WhiB6 is required for the secretion-dependent regulation of ESX-1 substrates in pathogenic mycobacteria.

Short title: Regulation of ESX-1 substrates by WhiB6 in pathogenic mycobacteria

Abdallah M. Abdallah[1¶*], E.M. Weerdenburg[2¶], Qingtian Guan[1], R. Ummels[2], S. Borggreve[2], S.A. Adroub[1], Tareq B. Malas[1], Raeece Naeem[1], Huoming Zhang[3], T.D. Otto[4], W. Bitter[2*&] & A. Pain[1*&]

[1]Pathogen Genomics Laboratory, BESE Division, King Abdullah University of Science and Technology (KAUST), Thuwal-Jeddah, Kingdom of Saudi Arabia;

[2]Department of Medical Microbiology and Infection Control, VU University Medical Center, Amsterdam, The Netherlands;

[3]Bioscience Core Laboratory, King Abdullah University of Science and Technology (KAUST), Thuwal-Jeddah, Kingdom of Saudi Arabia;

[4]Pathogen Genomics, The Wellcome Trust Sanger Institute, Hinxton, Cambridge, United Kingdom

[¶] Joint authors

[&] Joint authors

*Corresponding authors

Abdallah M. Abdallah (abdallah.abdallah@kaust.edu.sa)

Wilbert Bitter (w.bitter@vumc.nl)

Arnab Pain (arnab.pain@kaust.edu.sa)

26

Abstract

The mycobacterial type VII secretion system ESX-1 is responsible for the secretion of a number of proteins that play important roles during host infection. The regulation of the expression of secreted proteins is often essential to establish successful infection. Using transcriptome sequencing, we found that the abrogation of ESX-1 function in *Mycobacterium marinum* leads to a pronounced increase in gene expression levels of the *espA* operon during the infection of macrophages, suggesting an important role in ESX-1-mediated virulence during the early phase of infection. In addition, the disruption of ESX-1-mediated protein secretion also leads to a specific down-regulation of the ESX-1 substrates, but not of the structural components of this system, during growth in culture medium. This effect is observed in both *M. marinum* and *M. tuberculosis*. We established that down-regulation of ESX-1 substrates is the result of a regulatory process that is influenced by the putative transcriptional regulator *whib6*, which is located adjacent to the *esx-1* locus. In addition, the overexpression of the ESX-1-associated PE35/PPE68 protein pair resulted in a significantly increased secretion of the ESX-1 substrate EsxA, demonstrating a functional link between these proteins. Taken together, these data show that WhiB6 is required for the secretion-dependent regulation of ESX-1 substrates and that ESX-1 substrates are regulated independently from the structural components, both during infection and as a result of active secretion.

Introduction

Mycobacteria use several different type VII secretion systems (T7S) to transport proteins across their thick and waxy cell envelopes. One of these T7S systems, ESX-1, is responsible for the transport of a number of important virulence factors. Disruption of the

esx-1 gene cluster severely reduces the virulence of *Mycobacterium tuberculosis* [1], whereas restoration of *esx-1* in the *Mycobacterium bovis*-derived vaccine strain BCG, which lacks part of the *esx-1* region due to continuous passaging, leads to increased virulence [2]. Many studies have attempted to elucidate the function of ESX-1 substrates in virulence. In the case of pathogenic mycobacteria, such as *M. tuberculosis* and the fish pathogen *Mycobacterium marinum*, ESX-1 is responsible for the translocation of the bacteria from the phagolysosomal compartments to the cytosols of macrophages [3,4]. This translocation activity has been attributed to the ability of the secreted protein EsxA (also called ESAT-6) to lyse membranes [5,6]. Interestingly, a closely related homologue of this protein is also secreted by non-pathogenic and non-translocating mycobacteria such as *Mycobacterium smegmatis*. A recent report indicated that, although the EsxA proteins of *M. smegmatis* and *M. tuberculosis* are highly homologous, the membrane lysis potentials of these proteins are different [7]. In *M. smegmatis*, ESX-1 is involved in a completely different process, *i.e.*, conjugative DNA transfer [8]. The proposed functions of ESX-1 in pathogenic mycobacterial species include host cell entry and intercellular spread [9].

The ESX-1 substrates identified to date are mostly encoded by genes of the *esx-1* locus, such as EsxA, EsxB, EspE and EspB. The exceptions are EspA and EspC [10,11], which are both part of the *espA* operon, which is located elsewhere in the genome; however, these genes are homologous to the *espE* and *espF* genes, respectively, which belong to the *esx-1* locus . A peculiar characteristic of ESX-1 substrates is that these substrates are mutually dependent, *i.e.*, the secretion of each of these substrates is dependent on the secretion of the other substrates [10]. The secreted ESX proteins contain a conserved WxG amino acid motif located between two α-helices [12]. Recently, an additional conserved secretion signal, present in all secreted protein pairs, was identified. This C-terminal YxxxD/E motif can target proteins for secretion, but does not determine the specificity for a

particular type VII system [13]. Therefore, it remains difficult to bioinformatically predict novel ESX-1 substrates.

To establish successful infection, mycobacteria need regulatory mechanisms to express the right proteins at the right time. In different environments, mycobacteria require specific transcriptional responses to successfully respond to the stress conditions encountered. During the first stages of infection, ESX-1-mediated protein secretion is one of the most important virulence mechanisms of pathogenic mycobacteria [4,9,10,14,15]. Consequently, the tight transcriptional regulation of *esx-1* and the associated genes is required. The transcriptional regulator PhoP of the two-component system PhoPR positively regulates the transcription of many *esx-1*-associated genes, including genes in the *espA* operon [16,17]. It has been proposed that PhoP regulation is dependent on environmental pH [18], which could indicate that the acidic environment of the phagosome induces *esx-1* gene transcription via PhoP, leading to bacterial escape from this compartment. Other studies have shown that the *espA* operon is, in addition to PhoP, also regulated by the transcription factors EspR and MprAB and the repressors CRP and Lsr2, indicating that tight regulation of this operon is essential and, furthermore, suggesting that the *espA* operon may be regulated in a manner distinct from the regulation of other ESX-1 substrates [19-22].

Since ESX-1 is crucial for virulence, inactivation of this secretion system would be expected to have a large impact on gene regulation processes in mycobacteria. Here, we apply RNA-seq and quantitative proteomics to determine the gene expression and proteomic profiles of the pathogenic mycobacteria *M. marinum* and *M. tuberculosis* in the absence of a functional ESX-1 secretion system. During short-term infection of macrophages, we observed highly increased transcript levels of the *espA* operon. In contrast, during in vitro growth in culture medium, transcription of most ESX-1 substrates and some putative new substrates was seen to be decreased. Based on these gene transcription levels, we confirmed a regulatory

role for the putative transcriptional regulator WhiB6 in the gene expression of ESX-1 substrates.

Results

Global features of the M. marinum esx-1 mutant transcriptome and proteome

To investigate the effect of ESX-1 disruption on gene expression and protein production, RNA and protein were extracted from three independent exponential phase cultures of the *M. marinum* E11 strain and the isogenic *esx-1*-mutant during growth in 7H9 culture medium to characterize the transcriptome and proteome. Using transcriptomics (RNA-seq) and mass spectrometry (MS)-based proteomics with isobaric labelling for quantification, we captured the expression dynamics of the transcripts and proteomes of the *esx-1* mutant. Data quality was assessed using Euclidean distance matrices for RNA (Figure S1) and principal component analysis (PCA) for protein (Figure S2), which demonstrated high levels of reproducibility between biological replicates. After filtering (see Materials and Methods for details), a total of 823 genes were identified as being differentially expressed (DE) as messenger RNA, of which 525 were classified as down-regulated and 298 as up-regulated (Figure 1A, Table S1). To determine parallel changes in protein levels, 1,657 proteins were identified by the presence of 2 or more peptides, of which 576 proteins passed our filter and we classified them as DE. Of these, 412 proteins were found to be down-regulated and 164 were up-regulated (Table S2), and 482 protein-coding genes were shared and were identified in both the RNA-seq and quantitative proteomic datasets (Figure 1C).

Fig 1. Global Features of the Transcriptomes and Proteomes of the *M. marinum* and *M. tuberculosis* Esx-1 Mutant Strains.

Volcano plots obtained from RNA-seq analysis of the wild-type *M. marinum* E11 strain vs. the *eccCb$_1$* transposon mutant (A) and of *M. tuberculosis* mc^{2}6020 vs. the ESX-1 mutant strain (B). Each dot indicates the expression value of a gene. Red dots indicate statistical significance ($q < 0.05$), and black dots indicate a lack of statistical significance. Selected genes that are most down- or up-regulated in the *ESX-1* mutant strains are highlighted. (C) Venn diagram of the number of differentially expressed transcripts and proteins quantified using RNA-seq and quantitative proteomics, respectively. Scatterplots of the relationship between differentially expressed genes of *M. marinum eccCb$_1$* transposon mutant and those of the isogenic wild-type strain E11, quantified in both data sets and classified into the following categories: (D) lipid metabolism, (E) regulatory proteins and (F) cell wall and cell process. Scatterplots and bar chart show the rectilinear equation and the Pearson correlation coefficient (R^2).

The degree of global correlation between the gene expression and protein abundance scores among the shared genes was relatively low (Figure S3A), which has also been noted in other bacterial studies [23]. However, within certain classes of *M. marinum* functional categories (http://mycobrowser.epfl.ch/marinolist.html), the degree of correlation was much higher than that in other classes, with R^2 exceeding 0.8 for the categories such as lipid metabolism (Figure 1D), regulation (Figure 1E) and cell wall and cell processes (Figure 1F). Of the DE genes at the RNA and protein levels, 28% were in the intermediary metabolism and respiration category, 18% were in the cell wall and cell process category, 15% were in the information pathways category and 14% were in the lipid metabolism category (Figure S4).

Transcriptional profiling analysis of the double auxotrophic *M. tuberculosis* $mc^2$6020 mutant strains [24] and their isogenic *esx-1* mutants during growth was carried out to identify genes for which expression was dependent on ESX-1 disruption (Figure 1B, Table S3). For this species, the same trends could be identified as for *M. marinum*.

