DISS. ETH NO. 26776

Unraveling host-pathogen interactions upon *Mycobacterium tuberculosis* infections using quantitative mass spectrometry

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES of ETH ZURICH

(Dr. sc. ETH Zurich)

Presented by

Charlotte Marie Nicod

M.Sc., ETH Zurich

Born on February 6, 1992

Citizen of Switzerland

Place origin: Lausanne, Vaud

Accepted on the recommendation of

Prof. Dr. Ruedi Aebersold

Dr. Ben Collins

Prof. Dr. Sebastien Gagneux

Prof. Dr. Wolf-Dietrich Hardt

2020

This thesis is dedicated to

Pips

my beloved sister

who one day,
will become a "real" doctor

Summary

Tuberculosis (TB) has been a burden on humanity for more than centuries and still remains the leading killer caused by single-agent infections. TB is caused by a bacterial pathogen called *Mycobacterium tuberculosis (Mtb)*. Despite the advances in anti-microbial chemotherapies and the discovery of a vaccine a hundred years ago, albeit with moderate effectiveness in preventing TB, 10 million TB cases arise annually and 2 million patients succumb to the infection. Furthermore, it is estimated that up to 2 billion people are asymptomatically and latently infected with the bacilli, forming an inexhaustible reservoir for TB.

Being an airborne pathogen, *Mycobacterium tuberculosis* infects its host via the respiratory tract to end up internalized into alveolar macrophages. Through thousands of years of co-evolution with humans, *Mtb* has developed many intricate strategies to not only persist in its host macrophages, but to use its host resources, prevent its clearance and even replicate intracellularly. One of the keys to its successful virulence is the encoding of protein secretion apparatuses in its genome. By secreting proteins that will directly and physically interact with its host environment, *Mtb* can directly intervene and modulate host responses.

In the first part of this thesis, we aimed at quantifying the effect of the virulence-associated protein secretion system on global host proteomic responses. To achieve this, we infected in parallel primary host cells with wild type *Mycobacterium tuberculosis* in comparison with a strain lacking this virulence-associated protein secretion system called ESX-1. By using state-of-the-art mass spectrometry-based approaches, we monitor at an unprecedented depth nearly 6,000 host proteins along a dynamic infection time-course. This enables us to describe the immuno-modulatory effect caused by the protein interactions mediated by the secretion system ESX-1. We could then attribute some of the known general virulence patterns of *Mtb* to its functional protein secretion system. These include modulation of Rab mediated intracellular trafficking, suppression of cell adhesion molecules or dysregulation of complement and coagulation cascades.

After studying the global host response to the secreted bacterial proteins, also referred as indirect host-pathogen interactions, we set the challenge of mapping the direct physical interactions. To map the host-pathogen protein-protein interaction landscape, we employ the robust and sensitive method called Affinity-Purifications coupled to Mass Spectrometry (AP-MS). The method is based on the generation of transgenic host cell lines expressing individual baits. We then co-purify the bacterial baits with their host interactors to identify high-confidence host-pathogen protein-protein interactions. After mapping our results than 202 novel *Mtb*-human

interactions with known *Mtb*-human interactions, we perform a global analysis on the directly targeted host factors. Our results suggest that *Mtb* alters the Rab functions not only by affecting their expression levels, but also by physically interacting with them. In a further step, by overlapping our results with known host factors affecting the intracellular survival, we identify potential intervention targets suitable for host-directed therapies.

In the third part of the thesis, we present a method that could unravel global host-pathogen protein-protein interactions in the context of real infections. This approach is based on the co-fractionation of cell lysates followed by a mass spectrometry-enabled identification of the proteins in each fraction. Based on their co-elution profiles, we can extract evidence for interacting proteins and larger protein complexes. As a proof-of-concept, we apply this highly optimized workflow to study the global organization of the THP-1 cells' proteome when undergoing a differentiation into macrophages and finally when subjected to a bacterial-like stimulation using *E. coli* extracted lipopolysaccharides. Results suggest that the differentiation from monocytic precursor cells into macrophages induces a much higher proteome-wide reorganization than the LPS stimulation.

Finally, I summarize our findings and the lessons learned from studying *Mtb*-human interactions and how we could go further and leverage this information with the aim of developing novel treatments strategies to fight pandemic tuberculosis. I then conclude my thesis by providing an outlook on the field of host-pathogen protein-protein interactions.

Résumé

La tuberculose (TB) représente un défi pour la santé publique depuis des milliers d'années. Elle est plus grande cause de décès liée à un agent pathogène spécifique. La TB est causée par la bactérie *Mycobacterium tuberculosis* Mtb. Malgré les avancées des antibiotiques et la découverte il y a 100 ans d'un vaccin, dont l'efficacité est de loin pas suffisante, il y a 10 millions de cas avérés de TB annuellement dans le monde dont 2 millions décèdent suite à leur infection. De plus, 2 milliards d'individus non symptomatiques sont probablement infectés par les bacilles et représentent un formidable réservoir pour la TB.

Le Mtb est un pathogène qui se transmet par la voie des airs. Il pénètre dans les organismes sains par le système respiratoire pour finir dans les macrophages alvéolaires des poumons. Des milliers d'années d'évolution commune avec l'homme ont permis au Mtb de développer des stratégies qui renforcent son action et sa persistance. Le Mtb peut non seulement persister dans les macrophages qui lui servent d'hôte, mais il peut également utiliser leurs ressources pour éviter d'être éliminé. Il peut même se multiplier au sein des cellules hôtes. La capacité du Mtb d'encoder dans son génome un mécanisme de sécrétion de protéines, lui permet de sécréter des protéines qui vont interagir directement avec leur hôte et influencer leur comportement.

La première partie de la thèse quantifie la virulence associée au système de sécrétion de protéines dans les cellules hôtes. Nous avons infecté en parallèle des cellules hôtes avec des lignées de Mtb avec et sans le système de sécrétion de protéines appelé ESX-1. Des appareils de spectrométrie de masse de dernière génération nous ont permis d'examiner de manière détaillée l'évolution de 6'000 protéines au cours du processus d'infection. Ceci nous a permis de décrire l'effet immunitaire induit par l'interaction des protéines sécrétées par l'ESX-1. Nous avons de ce fait pu attribuer des modèles de virulence du Mtb bien référencés à son système de sécrétion de protéines. Ceci inclus la modulation de trafic intracellulaire assurée par le Rab, la suppression des molécules permettant l'adhésion des cellules ou encore la dérégulation des cascades de de coagulation.

Après avoir étudié la réponse de l'hôte aux protéines sécrétées par les bactéries, souvent référencée comme l'interactions hôte-pathogène, nous nous sommes fixés le défi de décrire les interactions physiques directes. Nous avons utilisé la méthode de "affinity purification" en parallèle à de la spectrométrie de masse (AP-MS) pour décrire les interactions de protéine à protéine entre l'hôte et son pathogène. Nous utilisons des cellules hôtes transgéniques qui

expriment la protéine bactérienne fusée avec un « affinity tag ». Ensuite nous purifions les protéines bactériennes et leurs appâts humain pour obtenir avec un haut degré de confiance des liens entre les protéines des hôtes et des pathogènes. Nous avons ainsi obtenu un mapping de 202 nouvelles interactions Mtb - hôtes et les avons comparées aux interactions déjà référencées. Ceci nous a permis de faire une analyse des facteurs de cible de l'hôte. Nos résultats indiquent que le Mtb altère les fonctions Rab, non seulement en influençant leur expression mais également en interagissant physiquement avec elles. Nous avons ensuite identifié des cibles pour des traitements des cellules hôtes en superposant nos résultats avec des facteurs connus influençant la survie des hôtes.

Dans la troisième partie de la thèse, nous examinons des manières de démêler les interactions entre les protéines pathogènes et celles de leurs hôtes dans le contexte de réelles infections. Le concept est basé sur le fractionnement de lysat cellulaire de cellules infectées suivi d'une identification par spectrométrie de masse et constituent la base de notre analyse. Pour tester la rigidité et première application de la méthode, nous étudions l'organisation du protéome des cellules THP-1 lorsqu'elles sont soumises à une différenciation pour devenir des macrophages. Nous les assujettissons également à une stimulation utilisant des lipopolysacharides (LPS) extraits des *E. coli*. Les résultats suggèrent que la réorganisation du protéome est bien plus forte si elle est induite par la différenciation de cellules monocytiques dans les macrophages que si elle est induite par une stimulation par les LPS.

Finalement, je propose des pistes de traitements pour combattre la pandémie de TB en utilisant notre analyse détaillée des interactions entre le Mtb et les cellules hôtes dans l'homme. Je conclus ma thèse avec quelques hypothèses sur l'évolution de la recherche dans les interactions entre les protéines pathogènes et celles de leurs hôtes sains.

Chapter 1

General Introduction

Parts of this chapter were published in two separate reviews published by Charlotte Nicod and Amir Banaei-Esfahani. The rest of the chapter was exclusively written by Charlotte Nicod.

Nicod C, Banaei-Esfahani A, Collins BC
Elucidation of host–pathogen protein–protein interactions to uncover mechanisms of host cell rewiring.
Curr. Opin. Microbiol. 2017, **39**:7–15.

Banaei-Esfahani A, Nicod C, Aebersold R, Collins BC
Systems proteomics approaches to study bacterial pathogens: application to Mycobacterium tuberculosis.
Curr. Opin. Microbiol. 2017, **39**:64–72.

1.1 General introduction to *Mycobacterium tuberculosis*

Mycobacterium tuberculosis (*Mtb*) is the causative agent of the pulmonary tuberculosis (TB) disease and was first discovered in 1882 by Robert Koch. It is hypothesized to be the deadliest infectious disease in the history of humanity with an estimated number of over one billion human causalities in the past 200 years [1]. Despite the discovery and development of the first and only vaccine strain of *Mycobacterium bovis* named Bacille Calmette-Guérin (BCG) after the two bacteriologist in 1921 and the advances in antibiotics, it remains even nowadays a leading killer caused by single-agent infections [2].

Once infected by the bacilli, patients carrying *Mtb* may be classified into two main categories: the clinically asymptomatic and non-transmissible latent infections which represents about 90% of the infected population, and the active and transmissible pulmonary tuberculosis infections. It has been reported that up to a quarter of the world population is latently infected and amongst which, 10 million people annually fall sick thus developing the active tuberculosis disease [2]. Desolately, most of *Mtb*-infected patients live in poor high burden countries with 60% of them living in Asia (predominantly in India) and in 24% in African countries [2]. This tuberculosis demography renders the treatment and prevention of this global epidemic even more challenging, and corroborates the fact that tuberculosis is fundamentally a disease of the poverty due to lack of effective health care, malnutrition and unsanitary living conditions [3]. Moreover, the treatment to achieve sterilization in patients is exceedingly costly [4] and lengthy, especially considering the demographics of the affected and the rise in antibiotic resistant strains. The recommended treatment for drug-susceptible tuberculosis consists of an intensive 2 month-long phase of four antibiotics (namely isoniazid INH, rifampin RIF, pyrazinamide PZA and ethambutol EMB) followed by a 4-months continuation phase with INH and RIF [5]. Although generally treatable with this cocktail of antibiotics, the number of patients carrying Multi-Drug and Extended-Drug Resistant (MDR and XDR respectively) strains is alarmingly increasing and reached 500’000 new cases in 2018, causing a major threat to global health [2,3,4]. In some of the high burden countries like Kazakhstan and Kyrgyzstan, the rate of multi-drug resistance has even surpassed 25% of all patients [8].

Mycobacterium tuberculosis is an airborne pathogen and infects its host by entering the lungs via the respiratory tract to end up internalized into alveolar macrophages via receptor-mediated phagocytosis. The macrophages then translocate into the subtending epithelial layer which as a consequence induces a localized innate immune response. Further mononuclear cells including lymphocytes are then recruited to the local inflammation, providing new host niches for the

bacilli and start organizing themselves into stratified structures called granulomas [9], a hallmark of the tuberculosis disease. Once internalized and engulfed into early-stage phagosomes, Mtb can employ numerous strategies to prevent its clearance by interfering with the macrophages' cellular modules. One of its main strategies consists of blocking the fusion of its bacilli-containing phagosomes with the lysosomes [10,11] and inducing the rupture of the phagosomal membrane to facilitate its escape in the hosts' cytosol and delivery of Mtb proteins in the host. Additionally, Mtb manages to modulate the apoptotic and autophagy pathways [12], to intervene in the MHC class II presentation [13], to impair dendritic cell maturation [14] and to modulate T cell mediated immunity [15]. Most of these hallmarks are however only valid for virulent strains, including the lab strain H37Rv. Indeed, attenuated lab strains such as H37Ra or the avirulent vaccine strain, *Mycobacterium bovis* BCG, undergo full clearance in infection models. The disparity in the phenotypes between virulent and avirulent strains is primarily due to the deletion of a genomic locus called the Region of Difference 1 (RD1) [16–19].

1.2 Introduction to the Region of Difference 1 and the Type VII Secretion Systems

The genomic locus RD1 was first discovered through a genomic comparative analysis between *M. tuberculosis* and attenuated *M. bovis* strain due to its absence in the BCG genome [16]. It encodes one out of five homologous specialized type VII secretion systems, namely ESX-1. Type VII secretion systems, titled ESX-1 to 5, are hypothesized to be responsible for the transport of specific protein substrates across the mycobacterial envelope. However only ESX-1, ESX-3 and ESX-5 systems have been demonstrated to secrete protein substrates across the mycolic wall [20,21].

Each ESX secretion system is composed of five conserved core membrane components, namely EccB, EccC, EccD, EccE and MycP. Although predicted to secrete a variety of protein substrates, two themes are found amongst all ESX systems. First, the *esx* loci encode small proteins of roughly 100 amino acids that have a Trp-X-Gly (WXG) motif, resulting in a helix-turn-helix tertiary structure [22] which often require homo- or hetero-dimerization for their secretion. Secondly, they also encode a transmembrane ATPase protein [23] as member of the secretion apparatus. Besides secreting WXG proteins, mycobacteria also frequently secrete Pro-Glu (PE) and Pro-Pro-Glu (PPE) proteins as monomers or dimers.

These ESX 1-5 systems are crucial for mycobacterial pathogenesis, as at least 3 of these ESX systems including ESX-1 and ESX-5 are necessary for full virulence [20,24]. Specifically, the deletion of the ESX-1 and ESX-5 secretion system leads to attenuated pathogenicity and an overall

decrease in intracellular bacterial replication in *in vitro* cellular models and in animal studies [17,25–27].

1.2.1 The ESX-1 Type VII Secretion System

The RD1 locus has a length of 9.5 kb and encodes 9 genes, including components of the ESX-1 secretion system and two of the best characterized co-secreted WXG-type virulence factors ESAT-6 (6 kDa Early Secreted Antigen Target, also called EsxA) and CFP-10 (10-kDa Culture Filtrate Protein, also called EsxB) [28]. It was until recently suggested that EsxA alone was responsible for the phagosomal permeability and subsequent bacterial translocation into the cytosol and mediated cytosolic interactions [17]. However, this theory was recently challenged and the claimed membrane-lysing properties of EsxA were in fact a direct consequence of the presence of detergents from the experimental preparations [29]. Nonetheless, EsxA-EsxB secretion within the context of an intact ESX-1 system does seem to still confer phagosomal lytic properties in a contact-dependent context [29] and is now undisputed. It was further demonstrated that the RD1 locus was involved in the necrosis induction of host macrophages [30], the granuloma formation [31] and the activation of caspase-1 and subsequent secretion of IL-18 and IL-1ß in infected macrophages [32]. In mice, the presence of RD1 in BCG is enough to cause bacterial replication in the lung and spleen and induce the formation of granulomas [19]. In reverse, the deletion of the Region of Difference 1 from *Mycobacterium tuberculosis* prevents the spread of the bacterial replication in mice and leads to survival [17,18].
Although it is now irrefutable that ESX-1 along with its secreted EsxA and EsxB proteins mediate some of the Mtb pathogenicity by permeabilizing the phagosomal membrane and subsequently leading towards direct host cytosol-pathogen interactions [17,33], their exact molecular mechanisms or host cellular targets remain unclear.

1.2.2 The ESX-3 Type VII Secretion System

The ESX-3 Type VII Secretion System has been demonstrated to play two distinct roles. First, it has been described to be involved in metal homeostasis and iron acquisition via a mechanism dependent on the secretion of its substrates PE5-PPE4 [34–36]. Secondly, it mediates its virulence by secreting its heterodimer EsxG-EsxH substrates which consequently impairs the host phagosomal maturation by interacting with the host Endosomal Sorting Complex Required for Transport ESCRT [37]. Additionally, the EsxG-EsxH dimer has been reported to elicit CD4 and CD8 T-cell responses in both humans and mice [38,39]

1.2.3 The ESX-5 Type VII Secretion System

The ESX-5 Secretion System is the most recently evolved of the Type VII Secretion Systems [40] and is activated at the transcriptional level by phosphate-depleted environments [41]. ESX-5 is predicted to secrete EsxN [42] and most of PE/PPE proteins, a class of proteins representing nearly 10% of the *M. tuberculosis* genome amounting to 167 putative secreted substrates [43]. ESX-5 has different roles in the biology of tuberculosis. Through the secretion of PPE10, it plays an important role in the maintenance of the mycobacterial capsule integrity [44]. Upon infections, the *Mycobacterium marinum* ESX-5 does not interfere with the phagosomal maturation within human host macrophages, but strongly affects the host immune response. It downregulates the production of proinflammatory cytokines including TNF-alpha and IL-6 [45]. It further represses TLR signaling dependent innate immune cytokine secretion whilst increasing the production of IL-1ß [45].

Considering the aforementioned phenotypes upon the individual ESX secretion systems deletions, it is undeniable that they play crucial roles in the mycobacterium virulence by somehow interacting with host components. We thus hypothesize that identifying the Host-Pathogen Protein-Protein Interactions (HP-PPI) could potentially clarify their underlying molecular mechanisms of the individual virulence factors, altogether elucidating the overall pathogenicity mechanisms.

1.2 Introduction to direct Host-Pathogen Protein-Protein Interactions

Infectious diseases reflect the evolutionary balance between a host and its pathogen. In order to ensure their survival and propagation, pathogens have developed numerous intricate tools to subvert their hosts' defense mechanisms. Understanding how pathogens actively rewire host cell defenses is of particular interest in infectious disease research. Ultimately by identifying host-directed targets for pharmacological intervention, this field of research may contribute to eradicate the public health burden caused by infectious agents.

The molecular mechanisms underlying pathogenic rewiring of host cells are widely varied. However, as protein complexes and their interaction networks into which they are organized comprise the primary functional modules of the cell [46], we can predict that the disruption of these host networks are likely to be a key strategy for manipulation by pathogens. Re-wiring of the host's proteome, also known as pathogenic hijacking, generally includes intervention at multiple stages of signaling pathways and cellular functions to ensure the robustness of the virulent intervention [47]. This hijacking by protein-protein interactions may be carried out by

evolutionarily derived partial molecular mimicry [48], which consists of virulent proteins having evolved similar structures or motifs to the host proteins to mediate such HP-PPI. It has further been proposed that the phenotypic impact of a pathogen is directly proportional to its ability to rewire the host interactome, and that the impacts of individual virulent proteins are linked to their number of interactions with host proteins [49]. Thus, mapping the host-pathogen protein interactome may provide valuable insights into the biological functions of virulence factor proteins, highlight interactions critical to the pathogens' progression and spread, and improve our overall understanding on the molecular basis of pathogenicity.

The field of HP-PPI has been highly successful in deciphering the virulence mechanisms of various pathogens, however mostly for viruses. This can be explained due to their minimal genomes and by being obligate parasites, viruses must rely on HP-PPI as a mean to carry out the pleiotropic functions of their proteins by hijacking various host protein modules to either avoid their clearance or enable their spread. For example, by mapping the Influenza A – human PPI network, viral proteins were reported to be highly inter-connected thus forming functional modules, and to interact with a greater number of host proteins compared to the average degree of connectivity within the human interactome [50]. The HP-PPI map further enabled the identification of multiple molecular mechanisms employed by the virus to manipulate its host, including how Influenza proteins intervene in the WNT/ß-catenin pathways as a mean to modulate the host's interferon production [50].

Unfortunately, the characterization of bacterial HP-PPI has lagged behind. The reason for this disparity most likely reflects differences in feasibility. That is, testing all proteins produced by a viral genome for interactions with a host proteome requires significantly less effort than that for bacterial genomes due to their increased genomic complexity. Nonetheless there is increasing amount of evidence that bacteria also rewire host cellular pathways via HP-PPI [47]. Pathogenic bacteria can interact with their host's proteome by three main mechanisms. First, bacterial membrane proteins are an obvious interaction point, as they are located at the physical interface between both organisms. Secondly, bacteria might secrete effector proteins (also known as virulence factors) into the host cell where they can interact with the host proteome. Secreted effector proteins are of particular interest as they are frequently required for full virulence [51]. Additionally, some bacterial pathogens such as certain *Shigella dysenteriae* or *Escherichia Coli* strains express Shiga toxins generally during their lytic cycle [52] or release these toxins through Outer Membrane Vesicles during their growth phase [53], leading to the inhibition of protein synthesis or activation of the apoptotic pathways of their host cells. As the number of bacterial host-pathogen interaction studies increases, they demonstrate that while bacteria generally do

not rely on host cell machinery for the purpose of replication as directly as viruses do, they do seem to disrupt the immune response [54] and interact preferentially with the hosts' cytoskeleton as a mean of motility, invasion of the host tissues [55] and escape of phagocytic cells [56]. For instance, *Mycobacterium tuberculosis*, is known to modulate the host's immune response and prevent its bacterial clearance by suppression of autophagy. Recent work has shown that a secreted Mtb factor, PE_PGRS47, locates in the host's cytosol and inhibits the Major Histocompatibility Complex II mediated antigen presentation, thereby partially suppressing the autophagy of the Mtb containing macrophages in chronic stages of infections [57]. By mapping such host interactors, HP-PPI studies could hint us towards the molecular mechanisms behind certain virulence factors like this PE_PGRS47.

Although the concept of cellular hijacking of host modules by viral proteins has been widely studied and approved, host hijacking by bacterial pathogens has less supporting evidence. However, in the last few years, increasing amounts of evidence has led to the conclusion that bacterial pathogens are also required to hijack cellular modules [37,54,58–65], such as the ubiquitin system [66–68] or the host cytoskeleton [55], mediated by PPIs to prevent their clearance or promote their replication and spread. Thus, multiple studies have focused their efforts on characterizing HP-PPIs by employing common PPI detection methods such as Yeast Two Hybrid (Y2H) [37,58–62] or Affinity-Purification coupled to Mass Spectrometry (AP-MS) [63–65,69].

Recently, in the context of *Mycobacterium tuberculosis* infections, Mehra A *et al.* used a high stringency Y2H study to screen for interactions between bacterial secreted virulence factors and 12'000 human ORFs. It enabled the discovery of a molecular mechanism by which a Type VII secreted effector protein, named EsxH, targets the endosomal sorting complex required for transport ESCRT machinery, thereby impairing the phagosomal maturation and fusion with the lysosomes [37]. With further validations, they subsequently showed that the EsxH-ESCRT interaction also impaired the major histocompatibility complex class II antigen presentation thereby inhibiting CD4+ T-cell activation [70]. Even though this study was a successful proof of concept to detect and subsequently test Host-Pathogen Protein-Protein Interactions in the context of *Mycobacterium tuberculosis*, Y2H-based methods have some considerable pitfalls. For instance, the necessity for exhaustive screens hampers its feasibility and the high false negatives or false positive rates are usually caused by technical challenges such as the non-physiological expression system [71]. Affinity-Purification coupled to Mass-Spectrometry, on the other hand, has been proven to overcome most of these challenges and has become a state-of-the-art method

to detect entire protein complexes in their physiological environment while keeping comparatively low rates of false positives or negatives.

1.3 Introduction to Protein-Protein Interactions Methods

In the following sections, we aim to summarize the methods available to characterize HP-PPI, consider their utility by providing biological insights, and present some outlook into the how the field may develop going forward. Even though we are primarily concerned with the possibilities of characterizing HP-PPI from the perspective of bacterial pathogens in particular *Mycobacterium tuberculosis*, a survey of the literature indicates that significantly more work has been done for viruses in this area [72]. As such, an examination of lessons learned from studies of interactions between viruses and hosts should also be instructive.

1.3.1 Yeast Two Hybrid

Historically first, the Yeast Two Hybrid method has been extensively used to detect direct physical interactions between two ectopically expressed tagged proteins in yeasts [73]. Although this method generates direct binary interaction datasets at high throughput, the need for exhaustive screens hampers its feasibility, and its technical challenges such as the non-physiological expression system provokes high rates of false negatives [71]. Nonetheless, many studies in the field of infectious diseases have successfully employed Y2H screens to investigate (near) genome-wide virus-host interactions [74–81], to compare homologous viral proteins from various strains [82,83], or to systematically map bacterial effector proteins - host interactions [37,58–62]. In the context of *Mycobacterium tuberculosis* infections, a Y2H screen along with functional validations, enabled the discovery of a molecular mechanism by which an effector protein, named EsxH, targets the Endosomal Sorting Complexes Required for Transport necessary for endosomal membrane trafficking, thereby impairing the phagosomal maturation and fusion with the lysosomes [37].

1.3.2 Affinity Purification coupled to Mass Spectrometry

While Yeast2Hybrid studies have proven themselves very useful to detect direct binary Protein-Protein Interactions, Affinity Purification coupled to Mass Spectrometry has emerged has a powerful, high throughput and sensitive alternative method, as it detects in an unbiased manner direct and indirect PPIs in near physiological conditions while maintaining low rates of false positives [84]. Most commonly, it consists of fusing an affinity epitope tag to a bait protein,

followed by a single or double biochemical affinity- or immuno- purification (AP or IP) steps in native lysis conditions. The purified bait, along with the non-covalently bound interacting proteins or macromolecular protein complexes (preys), are then identified and quantified via standard bottom up proteomics. To filter out non-specific interactions, this strategy relies on quantitative comparisons with control purifications.

In the field of infectious diseases, AP-MS is commonly applied to systematically map the interactome of individual virulent proteins ectopically expressed in the host's environment [63,64,69,82,85–94], to monitor single virulent proteins [95] or upon infection [96]. A related strategy uses immobilized recombinant bacterial effectors on beads [97] combined with AP-MS from their incubation with human plasma.

Although expressing single virulent genes in host environments is informative, it is believed that during the course of infections, the host-pathogen interactomes undergo infection stage-dependent dynamic changes [98], influenced by the hosts' responses and by the other co-expressed virulent proteins. Thus, some groups have generated replication-competent, epitope tagged viruses which enabled the spatio-temporal monitoring of empirical and quantitative changes upon viral infectious of host cells, including for Alphavirus Sindbis [98] and most recently HIV infected human cells [99]. Based on similar principles, Mousnier A *et al.* and subsequently So EC *et al.* respectively developed and applied a double purification-based method coupled to mass spectrometry to enable the identification of HP-PPI of bacterial effector proteins in host cells upon infections of *Legionella pneumophila.* This study, amongst others findings, described how three effector proteins may target up to 25 Rab GTPases individually during the course of infections [65,100].

For the first time to our knowledge, it was also applied to systematically map the HP-PPIs of secreted bacterial effector proteins to gain insights into *Chlamydia trachomatis* pathogenesis [63], which led to predicted functions of certain secreted proteins and insights into how host cells may restrict *Chlamydial* infections. Most recently, this AP-MS approach used to identify HP-PPI was applied in the case of *Mycobacterium tuberculosis* [69] and the results will be discussed later in this thesis.

1.3.3 Proximity Dependent Labelling coupled to Mass Spectrometry

BioID has recently emerged as a new possibility to detect transient and weaker PPI [101] as it doesn't necessitate the preservation of the PPIs during the cell lysis and can thus be used as a complementary method to AP-MS [102]. This method relies on the fusion of a mutated promiscuous biotin ligase BirA* to the bait protein. During an incubation with high biotin concentrations, neighbouring proteins to the fused BirA*-bait protein undergo proximity

dependent biotinylation reactions. Biotin-conjugated proteins, potential direct or indirect interactors of the bait, can then be affinity purified using streptavidin-coated affinity matrices and quantified by mass spectrometry (see Figure 1). Because the identification of interactions does not depend on the native purification conditions, weak, transient and insoluble interactions such as for membrane proteins can be readily identified [101]. BioID has been applied as a mean to obtain comprehensive interactome information of selected bacterial proteins [103] belonging to the human pathogen, *Chlamydia psittaci*. A variation of this proximity labelling strategy, called APEX based on the enzymatic activity of Ascorbic Acid Peroxidase, enables much faster reaction times (~30 seconds), and opens up the possibility of time-resolved proximity measurements [104].

1.3.4 Chemical Crosslinking coupled to Mass Spectrometry

Chemical Crosslinking coupled to Mass Spectrometry (XL-MS) consists of chemically crosslinking proximal reactive side chains of exposed specific amino acids from native proteins in monomeric states or in protein complexes, followed by an MS based, bottom up approach to identify the crosslinked peptides and infer their proteins. XL-MS thus yields fixed distance restraints between bound residues, suggesting direct physical intra-protein or inter-protein interactions between crosslinked peptides belonging to the same or distinct proteins respectively [105] (see Figure 1). Chemical crosslinking reactions can be performed on purified protein samples [106] using GFP epitope tags [107], on cell lysates [108] or on living cells such as on the pathogen *Pseudomonas aeruginosa* [109]. Although having gained popularity in recent years to study the topology of protein networks, decipher the architecture of macromolecular complexes, and provide insights into domain-resolution protein interactions, XL-MS has not yet been widely applied to study HP interactions due to its challenging utilization. One exception is the unbiased study of live human epithelial H292 cells infected with *A. baumannii* which led to the identification of 46 HP-PPI [110]. However informative by identifying protein-protein interactions and providing information about their three-dimensional structures, XL-MS experiments remains challenging at the bioinformatic level. The principal bottleneck lies in the identification of cross-linked peptides due to the combinatorial search space, as any peptide from a sequence database can be crosslinked to any other peptide from the same database [111].

1.3.5 Co-Fractionation coupled to Mass Spectrometry

To overcome the biases and limitations caused by the need for genetic tagging or availabilities of antibodies using in approaches like AP/IP-MS or BioID, methods like Co-Fractionation Mass

Spectrometry (CoF-MS) are gaining in popularity [112–114]. CoF-MS is based on the mild lysis of cells and a near-native extraction of the proteome in order to preserve the protein-protein interactions. The lysates containing the interacting proteins and larger protein complexes are then fractionated according to the desired physicochemical properties using approaches such as Ion-Exchange Chromatography or Size Exclusion Chromatography. The proteins contained in each fraction are subsequently identified and quantified by mass spectrometry. Protein-Protein Interactions and the protein complexes are then inferred based on their correlating elution profiles (or co-elution) along the chromatographic dimensions [115]. Theoretically, CoF-MS methods enable the global parallel and unbiased identification of all protein complexes present in the cells.

We hypothesize that such global and unbiased approaches could solve some challenges imposed by the systematic study of bacterial-host interactions. The first challenge is to identify all secreted proteins upon infection, where *in silico* predictions and experimental findings don't always corroborate [116]. Secondly, due to their increased genomic complexity compared to viruses, the generation of transgenic cell lines to ectopically express each putative secreted protein would be highly time-consuming. Thirdly, bacterial systems generally lack adequate genetic tools preventing endogenous tagging of their secreted proteins. Thusly, we hypothesize that more global approaches for bacterial - host PPI detection may be useful. To our knowledge, a CoF-MS based method was only applied once in the context of infected host cells with bacterial pathogens [115]. Although no Host-Pathogen Protein-Protein Interaction could be detected, probably caused by the asymmetric ratios in bacterial effector proteins abundance in comparison with the abundance of the host proteome, the study provided some insights into host protein complex rearrangement cause the *Salmonella enterica* infection.

Despite having many benefits, these global methods are hampered by some drawbacks. These include the relatively low chromatographic resolutions given the large number of eluting proteins, the ability to reliably identify and quantify all proteins in each fraction, and the fact that weak or transient protein interactions usually fall apart during the chromatographic separations and only more stable macromolecular protein complexes can thusly be identified.

