Published in final edited form as:

Nat Microbiol. 2023 March ; 8(3): 548–561. doi:10.1038/s41564-022-01313-7.

The *Mycobacterium tuberculosis* protein *O*-phosphorylation landscape

Andrew Frando[1,2], **Vishant Boradia**[1], **Marina Gritsenko**[3], **Claude Beltejar**[1], **Le Day**[3], **David R. Sherman**[4], **Shuyi Ma**[1,2,5,6], **Jon M. Jacobs**[3], **Christoph Grundner**[1,2,6]

[1]Center for Global Infectious Disease Research, Seattle Children's Research Institute, Seattle, WA, USA.

[2]Department of Global Health, University of Washington, Seattle, WA, USA.

[3]Pacific Northwest National Laboratory, Richland, WA, USA.

[4]Department of Microbiology, University of Washington, Seattle, WA, USA.

[5]Department of Chemical Engineering, University of Washington, Seattle, WA, USA.

[6]Department of Pediatrics, University of Washington, Seattle, WA, USA.

Abstract

Bacterial phosphosignalling has been synonymous with two-component systems and their histidine kinases, but many bacteria, including *Mycobacterium tuberculosis* (*Mtb*), also code for Ser/Thr protein kinases (STPKs). STPKs are the main phosphosignalling enzymes in eukaryotes but the full extent of phosphorylation on protein Ser/Thr and Tyr (*O*-phosphorylation) in bacteria is untested. Here we explored the global signalling capacity of the STPKs in *Mtb* using a panel of STPK loss-of-function and overexpression strains combined with mass spectrometry-based phosphoproteomics. A deep phosphoproteome with >14,000 unique phosphosites shows that *O*-phosphorylation in *Mtb* is a vastly underexplored protein modification that affects >80% of the

Correspondence and requests for materials should be addressed to Christoph Grundner. Christoph.grundner@seattlechildrens.org.

Author contributions

A.F. designed the work, was involved in sample generation, data acquisition, analysis and interpretation, and wrote the manuscript. V.B., M.G. and L.D. were ivolved in data acquisition. C.B., D.R.S. and S.M. were involved in data interpretation and discussion. J.M.J. was involved in data acquisition, analysis, interpretation and discussion. C.G. designed the work, was involved in data interpretation, organized the ideas, wrote the manuscript and was involved in the discussion.

Code availability

To assess the relative abundance of transcripts, we implemented a custom analysis pipeline in R v.3.3.0. The code used to process the RNA-seq reads is available at https://github.com/robertdouglasmorrison/DuffyNGS and https://github.com/robertdouglasmorrison/DuffyTools.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41564–022-01313–7.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41564–022-01313–7.

Peer review information *Nature Microbiology* thanks Yossef Av-Gay, Simone Lemeer and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

proteome and extensively interfaces with the transcriptional machinery. *Mtb O*-phosphorylation gives rise to an expansive, distributed and cooperative network of a complexity that has not previously been seen in bacteria and that is on par with eukaryotic phosphosignalling networks. A resource of >3,700 high-confidence direct substrate–STPK interactions and their transcriptional effects provides signalling context for >80% of *Mtb* proteins and allows the prediction and assembly of signalling pathways for mycobacterial physiology.

The histidine kinases of two-component systems (TCSs) have long been considered the mainstay of bacterial phosphosignalling, but phosphorylation on protein Ser/Thr and Tyr (*O*-phosphorylation) is emerging as another important bacterial phosphosignalling mechanism catalysed by Ser/Thr protein kinases (STPKs)[1,2]. Bacterial STPKs are termed 'eukaryotic-like' because of their similarity to the primary phosphosignalling enzymes in eukaryotes[3–5]. However, the pervasive presence of bacterial STPKs throughout the bacterial kingdom[6] and in archaea[7] and the presence of archetypical STPK sequences in bacteria suggest a common origin at the root of the tree of life. *Mycobacterium tuberculosis* (*Mtb*) encodes 11 STPKs, nearly as many as TCSs[8,9]. All but two STPKs are transmembrane receptors that are likely detecting changes in the environment[8]. In this way, the *Mtb* STPKs are positioned to be a major conduit for turning extracellular signals into bacterial adaptations.

Many studies have defined physiological substrates and functions for several of the *Mtb* STPKs related to transcription, metabolism[10,11], replication and cell division[12]. Phosphoproteomic studies provided the first global insights into *Mtb O*-phosphorylation[13,14]. These and other studies point to a more dominant role of *O*-phosphorylation in *Mtb* and perhaps other bacteria than is currently appreciated. Despite this emerging picture of *Mtb O*-phosphorylation as an important arm of bacterial and especially *Mtb* cell signalling, the full scope of *O*-phosphorylation is not known and most STPK cellular functions, signalling pathway organization, regulatory mechanisms and overall signalling network architecture remain to be characterized.

The similarity between eukaryotic and bacterial STPKs has made the better understood eukaryotic STPKs a template for understanding prokaryotic STPK signalling[15,16]. Eukaryotic STPKs engage in often extensive crosstalk that can modulate pathway output; unlike bacterial TCSs that typically only phosphorylate their specific response regulator, they can form branched signalling networks with multiple substrates[17]. Bacteria have between none and hundreds of STPKs. For example, *Sorangium cellulosum* codes for 317 STPKs[18], a number more similar to eukaryotic organisms that perhaps also implies a similar network organization. However, a system-wide analysis of a larger bacterial kinome is currently lacking. *Mtb* expresses 11 STPKs, a number that is small enough to be experimentally tractable yet large enough to potentially give rise to an *O*-phosphorylation signalling network. A more complex *O*-phosphorylation network structure in *Mtb* was previously suggested by a study of in vitro kinase–kinase interactions[19] and initial phosphoproteomic studies using single STPK perturbations[10,11]; however, the presence and architecture of a higher-order bacterial STPK *O*-phosphorylation system is undetermined. Importantly, STPKs are one of the major classes of drug targets in humans, with >60 drugs in the clinic[20]. A deeper understanding of *Mtb* STPK function could help translate these

successes to this human pathogen that is fuelling an enduring global pandemic for which few effective countermeasures are available.

To understand the full ensemble of *Mtb* STPKs and the signalling network they establish, we generated a panel of loss-of-function (LOF) and overexpression (OE) STPK strains and comprehensively characterized the phosphoproteomic and transcriptional effects of STPKs. We show that >70% of *Mtb* proteins are *O*-phosphorylated and draw a detailed map of >3,700 high-confidence kinase–substrate interactions. Remarkably, approximately 30% of *Mtb* gene expression was altered by STPK perturbation. This large impact of STPK signalling on transcription could be explained by the extensive interactions between STPKs, TCSs, and transcription factors (TFs). These data provide the deepest bacterial *O*-phosphoproteome to date, provide an expansive and detailed resource of thousands of *Mtb* STPK–substrate relationships, identify a large interface between *O*-phosphorylation and transcription in *Mtb*, predict and validate functional STPK–TF interactions, and together establish *O*-phosphorylation as a dominant signalling mechanism of a complexity not previously seen in bacteria.

Results

An STPK mutant panel

We sought to comprehensively characterize the *Mtb* phosphoproteome and test the effects of individual STPKs by measuring changes in phosphorylation and gene expression in STPK mutant strains. Nine *Mtb* STPKs have an extracellular sensor domain that reaches into the periplasmic space[21]. We anticipated that several have low basal activity in broth culture and are only fully activated by signals that are likely absent in vitro. The loss of such an inactive enzyme would not produce measurable effects. STPKs are typically activated by ligand-mediated oligomerization, a mode of activation that appears to be conserved in *Mtb*[15] and may be replicated by STPK OE. Thus, we employed a complementary perturbation approach that combines STPK LOF (knockdown or knockout) and OE to reveal substrates for all STPKs irrespective of their basal activation state.

All strains were generated in the *Mtb* H37Rv background. OE strains were engineered by inserting a plasmid expressing the complete STPK coding region under control of a tetracycline-inducible promoter. LOF strains were engineered by replacing all STPK genes except for the in vitro essential *pknA* and *pknB* by a hygromycin cassette using recombineering[22]. Expression of PknB was reduced by CRISPR interference (CRISPRi)[23]. PknA was not amenable to our complementary strategy due to toxicity of OE, prompting us to exclude PknA from our study. We grew all cultures to early stationary phase because slowed replication under nutrient limitation arguably mimics restrictive in vivo growth conditions more faithfully than exponential growth and because the stationary phase was previously associated with a maximum of *O*-phosphorylation in *Escherichia coli*[24]. We confirmed STPK knockout and induction by quantitative mass spectrometry (MS) using selected reaction monitoring (SRM). We observed a range of STPK abundances in wild-type (WT) strains, with several STPKs present at low concentrations (Extended Data Fig. 1a). All STPKs were absent or reduced to below the level of detection in the LOF strains, confirming knockout or knockdown, and all OE mutants showed robust induction (Extended Data Fig.

1a). We adjusted OE and knockdown so that none of the perturbations resulted in substantial effects on growth or survival (Extended Data Fig. 1b). In total, we generated 20 STPK mutant strains, a complete LOF/OE panel except for PknA.

A deep *Mtb* phosphoproteome

To define a comprehensive *Mtb O*-phosphoproteome, we analysed WT and mutant strains by MS-based phosphoproteomics. To maximize the number of total phosphosites in combination with precise quantitation, we measured all samples by MS using a label-free and tandem mass tag (TMT) labelling approach. Together, we analysed total protein abundance and phosphorylation in 3 biological replicates for each of the 20 mutant and WT strains, >150 phosphoproteomes and >10 million tandem MS (MS/MS) spectra in total. We only considered phosphosites that had a false discovery rate (FDR) < 0.01% at the peptide level, a posterior error probability (PEP) score < 0.05, a MaxQuant localization score > 0.99 (label-free) or an Ascore > 19 (TMT), corresponding to 99% confidence in phosphosite localization. An overview of the MS workflow is given in Extended Data Fig. 2.

We detected 3,232 of the approximate 4,200 predicted *Mtb* proteins, providing high coverage compared to previous *Mtb* proteogenomic studies[25] and in light that only approximately 80% of *Mtb* genes are expressed at a time[26]. We detected 14,452 unique *O*-phosphorylation sites on 2,617 *Mtb* proteins that met our criteria, or 80% of detected *Mtb* proteins. These data show a degree of *O*-phosphorylation in *Mtb* similar to or even exceeding that of eukaryotes, where between 50% and 70% of proteins are typically phosphorylated[27–29]. Our study identified over ten-fold more unique phosphosites than the next largest study of *Mtb*[11], five times more than all previously published *Mtb* phosphoproteome studies combined (Fig. 1a and Supplementary Table 1) and represents the largest bacterial phosphoproteome to date. The label-free and TMT approaches each contributed a large number of unique phosphosites, 4,208 and 6,125, respectively (Fig. 1b). The distribution of localization scores for sites detected by label-free and TMT data showed that most sites could be assigned with high confidence, with approximately 50% of assignments reaching >99% confidence (Fig. 1c). The relative distribution of pSer, pThr and pTyr detected in our study was 40% (5,816 sites), 52% (7,533 sites) and 8% (1,103 sites), a distinct phosphorylation profile compared to the eukaryotic or *E. coli* phosphoproteomes, where pSer is the most common modification (Fig. 1d). The share of pTyr was higher than that previously detected in *Mtb*[13] and more than ten-fold higher than in humans[29], indicating a more prominent role of pTyr in *Mtb* (Fig. 1d). Most phosphoproteins (>500) had a single phosphosite, while 1,743 proteins had between 2 and 10 sites. The number of proteins with more than 10 sites was small but one group of proteins had up to 84 phosphorylation sites (Fig. 1e). These highly phosphorylated proteins suggest a function of these sites outside of canonical phospho-switches. We next determined the number of phosphoproteins that are essential as defined by in vitro transposon mutagenesis[30]. Essential gene products were approximately 1.5-fold enriched in the phosphoproteome overall. However, a strong correlation between essentiality and the number of phosphorylation sites was apparent. For example, approximately 88% of those with 31–40 sites were essential ($P < 0.0001$), compared to approximately 20% essential genes in the genome overall (Fig. 1f).

The *Mtb* kinase–substrate relationships

We next sought to create a directory of individual STPKs and their cellular substrates. We defined substrates of a given STPK as either hypophosphorylated in the LOF of that kinase or hyperphosphorylated in the OE mutant strain or both. Proteins with phosphorylation changes in the opposite direction were considered indirect substrates. We compared the phosphorylation pattern of each site in every STPK mutant to that of the WT strain and calculated the changes in their abundance. Because of higher-quality quantitation data, we used TMT data for this analysis. We defined a significant change as one detected in all three replicates, changed by at least two-fold compared to WT and with $P < 0.005$. We also included on/off changes with no detected peptides in one set of replicates, although we could not calculate P values for these sites. We identified 3,503 significant changes in unique phosphosites across all mutants or 24% of all phosphosites (Fig. 2). The overall shifts in the phosphoproteome in response to kinase perturbation were largely as expected, with OE leading predominantly to increases in phosphorylation and LOF to decreases. The number of substrates varied largely by STPK from 15 to 968 (Fig. 2), suggesting widely varying signalling capacity of the different STPKs.

Consistent with our premise, many STPKs had low basal activity in the WT strain in broth culture. PknD, F, I, K and L showed little hypophosphorylation (<30 phosphosites) on deletion (Figs. 2 and 3a). However, these silent kinases could be effectively induced to phosphorylate substrates in the OE strains (Figs. 2 and 3a), allowing for the definition of a substrate set for each STPK (Supplementary Table 2). We next tested if STPK induction in the OE strains may lead to off-target phosphorylation. We found that the average magnitude and number of phosphorylation changes did not correlate with the abundance of the respective kinase (Extended Data Fig. 3a), which is consistent with OE leading to predominantly physiological phosphorylation events. To test whether OE of an STPK has effects on the phosphoproteome independent of its kinase activity, we generated a kinase-dead OE strain of PknK, which showed the second largest number of hyperphosphorylated substrates on OE of the WT strain. The kinase-dead PknK resulted in only 38 significant phosphorylation changes compared to 1,185 in the PknK WT OE strain, indicating that most phosphorylation changes are indeed due to kinase activity, not protein overproduction (Extended Data Fig. 3b).

Differential phosphorylation can result from changes in phosphoprotein abundance and/or from changes in occupancy of phosphosites. To understand this relationship, we determined the total abundance of proteins in all strains by MS and correlated abundance with phosphorylation (Fig. 3b). There was no significant correlation ($r = 0.09$) indicating that changes in phosphorylation primarily occur at the level of site occupancy. Protein abundance changes are given in Supplementary Table 2 for all proteins. Kinases interact with each other more than with other protein families and can profoundly affect each other's activity by phosphorylation[17,19]. A corollary of these interactions is that phosphorylation events detected on kinase perturbation may be indirect, which is difficult to discriminate from direct effects[31]. To identify indirect phosphorylation events, we examined the number of hyperphosphorylated sites in the LOF mutants and hypophosphorylated sites in the OE mutants because these cannot be explained by a direct STPK–substrate interaction (Fig.

3c). While all STPKs showed some of these indirect effects, their number was only approximately 8% of all altered phosphosites (Fig. 3c and Supplementary Table 2).