Major effects of ESX1 mutation on genes encoding ESX-1 substrates and biosynthetic pathways

Analysis of differential expression in *M. marinum* identified changes in genes involved in a variety of cellular processes (Figure 2), although a majority of the most differentially regulated genes were associated with cell wall and cell processes and lipid metabolism. We noted that a substantial number of *esx-1*-associated genes were down-regulated in the mutant strains during growth in culture medium, including 11 genes that were located within or directly adjacent to the *esx-1* gene cluster. Among these down-regulated genes were those coding for known ESX-1 substrates, such as EsxA, EsxB, EspE and EspB. Remarkably, mRNA levels of core components of the ESX-1 secretion system, *i.e.*, those encoding members of the type VII secretion complex, such as $EccB_1$, $EccD_1$, $EccE_1$ and $MycP_1$, remained unchanged, even though the corresponding genes are interspersed with genes encoding ESX-1 substrates. In contrast to the mRNA levels, we noted a strong increase in the protein levels of EsxA and EsxB, probably reflecting the accumulation of these proteins in the cell due to the secretion defect (Figure 2). Our data also indicate a significant effect of *esx-1* disruption on genes associated with lipid metabolism (Figure 2), including genes associated with the synthesis of mycolic acids. Strong down-regulation was observed at the mRNA and protein levels for several polyketide synthases, including genes involved in phthiocerol dimycocerosate synthesis and mycolic acid biosynthesis, such as *umaA*, *mmaA3*, *accD5*, *accD6*, and *pks15/1*, which encode components of the lipid biosynthesis pathway

(Figure 2, Table S1, S2). The changes observed in *esx-1* and lipid-metabolism-associated genes at the mRNA and protein levels were not unexpected; it has been reported previously that ESX-1-dependent protein secretion and mycolic acid synthesis are critically linked [25]. However, we also noted a surprisingly broad impact of ESX-1 mutation on major biosynthetic pathways, including ribosomal protein synthesis and DNA biosynthesis (Table S1, S2). Down-regulation was observed at the mRNA and protein levels for several genes encoding ribosomal proteins and DNA gyrase and a ribonucleotide-diphosphate reductase, which are components of protein and DNA biosynthesis, respectively. We also identified changes at both the mRNA and protein levels in genes involved in general stress response (*grpE, dnaK, groES, groEL1*), genes involved in stress response regulation (*sigA, sigB, devS*), members of the WhiB family (*whiB2, whiB4, whiB6*) and several PE_PGRS genes (Figure 2). A similar trend was observed for *M. tuberculosis* (Figure 2). In fact, of all the genes, the *esx-1* genes encoding the substrates EsxA, EsxB and EspK were the most significantly down-regulated in the mutant strain (Figure 2, Table S3). However, in contrast to the levels in the *M. marinum* mutant, the gene expression levels of $eccB_1$ and $eccD_1$ were also somewhat decreased in the *M. tuberculosis* mutant (Figure 2). On the other hand, the *M. tuberculosis esx-1* mutation did not seem to have a significant effect on the expression of genes involved in lipid metabolism compared to the effect seen in *M. marinum* (Figure 2, Table S1, S3). Finally, a significant number of genes that are associated with information pathways, including genes encoding ribosomal proteins, were up-regulated at the mRNA level in the *esx-1* mutant (Figure 2). Taken together, the observed changes in the transcriptome and proteome of the *esx-1* mutant reflect the role of the *esx-1* cluster employed by mycobacteria for major biochemical pathways.

Fig 2. Top Differentially Expressed Genes of *M. marinum* and *M. tuberculosis*, when Grown in Culture Medium, Grouped into Broad Functional Categories.

Within each group, genes are ranked in ascending order by *p*-value. (Red) Top 100 annotated *M. marinum* E11 genes that exhibit greatest differential expression in the *M. marinum eccCb*$_1$ transposon mutant compared to the isogenic wild-type strain E11 during growth in 7H9 culture medium. Bar chart of log2-fold change for individual genes (RNA, blue; protein, red; locus tags, outer). (Green) Top 100 annotated *M. tuberculosis* genes that exhibit greatest differential expression in the auxotrophic *M. tuberculosis* RD1 deletion mutant strain mc^{2}6030 compared to the isogenic control strain mc^{2}6020 during growth in 7H9 culture medium. Bar chart of log2-fold change for individual genes. The genes Rv3872-Rv3878 are not included as these genes are deleted in the RD1 mutant strain.

Global transcriptional profiling of intraphagosomal M. marinum and the esx-1 mutant

We next determined the effect of ESX-1 abrogation in *M. marinum* on gene transcription during infection of primary macrophages. Using a PMA-differentiated THP-1 cell line as a model of primary macrophages, we analysed the global gene expression of the wild-type and *esx-1* mutant strains of *M. marinum* after 6 hours of infection. Wild-type mycobacteria can escape the phagosome within two hours after infection, whereas *esx-1* mutants are known to be limited to the phagosomal compartment. The intraphagosomal transcriptome of the *esx-1* mutant was compared with the intracellular transcriptome of wild-type *M. marinum*. Furthermore, these intracellular transcriptomes were also compared with the transcriptome of wild-type *M. marinum* grown in standard broth culture. We identified 720 ($p<0.05$) genes in the *esx-1* mutant that exhibited significant changes in expression after THP-1 infection compared to the expression levels in the wild-type strain. Of these genes,

465 were down-regulated and 255 were up-regulated (Table S4, Figure S5). Remarkably, none of the genes within the *esx-1* region were significantly differentially expressed in the *esx-1*-mutant compared to the wild-type strain. However, we found a specific and pronounced increase in the transcript levels of the *espA* operon in the intraphagosomal transcriptome of the *esx-1* mutant compared with the levels in the in vitro transcriptomes (Figure 3A). During growth in culture medium, the mRNA levels of *espA* did not differ between the wild-type and *esx-1*-deficient *M. marinum* strains, which was confirmed by quantitative RT-PCR (qRT-PCR) (Figure 3B). Therefore, these data suggest that proteins encoded by the *espA* operon, *i.e.*, EspA, EspC and EspD, play an important role in ESX-1-specific processes during the first stages of macrophage infection. The *espA* operon was also induced in the wild-type bacteria inside macrophages, albeit at a lower level. Perhaps this difference exists because the wild-type bacteria are able to escape from the phagosome, whereas the *esx-1* mutants are not.

Fig 3. Effect of ESX-1 Disruption (*eccCb*$_1$ Transposon Mutant) on Gene Transcription During Infection (Indicated as ínt') and Growth in Culture Medium in *M. marinum* Compared to that in the Wild-Type Strain E11 during Growth in 7H9 Culture Medium. (A) Relative transcript expression levels of the ESX-1 secretion system-associated genes, including the main ESX-1 locus as well as the EspR regulator and accessory factors in the EspA operon, which is encoded outside the RD-1 region. (B) Gene expression levels, as measured by qRT-PCR, were compared to those of the wild-type strain E11 grown in similar conditions. Values represent mean ± standard error of the mean of two biological replicates. (C, D) Regulation of genes associated with cell wall synthesis, including genes involved in mycolic acid synthesis (C) and PDIMs (D).

Further analysis showed that a significant number of genes that code for proteins involved in cell wall and cell processes were differentially regulated by intracellular wild-type *M. marinum* and the ESX-1-deficient strain in comparison with their counterparts grown in culture medium (Table S5, S6). *M. marinum* genes involved in mycolic acid synthesis, phthiocerol dimycocerosate (PDIM) synthesis and transport to the cell surface, such as *fabG1*, *accDs*, *ppsC*, *ppsD*, *pks11_1*, *pks13*, as well as genes coding for polyketide synthases and the mycolic acid methyltransferase *umaA* were differentially expressed during infection of THP-1 cells (Figure 3C, D). Furthermore, *cpsY*, a gene that encodes UDP-glucose 4-epimerases and is essential for linking peptidoglycans and mycolic acid [26], exhibited a pronounced increase in mRNA level in the intracellular *esx-1* mutant (Table S4, S5, S6, S7). We also found that many genes associated with cell division and peptidoglycan assembly, such as *ftsE*, *ftsH*, *ftsW*, *murC*, and *murG* [27,28], were down-regulated by intracellular bacteria (Table S4, S5, S6, S7).

A significant number of genes that code for proteins associated with lipid metabolism and metabolic adaptation were differentially regulated in macrophages (Figure S6A). This subset includes genes involved in fatty acid metabolism such as isocitrate lyase (*icl*), an enzyme necessary for the glyoxylate cycle and required for intracellular survival [29,30], and *pckA*, which encodes the phosphoenolpyruvate carboxykinase and is essential for mycobacterial survival in both macrophages and mice [31,32] and is involved in energy metabolism (Figure S6B), and the KstR-dependent cholesterol regulon (Figure S6C), which is involved in lipid degradation and carbon metabolism [33]. We also observed significant effects of a number of genes involved in general stress response (*groES*, *groEL1*, *hsp*, *ahp*, *dnaK*), genes involved in stress response regulation (*sigB*, *devR*, *devS*, *hspR*, *kstR*), members of the WhiB family (*whiB2*, *whiB3*, *whiB4*, *whiB*, *whiB6*, *whiB7*) and alternative sigma factors (*sigE*, *sigL*, *sigM*) in the *esx-1* mutant during infection to macrophages. This pattern is

illustrated in Figure S6D and is probably associated with stressful intraphagosomal conditions.

Different M. marinum esx-1 transposon mutants have similar gene transcription profiles

The ESX-1-deficient strain of *M. marinum* used for RNA sequencing contains a transposon in the *eccCb*$_1$ gene. To confirm that the observed gene transcription effects were due to a defective ESX-1 system and not due to a side effect of this particular mutation, we analysed several mutants containing transposon insertions in different genes from the *esx-1* gene cluster and compared the mRNA levels of the selected genes by qRT-PCR. Our results showed decreased transcript levels of the known ESX-1 substrate *esxA* and other *esx-1* secretion-associated (*esp*) genes, namely, *espL*, *espK* and *espJ*, for all tested *esx-1* mutants, whereas the transcript levels of *eccD*$_1$, which encodes a structural component of the ESX-1 system, did not differ from the transcript levels in wild-type *M. marinum* (Figure 4). These gene expression patterns in the *eccB*$_1$, *eccCa*$_1$, *eccD*$_1$ and *eccE*$_1$ transposon mutants were similar to the RNA sequencing results obtained for the *eccCb*$_1$ mutant. The only exception was that for the mutant containing a transposon insertion in *eccD*$_1$, we observed an increase of *eccD*$_1$ transcription itself and, to a lesser extent, of the adjacent gene *espJ* (Figure 4). However, this increase was most likely due to the presence of a strong promoter on the transposon, driving the transcription of the kanamycin resistance cassette, as the measured mRNA is transcribed from sequences directly downstream of this promoter. Altogether, our results demonstrate that inactivation of the ESX-1 secretion system leads to down-regulation of the transcription of ESX-1 substrates and associated proteins.

Fig 4. *Esx-1* Transposon Mutants have Similar Gene Transcription Profiles.

Gene expression levels for *M. marinum* $eccB_1$, $eccCa_1$, $eccCb_1$, $eccD_1$ and $eccE_1$ transposon mutants as measured by qRT-PCR. All strains were grown in 7H9 culture medium, and gene expression levels were compared to those of the wild-type strain E11. Values represent mean ± standard error of the mean of at least three biological replicates.

ESX-1 substrate gene transcription is reduced by a regulatory mechanism

We next sought to determine the molecular mechanism underlying the down-regulation of specific transcripts in *esx-1* mutant strains of *M. marinum*. It is possible that the decrease in mRNA levels is due to a regulatory effect at the transcriptional level. Alternatively, mRNA derived from specific sequences may be degraded via a post-transcriptional mechanism. To investigate these possibilities, we expressed an extra copy of the *espL* gene under the control of a constitutively active promoter in the wild-type and $eccCb_1$ mutant strains of *M. marinum* and determined the *espL* gene transcript levels. We found a similar increase in *espL* transcripts in both the wild-type and $eccCb_1$ mutant strains, indicating that degradation of specific mRNA is probably not the cause of the decreased mRNA levels in the mutant strain (Figure 5A). Expression levels of the downstream gene *espK* were not affected by the introduction of *espL*. These results indicate that there is a regulatory mechanism that prevents the transcription of genes encoding ESX-1 substrates and associated proteins in the absence of a functionally active ESX-1.

Fig 5. Regulation of the ESX-1 Secretion System.

(A) Down-regulation of *espL* is the result of a regulatory process. A functional copy of *espL* was introduced into wild-type and *eccCb1* mutant strains of *M. marinum*, and the espK and

espL gene expression levels were measured by qRT-PCR. Gene expression levels were compared to those of the wild-type strain E11. Values represent mean ± standard error of the mean of two biological replicates. (B) Introduction of PE35/PPE68_1 results in increased EsxA secretion but not in gene regulation. Pellet (p), cell wall extract (cw), and supernatant (s) fractions of the wild-type and *eccCb1* mutant strains of *M. marinum* expressing PE35/PPE68_1; PE35/PPE68, containing a C-terminal deletion of PPE68_1; or PE35/PPE68_1, containing a 15-amino-acid C-terminal deletion of PE35, were analysed for the presence of EspE, EsxA and the introduced PE35 by Western blotting. Fractions represent 0.5, 1 or 2 OD units of original culture. In all cases, PE35 contained a C-terminal HA tag. (C) EspG1, EspI and PE35/PPE68_1 do not regulate the transcription of selected esx-1-associated genes. EspG1, EspI or PE35/PPE68_1 were overexpressed in the *M. marinum* eccCb1 mutant strain, and the expression levels of *espK, espL, esxA, pe_pgrs1* and *eccD1* were measured by qRT-PCR. Gene expression levels were compared to those of the wild-type strain E11. Values represent mean ± standard error of the mean of at least two biological replicates. (D) WhiB6 is involved in transcriptional regulation of ESX-1 substrates and associated genes. The *whib6* gene was overexpressed in the *M. marinum eccCb1* mutant strain, and transcript levels of *espK, espL, esxA, pe_pgrs1* and *eccD1* were measured by qRT-PCR. Gene expression levels were compared to those of the *eccCb1* mutant strain. Values represent mean ± standard error of the mean of two biological replicates.