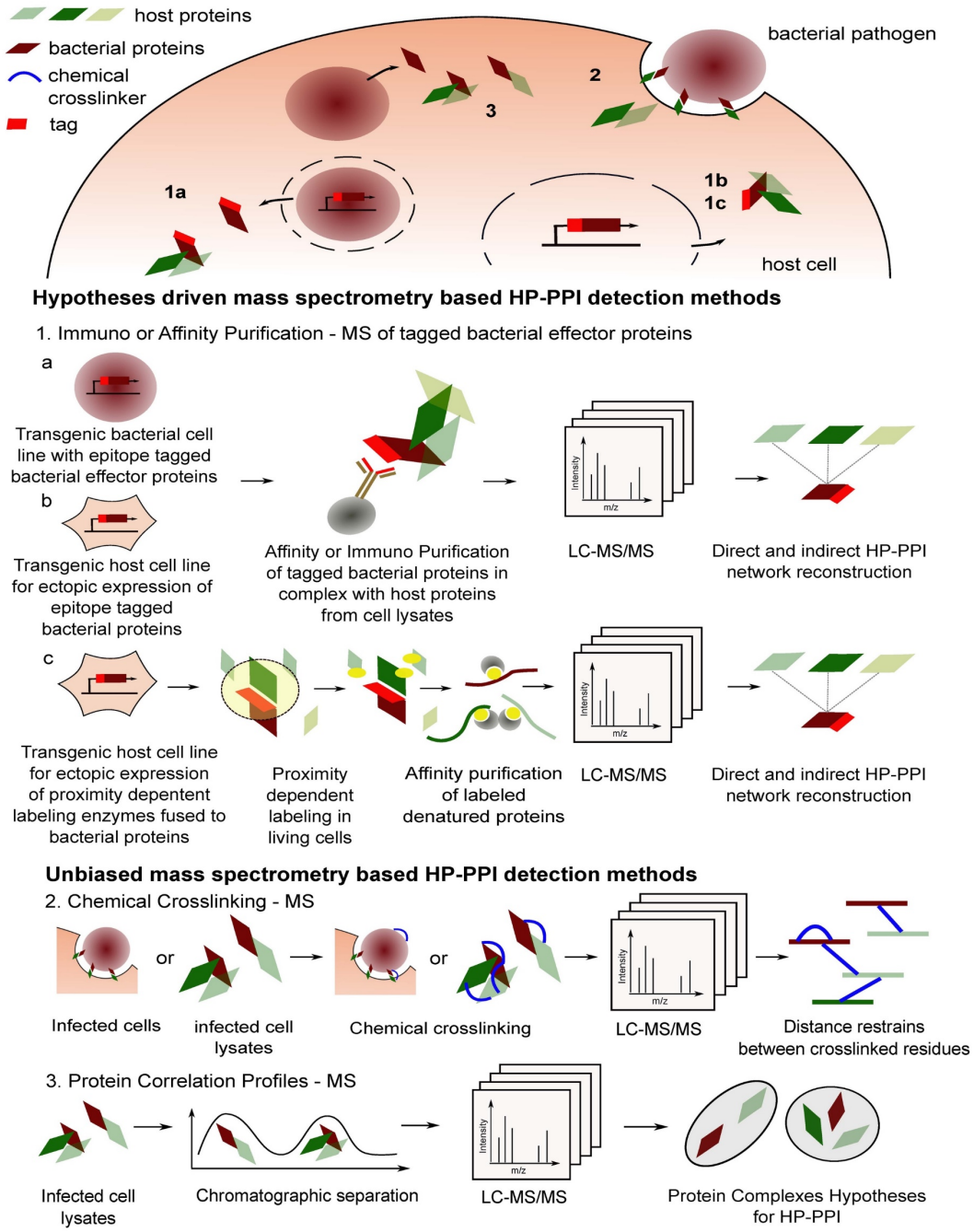


Figure 1 Mass-spectrometry based methods for Host-Pathogen Protein-Protein Interactions detections in the context of bacterial infections. AP/IP-MS from epitope tagged bacterial effector proteins (1a) post infections of their host cells enables the identification of physiological HP-PPI. Along with other hypotheses driven methods such as ectopic expressions of tagged bacterial proteins in the host environment (1b and 1c), they can lead to near comprehensive identifications of HP-PPI. However, because they rely on the prior tagging of the proteins of interest, they are limited by the number of proteins that can be cloned and expressed in the relevant cellular systems. Chemical Crosslinking – MS (2) and Protein Correlation Profiles – MS (3) methods, although less sensitive, do not require prior knowledge and tagging of bacterial proteins and thus allow *de novo* discovery of physiological and endogenous HP-PPI directly from infected cells (2 and 3).

1.4 Introduction to Global Host Response Profiling

Host-Pathogen Protein-Protein Interactions may provide valuable insights into the underlying molecular mechanisms in the pathogenicity carried over by virulence factors. However, they provide little information on the global phenotypic consequences of the interactions nor do they capture any indirect interactions they may have. To measure the cell-wide overall effect of secreted virulent proteins, it could be thus beneficiary to merge these HP-PPI interaction studies with global analysis of host response to differential infections with pathogens having the Wild Type strains or the deleted virulence factors strains.

Quantification of host responses to differential infections can be done at multiple levels, including at the transcriptomics [117], proteomics and specific signaling pathways analysis. Identifying and quantifying the host response to its pathogens can provide valuable information associated concerning both the infection and host defense mechanisms. Indeed, disease states including infectious diseases are generally characterized with subtle or large differences in the gene expression profiles and their protein products.

1.5 Introduction to Transcriptomics

The field of transcriptomics is the comprehensive identification and quantification of RNA transcripts within a biological sample. It has become a robust high throughput and affordable technology which enables the quantitative analysis of RNA molecules including protein-coding messenger RNAs from low amounts of cells, even down to the single cell level. It became an omnipresent tool, mostly used to quantify Differential Gene Expression (DEG) profiles between two or more conditions. Transcriptomics methods based on Next Generation Sequencing, most frequently Illumina, have the capacity to present multiple valuable benefits such as having single base-pair resolution, can sensitively detect alternative splicing variants, can cover multiple dynamic ranges and have a read depth of 10-30 million reads per sample. Yet, transcriptomics do not systematically predict their final product protein abundances [118,119] or final proteoforms. This unreliable correlation between transcript and protein level can be explained by existence of ubiquitous and essential biological process. First of all, protein-coding messenger RNA are subject to alternative splicing, increasing the diversity at the transcript level. Secondly proteins are subject to post-translational modifications introduced after the translation of the mRNA which can lead changes in the protein interactomes and functions [120]. We thus hypothesize that in order to get a comprehensive snapshot of the functional state of biological sample, it is necessary to identify and quantify the proteome content and its spatio-temporal regulation. In the case of

Mycobacterium tuberculosis infections, host transcriptional analysis were recently reviewed here [121].

1.6 Introduction to Proteomics

The field of proteomics is the large-scale analysis of proteins. Conceptually, Mass Spectrometry based protein analysis may be separated into two main approaches: top-down and bottom-up proteomics. Top-down proteomics enables the analysis of intact proteins and is an ideal method to study with high resolutions their proteoforms and their specific post-translational modifications. Unlike bottom-up proteomics, top-down do not require the proteolytic cleavage of the proteins, but can infer the protein sequence from an accurate experimentally determined intact mass [122,123]. It may also determine the exact amino-acid position of the PTM by employing tandem mass spectrometry analysis by ionizing [124] and subsequently fragmenting [125] the proteins using ElectroSpray Ionization (ESI) Mass Spectrometry ESI-MS [122]. Although highly sensitive for small diversity of proteoforms present in a single sample, top-down proteomics do not allow for the identification nor quantification of more complex mixtures of proteins. To do so, bottom-up proteomic approaches are needed.

Modern bottom-up proteomics based on mass spectrometry first consist of an enzymatic digestion, usually Trypsin, of the proteins extracted from the biological samples. The smaller tryptic peptides are then separated on a reverse-phase high-pressure Liquid Chromatography (LC) column coupled directly to a mass spectrometer. The peptides are ionized via an ESI prior to entering the mass spectrometer for the data acquisition, where the mass over charge (m/z) ratios are measured. Depending on the aim of the experiment and the nature of the samples such as its complexity, different acquisition strategies can be employed.

1.6.1 Data Dependent Acquisition Mass Spectrometry

Data Dependent Acquisition (DDA) MS is the most frequently used type of data acquisition scheme and is also known as shotgun or discovery proteomics. During a DDA run, the mass spectrometer performs a MS1 scan, which will measure the m/z values for the precursor ions (the tryptic peptides) present at a given point in the chromatographic dimension. The most abundant n peptides, typically around 10, are selected for subsequent fragmentation within a collision cell the and the m/z of the fragment products are then measured during a second scan

called MS2 (or MS/MS). The resulting ion spectra are then matched to an *in silico* digested protein sequence databases in order to identify the protein composition of the sample [126,127]. As for the quantification, the most commonly used strategy is to integrate the MS1 signals of the precursor ions along the chromatographic dimension. Although widely used, DDA strategies present some downsides such as low reproducibility and quantitative accuracy.

1.6.2 Data Independent Acquisition Mass Spectrometry

Compared to the classical semi-stochastic Data Dependent Acquisition (DDA) schemes, SWATH-MS [128] (short for Sequential Window Acquisition of All Theoretical Mass Spectra) is based on a Data Independent Acquisition (DIA) scheme that was developed in our group by Gillet *et al*. It allows to detect and measure in a more robust, sensitive, data-complete and quantitatively accurate manner a deeper proportion of the peptides across multiple samples in a linear dynamic range [129]. In order to circumvent the stochasticity and low reproducibility of DDA schemes, SWATH-MS relies on the fragmentation and MS2 acquisition of all precursor ions present in predefined MS1 windows. This conceptually enables the identification and quantification of all precursor ions present in the biological sample and results in a much deeper coverage. However, the hurdle of this approach relies in the data analysis due to the highly multiplexed and convoluted resulting MS2 spectra caused by the co-isolation and fragmentation of all precursor ions present in each MS1 extraction window. So to identify the present peptides, SWATH relies on the prior knowledge from previously generated fragment ion spectral libraries in combination with chromatographic coordinates to extract in a targeted fashion MS/MS signals [130] and enables MS2 based quantification for a higher accuracy.

Although the SWATH-MS acquisition scheme theoretically enables the fragmentation and MS2 acquisition of all precursor ions from each MS1 isolation window per cycle, the ion sampling efficiency by the quadrupole is highly limited considering the typical MS1 isolation window sizes [131]. Combined with the highly convoluted MS2 spectral complexity, this decreases drastically the effective number of precursor ions identified from the MS2 fragmentations. To overcome these limitations, recent technological advances introduced an additional layer for precursor ion separations based on their ion mobility by adding a Trapped Ion Mobility Spectrometer (TIMS) device. This not only increases the sensitivity by scanning all of the precursor ions by performing this parallel accumulation based on the ion mobility, but together with an extended targeted data extraction workflow that includes the ion mobility dimension that reduces the spectral

complexity, this novel data acquisition scheme allows us to reach a complete precursor ion sampling and identification from 10 fold fewer input material than classical SWATH-MS[131].

Together with Affinity Purification, SWATH-MS has been successfully applied to obtain temporally and quantitatively accurate representations of protein complexes in different biological conditions [132,133]. By replacing DDA with SWATH-MS, AP-SWATH studies have been able to sensitively quantify dynamic rearrangements of protein interactions, even for interactors so-far undetectable with DDA approaches as the dynamic range of abundance spanned over four orders of magnitude [132].

1.7 Introduction to *Mycobacterium tuberculosis* proteomics

A decade ago, shotgun-MS was the only viable mass spectrometry-based proteomic method and consequently most of the available proteomic data was in this form. In addition to providing important insights into the composition of the respective sample, the resulting fragment ion spectra provided the basis for specific measurement assays for targeting MS and DIA methods. This progression from proteome discovery to serial proteome quantification by targeting MS and DIA methods has been robustly implemented for *Mtb*. Specifically, *Mtb* is one of only a few species for which reference fragment ion spectra have been generated and made publicly accessible for proteins from every ORF of the genome [134–136]. In the following, we discuss the biological insights into the proteome of *Mtb* gained from different proteomic methods.

Comparison of proteomic datasets on H37Rv and H37Ra, a virulent and avirulent strain of *Mtb*, identified 29 significant changes of membrane associated proteins, including the possible protein export membrane protein SecF and three ABC-transporter proteins, that were upregulated in H37Rv. This suggested that the bacterial secretion and transporter systems might be significant determinants for the virulence of the bacilli [137]. A similar approach applied to the membrane associated proteins of H37Rv and BCG, the vaccine strain of Mtb, revealed the significance of membrane proteins in causing the disease. Analyzing the proteome of H37Rv, H37Ra, BND and JAL strains highlighted the distinct protein expression patterns of Esx and mce1 operon proteins in the JAL and BND strain, respectively, suggesting EsxA as a potential virulence factor. Proteins MmpL4, Rv1269c, Rv3137, and SseA have been reported as major differences between the ancient and modern Beijing strains which might clarify the increased virulence and success of the modern Beijing strains [138]. SseA, a predicted thiol-oxidoreductase, together with SodA and DoxX constitute a membrane-associated oxidoreductase complex (MRC) and lack of any MRC

subunit results in the defective recycling of mycothiol as a functional analogue of glutathione [139,140]. The low level of SseA in the modern Beijing strains most likely results in increased DNA oxidation damage which explains the higher rate of mutation and accelerated acquisition of drug resistance compared to more ancient strains [141]. Up-regulation of enzymes responsible for long-chain fatty acid biosynthesis and HsaA implicated in steroid degradation and down-regulation of long-chain fatty acid degrading enzymes have been observed in Beijing B0/W148 strains in compared to the reference strain termed H37Rv [139]. The differential expression of 23 proteins implicated in virulence were confirmed by targeted proteomics in seven clinically relevant strains showing various degrees of pathogenicity [142]. Proteomic and transcriptional analyses also generated some insights on metabolic remodeling between different BCG strains which might be manifested by various degrees of immunogenicity and potentially vaccine efficacy [143]. The mycobacterial protein analysis of mono-infected and HIV co-infected macrophages revealed 92 significant changes which belong to various functional categories such as toxin–antitoxin (TA) modules, cation transporters and type VII (Esx) secretion systems [144].

Proteomic studies have also increased the depth of our knowledge about the significant regulatory pathways of Mtb. PhoP as a virulence factor regulates a small non-coding RNA (ncRNA) namely Mcr7 which affects the activity of the Twin Arginine Translocation (Tat) protein secretion system through TatC modulation. Consequently, the secretion of BlaC and the antigen 85 complex (Ag85), a key player in the pathogenicity, changes significantly [145]. To decipher the role of SecA2 dependent export pathway, the cell wall and cytosolic proteome of a SecA2 mutant were compared to the wild type introducing the association of the pathway with DosR regulon and the Mce1 and Mce4 lipid transporters [146]. Proteomic analyses of culture filtrate on Mtb revealed EsxG and EsxH, secreted co-dependently, facilitate the secretion of several members of the proline–glutamic acid and proline–proline–glutamic acid protein families such as PE5 [35].

Drug resistant strains of Mtb are a growing problem for healthcare systems and have been investigated using proteomic methods. Bedaquiline (BDQ), approved for the treatment of multidrug-resistant TB, inhibits ATP synthesis inducing a bacteriostatic state for 3-4 days after drug exposure. The induction of the DosR regulon as well as the activation of ATP-generating pathways promote bacterial viability during this initial drug exposure, explaining in part why BDQ is more effective when the bacilli have access to only non-fermentable energy sources such as lipids [147]. Studying the proteome of ofloxacin (OFX) resistant strains showed fourteen proteins up-regulated in respect to the OFX susceptible strains. Further docking analysis on four of the proteins elucidated conserved motifs and domains interacting with OFX as a second-line drug against MDR-TB [148]. A study showed that the abundance of several proteins responsible for the maintenance of cell-envelope permeability barrier changed significantly in Mtb exposed

to thioridazine. Thioridazine increases cell-envelope permeability and thereby facilitates components uptake [149].

With the recent developments in the DIA/SWATH-MS field, we can quantify ~2700 proteins of Mtb at semi-high-throughput (14 samples/day) consistently from 1 µg of total peptide mass over many samples and conditions. To support the quantitative measurement of the Mtb proteome by DIA/SWATH-MS, an Mtb proteome spectral library has been generated, validated and made publicly accessible [135]. It contains 97% of the annotated Mtb proteins and has paved the way to study the Mtb proteome under many different conditions. According to shotgun-MS and deep RNASeq experiment, we presume that 3488 proteins are expressed in Mtb cumulatively and that protein concentrations range from 0.1 to 1000 fmol/µg (10–44 632 estimated protein copies per cell), spanning four order of magnitude. GroEL1/2, MihF, GroES and Tuf are the most abundant proteins. Furthermore, 29 previously unannotated proteins have been identified by MS-based proteomics which emphasizes that the genome annotation of Mtb still needs to be further refined [135]. In a prototypical study, the absolute protein concentrations of the Mtb proteome and its reorganization after exposure to hypoxia was determined in a time course experiment. The results showed that whereas ribosomal proteins remain largely unchanged, products of DosR regulon genes were strongly induced to constitute 20% of the cellular protein content during dormancy. A quarter of 631 differentially expressed proteins had metabolic functions and 80% of them constituted connected metabolic pathways with at least four enzymes [136].

1.8 Introduction *Mycobacterium tuberculosis* host proteomics

Despite the advances in quantitative mass-spectrometry based proteomics that enable a sensitive and deep coverage of any given biological samples, comprehensive profiling of host proteomics responses to virulent *Mycobacterium tuberculosis* infection have lagged behind and remain to-date scarce. Consequently, the virulence-associated effects of *Mtb* infections on global host responses remain unexplored. Indeed, most proteomic studies quantifying the responses to live *Mtb* infections were predominantly done on fractionated organelles using various peptide labelling strategies including on the endoplasmic reticulum (ER) in response to either H37Rv or H37Ra after 4 hours of infections to measure a total of 133 ER proteins [150], on exosomes from *Mtb*-infected macrophages [151], to quantify newly synthesized proteins secreted by macrophages upon the first 26 hours of infections with 4 different virulent *Mtb* strains [152]. However, due to the more physiologically relevant sub-cellular compartments in *Mtb* infections, many have quantified the proteomic composition of bacilli-containing phagosomes with the aim of shedding light on the *Mtb*-driven phagosomal maturation arrest [153–155].

Whole-cell proteomic studies quantifying the host proteomic responses include an analysis of infected lungs to identify *Mtb*-specific antigens [156] and a comparative study to identify putative *in vitro* biomarkers for the clearance of *Mtb* infections after antimicrobial treatment [157]. However, more recent publications aimed at analyzing the more global host responses of macrophages infected by *Mtb* and *M. bovis* comparative infections by quantitatively measuring 2'000 proteins belonging to the host enabled them to pin-point strain specific host proteomic responses. Yet, the only paper to our knowledge to study the virulence-specific proteomic response to differential virulent *Mtb* strains quantified the response to H37Rv and H37Ra. By quantifying the altered protein expression profiles in response to both strains from the 6'700 host proteins measured, they could identify the 235 proteins significantly altered after 12 hours of infections. Results showed that more virulent *Mtb* infected macrophages resulted in differentially regulated proteins involved in oxidative phosphorylation, nucleosome assembly, vesicular formation, apoptosis and blood coagulation [158].

1.9 References

1. Paulson T: **A mortal foe**. *Nature* 2013, **502**.
2. *Status of the TB epidemic*. 2019.
3. Shete PB, Reid M, Goosby E: **Message to world leaders: we cannot end tuberculosis without addressing the social and economic burden of the disease**. *Lancet Glob. Heal.* 2018, **6**:e1272–e1273.
4. Oh P, Pascopella L, Barry PM, Flood JM: **A systematic synthesis of direct costs to treat and manage tuberculosis disease applied to California, 2015 [Internet]**. *BMC Res. Notes* 2017, **10**:434.
5. **Treatment Highlights of Drug-Susceptible Tuberculosis Guidelines | Treatment | TB | CDC [Internet]**. [date unknown], [no volume].
6. **WHO | Global tuberculosis report 2016 [Internet]**. *WHO* 2017, [no volume].
7. Pai M, Behr MA, Dowdy D, Dheda K, Divangahi M, Boehme CC, Ginsberg A, Swaminathan S, Spigelman M, Getahun H, et al.: **Tuberculosis [Internet]**. *Nat. Rev. Dis. Prim.* 2016, **2**:16076.
8. Lange C, Dheda K, Chesov D, Mandalakas AM, Udwadia Z, Horsburgh CR: **Management of drug-resistant tuberculosis**. *Lancet* 2019, **394**:953–966.
9. Russell DG, Barry CE, Flynn JL: **Tuberculosis: What we don't know can, and does, hurt us**. *Science (80-.).* 2010, **328**:852–856.
10. Kaufmann SHE: **How can immunology contribute to the control of tuberculosis? [Internet]**. *Nat. Rev. Immunol.* 2001, **1**:20–30.
11. Armstrong JA, Hart PD: **Phagosome-lysosome interactions in cultured macrophages infected with virulent tubercle bacilli. Reversal of the usual nonfusion pattern and observations on bacterial survival. [Internet]**. *J. Exp. Med.* 1975, **142**:1–16.
12. Moraco AH, Kornfeld H: **Cell death and autophagy in tuberculosis. [Internet]**. *Semin. Immunol.* 2014, **26**:497–511.
13. Chang ST, Linderman JJ, Kirschner DE: **Multiple mechanisms allow Mycobacterium tuberculosis to continuously inhibit MHC class II-mediated antigen presentation by macrophages. [Internet]**. *Proc. Natl. Acad. Sci. U. S. A.* 2005, **102**:4530–5.
14. Hanekom WA, Mendillo M, Manca C, Haslett PAJ, Siddiqui MR, Barry III C, Kaplan G: ***Mycobacterium tuberculosis* Inhibits Maturation of Human Monocyte-Derived Dendritic Cells In Vitro [Internet]**. *J. Infect. Dis.* 2003, **188**:257–266.
15. Jasenosky LD, Scriba TJ, Hanekom WA, Goldfeld AE: **T cells and adaptive immunity to *Mycobacterium tuberculosis* in humans [Internet]**. *Immunol. Rev.* 2015, **264**:74–87.

16. Mahairas GG, Sabo PJ, Hickey MJ, Singh DC, Stover CK: **Molecular analysis of genetic differences between Mycobacterium bovis BCG and virulent M. bovis. [Internet]**. *J. Bacteriol.* 1996, **178**:1274–82.
17. Hsu T, Hingley-Wilson SM, Chen B, Chen M, Dai AZ, Morin PM, Marks CB, Padiyar J, Goulding C, Gingery M, et al.: **The primary mechanism of attenuation of bacillus Calmette–Guérin is a loss of secreted lytic function required for invasion of lung interstitial tissue**. *Proc. Natl. Acad. Sci.* 2003, **100**.
18. Lewis KN, Liao R, Guinn KM, Hickey MJ, Smith S, Behr MA, Sherman DR: **Deletion of RD1 from Mycobacterium tuberculosis mimics bacille Calmette-Guérin attenuation. [Internet]**. *J. Infect. Dis.* 2003, **187**:117–23.
19. Pym AS, Brodin P, Brosch R, Huerre M, Cole ST: **Loss of RD1 contributed to the attenuation of the live tuberculosis vaccines Mycobacterium bovis BCG and Mycobacterium microti**. *Mol. Microbiol.* 2002, **46**:709–717.
20. Bosserman RE, Champion PA: **Esx systems and the mycobacterial cell envelope: What's the connection?** *J. Bacteriol.* 2017, **199**.
21. Shah S, Briken V: **Modular organization of the ESX-5 secretion system in Mycobacterium tuberculosis**. *Front. Cell. Infect. Microbiol.* 2016, **6**.
22. Renshaw PS, Lightbody KL, Veverka V, Muskett FW, Kelly G, Frenkiel TA, Gordon S V, Hewinson RG, Burke B, Norman J, et al.: **Structure and function of the complex formed by the tuberculosis virulence factors CFP-10 and ESAT-6 [Internet]**. *EMBO J.* 2005, **24**:2491–2498.
23. Pallen MJ: **The ESAT-6/WXG100 superfamily - And a new Gram-positive secretion system? [Internet]**. *Trends Microbiol.* 2002, **10**:209–212.
24. Gröschel MI, Sayes F, Simeone R, Majlessi L, Brosch R: **ESX secretion systems: mycobacterial evolution to counter host immunity [Internet]**. *Nat. Rev. Microbiol.* 2016, **14**:677–691.
25. Stanley SA, Raghavan S, Hwang WW, Cox JS: **Acute infection and macrophage subversion by Mycobacterium tuberculosis require a specialized secretion system**. *Proc. Natl. Acad. Sci. U. S. A.* 2003, **100**:13001–13006.
26. Guinn KM, Hickey MJ, Mathur SK, Zakel KL, Grotzke JE, Lewinsohn DM, Smith S, Sherman DR: **Individual RD1 -region genes are required for export of ESAT-6/CFP-10 and for virulence of Mycobacterium tuberculosis**. *Mol. Microbiol.* 2004, **51**:359–370.
27. Abdallah AM, Verboom T, Hannes F, Safi M, Strong M, Eisenberg D, Musters RJP, Vandenbroucke-Grauls CMJE, Appelmelk BJ, Luirink J, et al.: **A specific secretion system mediates PPE41 transport in pathogenic mycobacteria [Internet]**. *Mol. Microbiol.* 2006, **62**:667–679.
28. Zhang Q, Wang D, Jiang G, Liu W, Deng Q, Li X, Qian W, Ouellet H, Sun J: **EsxA membrane-permeabilizing activity plays a key role in mycobacterial cytosolic translocation and virulence: effects of single-residue mutations at glutamine 5 [Internet]**. *Sci. Rep.* 2016, **6**:32618.
29. Conrad WH, Osman MM, Shanahan JK, Chu F, Takaki KK, Cameron J, Hopkinson-Woolley D, Brosch R, Ramakrishnan L: **Mycobacterial ESX-1 secretion system mediates host cell lysis through bacterium contact-dependent gross membrane disruptions**. *Proc. Natl. Acad. Sci. U. S. A.* 2017, **114**:1371–1376.
30. Kaku T, Kawamura I, Uchiyama R, Kurenuma T, Mitsuyama M: **RD1 region in mycobacterial genome is involved in the induction of necrosis in infected RAW264 cells via mitochondrial membrane damage and ATP depletion**. *FEMS Microbiol. Lett.* 2007, **274**:189–195.
31. Volkman HE, Clay H, Beery D, Chang JCW, Sherman DR, Ramakrishnan L: **Tuberculous granuloma formation is enhanced by a Mycobacterium virulence determinant**. *PLoS Biol.* 2004, **2**.
32. Kurenuma T, Kawamura I, Hara H, Uchiyama R, Daim S, Dewamitta SR, Sakai S, Tsuchiya K, Nomura T, Mitsuyama M: **The RD1 locus in the Mycobacterium tuberculosis genome contributes to activation of caspase-1 via induction of potassium ion efflux in infected macrophages**. *Infect. Immun.* 2009, **77**:3992–4001.
33. van der Wel N, Hava D, Houben D, Fluitsma D, van Zon M, Pierson J, Brenner M, Peters PJ: **M. tuberculosis and M. leprae Translocate from the Phagolysosome to the Cytosol in Myeloid Cells [Internet]**. *Cell* 2007, **129**:1287–1298.
34. Siegrist MS, Unnikrishnan M, McConnell MJ, Borowsky M, Cheng TY, Siddiqi N, Fortune SM, Moody DB, Rubin EJ: **Mycobacterial Esx-3 is required for mycobactin-mediated iron acquisition**. *Proc. Natl. Acad. Sci. U. S. A.* 2009, **106**:18792–18797.
35. Tufariello JAM, Chapman JR, Kerantzas CA, Wong KW, Vilchèze C, Jones CM, Cole LE, Tinaztepe E, Thompson V, Fenyö D, et al.: **Separable roles for Mycobacterium tuberculosis ESX-3 effectors**

in iron acquisition and virulence. *Proc. Natl. Acad. Sci. U. S. A.* 2016, **113**:E348–E357.
36. Tinaztepe E, Wei JR, Raynowska J, Portal-Celhay C, Thompson V, Philipsa JA: **Role of metal-dependent regulation of ESX-3 secretion in intracellular survival of Mycobacterium tuberculosis**. *Infect. Immun.* 2016, **84**:2255–2263.
37. Mehra A, Zahra A, Thompson V, Sirisaengtaksin N, Wells A, Porto M, Köster S, Penberthy K, Kubota Y, Dricot A, et al.: **Mycobacterium tuberculosis Type VII Secreted Effector EsxH Targets Host ESCRT to Impair Trafficking [Internet]**. *PLoS Pathog.* 2013, **9**:e1003734.
38. Hervas-Stubbs S, Majlessi L, Simsova M, Morova J, Rojas MJ, Nouzé C, Brodin P, Sebo P, Leclerc C: **High frequency of CD4+ T cells specific for the TB10.4 protein correlates with protection against Mycobacterium tuberculosis infection**. *Infect. Immun.* 2006, **74**:3396–3407.
39. Skjøt RLV, Oettinger T, Rosenkrands I, Ravn P, Brock I, Jacobsen S, Andersen P: **Comparative evaluation of low-molecular-mass proteins from Mycobacterium tuberculosis identifies members of the ESAT-6 family as immunodominant T-cell antigens**. *Infect. Immun.* 2000, **68**:214–220.
40. Gey Van Pittius NC, Gamieldien J, Hide W, Brown GD, Siezen RJ, Beyers AD: **The ESAT-6 gene cluster of Mycobacterium tuberculosis and other high G+C Gram-positive bacteria.** *Genome Biol.* 2001, **2**.
41. Elliott SR, White DW, Tischler AD: **Mycobacterium tuberculosis requires regulation of ESX-5 secretion for virulence in irgm1-deficient mice**. *Infect. Immun.* 2019, **87**.
42. Houben ENG, Bestebroer J, Ummels R, Wilson L, Piersma SR, Jiménez CR, Ottenhoff THM, Luirink J, Bitter W: **Composition of the type VII secretion system membrane complex [Internet]**. *Mol. Microbiol.* 2012, **86**:472–484.
43. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon S V., Eiglmeier K, Gas S, Barry CE, et al.: **Deciphering the biology of mycobacterium tuberculosis from the complete genome sequence**. *Nature* 1998, **393**:537–544.
44. Ates LS, van der Woude AD, Bestebroer J, van Stempvoort G, Musters RJP, Garcia-Vallejo JJ, Picavet DI, Weerd R van de, Maletta M, Kuijl CP, et al.: **The ESX-5 System of Pathogenic Mycobacteria Is Involved In Capsule Integrity and Virulence through Its Substrate PPE10 [Internet]**. *PLOS Pathog.* 2016, **12**:e1005696.
45. Abdallah AM, Savage NDL, van Zon M, Wilson L, Vandenbroucke-Grauls CMJE, van der Wel NN, Ottenhoff THM, Bitter W: **The ESX-5 Secretion System of Mycobacterium marinum Modulates the Macrophage Response** . *J. Immunol.* 2008, **181**:7166–7175.
46. Hartwell LH, Hopfield JJ, Leibler S, Murray AW: **From molecular to modular cell biology. [Internet]**. *Nature* 1999, **402**:C47-52.
47. Bhavsar AP, Guttman JA, Finlay BB: **Manipulation of host-cell pathways by bacterial pathogens [Internet]**. *Nature* 2007, **449**:827–834.
48. Elde NC, Malik HS: **The evolutionary conundrum of pathogen mimicry [Internet]**. *Nat. Rev. Microbiol.* 2009, **7**:787–797.
49. Crua Asensio N, Muñoz Giner E, de Groot NS, Torrent Burgas M, Lotteau V: **Centrality in the host–pathogen interactome is associated with pathogen fitness during infection [Internet]**. *Nat. Commun.* 2017, **8**:14092.
50. Shapira SD, Gat-Viks I, Shum BOV, Dricot A, de Grace MM, Wu L, Gupta PB, Hao T, Silver SJ, Root DE, et al.: **A Physical and Regulatory Map of Host-Influenza Interactions Reveals Pathways in H1N1 Infection [Internet]**. *Cell* 2009, **139**:1255–1267.
51. Green ER, Mecsas J: **Bacterial Secretion Systems: An Overview. [Internet]**. *Microbiol. Spectr.* 2016, **4**.
52. Tesh VL: **Induction of apoptosis by Shiga toxins. [Internet]**. *Future Microbiol.* 2010, **5**:431–53.
53. Kunsmann L, Rüter C, Bauwens A, Greune L, Glüder M, Kemper B, Fruth A, Wai SN, He X, Lloubes R, et al.: **Virulence from vesicles: Novel mechanisms of host cell injury by Escherichia coli O104:H4 outbreak strain [Internet]**. *Sci. Rep.* 2015, **5**:13252.
54. Durmuş Tekir S, Cakir T, Ulgen KÖ: **Infection Strategies of Bacterial and Viral Pathogens through Pathogen-Human Protein-Protein Interactions. [Internet]**. *Front. Microbiol.* 2012, **3**:46.
55. Colonne PM, Winchell CG, Voth DE: **Hijacking Host Cell Highways: Manipulation of the Host Actin Cytoskeleton by Obligate Intracellular Bacterial Pathogens. [Internet]**. *Front. Cell. Infect. Microbiol.* 2016, **6**:107.
56. Mostowy S, Shenoy AR: **The cytoskeleton in cell-autonomous immunity: structural determinants of host defence [Internet]**. *Nat. Rev. Immunol.* 2015, **15**:559–573.
57. Saini NK, Baena A, Ng TW, Venkataswamy MM, Kennedy SC, Kunnath-Velayudhan S, Carreño LJ, Xu