Protein Tyr phosphorylation was long thought to be absent in *Mtb* but has now been demonstrated conclusively[13]. However, in the absence of canonical bacterial Tyr kinases in *Mtb*, the identity of the Tyr kinases remained an open question. Based on in vitro biochemical activity, STPKs with dual specificity have been suggested as candidate Tyr kinases[13]. Our data provided the opportunity to test the hypothesis of dual specificity of STPKs in a cellular context. While all STPKs showed a preference for phosphorylating Ser/Thr over Tyr residues, many also phosphorylated Tyr residues. The pTyr sites affected by STPK permutation account for 231 pTyr sites or 24% of all pTyr sites detected in our study (Extended Data Fig. 4). These data suggest that some *Mtb* STPKs have evolved dual activity similar to that in many eukaryotes to perform a protein modification that is typically executed by specialized bacterial Tyr kinases.

STPKs constitute a distributed network

We next analysed the STPK–substrate relationships globally to detect larger features of the network. While 469 substrate proteins were only phosphorylated by a single STPK, 935 were phosphorylated by more than 1 STPK (Fig. 3d). Of the proteins phosphorylated by more than 1 STPK, most were phosphorylated by 2 STPKs (292). The same analysis at the level of individual phosphosites is shown in Extended Data Fig. 5. The high redundancy of this phosphorylation network is reminiscent of eukaryotic phosphosignalling networks[32]. We next inspected the STPK substrates showing significant changes on kinase perturbation for enrichment in several protein families. Among the most significantly enriched were the proteasome and Clp protease systems (Fig. 3e) and RNA polymerase components (Fig. 3f), almost all of which were phosphorylated by multiple STPKs, suggesting coordinated regulation of these central cellular functions by several STPK inputs.

STPKs have global transcriptional effects

We identified 158 or 77% of *Mtb* TFs[33] to be phosphorylated overall and 112 TFs (54% of TFs) with STPK-dependent phosphorylation changes, suggesting widespread regulation of TF activity by *O*-phosphorylation (Fig. 4a)[34] and predicting large STPK transcriptional effects. To test this idea, we measured the downstream transcriptional effects of STPKs by RNA sequencing (RNA-seq) using samples matched to the samples used for phosphoproteomics. We used a cut-off of at least four-fold upregulation or downregulation compared to WT and a significance cut-off of $P < 0.01$ to define differentially expressed genes (DEGs).

Overall, 1,155 of 4,030 detected genes met our criteria for DEGs, representing approximately 30% of *Mtb* genes (Fig. 4b and Extended Data Fig. 6). All STPKs caused gene expression changes and two distinct kinase activation patterns were apparent: several STPKs showed a larger change in gene expression on LOF (PknB and G), suggesting a constitutive housekeeping function. In particular PknB showed 30-fold more DEGs in the LOF than in the OE strain (Fig. 4b and Extended Data Fig. 5). Another group of STPKs including PknD, F, H and J showed no or little expression changes on LOF but large changes

on OE (Fig. 4b and Extended Data Fig. 6). The genes most affected by STPK permutation are shown in Fig. 4c. Except for PknE and PknF, the number of differentially regulated phosphosites and the number of DEGs corresponded well ($r = 0.92$) (Extended Data Fig. 7a). The number of DEGs was not correlated with the magnitude of STPK induction ($r = 0.37$), indicating few non-specific transcriptional effects linked to OE (Extended Data Fig. 7b). To test if the transcriptional effects of STPK perturbation are indeed due to the kinase activity of the STPKs, we measured the transcriptional effects of a kinase-dead PknL OE mutant. The number of DEGs was reduced from 260 to 1 in the mutant, clearly linking DEGs to kinase activity (Extended Data Fig. 8a). PknK also has a predicted transcriptional regulator domain of the MalT family. To test if gene expression changes in the PknK OE strain also reflected the effects of the MalT domain, we compared DEGs in strains overexpressing PknK WT and kinase-dead mutant. The WT OE mutant showed changes in 181 genes and the kinase-dead mutant in 13, suggesting that <10% of DEGs in the PknK OE mutant are due to the activity of the MalT domain rather than the kinase domain (Extended Data Fig. 8b). To understand the higher-level organization of the transcriptional effects of STPKs, we visualized the DEGs regulated by the STPK LOF and OE mutants (Fig. 4d). While some genes were affected by a single STPK, approximately 40% of genes were affected by several STPKs, some by up to 7. The co-regulation of these genes showed that signalling by multiple *Mtb* STPKs converges on the same cellular pathways.

Prediction of functional TF phosphosites

The share of functional phosphorylation events in any phosphoproteome is unknown and their prediction is a major challenge. Because we determined individual STPK–TF substrate pairs and global functional effects for the STPKs by measuring transcriptional changes, our data offered an opportunity to predict functional TF phosphorylation. To this end, we integrated our phosphoproteomic data on STPK–TF kinase–substrate pairs, the matched STPK transcriptional data and a transcriptional dataset mapping the transcriptional effects of approximately 200 *Mtb* TFs[33]. A significant overlap between the regulon of an STPK and that of its cognate TF substrate suggests a functional phosphorylation site. To correlate the regulons of phosphorylated TFs and those of their cognate STPKs, we determined the genes altered by an STPK and its TF partner and plotted the percentage overlap (Fig. 5a). Fifty-two TFs showed significant regulon overlap ($P = 0.05$) with their cognate STPK regulon. Twenty-four TFs showed >50% overlap and the regulons of 3 TFs were fully contained in their STPK regulon. These data predict a set of TFs whose phosphorylation is likely regulatory and leads to downstream gene expression changes.

O-phosphorylation alters Zur TF activity

To validate the STPK–TF functional predictions and identify new signalling pathways, we explored the interaction between PknK and the Zinc uptake repressor Zur (Rv2359), a negative transcriptional regulator of metal homeostasis[35]. The PknK OE mutant showed significant Zur hyperphosphorylation on Thr67 (8.7-fold) compared to WT, which is consistent with a direct interaction, and the phosphorylation event was predicted to be functional based on our STPK–TF regulon analysis (Fig. 5a). We confirmed a direct PknK–Zur interaction using an in vitro kinase assay with recombinant PknK and Zur, and PknG as a negative control (Fig. 5b). PknK, but not PknG, efficiently phosphorylated Zur. To test

for a regulatory connection between PknK and Zur, we compared the previously identified Zur LOF and OE regulons[33,35] to PknK OE mutant DEGs and found that >80% of the Zur regulon is also regulated in the PknK OE mutant, suggesting a function in the same pathway (Fig. 5c). While the Zur OE mutant strain repressed its regulon, which is consistent with its role as a negative regulator, the PknK OE mutant strain led to induction of nearly all Zur regulon genes. The opposite effects of Zur and PknK OE predicted that Zur phosphorylation by PknK derepresses Zur. To test the effect of Zur Thr67 phosphorylation on DNA binding, we purified recombinant WT, a Thr67 phosphoablative (Thr67Ala) and phosphomimetic (Thr67Asp) mutant and tested DNA binding activity by electrophoretic mobility shift assay (EMSA) (Fig. 5d). WT Zur efficiently bound to DNA and the phosphoablative mutant did not differ from WT in DNA binding. The phosphomimetic mutant, however, showed decreased DNA binding, which is consistent with our prediction. Together, these data define a phosphoregulation pathway in which PknK phosphorylates Thr67 of Zur and derepresses the Zur regulon (Fig. 5e).

A global resource for studying kinase function

The matched kinome-wide phosphoproteomic and transcriptomic datasets presented in this article provide a comprehensive resource to connect STPKs to individual substrate phosphosites and genes. Conversely, these data allow for the identification of STPKs affecting a particular physiological response if that response is characterized by the expression of a gene set or a change in phosphorylation. An overview of the combined interactions is shown in Fig. 6. In particular the OE data furnished a large set of phosphosites that were not accessible before. The complete dataset showing all transcriptional and phosphorylation effects of the STPK OE and LOF mutants are presented in Supplementary Table 1 (all phosphosites, label-free and TMT), Supplementary Table 2 (regulated phosphosites) and Supplementary Table 3 (transcriptional changes). Together, these data show that a large swath of *Mtb* physiology is integrated with STPK signalling, provide a resource for assigning kinases to substrates and regulated genes and thus the foundation for detailed mechanistic understanding of this pervasive *Mtb* phosphosignalling mechanism.

Discussion

The contribution of protein *O*-phosphorylation to bacterial cell signalling is an open question and the relatively small number of *O*-phosphorylation sites previously reported in bacteria have reinforced the notion that bacterial *O*-phosphorylation systems are simpler and more rudimentary than their eukaryotic counterparts. In this study, we challenged this view with the deepest phosphoproteome measured in a bacterium yet, which showed that *Mtb O*-phosphorylation is on par or exceeds the level of *O*-phosphorylation in even the most highly phosphorylated eukaryotes. The kinome-wide systems view afforded by our mutant panel revealed a signalling network of a complexity thought to be exclusive to eukaryotes. These findings place *O*-phosphorylation at the centre of cellular signalling in the pandemic human pathogen *Mtb* and suggest that STPKs constitute the backbone of bacterial phosphosignalling in *Mtb*.

To further understand phosphosignalling beyond the general inventory of the phosphoproteome requires parsing of contributions of individual kinases and their substrates, or the kinase–substrate relationships. Assigning kinase–substrate relationships is a major experimental bottleneck. Although several substrates and partial functional pathways of the *Mtb* STPKs have been identified previously, they are mostly unknown. Global phosphoproteomic approaches are arguably best suited to identify kinase–substrate relationships. These global approaches typically use LOF perturbations and then measure the phosphoproteomic changes[36]. While the changes resulting from an LOF perturbation are highly informative, this approach requires baseline kinase activity, which was an untested premise for most *Mtb* kinases grown in broth culture. Given that *Mtb* STPKs are mostly sensor kinases that likely survey the host, it is not surprising that many are largely inactive in broth culture, where host-specific signals are absent. Our OE approach captured many of these previously intractable phosphorylation events and downstream transcriptional effects of these silent STPKs.

Our study expands the overall number of *Mtb O*-phosphorylation sites and candidate substrates for the individual STPKs. The higher number of phosphosites identified in our study is likely a result of high-sensitivity MS hardware, combining two quantitative approaches with extensive fractionation, and the analysis of OE mutants. The extensive substrate sets for the STPKs presented in this study provide a directory for identifying STPK pathways and functions. One global functional consequence of STPK signalling was the sweeping effect on gene expression, which affected approximately 30% of all *Mtb* genes. The most direct connection between STPKs and gene expression is through TF phosphorylation, whose DNA binding can change on phosphorylation. Indeed, we uncovered extensive TF phosphorylation by the STPKs. A key question for any phosphosite is whether it is functional, but functional predictions are exceedingly difficult. Our data, by measuring both STPK–TF pairs and transcriptional effects, allowed prediction of functional phosphorylation sites on TFs and identified putative kinase signalling pathways. These data provide a rare experimental data point for gauging the share of functional phosphorylation sites within a large protein family. The many predicted functional STPK–TF connections and the PknK–Zur interaction suggest a general model of STPK-mediated gene regulation in *Mtb*—direct regulation of TFs by STPKs. The number of these interactions was surprising given the few previous examples of STPK–TF interactions reported in *Mtb*. Other direct interactions between STPKs and the transcriptional machinery suggested by our data include the STPK-specific phosphorylation of sigma factors, RNA polymerase, but also of TCSs. TCSs and STPKs were thought to be largely insulated from one another but extensive histidine kinase and response regulator phosphorylation by STPKs suggest close integration of the two arms of phosphosignalling. The complex and highly redundant architecture of the STPK signalling network may better translate complex signals into effective and robust responses than the typically more linear TCS pathways. The remarkable interconnectivity between signal transduction and gene regulation underscores how understanding the interplay of these networks will be critical to dissect how *Mtb* and potentially other bacteria sense, process and respond to environmental cues.

MS data are probabilistic in nature. We sought to minimize the uncertainty inherent in our phosphoproteomic data by setting stringent cut-offs. One limitation of our study is

the analysis of the phosphoproteome in only one condition and the fact that compensatory gene knockout effects can obscure STPK functions. The large effects of the STPK OE mutants on phosphorylation and transcription also raised the possibility of experimental artefacts. However, we found no evidence for capricious phosphorylation: the frequency of phosphorylation did not correlate with STPK expression and most changes in the OE strains were in fact changes in the abundance of phosphosites that were also present in the WT strains. The likelihood of an inadvertent functional impact of erratic phosphosites is small: phosphosignals propagate through evolved protein–protein interactions that are unlikely to be replicated by erratic phosphorylation events; the low occupancy of erratic phosphorylation would further reduce phenotypic consequences. However, a phosphosite or phenotypic effect that can be attributed to an LOF mutation may ultimately be more reliable than one in an OE mutant; the corresponding effects of LOF and OE mutants can further increase the confidence in the detected change.

On a system level, this study shows a central and dominant place of *O*-phosphorylation in the important human pathogen *Mtb* and a cooperative signalling system of a complexity that is typically only associated with eukaryotes and that has not previously been seen in bacteria. These findings revise, at least for *Mtb*, the long-held view of TCSs as the canonical bacterial phosphosignalling system. As a resource, this study provides rich detail of this system through >14,000 unique phosphosites and >3,700 STPK-specific phosphosite changes and their transcriptional effects. These data lay a foundation for understanding the cellular functions of these enzymes and provide signalling context for a large part of *Mtb* physiology.

Methods

Mtb STPK mutant strain generation

LOF mutants were constructed by recombineering[22]. The recombineering substrate was generated by amplifying 500 bp upstream and downstream of each STPK's coding sequence by PCR and cloning them 5′ and 3′ of a hygromycin resistance cassette, respectively. A linear recombineering substrate was amplified by PCR, purified and transformed into *Mtb* strain H37Rv carrying pNIT-ETc[37]. Successful integration was confirmed with DNA sequencing. OE mutants were constructed using the episomal pDTCF plasmid encoding a tetracycline-inducible promoter, C-terminal FLAG tag and a hygromycin-selectable marker. STPKs were amplified from the genomic DNA of *Mtb* strain H37Rv, cloned into the pDTCF plasmid 5′ of the FLAG tag using Gibson assembly (cat. no. E5510S; New England Biolabs) cloning and confirmed with sequencing. Plasmids were transformed into *Mtb* and expression was confirmed by anti-FLAG western blot. The *pknB* CRISPRi knockdown strain was constructed using a pJR965 plasmid encoding a tetracycline-inducible dCas9, a kanamycin-selectable marker and a tetracycline-inducible single-guide RNA targeting the 5′ coding region of *pknB* (5′-GTTCGTAGCGGTCGGACAGG-3′). STPK LOF and OE were confirmed by DNA sequencing and SRM and label-free liquid chromatography (LC)–MS/MS.

Mtb culture

All strains were grown in Middlebrook 7H9 (cat. no. 90003–876; VWR) with (10%) OADC (cat. no. B12351; Thermo Fisher Scientific) and 0.2% glycerol at 37 °C in rolling culture. For experiments involving STPK OE, cultures were grown in 50 μg ml^{-1} hygromycin B (cat. no. H-270–2-1; Goldbio). For experiments involving STPK CRISPRi knockdown, cultures were grown in 30 μg ml^{-1} kanamycin. The WT and STPK LOF strains were grown to stationary phase in triplicate. The STPK OE and CRISPRi knockdown strains were grown in triplicate to OD_{600} = 0.4–0.6, induced with 100 ng ml^{-1} anhydrotetracycline and grown to stationary phase.