PE35 and PPE68 play an important role in ESX-1 secretion but not in gene regulation

Previously, PE35, which is located within the *esx-1* gene cluster, has been implicated in the regulation of *esxA/esxB* gene expression in *M. tuberculosis* [34]. In contrast to this proposed function, the PE35/PPE68_1 protein pair in *M. marinum* is secreted via ESX-1

348 [35,36]. To determine whether PE35 plays a role in the regulation of ESX-1 substrates, we
349 overexpressed the *pe35/ppe68_1* operon in *M. marinum*. Interestingly, although there was no
350 effect on gene transcription (Figure 5C), we noticed a substantial increase in EsxA secretion
351 in the wild-type strain (Figure 5B). This increased EsxA secretion does not seem to represent
352 a general increase in ESX-1 secretion, as protein levels of the cell-surface-localized EspE
353 remained similar (Figure 5B). To study this effect in more detail, we introduced PE35 with a
354 truncated version of PPE68_1 that contained only the PPE domain and was devoid of the C-
355 terminal portion. Although the introduced PE35 protein was expressed and secreted
356 efficiently by ESX-1 (Figure 5B), the levels of secreted EsxA were not increased, indicating
357 that the C-terminal portion of PPE68_1 plays a role in EsxA secretion. To determine whether
358 secretion of the PE35/PPE68_1 protein pair itself was important for this process, we also
359 determined the effect of removal of the last 15 amino acids of the PE protein, which
360 contained the general secretion signal. This small deletion not only abolished the secretion of
361 the introduced PE35 protein but also abolished EsxA secretion completely, despite the
362 presence of an intact chromosomal copy of the *pe35/ppe68_1* operon (Figure 5B). This result
363 suggests that the truncated form of PE35 somehow interferes with EsxA secretion. Together
364 these data show that, although PE35 and PPE68_1 do not seem to regulate the transcription of
365 genes encoding ESX-1 substrates, these proteins have a strong effect on EsxA, as previously
366 observed [34].
367

368 Increasing EspI and EspG$_1$ levels does not lead to altered esx-1 gene expression

370 A second candidate protein that might regulate gene expression levels of ESX-1
371 substrates is EspI. The gene encoding this *esx-1*-secretion-associated protein of unknown
372 function is located within the *esx-1* region and is down-regulated in *esx-1* mutants of both *M.*

373 *marinum* and *M. tuberculosis* (Figure 2). In contrast to the other Esp proteins, EspI contains a
374 putative nucleotide-binding domain. However, when we overexpressed this protein, we did
375 not observe a change in the down-regulation of *esx-1*-associated gene transcription in the *M.*
376 *marinum eccCb$_1$* transposon mutant, suggesting that EspI does not regulate this process
377 (Figure 5C). We next focused on EspG$_1$ as a candidate *esx-1* gene regulator. EspG$_1$, which is
378 a cytosolic protein that is not part of the membrane-bound secretion machinery, has recently
379 been shown to interact specifically with PE35/PPE68_1 in *M. marinum* [35]. It is conceivable
380 that EspG$_1$ might function as a sensor that measures protein levels of intracellular ESX-1
381 substrates. When substrate levels are low, unbound EspG$_1$ may act as a signal for the
382 induction of gene expression. In the absence of a functional ESX-1 system, accumulated
383 PE35/PPE68_1 or other substrates may bind to EspG$_1$, leading to reduced transcription of
384 *esx-1*-associated genes. To investigate the effect of EspG$_1$ on *esx-1*-associated gene
385 expression and protein levels, we increased EspG$_1$ levels by overexpressing the *espG$_1$* gene in
386 wild-type and ESX-1-deficient *M. marinum*. This overexpression did not result in altered
387 gene transcription (Figure 5C) or ESX-1 protein secretion (data not shown). Together, our
388 data show that EspI and EspG$_1$ do not appear to play key roles in *esx-1*-associated gene
389 regulation.
390

391 ## WhiB6 plays a role in the transcription of ESX-1 substrates

392 In addition to *espI*, another gene encoding a putative regulatory protein was down-
393 regulated in *esx-1* mutant strains of both *M. marinum* and *M. tuberculosis*, namely, *whiB6*
394 (Figure 2). WhiB proteins are actinobacteria-specific regulators that contain iron-sulfur
395 clusters and are thought to act as redox-sensing transcription factors that can cause both gene
396 activation and repression [37]. WhiB6 was suggested to be involved in the regulation of
397 EsxA secretion [38], and subsequent studies have confirmed this suggestion [39-41]. To

398 determine whether Whib6 had an effect on the expression levels of *esx-1* associated genes,
399 we overexpressed this protein in the ESX-1-deficient *M. marinum eccCb*$_1$ transposon mutant
400 strain. We found that particularly those genes that were already down-regulated in the mutant
401 strain, such as *esxA* and *espK*, showed even further transcriptional inhibition when *whib6*
402 levels were increased (Figure 5D). Furthermore, expression of *eccD1* was unaltered by *whib6*
403 overexpression, indicating that *whib6* is involved in the transcription of ESX-1 substrates and
404 associated genes but not of the components of the ESX-1 system. Surprisingly, *whiB6* is one
405 of the genes that is down-regulated upon abrogation of ESX-1-mediated protein secretion.
406 WhiB6 may be activated when the ESX-1 machinery is blocked and represses genes
407 encoding ESX-1 substrates as well as its own gene. Together, our data suggests that ESX-1
408 regulation is even more complex than previously thought.
409

410 ## WhiB6 is required for the regulation of the ESX-1 system

411 To determine whether WhiB6 is required for ESX-1 regulation, we constructed a deletion
412 mutant of *whiB6* both in the *M. marinum* WT strain and in the *esx-1* mutant background (*M.*
413 *marinum* M^{USA} *-ΔwhiB6*–and *M. marinum* M^{VU} *–ΔwhiB6*). First, we analyzed the effect of
414 this mutation on all genes, except for the genes of the *esx-1* locus. This analysis identified 34
415 genes ($p<0.05$) that were significantly downregulated in the *esx-1* mutant strains (Figure 6A).
416 Complementation of both mutants with the *whiB6* gene on a mycobacterial shuttle plasmid
417 reversed the upregulation of these genes and generally resulted in decreased expression levels
418 (Figure 6A, B). As expected, several genes that are associated with oxidative stress (*ahpC,*
419 *ahpD, rebU*) were found in the differently expressed gene pool. Also, the enrichment analysis
420 of the associated Gene Ontology terms for the differently expressed genes (*dnaB, dinP*)
421 reveal that WhiB6 may also regulate DNA replication or repair through regulating DNA-
422 directed DNA polymerase and DNA helicase (Figure S7). Another noteworthy gene affected

by *whiB6* deletion is *iniA*, which is associated with cell wall stress induced by specific antibiotics. Interestingly, *whiB7* is within the *whiB6-active* gene set, which implies that WhiB7 is active by, or works with WhiB6. Other than the *whiB6*-active gene set, 13 genes, which are involved in iron-sulfur cluster binding, cellular lipid metabolic processes are downregulated.

Fig 6. Gene Expression Profiles (Log2 Fold Change) of the Complementary Strains (*M. marinum* M^{USA}-Complementary and *M. marinum* M^{VU}-complementary) and Knock-Out Strains (*M. marinum* M^{USA} -⊿*whiB6* and *M. marinum* M^{VU}-⊿*whiB6*) Compared with those of the Corresponding Control Strains (*M. marinum* M^{USA}-Empty Vector Strain and *M. marinum* M^{VU}-empty vector). The green colour represents up-regulate genes and red colour represents down regulate genes compared with the control strains. The heat map of expression of the whiB6-activated gene set is shown in (A); Expression of the *esx-1* locus is shown in (B); and (C) shows the WhiB6-repressed gene set.

Remarkably, these genes are almost exclusively downregulated in the *whiB6 esx-1* double mutant, reinforcing a functional link between WhiB6 and the ESX-1 system. However, many of the downregulated genes are encoding hypothetical proteins and hence needed to be further characterized.

Separately, we analyzed the effect of the *whiB6* deletion on all *esx-1* genes. In line with our previous results, overexpression of *whiB6* in the *esx-1* mutant resulted in downregulation of many *esx-1* genes (Figure 6C), whereas deletion of *whiB6* did not have a strong affect in the *esx-1* mutant background. The effect was the opposite for the mutant with a functional ESX-1 system, there the whiB6 deletion had a strong positive effect on transcription of *esx-1* genes. (Figure 6C) and also this effect could be complemented. Only

the structural *eccE1* and *mycP1* genes behaved differently. There are six genes that show the same pattern of up- or down-regulation as most of the *esx-1* genes in the two different *whiB6* mutants and the complemented strains. Of these six genes, 4 encode putative ESX-1 substrates or ESX-1 chaperones, *i.e.* MMAR_2894 (PE34-like protein), MMAR0299 (PE_PGRS1), MMAR5414 ((EspA-like) and MMAR5432 (EspD-like). Together, these experiments show a strong linkage between WhiB6 and the regulation of different *esx-1* genes in response to secretion activity.

Discussion

In this study, we determined the transcriptomes of the *M. marinum* E11 wild-type and the double-auxotrophic *M. tuberculosis* $mc^{2}6020$ mutant strains and compared these transcriptomes with those of the corresponding isogenic *esx-1* mutants. We found that during growth in 7H9 culture medium, genes encoding ESX-1 substrates, such as EsxA and other ESX-1-associated proteins, were down-regulated in the mutant strains, whereas the transcription of genes encoding several structural components of the ESX-1 system remained unaffected. This specific decrease in transcription might function as a mechanism to avoid toxic accumulation of ESX-1 substrates. Interestingly, a similar decrease in substrate production has been shown for the ESX-5 secretion system, where the PE_PGRS substrates do not accumulate intracellularly when secretion is blocked [42,43]. However, for these PE_PGRS substrates, regulation was shown to occur post-transcriptionally [43], implying that a different mechanism is involved.

The most prominent change in gene expression that was observed upon host cell infection by the *M. marinum esx-1* mutant strain was the increase in transcription of the *espA* operon. The specific and pronounced transcriptional increase in the expression of this operon, and not of any other *esx-1* associated gene, indicates that transcription of the *espA* operon is

473 regulated independently of the other substrates during infection. Previously, it has been
474 shown that the *espA* operon is regulated by different transcription and regulation factors,
475 including EspR, MprAB and PhoPR [20,44,45]. Our new finding also suggests that EspA,
476 EspC and EspD are vital for the bacteria during the early phase of infection. Since ESX-1 has
477 been shown to be responsible for mycobacterial escape from the phagosome, which occurs
478 within the first few hours of infection with *M. marinum* [6], the proteins produced by the
479 *espA* operon may play an important role in this process. Consequently, the avirulent
480 phenotype of ESX-1-deficient mycobacteria might be partly attributable to the inability to
481 secrete EspA and/or EspC early in infection.