J, Chan J, Larsen MH, et al.: **Suppression of autophagy and antigen presentation by Mycobacterium tuberculosis PE_PGRS47 [Internet]**. *Nat. Microbiol.* 2016, **1**:16133.
58. Blasche S, Arens S, Ceol A, Siszler G, Schmidt MA, Häuser R, Schwarz F, Wuchty S, Aloy P, Uetz P, et al.: **The EHEC-host interactome reveals novel targets for the translocated intimin receptor. [Internet]**. *Sci. Rep.* 2014, **4**:7531.
59. Wallqvist A, Memišević V, Zavaljevski N, Pieper R, Rajagopala S V, Kwon K, Yu C, Hoover TA, Reifman J: **Using host-pathogen protein interactions to identify and characterize Francisella tularensis virulence factors. [Internet]**. *BMC Genomics* 2015, **16**:1106.
60. Memisević V, Zavaljevski N, Pieper R, Rajagopala S V, Kwon K, Townsend K, Yu C, Yu X, DeShazer D, Reifman J, et al.: **Novel Burkholderia mallei virulence factors linked to specific host-pathogen protein interactions. [Internet]**. *Mol. Cell. Proteomics* 2013, **12**:3036–51.
61. Dyer MD, Neff C, Dufford M, Rivera CG, Shattuck D, Bassaganya-Riera J, Murali TM, Sobral BW: **The Human-Bacterial Pathogen Protein Interaction Networks of Bacillus anthracis, Francisella tularensis, and Yersinia pestis [Internet]**. *PLoS One* 2010, **5**:e12089.
62. Yang H, Ke Y, Wang J, Tan Y, Myeni SK, Li D, Shi Q, Yan Y, Chen H, Guo Z, et al.: **Insight into bacterial virulence mechanisms against host immune response via the Yersinia pestis-human protein-protein interaction network. [Internet]**. *Infect. Immun.* 2011, **79**:4413–24.
63. Mirrashidi KM, Elwell CA, Verschueren E, Johnson JR, Frando A, Von Dollen J, Rosenberg O, Gulbahce N, Jang G, Johnson T, et al.: **Global Mapping of the Inc-Human Interactome Reveals that Retromer Restricts Chlamydia Infection [Internet]**. *Cell Host Microbe* 2015, **18**:109–121.
64. Auweter SD, Bhavsar AP, de Hoog CL, Li Y, Chan YA, van der Heijden J, Lowden MJ, Coombes BK, Rogers LD, Stoynov N, et al.: **Quantitative mass spectrometry catalogues Salmonella pathogenicity island-2 effectors and identifies their cognate host binding partners. [Internet]**. *J. Biol. Chem.* 2011, **286**:24023–35.
65. So EC, Schroeder GN, Carson D, Mattheis C, Mousnier A, Broncel M, Tate EW, Frankel G: **The Rab-binding Profiles of Bacterial Virulence Factors during Infection. [Internet]**. *J. Biol. Chem.* 2016, **291**:5832–43.
66. Hicks SW, Galán JE: **Exploitation of eukaryotic subcellular targeting mechanisms by bacterial effectors [Internet]**. *Nat. Rev. Microbiol.* 2013, **11**:316–326.
67. Zhou Y, Zhu Y: **Diversity of bacterial manipulation of the host ubiquitin pathways [Internet]**. *Cell. Microbiol.* 2015, **17**:26–34.
68. Ashida H, Kim M, Sasakawa C: **Exploitation of the host ubiquitin system by human bacterial pathogens [Internet]**. *Nat. Rev. Microbiol.* 2014, **12**:399–413.
69. Penn BH, Netter Z, Johnson JR, Von Dollen J, Jang GM, Johnson T, Ohol YM, Maher C, Bell SL, Geiger K, et al.: **An Mtb-Human Protein-Protein Interaction Map Identifies a Switch between Host Antiviral and Antibacterial Responses [Internet]**. *Mol. Cell* 2018, **71**:637-648.e5.
70. Portal-Celhay C, Tufariello JM, Srivastava S, Zahra A, Klevorn T, Grace PS, Mehra A, Park HS, Ernst JD, Jacobs WR, et al.: **Mycobacterium tuberculosis EsxH inhibits ESCRT-dependent CD4+ T-cell activation [Internet]**. *Nat. Microbiol.* 2016, **2**:16232.
71. Mehta V, Trinkle-Mulcahy L, Mehta V, Trinkle-Mulcahy L: **Recent advances in large-scale protein interactome mapping [Internet]**. *F1000Research* 2016, **5**:782.
72. Lum KK, Cristea IM: **Proteomic approaches to uncovering virus-host protein interactions during the progression of viral infection. [Internet]**. *Expert Rev. Proteomics* 2016, **13**:325–40.
73. Fields S, Song O: **A novel genetic system to detect protein□protein interactions [Internet]**. *Nature* 1989, **340**:245–246.
74. Calderwood MA, Venkatesan K, Xing L, Chase MR, Vazquez A, Holthaus AM, Ewence AE, Li N, Hirozane-Kishikawa T, Hill DE, et al.: **Epstein-Barr virus and virus human protein interaction maps. [Internet]**. *Proc. Natl. Acad. Sci. U. S. A.* 2007, **104**:7606–11.
75. de Chassey B, Navratil V, Tafforeau L, Hiet MS, Aublin-Gex A, Agaugué S, Meiffren G, Pradezynski F, Faria BF, Chantier T, et al.: **Hepatitis C virus infection protein network. [Internet]**. *Mol. Syst. Biol.* 2008, **4**:230.
76. Shapira SD, Gat-Viks I, Shum BOV, Dricot A, de Grace MM, Wu L, Gupta PB, Hao T, Silver SJ, Root DE, et al.: **A Physical and Regulatory Map of Host-Influenza Interactions Reveals Pathways in H1N1 Infection [Internet]**. *Cell* 2009, **139**:1255–1267.
77. Lee S, Salwinski L, Zhang C, Chu D, Sampankanpanich C, Reyes NA, Vangeloff A, Xing F, Li X, Wu T-T, et al.: **An Integrated Approach to Elucidate the Intra-Viral and Viral-Cellular Protein Interaction Networks of a Gamma-Herpesvirus [Internet]**. *PLoS Pathog.* 2011, **7**:e1002297.
78. Gulbahce N, Yan H, Dricot A, Padi M, Byrdsong D, Franchi R, Lee D-S, Rozenblatt-Rosen O, Mar JC, Calderwood MA, et al.: **Viral Perturbations of Host Networks Reflect Disease Etiology**

[Internet]. *PLoS Comput. Biol.* 2012, **8**:e1002531.
79. Khadka S, Vangeloff AD, Zhang C, Siddavatam P, Heaton NS, Wang L, Sengupta R, Sahasrabudhe S, Randall G, Gribskov M, et al.: **A physical interaction network of dengue virus and human proteins. [Internet]**. *Mol. Cell. Proteomics* 2011, **10**:M111.012187.
80. Mairiang D, Zhang H, Sodja A, Murali T, Suriyaphol P, Malasit P, Limjindaporn T, Finley RL: **Identification of New Protein Interactions between Dengue Fever Virus and Its Hosts, Human and Mosquito [Internet]**. *PLoS One* 2013, **8**:e53535.
81. Dolan PT, Zhang C, Khadka S, Arumugaswami V, Vangeloff AD, Heaton NS, Sahasrabudhe S, Randall G, Sun R, LaCount DJ, et al.: **Identification and comparative analysis of hepatitis C virus–host cell protein interactions [Internet]**. *Mol. Biosyst.* 2013, **9**:3199.
82. Rozenblatt-Rosen O, Deo RC, Padi M, Adelmant G, Calderwood MA, Rolland T, Grace M, Dricot A, Askenazi M, Tavares M, et al.: **Interpreting cancer genomes using systematic host network perturbations by tumour virus proteins. [Internet]**. *Nature* 2012, **487**:491–5.
83. Muller M, Jacob Y, Jones L, Weiss A, Brino L, Chantier T, Lotteau V, Favre M, Demeret C: **Large Scale Genotype Comparison of Human Papillomavirus E2-Host Interaction Networks Provides New Insights for E2 Molecular Functions [Internet]**. *PLoS Pathog.* 2012, **8**:e1002761.
84. Gingras A-C, Gstaiger M, Raught B, Aebersold R: **Analysis of protein complexes using mass spectrometry [Internet]**. *Nat. Rev. Mol. Cell Biol.* 2007, **8**:645–654.
85. White EA, Kramer RE, Tan MJA, Hayes SD, Harper JW, Howley PM: **Comprehensive analysis of host cellular interactions with human papillomavirus E6 proteins identifies new E6 binding partners and reflects viral diversity. [Internet]**. *J. Virol.* 2012, **86**:13174–86.
86. White EA, Sowa ME, Tan MJA, Jeudy S, Hayes SD, Santha S, Münger K, Harper JW, Howley PM: **Systematic identification of interactions between host cell proteins and E7 oncoproteins from diverse human papillomaviruses. [Internet]**. *Proc. Natl. Acad. Sci. U. S. A.* 2012, **109**:E260-7.
87. Kane JR, Stanley DJ, Hultquist JF, Johnson JR, Mietrach N, Binning JM, Jónsson SR, Barelier S, Newton BW, Johnson TL, et al.: **Lineage-Specific Viral Hijacking of Non-canonical E3 Ubiquitin Ligase Cofactors in the Evolution of Vif Anti-APOBEC3 Activity [Internet]**. *Cell Rep.* 2015, **11**:1236–1250.
88. Colpitts TM, Cox J, Nguyen A, Feitosa F, Krishnan MN, Fikrig E: **Use of a tandem affinity purification assay to detect interactions between West Nile and dengue viral proteins and proteins of the mosquito vector [Internet]**. *Virology* 2011, **417**:179–187.
89. Ramage HR, Kumar GR, Verschueren E, Johnson JR, Von Dollen J, Johnson T, Newton B, Shah P, Horner J, Krogan NJ, et al.: **A Combined Proteomics/Genomics Approach Links Hepatitis C Virus Infection with Nonsense-Mediated mRNA Decay [Internet]**. *Mol. Cell* 2015, **57**:329–340.
90. Jäger S, Cimermancic P, Gulbahce N, Johnson JR, McGovern KE, Clarke SC, Shales M, Mercenne G, Pache L, Li K, et al.: **Global landscape of HIV–human protein complexes [Internet]**. *Nature* 2011, **481**:365.
91. Watanabe T, Kawakami E, Shoemaker JE, Lopes TJS, Matsuoka Y, Tomita Y, Kozuka-Hata H, Gorai T, Kuwahara T, Takeda E, et al.: **Influenza Virus-Host Interactome Screen as a Platform for Antiviral Drug Development [Internet]**. *Cell Host Microbe* 2014, **16**:795–805.
92. Davis ZH, Verschueren E, Jang GM, Kleffman K, Johnson JR, Park J, Von Dollen J, Maher MC, Johnson T, Newton W, et al.: **Global Mapping of Herpesvirus-Host Protein Complexes Reveals a Transcription Strategy for Late Genes [Internet]**. *Mol. Cell* 2015, **57**:349–360.
93. Pichlmair A, Kandasamy K, Alvisi G, Mulhern O, Sacco R, Habjan M, Binder M, Stefanovic A, Eberle C-A, Goncalves A, et al.: **Viral immune modulators perturb the human molecular network by common and unique strategies [Internet]**. *Nature* 2012, **487**:486–490.
94. Germain M-A, Chatel-Chaix L, Gagné B, Bonneil É, Thibault P, Pradezynski F, de Chassey B, Meyniel-Schicklin L, Lotteau V, Baril M, et al.: **Elucidating novel hepatitis C virus-host interactions using combined mass spectrometry and functional genomics approaches. [Internet]**. *Mol. Cell. Proteomics* 2014, **13**:184–203.
95. Yamayoshi S, Noda T, Ebihara H, Goto H, Morikawa Y, Lukashevich IS, Neumann G, Feldmann H, Kawaoka Y: **Ebola Virus Matrix Protein VP40 Uses the COPII Transport System for Its Intracellular Transport [Internet]**. *Cell Host Microbe* 2008, **3**:168–177.
96. Malik-Soni N, Frappier L: **Proteomic profiling of EBNA1-host protein interactions in latent and lytic Epstein-Barr virus infections. [Internet]**. *J. Virol.* 2012, **86**:6999–7002.
97. Posner MG, Upadhyay A, Abubaker AA, Fortunato TM, Vara D, Canobbio I, Bagby S, Pula G: **Extracellular Fibrinogen-binding Protein (Efb) from Staphylococcus aureus Inhibits the Formation of Platelet-Leukocyte Complexes. [Internet]**. *J. Biol. Chem.* 2016, **291**:2764–76.

98. Cristea IM, Carroll J-WN, Rout MP, Rice CM, Chait BT, MacDonald MR: **Tracking and elucidating alphavirus-host protein interactions. [Internet]**. *J. Biol. Chem.* 2006, **281**:30269–78.
99. Luo Y, Jacobs EY, Greco TM, Mohammed KD, Tong T, Keegan S, Binley JM, Cristea IM, Fenyö D, Rout MP, et al.: **HIV–host interactome revealed directly from infected cells [Internet]**. *Nat. Microbiol.* 2016, **1**:16068.
100. Mousnier A, Schroeder GN, Stoneham CA, So EC, Garnett JA, Yu L, Matthews SJ, Choudhary JS, Hartland EL, Frankel G: **A New Method To Determine In Vivo Interactomes Reveals Binding of the Legionella pneumophila Effector PieE to Multiple Rab GTPases [Internet]**. *MBio* 2014, **5**:e01148-14-e01148-14.
101. Roux KJ, Kim DI, Raida M, Burke B: **A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells [Internet]**. *J. Cell Biol.* 2012, **196**:801–810.
102. Lambert J-P, Tucholska M, Go C, Knight JDR, Gingras A-C: **Proximity biotinylation and affinity purification are complementary approaches for the interactome mapping of chromatin-associated protein complexes [Internet]**. *J. Proteomics* 2015, **118**:81–94.
103. Mojica SA, Hovis KM, Frieman MB, Tran B, Hsia R, Ravel J, Jenkins-Houk C, Wilson KL, Bavoil PM: **SINC, a type III secreted protein of Chlamydia psittaci, targets the inner nuclear membrane of infected cells and uninfected neighbors. [Internet]**. *Mol. Biol. Cell* 2015, **26**:1918–34.
104. Lobingier BT, Hüttenhain R, Eichel K, Miller KB, Ting AY, von Zastrow M, Krogan NJ: **An Approach to Spatiotemporally Resolve Protein Interaction Networks in Living Cells [Internet]**. *Cell* 2017, **169**:350-360.e12.
105. Leitner A, Faini M, Stengel F, Aebersold R: **Crosslinking and Mass Spectrometry: An Integrated Technology to Understand the Structure and Function of Molecular Machines [Internet]**. *Trends Biochem. Sci.* 2016, **41**:20–32.
106. Makowski MM, Willems E, Jansen PWTC, Vermeulen M: **Cross-linking immunoprecipitation-MS (xIP-MS): Topological Analysis of Chromatin-associated Protein Complexes Using Single Affinity Purification [Internet]**. *Mol. Cell. Proteomics* 2016, **15**:854–865.
107. Shi Y, Pellarin R, Fridy PC, Fernandez-Martinez J, Thompson MK, Li Y, Wang QJ, Sali A, Rout MP, Chait BT: **A strategy for dissecting the architectures of native macromolecular assemblies [Internet]**. *Nat. Methods* 2015, **12**:1135–1138.
108. Rinner O, Seebacher J, Walzthoeni T, Mueller L, Beck M, Schmidt A, Mueller M, Aebersold R: **Identification of cross-linked peptides from large sequence databases [Internet]**. *Nat. Methods* 2008, **5**:315–8.
109. Navare AT, Chavez JD, Zheng C, Weisbrod CR, Eng JK, Siehnel R, Singh PK, Manoil C, Bruce JE: **Probing the Protein Interaction Network of Pseudomonas aeruginosa Cells by Chemical Cross-Linking Mass Spectrometry [Internet]**. *Structure* 2015, **23**:762–773.
110. Schweppe DK, Harding C, Chavez JD, Wu X, Ramage E, Singh PK, Manoil C, Bruce JE: **Host-Microbe Protein Interactions during Bacterial Infection [Internet]**. *Chem. Biol.* 2015, **22**:1521–1530.
111. Ji C, Li S, Reilly JP, Radivojac P, Tang H: **XLSearch: A probabilistic database search algorithm for identifying cross-linked peptides**. *J. Proteome Res.* 2016, **15**:1830–1841.
112. Kristensen AR, Gsponer J, Foster LJ: **A high-throughput approach for measuring temporal changes in the interactome [Internet]**. *Nat. Methods* 2012, **9**:907–909.
113. Havugimana PC, Hart GT, Nepusz TS, Yang H, Turinsky AL, Li Z, Wang PI, Boutz DR, Fong V, Phanse S, et al.: **A Census of Human Soluble Protein Complexes [Internet]**. *Cell* 2012, **150**:1068–1081.
114. Wan C, Borgeson B, Phanse S, Tu F, Drew K, Clark G, Xiong X, Kagan O, Kwan J, Bezginov A, et al.: **Panorama of ancient metazoan macromolecular complexes**. *Nature* 2015, **525**:339–344.
115. Scott NE, Brown LM, Kristensen AR, Foster LJ: **Development of a computational framework for the analysis of protein correlation profiling and spatial proteomics experiments**. *J. Proteomics* 2015, **118**:112–129.
116. Wang G, Xia Y, Song X, Ai L: **Common Non-classically Secreted Bacterial Proteins with Experimental Evidence [Internet]**. *Curr. Microbiol.* 2016, **72**:102–111.
117. Etna MP, Giacomini E, Pardini M, Severa M, Bottai D, Cruciani M, Rizzo F, Calogero R, Brosch R, Coccia EM: **Impact of Mycobacterium tuberculosis RD1-locus on human primary dendritic cell immune functions [Internet]**. *Sci. Rep.* 2015, **5**:17078.
118. Gygi SP, Rochon Y, Franza BR, Aebersold R: **Correlation between Protein and mRNA Abundance in Yeast**. *Mol. Cell. Biol.* 1999, **19**:1720–1730.
119. Anderson L, Seilhamer J: **A comparison of selected mRNA and protein abundances in human liver**. In *Electrophoresis*. . 1997:533–537.
120. Duan G, Walther D: **The Roles of Post-translational Modifications in the Context of Protein**

Interaction Networks. *PLoS Comput. Biol.* 2015, **11**.
121. Singhania A, Wilkinson RJ, Rodrigue M, Haldar P, O'Garra A: **The value of transcriptomics in advancing knowledge of the immune response and diagnosis in tuberculosis**. *Nat. Immunol.* 2018, **19**:1159–1168.
122. Donnelly DP, Rawlins CM, DeHart CJ, Fornelli L, Schachner LF, Lin Z, Lippens JL, Aluri KC, Sarin R, Chen B, et al.: **Best practices and benchmarks for intact protein analysis for top-down mass spectrometry**. *Nat. Methods* 2019, **16**:587–594.
123. Mann M, Kelleher NL: **Precision proteomics: The case for high resolution and high mass accuracy**. *Proc. Natl. Acad. Sci. U. S. A.* 2008, **105**:18132–18138.
124. Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM: **Electrospray ionization for mass spectrometry of large biomolecules**. *Science (80-.).* 1989, **246**:64–71.
125. Reid GE, McLuckey SA: **'Top down' protein characterization via tandem mass spectrometry**. *J. Mass Spectrom.* 2002, **37**:663–675.
126. Eng JK, Jahan TA, Hoopmann MR: **Comet: An open-source MS/MS sequence database search tool [Internet]**. *Proteomics* 2013, **13**:22–24.
127. Craig R, Beavis RC: **TANDEM: Matching proteins with tandem mass spectra [Internet]**. *Bioinformatics* 2004, **20**:1466–1467.
128. Gillet LC, Navarro P, Tate S, Röst H, Selevsek N, Reiter L, Bonner R, Aebersold R: **Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. [Internet]**. *Mol. Cell. Proteomics* 2012, **11**:O111.016717.
129. Collins BC, Hunter CL, Liu Y, Schilling B, Rosenberger G, Bader SL, Chan DW, Gibson BW, Gingras A-C, Held JM, et al.: **Multi-laboratory assessment of reproducibility, qualitative and quantitative performance of SWATH-mass spectrometry [Internet]**. *Nat. Commun.* 2017, **8**:291.
130. Röst HL, Rosenberger G, Navarro P, Gillet L, Miladinović SM, Schubert OT, Wolski W, Collins BC, Malmström J, Malmström L, et al.: **OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data [Internet]**. *Nat. Biotechnol.* 2014, **32**:219–223.
131. Meier F, Brunner A-D, Frank M, Ha A, Voytik E, Kaspar-Schoenefeld S, Lubeck M, Raether O, Aebersold R, Collins BC, et al.: **Parallel accumulation – serial fragmentation combined with data-independent acquisition (diaPASEF): Bottom-up proteomics with near optimal ion usage [Internet]**. *bioRxiv* 2019, doi:10.1101/656207.
132. Collins BC, Gillet LC, Rosenberger G, Röst HL, Vichalkovski A, Gstaiger M, Aebersold R: **Quantifying protein interaction dynamics by SWATH mass spectrometry: application to the 14-3-3 system [Internet]**. *Nat. Methods* 2013, **10**:1246–1253.
133. Caron E, Roncagalli R, Hase T, Wolski WE, Choi M, Menoita MG, Durand S, García-Blesa A, Fierro-Monti I, Sajic T, et al.: **Precise Temporal Profiling of Signaling Complexes in Primary Cells Using SWATH Mass Spectrometry. [Internet]**. *Cell Rep.* 2017, **18**:3219–3226.
134. de Keijzer J, Mulder A, de Ru AH, van Soolingen D, van Veelen PA: **Parallel reaction monitoring of clinical Mycobacterium tuberculosis lineages reveals pre-existent markers of rifampicin tolerance in the emerging Beijing lineage [Internet]**. *J. Proteomics* 2017, **150**:9–17.
135. Schubert OT, Mouritsen J, Ludwig C, Röst HL, Rosenberger G, Arthur PK, Claassen M, Campbell DS, Sun Z, Farrah T, et al.: **The Mtb proteome library: a resource of assays to quantify the complete proteome of Mycobacterium tuberculosis. [Internet]**. *Cell Host Microbe* 2013, **13**:602–12.
136. Schubert OT, Ludwig C, Kogadeeva M, Zimmermann M, Rosenberger G, Gengenbacher M, Gillet LC, Collins BC, Rost HL, Kaufmann SHE, et al.: **Absolute proteome composition and dynamics during dormancy and resuscitation of mycobacterium tuberculosis [Internet]**. *Cell Host Microbe* 2015, **18**:96–108.
137. Målen H, De Souza GA, Pathak S, Søfteland T, Wiker HG: **Comparison of membrane proteins of Mycobacterium tuberculosis H37Rv and H37Ra strains [Internet]**. *BMC Microbiol.* 2011, **11**:18.
138. Jhingan GD, Kumari S, Jamwal S V., Kalam H, Arora D, Jain N, Kumaar LK, Samal A, Rao KVS, Kumar D, et al.: **Comparative Proteomic Analyses of Avirulent, Virulent, and Clinical Strains of Mycobacterium tuberculosis Identify Strain-specific Patterns [Internet]**. *J. Biol. Chem.* 2016, **291**:14257–14273.
139. Bespyatykh J, Shitikov E, Butenko I, Altukhov I, Alexeev D, Mokrousov I, Dogonadze M, Zhuravlev V, Yablonsky P, Ilina E, et al.: **Proteome analysis of the Mycobacterium tuberculosis Beijing B0/W148 cluster**. *Sci. Rep.* 2016, **6**:1–11.

140. Nambi S, Long JE, Mishra BB, Baker R, Murphy KC, Olive AJ, Nguyen HP, Shaffer SA, Sassetti CM: **The Oxidative Stress Network of Mycobacterium tuberculosis Reveals Coordination between Radical Detoxification Systems [Internet]**. *Cell Host Microbe* 2015, **17**:829–837.
141. De Keijzer J, De Haas PE, De Ru AH, Van Veelen PA, Van Soolingen D: **Disclosure of selective advantages in the "modern" sublineage of the Mycobacterium tuberculosis Beijing genotype family by quantitative proteomics [Internet]**. *Mol. Cell. Proteomics* 2014, **13**:2632–2645.
142. Peters JS, Calder B, Gonnelli G, Degroeve S, Rajaonarifara E, Mulder N, Soares NC, Martens L, Blackburn JM: **Identification of quantitative proteomic differences between Mycobacterium tuberculosis lineages with altered virulence**. *Front. Microbiol.* 2016, **7**.
143. Abdallah AM, Hill-Cawthorne GA, Otto TD, Coll F, Guerra-Assunção JA, Gao G, Naeem R, Ansari H, Malas TB, Adroub SA, et al.: **Genomic expression catalogue of a global collection of BCG vaccine strains show evidence for highly diverged metabolic and cell-wall adaptations [Internet]**. *Sci. Rep.* 2015, **5**:15443.
144. Ganji R, Dhali S, Rizvi A, Rapole S, Banerjee S: **Understanding HIV-Mycobacteria synergism through comparative proteomics of intra-phagosomal mycobacteria during mono- and HIV co-infection**. *Sci. Rep.* 2016, **6**:1–14.
145. Solans L, Gonzalo-Asensio J, Sala C, Benjak A, Uplekar S, Rougemont J, Guilhot C, Malaga W, Martín C, Cole ST: **The PhoP-Dependent ncRNA Mcr7 Modulates the TAT Secretion System in Mycobacterium tuberculosis [Internet]**. *PLoS Pathog.* 2014, **10**:e1004183.
146. Feltcher ME, Gunawardena HP, Zulauf KE, Malik S, Griffin JE, Sassetti CM, Chen X, Braunstein M: **Label-free quantitative proteomics reveals a role for the Mycobacterium tuberculosis SecA2 pathway in exporting solute binding proteins and Mce transporters to the cell wall [Internet]**. *Mol. Cell. Proteomics* 2015, **14**:1501–1516.
147. Koul A, Vranckx L, Dhar N, Göhlmann HWH, Özdemir E, Neefs JM, Schulz M, Lu P, Mørtz E, McKinney JD, et al.: **Delayed bactericidal response of Mycobacterium tuberculosis to bedaquiline involves remodelling of bacterial metabolism**. *Nat. Commun.* 2014, **5**:3369.
148. Lata M, Sharma D, Deo N, Tiwari PK, Bisht D, Venkatesan K: **Proteomic analysis of ofloxacin-mono resistant Mycobacterium tuberculosis isolates [Internet]**. *J. Proteomics* 2015, **127**:114–121.
149. de Keijzer J, Mulder A, de Haas PEW, de Ru AH, Heerkens EM, Amaral L, van Soolingen D, van Veelen PA: **Thioridazine Alters the Cell-Envelope Permeability of *Mycobacterium tuberculosis* [Internet]**. *J. Proteome Res.* 2016, **15**:1776–1786.
150. Saquib NM, Jamwal S, Midha MK, Narain Verma H, Manivel V: **Quantitative Proteomics and Lipidomics Analysis of Endoplasmic Reticulum of Macrophage Infected with Mycobacterium tuberculosis [Internet]**. 2015, doi:10.1155/2015/270438.
151. Diaz G, Wolfe LM, Kruh-Garcia NA, Dobos KM: **Changes in the Membrane-Associated Proteins of Exosomes Released from Human Macrophages after Mycobacterium tuberculosis Infection**. *Sci. Rep.* 2016, **6**:1–10.
152. Kumar A, Jamwal S, Midha MK, Hamza B, Aggarwal S, Yadav AK, Rao KVS: **Dataset generated using hyperplexing and click chemistry to monitor temporal dynamics of newly synthesized macrophage secretome post infection by mycobacterial strains**. *Data Br.* 2016, **9**:349–354.
153. Herweg J-A, Hansmeier N, Otto A, Geffken AC, Subbarayal P, Prusty BK, Becher D, Hensel M, Schaible UE, Rudel T, et al.: **Purification and proteomics of pathogen-modified vacuoles and membranes. [Internet]**. *Front. Cell. Infect. Microbiol.* 2015, **5**:48.
154. Li Q, Singh CR, Ma S, Price ND, Jagannath C: **Label-free proteomics and systems biology analysis of mycobacterial phagosomes in dendritic cells and macrophages [Internet]**. *J. Proteome Res.* 2011, **10**:2425–2439.
155. Rao PK, Singh CR, Jagannath C, Li Q: **A systems biology approach to study the phagosomal proteome modulated by mycobacterial infections. [Internet]**. *Int. J. Clin. Exp. Med.* 2009, **2**:233–47.
156. Yu Y, Jin D, Hu S, Zhang Y, Zheng X, Zheng J, Liao M, Chen X, Graner M, Liu H, et al.: **A novel tuberculosis antigen identified from human tuberculosis granulomas**. *Mol. Cell. Proteomics* 2015, **14**:1093–1103.
157. Kaewseekhao B, Naranbhai V, Roytrakul S, Namwat W, Paemanee A, Lulitanond V, Chaiprasert A, Faksri K: **Comparative Proteomics of Activated THP-1 Cells Infected with Mycobacterium tuberculosis Identifies Putative Clearance Biomarkers for Tuberculosis Treatment [Internet]**. *PLoS One* 2015, **10**:e0134168.

158. Li H, Wei S, Fang Y, Li M, Li X, Li Z, Zhang J, Zhu G, Li C, Bi L, et al.: **Quantitative proteomic analysis of host responses triggered by Mycobacterium tuberculosis infection in human macrophage cells**. *Acta Biochim. Biophys. Sin. (Shanghai).* 2017, **49**:835–844.
159. Nicod C, Banaei-Esfahani A, Collins BC: **Elucidation of host–pathogen protein–protein interactions to uncover mechanisms of host cell rewiring [Internet]**. *Curr. Opin. Microbiol.* 2017, **39**:7–15.
160. Banaei-Esfahani A, Nicod C, Aebersold R, Collins BC: **Systems proteomics approaches to study bacterial pathogens: application to Mycobacterium tuberculosis [Internet]**. *Curr. Opin. Microbiol.* 2017, **39**:64–72.

Chapter 2

Motivation and outline of the thesis

This chapter was exclusively written by Charlotte Nicod.

2.0 Motivation and outline of the thesis

Mycobacterium tuberculosis persists as a global health issue that affects 10 million individuals annually across the world amongst whom 1.5 million victims succumb to the pulmonary tuberculosis disease. Despite the discovery of the vaccine strain named BCG a hundred years ago by the French bacteriologists Albert Calmette and Camille Guéri which has a limited efficacy in preventing tuberculosis, we still have not found a better performing prophylactic vaccine that could provide a protective immunity to the global population and help towards eradicating the disease. Furthermore, chemotherapies based on antibiotics are slowly losing in potency to combat TB infections as they are giving rise to alarming numbers of multi drug and extended drug resistant strains. Together, this supports the need to gain a better understanding of the virulence mechanisms behind *Mycobacterium tuberculosis* pathogenesis in order to develop novel treatment strategies.

In recent years, the discovery of the virulence associated Region of Difference 1 locus in the *M. tuberculosis* genome that encodes a protein secretion apparatus named ESX-1 has consolidated our hypotheses that host-pathogen interactions represent crucial interfaces by which pathogens immunomodulate their hosts to their own advantage. Although it is now undisputed that *M. tuberculosis* secretes proteins that directly interact with host cellular modules thus preventing adequate immune responses, their direct targets and global effects on the host response at the proteomic level remain obscure. In this thesis, we aim at not only quantifying the immunomodulatory effects of this protein secretion apparatus by analysing the host proteomic responses, but we also intent on discovering the exact physical interactions points between the pathogenic bacilli and their human host cells. We then set the challenge of developing a feasible method to be able to map such host-pathogen interaction points in the context of a real infection. To achieve these goals, we separated our research efforts into three main categories.

In chapter 3, we apply novel state-of-the-art quantitative mass-spectrometry based tools to reliably and robustly quantify primary host cell proteomic responses to infections with Wild Type (WT) *Mycobacterium tuberculosis* strains in comparison to strains lacking the virulence-associated RD1. In order to gain a better understanding of the dynamic response to the secreted bacterial proteins, we studied the proteomic responses of the host along a 5-day-long infection time course. This enabled us to shed light on the immediate and longer-term immunomodulatory effect of the protein secretion apparatus on the host.
The fourth chapter is dedicated towards identifying the exact *M. tuberculosis* interaction points with the host cells. To accomplish this, we generate human cell lines each expressing an individual

tagged bacterial secreted protein. By applying a method based on affinity-purifications of the bacterial protein, we identify via bottom-up proteomics the direct and indirect host interactors. Although it is unlikely that all identified host-pathogen protein-protein interactors are relevant within the context of an actual *M. tuberculosis* infection, by combining our results with previous knowledge of host factors restricting or enabling the intracellular bacterial replication, this enables us to identify potential human druggable targets that could be exploited as host-directed therapies.

In the fifth chapter, we present a novel method that could unravel in a global and unbiased manner protein-protein interactions that overcome that limitations imposed by current PPI mapping methodologies like affinity-based purifications coupled to mass spectrometry. Although the basic methodologies are already available, there are greatly limited in their feasibility due to lengthy and costly biochemical and computational frameworks. To overcome these bottlenecks, we develop a method that enable deep and quantitative profiling of proteome-wide protein-protein interactions, all achievable in under two days. This method is a comprehensive workflow, that optimizes the sample preparation, the MS acquisition methods and comprises an automated computational framework that enables rapid quantitative comparisons of proteome-wide organisation in more than one condition. To test the method as a proof-of-concept, we apply the method to study the proteome-wide spatial organisation upon the differentiation of a monocytic precursor cell into a macrophage-like phenotype and finally upon a bacterial infection-like stimulation.