Lysate production

Mtb cultures were pelleted at 4,000*g* for 5 min at 4 °C, washed with PBS and resuspended in lysis buffer (8 M urea; 50 mM Tris, pH 8.0, 75 mM NaCl, cOmplete protease inhibitor cocktail (cat. no. 5892791001; Sigma-Aldrich), 10 mM sodium fluoride (cat. no. 919; Sigma-Aldrich), 1% phosphatase inhibitor cocktail 2 (cat. no. P5726; Sigma-Aldrich), 1% phosphatase inhibitor cocktail 3 (cat. no. P0044; Sigma-Aldrich)). Lysates were generated by bead beating 3 times at 6 m s^{-1} for 30 s with cooling on ice between steps. Samples were centrifuged at 12,000*g* for 1 min and the supernatant was sterilized by filtration. Protein concentrations were determined by Bradford assay and the lysates were stored at −80 °C until further processing.

Proteomic and phosphoproteomic sample processing

A total of 1 mg of protein from each sample in 1 ml lysis buffer was incubated for 15 min at 4 °C. Protein concentrations were determined by bicinchoninic acid assay (BCA) (cat. no. 23227; Thermo Fisher Scientific). Proteins were reduced with 5 mM dithiothreitol (DTT) for 1 h at 37 °C and subsequently alkylated with 10 mM iodoacetamide for 45 min at 25 °C in the dark. Samples were diluted four-fold with 50 mM Tris-HCl, pH 8.0, to decrease a final concentration of urea below 2 M for digesting with Lys-C (Wako Chemicals) at 1:50 enzyme-to-substrate ratio. After 2 h of Lys-C digestion at 25 °C, sequencing grade-modified trypsin (cat. no. V5117; Promega Corporation) at 1:50 enzyme-to-substrate ratio was added to the samples and samples were incubated at 25 °C for 14 h. The reaction was stopped by acidifying the samples with 100% formic acid (Sigma-Aldrich) to a final concentration of 1% formic acid. Acidified samples were centrifuged for 15 min at 1,500*g* to clear digestion from possible precipitation. Tryptic peptides were desalted on 200-mg tC18 SepPak (cat. no. WAT054925; Waters) SPE and concentrated down using a Speed-Vac concentrator. The final peptide concentration was determined via BCA assay. Desalted peptide samples from each sample were labelled with 10-plex TMTs (Thermo Fisher Scientific) with one of the TMT channels (131) occupied with a pooled mixture of peptides from all the samples, which served as a reference to normalize across different sets of samples. Peptides (400 μg) were dissolved in 80 μl of 50 mM HEPES, pH 8.5 solution, and mixed with 400 μg of TMT reagent that was dissolved freshly in 20 μl of anhydrous acetonitrile. To quench the labelling reaction, 12 μl of 5% hydroxylamine was added to the samples for 15 min for incubation at room temperature. Peptides labelled by different TMT reagents were then mixed, dried in the Speed-Vac concentrator, reconstituted with 3% acetonitrile, 0.1% formic acid and

desalted on tC18 SepPak SPE columns. Then, 3.5 mg of 10-plex TMT-labelled sample was separated on a reversed-phase Agilent ZORBAX 300 Extend-C18 column (25 × 4.6 mm column containing 3.5 μm particles) using the Agilent 1200 HPLC System. Solvent A was 4.5 mM ammonium formate, pH 10, 2% acetonitrile and solvent B was 4.5 mM ammonium formate, pH 10, 90% acetonitrile. The flow rate was 1 ml min^{-1} and the injection volume was 900 μl. The LC gradient started with a linear increase of solvent B to 16% in 6 min, then increased linearly to 40% B in 60 min, 4 min to 44% B, 5 min to 60% B and another 14 of 60% solvent B. A total of 96 fractions were collected into a 96-well plate throughout the LC gradient. These fractions were concatenated into 6 fractions by combining 16 fractions into 1 for a total of 6 fractions. For proteome analysis, 5% of each concatenated fraction was dried and resuspended in 2% acetonitrile, 0.1% formic acid to a peptide concentration of 0.1 mg ml^{-1} for LC–MS/MS analysis. The rest of the fractions (95%) were subjected to immobilized metal affinity chromatography (IMAC) for phosphopeptide enrichment[38,39].

Phosphopeptide enrichment

Fe^{3+}-NTA agarose beads were freshly prepared using the Ni-NTA Superflow agarose beads (cat. no. 30410; QIAGEN) for phosphopeptide enrichment. For each of the 12 fractions (approximately 200 μg for each fraction) and for each label-free sample (100 μg), peptides were reconstituted to 0.5 μg $μl^{-1}$ in IMAC binding/wash buffer (80% acetonitrile, 0.1% trifluoroacetic acid) and incubated with 10 μl of the Fe^{3+}-NTA agarose beads for 30 min at room temperature. After incubation, the beads were washed 2 times each with 50 μl of wash buffer and once with 50 μl of 1% formic acid on the stage tip packed with 2 discs of Empore C18 material (Empore Octadecyl C18, 47 mm; cat. no. 98–0604-0217–3; CDS Analytical). Phosphopeptides were eluted from the beads on C18 using 70 μl of elution buffer (500 mM potassium phosphate buffer); 50% acetonitrile and 0.1% formic acid were used for the elution of phosphopeptides from the C18 stage tips. Samples were dried using Speed-Vac and later reconstituted with 12 μl of 3% acetonitrile and 0.1% formic acid containing 0.01% n-dodecyl-beta-maltoside for LC–MS/MS analysis.

Global and phosphoproteome analysis

Both TMT and label-free global and phosphopeptide-enriched samples were subjected to a custom high mass accuracy LC–MS/MS system as described previously[11], where the LC component consisted of automated reversed-phase columns prepared in house by slurry packing 3 μm Jupiter C18 (Phenomenex) into 35 cm × 360 μm o.d. × 75 μm i.d fused silica (Polymicro Technologies). The MS component consisted of a Q Exactive HF Hybrid Quadrupole-Orbitrap mass spectrometer (Thermo Fisher Scientific) outfitted with a custom electrospray ionization interface. Electrospray emitters were custom-made using a 360 μm o.d. × 20 μm i.d. chemically etched fused silica capillary. Analysis of the phosphoproteome samples applied similar conditions as used in the global proteome sample analysis, except that the spray voltage was 2.2 kV. All other instrument conditions were set as described previously[40]. Raw spectral data and analysis information is available via the MassIVE database under accession no. MSV000088254. LC–MS/MS raw data were converted into DTA files using Bioworks Cluster 3.2 (Thermo Fisher Scientific). The $MSGF^{+}$ algorithm was used to search the MS/MS spectra against the *Mtb* database (RefSeq H37Rv_uid57777_2014–08-14, 20,198 entries). Search parameters included 20

parts per million (ppm) tolerance for precursor ion masses, +2.5 Da and −1.5 Da window on fragment ion mass tolerances, no limit on missed cleavages, partial tryptic search, no exclusion of contaminants, dynamic oxidation of methionine (15.9949 Da), static iodoacetamide alkylation on cysteine (57.0215 Da) and static TMT modification of lysine and N termini (+144.1021 Da) for TMT analyses. After searching and mass correction, mass tolerances were limited to 5 ppm for the final identifications. The decoy database searching methodology[41,42] was used to control the FDR at the unique peptide level to <0.01% and subsequent protein level to <0.1%[43]. Quantification for TMT was based on initially summing to the peptide (phospho) or protein (global) level the sample-specific peptide reporter ion intensities captured for each channel across all 24 or 12 analytical fractions. The final data for the statistical analysis were the scaling and central tendency normalization of each peptide or protein summed value across all observations within each TMT10 experiment to adjust for experiment-specific variability. Label-free quantification used MaxQuant (v.1.6.2.10)[44] and generated label-free quantification values at the peptide (phospho) and protein (global) levels as described previously[45]. All collected protein and peptide abundance values were subjected to statistical comparison using WT as the basis for comparison. Primary comparisons were based on analysis of variance (ANOVA) at a $P <$ 0.005 threshold coupled with a fold change criterion minimum of ±2 between abundances.

SRM

SRM was performed on the panel of WT, OE and LOF mutants as described above. Crude heavy peptides labelled with $^{13}C/^{15}N$ on C-terminal lysine and arginine were purchased from vivitide. Trypsin-digested samples that had been stored at −80 °C until use were processed as described previously[46]. For each sample, the digested peptides were diluted to 0.25 μg μl^{-1} containing standards at a final concentration of 20 fmol μl^{-1}. All samples were analysed with a nanoACQUITY UPLC system (Waters Corporation) coupled online to a TSQ Altis triple quadrupole mass spectrometer (Thermo Fisher Scientific). The LC–SRM platform was configured and used as described previously[47]. The abundance of PknE and PknI in WT, PknE LOF mutant and PknI LOF mutant were too low to be detectable through the standard LC–SRM analysis. An extra PRISM procedure was applied to these low-abundance samples to enrich the targeted peptides for PknE and PknI before LC–SRM analysis[48]. SRM data acquired on the TSQ Altis were analysed with the Skyline-daily software (v.21.0.9.139)[49]. Peak detection and integration were determined based on retention time and the relative SRM peak intensity ratios across multiple transitions between light and heavy peptide standards[50]. All the data were manually inspected to ensure correct peak assignment and peak boundaries. The peak area ratios of endogenous light peptides and their heavy isotope-labelled internal standards (that is, L/H peak area ratios) were then automatically calculated by Skyline-daily and the average peak area ratios from all the transitions were used for quantitative analysis of the samples. For targets that had more than one surrogate peptide, correlation graphs were plotted to verify a strong correlation; ultimately, the peptide that had the most sensitive response was selected to obtain quantitative values.

RNA isolation and RNA-seq

Mtb cultures were pelleted at 4,000*g* for 5 min at 4 °C, resuspended in TRIzol (cat. no. 15596018; Thermo Fisher Scientific), transferred to a tube containing Lysing Matrix B (cat. no. 116911050-CF; MP Biomedicals) and lysed via bead beating using 30 s shaking at 6 m s^{-1} in a homogenizer three times, with cooling on ice. The samples were centrifuged at 21,000*g* for 1 min and the supernatant was transferred to a tube containing 300 µl chloroform and Heavy Phase Lock Gel (cat. no. 10847–802; VWR). The tubes were inverted for 2 min and centrifuged at 21,000*g* for 5 min. RNA in the aqueous phase was precipitated using 300 µl isopropanol and 300 µl high-salt solution (0.8 M Na citrate, 1.2 M NaCl). RNA was purified with an RNeasy kit (cat. no. 74004; QIAGEN) with one on-column DNase (cat. no. 79254; QIAGEN) treatment. Total RNA was quantified using a NanoDrop. Messenger RNA was enriched by first depleting ribosomal RNA from samples using the Ribo-Zero rRNA removal (bacteria) magnetic kit (cat. no. 20037135; Illumina). The purified mRNA was prepared for Illumina sequencing using the NEBNext Ultra RNA Library Prep Kit (cat. no. E7645L; New England Biolabs) for Illumina according to the manufacturer's instructions in combination with the AMPure XP reagent for clean-up of adaptor-ligated DNA. Each replicate was barcoded in the DNA library using the NEBNext Multiplex Oligos for Illumina (Dual Index Primers Set 1, cat. no. E7600S; New England Biolabs). Then, 30–40 libraries were multiplexed per sequencing run and libraries were quantified using the KAPA quantitative PCR quantification kit (cat. no. KK4844; Roche). The libraries were sequenced at the University of Washington Northwest Genomics Center with the Illumina NextSeq 500 High Output v2 Kit. Read alignment was performed using the custom processing pipeline that uses Bowtie 2 utilities, available at https://github.com/robertdouglasmorrison/DuffyNGS and https://github.com/robertdouglasmorrison/DuffyTools. Sequencing reads were mapped to the *Mtb* reference genome (NCBI GenBank accession no. AL123456.3) with the Bowtie 2 Alignment tool v.2.3.4.2. The BAM files were converted into read-depth wiggle tracks that record both uniquely mapped and multiply mapped reads to the genome at single-nucleotide resolution. Gene expression quantification was measured by summing the total reads located inside the annotated gene exon boundaries, expressed as reads per kilobase of transcript per million mapped reads, transcripts per million and raw read counts. Gene expression and differential gene expression analyses were performed using the R package DuffyNGS v.1.9.1. A panel of five differentially expressed tools was used: Round Robin (in house); Rank Product; Significance Analysis of Microarrays; EdgeR v.3.26.8; and DESeq2. All differentially expressed tools were compiled into a final result by combining fold changes, *P* values and rank positions from all five differentially expressed tools.

In vitro phosphorylation assay

The kinase domains of *pknB–pknL* were cloned from *Mtb* genomic DNA into an inducible pET28b protein expression plasmid using Gibson assembly cloning and confirmed by DNA sequencing. pET28b plasmids were transformed into *E. coli* BL21 for protein expression. Cultures were grown to 0.4–0.6 OD, induced with 250 µM isopropyl β-D-1-thiogalactopyranoside and grown at 20 °C for 18–24 h. Cultures were pelleted at 4,000*g* for 20 min, resuspended with lysis buffer (20 mM Tris-HCl, pH 7.5, 150 mM NaCl, 20 mM imidazole) and lysed by sonication. Lysates were centrifuged to remove the insoluble

fraction and purified using affinity chromatography with Ni-NTA. Proteins were further purified using size exclusion chromatography and stored in size exclusion buffer (20 mM Tris-HCl, 150 mM NaCl, 5% glycerol, pH 7.4). For the auto-phosphorylation assay, 1 μM STPK was combined with reaction buffer (50 μM $MnCl_2$, 50 μM $MgCl_2$, 1 mM DTT, 50 mM Tris, pH 8.0, 75 mM NaCl) and initiated with 50 μM ATP and 15 μCi ATP [γ−32P] (BLU002A2550UC; PerkinElmer). The reaction was incubated at 37 °C for 30 min. To terminate the reaction, 4× loading buffer with β-mercaptoethanol was added and incubated at 95 °C for 5 min. Proteins were separated by SDS–polyacrylamide gel electrophoresis (PAGE) in MES buffer and 4–12% Bis-Tris Plus gels (cat. no. NW04127BOX; Thermo Fisher Scientific). Gels were dried using a gel dryer at 80 °C for 3 h. Phosphor screens were exposed to the dried gels overnight and then imaged using a Sapphire biomolecular imager. For the Zur phosphorylation assay, 10 μM Rv2359 was combined with 1 μM STPK in reaction buffer (50 μM $MnCl_2$, 50 μM $MgCl_2$, 1 mM DTT, 50 mM Tris, pH 8.0, 75 mM NaCl), with 50 μM ATP and 15 μCi ATP [γ−32P], and then incubated at 37 °C for 30 min. Proteins were separated by SDS–PAGE and visualized as above.