482 To determine the mechanism via which ESX-1 substrate regulation is mediated, we
483 overexpressed proteins that may have a regulatory function. Overexpression of the *esx-1*-
484 encoded EspI and $EspG_1$ proteins did not have an effect on the reduced transcription of ESX-
485 1 substrates in ESX-1-deficient *M. marinum*. The putative regulatory protein WhiB6,
486 however, did affect the transcription of these genes. While the transcript levels of *whib6* itself
487 were decreased in *esx-1* mutants of *M. marinum* and *M. tuberculosis*, increasing WhiB6
488 levels by overexpression resulted in a further decrease in transcription of the ESX-1 substrate
489 in ESX-1-deficient *M. marinum*. This result clearly indicates that WhiB6 is involved in ESX-
490 1-associated gene regulation, as previously suggested [41]. Indeed, there is accumulating
491 evidence that WhiB proteins function as transcription factors that may play a role in survival
492 within the host (reviewed in [46]). Recently, other groups have also presented evidence
493 supporting a role for WhiB6 in the regulation of the transcription of *esx-1* genes [39-41].

494 A remarkable finding in this study was that overproduction of PE35/PPE68_1 resulted
495 in a large increase in EsxA secretion. Previously, deletion of *M. tuberculosis* PE35 had
496 already been shown to abolish *esxA* transcription and secretion of the corresponding gene
497 product [34]. Here, we found that EsxA and PE35 secretion are linked, as an increase in PE35

secretion resulted in a concomitant increase in EsxA secretion. The C terminus of PPE68_1 is required for this effect, indicating that this is a specific process, which is consistent with the fact that cell-surface localization of another ESX-1 substrate, namely, EspE, is unaffected by overproduction of PE35/PPE68_1. It is possible that the PPE68 proteins serve as chaperones to escort EsxA outside the bacterium, or these proteins, may be part of the secretion apparatus, making the secretion of specific substrates highly efficient.

During *M. marinum* infection of human macrophages, we found that transcription of many *pe_pgrs* and *ppe* family genes was strongly down-regulated when ESX-1 function was abrogated. As many as 50% of all genes with decreased transcript levels in the *esx-1* mutant strain belongs to one of these gene families (Table S4). Notably, in the wild-type strain, transcription of the *pe_pgrs* and *ppe* genes was decreased during infection in comparison to the levels observed during growth in 7H9 medium (Table S5). As part of an adaptive response to the macrophage environment, expression of these cell-wall-localized proteins may be down-regulated in order to evade immune recognition or to reduce cell permeability [47]. The fact that in the absence of a functional ESX-1 secretion system these genes are even further down-regulated suggests that there are functional links or shared transcriptional pathways between ESX-1 and (some of the) PE_PGRS and PPE proteins, which are generally ESX-5 substrates [48].

Taken together, our results show that transcription of the *espA* locus plays an important role in ESX-1 mediated processes during the first hours of infection. Furthermore, we established a functional link between PE35 and EsxA secretion and provided evidence of a regulatory role of WhiB6 in the transcription of ESX-1 substrates and associated genes.

Materials and Methods

Bacterial strains and growth conditions

The *esx-1* mutants of the *M. marinum* E11 wild-type strain used in this study contain transposon insertions in *eccB*$_1$, *eccCa*$_1$, *eccCb*$_1$, *eccD*$_1$ and *eccE*$_1$ [49]. For *M. tuberculosis*, the attenuated double-deletion strains mc^{2}6020 and mc^{2}6030 of H37Rv were used, with deletions of *lysA* and *panCD* and of *RD1* and *panCD*, respectively [24,50]. Bacterial strains were grown with shaking at 30°C (*M. marinum)* or 37°C (*M. tuberculosis*) in Middlebrook 7H9 culture medium supplemented with 10% ADC (albumin-dextrose-catalase, BD Biosciences) and 0.05% Tween-80. Culture medium containing the auxotrophic *M. tuberculosis* deletion strains was supplemented with 50 µg/ml pantothenic acid and, for mc^{2}6020, 100 µg/ml L-lysine.

Infection of human macrophages

THP-1 monocytes were cultured at 37°C in 5% CO_2 in RPMI-1640 with GlutaMAX-1 (Gibco) supplemented with 10% FBS, 100 µg/ml streptomycin and 100 U/ml penicillin. Cells were seeded at a density of 3×10^7 cells per T175 flask and differentiated into macrophages by 48 hours of incubation with 25 ng/ml PMA (Sigma-Aldrich). Then, 1.8×10^8 THP-1 cells were infected with *M. marinum* at a multiplicity of infection (MOI) of 20 for 2 hours, after which the cells were washed with PBS to remove extracellular bacteria. After 4 additional hours of infection at 33°C, the THP-1 cells were lysed with 1% Triton X-100. After a low-speed centrifugation step to remove cellular debris, mycobacteria were pelleted, after which RNA was extracted as described in the following section.

Genome sequence

We sequenced the *M. marinum* E11 strain with PacBio RSII single-molecule real-time (SMRT) sequencing technology [51]. The raw reads were assembled into two pieces (the core and the plasmid) with HGAP assembler [52] using the default parameters. The sequence was improved with iCORN2 [53] with three iterations, correcting 20 single base pair errors and 61 insertions and deletions. To transfer the annotation from the current reference, we used RATT [54] with the PacBio parameter. Gene models around gaps were manually improved on the new sequence. The updated genome annotation was resubmitted under the same accession numbers (HG917972 for the *M. marinum E11* main chromosome genome and HG917973 for the *M. marinum E11* pRAW plasmid; complete sequences).

RNA extraction and qRT-PCR

M. marinum and *M. tuberculosis* cultures were pelleted and bead beated in 1 ml of TRIzol (Invitrogen) with 0.1-mm zirconia/silica beads (BioSpec Products). After centrifugation, supernatants were extracted with chloroform, and RNA was precipitated with isopropanol. RNA pellets were washed with 80% ethanol and dissolved in RNAse-free water. Contaminant DNA was removed by incubation with DNAse I (Fermentas). For RT-PCR, cDNA was generated using a SuperScript VILO cDNA Synthesis Kit (Invitrogen). An equivalent of 5 ng of RNA was used in the quantitative PCRs. qRT-PCR was performed using SYBR GreenER (Invitrogen) and a LightCycler 480 (Roche) instrument. Transcript levels were normalized to the levels of the housekeeping gene *sigA* [55] using ΔΔCt analysis. All primer sequences used for qRT-PCR are listed in Table S8.

RNA preparation for Illumina sequencing

Total RNA was extracted with TRIzol (Invitrogen) and then purified on RNeasy spin columns (Qiagen) according to the manufacturer's instructions. RNA integrity (RNA

integrity score ≥ 6.8) and quantity were determined on an Agilent 2100 Bioanalyzer (Agilent; Palo Alto, CA, USA). As ribosomal RNA constitutes a vast majority of the extracted RNA population, depletion of these molecules via RiboMinus-based rRNA depletion was conducted. For mRNA enrichment, Invitrogen's RiboMinus Transcriptome Isolation Kit, bacteria was used according to manufacturer's instructions. Briefly, 2 µg of total RNA samples was hybridized with prokaryotic rRNA-sequence-specific 5′-biotin-labelled oligonucleotide probes to selectively deplete large rRNA molecules from total RNA. Then, these rRNA-hybridized, biotinylated probes were removed from the sample with streptavidin-coated magnetic beads. The resulting RNA sample was concentrated using the RiboMinus concentrate module according to the manufacturer's protocol. The final RiboMinus RNA sample was subjected to thermal mRNA fragmentation using the Elute, Prime, Fragment Mix from the Illumina TruSeq RNA Sample Preparation Kit v2 (Low-Throughput protocol). The fragmented mRNA samples were subjected to cDNA synthesis using the Illumina TruSeq RNA Sample Preparation Kit (low-throughput protocol) according to the manufacturer's protocol. Briefly, cDNA was synthesized from enriched and fragmented RNA using SuperScript III reverse transcriptase (Invitrogen) and the SRA RT primer (Illumina). The cDNA was further converted into double-stranded DNA using the reagents supplied in the kit, and the resulting dsDNA was used for library preparation. To this end, cDNA fragments were end-repaired and phosphorylated, followed by adenylation of 3′ ends and adapter ligation. Twelve cycles of PCR amplification were then performed, and the library was finally purified with AMPure beads (Beckman Coulter) as per the manufacturer's instructions. A small aliquot (1 µl) was analysed on an Invitrogen Qubit and an Agilent Bioanalyzer. The bar-coded cDNA libraries were pooled at equal concentrations before sequencing on an Illumina HiSeq2000 using the TruSeq SR Cluster Generation Kit v3 and TruSeq SBS Kit v3. Data were processed with Illumina Pipeline software v1.82.

RNA-seq analysis

The Illumina reads were mapped with SMALT (http://www.sanger.ac.uk/science/tools/smalt-0) (default parameters) against the new PacBio reference. From the read count, which was obtained with bedtools ([56], parameter multicov, with -D to include duplicates and -q 5 to exclude repetitive mapping reads), we performed a differential expression analysis with DESeq [57] using default parameters.

Plasmid construction

The *E. coli*-mycobacterial shuttle vector pSMT3 was used for the construction of all plasmids. To overexpress PE35-PPE68_1 (MMARE11_01740- MMARE11_01750), we used a previously described plasmid [13]. For construction of the plasmid containing $espG_1$, this gene was amplified from the *M. marinum* E11 genome by PCR using primers containing NheI and EcoRV restriction sites and a 3' HA epitope. The resulting PCR product and empty pSMT3 were digested with NheI and EcoRV followed by ligation of $espG_1$ into the vector by T4 ligase (Fermentas). For construction of the plasmid containing *whib6*, this gene was amplified from the *M. marinum* E11 genome by PCR using primers containing NheI and BamHI restriction sites. For the other construct, *espI* was amplified from the *M. marinum* E11 genome by PCR using primers containing NheI and BglII restriction sites. The PCR product was digested with NheI and BamHI. Empty pSMT3 was digested with NheI and BamHI, after which the PCR product was ligated into the vector. All plasmids were introduced into the *M. marinum* wild-type E11 and isogenic $eccCb_1$ mutant strains by electroporation. All primer sequences are listed in Table S8.

Analysis of protein expression and secretion

M. marinum cultures were grown to mid-logarithmic phase in 7H9 culture medium supplemented with 0.2% glycerol and 0.2% dextrose. Bacteria were pelleted, washed in PBS and incubated in 0.5% Genapol X-080 (Sigma-Aldrich) for 30 minutes to extract cell wall proteins. Genapol X-080-treated *M. marinum* cells were disrupted by sonication. Secreted proteins were precipitated from the culture supernatant by 10% trichloroacetic acid (TCA, Sigma-Aldrich). Proteins were separated according to molecular weight on 15% SDS-PAGE gels and subsequently transferred to nitrocellulose membranes (Amersham Hybond ECL, GE Healthcare Life Sciences). Immunostaining was performed with mouse monoclonal antibodies directed against the HA epitope (HA.11, Covance), EsxA (Hyb76-8), or rabbit polyclonal sera recognizing EspE [58].

LC-MS analysis

Peptide preparation from the *M. marinum* E11 and isogenic *esx-1* mutant strains was performed as previously described [59]. Approximately 100-µg protein digests of each sample were labelled with 4plex iTRAQ reagents (Applied Biosystems). The combined iTRAQ-labelled samples were fractionated using strong cation exchange chromatography. The eluted fractions were dried and desalted using a Sep-Pak C-18 SPE cartridge (Waters, Milford, MA, USA). LC-MS analysis as well as MS data processing was carried out following our published procedure [60]. Briefly, each fraction was analysed three times using an LTQ-Orbitrap Velos (Thermo Scientific). The MS spectra were recorded in the Orbitrap, whereas the MS2 spectra were recorded in the c-TRAP for HCD fragmentation and in the LTQ for the CID fragmentation. Both HCD and CID spectra were extracted separately using Proteome Discoverer software and processed by an in-house script before a Mascot search against the *M. marinum* E11 proteome. The Mascot results (.dat file) were processed by

Scaffold software for validation of protein identification and quantitative assessment. For protein identification, local false positive rates (FDR) were maintained below 1% for both protein and peptide identification (0.91% and 0.9% for peptides and proteins, respectively, for this dataset). Protein quantitation was processed using Scaffold Q+, which is based on the i-Tracker algorithm [61]. The iTRAQ quantitation using HCD is highly accurate, and a change of more than 2-fold was considered significant differential expression in this study.