Finally, I conclude this thesis in the sixth chapter by summarizing the key learnings about the host manipulation by *Mycobacterium tuberculosis* secreted proteins and the converging conclusions obtained from all three approaches. I then hypothesise that although classical host-pathogen protein-protein interaction methods provide valuable knowledge in discovering the physical interaction points between a host and its pathogen, methods that can identify host-pathogens in the physiologically more relevant context of a real infection will most likely become the ground truth for future host-pathogen interaction studies.

Chapter 3

Whole proteome analysis of WT and ΔRD1 *Mycobacterium tuberculosis* infected macrophages via Data Independent Acquisition – Parallel Accumulation Serial Fragmentation

Contributions:

Johannes Nemeth performed the *Mtb* infections
Ben Collins acquired the samples in the timsTOF pro and performed the SWATH analysis
Charlotte Nicod processed the samples, analyzed the data and exclusively wrote the chapter besides the materials and methods section for the MS acquisition and OpenSWATH analysis taken from:
Meier F, Brunner A-D, Frank M, Ha A, Voytik E, Kaspar-Schoenefeld S, Lubeck M, Raether O, Aebersold R, Collins BC, et al.: **Parallel accumulation – serial fragmentation combined with data-independent acquisition (diaPASEF): Bottom-up proteomics with near optimal ion usage**. bioRxiv 2019, doi:10.1101/656207.

Aimed to be published as

Nicod C, Nemeth J, Aebersold R, Collins BC
Whole proteome analysis of WT and ΔRD1 *Mycobacterium tuberculosis* infected macrophage via Data Independent Acquisition – Parallel Accumulation Serial Fragmentation

3.0 Abstract

Mycobacterium tuberculosis is the causative agent of tuberculosis and infects primarily alveolar macrophages. Virulent *Mtb* strains having an intact RD1 locus encoding the ESX-1 protein secretion system successfully immunomodulate their host macrophages to favor their intracellular survival and replications by interfering with host cellular processes. To investigate the effects caused by this secretion apparatus on the primary host mouse macrophages, we report the longitudinal host proteome responses along 5 days of infections to Wild Type and ΔRD1 *Mtb* strains via Data Independent Acquisition - Parallel Accumulation Serial Fragmentation. We quantify over 5822 proteins and calculate the differential expression profiles associated with the virulence locus RD1. We found that whereas immune-related biological processes and general responses to bacterial invasions are repressed by the RD1 region at the beginning of the infection, the RD1 mostly represses the ion transports and metabolism of *Mtb* infected cells. By overlapping our expression profiles with existing siRNA screens, we identify RD1-mediated differential regulation of host restrictive or permissive factors for the intracellular bacterial survival.

3.1 Introduction

Mycobacterium tuberculosis (*Mtb*) is an airborne pathogen and the causative agent tuberculosis. It is hypothesized to be the deadliest infectious disease in human history [1]. Despite the development of the attenuated vaccine *Mycobacterium bovis* strain Bacille-Calmette-Guérin (BCG) in 1921 and the onset of antibiotics, it still remains a leading killer caused by single-agent infections. It is estimated that roughly a quarter of the global population is currently latently infected with the bacilli, amongst whom 10% develop the active tuberculosis disease including 1% who never recover and succumb to the consequences of the bacterial infection [2]. Given the rise of antibiotic resistance to current treatments, amounting to half a million patients per year and still increasing [2], it is crucial that we gain a better understanding of the underlying molecular mechanisms of MTB virulence. This could potentially enable the identification of novel treatment strategies, including host-directed therapies to fight this persistent global epidemic.

Mtb enters the lungs with the respiratory track to be internalized into alveolar macrophages via receptor-mediated phagocytosis. Once engulfed into early stage phagosomes, *Mtb* prevents its clearance by interfering in the phagocytotic pathway and blocks the fusion between the lysosome and the bacilli-containing phagosomes [3,4]. Yet, this pathogen-driven immunomodulation of its host macrophages and blockage of the phagosomal maturation is a hallmark that holds true only

for virulent strains including the lab strain named H37Rv. Through comparative genomic sequencing and further differential infection studies between the attenuated BCG and Wild Type (WT) strains [5,6], this hallmark of virulence was attributed to the genomic locus named the Region of Difference 1 (RD1) [5,7–9] naturally deleted in BCG. The 9.5 kb RD1 locus comprises 9 genes encoding core components of one out of five paralogous specialized type VII secretion systems, namely ESX-1, and a subset of its secreted substrates including two well characterized virulence factors EsxA and EsxB (also called ESAT-6 and CFP-10). The paramount importance of the RD1 locus is highlighted by a decrease in intracellular bacterial replication in *in vitro* and *in vivo* models upon RD1 deletions [10–12]. In fact, the deletion of the RD1 from MTB not only prevents the spread of the bacterial replication in mice, but even leads to survival [7,13].
The ESX-1 system, through the secretion of its immunogenic antigens EsxA-EsxB dimers, is also necessary to induce the phagosomal rupture and consequent bacterial cytosolic translocation and release of bacterial effector proteins to counter host responses. [10,11,14,15]. The repercussions of this cytosolic translocation mediated by ESX-1 elicits innate immune responses and activates the inflammasome [16,17]. In a later stage, it induces cell death via the necrotic and apoptotic pathways rather than clearance by autophagy altogether facilitating the cell-to-cell spread of the bacilli [18–20],

Given the undisputed essentiality of the RD1 locus in the virulence of *Mycobacterium tuberculosis* infections and its ability to modulate host responses, we aimed here to study the dynamic host proteome remodeling to differential infections using H37Rv WT and H37Rv ΔRD1 strains along 5-day-long *in vitro* infections of primary mouse Bone Marrow Derived Macrophages (mBMDM). By analyzing the global host responses, we aim to identify the differential protein expression profiles caused by the presence or loss of the ΔRD1 locus. To achieve a deep coverage of the host proteome response, the samples were acquired using the recently developed diaPASEF strategy (data independent acquisition – parallel accumulation serial fragmentation) with a timsTOF pro mass spectrometer. Contrary to the semi-stochastic DDA (Data Dependent Acquisition) methods where precursor ions are sequentially selected, DIA approaches enable a robust, sensitive, data-complete and quantitatively accurate coverage of a deeper proposition of the peptides across multiple samples without requiring the usage of quantitative labelling strategies [21,22]. To achieve this, DIA schemes rely on the fragmentation and MS2 acquisition of theoretically all precursor ions present in predefined MS1 windows, which however leads to a low ion sampling efficiency and a highly convoluted fragment ion spectrum. These drawbacks can be circumvented by adding an ion mobility to the separation, leading to a higher coverage and reduced spectral complexity [23]. By combining 100 minute-long chromatographic gradients and the DIA method on the timsTOF pro, for the first time we successfully quantitatively monitored 5,822 of host

proteins at an unprecedented depth [24] along the infection time course in response to either WT or ΔRD1 H37Rv macrophage infections, the primary host cell type of *Mycobacterium tuberculosis.* Together, these results provide a better understanding of the immunomodulation upon infections of primary cells caused by the RD1 region present in virulent *Mtb* strains.

3.2 Results

In order to study the effect of the region of difference on the macrophage host proteomic responses, we performed comparative infections of primary mouse Bone Marrow Derived Macrophages (mBMDM) with WT and ΔRD1 H37Rv *Mycobacterium tuberculosis* strains at a MOI of 10 and extracted the whole proteome of uninfected mBMDM, and after one, three and five days following the infections. In parallel, we reported the cell death rates and Colony Forming Units to monitor the bacterial replication. In order to generate a high-quality spectral library [25], we first pooled a fraction of each sample together and performed high pH fractionation resulting into 8 fractions which we acquired using a long liquid chromatography (LC) and Data Dependent Acquisition scheme.

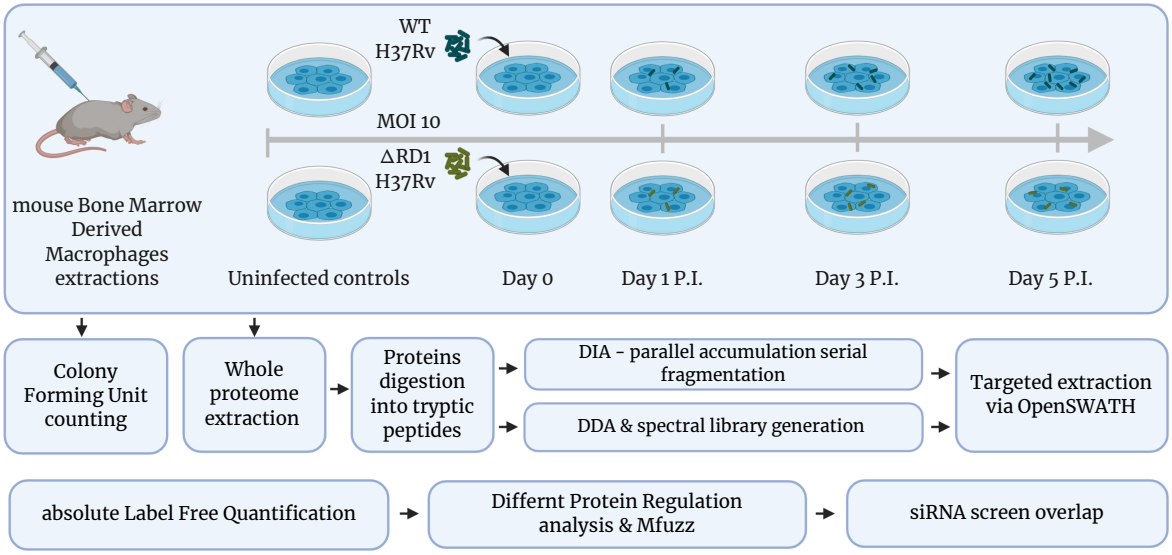


Figure 1 Workflow representation from the sample preparation to the computational analysis. After extracting and treating the mouse bone marrow derived macrophages, the cells where infected at a Multiplicity of Infection of 10 for 4 hours and the extracellular bacteria were washed away. The whole proteomes were extracted from uninfected controls, at 1 day post infection, 3 days post infection and 5 days post infection. The colony forming units were counted to monitor intracellular bacterial growth. The whole proteomes were digested into peptides which were acquired in DDA mode to build a high-quality spectral library and in DIA-PASEF mode. OpenSWATH was used for the targeted extraction. Proteins were quantified from the top 3 peptides using the aLFQ software. The differentially protein regulation profiles were analyzed to detect RD1 mediated host modulation. The Mfuzz fuzzy c-means clustering was used to

identify more subtle patterns. Finally, we overlapped our differential protein regulation profiles with siRNA screens.

The resulting high-quality spectral library generated with an adapted version of OpenSWATH resulted in 75,790 precursor ions amounting to 6,272 proteins of which 6168 were mouse and 91 were *Mtb*. The infection samples were then acquired during 100-minute-long gradients in diaS mode. By employing a targeted extraction using our OpenSWATH version and our newly generated spectral library, we reliably quantified 5822 proteins across all uninfected and infected samples, with 10.2% of missing values across the entire dataset and a technical Coefficient of Variation between technical injections of 0.15. at the peptide level. The missing values were imputed by sampling from a lower intensity background distribution (figure 2). Proteins were quantified by summing up the top 3 proteotypic peptides using the absolute Label Free Quantification (aLFQ) R package.

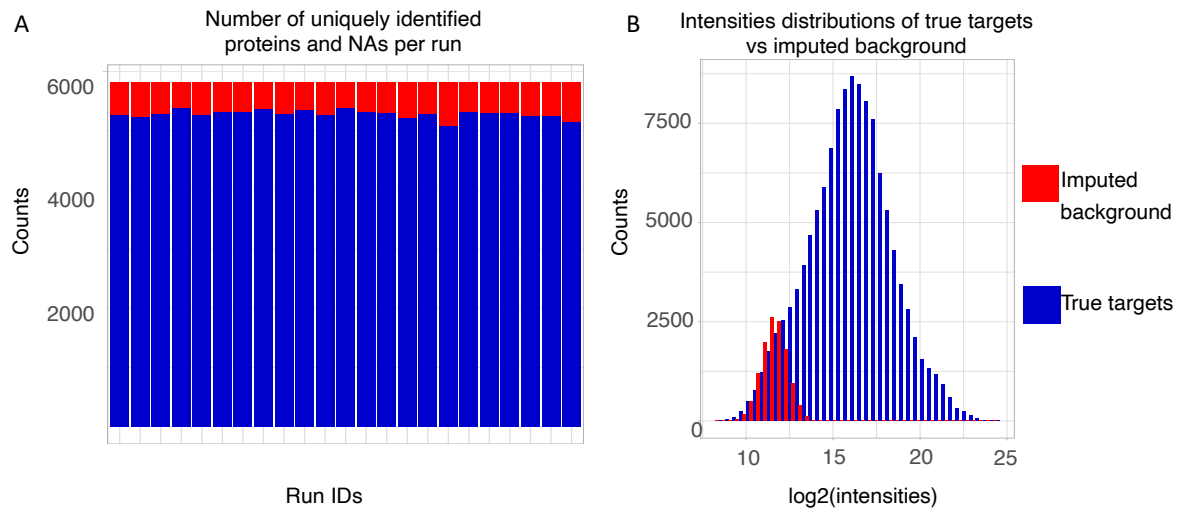


Figure 2 Panel 2A shows the number of uniquely identified proteins and missing values per MS run showing a cumulative number of 5822 across all runs. Panel 2B shows the intensity distribution of true targets and the imputed background.

From the 5822 identified proteins, 73 were annotated as *Mycobacterium tuberculosis* targets, including the ESX-1 secreted substrates EsxA and EsxB. The comparatively low number of MTB proteins identified relative to the host background is due to an asymmetric ratio in the overall protein abundance belonging to the bacteria versus host (figure 3).

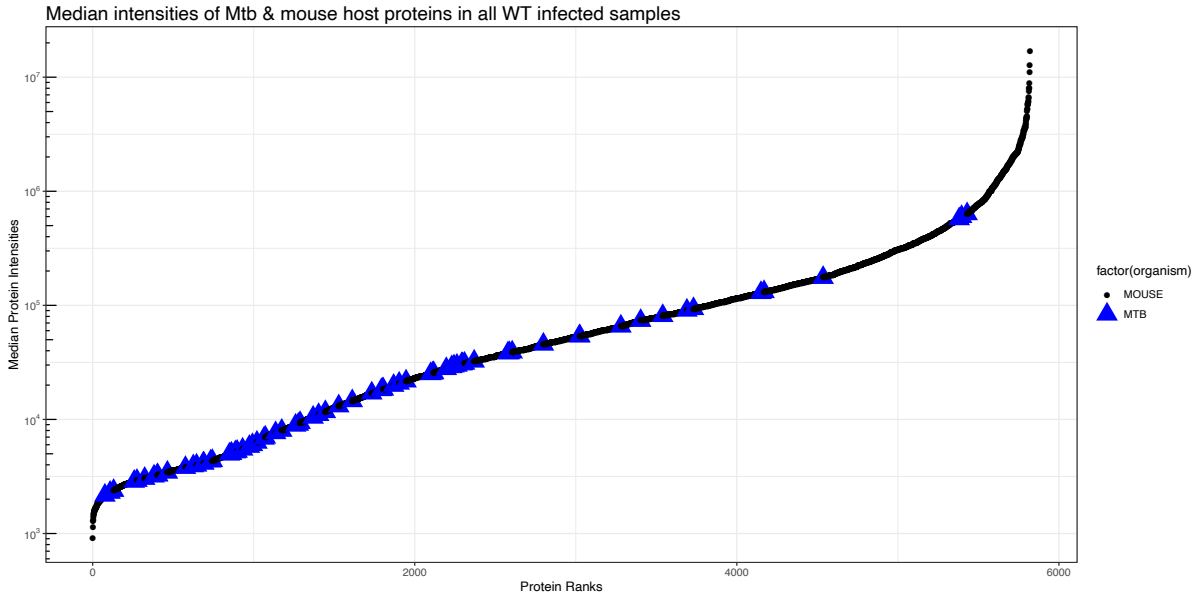


Figure 3 Dynamic range of all detected proteins based on the median intensities from WT infected samples at all time points. The log10 abundance of all detected proteins is plotted. The 73 *Mtb* proteins are highlighted in blue compared to the 5749 quantified mouse host proteins.

When performing a hierarchical clustering on all quantified proteins based on their log2 transformed intensities, as expected control samples cluster together and WT and ΔRD1 infections after 24 hours of infections form 2 sub-clusters. However, for samples infected longer than 24 hours, clusters form according to the strains and not the day of the infection, suggesting that the presence or loss of the RD1 locus has a greater influence on the global host proteomic responses than length of the infection past the first 24 hours.

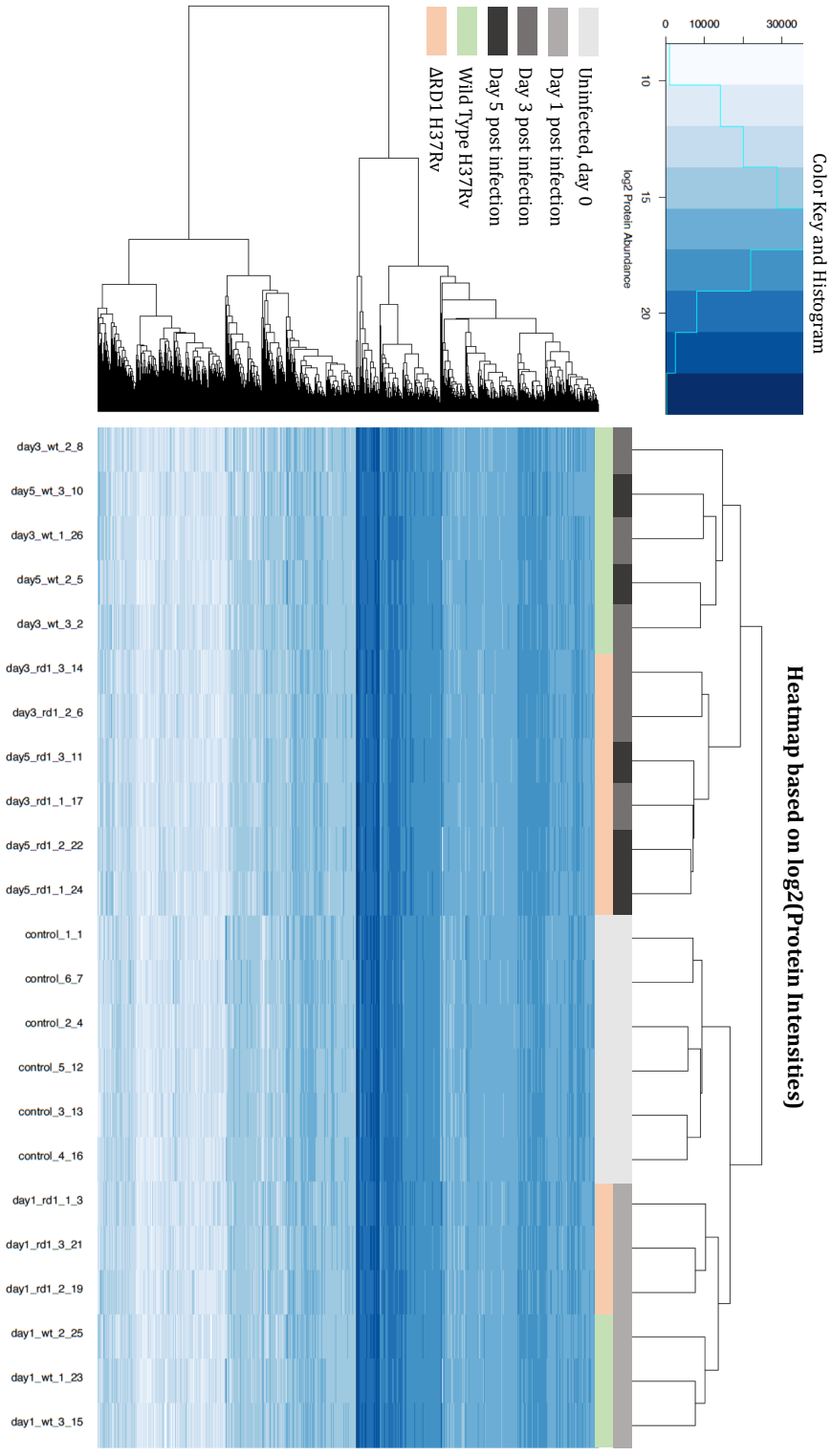


Figure 4 Hierarchical clustering on the 5822 identified proteins based on their log2-transformed intensities showing a greater effect of the type of strain used for the infection than the length of the infection past the first 24 hours.

In order to identify RD1 induced effects of the host proteomes, we calculated the number of Differentially Regulated Proteins (DRP) between not only the WT vs uninfected controls and the

ΔRD1 vs uninfected macrophages, but the Fold Change (FC) of every protein along the infection time course between the ΔRD1 and WT infected samples. This enables the identification and quantification of the RD1 mediated immunomodulation of the host. After establishing a cutoff of an absolute log2 FC greater than 1 with an assigned pvalue < 0.05, the number of DRP after 1, 3 and 5 days of infections were of 154, 143 and 188 respectively between ΔRD1 and WT infections.

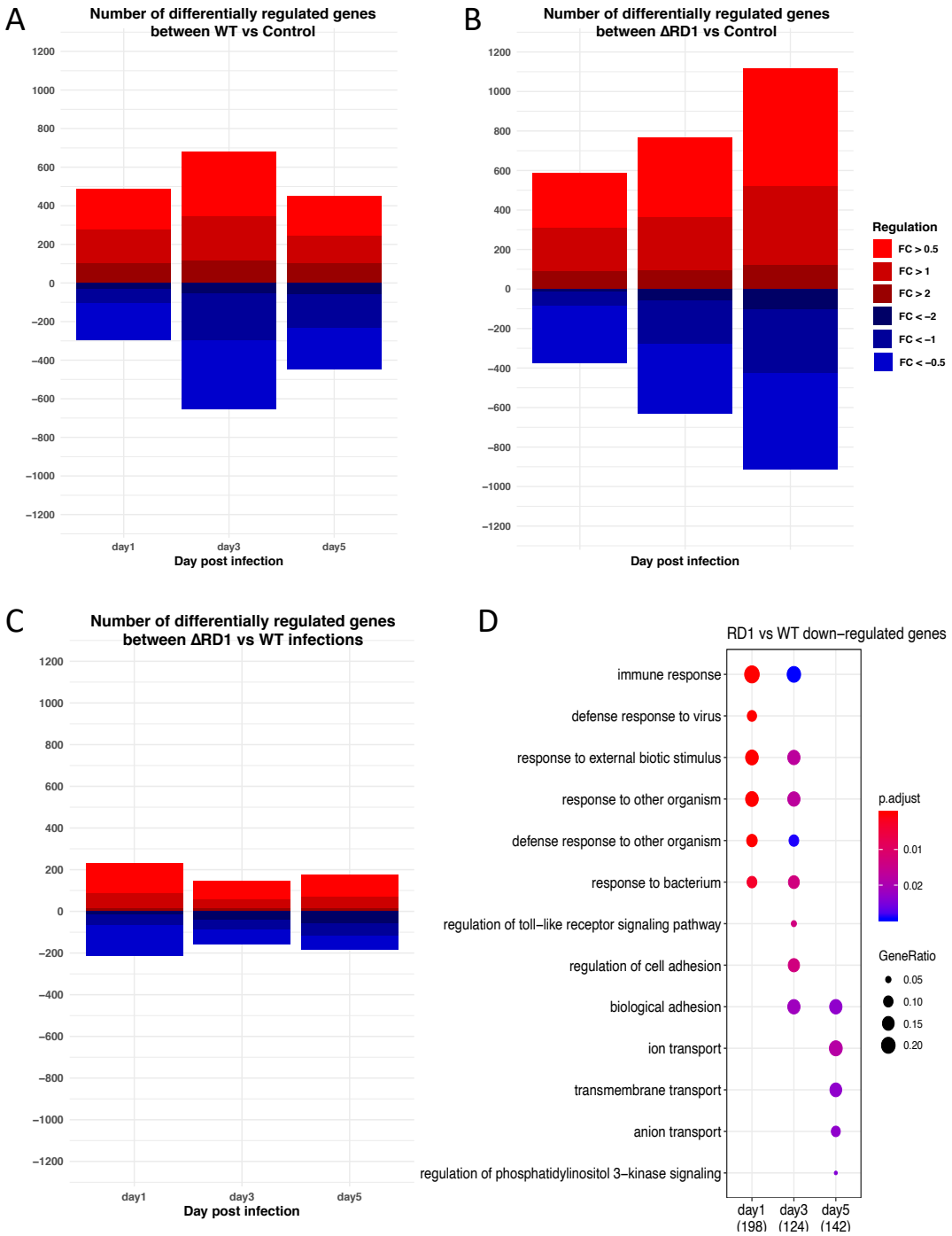


Figure 5 Number of differentially regulated proteins in all 3 comparisons for absolute fold changes of 0.5, 1 and 2 with pvalues < 0.05. Panel A shows the number of differentially regulated proteins along the three infection time points between WT infected macrophages and uninfected control macrophages. Panel B shows the number of differentially regulated proteins when comparing ΔRD1 infections with uninfected

macrophages. Panel C shows the number of up or down regulated proteins caused by the RD1 locus. Up regulated proteins (in red) are proteins higher expressed in the host upon the loss of the RD1 locus. Down-regulated proteins showed in blue are repressed upon the presence of the RD1 locus in WT compared to the ΔRD1 infections. Panel D represents the Gene Ontology Biological Processes enriched in host proteins downregulated by the presence of the RD1 locus (Benjamini pvalue < 0.05). This was calculated from proteins with a log2 FC < -0.5 (pvalue < 0.05) between RD1 and WT infected samples.

When comparing the global host responses between the WT and the ΔRD1 infections, WT strains seem to dampen the number of DRP particularly towards the end of the infection, where the number of up or down DRP in WT infections amounted to 473 compared to the 947 affected host proteins upon ΔRD1 infections (figure 5). This suggested that despite a 5-fold increase (supplementary figures) in the WT bacterial burden at this late-stage infection, WT induced proteomic dysregulation of protein levels is dampened by a factor of 2. In order to identify global biological processes dysregulated by the RD1 virulence associated region, we calculated Gene Ontology enrichments for Biological Processes and KEGG pathways. When applying a more stringent DRP cutoff greater than |2|, 41 MTB proteins and 81 mouse targets had significant altered expression profiles (pvalue <0.05). Although no biological processes were enriched in host cells lacking the RD1 region compared to WT infections, KEGG enrichments analysis showed an up-regulation of oxidative phosphorylation after 24 hours of infection. When performing the same enrichment analysis on host proteins that were repressed by the presence of the RD1 locus, significantly enriched KEGG pathways (Benjamini pval < 0.05) included a repression of cell adhesion molecules, Herpes simplex virus 1 infection, protein digestion and absorption and the complement and coagulation cascades. This confirms the previous findings stating that virulent *Mtb* strains greatly interfere in blood coagulation pathway [26] compared to strains with attenuated virulence. The interference in the blood coagulation pathway upon tuberculosis infections is hypothesized contribute to the formation of the granuloma *in vivo* and in the chemokine and cytokine production via abhorrent fibrin regulation [27–29]. Based on this data, we can hypothesize that the RD1 locus leads to the down-regulation of Complement Receptors 1, 3 and 4 (CR1, CR3 and CR4) mediated phagocytosis despite the 4.9-fold decrease in bacterial burden in ΔRD1 infections compared to WT strains (supplementary materials).

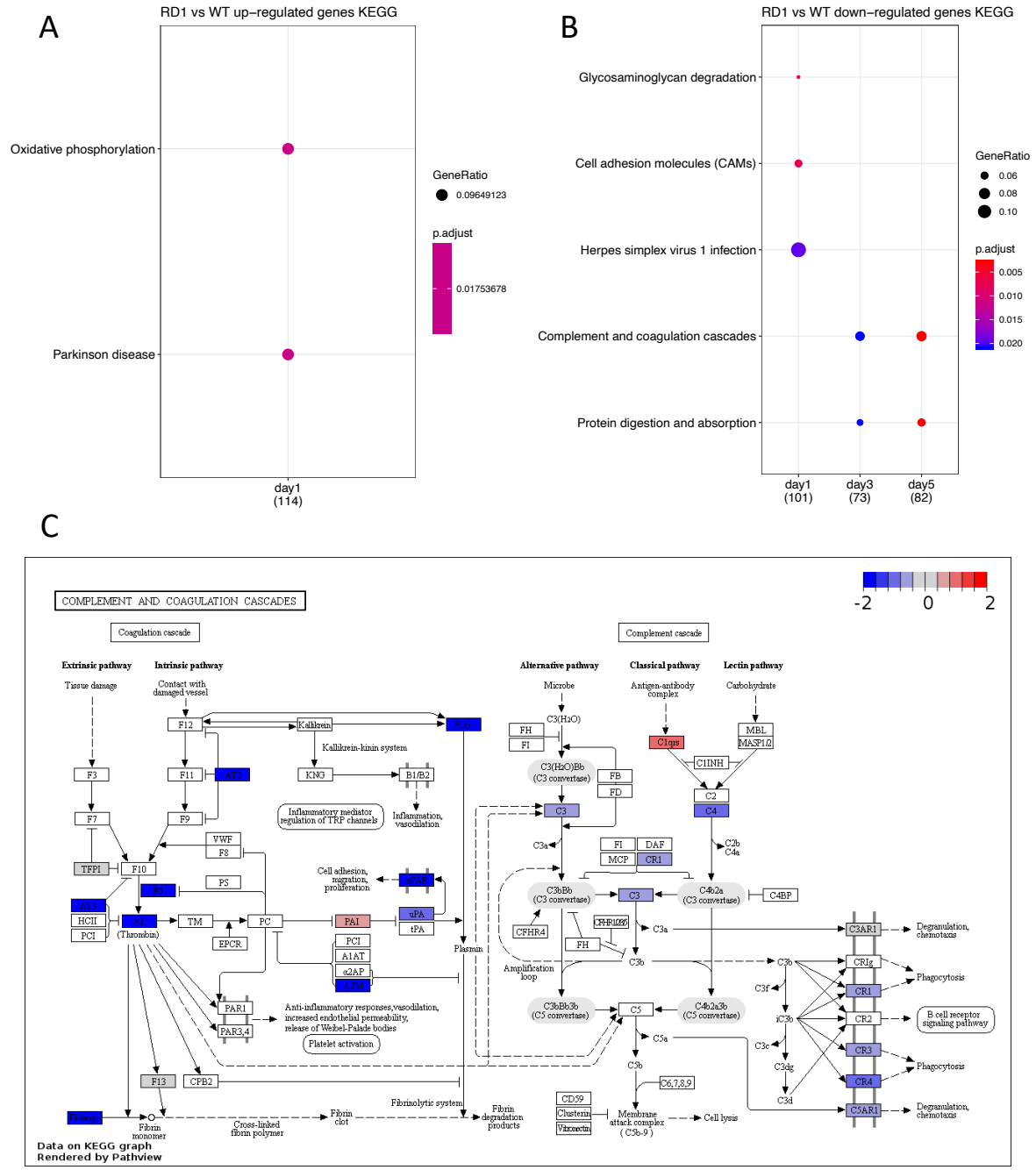


Figure 6 KEGG enrichment analysis on host proteins differentially expressed upon ΔRD1 compared to WT *Mtb* infections. Panel A shows KEGG pathways that are up-regulated upon the loss of the RD1 region. Panel B shows KEGG pathways repressed in the host upon the presence of the RD1 locus in the WT infections, otherwise said the repressive immuno-modulation on the host pathways caused by the RD1 virulence associated locus. Panel C represents the complement and coagulation cascade KEGG pathway with the embedded fold changes between host cells infected with ΔRD1 and WT *Mtb* strains at the fifth day post infection.

While analyzing host targets downregulated upon WT infections in comparison with ΔRD1, many infection-related Biological Processes are significantly repressed. These including immune responses, Toll-like receptor signaling and responses to bacterium during the first days of the

infections despite the fact that WT infections result in higher bacterial burden (supplementary figure). More precisely, it was found that WT represses STAT1 comparatively to ΔRD1 infections, a known Mendelian susceptibility factor for MTB infections if deleted [30]. Together, this corroborates previous finding that virulent MTB strains have immuno-suppressive effects on the host in order to promote its bacterial spread and prevent its clearance. Future experiments could aim at increasing the activity of such host factors to increase their immunity functions and clear the bacterial infection [31].

In order to identify subtle and dynamic patters over the entire course of the infection in the RD1 locus driven immunomodulation of the host, we then performed fuzzy c-means clustering on the macrophage responses from the host protein fold changes between the ΔRD1 and WT infections (figure 7).