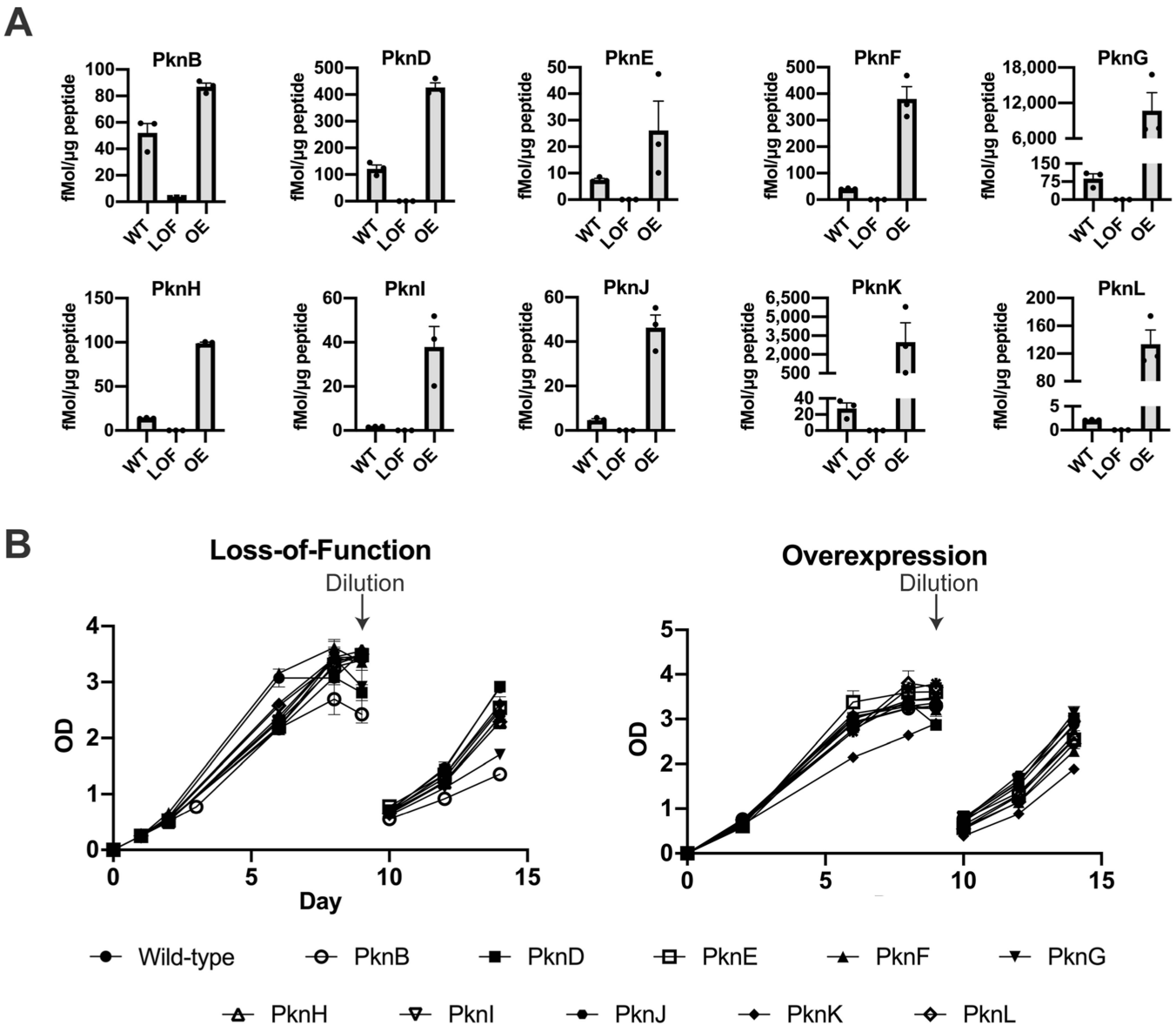
Zur mutagenesis

Zur phosphomimetic (Thr67Asp) and phosphoablative (Thr67Ala) mutants were generated by QuikChange site-directed mutagenesis. In summary, overlapping primers containing the desired mutation were used to linearize and introduce the mutation into the pET28b plasmid containing Zur. PCR products were gel-extracted and DpnI-digested, followed by transformation into competent *E. coli*. Mutagenesis was confirmed by DNA sequencing.

EMSA

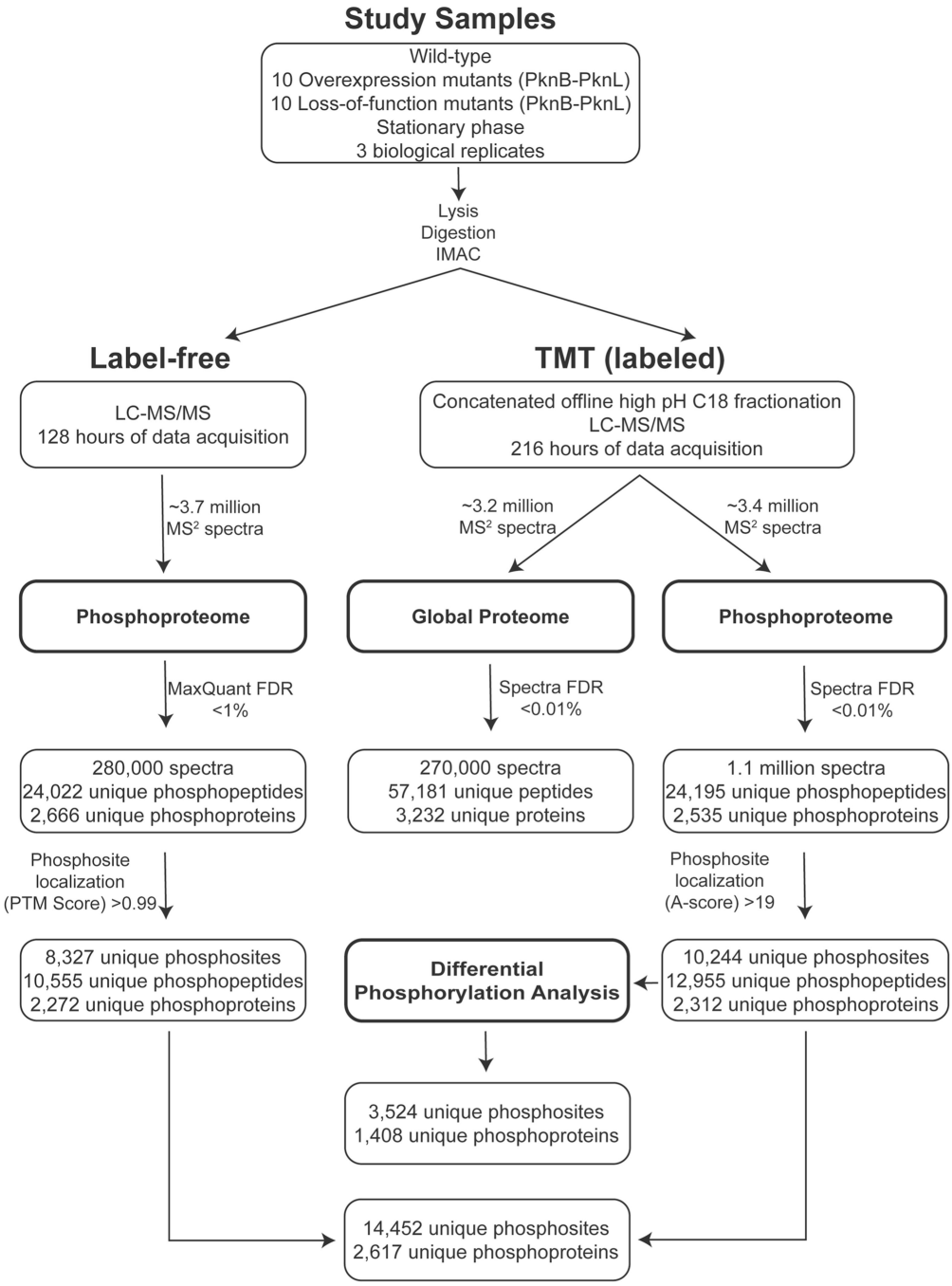
Cy5-labelled DNA probe was made by resuspending 3 oligonucleotides (Integrated DNA Technologies) to 100 μM in double-stranded DNA annealing buffer (10 mM Tris-HCl, pH 7.5, 100 mM NaCl, 1 mM EDTA). Oligonucleotide 1 corresponded to a 100–200 bp binding region of Rv2359. Oligonucleotide 2 corresponded to the reverse complement of oligonucleotide 1, as well as a 12-nucleotide sequence at the 3′ end to which oligonucleotide 3 was the reverse complement. Oligonucleotide 3 was a 12-*mer* with Cy5 covalently coupled to the 5′ end. The 3 oligonucleotides were combined to a final concentration of 100 μM, vortexed and heated to 95 °C for 10 min. The reaction was cooled to room temperature over 3–4 h protected from light and used for EMSA. For the EMSA, various concentrations of purified recombinant Rv2359 WT or mutant protein were combined with 2 μM final DNA probe in reaction buffer (20 mM Tris-HCl, pH 8.0, 75 mM NaCl, 1 mM DTT, 50 μM $MgCl_2$, 50 μM $MnCl_2$, 5% glycerol, 1% zinc sulphate, 50 μg ml^{-1} bovine serum albumin, 50 μg ml^{-1} salmon sperm DNA) and incubated for 30 min at room temperature, protected from light. After incubation, 4 μl native sample loading buffer was added to each sample. The whole reaction was loaded on a native 10% Tris-Glycine gel (cat. no. XP00105BOX; Thermo Fisher Scientific) and the gel was maintained on ice and ran at a constant 100 V for 2 h. The gel was rinsed with water and visualized on a LI-COR Odyssey scanner v.1.0.20. Bands were quantified using the Odyssey and Image Studio Acquisition software v.5.2.

Extended Data

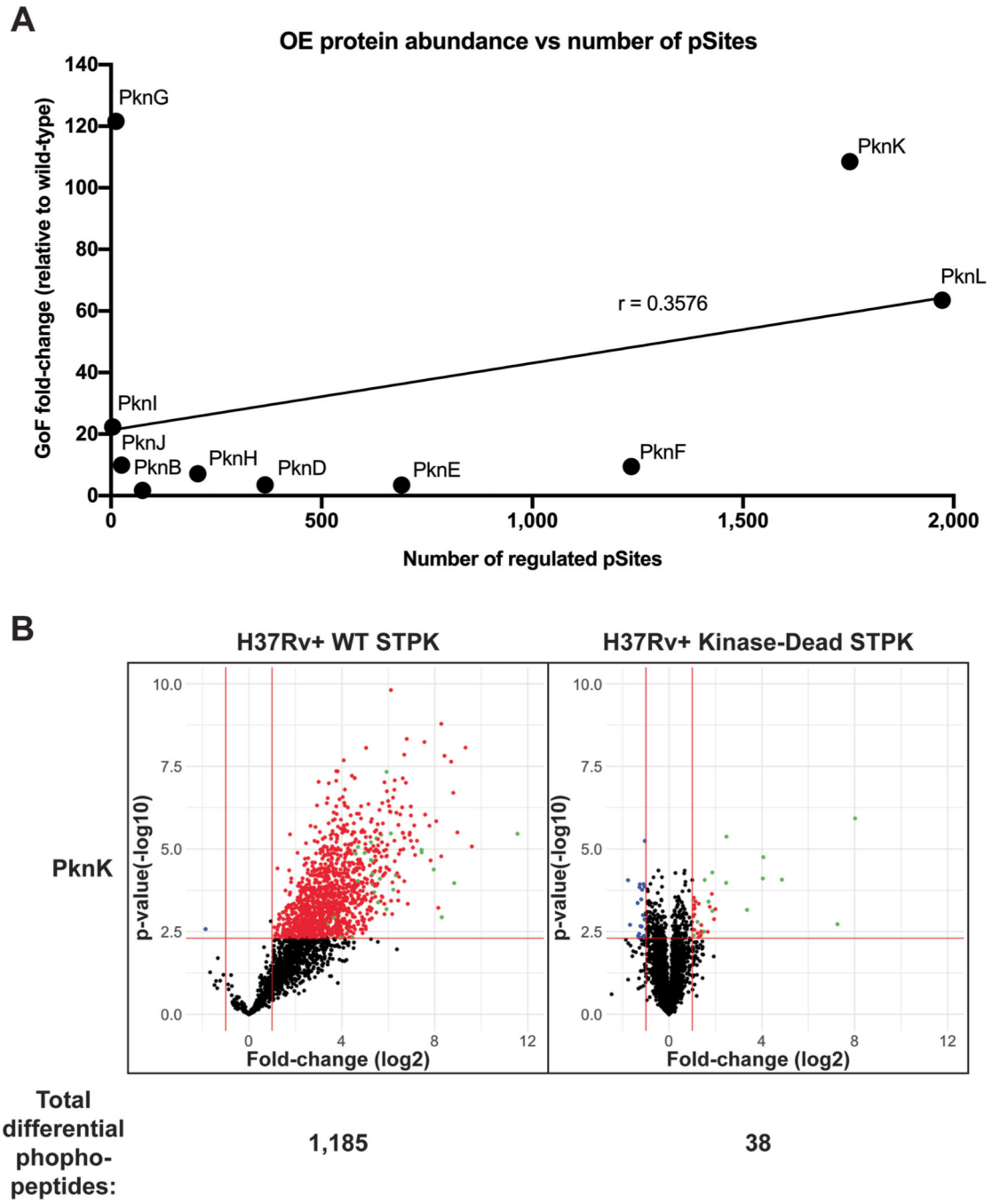


Extended Data Fig. 1 |. Characterization of STPK LOF and OE strains.

(**a**) Quantitation of mean STPK protein levels in LOF and OE *Mtb* mutant strains by selected reaction monitoring (n = 3 biologically independent replicates). PknB LOF is a knockdown strain, all other LOF strains are knockouts. Error bars show standard error of the mean. (**b**) Growth analysis of strains show only small growth effects of STPK perturbation and viability of strains in stationary phase (shown by regrowth after dilution). Error bars show standard error of the mean (n = 2 biologically independent replicates).