Accession codes

Sequencing reads have been submitted to the EMBL-EBI European Nucleotide Archive (ENA) Sequence Read Archive (SRA) under the study accession no. PRJEB8560.

Acknowledgements

We thank Astrid van der Sar and Esther Stoop for providing the *M. marinum E11* ESX-1 mutants. Work in AP's laboratory is supported by the KAUST faculty baseline fund (BAS/1/1020-01-01). The authors thank members of the Bioscience Core Lab (BCL) at KAUST for sequencing the RNA-seq libraries on the Illumina Hiseq platform and for running protein samples through the quantitative proteomics workflow with the LTQ-Orbitrap Velos instrument (Thermo Scientific).

References

1. Lewis KN, Liao R, Guinn KM, Hickey MJ, Smith S, Behr MA, et al. Deletion of RD1 from Mycobacterium tuberculosis mimics bacille Calmette-Guerin attenuation. J Infect Dis. 2003;187: 117-123.

668 2. Pym AS, Brodin P, Brosch R, Huerre M, Cole ST. Loss of RD1 contributed to the
669 attenuation of the live tuberculosis vaccines Mycobacterium bovis BCG and
670 Mycobacterium microti. Mol Microbiol. 2002;46: 709-717.
671 3. Stamm LM, Morisaki JH, Gao LY, Jeng RL, McDonald KL, Roth R, et al.
672 Mycobacterium marinum escapes from phagosomes and is propelled by actin-based
673 motility. J Exp Med. 2003;198: 1361-1368.
674 4. van der Wel N, Hava D, Houben D, Fluitsma D, van Zon M, Pierson J, et al. M.
675 tuberculosis and M. leprae translocate from the phagolysosome to the cytosol in
676 myeloid cells. Cell. 2007;129: 1287-1298.
677 5. de Jonge MI, Pehau-Arnaudet G, Fretz MM, Romain F, Bottai D, Brodin P, et al.
678 ESAT-6 from Mycobacterium tuberculosis dissociates from its putative chaperone
679 CFP-10 under acidic conditions and exhibits membrane-lysing activity. J Bacteriol.
680 2007;189: 6028-6034.
681 6. Houben D, Demangel C, van Ingen J, Perez J, Baldeon L, Abdallah AM, et al. ESX-1-
682 mediated translocation to the cytosol controls virulence of mycobacteria. Cell
683 Microbiol. 2012;14: 1287-1298.
684 7. De Leon J, Jiang G, Ma Y, Rubin E, Fortune S, Sun J. Mycobacterium tuberculosis
685 ESAT-6 exhibits a unique membrane-interacting activity that is not found in its
686 ortholog from non-pathogenic Mycobacterium smegmatis. J Biol Chem. 2012;287:
687 44184-44191.
688 8. Coros A, Callahan B, Battaglioli E, Derbyshire KM. The specialized secretory
689 apparatus ESX-1 is essential for DNA transfer in Mycobacterium smegmatis. Mol
690 Microbiol. 2008;69: 794-808.

691 9. Gao LY, Guo S, McLaughlin B, Morisaki H, Engel JN, Brown EJ. A mycobacterial
692 virulence gene cluster extending RD1 is required for cytolysis, bacterial spreading and
693 ESAT-6 secretion. Mol Microbiol. 2004;53: 1677-1693.
694 10. Fortune SM, Jaeger A, Sarracino DA, Chase MR, Sassetti CM, Sherman DR, et al.
695 Mutually dependent secretion of proteins required for mycobacterial virulence. Proc
696 Natl Acad Sci USA. 2005;102: 10676-10681.
697 11. MacGurn JA, Raghavan S, Stanley SA, Cox JS. A non-RD1 gene cluster is required
698 for Snm secretion in Mycobacterium tuberculosis. Mol Microbiol. 2005;57: 1653-
699 1663.
700 12. Pallen MJ. The ESAT-6/WXG100 superfamily -- and a new Gram-positive secretion
701 system? Trends Microbiol. 2002;10: 209-212.
702 13. Daleke MH, Ummels R, Bawono P, Heringa J, Vandenbroucke-Grauls CM, Luirink J,
703 et al. General secretion signal for the mycobacterial type VII secretion pathway. Proc
704 Natl Acad Sci USA. 2012;109: 11342-11347.
705 14. McLaughlin B, Chon JS, MacGurn JA, Carlsson F, Cheng TL, Cox JS, et al. A
706 mycobacterium ESX-1-secreted virulence factor with unique requirements for export.
707 PLoS Pathog. 2007;3: e105.
708 15. Tan T, Lee WL, Alexander DC, Grinstein S, Liu J. The ESAT-6/CFP-10 secretion
709 system of Mycobacterium marinum modulates phagosome maturation. Cell
710 Microbiol. 2006;8: 1417-1429.
711 16. Asensio JA, Arbues A, Perez E, Gicquel B, Martin C. Live tuberculosis vaccines
712 based on phoP mutants: a step towards clinical trials. Expert Opin Biol Ther. 2008;8:
713 201-211.

714 17. Walters SB, Dubnau E, Kolesnikova I, Laval F, Daffe M, Smith I. The
715 Mycobacterium tuberculosis PhoPR two-component system regulates genes essential
716 for virulence and complex lipid biosynthesis. Mol Microbiol. 2006;60: 312-330.
717 18. Abramovitch RB, Rohde KH, Hsu FF, Russell DG. aprABC: a Mycobacterium
718 tuberculosis complex-specific locus that modulates pH-driven adaptation to the
719 macrophage phagosome. Mol Microbiol. 2011;80: 678-694.
720 19. Gordon BRG, Li YF, Wang LR, Sintsova A, van Bakel H, Tian SH, et al. Lsr2 is a
721 nucleoid-associated protein that targets AT-rich sequences and virulence genes in
722 Mycobacterium tuberculosis (vol 107, pg 5154, 2010). Proc Natl Acad Sci USA.
723 2010;107: 18741-18741.
724 20. Pang XH, Samten B, Cao GX, Wang XS, Tvinnereim AR, Chen XL, et al. MprAB
725 regulates the espA operon in mycobacterium tuberculosis and modulates ESX-1
726 function and host cytokine response. J Bacteriol. 2013;195: 66-75.
727 21. Raghavan S, Manzanillo P, Chan K, Dovey C, Cox JS. Secreted transcription factor
728 controls Mycobacterium tuberculosis virulence. Nature. 2008;454: 717-721.
729 22. Rickman L, Scott C, Hunt DM, Hutchinson T, Menendez MC, Whalan R, et al. A
730 member of the cAMP receptor protein family of transcription regulators in
731 Mycobacterium tuberculosis is required for virulence in mice and controls
732 transcription of the rpfA gene coding for a resuscitation promoting factor. Mol
733 Microbiol. 2005;56: 1274-1286.
734 23. Waldbauer JR, Rodrigue S, Coleman ML, Chisholm SW. Transcriptome and
735 proteome dynamics of a light-dark synchronized bacterial cell cycle. PLoS One.
736 2012;7: e43432.
737 24. Sambandamurthy VK, Derrick SC, Jalapathy KV, Chen B, Russell RG, Morris SL, et
738 al. Long-term protection against tuberculosis following vaccination with a severely

739 attenuated double lysine and pantothenate auxotroph of Mycobacterium tuberculosis.
740 Infect Immun. 2005;73: 1196-1203.
741 25. Joshi SA, Ball DA, Sun MG, Carlsson F, Watkins BY, Aggarwal N, et al. EccA1, a
742 component of the Mycobacterium marinum ESX-1 protein virulence factor secretion
743 pathway, regulates mycolic acid lipid synthesis. Chem Biol. 2012;19: 372-380.
744 26. Weston A, Stern RJ, Lee RE, Nassau PM, Monsey D, Martin SL, et al. Biosynthetic
745 origin of mycobacterial cell wall galactofuranosyl residues. Tuber Lung Dis. 1997;78:
746 123-131.
747 27. Pastoret S, Fraipont C, den Blaauwen T, Wolf B, Aarsman ME, Piette A, et al.
748 Functional analysis of the cell division protein FtsW of Escherichia coli. J Bacteriol.
749 2004;186: 8370-8379.
750 28. Schmidt KL, Peterson ND, Kustusch RJ, Wissel MC, Graham B, Phillips GJ, et al. A
751 predicted ABC transporter, FtsEX, is needed for cell division in Escherichia coli. J
752 Bacteriol. 2004;186: 785-793.
753 29. McKinney JD, Honer zu Bentrup K, Munoz-Elias EJ, Miczak A, Chen B, Chan WT,
754 et al. Persistence of Mycobacterium tuberculosis in macrophages and mice requires
755 the glyoxylate shunt enzyme isocitrate lyase. Nature. 2000;406: 735-738.
756 30. Munoz-Elias EJ, McKinney JD. Mycobacterium tuberculosis isocitrate lyases 1 and 2
757 are jointly required for in vivo growth and virulence. Nat Med. 2005;11: 638-644.
758 31. Liu KY, Yu JZ, Russell DG. pckA-deficient Mycobacterium bovis BCG shows
759 attenuated virulence in mice and in macrophages. Microbiology-Sgm. 2003;149:
760 1829-1835.
761 32. Marrero J, Rhee KY, Schnappinger D, Pethe K, Ehrt S. Gluconeogenic carbon flow of
762 tricarboxylic acid cycle intermediates is critical for Mycobacterium tuberculosis to
763 establish and maintain infection. Proc Natl Acad Sci USA. 2010;107: 9819-9824.

764 33. Kendall SL, Withers M, Soffair CN, Moreland NJ, Gurcha S, Sidders B, et al. A
765 highly conserved transcriptional repressor controls a large regulon involved in lipid
766 degradation in Mycobacterium smegmatis and Mycobacterium tuberculosis. Mol
767 Microbiol. 2007;65: 684-699.
768 34. Brodin P, Majlessi L, Marsollier L, de Jonge MI, Bottai D, Demangel C, et al.
769 Dissection of ESAT-6 system 1 of Mycobacterium tuberculosis and impact on
770 immunogenicity and virulence. Infect Immun. 2006;74: 88-98.
771 35. Daleke MH, van der Woude AD, Parret AHA, Ummels R, de Groot AM, Watson D,
772 et al. Specific chaperones for the type VII Protein secretion pathway. J Biol Chem.
773 2012;287: 31939-31947.
774 36. Sani M, Houben EN, Geurtsen J, Pierson J, de Punder K, van Zon M, et al. Direct
775 visualization by cryo-EM of the mycobacterial capsular layer: a labile structure
776 containing ESX-1-secreted proteins. PLoS Pathog. 2010;6: e1000794.
777 37. Zheng F, Long Q, Xie J. The function and regulatory network of WhiB and WhiB-
778 like protein from comparative genomics and systems biology perspectives. Cell
779 Biochem Biophys. 2012;63: 103-108.
780 38. Das C, Ghosh TS, Mande SS. Computational analysis of the ESX-1 region of
781 Mycobacterium tuberculosis: insights into the mechanism of type VII secretion
782 system. PLoS One. 2011;6: e27980.
783 39. Bosserman RE, Nguyen TT, Sanchez KG, Chirakos AE, Ferrell MJ, Thompson CR, et
784 al. WhiB6 regulation of ESX-1 gene expression is controlled by a negative feedback
785 loop in Mycobacterium marinum. Proc Natl Acad Sci USA. 2017;114: E10772-
786 E10781.