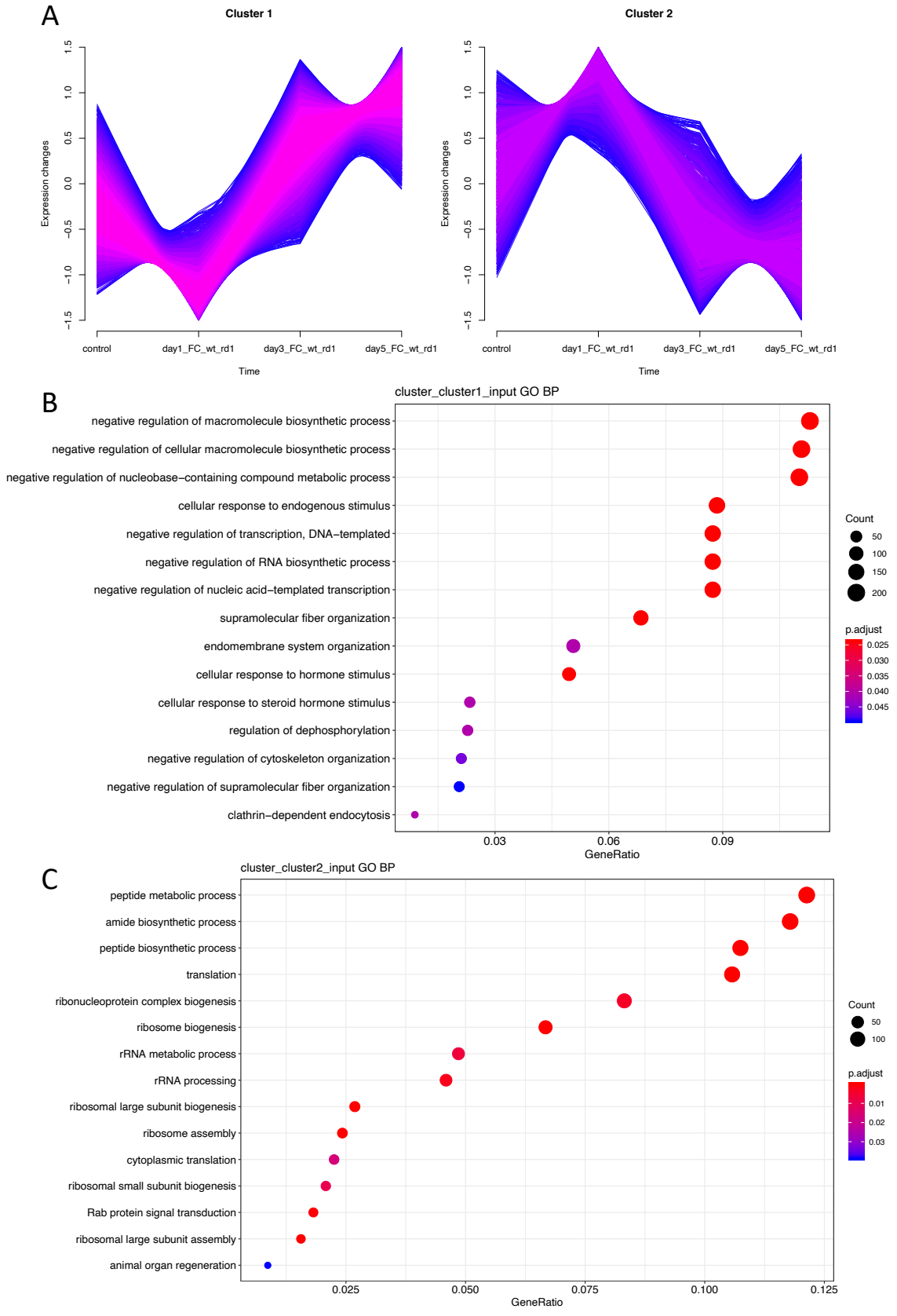


Figure 7 Panel A represent the two resulting clusters from the fuzzy c-means clustering using the Mfuzz

package on the host protein fold changes between ΔRD1 and WT infections. All quantified proteins across the infection time courses were included in the clustering analysis and were not filtered for minimum standard deviation nor fold change thresholds. The clusters were filtered to contain only proteins having a minimum membership values greater than 0.7.

The cluster 1 represents host proteins that are up-regulated in the absence of the RD1 locus, while the second cluster represent a group of host targets that are increasingly down-regulated overtime upon the WT infection respective to the infections lacking the RD1 locus. This fuzzy c-means clustering enabled the detection of much more subtle immunomodulation patterns. One of the findings by analyzing the proteins present in the second clusters was the relative suppression of Rab signaling proteins upon WT infections mostly at 1 day post infection (figure 7C). Out of the 34 Rab proteins quantified in the dataset, 7 were found to be differentially regulated at one infection time point between ΔRD1 and WT infections (supplementary materials). Out of the 7 significantly differentially regulated, 5 of them were repressed by the presence of the RD1 locus after 24 hours of infections, but showed no significant changes over time. These 5 Rab proteins include Rab21, Rab2A, Rab8A, Rab31 and Rab32. Rab32 is known degradation target of a *Salmonella* effector protein, and the silencing of Rab32 leads to an increased increases intracellular survival of the pathogen suggesting that Rab32 is a restriction host factor for the intracellular pathogen. By down-regulating the Rab32 expression in WT *Mtb*, with further experimental validations we could hypothesize that the RD1 locus downregulates this host restriction factor [32]. This confirms that Rab proteins are affected by virulent *Mtb* infections and are subject to RD1-mediated modulation. Surprisingly, both key Rab proteins known to be strongly immunomodulated by the RD1 responsible for the phagosomal maturation arrest [33], namely Rab7 and Rab5, were not significantly affected at the whole proteome level which suggests that they are affected by the RD1 locus via a different mechanism than at the protein intensity levels (will be discussed in chapter 4).

In order to identify strain-specific host regulation that have direct impact on the bacterial replication capacity, a proxy for the bacterial infectivity, we overlaid our differential expression protein data with a genome-wide siRNA screen carried out by J.-P. Carrlot and P.Brodin (unpublished) with the aim to identify host targets that either support or restrict intracellular H37Rv growth after an MOI of 5. This could enable to pin-point RD1 mediated repression of host restriction factors, or conversely up-regulation of supportive host factors for intracellular bacterial replication. Overlaying the siRNA target hits with the whole proteome resulted in significant enrichments of siRNA targets in our differentially regulated set of proteins between both ΔRD1 and WT infections and between WT and uninfected cells (both pval < 0.01) when performing hypergeometric enrichment tests. Results show an overlap of 18 targets (tabled 1)

in our differentially regulated host proteins upon WT infections respective to uninfected cells, and 5 when overlapping the siRNA screen with host DRP between ΔRD1 and WT infections (table 2).

Uniprot ID	Gene symbol	Regulation direction between WT - C	Host factor type
Q6PDI6	Mindy2	down	Restricts intracellular bacterial replication
Q9D0F1	Kntc2	down	Restricts intracellular bacterial replication
Q9R098	Hgfac	down	Restricts intracellular bacterial replication
P13597	Icam1	up	Facilitates intracellular bacterial replication
P49442	Inpp1	up	Facilitates intracellular bacterial replication
P61226	Rap2b	up	Facilitates intracellular bacterial replication
P84096	Rhog	up	Facilitates intracellular bacterial replication
Q60710	Samhd1	up	Facilitates intracellular bacterial replication
Q6P1Y8	Inpp4b	up	Facilitates intracellular bacterial replication
Q8BYI6	Lpcat2	up	Facilitates intracellular bacterial replication
Q8CAS9	Parp9	up	Facilitates intracellular bacterial replication
Q9WTP7	Ak3l1	up	Facilitates intracellular bacterial replication
P06869	Plau	down	Facilitates intracellular bacterial replication
P13864	Dnmt1	down	Facilitates intracellular bacterial replication
Q9JHC9	Elf2	down	Facilitates intracellular bacterial replication
Q9JJY4	Ddx20	down	Facilitates intracellular bacterial replication
Q9Z2L6	Minpp1	down	Facilitates intracellular bacterial replication
Q9JJF9	Sppl2a	up	Facilitates bacterial entry

Table 1 List of proteins overlapped in our differentially expressed data across time and the siRNA target list. The positive or negative regulation of the host proteins upon WT infections compared to uninfected controls indicated in the third column. The host factor type column indicates whether the host factor was supportive or restrictive of the intracellular *Mtb* replication based on the P. Brodin siRNA screen.

This supports the hypothesis that virulent *Mycobacterium tuberculosis* strains modulate host targets favorably for their replication and survival. The overlap shows that *Mtb* down-regulates 3 restrictive host factors and up-regulates 9 supportive host factors for mycobacterial intracellular replication and increases the protein level of a host factor which facilitates bacterial entry. However, WT infections also seem to repress 5 host supportive factors. Amongst the 18 targets identified in the siRNA screen at DRP between the WT and uninfected control cells, ICAM1 was identified to decrease *Mycobacterial replication* and was concurrently up-regulated in the WT infections, consolidating previous findings stating that ICAM1 is a bacterial promoting factor [34]. While examining the immuno-modulatory effects of RD1, 5 host targets repressed by WT strains comparatively to ΔRD1 infections were identified in the overlap: a mitochondrial phosphotransferase AK3, the hepatocyte growth factor activator HGFA, the retinoblastoma-like protein 2 RBL2, the 40S ribosomal protein S15a necessary for erythropoiesis, and the large neutral amino acids transporter small subunit 2 slc7a8 (table 2).

Uniprot ID	Gene symbol	Regulation direction between ΔRD1 - WT	Host factor type
P62245	Rps15a	up	Facilitates bacterial entry
Q64700	Rbl2	up	Facilitates bacterial entry
Q9QXW9	Slc7a8	down	Facilitates bacterial entry
Q9R098	Hgfac	down	Restricts intracellular bacterial replication
Q9WTP7	Ak3l1	up	Facilitates intracellular bacterial replication

Table 2 List of proteins overlapped in our differentially expressed data across time and the siRNA target list. The positive or negative regulation of the host proteins upon ΔRD1 compared to WT infections compared are indicated in the third column. The host factor type column indicates whether the host factor was supportive or restrictive of the intracellular *Mtb* replication based on the P. Brodin siRNA screen.

The HGFA was identified to increase bacterial growth upon *in vitro* infections upon RNA interference, and was correspondingly repressed in WT infections in regards of the ΔRD1 strains. Although the molecular link and explanation remain unexplored, recent studies show that activation of the HGF pathways *in vivo* decreases the bacterial burden and improve the overall survival of mice when challenged with multi-drug resistant tuberculosis strains [35].

3.3 Discussion and conclusions

In this study, we followed the dynamic response of primary host macrophages in response to both virulent H37Rv WT and avirulent H37Rv ΔRD1 strains with the objective of identifying host targets immuno-modulated by the ESX-1 Type VII secretion system. We measured the quantitative dynamic response of 5822 proteins, mostly belonging to the host mouse macrophages. By following the protein responses, we uncovered the biological processes targeted by the ESX-1 system hypothesized to promote bacterial replication and required for the overall virulence. These included repressive modulation of host targets involved in blood coagulation and complement cascade, immune responses, Toll-like receptor signaling and responses to bacterium. Although multiple studies have been conducted to study the comparative response at the transcriptomic levels upon infections with strains expressing or lacking virulence regions [36,37], we report for the first time protein-based immunomodulation caused by the virulence-associated RD1. Our findings indicate that only a partial regulation of known modulation of Rab proteins by virulent *Mtb* strains is done at the protein intensities' levels, suggesting that Rab-modulation is based on other mechanisms such as post-translational modification or spatial sequestering. Our results show that by combining differential protein level analysis and RNA interference screens, we can identify host proteins targeted by the virulence region RD1 that either increase or decrease mycobacterial replication. Combined with additional validation studies, this is could unravel the molecular mechanisms by which *Mycobacterium*

tuberculosis achieves the RD1-mediated virulence by favoring its intracellular bacterial replication.

3.4 Materials and methods

Mouse Bone Marrow Derived Macrophages extraction and preparation

Mouse Bone Marrow Derived Macrophages were extracted from adult (between 8–10 weeks old) male C5B16 mice. After euthanasia, femurs were extracted and excess tissue was removed and the femurs were rinsed with RPMI media supplemented with Penicillin and Streptomycin and stored on ice. Bone marrow cells were flushed from bone shafts into fresh RPMI media supplemented with 1 % penicillin/streptomycin and 10 % FBS, L-glutamine and rh-MCF (cRPMI) at a final concentration of 50ng/ml and then manually separated with 26 5/8-gauge needles. Cells were separated by using a nylon cell strainer, rinsed twice with cRPMI and 5x10^6 cells per 150x15mm plate were plated. After 3 days at 37°C and 5% CO2, rinse the cells with plain warm RPMI media and then add cRPMI. At the 6th day after the initial cell plating, the cells were washed thrice with pre-warmed PBS. After the third wash, 10ml PBS supplemented with 2mM EDTA was added to the plates which were then incubate at 37°C for 10 minutes. The detached cells were collected and pelleted at 300 g for 5 minutes. The media was replaced with fresh cRPMI media and re-plated into 6 well TC-treated plates until they reached 7 after the bone marrow extraction.

Bacterial preparation

Bacterial cultures of H37Rv WT and H37Rv ΔRD1 were grown in 7H9 media supplemented with OADC and TWEEN 80 0.01% and grown on rotors at 37°C and were considered ready once they reached an OD600 at 0.1-0.3.

MTB infection of mBMDM

The mBMDM cells were infected with the calculated amount of MTB required to reach a Multiplicity of Infection of 10. The bacteria were pelleted at 3000 rpm for 6 minutes at room temperature. The supernatants were aspirated and replaced with cRPMI (depleted of Pen/Strep) and the bacterial pellets resuspended at a concentration of 10^7 bacilli per ml. The mBMDM media was removed and replaced with fresh cRPMI (depleted of Pen/Strep) with the resuspended bacteria. The infection plates were incubated at 37°C with 5% CO2 for 2 hours. After the 2-hour infection period, the infected mBMDM were washed thrice with pre-warmed RPMI media, before replacing it with new complete RPMI culture (depleted of Pen/Strep).

Proteome harvesting and digestion

The infected cells were pelleted via centrifugation at 300 rpm and the media was discarded. The cell pellets were lysed in TRizol Reagent and incubated for 5 minutes at room temperature before being supplemented with chloroform in a 3.75:1 ratio and incubated for 3 minutes at room temperature. The protein phase was separated from the RNA and DNA phases by a 15-minute centrifugation at 12'000g. After the removal of the upper and interphase, the phenol-ethanol phase was mixed with isopropanol in a 1:2 ratio and incubated for 10 minutes at room temperature. The proteins were pelleted upon a 10-minute centrifugation at 12'000g for 10 minutes at 4°C. The protein pellet was then washed three times with 0.3M guanidine hydrochloride in 95% ethanol and incubated for 20 minutes at room temperature, where the clean protein pellet was separated via a 5-minute centrifugation at 4°C. Following the last centrifugation step, the pellets were washed with 2 ml pure ethanol while being incubated for 20 minutes at room temperature. The proteins were again pelleted upon a 5-minute centrifugation step at 4°C and the ethanol was discarded prior to drying up the pellets in a SpeedVac. After reaching dryness, the protein pellets were resuspended in lysis buffer (8M Urea, 0.1% RapiGest, 0.1M Ambic) and incubated at room temperature while shaking at 200rpm. Insoluble materials were discarded after a 10-minute centrifugation at 4°C. The proteins were then reduced and alkylated with 0.2M TCEP and 0.4M iodoacetamide respectively for 30 minutes at 37°C and room temperature respectively. The protein pellets were then precipitated upon the addition of ice-cold acetone:ethanol 1:1 solution overnight at -20°C and recovered upon 30 minutes of centrifugation at 15000g at 4°C. The acetone:ethanol solution was then discarded and the protein pellet dried before being resuspended in a lysis buffer composed of 8M Urea, 0.1% RapiGest and 0.1M ammonium bicarbonate by a brief sonication of 3 minutes and an incubation at room temperature for 2 hours. The samples were then diluted with 0.1M ammonium bicarbonate to reach a concentration of Urea of 1.6M. The proteins were digested overnight at 37°C while shaking at 300rpm upon the addition of the proteolytic enzyme Trypsin in a 1:50 enzyme-to-substrate ratio. The digestion was stopped on the next day by adding 50% TFA until reaching a pH below 2 and the detergent present from RapiGest were precipitated

peptides were collected the next day by 15'000g centrifugation for 10 minutes. The peptides were then desalted using C18 columns (The Nest Group, #SEM SS18V) and were resuspended in 5% acetonitrile, 0.1% formic acid with iRT peptides (Biognosys, Ki-3002).

Liquid Chromatography

Nanoflow reversed-phase chromatography was one on an a nanoElute liquid chromatography system (Bruker Daltonics). Peptides were separated within 100 min at a flow rate of 400 nL/min

on a commercially available reversed-phase C18 column with an integrated CaptiveSpray Emitter (25 cm x 75µm, 1.6 µm, IonOpticks, Australia). Mobile phases A and B were composed of 0.1% formic acid in water and 0.1% formic acid in CAN respectively. The fraction of B was linearly increased from 2 to 25% within 90 min, followed by an increase to 35% within 10 min and a further increase to 80% before re-equilibration.

DDA acquisition on the timsTOF pro for generating a high-quality spectral library

LC was coupled online to a hybrid TIMS quadrupole time-of-flight mass spectrometer (Bruker timsTOF Pro) via a CaptiveSpray nano-electrospray ion source. The dual TIMS analyzer was operated at a fixed duty cycle close to 100% using equal accumulation and ramp times of 100 ms each. We performed data dependent data acquisition in PASEF mode with 10 PASEF scans per top10 acquisition cycle. Singly charged precursors were excluded by their position in the m/z-ion mobility plane and precursors that reached a 'target value' of 20,000 a.u. were dynamically excluded for 0.4 min. The quadrupole isolation width was set to 2 Th for $m/z < 700$ and 3 Th for $m/z > 700$.

DIA acquisition on the timsTOF pro

To perform data-independent acquisition, we extended the instrument control software to define quadrupole isolation windows as a function of the TIMS scan time (diaPASEF). The instrument control electronics were modified to allow seamless and synchronous ramping of all applied voltages.

Spectral library generation

To generate spectral libraries for targeted data extraction, we first analyzed the high pH reversed-phase fraction acquired in DDA mode with MaxQuant version 1.6.5.0, which extracts four-dimensional features on the MS1 level (retention time, m/z, ion mobility and intensity) and links them to peptide spectrum matches. The maximum precursor mass tolerance of the main search was set to 20 ppm and deisotoping of fragment ions was deactivated. Other than that, we used the default 'TIMS-DDA' parameters. MS/MS spectra were matched against an in silico digest of the Swiss-Prot reference proteome (human 20,414 entries, yeast 6,721 entries, downloaded July 2019) and a list of common contaminants. The minimum peptide length was set to 7 amino acids and the peptide mass was limited to 4,600 Da. Carbamidomethylation of cysteine residues was defined as a fixed modification, methionine oxidation and acetylation of protein N-termini were

defined as variable modifications. The false discovery rate was controlled 400 were included in the library and required to have a minimum of 6 fragment ions with m/z >350 and outside the precursor mass isolation range.

Spectral library generation

To analyze diaPASEF data, we developed an ion mobility DIA analysis kit (Mobi-DIK) that extracts fragment ion traces from the fourdimensional data space as detailed in the main text. Raw data were automatically recalibrated using curated reference values in m/z, retention time and ion mobility dimensions (387 peptides for linear and 3,184 peptides for non-linear alignment). We applied an outlier detection in each dimension before calculating the final fit function to increase robustness. Peak picking and sub-sequent scoring functionalities in the Mobi-DIK software build on OpenSWATH modules. For diaPASEF, we extended these modules to also take into account the additional ion mobility dimension. OpenSWATH was run with following parameters: min_coverage = 0.1, RTNormalization:alignmentMethod = lowess, RTNormalization:lowess:span = 0.01, Scoring:TransitionGroupPicker:PeakPicker MRM:sgolay_frame_length = 11, Scoring:stop_report_after_feature = 5, rt_extraction_window = 250, Scoring:Scores:use_ion_mobility_scores, mz_correction_function = quadratic_regression_delta_ppm, use_ms1_traces, mz_extraction_window = 25, mz_extraction_window_unit = ppm, mz_extraction_window_ms1 = 25, mz_extraction_window_ms1_unit = ppm, irt_mz_extraction_window_unit = ppm, irt_mz_extraction_window = 40, Calibration:ms1_im_calibration, ion_mobility_window = 0.06, irt_im_extraction_window = 99, RTNormalization:NrRTBins = 8, RTNormalization:MinBinsFilled = 4. All other parameters were set to default. PyProphet was used to train an XGBoost classifier for target-decoy separation by first creating one concatenated and sub-sampled OpenSwath output for each set of three replicate injections of the same acquisition strategy and sample amount. The classifier was subsequently applied for scoring all samples, controlling the FDRs

3.5 References

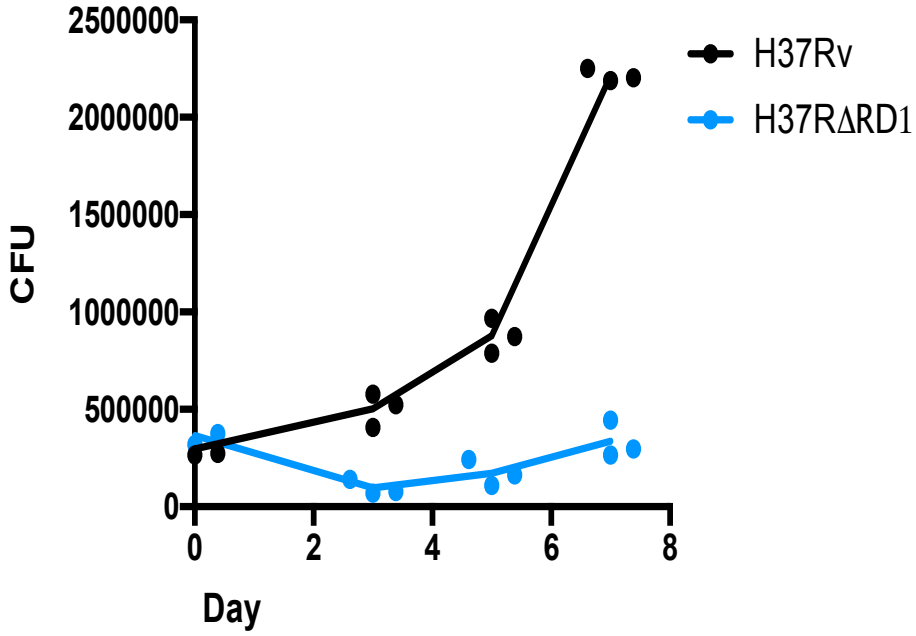
1. Paulson T: **A mortal foe**. *Nature* 2013, **502**.
2. *Status of the TB epidemic*. 2019.

3. Kaufmann SHE: **How can immunology contribute to the control of tuberculosis? [Internet]**. *Nat. Rev. Immunol.* 2001, **1**:20–30.
4. Armstrong JA, Hart PD: **Phagosome-lysosome interactions in cultured macrophages infected with virulent tubercle bacilli. Reversal of the usual nonfusion pattern and observations on bacterial survival. [Internet]**. *J. Exp. Med.* 1975, **142**:1–16.
5. Mahairas GG, Sabo PJ, Hickey MJ, Singh DC, Stover CK: **Molecular analysis of genetic differences between Mycobacterium bovis BCG and virulent M. bovis. [Internet]**. *J. Bacteriol.* 1996, **178**:1274–82.
6. Behr MA, Wilson MA, Gill WP, Salamon H, Schoolnik GK, Rane S, Small PM: **Comparative genomics of BCG vaccines by whole-genome DNA microarray**. *Science (80-.).* 1999, **284**:1520–1523.
7. Hsu T, Hingley-Wilson SM, Chen B, Chen M, Dai AZ, Morin PM, Marks CB, Padiyar J, Goulding C, Gingery M, et al.: **The primary mechanism of attenuation of bacillus Calmette-Guérin is a loss of secreted lytic function required for invasion of lung interstitial tissue**. *Proc. Natl. Acad. Sci. U. S. A.* 2003, **100**:12420–12425.
8. Lewis KN, Liao R, Guinn KM, Hickey MJ, Smith S, Behr MA, Sherman DR: **Deletion of RD1 from Mycobacterium tuberculosis Mimics Bacille Calmette-Guérin Attenuation** . *J. Infect. Dis.* 2003, **187**:117–123.
9. Pym AS, Brodin P, Brosch R, Huerre M, Cole ST: **Loss of RD1 contributed to the attenuation of the live tuberculosis vaccines Mycobacterium bovis BCG and Mycobacterium microti [Internet]**. *Mol. Microbiol.* 2002, **46**:709–717.
10. Stanley SA, Raghavan S, Hwang WW, Cox JS: **Acute infection and macrophage subversion by Mycobacterium tuberculosis require a specialized secretion system**. *Proc. Natl. Acad. Sci. U. S. A.* 2003, **100**:13001–13006.
11. Guinn KM, Hickey MJ, Mathur SK, Zakel KL, Grotzke JE, Lewinsohn DM, Smith S, Sherman DR: **Individual RD1 -region genes are required for export of ESAT-6/CFP-10 and for virulence of Mycobacterium tuberculosis**. *Mol. Microbiol.* 2004, **51**:359–370.
12. Abdallah AM, Verboom T, Hannes F, Safi M, Strong M, Eisenberg D, Musters RJP, Vandenbroucke-Grauls CMJE, Appelmelk BJ, Luirink J, et al.: **A specific secretion system mediates PPE41 transport in pathogenic mycobacteria [Internet]**. *Mol. Microbiol.* 2006, **62**:667–679.
13. Lewis KN, Liao R, Guinn KM, Hickey MJ, Smith S, Behr MA, Sherman DR: **Deletion of RD1 from Mycobacterium tuberculosis mimics bacille Calmette-Guérin attenuation. [Internet]**. *J. Infect. Dis.* 2003, **187**:117–23.
14. Hsu T, Hingley-Wilson SM, Chen B, Chen M, Dai AZ, Morin PM, Marks CB, Padiyar J, Goulding C, Gingery M, et al.: **The primary mechanism of attenuation of bacillus Calmette–Guérin is a loss of secreted lytic function required for invasion of lung interstitial tissue**. *Proc. Natl. Acad. Sci.* 2003, **100**.
15. van der Wel N, Hava D, Houben D, Fluitsma D, van Zon M, Pierson J, Brenner M, Peters PJ: **M. tuberculosis and M. leprae Translocate from the Phagolysosome to the Cytosol in Myeloid Cells [Internet]**. *Cell* 2007, **129**:1287–1298.
16. Koo IC, Wang C, Raghavan S, Morisaki JH, Cox JS, Brown EJ: **ESX-1-dependent cytolysis in lysosome secretion and inflammasome activation during mycobacterial infection [Internet]**. *Cell. Microbiol.* 2008, **10**:1866–1878.
17. Carlsson F, Kim J, Dumitru C, Barck KH, Carano RAD, Sun M, Diehl L, Brown EJ: **Host-detrimental role of Esx-1-mediated inflammasome activation in Mycobacterial infection [Internet]**. *PLoS Pathog.* 2010, **6**:1–12.
18. Aguilo JI, Alonso H, Uranga S, Marinova D, Arbués A, de Martino A, Anel A, Monzon M, Badiola J, Pardo J, et al.: **ESX-1-induced apoptosis is involved in cell-to-cell spread of Mycobacterium tuberculosis**. *Cell. Microbiol.* 2013, **15**:1994–2005.
19. Simeone R, Bobard A, Lippmann J, Bitter W, Majlessi L, Brosch R, Enninga J: **Phagosomal rupture by Mycobacterium tuberculosis results in toxicity and host cell death [Internet]**. *PLoS Pathog.* 2012, **8**:e1002507.
20. Kaku T, Kawamura I, Uchiyama R, Kurenuma T, Mitsuyama M: **RD1 region in**

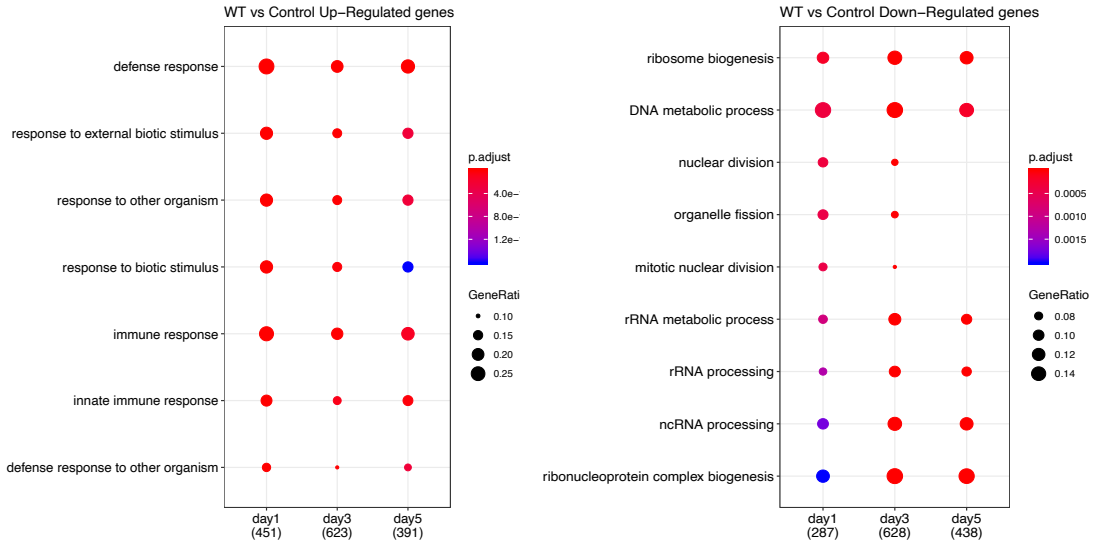
mycobacterial genome is involved in the induction of necrosis in infected RAW264 cells via mitochondrial membrane damage and ATP depletion. *FEMS Microbiol. Lett.* 2007, **274**:189–195.
21. Gillet LC, Navarro P, Tate S, Röst H, Selevsek N, Reiter L, Bonner R, Aebersold R: **Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. [Internet]**. *Mol. Cell. Proteomics* 2012, **11**:O111.016717.
22. Collins BC, Hunter CL, Liu Y, Schilling B, Rosenberger G, Bader SL, Chan DW, Gibson BW, Gingras A-C, Held JM, et al.: **Multi-laboratory assessment of reproducibility, qualitative and quantitative performance of SWATH-mass spectrometry [Internet]**. *Nat. Commun.* 2017, **8**:291.
23. Meier F, Brunner A-D, Frank M, Ha A, Voytik E, Kaspar-Schoenefeld S, Lubeck M, Raether O, Aebersold R, Collins BC, et al.: **Parallel accumulation – serial fragmentation combined with data-independent acquisition (diaPASEF): Bottom-up proteomics with near optimal ion usage [Internet]**. *bioRxiv* 2019, doi:10.1101/656207.
24. Li P, Wang R, Dong W, Hu L, Zong B, Zhang Y, Wang X, Guo A, Zhang A, Xiang Y, et al.: **Comparative proteomics analysis of human macrophages infected with virulent Mycobacterium bovis**. *Front. Cell. Infect. Microbiol.* 2017, **7**.
25. Schubert OT, Gillet LC, Collins BC, Navarro P, Rosenberger G, Wolski WE, Lam H, Amodei D, Mallick P, Maclean B, et al.: **Building high-quality assay libraries for targeted analysis of SWATH MS data [Internet]**. *Nat. Protoc.* 2015, **10**:426–441.
26. Li H, Wei S, Fang Y, Li M, Li X, Li Z, Zhang J, Zhu G, Li C, Bi L, et al.: **Quantitative proteomic analysis of host responses triggered by Mycobacterium tuberculosis infection in human macrophage cells**. *Acta Biochim. Biophys. Sin. (Shanghai).* 2017, **49**:835–844.
27. Szaba FM, Smiley ST: **Roles for thrombin and fibrin(ogen) in cytokine/chemokine production and macrophage adhesion in vivo**. *Blood* 2002, **99**:1053–1059.
28. Izaki S, Goldstein SM, Fukuyama K, Epstein WL: **Fibrin deposition and clearance in chronic granulomatous inflammation: Correlation with T-cell function and proteinase inhibitor activity in tissue**. *J. Invest. Dermatol.* 1979, **73**:561–565.
29. Kothari H, Rao LVM, Vankayalapati R, Pendurthi UR: **Mycobacterium tuberculosis Infection and Tissue Factor Expression in Macrophages**. *PLoS One* 2012, **7**.
30. Mogensen TH: **IRF and STAT transcription factors - From basic biology to roles in infection, protective immunity, and primary immunodeficiencies**. *Front. Immunol.* 2019, **10**:3047.
31. Miklossy G, Hilliard TS, Turkson J: **Therapeutic modulators of STAT signalling for human diseases**. *Nat. Rev. Drug Discov.* 2013, **12**:611–629.
32. Spanò S, Galán JE: **A Rab32-dependent pathway contributes to Salmonella typhi host restriction [Internet]**. *Science (80-.).* 2012, **338**:960–963.
33. Via LE, Deretic D, Ulmer RJ, Hibler NS, Huber LA, Deretic V: **Arrest of mycobacterial phagosome maturation is caused by a block in vesicle fusion between stages controlled by rab5 and rab7**. *J. Biol. Chem.* 1997, **272**:13326–13331.
34. Bhalla K, Chugh M, Mehrotra S, Rathore S, Tousif S, Dwivedi VP, Prakash P, Samuchiwal SK, Kumar S, Singh DK, et al.: **Host ICAMs play a role in cell invasion by Mycobacterium tuberculosis and Plasmodium falciparum**. *Nat. Commun.* 2015, **6**:1–13.
35. Bello-Monroy O, Mata-Espinosa D, Enríquez-Cortina C, Souza V, Miranda RU, Bucio L, Barrios-Payán J, Marquina-Castillo B, Rodríguez-Ochoa I, Rosales P, et al.: **Hepatocyte growth factor enhances the clearance of a multidrug-resistant *Mycobacterium tuberculosis* strain by high doses of conventional chemotherapy, preserving liver function [Internet]**. *J. Cell. Physiol.* 2020, **235**:1637–1648.
36. Etna MP, Giacomini E, Pardini M, Severa M, Bottai D, Cruciani M, Rizzo F, Calogero R, Brosch R, Coccia EM: **Impact of Mycobacterium tuberculosis RD1-locus on human primary dendritic cell immune functions [Internet]**. *Sci. Rep.* 2015, **5**:17078.