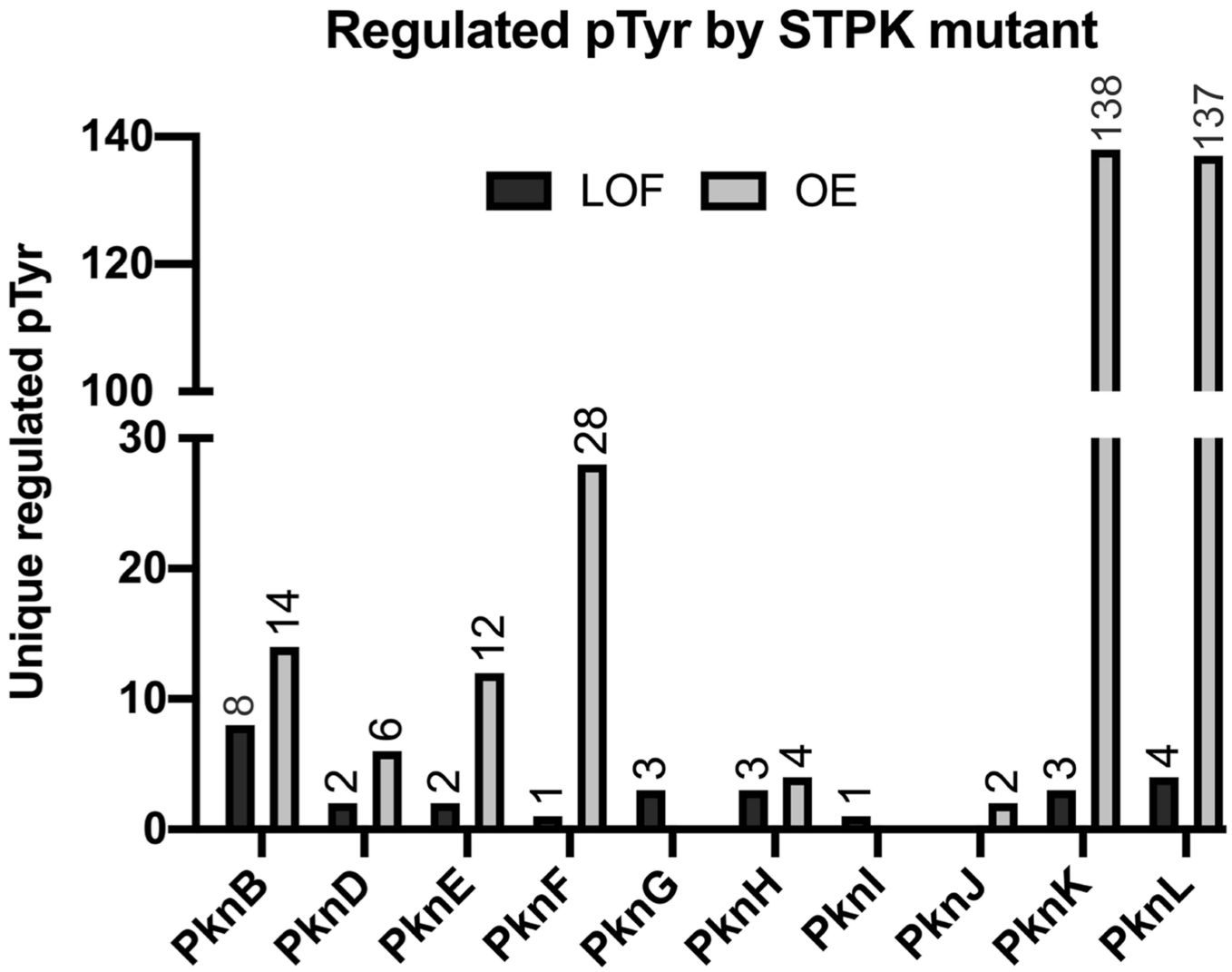


Extended Data Fig. 2 |. Overview of the MS experimental workflow.

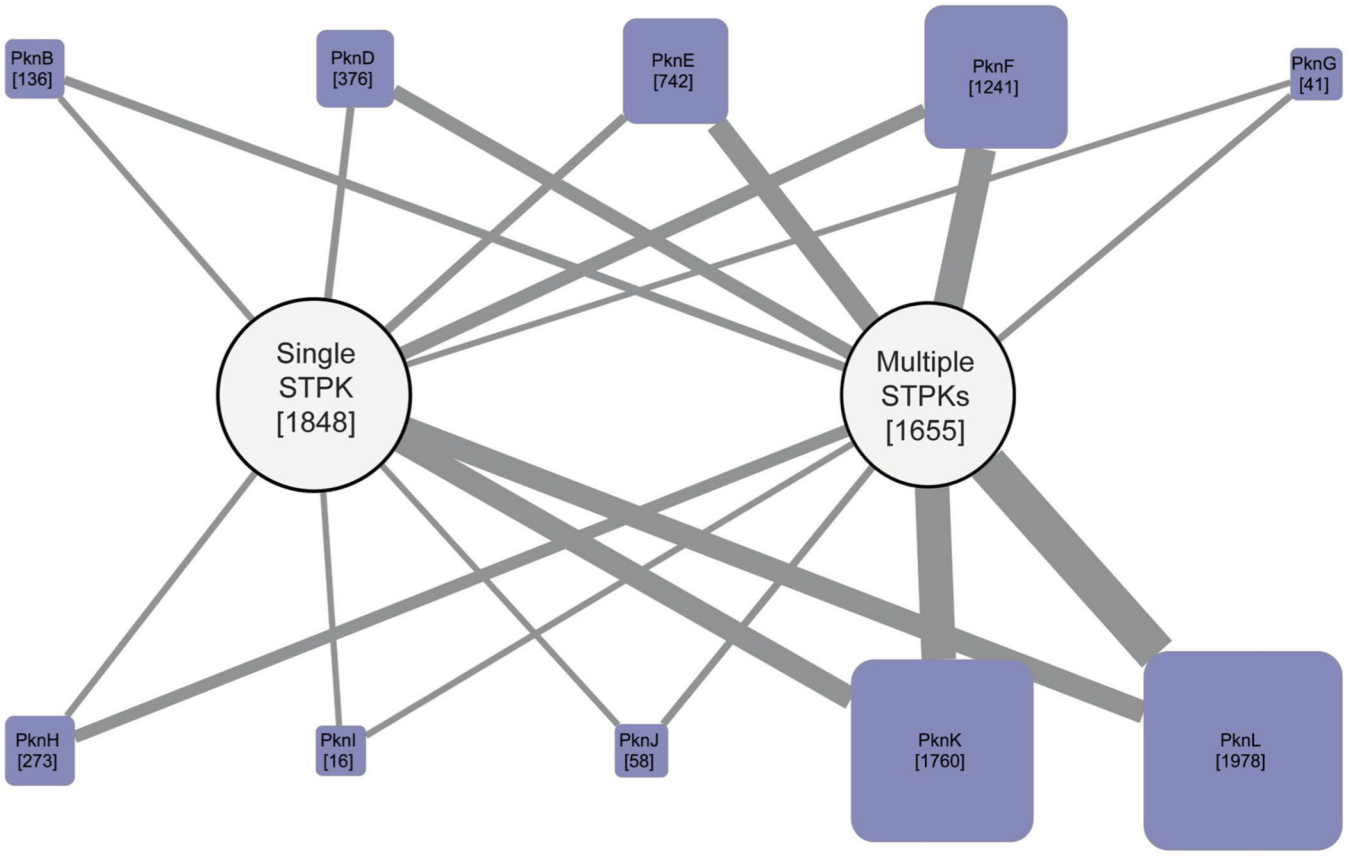


Extended Data Fig. 3 |. Correlation of phosphosites and STPK abundance and kinase-independent effects on phosphoproteome.

(**a**) The abundance of STPKs in the OE mutants does not overall correlate with the number of phosphosites detected in these strains. Statistically significant correlation was evaluated by calculating a Pearson correlation coefficient and comparing against a Student's *t*-distribution. (**b**) PknK kinase-dead OE causes only few changes in the phosphoproteome. Green dots indicate PknK peptides and show OE of PknK. Significance was determined by one-way ANOVA.

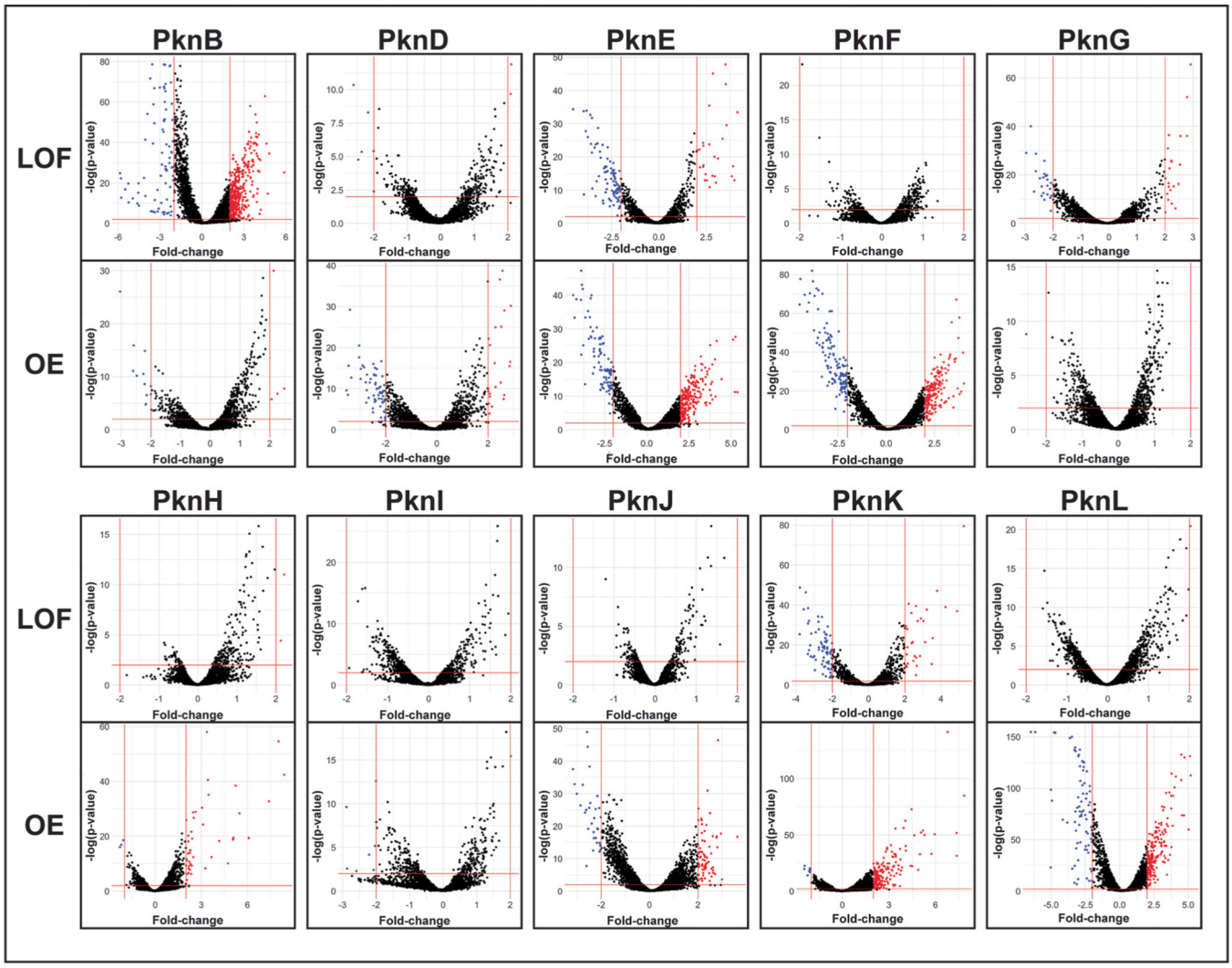


Extended Data Fig. 4 |. Contribution of STPKs to *Mtb* protein Tyr phosphorylation.
Number of Tyr phosphosites changed by STPK perturbation compared to WT.



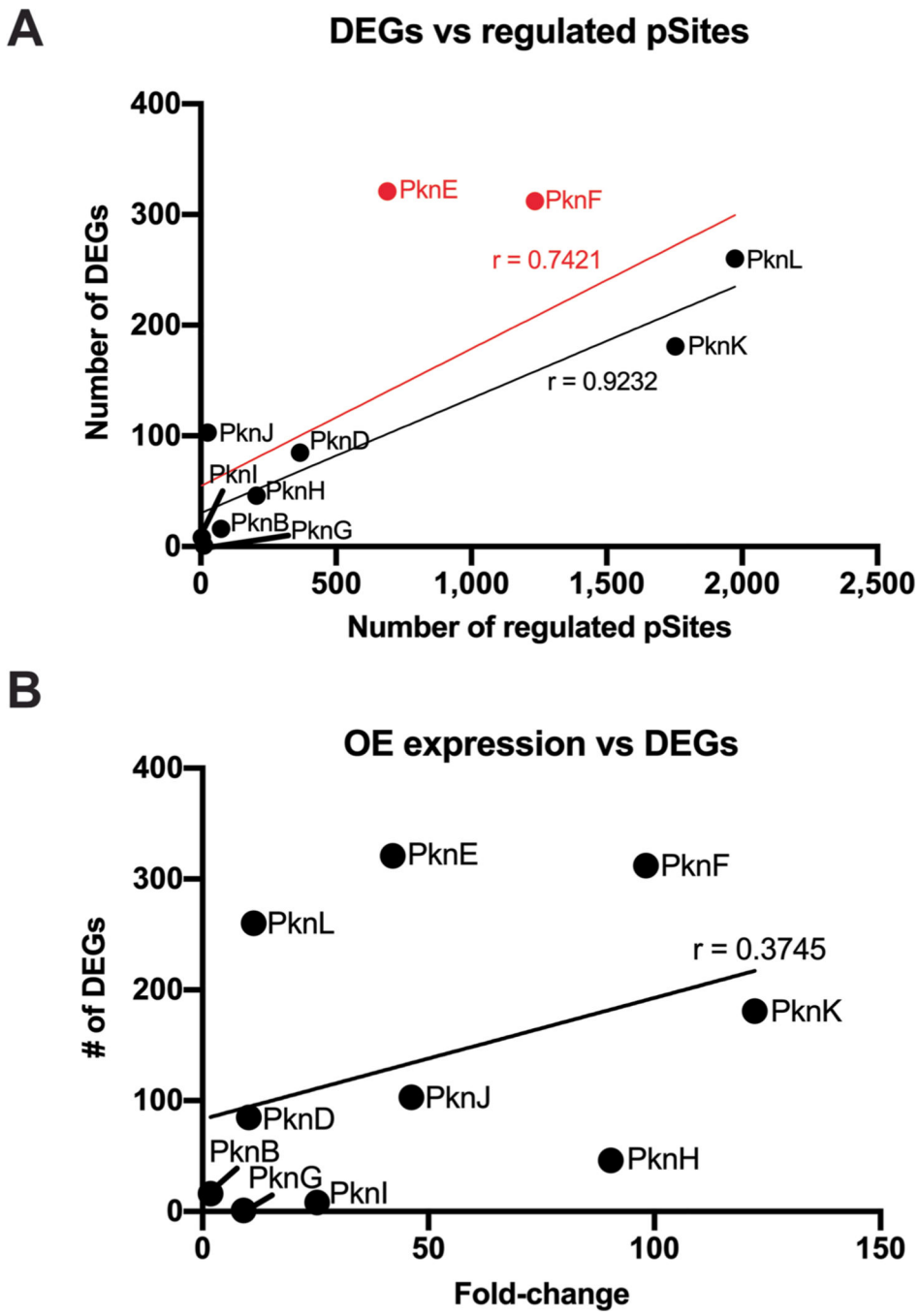
Extended Data Fig. 5 |. Phosphosite-level overview of STPK substrates and overlap in substrate phosphorylation.

The purple squares represent the STPKs. The size of the square is proportional to the number of that STPK's substrates, which is also given in parentheses. The round nodes show the number of sites phosphorylated by one or multiple STPKs. The thickness of the edge corresponds to the number of sites phosphorylated by the respective STPK.