787 40. Chen Z, Hu Y, Cumming BM, Lu P, Feng L, Deng J, et al. Mycobacterial WhiB6
788 differentially regulates ESX-1 and the Dos Regulon to modulate granuloma formation
789 and virulence in zebrafish. Cell Rep. 2016;16: 2512-2524.
790 41. Solans L, Aguilo N, Samper S, Pawlik A, Frigui W, Martin C, et al. A specific
791 polymorphism in mycobacterium tuberculosis H37Rv causes differential ESAT-6
792 expression and identifies WhiB6 as a novel ESX-1 component. Infect Immun.
793 2014;82: 3446-3456.
794 42. Abdallah AM, Verboom T, Weerdenburg EM, van Pittius NCG, Mahasha PW,
795 Jimenez C, et al. PPE and PE_PGRS proteins of Mycobacterium marinum are
796 transported via the type VII secretion system ESX-5. Mol Microbiol. 2009;73: 329-
797 340.
798 43. Ates LS, Dippenaar A, Ummels R, Piersma SR, van der Woude AD, van der Kuij K,
799 et al. Mutations in ppe38 block PE_PGRS secretion and increase virulence of
800 Mycobacterium tuberculosis. Nat Microbiol. 2018;3: 181-188.
801 44. Cao G, Howard ST, Zhang P, Wang X, Chen XL, Samten B, et al. EspR, a regulator
802 of the ESX-1 secretion system in Mycobacterium tuberculosis, is directly regulated by
803 the two-component systems MprAB and PhoPR. Microbiology. 2015;161: 477-489.
804 45. Hunt DM, Sweeney NP, Mori L, Whalan RH, Comas I, Norman L, et al. Long-range
805 transcriptional control of an operon necessary for virulence-critical ESX-1 secretion
806 in Mycobacterium tuberculosis. J Bacteriol. 2012;194: 2307-2320.
807 46. den Hengst CD, Buttner MJ. Redox control in actinobacteria. Biochim Biophys Acta.
808 2008;1780: 1201-1216.
809 47. Ates LS, van der Woude AD, Bestebroer J, van Stempvoort G, Musters RJ, Garcia-
810 Vallejo JJ, et al. The ESX-5 system of pathogenic mycobacteria is involved in capsule

integrity and virulence through its substrate PPE10. PLoS Pathog. 2016;12: e1005696.

48. Abdallah AM, Verboom T, Weerdenburg EM, van Pittius NCG, Mahasha PW, Jimenez C, et al. PPE and PE_PGRS proteins of Mycobacterium marinum are transported via the type VII secretion system ESX-5. Molecular Microbiology. 2009;73: 329-340.

49. van der Sar AM, Abdallah AM, Sparrius M, Reinders E, Vandenbroucke-Grauls CMJE, Bitter W. Mycobacterium marinum strains can be divided into two distinct types based on genetic diversity and virulence. Infect Immun. 2004;72: 6306-6312.

50. Sambandamurthy VK, Derrick SC, Hsu T, Chen B, Larsen MH, Jalapathy KV, et al. Mycobacterium tuberculosis DeltaRD1 DeltapanCD: a safe and limited replicating mutant strain that protects immunocompetent and immunocompromised mice against experimental tuberculosis. Vaccine. 2006;24: 6309-6320.

51. Levene MJ, Korlach J, Turner SW, Foquet M, Craighead HG, Webb WW. Zero-mode waveguides for single-molecule analysis at high concentrations. Science. 2003;299: 682-686.

52. Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, et al. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nature Methods. 2013;10: 563-+.

53. Otto TD, Sanders M, Berriman M, Newbold C. Iterative Correction of Reference Nucleotides (iCORN) using second generation sequencing technology. Bioinformatics. 2010;26: 1704-1707.

54. Otto TD, Dillon GP, Degrave WS, Berriman M. RATT: rapid annotation transfer tool. Nucleic Acids Res. 2011;39: e57.

835 55. Manganelli R, Dubnau E, Tyagi S, Kramer FR, Smith I. Differential expression of 10
836 sigma factor genes in Mycobacterium tuberculosis. Mol Microbiol. 1999;31: 715-724.
837 56. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic
838 features. Bioinformatics. 2010;26: 841-842.
839 57. Anders S, Huber W. Differential expression analysis for sequence count data.
840 Genome Biol. 2010;11: R106.
841 58. Carlsson F, Joshi SA, Rangell L, Brown EJ. Polar localization of virulence-related
842 Esx-1 secretion in mycobacteria. PLoS Pathog. 2009;5: e1000285.
843 59. Abdallah AM, Hill-Cawthorne GA, Otto TD, Coll F, Guerra-Assuncao JA, Gao G, et
844 al. Genomic expression catalogue of a global collection of BCG vaccine strains show
845 evidence for highly diverged metabolic and cell-wall adaptations. Sci Rep. 2015;5:
846 15443.
847 60. Zhu J, Zhang HM, Guo TN, Li WY, Li HY, Zhu Y, et al. Quantitative proteomics
848 reveals differential biological processes in healthy neonatal cord neutrophils and adult
849 neutrophils. Proteomics. 2014;14: 1688-1697.
850 61. Shadforth IP, Dunkley TPJ, Lilley KS, Bessant C. i-Tracker: For quantitative
851 proteomics using iTRAQ (TM). BMC Genomics. 2005;6: 145.

852
853
854 **Supplementary table and figure legends**

855
856 **Table S1.**

857 Complete list of genes for which the expression levels changed significantly in the *M.*
858 *marinum eccCb*$_1$ transposon mutant compared to the levels in the isogenic wild-type strain
859 E11 during growth in 7H9 culture medium. P<0.05.

860

861 **Table S2.**

Complete list of proteins for which the expression levels changed in *M. marinum eccCb*$_1$ transposon mutant compared to the levels in the isogenic wild-type strain E11 during growth in 7H9 culture medium. Proteins with greater than 2-fold change were considered significantly differentially expressed.

Table S3.

Complete list of genes for which the expression levels changed significantly ($p<0.05$) in the auxotrophic *M. tuberculosis* RD1 deletion mutant strain mc^{2}6030 compared to the levels in the isogenic control strain mc^{2}6020 during growth in 7H9 culture medium.

Table S4.

Complete list of genes for which the expression levels changed significantly ($p<0.05$) in the *M. marinum eccCb*$_1$ transposon mutant strain compared to the levels in the wild-type strain E11 during infection of human THP-1 macrophages.

Table S5.

Complete list of genes for which the expression levels changed significantly ($p<0.05$) in the *M. marinum* wild-type strain during infection of macrophages compared to the levels during growth in 7H9 culture medium.

Table S6.

Complete list of genes for which expression levels changed significantly ($p<0.05$) in the *M. marinum eccCb*$_1$ transposon mutant strain during infection of macrophages compared to the levels in the wild-type strain E11 during growth in 7H9 culture medium.

887 **Table S7.**

888 Complete list of genes for which the expression levels changed significantly ($p<0.05$) in the
889 *M. marinum eccCb$_1$* transposon mutant strain during infection of macrophages compared to
890 the levels during growth in 7H9 culture medium.

891

892 **Table S8.**

893 Primers used in this study. Restriction sites are shown in bold.

894

895 **Figure S1**

896 Euclidean distance matrices of RNA-seq transcriptome data showing clustering of *M.*
897 *marinum* wild-type (E11) and *eccCb$_1$* transposon mutant (ESX-1) strains grown in culture
898 medium (three biological replicates) or during infection of THP-1 cells (indicated as 'int').

899

900 **Figure S2**

901 Principal component analysis (PCA) of biological replicates of proteome data showing
902 clustering of *M. marinum* wild-type (E11) and *eccCb$_1$* transposon mutant (ESX-1) strains.
903 PCA mapping showed clustering of biological replicates of the E11 wild-type and *esx-1*
904 mutant strains.

905

906 **Figure S3**

907 Correlation between protein and mRNA expression of the *M. marinum eccCb$_1$* transposon
908 mutant and the isogenic wild-type strain E11 during growth in 7H9 culture medium. (A)
909 Scatterplot of the relationship between differentially expressed genes quantified in both data
910 sets. (B-F) Scatterplots for protein and gene transcript expression classified by functional

911 categories. Scatterplots display the rectilinear equation and the Pearson correlation coefficient
912 (R^2).
913
914 **Figure S4**
915 Functional categories of genes that are significantly changed in the transcriptome and
916 proteome of the *M. marinum eccCb$_1$* transposon mutant compared to the isogenic wild-type
917 strain E11 during growth in 7H9 culture medium. Genes exhibiting differential expression at
918 the RNA and protein levels were grouped according to the MarinoList classification
919 (http://mycobrowser.epfl.ch/marinolist.html).
920
921 **Figure S5**
922 Most differentially expressed genes of the *M. marinum eccCb$_1$* transposon mutant compared
923 to the isogenic wild-type strain E11 during infection of primary macrophages, grouped into
924 broad functional categories. Within each group, genes are ranked in ascending order by *P*-
925 value. (A). Top 100 annotated genes from the *M. marinum E11* strain that were the most
926 differentially expressed in the *M. marinum* wild-type strain E11 during infection of primary
927 macrophages. Bar chart of log2-fold changes for individual genes (tags, left). (B). Top 100
928 annotated genes from the *M. marinum E11* strain that were the most differentially expressed
929 in the *M. marinum eccCb$_1$* transposon mutant compared to the isogenic wild-type strain E11
930 during infection of primary macrophages (tags, left). Bar chart of log2-fold changes for
931 individual genes.
932
933 **Figure S6:**
934 Regulation of genes encoding proteins predicted to be involved in metabolic adaptation,
935 energy metabolism and transcriptional regulatory processes in the *M. marinum eccCb$_1$*

936 transposon mutant grown in 7H9 culture medium as well as in the wild-type and *eccCb*$_1$
937 transposon mutant strains during infection in human THP-1 macrophages (indicated as 'int')
938 compared to the that in the wild-type strain E11 during growth in 7H9 culture medium. (A)
939 *Catabolism of fatty acids*. Genes were selected based on their annotation and ordered based
940 on expression. (B) Energy generation and NAD^+ regeneration. Genes were selected based on
941 their annotation and ordered based on expression. (C) Genes of the *kstR* regulon, which are
942 required for uptake and metabolism of cholesterol (1, 2). (D) Transcriptional regulation.
943 Genes were selected based on their annotation and ordered based on expression.

944 **Figure S7**

945 The enriched Gene Ontology (GO) terms of the gene set activated (excluding the *esx-1* locus
946 genes) or repressed by WhiB6. The molecular function GO terms are in red, while the
947 biological process terms are in blue.