37. Lee J, Lee SG, Kim KK, Lim YJ, Choi JA, Cho SN, Park C, Song CH: **Characterisation of genes differentially expressed in macrophages by virulent and attenuated Mycobacterium tuberculosis through RNA-Seq analysis**. *Sci. Rep.* 2019, **9**:1–9.

3.6 Supplementary Information

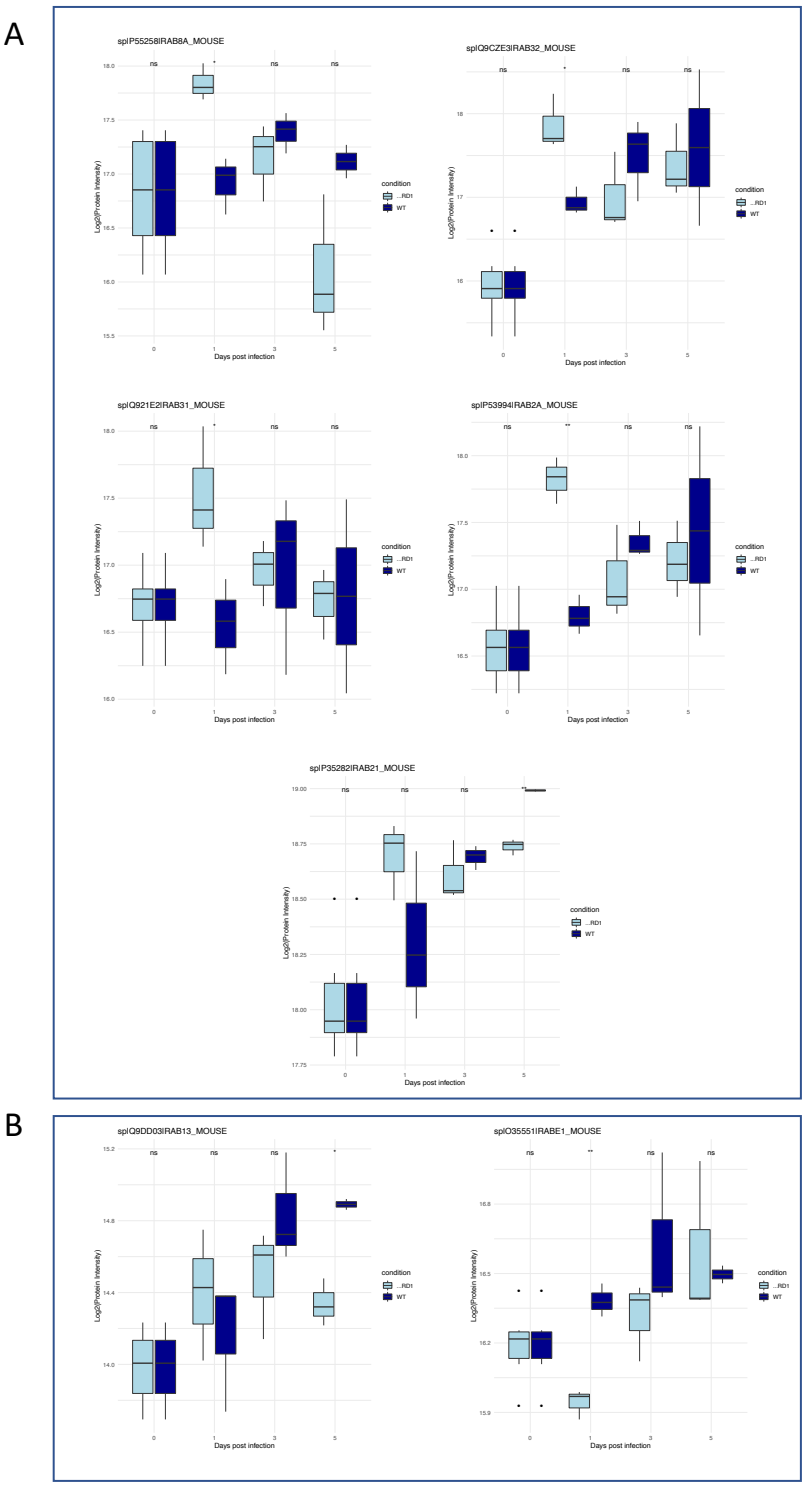


Supplementary Figure 1 Colony Forming Unit counts after 4 hours of infection, after 3, 5 and 7 days of infections using WT H37Rv or ΔRD1 H37Rv *Mtb* strains in mouse Bone Marrow Derived Macrophages.



Supplementary Figure 2 Gene Ontology Biological Processes (Benjamini pvalue < 0.05) enrichment on mouse host proteins upon WT infection at days 1, 3 and 5 post infections based on proteins having an absolute fold change greater than 0.5 with a pvalue < 0.05. the left panels represents GO BP enriched in host proteins that are up regulated upon WT infection in comparison to uninfected control macrophages.

Oppositely, the right panel shows GO BP enriched in host proteins that are down-regulated upon the WT *Mtb* infections.



Supplementary Figure 3 Panel A shows all Rab proteins are significantly repressed (* = pvalue < 0.05, ** = pvalues < 0.01) at 24 hours post infection upon WT infections in comparison to ΔRD1 infections. Panel B shows Rab proteins that are up-regulated upon WT infections in comparison to ΔRD1 infections. For both panels, the X axis represents the infection time points and the Y axis represents the Log2 protein intensities.

Chapter 4

An *Mtb*-Human Protein-Protein Interaction map to identify host proteins targeted by secreted bacterial virulence-associated proteins

Contributions

Charlotte Nicod generated most HEK293 cell lines, prepared the samples and acquired the MS runs, performed the analysis and wrote this chapter
Audrey van Drogen contributed greatly to the generation of the stable cell-lines
Jun Cai contributed to processing some of the AP-MS samples
Shruti Jha generated the THP-1 cell lines
Ben Collins provided critical input and guidance in the project

Aimed to be partially published as
Nicod C, van Drogen A, Cai J, Jha S, Mehnert M, Collins BC

An *Mtb*-Human Protein-Protein Interaction map to identify host proteins targeted by secreted bacterial virulence-associated proteins

4.0 Abstract

Mycobacterium tuberculosis is the causative agent of tuberculosis and remains a leading killer caused by single-agent infections. Its ability to modulate and interfere with the host immunological processes enables its intracellular survival and replication. This host modulation is mainly attributed to its protein secretion systems which provide means to physically interact with host cellular modules and intervene in host pathways. Although the phenotypic loss of virulence upon the deletions of the protein secretion apparatuses is well characterized, the exact human host proteins targeted by the bacterial proteins remain largely unknown. Here, we employ the sensitive and robust approach of affinity-purification coupled to mass spectrometry to map the exact interaction protein-protein interactions between secreted bacterial proteins and their human host targets. From our results, we hypothesize that *Mtb* physically interacts with Rab proteins in order prevent the phagosomal maturation machinery. We further overlap our results with siRNA screens dedicated to identify host factors that are restrictive or supportive of intracellular bacterial replication in order to identify direct host-pathogen interactors that have a direct impact on intracellular bacterial growth.

4.1 Introduction

Mycobacterium tuberculosis (Mtb) is an airborne pathogen and the causative agent of tuberculosis (TB) disease. It is hypothesized to be the deadliest infectious disease in human history with an estimated number of over one billion human causalities in the past 200 years[1]. Despite the advances in antibiotics and the discovery of the only vaccine strain of *Mycobacterium bovis* named Bacille Calmette-Guérin (BCG) a hundred years ago, albeit proven to have a limited efficacy in preventing pulmonary TB, it remains a leading killer caused by single-agent infections [2]. The conventional treatment strategy to clear tuberculosis infections consists of an intensive 2 month-long phase of four antibiotics (namely isoniazid INH, rifampin RIF, pyrazinamide PZA and ethambutol EMB) followed by a 4-months continuation phase with INH and RIF [3]. Although mostly effective, the aforementioned treatment strategy is evidently lengthy and costly, rendering it a logistical challenge and reduces patient compliance. Additionally, it may lead to severe adverse effects and irreversible toxicity in patients including hepatotoxicity [4,5]. More importantly, antibiotic-based treatments are the root cause of an alarming rise of multi-drug and extended-drug resistance tuberculosis strains [6,7] (MDR and XDR respectively), for which they are no longer effective and are consequently linked to poor patient outcomes [8]. It is thus crucial that we aim at developing novel treatments strategies that would either complement or replace bacilli-directed drugs, together improving patient outcomes including for those suffering from

MDR- and XDR-TB. The novel treatment strategies should ideally aspire to circumvent drug resistance issues, shorten current treatment length and have comparatively reduced cytotoxic effects and should inevitably avoid the development of bacterial resistance.

Host-Directed Therapies (HDT), by modifying specific host immune pathways towards the host's advantage, represent viable candidates in these regards. HDT aim to intervene with host factors that facilitate the bacterial survival or replication, to improve protective immunity or balance immune activities at the site of the infection [9]. HDT can be separated theoretically into two concepts: either by promoting host restriction factors to bacterial growth, or restricting host factors facilitating the bacterial replication. Because it is well established that *Mycobacterium tuberculosis* successfully immunomodulates its host cells, HDT therapies could prevent this bacterial-mediated host immunomodulatory effect by preventing the causative host-pathogen interactions [9].

Mtb immunomodulates its host cells, primarily alveolar macrophages, to favor its intracellular survival, immune evasion and replication. For instance, *Mtb* manages to modulate the apoptotic and autophagy pathways [10], to intervene in the MHC class II presentation [11], to impair dendritic cell maturation [12], to block the phagosomal maturation [13] and even escapes from bacilli-containing phagosomes into the cytosol [14,15] and modulates T cell mediated immunity [16]. To achieve most of these virulence hallmarks *Mtb* secretes, via different protein secretion apparatuses, substrates which will interact with and interfere with host cellular modules, a phenomenon also known as pathogenic hijacking mostly associated with viruses. Specifically, *Mtb* encodes a functional general secretion Sec pathway, an accessory SecA2 export system, a twin-arginine translocation Tat pathway, and five homologous Type VII Secretion Systems (T7SS), namely ESX-1 through 5. Whereas the Tat and Sec protein secretion pathways are necessary to transport proteins across the cytoplasmic membranes and are highly conserved across gram positive and negative bacteria [17], the ESX systems seem to be functionally involved in virulence. Indeed, the ESX-1, -3 and-5 systems are necessary for full virulence [18,19] although ESX-2 and ESX-4 have yet to be proven as functional secretion systems and as essential. The ESX-1 T7SS was first discovered through genomic studies comparing the attenuated vaccine strain BCG to the virulent lab strain H37Rv. The deletion of the ESX-1 leads to a decrease in intracellular bacterial replication in *in vitro* and *in vivo* infections [20–22]. It was further demonstrated that the blockage of the phagosomal maturation is achieved only upon successful secretion of their substrates EsxA/EsxB [23,24]. Whereas the ESX-1 system is necessary for the phagosomal escape and cytosolic translocation, the ESX-3 has been demonstrated to play two distinct roles. First, it has been described to be involved in metal homeostasis and iron acquisition via a mechanism dependent on the secretion of its substrates Pe5/Ppe4 [25–27]. Secondly, it mediates its virulence by secreting its heterodimer EsxG/EsxH substrates which consequently impairs the host

phagosomal maturation by interacting with the host Endosomal Sorting Complex Required for Transport ESCRT [28]. Diversely, ESX-5 is predicted to secrete EsxN [29] and most of PE/PPE proteins, a class of proteins representing nearly 10% of the *M. tuberculosis* genome amounting to 167 putative secreted substrates [30] and which has been characterized to strongly affect host immune responses. In the case of *Mycobacterium marinium,* ESX-5 downregulates the production of proinflammatory cytokines including TNF-alpha and IL-6 [31] and further represses TLR signaling dependent innate immune cytokine secretion whilst increasing the production of IL-1ß [31].

Although the phenotypic effects of these protein secretion systems and their importance in the virulence for *Mtb* are no longer neglected, their molecular mechanisms remain unclear and most host targets are yet to be discovered. By mapping these specific host-pathogen protein-protein interactions, it could lead to the discovery of potential human druggable targets, via HDT, to counter the immunomodulatory effects of *Mycobacterium tuberculosis,* altogether preventing the hijacking of the host cellular modules for its survival and replication.
To build the *Mtb*-Human Protein-Protein Interaction map, we employed a strategy that has already been successfully applied to study other host-pathogen interactions in particular for viral pathogens [32–35]. The strategy is called Affinity Purification coupled to Mass Spectrometry (AP-MS). This AP-MS approach consists of expressing predicted secreted bacterial proteins fused to an affinity tag (called baits) in the host cellular environment, and to identify their host interacting proteins (called preys) via Affinity Purification coupled to Mass Spectrometry (Figure 1A). Such studies mapping host-pathogen protein-protein interactions landscapes have been successful as they have not only shed light on the underlying molecular mechanisms of virulence, but also provided the identification of potential human druggable targets when no other therapeutics were available[36].

Recently, this large-scale AP-MS approach was further employed to identify PPI between *Chlamydia trachomatis* Inc secreted proteins and the host to decode its molecular mechanisms to sustain intracellular survival an replication [37]. While conducting the experiments for this study to map the *Mtb*-Human Protein-Protein Interaction map, a similar network was published so we leveraged the available information to study the overlap in findings and draw common conclusions [38]. Here, we will present our interaction network from 28 selected bait bacterial proteins with their human host prey proteins. We will further overlap our findings with the previously published manuscript by Penn *et al* [38], as well as with two siRNA screens to identify host targets that have been confirmed to be either restrictive or beneficial factors for the intracellular replication of *Mtb.*

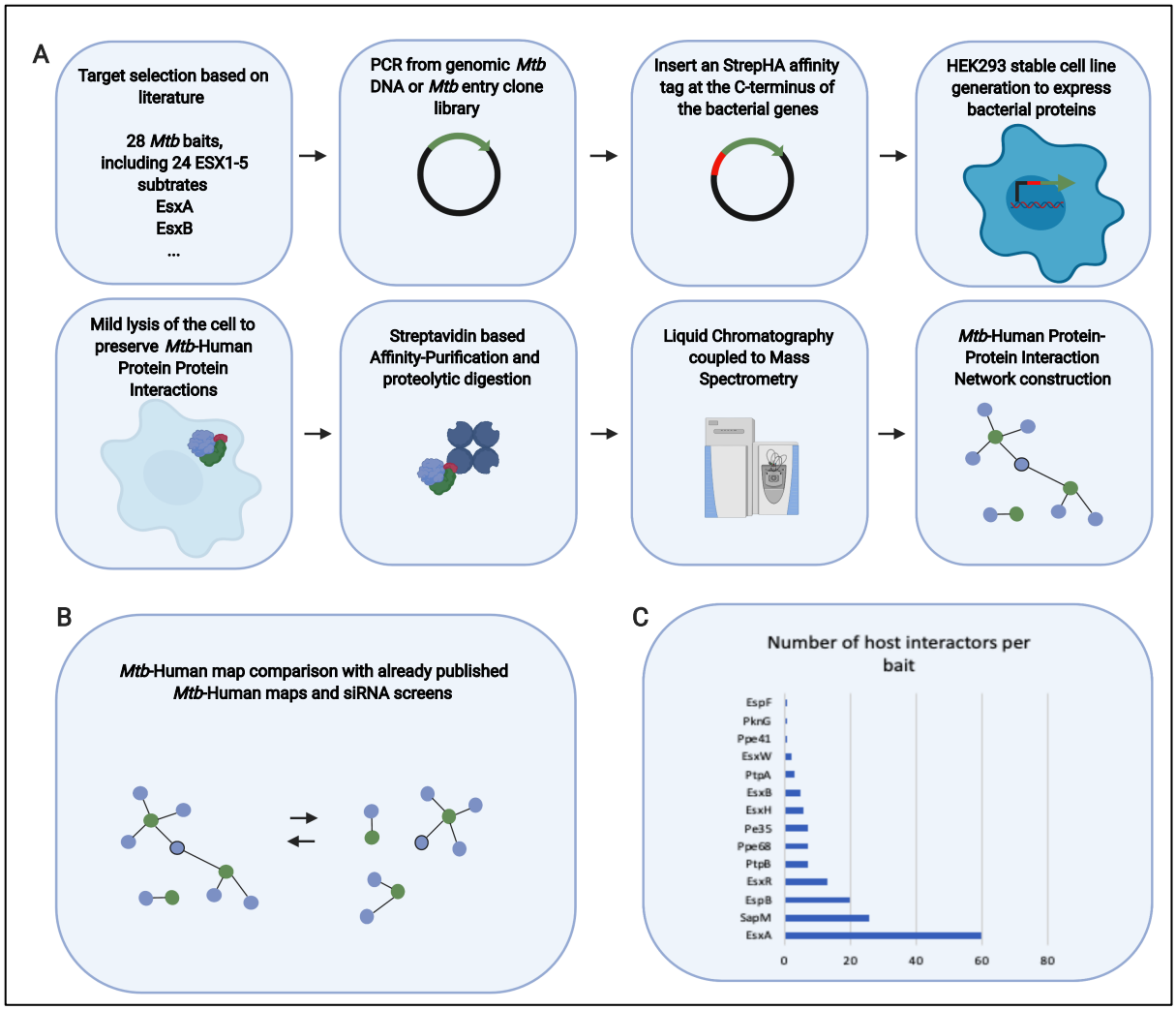


Figure 1 Panel A shows the workflow employed to map *Mtb*-human interactions. Once the bacterial proteins to study were selected, they were amplified from genomic DNA or from an *Mtb* entry clone library. The genes were then fused at their C terminus with a StrepHA tag and which were used to generate each of the 28 transgenic HEK293 cell lines. After inducing the expression of the bacterial gene upon Doxycycline addition, the cells were mildly lysed in order to preserve protein-protein interactions. The tagged bacterial baits were purified and along with the co-purified proteins, were subsequently digested into peptides. The peptides were analyzed by LC/MS using a Data Dependent Acquisition scheme. The host-pathogen interaction network was constructed using SAINT express analysis to identify high-confidence interactions. Panel B: our AP-MS results were overlapped with an existing *Mtb*-human protein-protein interaction map to calculate the overlapping results. Panel C represents the number of interactors identified by bacterial baits. Out of the 28 tested bacterial baits, only 50% were found to have host interactors.

4.2 Results

The first step towards selecting the putative secreted bacterial proteins was to find evidence of the secretion of the targets in the literature, more specifically in publications that analyzed the culture filtrates (CF) via bottom-up proteomics for secretion of the substrate to corroborate bioinformatic secretion predictions [39–41]. Next we looked for evidence that the secreted bacterial proteins elicited T or B cell responses or a protective immunity [42]. We then scouted for evidence suggesting they are strongly involved in the virulence of *Mycobacterium tuberculosis,*

mostly while observing marked phenotypes upon gene deletions [24,28,43–56] (summarized in table 1). The final target selection was composed of 10 ESX-1 secreted substrates, 4 ESX-3, 10 ESX-5, 3 SecA2 and 1 SecA1 secreted substrate. They included 14 common selected targets when comparing with the previously published AP-MS *Mtb*-Human PPI [38] which were all ESX secreted substrates.

Rv ID	Protein Name	Product description	Attenuated virulence upon KO *Mtb*	Essential for *in vitro* growth	Secretion System	Co-secretion	Remarks on the role of the bacterial product on host upon infections	Key publications
Rv3875	esxA	6 kDa early secretory antigenic target EsxA (ESAT-6)	yes	No	ESX-1	EsxA/EsxB	Necessary for phagosomal membran permability and cytosolic escape	Conrad WH *et al,* 2017
Rv3874	esxB	10 kDa culture filtrate antigen EsxB (CFP10)	yes	No	ESX-1	EsxA/EsxB	Necessary for phagosomal membran permability and cytosolic escape	Conrad WH *et al,* 2017
Rv0288	esxH	Low molecular weight protein antigen 7 EsxH (protein TB10.4)	yes	Yes	ESX-3	EsxH/EsxG	Inhibits ESCRT-dependent trafficking of receptors to the lysosome	Mehra A *et al,* 2013
Rv0287	esxG	ESAT-6 like protein EsxG (conserved protein TB9B)	yes	Yes	ESX-3	EsxH/EsxG	Inhibits ESCRT-dependent trafficking of receptors to the lysosome	Mehra A *et al,* 2013
Rv1197	esxK	ESAT-6 like protein EsxK (ESAT-6 like protein 3)	yes	No, disruption of this gene provides even growth advtanage to H37Rv	ESX-5	EsxK/EsxL	ΔesxKL mutants have reduced ability to colonize THP-1 cells	Bukka A *et al,* 2012
Rv1198	EsxL	Putative ESAT-6 like protein EsxL (ESAT-6 like protein 4)	yes	No, disruption of this gene provides even growth advtanage to H37Rv	ESX-5	EsxK/EsxL	ΔesxKL mutants have reduced ability to colonize THP-1 cells	Bukka A *et al,* 2012
Rv2431c	PE25	PE family protein PE25	yes	No	ESX-5	PE25/PPE41	Induces cell necrosis or host cell lysis, but not apoptosis	Smanla T *et al,* 2014
Rv2430c	PPE41	PPE family protein PPE41	yes	No	ESX-5	PE25/PPE41	Induces cell necrosis or host cell lysis, but not apoptosis	Smanla T *et al,* 2014
Rv3477	PE31	PE family protein PE31	yes	No	ESX-5	PE31/PPE18	Attenuates host cell apoptosis	Ali MD *et al,* 2020
Rv3872	PE35	PE family-related protein PE35	-	No	ESX-1	PE35/PPE68	Plays an Immunomodulatory role on host macrophages	Tiwari B *et al,* 2014
Rv3873	PPE68	PPE family protein PPE68	-	No	ESX-1	PE35/PPE68	Plays an Immunomodulatory role on host macrophages	Tiwari B *et al,* 2014
Rv3865	EspF	ESX-1 secretion-assodated protein EspF	yes	No	ESX-1	none	Attenuated bacterial replicatin *in vivo* with ΔEspF MTB	Bottai D *et al,* 2011
Rv3019c	EsxR	Secreted ESAT-6 like protein EsxR (TB10.3)	-	No	ESX-3	EsxR/EsxS	Causes strong humoral immunity	Arlehamn CL *et al,* 2013
Rv3020c	EsxS	ESAT-6 like protein EsxS	-	No	ESX-3	EsxR/EsxS	Causes strong humoral immunity	Arlehamn CL *et al,* 2013
Rv3864	EspE	ESX-1 secretion-associated protein EspE	yes	No	ESX-1	none	Leads to increased bacterial load and splenomegaly if complemented into avirulent strains	Brodin P *et al,* 2006
Rv3881c	EspB	Secreted ESX-1 substrate protein B, EspB	yes	No, disruption of this gene provides even growth advtanage to H37Rv	ESX-1	none	Required for host cell death, lack of EspB leads to decrease *in vivo* virulence	Ohol YM *et al,* 2010
Rv2347c	esxP	Putative ESAT-6 like protein EsxP	-	No, disruption of this gene provides even growth advtanage to H37Rv	ESX-5	EsxO/EsxP	EsxO promotes bacillary survival by inducing oxidative stress mediated genomic instability in macrophages.	Mohanty S *et al,* 2016
Rv2346c	esxO	Putative ESAT-6 like protein EsxO	-	No, disruption of this gene provides even growth advtanage to H37Rv	ESX-5	EsxO/EsxP	EsxO promotes bacillary survival by inducing oxidative stress mediated genomic instability in macrophages.	Mohanty S *et al,* 2016
Rv3620c	esxW	Putative ESAT-6 like protein EsxW	-	No	ESX-5	EsxW/EsxV	Could enhace the transmission of Beijing lineage in hosts	Holt K *et al,* 2018
Rv3619c	esxV	Putative ESAT-6 like protein EsxV	-	No	ESX-5	EsxW/EsxV	Could enhace the transmission of Beijing lineage in hosts	Holt K *et al,* 2018
Rv3878	EspJ	ESX-1 secretion-associated protein EspJ	-	No, disruption of this gene provides even growth advtanage to H37Rv	ESX-1	none	Phosphorylated EspJ seems to negatively affect bacterial growth upon infections	Singh P *et al,* 2015
Rv3849	espR	ESX-1 transcriptional regulatory protein EspR	-	No	ESX-1	none	EspR is a key regulated of ESX-1 expression, activates during early phagocytosis	Rosenberg O *et al,* 2011
Rv3867	espH	ESX-1 secretion-assodated protein EspH	yes	No	ESX-1	none	EspH is necessary to retain ESX-1 mediated substrate secretion and virulence	Phan T *et al,* 2018
Rv0410c	pknG	Serine/threonine-protein kinase PknG (protein kinase G)	yes	Disputed, likely not	SecA2	none	promotes MTB survival upon infections, and blocks phagosomal maturation	Kanehiro Y *et al,* 2018
Rv0153c	ptpB	Phosphotyrosine protein phosphatase PTPB	-	No	SecA2	none	PtpB attenuates host immune defenses by interferring with signal transduction pathways in macrophages	Mascarello A *et al,* 2013
Rv2234	ptpA	Phosphotyrosine protein phosphatase PtpA	-	No	SecA2	none	PtpA disrupts key components of the endocytic pathway, resulting in the arrest of phagosome maturation	Wong D *et al,* 2013
Rv3310	sapM	Acid phosphatase (acid phosphomonoesterase)	-	No	SecA1	none	SapM mediates the arrest of phagosome maturation by disrupting the recruitment of PI3P effector proteins such as early endosome antigen 1 (EEA1).	Wong D *et al,* 2013
Rv0256c	PPE2	PPE family protein PPE2	-	No	ESX-5		The PPE2 protein of *Mycobacterium tuberculosis* translocates to host nucleus and inhibits nitric oxide production	Bhat KH *et al,* 2017

Table 1 This table lists all the bacterial secreted proteins used in our screen. The fourth column describes whether *Mtb* KO strains of the secreted bacterial protein affects intracellular bacterial growth and overall pathogenicity. The fifth column states whether the KO of the bacterial gene affect the *in vitro* growth of *Mtb* and whether it is essential. The co-secretion system column states the co-secretion factor.

We hypothesized that because *Mtb* has been proven to permeabilize bacilli-containing phagosomes and eventually translocate into the host cytosol [15,57–59], any secreted bacterial protein could theoretically be in contact with the host and, as such, we could express them in human host cytosols. Initially, our aim was to express each of them tagged at their C-terminus with a StrepHA tag in HEK293 cells and then in the more physiological context-relevant monocytic precursor cells THP-1 differentiated into macrophages using a retroviral system [60]. However, due to the low expression levels within differentiated THP-1 of our tagged baits and consequent poor-quality affinity purifications (supplementary information), we will present only the HEK293 based purification results. In order to increase the expression levels of the bacterial genes within the host environment and consequently enhance the identifications of potential host preys, we performed each purification in biological triplicates, and included additional triplicates from HEK293 cells treated with 8 hours of Bortezomib, a potent protease inhibitor (supplementary information). In order to identify High-Confidence Interactors (HCI) and filter out non-specific interactions and contaminants, we performed in parallel control purifications using HEK293 cells expressing GFP with a StrepHA fused to its C-terminus, amounting to 42 control GFP purifications, and performed a SAINT analysis using a SAINT Probability Score filtering cut-off of 0.9 [61]. The interaction network yielded 207 *Mtb*-Human Protein-Protein Interactions between 142 host prey proteins and only 14 *Mtb* bait proteins. This suggested that only 50% of the tested bacterial baits had HCI with host proteins. When comparing these results with similar studies performed for viral pathogens, where 89% of HIV proteins [32], 100% of HCV [34] and 75% of KSHV [62] tested proteins interacted with human host proteins, this suggested that not all putative secreted interact directly with the host and might carry out their virulence functions by other mechanisms. However, the 46% of interactions between tested *Mtb* baits with human hosts where found to be similar when comparing with other bacterial AP-MS studies, including one for *C. trachomatis*-human [37] and one for *Mtb*-human [38] where 66% and 34% respectively of tested bacterial baits interacted with host proteins. One hypothesis that could explain a lower interaction rate between tested bacterial proteins compared to viral proteins with their hosts, is that viruses are obligated intracellular pathogens with minimal genomes and so limited in redundancy, whereas bacterial pathogens may survive outside of their host and have large genomes with hundreds of homologous predicted secreted proteins. Therefore, most viral proteins carry out pleiotropic functions and tend to interact with a higher degree to host proteins [63,64]. Another hypothesis could state that the function of some effectors is not mediated by stable protein-protein interactions, but rather by enzymatic activity.

With the aim to study all known *Mtb*-human PPIs, from this point we merged our network with the previously published one [38], thus allowing us to analyze in a more comprehensive and global manner all putative *Mtb*-human PPIs (figure 2). After merging both host-pathogen

networks, from the 14 common bacterial bait proteins, only 3 of them were found to have high-confidence host interactors identified in both datasets, namely EsxA, Pe35 and EspB. From those 3 interactors, only EspB baits co-purified specifically with some of the same preys including members of the nuclear pore complex NUP88 and NUP214, as well as YTHDF1 and DAZAP1, two RNA binding proteins and HNRPUL1, the heterogenous nuclear ribonucleoprotein 1 involved in alternative RNA splicing and a known basic transcription repressed upon viral infections [65]. From the remaining 11 common bacterial bait tested in both datasets, Pe25 and EspR were shown to interact with host components exclusively in the Penn dataset whereas EspF, EsxH and EsxB were found to interact with host proteins exclusively in our AP-MS screen. The remaining 6 bacterial baits tested in both datasets were found to interact with no host proteins. The discrepancy between the host interactors found in both datasets might be attributed to the different scoring systems but mostly due to the different cellular models utilized and lysate extraction methods. In the global *Mtb*-human map composed of the 314 host proteins, 26 host proteins were targeted by more than one bacterial bait. With the exception of 5 EspB interactors, 21 host proteins were targeted by 2 distinct bacterial baits. This could suggest that *Mycobacterium tuberculosis* could target the same host modules multiple times in order to ensure robustness of the intervention, especially considering that many secreted bacterial effector have high sequence homology.

To first identify which types of proteins the *Mtb* baits target within the host cellular environment, we performed Gene Ontology enrichment tests to retrieve statistically over represented Biological Processes (BP) amongst the interactors (figure 3, Benjamini pval < 0.05). Amidst the enriched biological processes, expected groups often targeted by both bacterial and viral pathogens were over-represented such as the transcription and translational machineries [66]. Further BP more specific to intracellular bacterial pathogens were found and included microtubules-based processes for their intracellular invasion and motility by interacting with microtubules or small GTPases [67] and proteasome-mediated and ubiquitin-dependent protein catabolism [68]. Host targets more specific to *Mtb* infections were enriched in RNA splicing via the Spliceosome [69], antigen processing and presenting via MHC II molecules, regulation of type I interferon production [70] and COPII vesicle coating [38]. In order to establish the topology of the network, we merged the host interactors with known core complexes found in the Corum database [71] (highlighted in dotted lines in Figure 2). After annotating host interactors with the Corum database, multiple complexes clustered. Those included 2 out of 3 mitochondrial intermembrane transport complex TIMM8A-TIMM13-TIMM23 complex, 16 members of the spliceosome complex, two members of the RAF1-PPP2-PIN1 complex involved in the transmembrane receptor protein serine/threonine signaling pathway, 4 members of the PI4K2A-

WASH complex necessary for the early endosome formation, 4 components of the nuclear pore complex, 12 members of the Nop56p-associated pre-rRNA complex for the ribosomal genesis, 50% of the MLL1 complex, 66% of the MCM complex, the entire IRF2BP2-IRF2BP1-IRF2BPL which is an interferon-dependent transcriptional corepressor complex, the endosomal targeting complex VIPAS39-VPS33B which already proven to interact with *Mtb* secreted bacterial effect proteins [72] and 7 out of 8 the chaperonin containing TCP complex. Three bacterial proteins, namely EsxA, Ppe68 and Tb8.4 were found to interact with 5 Rab proteins: Rab1A, Rab1B, Rab5A, Rab7A and Rab11B. The roles of Rab1a and Rab1b in the context of *Mycobacterium tuberculosis* infections have yet to be clearly understood, but publications studying the infection patterns of the intracellular *Legionella pneumophila* pathogen have shown that the bacilli recruits Rab1 proteins to their early-stage phagosomes and is required for its intracellular growth [73]. Rab5 on the other hand is undeniably necessary for the phagosomal maturation and is necessary to recruit Rab7 to ensure the progression of the maturation towards a phagolysosome [74]. By interacting with and thus spatially sequestering both Rab5 and Rab7, we could hypothesized that *Mycobacterium tuberculosis* prevents the maturation of its bacilli-containing phagosomes which could explain why only virulent *Mtb* strains with fully functional ESX secretion systems selectively modulate the spatial organization of Rab7 [75]. This corroborates our findings that *Mtb* does not modulate Rab5 and Rab7 activities by regulating their total abundance (cf chapter 3) but rather on their spatial sequestration.