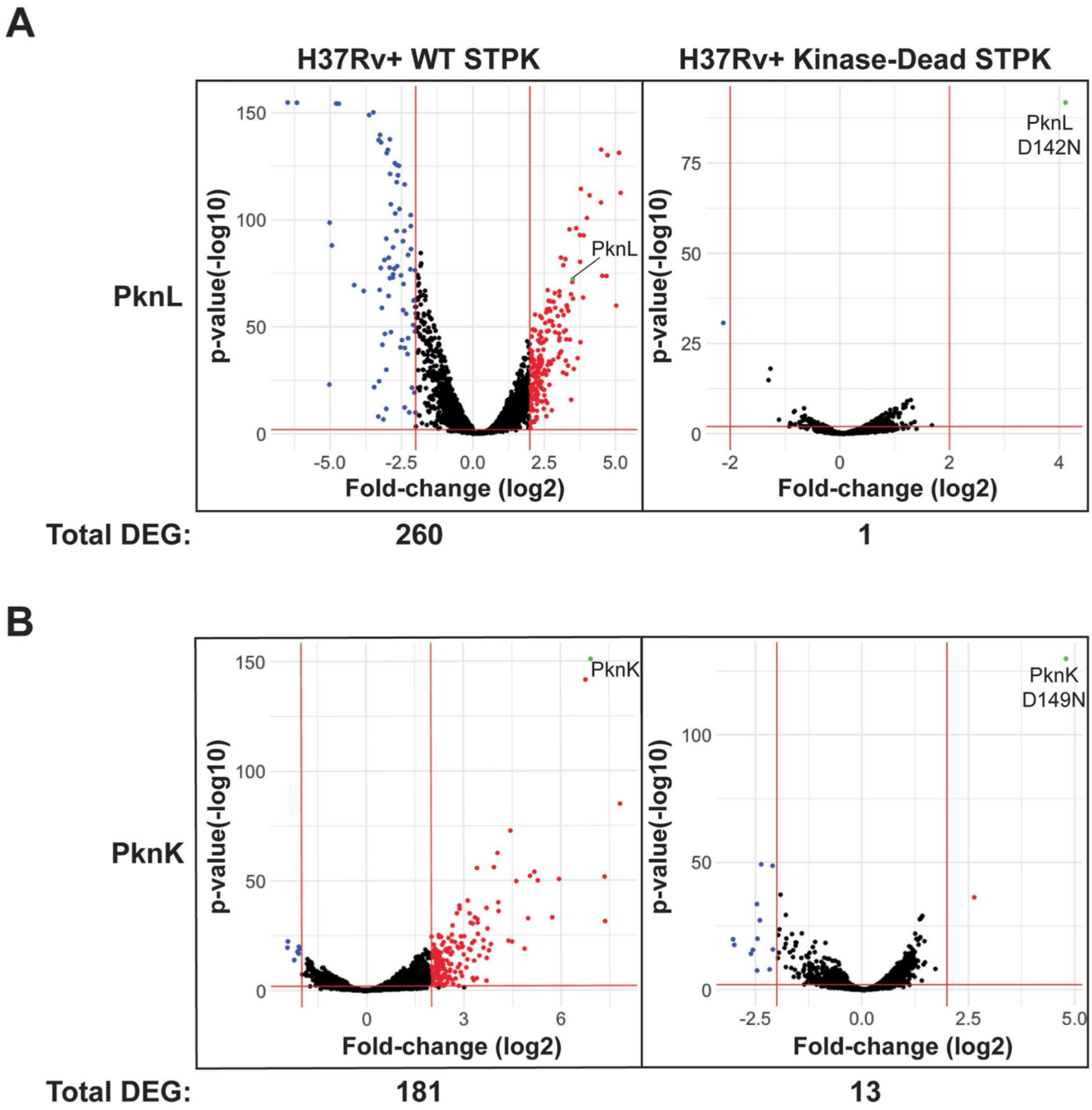


Extended Data Fig. 6 |. Transcriptional effects of STPK perturbation.

Volcano plots show the DEGs in STPK mutant strains, direction of change, magnitude, and *P* value associated with the change. Horizontal and vertical red lines show significance cutoffs (>4-fold, $P < 0.01$) used for further analysis of DEGs. The STPKs that were altered in the respective strain were removed from the data as they showed the largest changes and distorted the plots. Plots are shown with different scales to avoid compression of data points. Significance was determined by calculating the geometric mean of five DE tools.



Extended Data Fig. 7 |. Relationship between STPK induction, phosphosites, and DEGs.
(**a**) More phosphorylation sites in the OE mutant strains led to more DEGs. PknJ affects the expression of a large number of genes through few phosphosites. Red line shows linear regression if outliers PknE and PknF are excluded. (**b**) STPK abundance is not correlated with the number of DEGs in the OE strains. Statistically significant correlation was evaluated by calculating a Pearson correlation coefficient and comparing against a Student's *t*-distribution.



Extended Data Fig. 8 |. Transcriptional effects of a PknL and PknK kinase-dead mutant.

(**a**) The PknL kinase-dead mutant reduces the DEGs compared to WT OE from 260 to 1. PknL WT and kinase-dead transcripts are highlighted in green, showing induction in both cases. (**b**) The PknK kinase-dead mutant reduces DEGs compared to WT OE from 181 to 13, suggesting a small transcriptional effect of the PknK malT domain in the OE. PknK WT and kinase-dead transcripts are highlighted in green, showing induction in both cases. Horizontal and vertical red lines show significance cutoffs (>4-fold, $P < 0.01$) used for

further analysis of DEGs. Significance was determined by calculating the geometric mean of five DE tools.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work was supported by National Institutes of Health (NIH) grant nos. R01AI117023, R01AI158159, R21AI137571 and R03AI131223 and by a grant from the American Lung Association to C.G. A.F. was supported by the Interdisciplinary Program in Bacterial Pathogenesis no. 5T32AI053396. Portions of this research were supported by NIH National Institute of General Medical Sciences GM103493. Some of the work was performed in the Environmental Molecular Sciences Laboratory, a US Department of Energy Office of Biological and Environmental Research national scientific user facility located at Pacific Northwest National Laboratory. The Pacific Northwest National Laboratory is operated by Battelle for the US Department of Energy under contract no. DE-AC05–76RLO 1830.

Data availability

The mass spectrometry data reported in this paper are available in Supplementary Tables 1 and 2. Raw files are available under MassIVE accession no. MSV000088254 at https://massive.ucsd.edu/ProteoSAFe/dataset.jsp?task=e7a33adf4a424aef9eaad14ec5ddc5b4. The mass spectrometry data were searched against the *Mtb* database RefSeq H37Rv_uid57777_2014–08-14 with 20,198 entries. RNA-seq data are available in Supplementary Table 3. The RNA-seq data are deposited in the Gene Expression Omnibus under accession no. GSE195959. Sequencing reads were mapped to the *Mtb* (NCBI GenBank accession no. AL123456.3). The *Mtb* STPK mutant strains are available upon request. Source data are provided with this paper.

References

1. Dworkin J. Ser/Thr phosphorylation as a regulatory mechanism in bacteria. Curr. Opin. Microbiol. 24, 47–52 (2015). [PubMed: 25625314]
2. Stock AM, Robinson VL & Goudreau PN Two-component signal transduction. Annu. Rev. Biochem. 69, 183–215 (2000). [PubMed: 10966457]
3. Kannan N, Taylor SS, Zhai Y, Venter JC & Manning G. Structural and functional diversity of the microbial kinome. PLoS Biol. 5, e17 (2007). [PubMed: 17355172]
4. Krupa A. & Srinivasan N. Diversity in domain architectures of Ser/Thr kinases and their homologues in prokaryotes. BMC Genomics 6, 129 (2005). [PubMed: 16171520]
5. Leonard CJ, Aravind L. & Koonin EV Novel families of putative protein kinases in bacteria and archaea: evolution of the ‘eukaryotic’ protein kinase superfamily. Genome Res. 8, 1038–1047 (1998). [PubMed: 9799791]
6. Stancik IA et al. Serine/threonine protein kinases from bacteria, archaea and eukarya share a common evolutionary origin deeply rooted in the tree of life. J. Mol. Biol. 430, 27–32 (2018). [PubMed: 29138003]
7. Kennelly PJ Protein Ser/Thr/Tyr phosphorylation in the Archaea. J. Biol. Chem. 289, 9480–9487 (2014). [PubMed: 24554702]
8. Av-Gay Y. & Everett M. The eukaryotic-like Ser/Thr protein kinases of Mycobacterium tuberculosis. Trends Microbiol. 8, 238–244 (2000). [PubMed: 10785641]
9. Cole ST et al. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. Nature 393, 537–544 (1998). [PubMed: 9634230]

10. Carette X. et al. Multisystem analysis of Mycobacterium tuberculosis reveals kinase-dependent remodeling of the pathogen–environment interface. mBio 9, e02333–17 (2018). [PubMed: 29511081]
11. Zeng J. et al. Protein kinases PknA and PknB independently and coordinately regulate essential Mycobacterium tuberculosis physiologies and antimicrobial susceptibility. PLoS Pathog. 16, e1008452 (2020). [PubMed: 32255801]
12. Molle V. & Kremer L. Division and cell envelope regulation by Ser/Thr phosphorylation: Mycobacterium shows the way. Mol. Microbiol. 75, 1064–1077 (2010). [PubMed: 20487298]
13. Kusebauch U. et al. Mycobacterium tuberculosis supports protein tyrosine phosphorylation. Proc. Natl Acad. Sci. USA 111, 9265–9270 (2014). [PubMed: 24927537]
14. Prisic S. et al. Extensive phosphorylation with overlapping specificity by Mycobacterium tuberculosis serine/threonine protein kinases. Proc. Natl Acad. Sci. USA 107, 7521–7526 (2010). [PubMed: 20368441]
15. Alber T. Signaling mechanisms of the Mycobacterium tuberculosis receptor Ser/Thr protein kinases. Curr. Opin. Struct. Biol. 19, 650–657 (2009). [PubMed: 19914822]
16. Sherman DR & Grundner C. Agents of change—concepts in Mycobacterium tuberculosis Ser/Thr/Tyr phosphosignalling. Mol. Microbiol. 94, 231–241 (2014). [PubMed: 25099260]
17. Breitkreutz A. et al. A global protein kinase and phosphatase interaction network in yeast. Science 328, 1043–1046 (2010). [PubMed: 20489023]
18. Schneiker S. et al. Complete genome sequence of the myxobacterium Sorangium cellulosum. Nat. Biotechnol. 25, 1281–1289 (2007). [PubMed: 17965706]
19. Baer CE, Iavarone AT, Alber T. & Sassetti CM Biochemical and spatial coincidence in the provisional Ser/Thr protein kinase interaction network of Mycobacterium tuberculosis. J. Biol. Chem. 289, 20422–20433 (2014). [PubMed: 24928517]
20. Roskoski R Jr. Properties of FDA-approved small molecule protein kinase inhibitors: a 2021 update. Pharmacol. Res. 165, 105463 (2021). [PubMed: 33513356]
21. Prisic S. & Husson RN Mycobacterium tuberculosis serine/threonine protein kinases. Microbiol. Spectr. 10.1128/microbiolspec.MGM2-0006-2013 (2014).
22. van Kessel JC & Hatfull GF Recombineering in Mycobacterium tuberculosis. Nat. Methods 4, 147–152 (2007). [PubMed: 17179933]
23. Rock JM et al. Programmable transcriptional repression in mycobacteria using an orthogonal CRISPR interference platform. Nat. Microbiol. 2, 16274 (2017). [PubMed: 28165460]
24. Soares NC, Spät P, Krug K. & Macek B. Global dynamics of the Escherichia coli proteome and phosphoproteome during growth in minimal medium. J. Proteome Res. 12, 2611–2621 (2013). [PubMed: 23590516]
25. Kelkar DS et al. Proteogenomic analysis of Mycobacterium tuberculosis by high resolution mass spectrometry. Mol. Cell. Proteom. 10, M111 011627 (2011).
26. Arnvig KB et al. Sequence-based analysis uncovers an abundance of non-coding RNA in the total transcriptome of Mycobacterium tuberculosis. PLoS Pathog. 7, e1002342 (2011).
27. Hornbeck PV et al. PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. Nucleic Acids Res. 43, D512–D520 (2015). [PubMed: 25514926]
28. Olsen JV et al. Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. Sci. Signal. 3, ra3 (2010).
29. Sharma K. et al. Ultradeep human phosphoproteome reveals a distinct regulatory nature of Tyr and Ser/Thr-based signaling. Cell Rep. 8, 1583–1594 (2014). [PubMed: 25159151]
30. Griffin JE et al. High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. PLoS Pathog. 7, e1002251 (2011).
31. Kanshin E, Giguere S, Jing C, Tyers M. & Thibault P. Machine learning of global phosphoproteomic profiles enables discrimination of direct versus indirect kinase substrates. Mol. Cell. Proteomics 16, 786–798 (2017). [PubMed: 28265048]
32. Bodenmiller B. et al. Phosphoproteomic analysis reveals interconnected system-wide responses to perturbations of kinases and phosphatases in yeast. Sci. Signal. 3, rs4 (2010).

33. Rustad TR et al. Mapping and manipulating the Mycobacterium tuberculosis transcriptome using a transcription factor overexpression-derived regulatory network. Genome Biol. 15, 502 (2014). [PubMed: 25380655]
34. Hunter T. & Karin M. The regulation of transcription by phosphorylation. Cell 70, 375–387 (1992). [PubMed: 1643656]
35. Maciag A. et al. Global analysis of the Mycobacterium tuberculosis Zur (FurB) regulon. J. Bacteriol. 189, 730–740 (2007). [PubMed: 17098899]
36. Baros SS, Blackburn JM & Soares NC Phosphoproteomic approaches to discover novel substrates of mycobacterial Ser/Thr protein kinases. Mol. Cell. Proteomics 19, 233–244 (2020). [PubMed: 31839597]
37. van Kessel JC & Hatfull GF Mycobacterial recombineering. Methods Mol. Biol. 435, 203–215 (2008). [PubMed: 18370078]
38. Wang Y. et al. Reversed-phase chromatography with multiple fraction concatenation strategy for proteome profiling of human MCF10A cells. Proteomics 11, 2019–2026 (2011). [PubMed: 21500348]
39. Mertins P. et al. Reproducible workflow for multiplexed deep-scale proteome and phosphoproteome analysis of tumor tissues by liquid chromatography–mass spectrometry. Nat. Protoc. 13, 1632–1661 (2018). [PubMed: 29988108]
40. Zhang H. et al. Integrated proteogenomic characterization of human high-grade serous ovarian cancer. Cell 166, 755–765 (2016). [PubMed: 27372738]
41. Elias JE & Gygi SP Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nat. Methods 4, 207–214 (2007). [PubMed: 17327847]
42. Qian W-J et al. Probability-based evaluation of peptide and protein identifications from tandem mass spectrometry and SEQUEST analysis: the human proteome. J. Proteome Res. 4, 53–62 (2005). [PubMed: 15707357]
43. Kim S, Gupta N. & Pevzner PA Spectral probabilities and generating functions of tandem mass spectra: a strike against decoy databases. J. Proteome Res. 7, 3354–3363 (2008). [PubMed: 18597511]
44. Cox J. & Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. 26, 1367–1372 (2008). [PubMed: 19029910]
45. Clair G. et al. Spatially-resolved proteomics: rapid quantitative analysis of laser capture microdissected alveolar tissue samples. Sci. Rep. 6, 39223 (2016). [PubMed: 28004771]
46. Shi T. et al. Targeted quantification of low ng/ml level proteins in human serum without immunoaffinity depletion. J. Proteome Res. 12, 3353–3361 (2013). [PubMed: 23763644]
47. Nielson CM et al. Free 25-hydroxyvitamin D: impact of vitamin D binding protein assays on racial-genotypic associations. J. Clin. Endocrinol. Metab. 101, 2226–2234 (2016). [PubMed: 27007693]
48. Shi T. et al. Antibody-free, targeted mass-spectrometric approach for quantification of proteins at low picogram per milliliter levels in human plasma/serum. Proc. Natl Acad. Sci. USA 109, 15395–15400 (2012). [PubMed: 22949669]
49. MacLean B. et al. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics 26, 966–968 (2010). [PubMed: 20147306]
50. He J. Antibody-independent targeted quantification of TMPRSS2-ERG fusion protein products in prostate cancer. Mol. Oncol. 8, 1169–1180 (2014). [PubMed: 25266362]