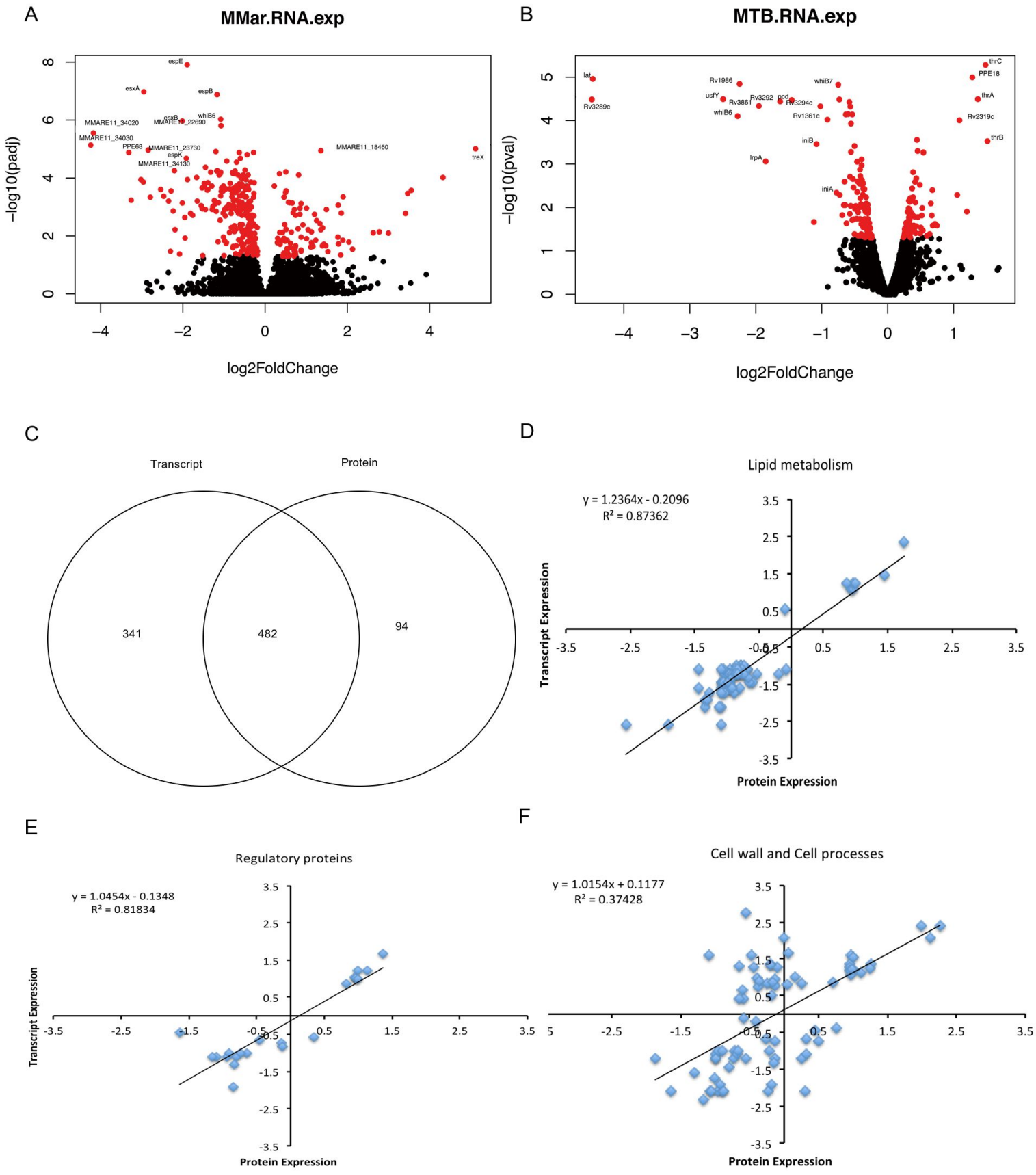


Figure.1

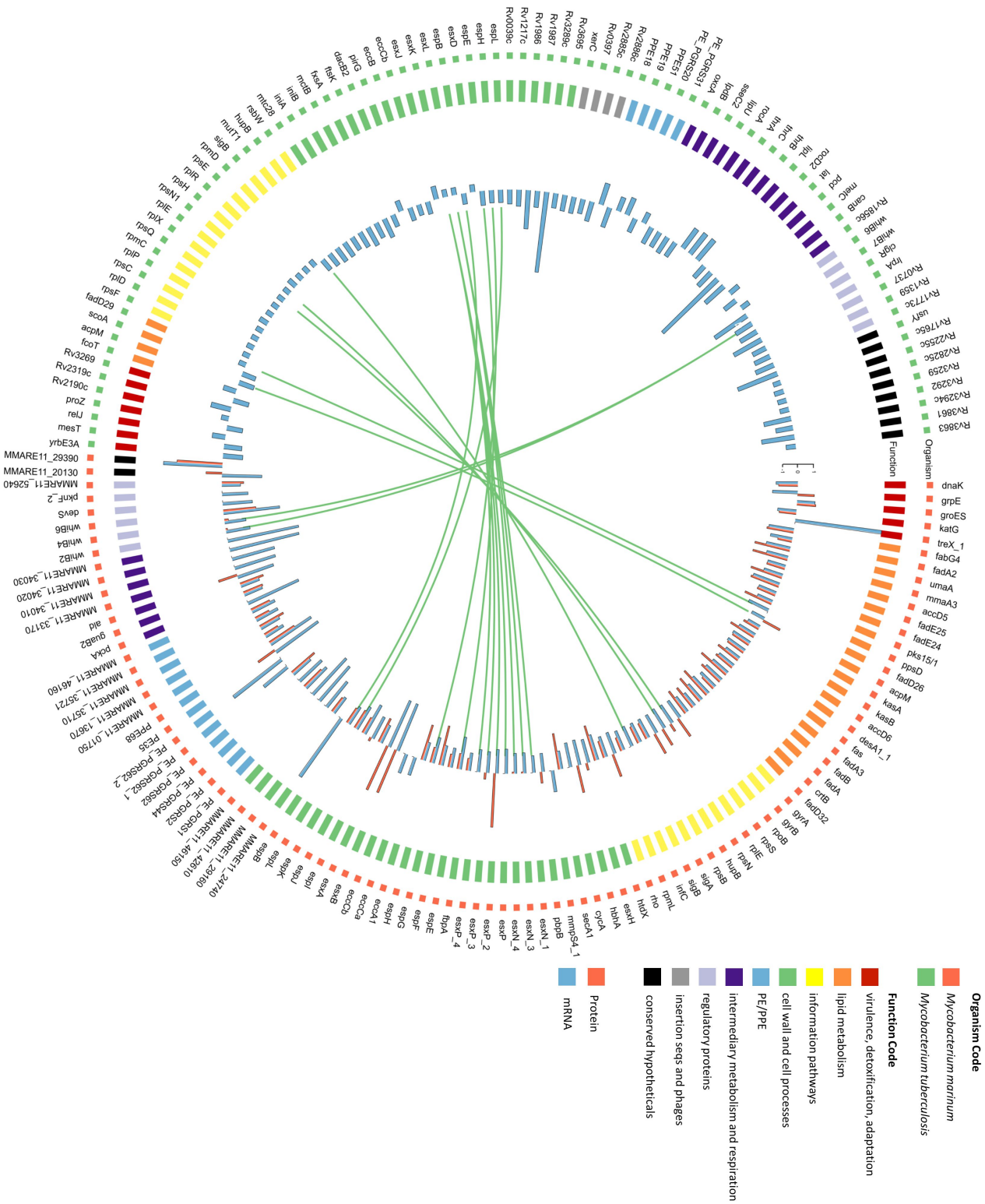


Figure.2

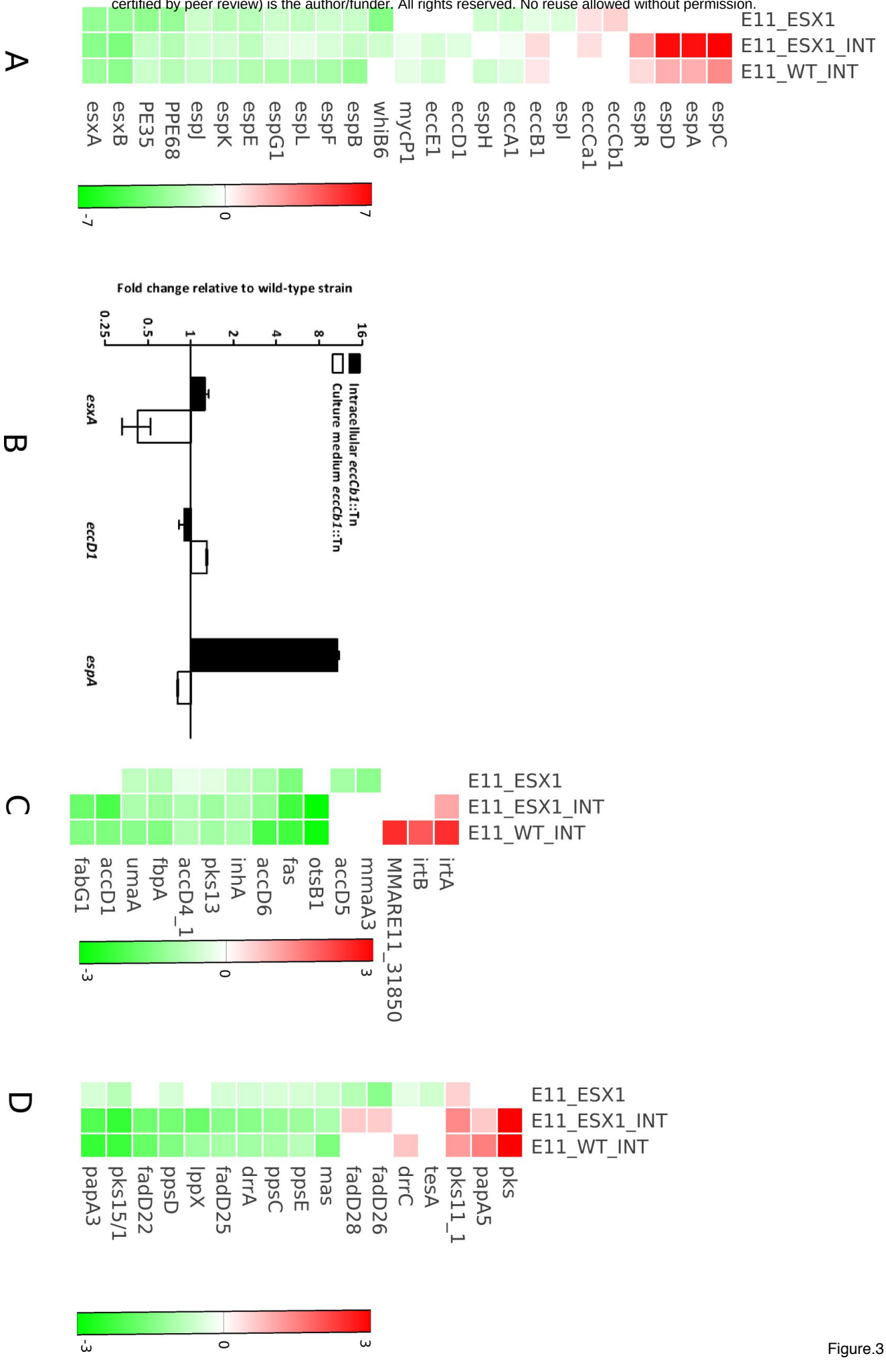


Figure.3

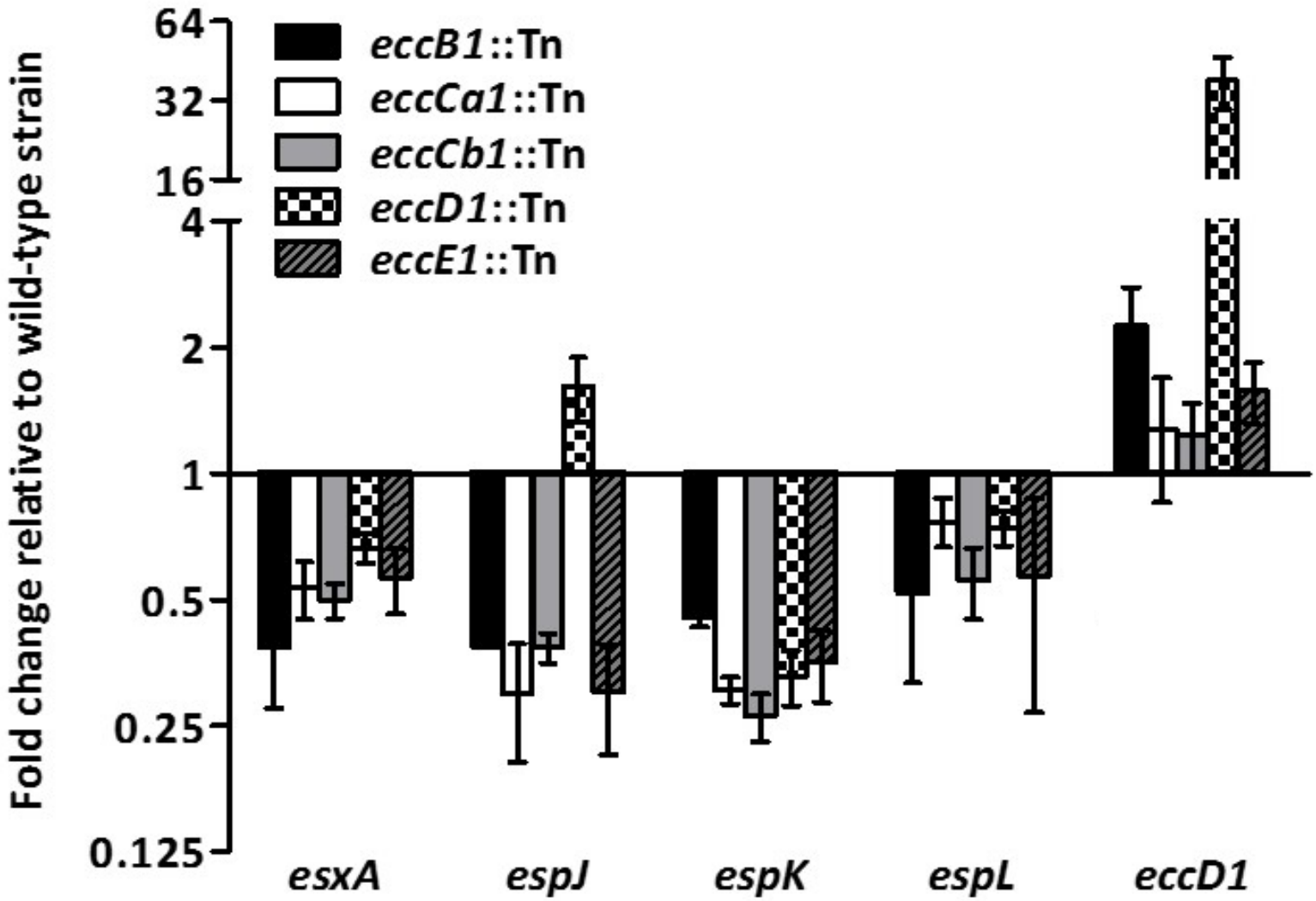
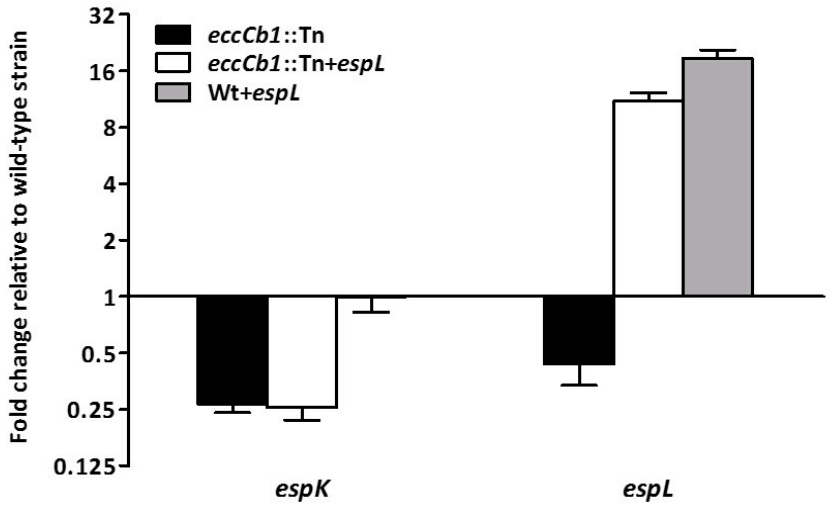