It was recently shown that virulent *Mtb* strains containing fully functioning ESX secretion systems modulate alternative splicing in host macrophages in order to shape host response to *Mtb* infections [69]. By interfering with alternative splicing, *Mtb* could thus influence the transcripts' stabilities or structures, and indirectly on the functions and localizations of their protein products. Due to the fact that *Mtb* has been reported to modulate alternative splicing and the fact that we unraveled so many interactions between 16 spliceosome-related proteins and 8 secreted bacterial proteins, we sought to validate the *Mtb*-spliceosome interactions. To accomplish this, we infected THP-1 cells with H37Rv at a Multiplicity of Infection of 10 and performed an immuno-purification 48 hours post infection using the Active Motif Nuclear Complex Co-IP kit (catalog No.54001). Unfortunately, the results were not conclusive as the quality of the purification within the BSL3 conditions probably yielded no recovery of the spliceosome complex due to the filter-sterilization steps.

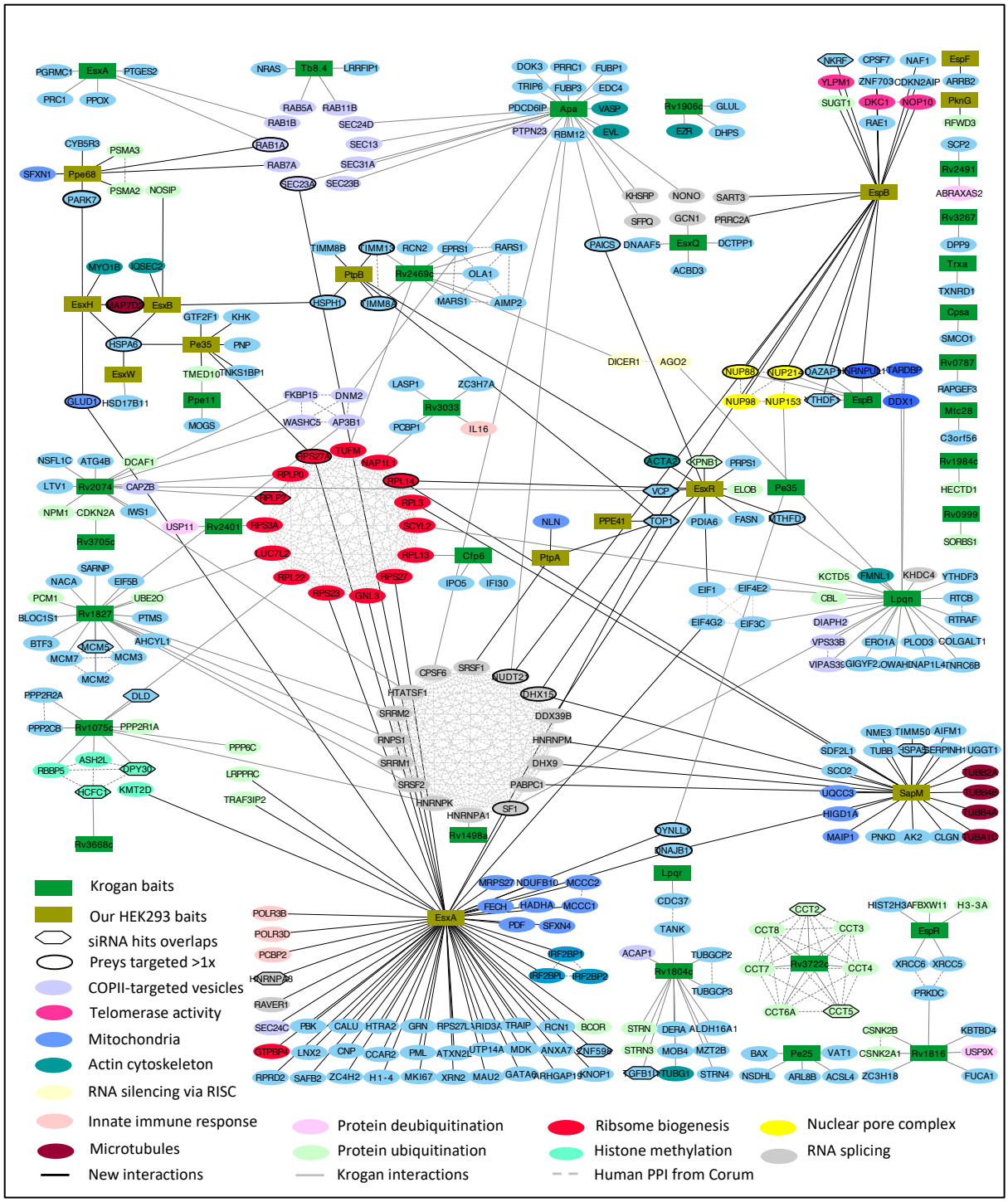


Figure 3 This figure represents the merged *Mtb*-human interaction map obtained from our AP-MS study, and the previously published one by Penn *et al.* Edges in black are interactions found in our dataset. Edges in grey those reported in the Penn *et al* dataset. Dashed edges are those reported from the Corum database. Dark green rectangles are the baits used in the Penn study. Light green rectangles are those tested in our AP-MS study. Host interactors circled in black are those identified to interact with more than one bait regardless of the study. Host interactors in the shape of hexagons signify that they are either host restrictive or supportive of intracellular bacterial replication based on siRNA screens. Color coding represent the belonging biological process extracted from Corum. If the target host proteins were not listed in the Corum database, their biological process is not highlighted.

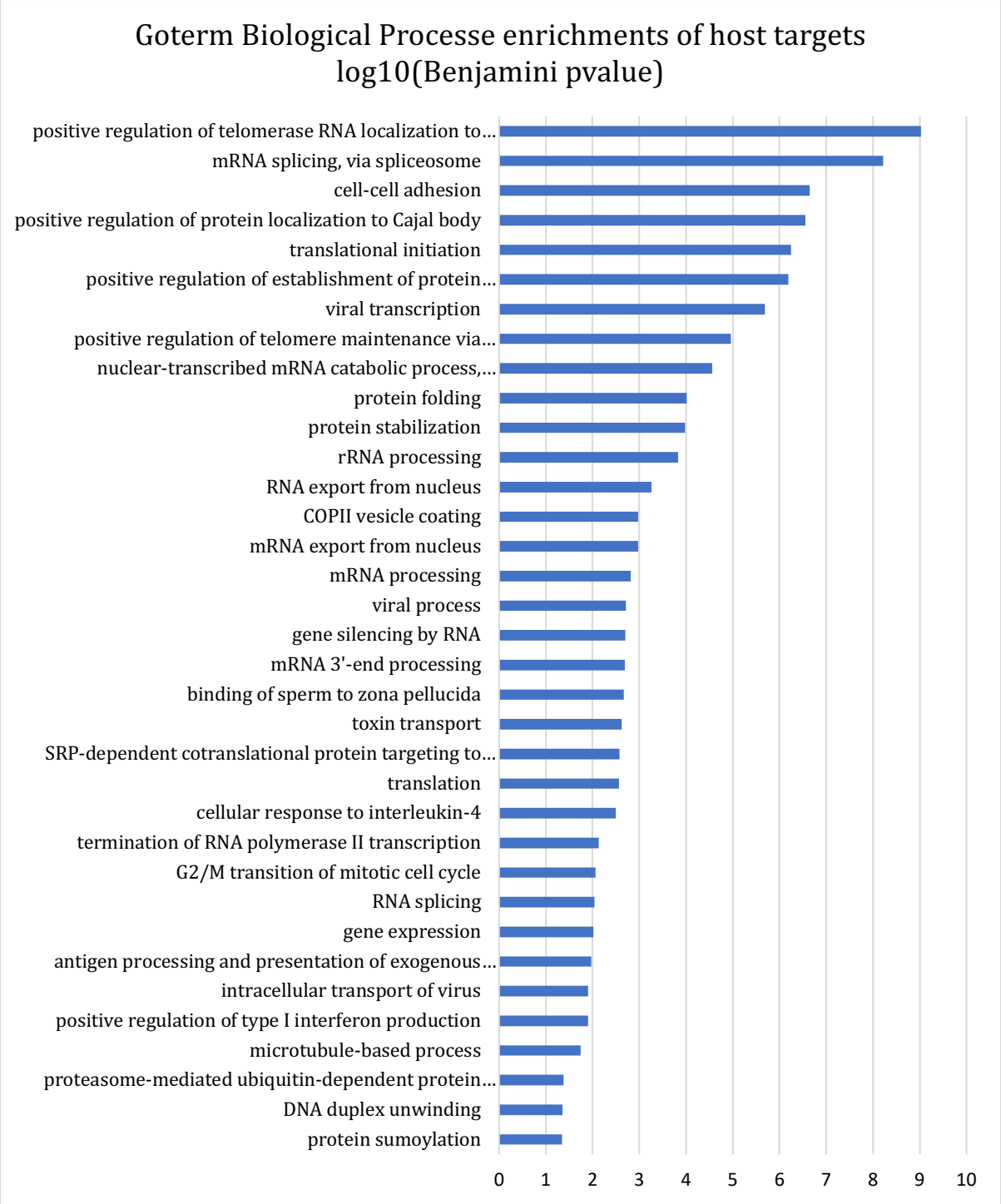


Figure 4 This figure shows the Gene Ontology Biological Process enriched in the host interactors found in the merged AP-MS *Mtb*-human interaction map. The adjusted pvalue was used to set a threshold (Benjamini pvalue < 0.05) for the BP.

survival and replication, we overlapped our *Mtb*-human PPI network with 2 siRNA screens: one recuperated from the published dataset from D. Kumar and colleagues [76] and one provided by Prof. Priscille Brodin (unpublished). Both datasets were produced using a combination of RNA interference screens with visual screens monitoring intracellular GFP-*Mtb* growth to allow the identification of candidate host factors that would either support or restrict intracellular

mycobacterial growth. The overlap consisted of 16 targets identified as direct *Mtb*-human interactions are that affected bacterial growth upon an H37Rv infection (Figure 2). Notable overlaps included two components of the CCT complex, which upon host interference induced a decreased bacterial replication suggesting that the CCT facilitates intracellular bacterial replication. Recently, it was demonstrated that the CCT chaperonin complex was necessary to the carry out the toxicity of *Clostridium difficile* by interacting with two of its toxins, suggesting here again to be a facilitator of the intracellular bacterial replication. Further notable overlaps with the siRNA screen demonstrated that two out of the four of the components of the MLL1 complex interacting with bacterial proteins, namely Hcfc1 and Dpy30 both decreased the bacterial loads upon siRNA treatments in THP-1 macrophages. Additionally, the NF-kappa-B-repressing factor was found to interact with EspB and to be a facilitator to mycobacterial growth from the siRNA screens. Given that the exact roles of the NF-kappa-B activities upon *Mtb* infections remain highly intricate and disputed [77–79], the functional relevance of the of the EspB-NFKBRF should be further examined and could potentially lead to a HDT.

4.3 Conclusions

In this study, we completed and confirmed existing *Mycobacterium tuberculosis* – human protein-protein interaction maps. Our merged networks with previously published Mtb-human interactions yielded *Mtb*-Human Protein-Protein Interactions between 316 unique host prey proteins and only 47 *Mtb* bait proteins. Findings supported previous knowledge that *Mtb* manipulates proteins involved in Rab-mediated phagosomal maturation including Rab5 and Rab7, antigen presentation via MHC class II molecules and most likely interferes with canonical spliceosome functions. Although we did not verify the mentioned interactions, previous studies utilizing AP-MS screens to map host-pathogen interactions teach us that most tested via reciprocal purifications have been validated [38]. By overlapping direct host interactors with siRNA screens aimed at identifying host factors that either promote or restrict intracellular bacterial growth upon infection, we could gain a better view of potential molecular mechanisms of virulence carried out by bacterial secreted proteins. Although further studies should be carried out to understand the molecular mechanisms, targeting certain host factors could lead towards the development of host-directed therapies. Potential targets could be the modulation of the NF-Kappa-B repression factor, the CCT chaperonin complex or the MLL1 complex.

4.4 Materials and Methods

Cloning

The *Mycobacterium tuberculosis* gene collection was obtained from beiresources.org. The bacteria containing the plasmid of interest were picked and grown overnight in LB Media with 50 ug/mL of Kanamycin at 37°C while shaking at 300 rpm. The bacteria were pelleted the following morning by a centrifugation step at 10,000g for 5 minutes. The supernatants were discarded and the plasmids were purified using the QIAprep Spin Miniprep Kit (cat. No 27104) following the kit instructions. The LR (ligase reaction) to fuse the strepHA tags were done as following. 100ng of the entry clone and 100ng of the destination vector were incubated with 1uL of Gateway LR Clonase II ® enzyme mix (from Thermo Fisher Scientific, cat no 11791020) for 1 hour at 37°C while shaking at 300 rpm.

Bacterial transformation

The LR cloning reaction was added to 50 ul of DH5 alpha competent cells (Thermo Fischer Scientific, cat no 18263012) and incubated on ice for 30 minutes. They were then heat-shocked at 42°C for 2 minutes before being incubated on ice for 1 minute. They were then added to 250uL of LB Medium and incubated for 37°C for 45 minutes while shaking at 300 rpm. The bacteria were spun down for 1 minute at 2,000g and the supernatants were discarded. The bacteria were then plated onto agar plates supplemented with selective 50ug/mL of ampicillin and incubated overnight at 37°C. The bacterial colonies were picked and grown in LB media supplemented of selective 50ug/mL ampicillin. Finally, the plasmids were purified again using the QIAprep Spin Miniprep Kit (cat. No 27104) following the kit instructions.

HEK293 stable cell line generation

HEK Flp-In 293 T-Rex cells (Life Technologies) containing a single genomic FRT site and stably expressing the tet repressor were cultured in DMEM medium (4.5 g/l glucose, 2 mM L-glutamine; Life Technologies) supplemented with 10% FCS, 50 ug/ml penicillin, 50 ug/ml streptomycin and 15 lg/ml blasticidin. The medium was exchanged with DMEM medium (10% FCS, 50 lg/ml penicillin, 50 lg/ml streptomycin) before transfection. For cell line generation, Flp-In HEK293 cells were co-transfected with the corresponding expression plasmids containing and the pOG44 vector (Life Technologies) for co-expression of the Flp-recombinase using the FuGENE 6 transfection reagent (Promega, Fitchburg, WI, USA). Two days after transfection, cells were selected in hygromycin containing medium (100 lg/ml) for 2–3 weeks.

Protein expression in stable HEK293 stable cell lines

Cells were induced to express the tagged bacterial baits or GFP upon reaching a confluency of 60%. The media was replaced with fresh DMEM media supplemented with 10% FCS, 1% Penicillin and Streptomycin as well as 2ug/ml of doxycyline (Sigma Aldrich) dissolved in 70% ethanol. Half of the samples were supplemented with 5 uM Bortezomib dissolved in DMSO for 8 hours prior to cell harvesting.

Cell harvesting

The adherent HEK293 cells were washed with ice-cold PBS (Gibco) and were then harvested in PBS supplemented with 10 mM EDTA by pipetting. The cells were then spun down upon 5 minutes of centrifugation at 300g at 4°C, before being washed again with PBS and spun again. The supernatants were then discarded, and the cells were snap-frozen in liquid nitrogen and stored at -80°C.

DSS-MagStrep XT beads preparation

MagStrep III XT beads (IBA) were crosslinkinged with 1 mM DSS dissolved in DMF (Sigma) for 1 hour at 37°C on a slow motion rotor in 50 mM HEPES, 150 mM NaCl pH 7.8. After the crosslinking, the beads were incubated with 50 mM Glycine for 1 hour, and washed 3 times with HEPES /NaCl before being stored at 4°C.

Protein purification and on-beads digestion

The equivalent of 6x 150mm plates at a confluency of 80% as frozen cell pellets of stable isogenic cell lines were resuspended in 4 ml HNN lysis buffer [50 mM HEPES pH 7.5, 150 mM NaCl, 50 mM NaF, 0.5% Igepal CA-630 (Nonidet P-40 Substitute), 200 lM Na3VO4, 1 mM PMSF, 20 lg/ml Avidin and 1x Protease Inhibitor mix (SigmaAldrich, St. Louis, MO, USA)] and incubated on a slow rotor for 15 minutes at 4°C. Insoluble material was removed by centrifugation for 30 minutes at 4°C at 13,000g. Cleared lysates were loaded on prewashed and equilibrated in HNN lysis buffer DSS-MagStrep beads and the beads-cell lysate mix was transferred to 4°C on a slow-motion rotor for 45 minutes. After the binding incubation, the beads were washed twice with HNN lysis buffer and three tiems with HNN buffer (50 mM HEPES pH 7.5, 150 mM NaCl, 50 mM NaF). After the third wash, the beads with the bound bacterial proteins and interactors were resuspended in 50ul of 50 mM Ammonium Bicarbonate and 5M Urea. They were then reduced with a final concentration of 2.5 mM TCEP (Sigma) for 30 minutes at 37°C. The reduced cysteine residues were then alkylated for 30 minutes at room temperature in the dark with a 4 mM final Iodoacetamide concentration (Sigma). The urea-containing solution was then diluted to a final concentration of 1M Urea by adding 50 mM ammonium bicarbonate. The proteins were digested for 1 hour at 37°C

with LysC endopeptidase (Pierce LysC Protease) in a 1:100 enzyme-to-substrate ratio and then overnight with Trypsin (Promega) at 37°C on a slow-motion rotor. The following morning, the digestion was stopped by adding formic acid until reaching pH 2. The peptides were desalted using a reverse-phase peptide clean-up step using C18 MicroSpin, 5-60ug capacity from the Nest Group. The C18 columns were first activated with pure acetonitrile (ACN), and then equilibrated with 2% ACN and 0.1% formic acid. The acidified peptide-containing solutions were then loaded onto the C18 columns, and subsequently washed 3 times with 2% ACN, 0.1% formic acid. Peptides were eluted from the desalting column using 50% ACN, 0.1% FA and dried in acid-resistance centrivap concentrator (Labconco). Finally the peptides were resuspended in 20 ul of 2% CAN, 0.1% formic acid with 1:20 iRT peptides (Biognosys), sonicated for 10 minutes and transferred to MS-compatible vials.

Peptide analysis via LC-MS/MS

LC-MS/MS analysis of the peptides was performed on an LTQ Orbitrap XL mass spectrometer (Thermo Fisher Scientific) which was coupled to a a Proxeon EASY-nLC II liquid chromatography system (Thermo Fisher Scientific) connected to an Reverse Phase -High Pressure Liquid Chromatography column (15 cm packed with Reprosil Pur 120 C18-AQ) for the chromatographic peptide separation. Solvent A was used as RP-HPLC stationary phase (0.1% formic acid, 2% acetonitrile). Solvent B (mobile phase; 0.1% formic acid, 98% acetonitrile) was used to run a linear gradient from 5 to 35% over 90 min at a flow rate of 300 nl/min. The 6 most abundant ions from the first MS scan were fragmented by collisioninduced dissociation (CID) and MS/MS fragment ion spectra were acquired in the linear trap quadrupole (LTQ). Charge state screening was enabled and unassigned or singly charged ions were rejected. The dynamic exclusion window was set to 15 s and limited to 300 entries. Only MS precursors that exceeded a threshold of 150 ion counts were allowed to trigger MS/MS scans. The ion accumulation time was set to 500 ms and 250 for the MS1 and MS2 scans respectively.

Protein identification

The centroided RAW acquired spectra via DDA LC-MS/MS were first converted to mzXML files. They were then searched with the against the canonical human proteome reference dataset (http://www.uniprot.org/) supplemented with the bacterial protein sequences and extended with reverse decoy sequences for all entries using the Comet version 2017.01.03 search engine (comet-ms.sourceforge.net) with the following parameters: carbamido-methylation as fixed modification and Oxidation on methionine and phosphorylation on Serine, Threonine and Tyrosine as variable modifications. The search parameters were set to include fully tryptic peptides and containing up to two missed cleavages. The precursor mass tolerance was set to 25

ppm and the fragment mass error tolerance to 0.5 Da. The obtained peptide spectrum matches were statistically evaluated using PeptideProphet and protein inference by ProteinProphet, both part of the Trans Proteomic Pipeline (TPP). A minimum protein probability of 0.9 was set to match a false discovery rate (FDR) of < 1%.

Evaluation of high confidence interacting proteins and network analysis

Protein matrix with the spectral counts were subjected to SAINT express analysis on the CRAPome platform (crapome.org) using the 43 aggregated GFP control purification to filter out non-specific interactions. HCI were filtered by using a SAINT Probability Score above 0.9. The network visualization was done by using Cytoscape software version 3.7.2. Functional gene enrichment analysis was performed on DAVID (Database for Annotation, Visualization and Integrated Discovery, version 6.8, https://david.ncifcrf.gov/) using an adjusted Benjamini pvalue cutoff of 0.05.

4.5 References

1. Paulson T: **A mortal foe**. *Nature* 2013, **502**.
2. *Status of the TB epidemic*. 2019.
3. **Treatment Highlights of Drug-Susceptible Tuberculosis Guidelines | Treatment | TB | CDC [Internet]**. [date unknown], [no volume].
4. Sarkar S, Ganguly A, Sunwoo HH: **Current Overview of Anti-Tuberculosis Drugs: Metabolism and Toxicities**. 2016, doi:10.4172/2161-1068.1000209.
5. Mirlohi M-S, Ekrami A, Shirali S, Ghobeishavi M, Pourmotahari F: **Hematological and liver toxicity of anti-tuberculosis drugs**. *Electron. Physician* 2016, **8**:3005.
6. Palomino JC, Martin A: **Drug resistance mechanisms in Mycobacterium tuberculosis**. *Antibiotics* 2014, **3**:317–340.
7. Coll F, Phelan J, Hill-Cawthorne GA, Nair MB, Mallard K, Ali S, Abdallah AM, Alghamdi S, Alsomali M, Ahmed AO, et al.: **Genome-wide analysis of multi- and extensively drug-resistant Mycobacterium tuberculosis**. *Nat. Genet.* 2018, **50**:307–316.
8. Shean K, Streicher E, Pieterson E, Symons G, van Zyl Smit R, Theron G, Lehloenya R, Padanilam X, Wilcox P, Victor TC, et al.: **Drug-Associated Adverse Events and Their Relationship with Outcomes in Patients Receiving Treatment for Extensively Drug-Resistant Tuberculosis in South Africa [Internet]**. *PLoS One* 2013, **8**:e63057.
9. Kaufmann SHE, Dorhoi A, Hotchkiss RS, Bartenschlager R: **Host-directed therapies for bacterial and viral infections**. *Nat. Rev. Drug Discov.* 2018, **17**:35–56.
10. Moraco AH, Kornfeld H: **Cell death and autophagy in tuberculosis. [Internet]**. *Semin. Immunol.* 2014, **26**:497–511.
11. Chang ST, Linderman JJ, Kirschner DE: **Multiple mechanisms allow Mycobacterium tuberculosis to continuously inhibit MHC class II-mediated antigen presentation by macrophages. [Internet]**. *Proc. Natl. Acad. Sci. U. S. A.* 2005, **102**:4530–5.
12. Hanekom WA, Mendillo M, Manca C, Haslett PAJ, Siddiqui MR, Barry III C, Kaplan G: ***Mycobacterium tuberculosis* Inhibits Maturation of Human Monocyte-Derived Dendritic Cells In Vitro [Internet]**. *J. Infect. Dis.* 2003, **188**:257–266.
13. Armstrong JA, Hart D: **Response of cultured macrophages to Mycobacterium Tuberculosis, with observations on fusion of lysosomes with phagosomes [Internet]**. *J. Exp. Med.* 1971, **134**:713–740.
14. Houben D, Demangel C, van Ingen J, Perez J, Baldeón L, Abdallah AM, Caleechurn L, Bottai D, van Zon M, de Punder K, et al.: **ESX-1-mediated translocation to the cytosol controls virulence of mycobacteria [Internet]**. *Cell. Microbiol.* 2012, **14**:1287–1298.

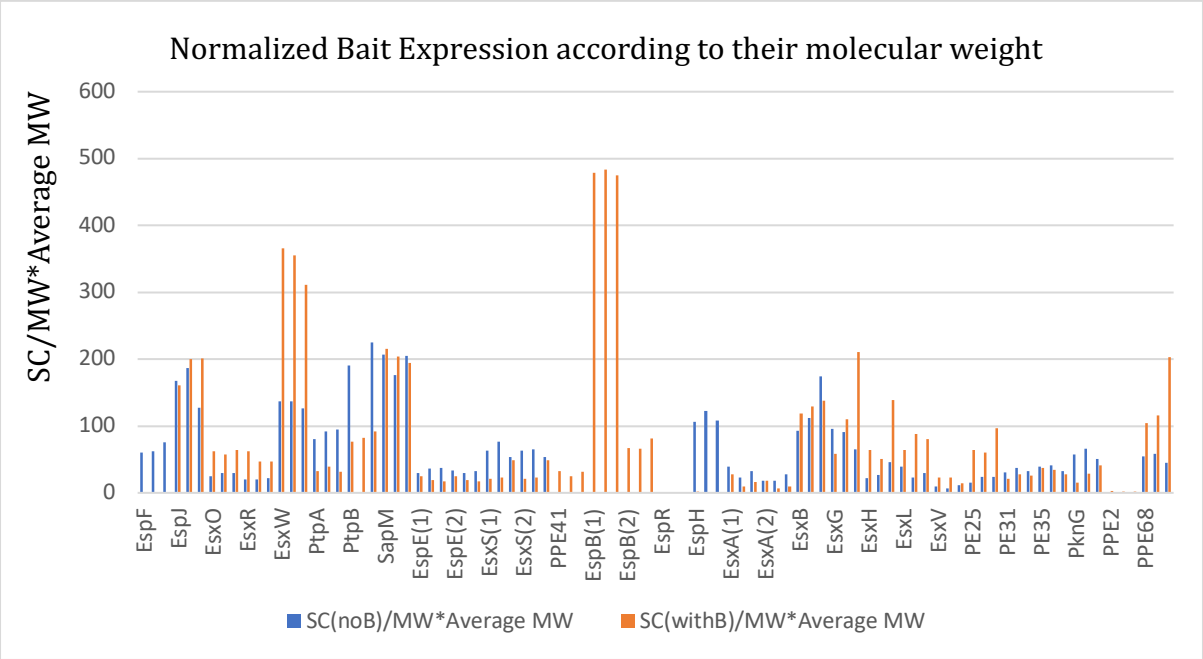
15. van der Wel N, Hava D, Houben D, Fluitsma D, van Zon M, Pierson J, Brenner M, Peters PJ: **M. tuberculosis and M. leprae Translocate from the Phagolysosome to the Cytosol in Myeloid Cells [Internet]**. *Cell* 2007, **129**:1287–1298.
16. Jasenosky LD, Scriba TJ, Hanekom WA, Goldfeld AE: **T cells and adaptive immunity to *Mycobacterium tuberculosis* in humans [Internet]**. *Immunol. Rev.* 2015, **264**:74–87.
17. Green ER, Mecsas J: **Bacterial Secretion Systems: An Overview. [Internet]**. *Microbiol. Spectr.* 2016, **4**.
18. Bosserman RE, Champion PA: **Esx systems and the mycobacterial cell envelope: What's the connection?** *J. Bacteriol.* 2017, **199**.
19. Gröschel MI, Sayes F, Simeone R, Majlessi L, Brosch R: **ESX secretion systems: mycobacterial evolution to counter host immunity [Internet]**. *Nat. Rev. Microbiol.* 2016, **14**:677–691.
20. Stanley SA, Raghavan S, Hwang WW, Cox JS: **Acute infection and macrophage subversion by Mycobacterium tuberculosis require a specialized secretion system**. *Proc. Natl. Acad. Sci. U. S. A.* 2003, **100**:13001–13006.
21. Guinn KM, Hickey MJ, Mathur SK, Zakel KL, Grotzke JE, Lewinsohn DM, Smith S, Sherman DR: **Individual RD1 -region genes are required for export of ESAT-6/CFP-10 and for virulence of Mycobacterium tuberculosis**. *Mol. Microbiol.* 2004, **51**:359–370.
22. Abdallah AM, Verboom T, Hannes F, Safi M, Strong M, Eisenberg D, Musters RJP, Vandenbroucke-Grauls CMJE, Appelmelk BJ, Luirink J, et al.: **A specific secretion system mediates PPE41 transport in pathogenic mycobacteria [Internet]**. *Mol. Microbiol.* 2006, **62**:667–679.
23. Hsu T, Hingley-Wilson SM, Chen B, Chen M, Dai AZ, Morin PM, Marks CB, Padiyar J, Goulding C, Gingery M, et al.: **The primary mechanism of attenuation of bacillus Calmette–Guérin is a loss of secreted lytic function required for invasion of lung interstitial tissue**. *Proc. Natl. Acad. Sci.* 2003, **100**.
24. Conrad WH, Osman MM, Shanahan JK, Chu F, Takaki KK, Cameron J, Hopkinson-Woolley D, Brosch R, Ramakrishnan L: **Mycobacterial ESX-1 secretion system mediates host cell lysis through bacterium contact-dependent gross membrane disruptions**. *Proc. Natl. Acad. Sci. U. S. A.* 2017, **114**:1371–1376.
25. Siegrist MS, Unnikrishnan M, McConnell MJ, Borowsky M, Cheng TY, Siddiqi N, Fortune SM, Moody DB, Rubin EJ: **Mycobacterial Esx-3 is required for mycobactin-mediated iron acquisition**. *Proc. Natl. Acad. Sci. U. S. A.* 2009, **106**:18792–18797.
26. Tufariello JAM, Chapman JR, Kerantzas CA, Wong KW, Vilchèze C, Jones CM, Cole LE, Tinaztepe E, Thompson V, Fenyö D, et al.: **Separable roles for Mycobacterium tuberculosis ESX-3 effectors in iron acquisition and virulence**. *Proc. Natl. Acad. Sci. U. S. A.* 2016, **113**:E348–E357.
27. Tinaztepe E, Wei JR, Raynowska J, Portal-Celhay C, Thompson V, Philipsa JA: **Role of metal-dependent regulation of ESX-3 secretion in intracellular survival of Mycobacterium tuberculosis**. *Infect. Immun.* 2016, **84**:2255–2263.
28. Mehra A, Zahra A, Thompson V, Sirisaengtaksin N, Wells A, Porto M, Köster S, Penberthy K, Kubota Y, Dricot A, et al.: **Mycobacterium tuberculosis Type VII Secreted Effector EsxH Targets Host ESCRT to Impair Trafficking [Internet]**. *PLoS Pathog.* 2013, **9**:e1003734.
29. Houben ENG, Bestebroer J, Ummels R, Wilson L, Piersma SR, Jiménez CR, Ottenhoff THM, Luirink J, Bitter W: **Composition of the type VII secretion system membrane complex [Internet]**. *Mol. Microbiol.* 2012, **86**:472–484.
30. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon S V., Eiglmeier K, Gas S, Barry CE, et al.: **Deciphering the biology of mycobacterium tuberculosis from the complete genome sequence**. *Nature* 1998, **393**:537–544.
31. Abdallah AM, Savage NDL, van Zon M, Wilson L, Vandenbroucke-Grauls CMJE, van der Wel NN, Ottenhoff THM, Bitter W: **The ESX-5 Secretion System of Mycobacterium marinum Modulates the Macrophage Response** . *J. Immunol.* 2008, **181**:7166–7175.
32. Jäger S, Cimermancic P, Gulbahce N, Johnson JR, McGovern KE, Clarke SC, Shales M, Mercenne G, Pache L, Li K, et al.: **Global landscape of HIV–human protein complexes [Internet]**. *Nature* 2011, **481**:365.
33. Davis ZH, Verschueren E, Jang GM, Kleffman K, Johnson JR, Park J, Von Dollen J, Maher MC, Johnson T, Newton W, et al.: **Global Mapping of Herpesvirus-Host Protein Complexes Reveals a Transcription Strategy for Late Genes [Internet]**. *Mol. Cell* 2015, **57**:349–360.
34. Ramage HR, Kumar GR, Verschueren E, Johnson JR, Von Dollen J, Johnson T, Newton B, Shah P, Horner J, Krogan NJ, et al.: **A Combined Proteomics/Genomics Approach Links Hepatitis C Virus Infection with Nonsense-Mediated mRNA Decay [Internet]**. *Mol. Cell* 2015, **57**:329–340.
35. Kane JR, Stanley DJ, Hultquist JF, Johnson JR, Mietrach N, Binning JM, Jónsson SR, Barelier S,

