding, M. Gingery, D. Eisenberg, R. G. Russell, S. C. Derrick, F. M. Collins, S. L. Morris, C. H. King, and W. R. Jacobs, Jr.** 2003. The primary mechanism of attenuation of bacillus Calmette-Guerin is a loss of secreted lytic function required for invasion of lung interstitial tissue. Proc. Natl. Acad. Sci. USA **100:**12420–12425.
17. **Iyer, L. M., K. S. Makarova, E. V. Koonin, and L. Aravind.** 2004. Comparative genomics of the FtsK-HerA superfamily of pumping ATPases: implications for the origins of chromosome segregation, cell division and viral capsid packaging. Nucleic Acids Res. **32:**5260–5279.
18. **Kaufmann, S. H.** 2000. Is the development of a new tuberculosis vaccine possible? Nat. Med. **6:**955–960.
19. **Lewthwaite, J. C., A. R. M. Coates, P. Tormay, M. Singh, P. Mascagni, S. Poole, M. Roberts, L. Sharp, and B. Henderson.** 2001. *Mycobacterium tuberculosis* chaperonin 60.1 is a more potent cytokine stimulator than chaperonin 60.2 (Hsp 65) and contains a CD14-binding domain. Infect. Immun. **69:**7349–7355.
20. **Liu, G., G. C. Draper, and W. D. Donachie.** 1998. FtsK is a bifunctional protein involved in cell division and chromosome localization in Escherichia coli. Mol. Microbiol. **29:**893–903.
21. **Mahon, B. P., and K. H. Mills.** 1999. Interferon-gamma mediated immune effector mechanisms against Bordetella pertussis. Immunol. Lett. **68:**213–217.
22. **Marcotte, E. M., M. Pellegrini, H. L. Ng, D. W. Rice, T. O. Yeates, and D. Eisenberg.** 1999. Detecting protein function and protein-protein interactions from genome sequences. Science **285:**751–753.
23. **Marcotte, E. M., M. Pellegrini, M. J. Thompson, T. O. Yeates, and D. Eisenberg.** 1999. A combined algorithm for genome-wide prediction of protein function. Nature **402:**83–86.
24. **Overbeek, R., M. Fonstein, M. D'Souza, G. D. Pusch, and N. Maltsev.** 1999. The use of gene clusters to infer functional coupling. Proc. Natl. Acad. Sci. USA **96:**2896–2901.
25. **Pellegrini, M., E. M. Marcotte, M. J. Thompson, D. Eisenberg, and T. O. Yeates.** 1999. Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. Proc. Natl. Acad. Sci. USA **96:**4285–4288.
26. **Phetsuksiri, B., M. Jackson, H. Scherman, M. McNeil, G. S. Besra, A. R. Baulard, R. A. Slayden, A. E. DeBarber, C. E. Barry III, M. S. Baird, D. C. Crick, and P. J. Brennan.** 2003. Unique mechanism of action of the thiourea drug isoxyl on Mycobacterium tuberculosis. J. Biol. Chem. **278:**53123–53130.
27. **Pym, A. S., P. Brodin, L. Majlessi, R. Brosch, C. Demangel, A. Williams, K. E. Griffiths, G. Marchal, C. Leclerc, and S. T. Cole.** 2003. Recombinant BCG exporting ESAT-6 confers enhanced protection against tuberculosis. Nat. Med. **9:**533–539.
28. **Robin, A., D. Joseleau-Petit, and R. D'Ari.** 1990. Transcription of the *ftsZ* gene and cell division in *Escherichia coli*. J. Bacteriol. **172:**1392–1399.
29. **Schnappinger, D., S. Ehrt, M. I. Voskuil, Y. Liu, J. A. Mangan, I. M. Monahan, G. Dolganov, B. Efron, P. D. Butcher, C. Nathan, and G. K. Schoolnik.** 2003. Transcriptional adaptation of Mycobacterium tuberculosis within macrophages: insights into the phagosomal environment. J. Exp. Med. **198:**693–704.
30. **Schuchhardt, J., D. Beule, A. Malik, E. Wolski, H. Eickhoff, H. Lehrach, and H. Herzel.** 2000. Normalization strategies for cDNA microarrays. Nucleic Acids Res. **28:**E47.
31. **Seiler, P., T. Ulrichs, S. Bandermann, L. Pradl, S. Jorg, V. Krenn, L. Morawietz, S. H. Kaufmann, and P. Aichele.** 2003. Cell-wall alterations as an attribute of Mycobacterium tuberculosis in latent infection. J. Infect. Dis. **188:**1326–1331.

32. **Sherman, D. R., K. M. Guinn, M. J. Hickey, S. K. Mathur, K. L. Zakel, and S. Smith.** 2004. Mycobacterium tuberculosis H37Rv: Delta RD1 is more virulent than M. bovis bacille Calmette-Guerin in long-term murine infection. J. Infect. Dis. **190:**123–126.
33. **Strong, M., T. G. Graeber, M. Beeby, M. Pellegrini, M. J. Thompson, T. O. Yeates, and D. Eisenberg.** 2003. Visualization and interpretation of protein networks in Mycobacterium tuberculosis based on hierarchical clustering of genome-wide functional linkage maps. Nucleic Acids Res. **31:**7099–7109.
34. **Strong, M., P. Mallick, M. Pellegrini, M. J. Thompson, and D. Eisenberg.** 2003. Inference of protein function and protein linkages in Mycobacterium tuberculosis based on prokaryotic genome organization: a combined computational approach. Genome Biol. **4:**R59.
35. **Talaat, A. M., P. Hunter, and S. A. Johnston.** 2000. Genome-directed primers for selective labeling of bacterial transcripts for DNA microarray analysis. Nat. Biotechnol. **18:**679–682.
36. **Talaat, A. M., R. Lyons, S. T. Howard, and S. A. Johnston.** 2004. The temporal expression profile of Mycobacterium tuberculosis infection in mice. Proc. Natl. Acad. Sci. USA **101:**4602–4607.
37. **Timm, J., F. A. Post, L. G. Bekker, G. B. Walther, H. C. Wainwright, R. Manganelli, W. T. Chan, L. Tsenova, B. Gold, I. Smith, G. Kaplan, and J. D. McKinney.** 2003. Differential expression of iron-, carbon-, and oxygen-responsive mycobacterial genes in the lungs of chronically infected mice and tuberculosis patients. Proc. Natl. Acad. Sci. USA **100:**14321–14326.
38. **Tusher, V. G., R. Tibshirani, and G. Chu.** 2001. Significance analysis of microarrays applied to the ionizing radiation response. Proc. Natl. Acad. Sci. USA **98:**5116–5121.
39. **Ulrichs, T., G. A. Kosmiadi, V. Trusov, S. Jorg, L. Pradl, M. Titukhina, V. Mishenko, N. Gushina, and S. H. Kaufmann.** 2004. Human tuberculous granulomas induce peripheral lymphoid follicle-like structures to orchestrate local host defence in the lung. J. Pathol. **204:**217–228.
40. **Waddell, S. J., R. A. Stabler, K. Laing, L. Kremer, R. C. Reynolds, and G. S. Besra.** 2004. The use of microarray analysis to determine the gene expression profiles of Mycobacterium tuberculosis in response to antibacterial compounds. Tuberculosis (Edinburgh) **84:**263–274.
41. **Wayne, L. G., and L. G. Hayes.** 1998. Nitrate reduction as a marker for hypoxic shiftdown of Mycobacterium tuberculosis. Tuber. Lung Dis. **79:**127–132.

Editor: J. L. Flynn