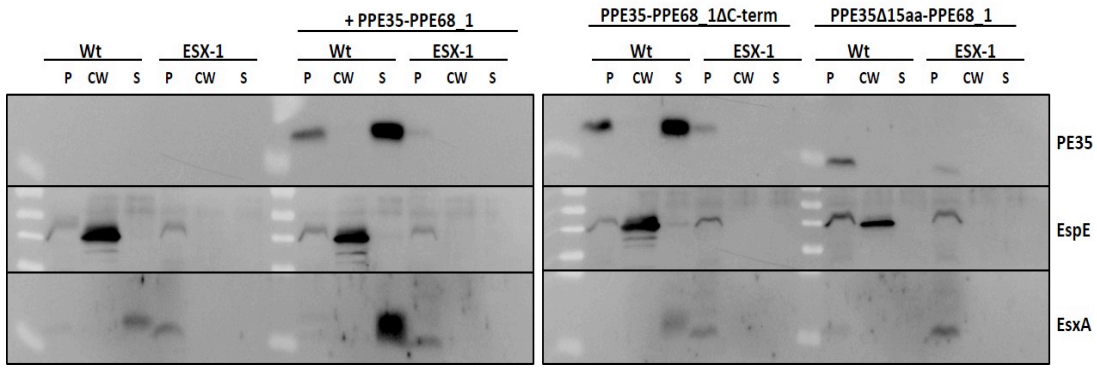


Figure.4

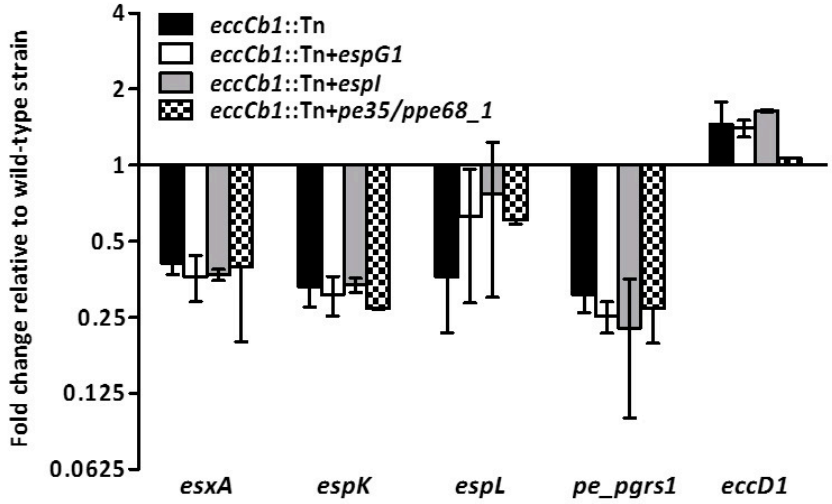
A



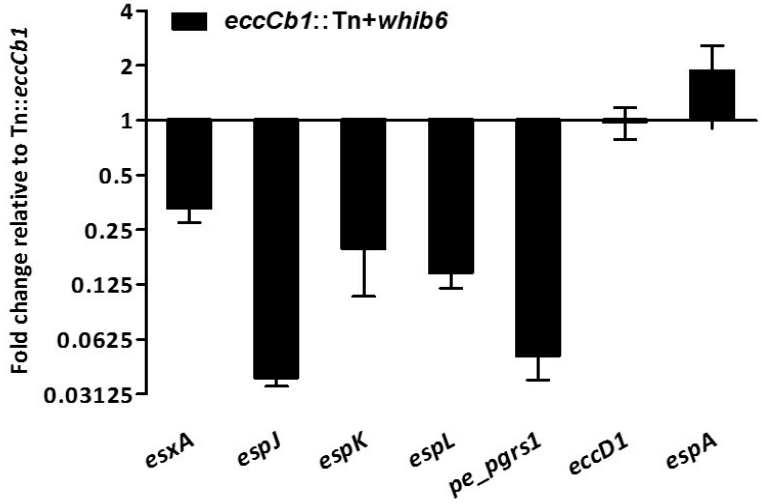
B



C



D



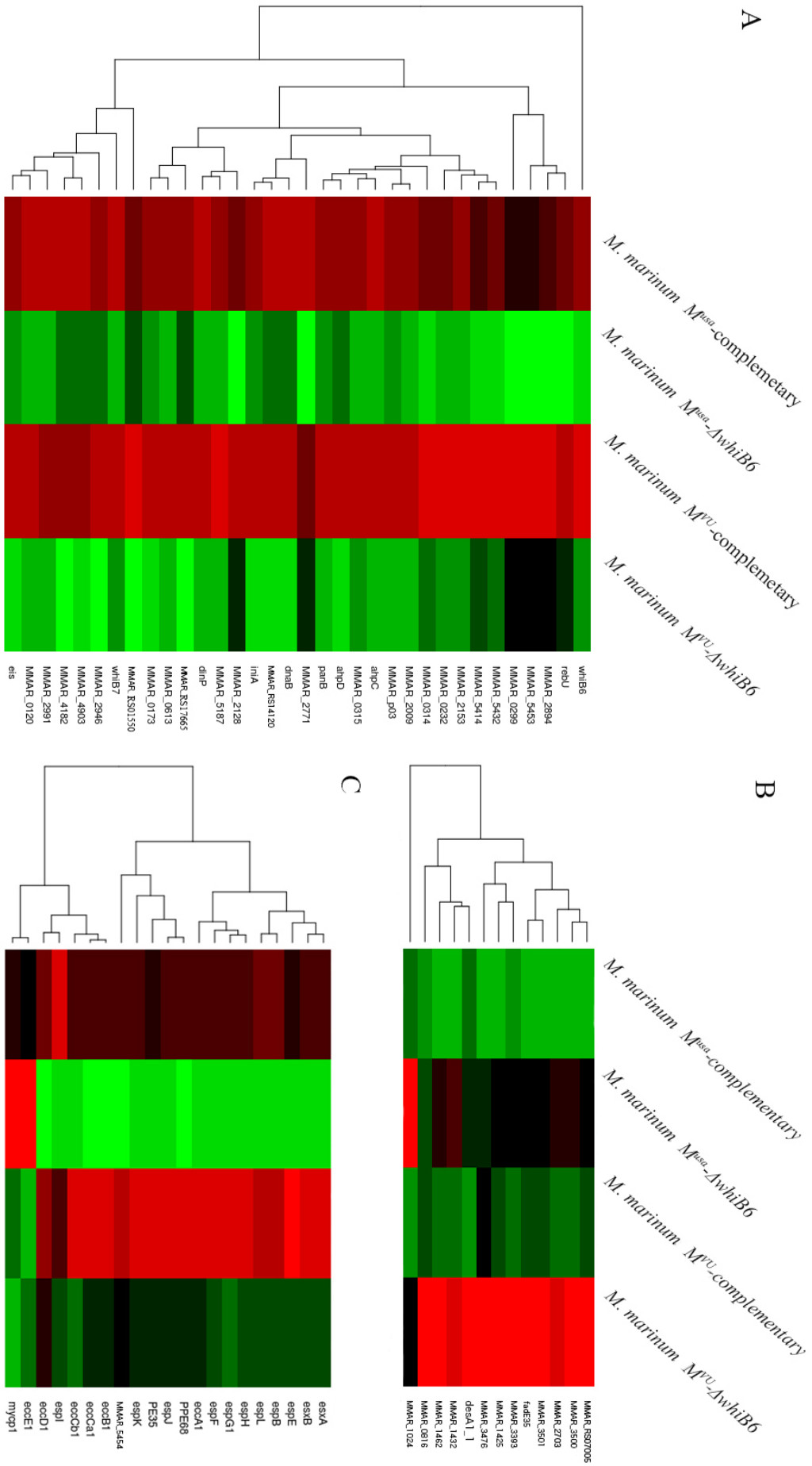


Figure.6