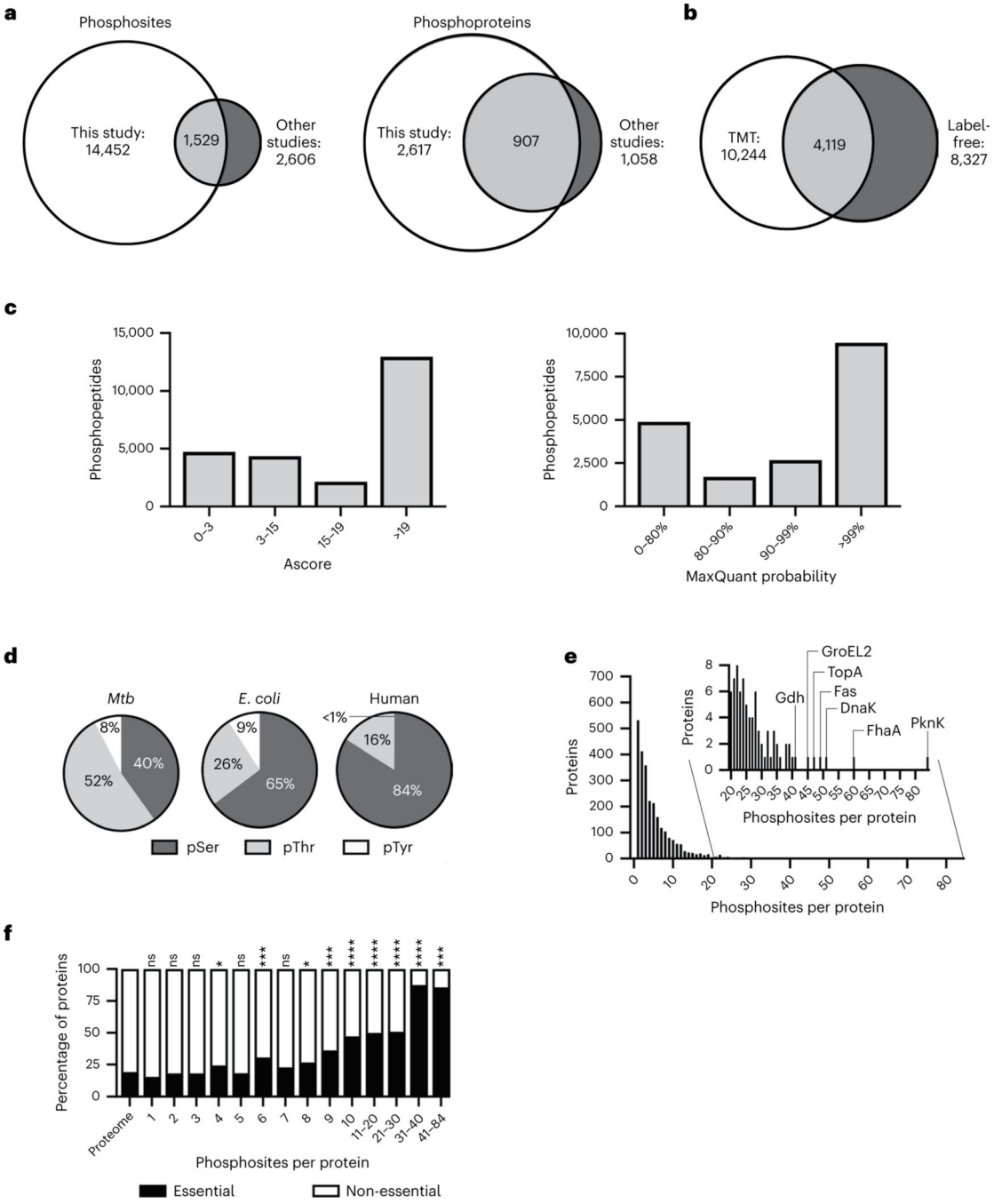


Fig. 1 |. A deep *Mtb* phosphoproteome.

a, Comparison of the number of unique protein *O*-phosphorylation sites (left) and proteins (right) identified in this study and previously published. **b**, Number of unique phosphosites identified by the TMT and label-free approaches. **c**, Distribution of phosphosite localization Ascore (for the TMT data) and MaxQuant probability (for the label-free data). An Ascore of 19 corresponds to 99% probability of correct localization. **d**, Relative share of pSer, pThr and pTyr sites in *Mtb*, *E. coli* and human cells. **e**, Number of phosphorylation sites per protein and highly phosphorylated proteins with >20 phosphosites (inset). **f**, Share of

phosphorylated essential gene products stratified by the number of phosphosites. Essentiality was based on transposon mutagenesis (30). *P* values were calculated using a hypergeometric test for enrichment and are $*P < 0.05$, $**P < 0.01$, $***P < 0.001$ and $****P < 0.0001$.

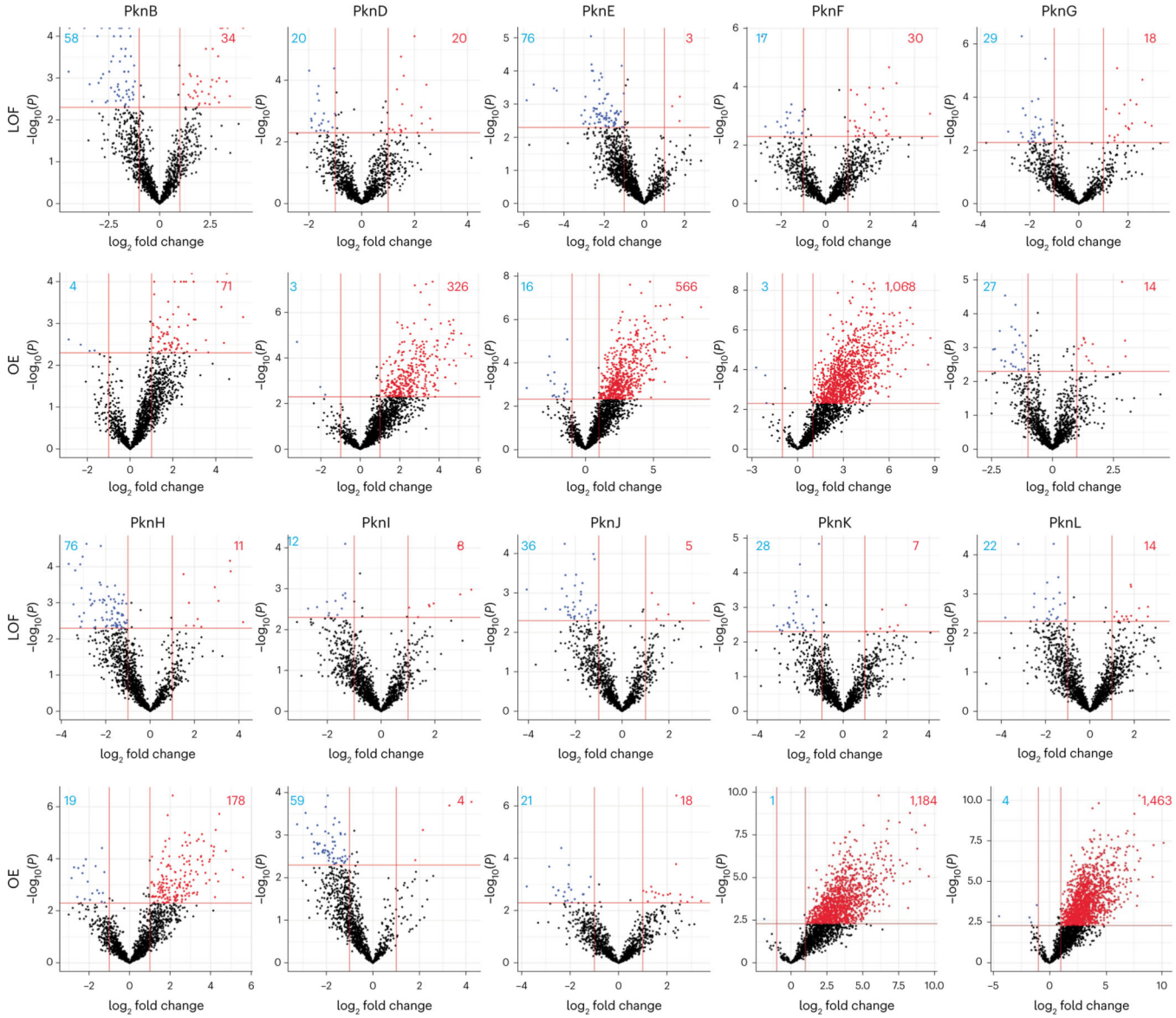


Fig. 2 |. STPK perturbation defines kinase–substrate interactions.

Magnitude and direction of changes in phosphorylation sites in LOF (rows 1 and 3) and OE strains (rows 2 and 4). The volcano plots show $\log_2$ fold changes in phosphorylation levels by phosphosite (x axis) versus the $-\log_{10}$ P value of the change (y axis). The horizontal and vertical lines show significance cut-offs applied to the data for further analysis (greater than two-fold change in abundance, $P < 0.005$). Significantly changing phosphosites are shown in red and blue and their numbers are given in the respective quadrant. STPKs altered in the respective strain were not included because they showed the largest changes and distorted the plots. Plots are shown with different scales to avoid compression of data points. Significance was determined by one-way ANOVA.

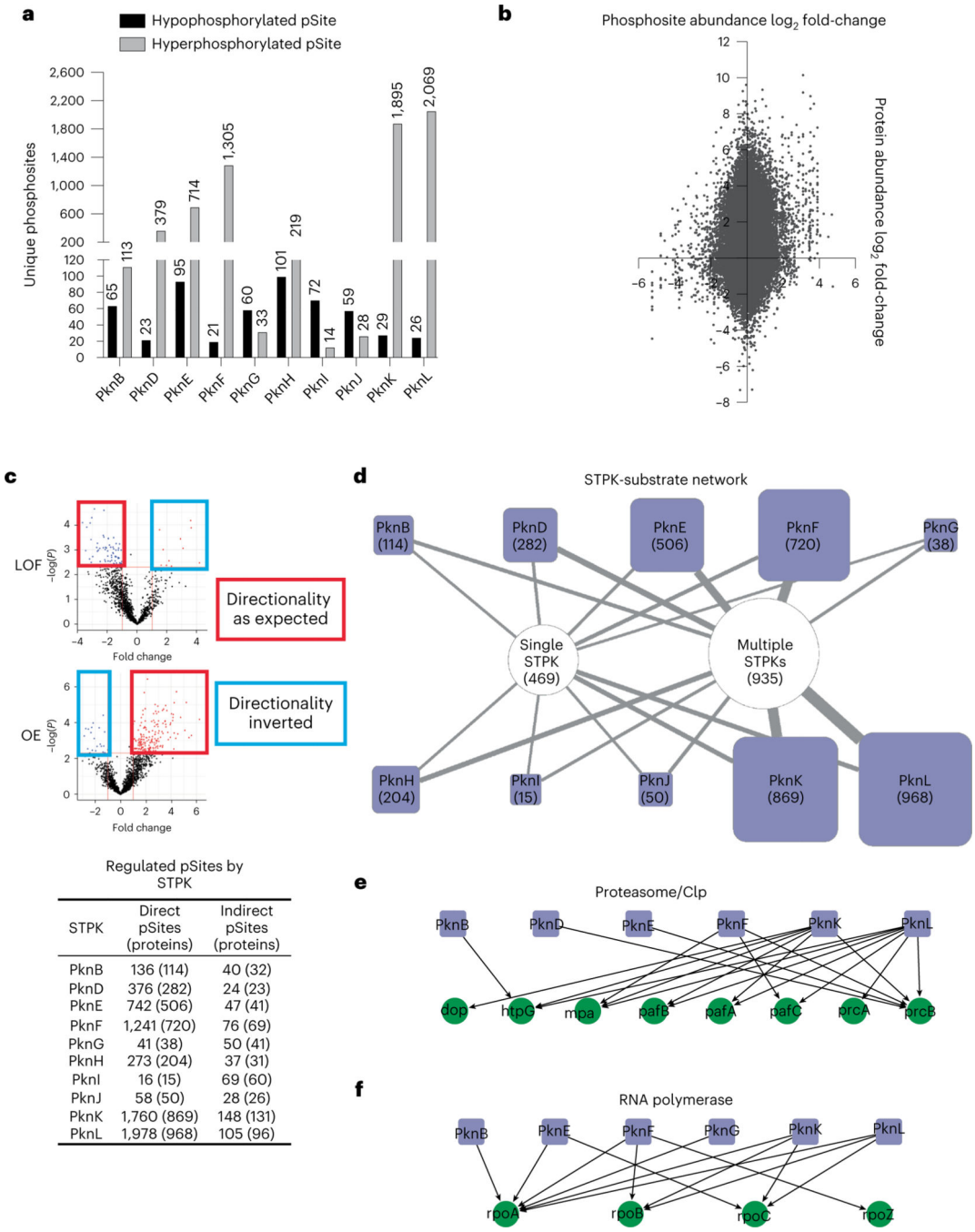
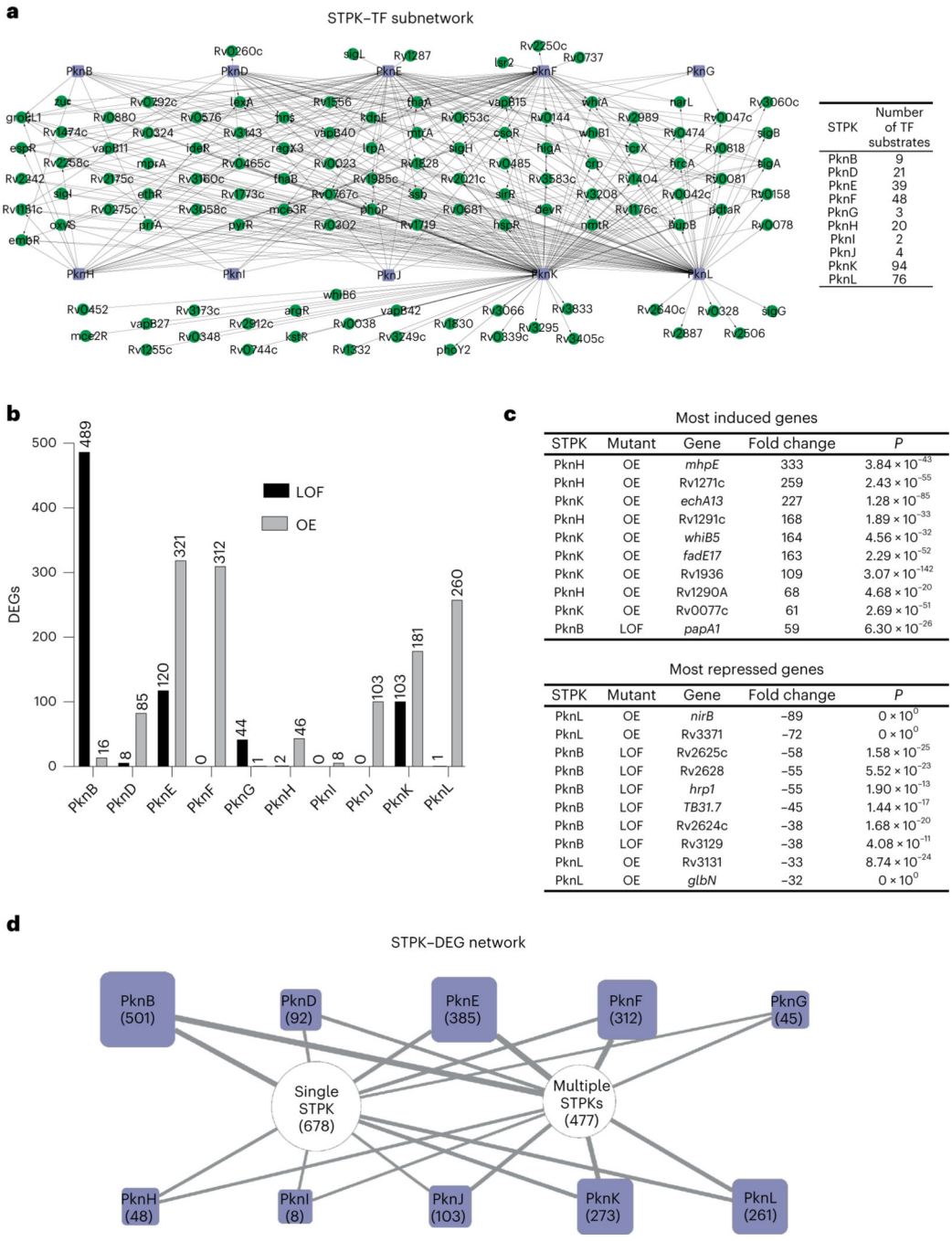


Fig. 3 |. *Mtb* protein *O*-phosphorylation establishes a network.

a, Effects of STPK perturbation on the abundance of unique phosphosites by STPK compared to WT. Changes in LOF and OE strains were combined for each STPK. **b**, Correlation between global protein abundance and phosphosite abundance. Quantitative TMT data of total protein abundance were plotted against the phosphosite abundance for each protein, showing that most increases in phosphorylation are not concomitant with higher protein abundance. **c**, Directionality of phosphorylation changes in response to STPK perturbation. The volcano plot example shows assignment of direct (hypophosphorylation

in LOF, hyperphosphorylation in OE) and indirect phosphorylation events with inverted directionality. The table shows putative direct and indirect phosphosites (phosphoproteins) in our dataset by STPK. Significance was determined by one-way ANOVA. **d**, Overview of the STPK substrates and overlap in substrate phosphorylation. The purple squares represent the STPKs. The size of the square is proportional to the number of substrates for that STPK, which is also given in parentheses. The round nodes show the number of substrates phosphorylated by one or multiple STPKs. The thickness of the edge corresponds to the number of substrates phosphorylated by the respective STPK. **e**, Phosphorylation of members of the proteasome/Clp protein degradation systems by different STPKs. **f**, Phosphorylation of RNA polymerase subunits by multiple STPKs.



STPK	Number of TF substrates
PknB	9
PknD	21
PknE	39
PknF	48
PknG	3
PknH	20
PknI	2
PknJ	4
PknK	94
PknL	76

Most induced genes

STPK	Mutant	Gene	Fold change	*P*
PknH	OE	*mhpE*	333	3.84×10^{-43}
PknH	OE	Rv1271c	259	2.43×10^{-55}
PknK	OE	*echA13*	227	1.28×10^{-85}
PknH	OE	Rv1291c	168	1.89×10^{-33}
PknK	OE	*whiB5*	164	4.56×10^{-32}
PknK	OE	*fadE17*	163	2.29×10^{-52}
PknK	OE	Rv1936	109	3.07×10^{-142}
PknH	OE	Rv1290A	68	4.68×10^{-20}
PknK	OE	Rv0077c	61	2.69×10^{-51}
PknB	LOF	*papA1*	59	6.30×10^{-26}

Most repressed genes

STPK	Mutant	Gene	Fold change	*P*
PknL	OE	*nirB*	−89	0×10^{0}
PknL	OE	Rv3371	−72	0×10^{0}
PknB	LOF	Rv2625c	−58	1.58×10^{-25}
PknB	LOF	Rv2628	−55	5.52×10^{-23}
PknB	LOF	*hrp1*	−55	1.90×10^{-13}
PknB	LOF	*TB31.7*	−45	1.44×10^{-17}
PknB	LOF	Rv2624c	−38	1.68×10^{-20}
PknB	LOF	Rv3129	−38	4.08×10^{-11}
PknL	OE	Rv3131	−33	8.74×10^{-24}
PknL	OE	*glbN*	−32	0×10^{0}

Fig. 4 |. *O*-phosphorylation has large transcriptional effects.

a, Overview of TF phosphorylation by STPKs. STPKs are shown as purple squares and TFs are shown as green circles. The edges represent kinase–substrate interactions. No Ascore cut-off was applied because the analysis used phosphoproteins, not phosphosites. The table shows the number of TF substrates for each STPK. **b**, Overall number of DEGs on STPK perturbation by STPK and by type of perturbation. DEGs are defined as greater than a four-fold change with $P < 0.01$. Significance was determined by calculating the geometric mean of five differentially expressed tools. **c**, The most induced and repressed genes by

STPK and type of perturbation. Significance was determined by calculating the geometric mean of five differentially expressed tools. **d**, Overview of the effect of STPKs on gene expression. The purple squares represent the STPKs. The size of the square is proportional to the number of that STPK's regulon, which is also given in parentheses. The round nodes show the number of genes regulated by one or multiple STPKs. The thickness of the edge corresponds to the number of substrates phosphorylated by the respective STPK.

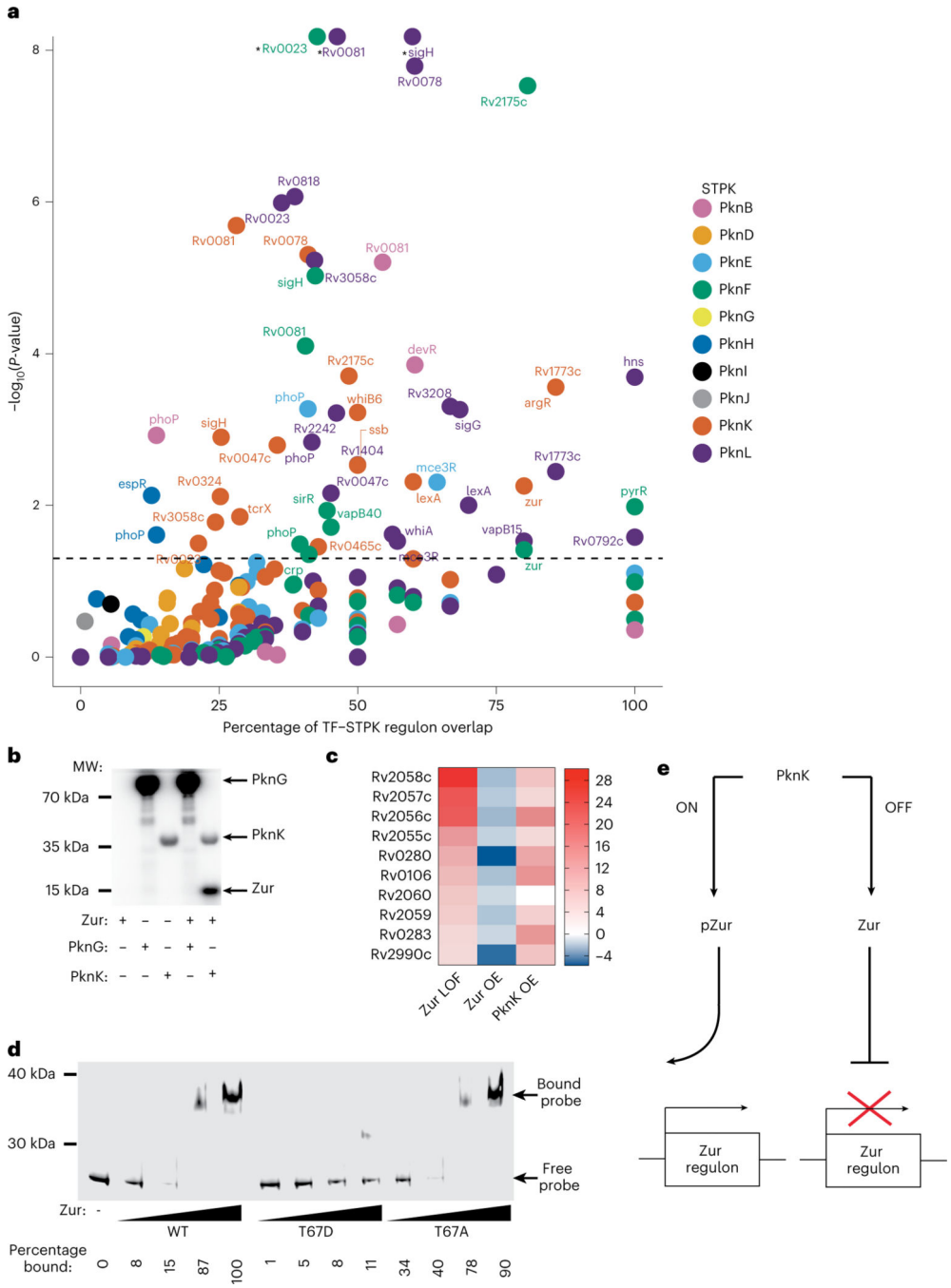


Fig. 5 |. Prediction of functional TF phosphorylation events.

a, Genes regulated by both TFs and their cognate STPK were plotted as the percentage regulon overlap against the significance of the overlap. The colours denote different STPKs. The dotted line indicates $P = 0.05$. These data are based on protein-level phosphorylation and no localization score cut-off was applied. P values were calculated using a hypergeometric test for enrichment. TFs with a single asterisk have $P < 1 \times 10^{-13}$. **b**, In vitro phosphorylation assay showing the specific phosphorylation of recombinant Zur by PknK but not PknG. The data shown are representative of three independent experiments.

c, Transcriptional response of Zur LOF, Zur OE and PknK OE mutations on the Zur operon shows overlapping but opposite effects of PknK and Zur OE, predicting derepression of Zur activity by PknK phosphorylation. **d**, Effect of Zur Thr67 phosphoablative and phosphomimetic mutations on DNA binding activity. The data shown are representative of three independent experiments. **e**, Model of the PknK–Zur pathway.

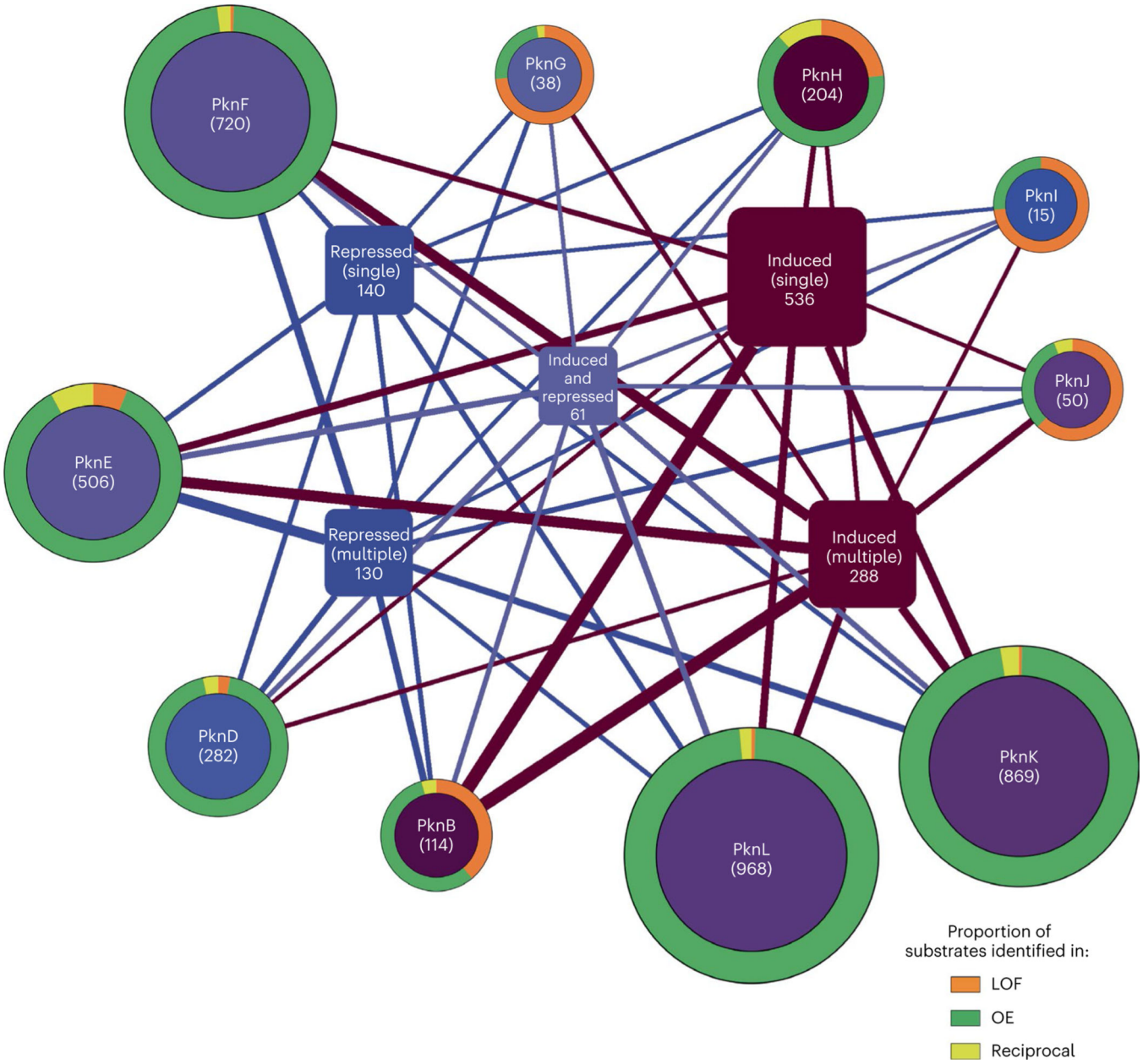


Fig. 6 |. Overview of the *Mtb* STPK *O*-phosphoproteome and transcriptional network.

The phosphoproteomic and transcriptional effects of STPK mutation are shown, illustrating the strongly overlapping nature of the network. STPKs are shown as round nodes, with their total number of phosphorylated substrates in parentheses. The outer circle of each STPK node shows the relative share of substrates observed in LOF (orange), OE (green) or both (down in LOF, up in OE, yellow). The colour of the inner STPK circle indicates whether the effect of the STPKs on DEGs on average is induction (red), repression (blue) or a combination (purple). The inner square nodes represent genes regulated by the STPKs, with the thickness of edges connecting the STPKs to genes corresponding to the number of genes regulated by an STPK. Single and multiple refer to the number of STPKs that have a particular effect on the group of genes (for example, 130 genes are repressed by more than 1

STPK). Interestingly, the expression of 61 genes is activated by some STPKs and repressed by others.