Newton BW, Johnson TL, et al.: **Lineage-Specific Viral Hijacking of Non-canonical E3 Ubiquitin Ligase Cofactors in the Evolution of Vif Anti-APOBEC3 Activity [Internet]**. *Cell Rep.* 2015, **11**:1236–1250.
36. Gordon DE, Jang GM, Bouhaddou M, Xu J, Obernier K, O'Meara MJ, Guo JZ, Swaney DL, Tummino TA, Huettenhain R, et al.: **A SARS-CoV-2-Human Protein-Protein Interaction Map Reveals Drug Targets and Potential Drug-Repurposing**. *bioRxiv* 2020, doi:10.1101/2020.03.22.002386.
37. Mirrashidi KM, Elwell CA, Verschueren E, Johnson JR, Frando A, Von Dollen J, Rosenberg O, Gulbahce N, Jang G, Johnson T, et al.: **Global Mapping of the Inc-Human Interactome Reveals that Retromer Restricts Chlamydia Infection [Internet]**. *Cell Host Microbe* 2015, **18**:109–121.
38. Penn BH, Netter Z, Johnson JR, Von Dollen J, Jang GM, Johnson T, Ohol YM, Maher C, Bell SL, Geiger K, et al.: **An Mtb-Human Protein-Protein Interaction Map Identifies a Switch between Host Antiviral and Antibacterial Responses [Internet]**. *Mol. Cell* 2018, **71**:637-648.e5.
39. Målen H, Berven FS, Fladmark KE, Wiker HG: **Comprehensive analysis of exported proteins fromMycobacterium tuberculosis H37Rv [Internet]**. *Proteomics* 2007, **7**:1702–1718.
40. de Souza GA, Leversen NA, Målen H, Wiker HG: **Bacterial proteins with cleaved or uncleaved signal peptides of the general secretory pathway [Internet]**. *J. Proteomics* 2011, **75**:502–510.
41. Målen H, Pathak S, Søfteland T, de Souza GA, Wiker HG: **Definition of novel cell envelope associated proteins in Triton X-114 extracts of Mycobacterium tuberculosis H37Rv. [Internet]**. *BMC Microbiol.* 2010, **10**:132.
42. Sampson SL: **Mycobacterial PE/PPE proteins at the host-pathogen interface. [Internet]**. *Clin. Dev. Immunol.* 2011, **2011**:497203.
43. Holt KE, McAdam P, Thai PVK, Thuong NTT, Ha DTM, Lan NN, Lan NH, Nhu NTQ, Hai HT, Ha VTN, et al.: **Frequent transmission of the Mycobacterium tuberculosis Beijing lineage and positive selection for the EsxW Beijing variant in Vietnam**. *Nat. Genet.* 2018, **50**:849–856.
44. Rosenberg OS, Dovey C, Tempesta M, Robbins RA, Finer-Moore JS, Stroud RM, Cox JS: **EspR, a key regulator of Mycobacterium tuberculosis virulence, adopts a unique dimeric structure among helix-turn-helix proteins**. *Proc. Natl. Acad. Sci. U. S. A.* 2011, **108**:13450–13455.
45. Phan TH, van Leeuwen LM, Kuijl C, Ummels R, van Stempvoort G, Rubio-Canalejas A, Piersma SR, Jiménez CR, van der Sar AM, Houben ENG, et al.: **EspH is a hypervirulence factor for Mycobacterium marinum and essential for the secretion of the ESX-1 substrates EspE and EspF**. *PLoS Pathog.* 2018, **14**.
46. Kanehiro Y, Tomioka H, Pieters J, Tatano Y, Kim H, Iizasa H, Yoshiyama H: **Identification of Novel Mycobacterial Inhibitors Against Mycobacterial Protein Kinase G [Internet]**. *Front. Microbiol.* 2018, **9**:1517.
47. Wong D, Chao JD, Av-Gay Y: **Mycobacterium tuberculosis-secreted phosphatases: From pathogenesis to targets for TB drug development**. *Trends Microbiol.* 2013, **21**:100–109.
48. Bhat KH, Srivastava S, Kotturu SK, Ghosh S, Mukhopadhyay S: **The PPE2 protein of Mycobacterium tuberculosis translocates to host nucleus and inhibits nitric oxide production**. *Sci. Rep.* 2017, **7**:1–13.
49. Bukka A, Price CTD, Kernodle DS, Graham JE: **Mycobacterium tuberculosis RNA expression patterns in sputum bacteria indicate secreted Esx factors contributing to growth are highly expressed in active disease [Internet]**. *Front. Microbiol.* 2012, **2**:266.
50. Tundup S, Mohareer K, Hasnain SE: **Mycobacterium tuberculosis PE25/PPE41 protein complex induces necrosis in macrophages: Role in virulence and disease reactivation?** *FEBS Open Bio* 2014, **4**:822–828.
51. Ali MK, Zhen G, Nzungize L, Stojkoska A, Duan X, Li C, Duan W, Xu J, Xie J: **Mycobacterium tuberculosis PE31 (Rv3477) Attenuates Host Cell Apoptosis and Promotes Recombinant M. smegmatis Intracellular Survival via Up-regulating GTPase Guanylate Binding Protein-1 [Internet]**. *Front. Cell. Infect. Microbiol.* 2020, **10**:40.
52. Tiwari B, Soory A, Raghunand TR: **An immunomodulatory role for the Mycobacterium tuberculosis region of difference 1 locus proteins PE35 (Rv3872) and PPE68 (Rv3873). [Internet]**. *FEBS J.* 2014, **281**:1556–70.
53. Lindestam Arlehamn CS, Gerasimova A, Mele F, Henderson R, Swann J, Greenbaum JA, Kim Y, Sidney J, James EA, Taplitz R, et al.: **Memory T Cells in Latent Mycobacterium tuberculosis Infection Are Directed against Three Antigenic Islands and Largely Contained in a CXCR3+CCR6+ Th1 Subset [Internet]**. *PLoS Pathog.* 2013, **9**:e1003130.
54. Brodin P, Majlessi L, Marsollier L, De Jonge MI, Bottai D, Demangel C, Hinds J, Neyrolles O, Butcher PD, Leclerc C, et al.: **Dissection of ESAT-6 system 1 of Mycobacterium tuberculosis and impact on immunogenicity and virulence**. *Infect. Immun.* 2006, **74**:88–98.

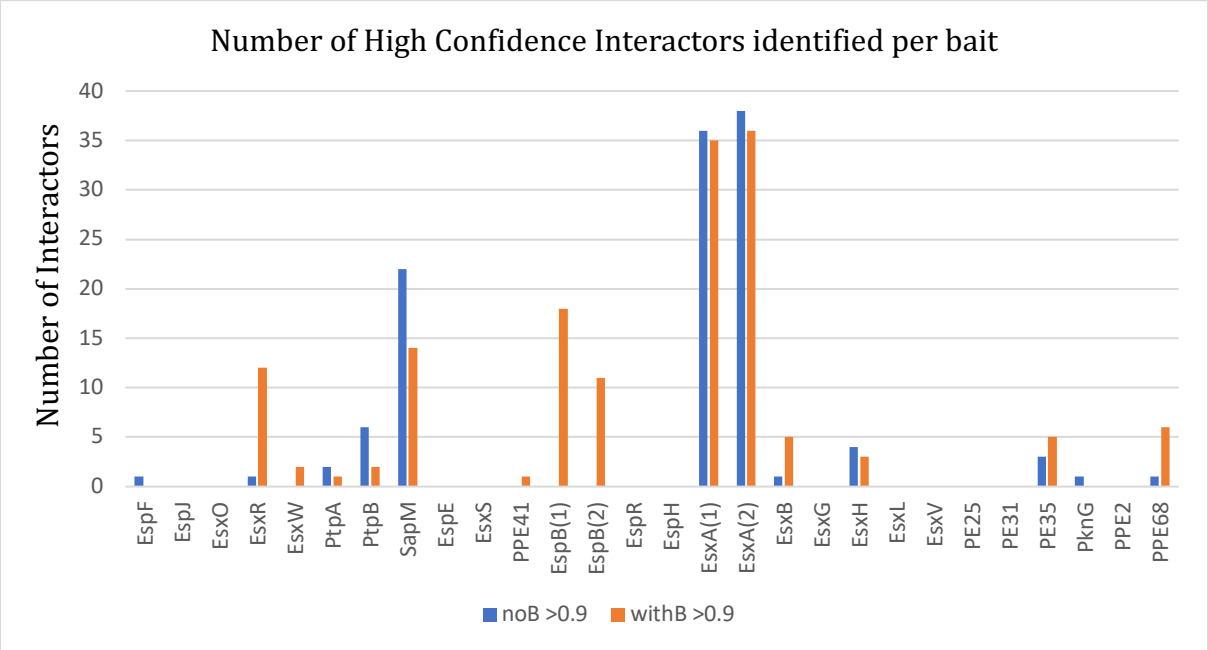
55. Ohol YM, Goetz DH, Chan K, Shiloh MU, Craik CS, Cox JS: **Mycobacterium tuberculosis MycP1 Protease Plays a Dual Role in Regulation of ESX-1 Secretion and Virulence**. *Cell Host Microbe* 2010, **7**:210–220.
56. Mohanty S, Dal Molin M, Ganguli G, Padhi A, Jena P, Selchow P, Sengupta S, Meuli M, Sander P, Sonawane A: **Mycobacterium tuberculosis EsxO (Rv2346c) promotes bacillary survival by inducing oxidative stress mediated genomic instability in macrophages**. *Tuberculosis* 2016, **96**:44–57.
57. Manzanillo PS, Shiloh MU, Portnoy DA, Cox JS: **Mycobacterium tuberculosis activates the DNA-dependent cytosolic surveillance pathway within macrophages [Internet]**. *Cell Host Microbe* 2012, **11**:469–480.
58. Novikov A, Cardone M, Thompson R, Shenderov K, Kirschman KD, Mayer-Barber KD, Myers TG, Rabin RL, Trinchieri G, Sher A, et al.: **Mycobacterium tuberculosis Triggers Host Type I IFN Signaling To Regulate IL-1β Production in Human Macrophages [Internet]**. *J. Immunol.* 2011, **187**:2540–2547.
59. Simeone R, Bobard A, Lippmann J, Bitter W, Majlessi L, Brosch R, Enninga J: **Phagosomal rupture by Mycobacterium tuberculosis results in toxicity and host cell death [Internet]**. *PLoS Pathog.* 2012, **8**:e1002507.
60. Bigenzahn JW, Fauster A, Rebsamen M, Kandasamy RK, Scorzoni S, Vladimer GI, Müller AC, Gstaiger M, Zuber J, Bennett KL, et al.: **An Inducible Retroviral Expression System for Tandem Affinity Purification Mass-Spectrometry-Based Proteomics Identifies Mixed Lineage Kinase Domain-like Protein (MLKL) as an Heat Shock Protein 90 (HSP90) Client**. *Mol. Cell. Proteomics* 2016, **15**:1139–1150.
61. **Main [Internet]**. [date unknown], [no volume].
62. Davis ZH, Verschueren E, Jang GM, Kleffman K, Johnson JR, Park J, Von Dollen J, Maher MC, Johnson T, Newton W, et al.: **Global Mapping of Herpesvirus-Host Protein Complexes Reveals a Transcription Strategy for Late Genes [Internet]**. *Mol. Cell* 2015, **57**:349–360.
63. Dyer MD, Murali TM, Sobral BW, Wille A, Buhlmann P: **The Landscape of Human Proteins Interacting with Viruses and Other Pathogens [Internet]**. *PLoS Pathog.* 2008, **4**:e32.
64. Durmuş Tekir S, Cakir T, Ulgen KÖ: **Infection Strategies of Bacterial and Viral Pathogens through Pathogen-Human Protein-Protein Interactions. [Internet]**. *Front. Microbiol.* 2012, **3**:46.
65. Wuchty S, Siwo G, Ferdig MT: **Viral Organization of Human Proteins [Internet]**. *PLoS One* 2010, **5**:e11796.
66. Dyer MD, Neff C, Dufford M, Rivera CG, Shattuck D, Bassaganya-Riera J, Murali TM, Sobral BW: **The Human-Bacterial Pathogen Protein Interaction Networks of Bacillus anthracis, Francisella tularensis, and Yersinia pestis [Internet]**. *PLoS One* 2010, **5**:e12089.
67. Radhakrishnan GK, Splitter GA: **Modulation of host microtubule dynamics by pathogenic bacteria**. *Biomol. Concepts* 2012, **3**:571–580.
68. Angot A, Vergunst A, Genin S, Peeters N: **Exploitation of Eukaryotic Ubiquitin Signaling Pathways by Effectors Translocated by Bacterial Type III and Type IV Secretion Systems [Internet]**. *PLoS Pathog.* 2007, **3**:e3.
69. Kalam H, Fontana MF, Kumar D: **Alternate splicing of transcripts shape macrophage response to Mycobacterium tuberculosis infection [Internet]**. *PLOS Pathog.* 2017, **13**:e1006236.
70. Lúcia Moreira-Teixeira, Mayer-Barber K, Sher A, O'Garra A: **Type I interferons in tuberculosis: Foe and occasionally friend**. *J. Exp. Med.* 2018, **215**:1273–1285.
71. Ruepp A, Waegele B, Lechner M, Brauner B, Dunger-Kaltenbach I, Fobo G, Frishman G, Montrone C, Mewes HW: **CORUM: The comprehensive resource of mammalian protein complexes-2009 [Internet]**. *Nucleic Acids Res.* 2009, **38**:D497-501.
72. Bach H, Papavinasasundaram KG, Wong D, Hmama Z, Av-Gay Y: **Mycobacterium tuberculosis Virulence Is Mediated by PtpA Dephosphorylation of Human Vacuolar Protein Sorting 33B [Internet]**. *Cell Host Microbe* 2008, **3**:316–322.
73. Gutierrez MG: **Functional role(s) of phagosomal Rab GTPases**. *Small GTPases* 2013, **4**:148.
74. Vieira O V., Bucci C, Harrison RE, Trimble WS, Lanzetti L, Gruenberg J, Schreiber AD, Stahl PD, Grinstein S: **Modulation of Rab5 and Rab7 Recruitment to Phagosomes by Phosphatidylinositol 3-Kinase**. *Mol. Cell. Biol.* 2003, **23**:2501–2514.
75. Chandra P, Ghanwat S, Matta SK, Yadav SS, Mehta M, Siddiqui Z, Singh A, Kumar D: **Mycobacterium tuberculosis Inhibits RAB7 Recruitment to Selectively Modulate Autophagy Flux in Macrophages [Internet]**. *Sci. Rep.* 2015, **5**:16320.
76. Kumar D, Nath L, Kamal MA, Varshney A, Jain A, Singh S, Rao KVS: **Genome-wide Analysis of the**

Host Intracellular Network that Regulates Survival of Mycobacterium tuberculosis [Internet]. *Cell* 2010, **140**:731–743.
77. Kumar M, Kundu M, Correspondence JB: **MicroRNA let-7 Modulates the Immune Response to Mycobacterium tuberculosis Infection via Control of A20, an Inhibitor of the NF-κB Pathway [Internet]**. *Cell Host Microbe* 2015, **17**:345–356.
78. Yamada H, Mizuno S, Reza-Gholizadeh M, Sugawara I: **Relative importance of NF-κB p50 in mycobacterial infection**. *Infect. Immun.* 2001, **69**:7100–7105.
79. Bai X, Feldman NE, Chmura K, Ovrutsky AR, Su WL, Griffin L, Pyeon D, McGibney MT, Strand MJ, Numata M, et al.: **Inhibition of Nuclear Factor-Kappa B Activation Decreases Survival of Mycobacterium tuberculosis in Human Macrophages**. *PLoS One* 2013, **8**.

4.6 Supplementary information



Supplementary figure 1 This graph shows the normalized bait expression levels based on the sum of the spectral counts per protein divided by their molecular weight and multiplied by the average molecular weight of all bacterial proteins tested. Blue bars represent normalized expression in the absence of bortezomib, and the orange bars represent the generally increased expression levels upon the addition of the protease inhibitor.



Supplementary figure 2 This graph shows the number of host protein interactors filtered for a Saint Probability scores greater than 0.9. Bars in blue represent the interactors in the absence of bortezomib, in orange the interactors identified upon the addition of bortezomib.

Chapter 5

Differential complex-centric proteome profiling in two days

Contributions

This thesis chapter is the result of a highly collaborative project:
I.B. developed the computational framework and performed the CCprofiler analysis. I.B wrote the materials and methods section dedicated to CCprofiler
C.N. performed the THP-1 experiments, the SEC, contributed to the data analysis and wrote this chapter
C.M. developed and implemented the Evosep LC system, the SEC, and performed the LC-MS/MS measurements and wrote parts of the materials and methods section
F.F., F.A. and F.U. developed the FASP-aided protocol
M.H. created the web-based visualization tool
B.C. and R.A. supervised the project are provided critical input

Aimed to be adapted and published as

Isabell Bludau *, Charlotte Nicod *, Claudia Martelli *, Fabian Frommelt, Fossati Andrea, Federico Uliana, Moritz Heusel, Ben Collins and Ruedi Aebersold.
Differential complex-centric proteome profiling in two days

5.0 Abstract

Protein-protein interactions and protein complexes, together constituting the complexome, build the primary functional modules of the cell. To respond to stimulations or perturbations, the complexomes undergo qualitative and quantitative changes. To understand the functional state of biological samples, it is thus important to capture the dynamic rearrangements of the complexomes. We recently presented a method based on Size Exclusion Chromatography coupled to DIA/SWATH-MS which enables the global identification and relative quantification of proteome-wide organization. Although achievable, proteome-wide studies based on such Co-Fractionation paired with Mass Spectrometry are limited in feasibility due to extremely lengthy workflows. Here, we present a highly optimized and integrated workflow that enables the identification and quantification of global complexomes for multiple samples in two days. The workflow entails a Filter Aided Sample Preparation step, 21-minute long chromatographic gradients proceeding the acquisition using DIA/SWATH-MS and an automated software to quantify changes in the global complexomes. We then apply the workflow to study the global rearrangements of the THP-1 complexomes undergoing a macrophage differentiation and a mimicked bacterial infection.

5.1 Introduction

Protein-protein interactions (PPI) and in extension protein complexes, together embodying the complexome, constitute the primary functional modules of the cell [1] and play crucial roles in nearly all cellular processes. Indeed, most molecular functions are carried out by multiple interacting proteins rather than individual proteins. It is estimated that the majority of the total proteome mass is assembled in stable macromolecular protein complexes [2]. Moreover, proteins may dynamically change their interaction networks either qualitatively or quantitatively in order to respond to specific environmental stimuli. It is thus crucial to not only capture context-dependent snapshots of the global proteome composition, but also its organization into protein complexes to understand the underlying functional states of given cells.

The mostly commonly used methods to map protein-protein interactions are based on affinity-purifications coupled to mass spectrometry (AP-MS). They consist of fusing an affinity tag to the carboxyl or amine end to the protein of interest and overexpressing in the desired cellular environment [3]. After the co-expression, the tagged bait is co-purified along with its direct and indirect interacting proteins. The coelutions are then digested into peptides and analyzed via LC-MS/MS. Although highly sensitive and robust, the AP-MS entails a few hurdles and limitations.

First, fusing an affinity tag to the protein of interest might affect its structure or impair certain interactions[4]. Secondly, AP-MS requires the genetic engineering for each cell line expressing the tagged protein, limiting its applicability to large scale studies thus not suitable to monitor dynamic global changes of proteome organizations across multiple biological samples. Additionally, the need for transgenic cell lines can prevent its usability to study biological materials where genetic engineering isn't available. Thirdly, it provides no stochiometric information nor does it provide topological information unless coupled with additional baits [5].

Methods based on the principles of co-fractionation of native proteome extractions coupled to mass-spectrometry [6–8] (CoFrac-MS) are promising as they could overcome the limitations associated to AP-MS based approaches. CoFrac-MS methods rely on the biochemical fractionations of cell lysates to separate protein complexes based on the chosen physio-chemical properties, and the proteins contained in every fraction are subsequently identified via LC-MS/MS. In principle, they can capture a quantitative snapshot of the proteome-wide organization of any given biological sample circumventing the need to genetic engineering. Although achievable, current CoFrac-MS biochemical workflows require high amounts of primary biological material, generally rely on cell based or peptide labelling strategies to achieve quantitative accuracy and require weeks of acquisition time. To then detect protein complexes from such CoFrac-MS datasets, the most commonly used computational approaches rely on the *de novo* strategies, where protein complexes are inferred by the co-elution profiles of two or more proteins as evidence that they interact [7,9–13]. However, a considerable drawback from such *do novo* strategies is the propensity to report erroneous assignment of co-eluting proteins to complexes, due to the combination of low separative resolution of fractionation techniques and high number of eluting analytes [2].

To overcome challenges caused by *de novo* protein interaction detection strategies, we recently presented a complex-centric proteome profiling by SEC-SWATH-MS for investigating the modularity of the protein complex landscape in a single condition [2]. The workflow is based on the native extraction of protein complexes from a mild lysis of cells which are subsequently separated by high-resolution Size Exclusion Chromatography (SEC) according to their hydrodynamic radius. The proteins contained in each fractions are then identified and quantified via SWATH-MS [14], a Data Independent Acquisition (DIA) scheme enabling a highly reproducible, robust and sensitive quantification of peptides across multiple samples [15] circumventing the need for cell-based or peptide labelling strategies. The quantitative, peptide-level SEC-SWATH-MS data is subsequently analyzed by a targeted, complex-centric strategy that is implemented in the computational framework CCprofiler. Here, prior protein connectivity information from

protein complex or PPI databases such as Corum [16] or String [17] are employed to generate target protein complex queries. CCprofiler subsequently performs automated detection and quantification of these target complexes, while controlling the error-rate by using a target-decoy based statistical model. This enables the detection of hundreds of protein complexes at unprecedented throughput, resolution and specificity. However, both the biochemical sample preparation and the computation workflow were still suboptimal to determine quantitative differences in protein assembly states across multiple biological conditions and replicates. Despite these challenges, we recently provided insights into changes in protein assembly states across two different cell cycle stages [18], although at a high cost of sample processing and acquisition time.

Here, we present a highly optimized workflow that addresses the prevailing needs for quantitatively assessing protein complex assembly states on a proteome wide scale and across multiple conditions, overcoming previous limitations of CoFrac-MS analyses. We exemplify the power of our workflow by investigating rearrangements in the protein complex landscape of untreated THP-1 human monocytic precursor cells, when undergoing a PMA-induced differentiation into a macrophage-like cells [19], and upon lipopolysaccharide (LPS) [20] (3 biological replicates of 3 distinct conditions, Figure 1A).

5.2 Results – workflow structure

The integrated high throughput workflow starts from a mild lysis and native proteome extraction where the protein complexes undergo a SEC fractionation requiring less than a gram per run. The proteins contained in each fraction undergo parallelized sample preparations including proteolytic digestions using 96-well Filter-Aided Sample Preparation (FASP) [21] plates to ensure robustness and comparability while significantly reducing sample handling times. The peptides are then loaded directly on EvoSep tips, omitting a reverse phase-based desalting step which will take place in the HLPC system. Next, the acquisition strategy comprises a 21-minute (24 minute total run time) short LC gradient using an EvoSep One HPLC system that relies on embedded gradients to reduce overhead, altogether enabling the acquisition nearly one SEC dimension of 60 fractions per day while minimizing the loss in sensitivity [22] (Figure 1A, Supplementary Figures 1&2). The content of each fraction is then acquired in DIA/SWATH mode using parameters optimized for short gradient analysis (Supplementary Figures 1&2), to assure deep coverage and quantitative accuracy across the multiple samples. The computational part of the workflow comprises an upgrade of our software CCprofiler [2] to systematically and

automatically investigate changes in proteome assembly across multiple conditions or cellular states (Figure 1B). Importantly, CCprofiler includes several pre-processing functionalities to align SEC profiles, to compute missing values and to normalize intensities between replicates and conditions (described in materials and methods, visualized in supplementary figures). The extended CCprofiler version further enables the qualitative and quantitative detections of three different aspects of proteome organization.

The first differential analysis module in CCprofiler is tailored towards detecting proteins that differ in their global assembly state, meaning that the relative distribution between monomeric and assembled states is different across the conditions (Figure 1B panel 4.I). For this analysis, we first exploit the log-linear relationship between the SEC elution fractions and their apparent molecular weight (supplementary figure), thereby enabling the assignment of a monomeric and assembled SEC elution range specific for each protein. The fraction of observed protein mass in the assembled SEC elution range is represented by the Assembled Mass Fraction (AMF) (c.f. materials and methods). The differential module assesses whether a protein undergoes a significant change in AMF across the different conditions, meaning that it changes from assembled states to monomeric states or vice versa.

Secondly, CCprofiler contains a protein-centric analysis module and evaluates the number of distinct assembly states each protein is observed in. We define a distinct assembly state as a distinctly resolved peptide co-elution peak group of a protein along the SEC chromatographic dimension, referred to as 'protein feature'. Recently, we extended the protein-centric analysis to quantitatively compare protein features across different conditions [18]. In contrast to a standard differential protein expression analysis, abundance fold-changes and p-values are provided for each distinct protein feature, thereby capturing not only changes in overall protein expression, but abundance changes of specific assembly states. In addition to the feature-specific differential analysis, global differential assessment is performed by comparing integrated intensities across the entire fractionation dimension instead of restricting the analysis to a feature-specific range. The same strategies as for feature-specific estimation of log2-fold-changes and p-values are performed. Additionally, we provide the opportunity to compare the relative distribution of protein mass across these various assembly states (Figure 1B panel 4.II), represented by a relative Feature-specific Mass Fraction (FMF). Here, a change in FMF across conditions indicates that the protein changes its relative distribution across different assembly states. The protein-centric differential analysis yields a fine-grained view of individual assembly states of each protein but also enables more global assessments of the overall degree of higher order assembly observed in each biological condition.

Finally, CCprofiler quantitatively compares the abundances and compositions of protein complexes across different biological conditions in an automated and error-controlled manner (Figure 1B panel 4.III). The complex-centric analysis module first relies on prior protein connectivity information to query the data in a targeted fashion and to extract protein complexes based on their co-elution profiles under a controlled FDR (materials and methods). Then, CCprofiler carries out a differential analysis step by comparing the signal intensity for each protein complex feature across all pairwise biological conditions. This analysis enables the consistent detection and quantitative comparison of hundreds of protein complexes across different biological conditions at unprecedented sensitivity and specificity.

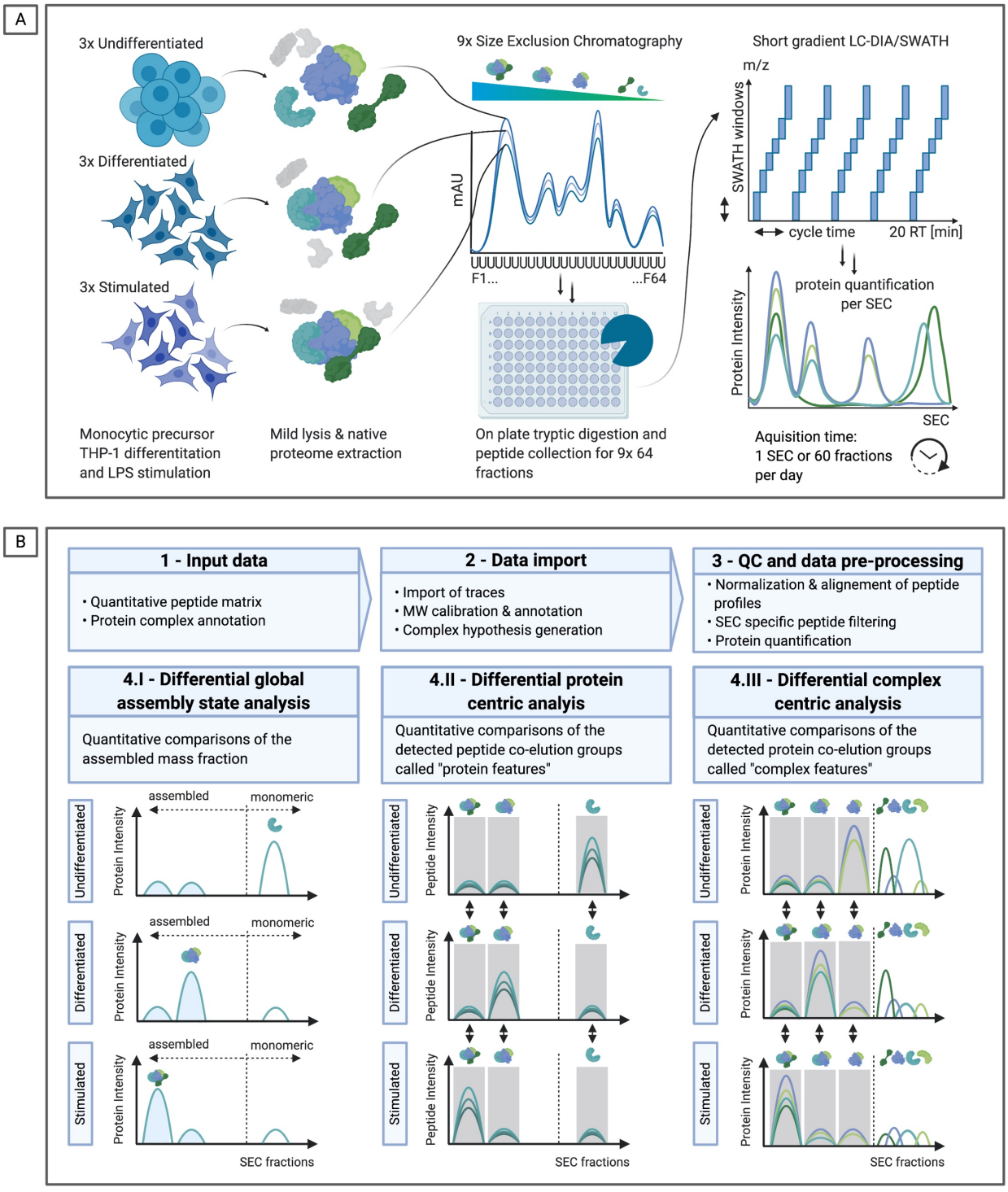


Figure 1 Panel A represents the main steps in the sample processing workflow and the three biological conditions used in biological triplicates as proof-of-concept. The first biological conditions entailed untreated and undifferentiated monocytic precursor THP-1 cells. The second condition were THP-1 cells subjected to PMA-induced differentiation into macrophages. The third condition was an LPS-stimulation of the differentiated THP-1 cells. Each biological condition was injected with biological triplicates in the SEC fractionating each sample into 64 fractions. The collected fractions were processed to peptides using a Filter Aided Sample Preparation (FASP) protocol. The peptides were loaded onto Evosep tips without any desalting steps and added to the autosampler coupled to a TripleTOF® 6600 QTOF mass spectrometer (Sciex). Peptides were analyzed upon 21-minute long liquid chromatographic gradients in DIA mode,

together enabling the acquisition of 60 samples per day. Panel B represents the computational workflow embedded in CCprofiler. The CCprofiler workflow necessitates a quantitative peptide matrix along with its protein annotation. Second, it imports the annotated MW calibration curved for the SEC and the protein complex hypotheses. In a third step, it normalizes and aligns the runs to make them comparable and filters out peptides based on their identifications in consecutive fractions. Finally, it quantifies the proteins based on the peptide traces, representing the filtered peptides per SEC run along the chromatographic dimension. The first differential analysis module consists of the differential global assembly state analysis, which reports for each protein the relative assembled fraction compared to their monomeric state. The protein-centric analysis, in panel B 4.II reports quantitative comparisons of all detected peptide co-elution groups called protein features. Panel B 4.III depicts the last CCprofiler module, namely the differential complex centric analysis. This module performs pair-wise quantitative comparisons of the detected protein co-elution groups called "complex features" between all the biological conditions.

5.3 Results - Differential complexomes analysis of stimulated THP-1 cells

We applied the highly optimized workflow for fast and differential complexome profiling by SEC-SWATH-MS to compare the underlying complexomes associated with the phenotypes of untreated THP-1 human monocytic precursor cells, when undergoing a PMA-induced differentiation into a macrophage-like cells [19], and a lipopolysaccharide (LPS) stimulation to elicit an immune-like response [20]. Upon induction with PMA, the suspension THP-1 cells which are often used as a model to study human macrophage functions, undergo a differentiation into adherent macrophage like cells with phagocytic phenotype and lose their proliferative capabilities [23].

The mild cellular extracts from all 3 biological conditions of THP-1 cells were injected with 3 biological replicates in the SEC, amounting to 9 SEC runs of 64 fractions each, leading to a total of 576 MS runs injected within 9.5 days. The resulting dataset was analyzed using the OpenSWATH computational pipeline using a THP-1 specific spectral library. The quantitative output matrix contained 52,285 proteoytpic peptides inferring the detection of 5,889 unique proteins across the dataset at a 10% run-specific peak-group FDR, 5% global peptide FDR and 5% global protein FDR. Within CCprofiler for quality control, we first assessed the consistency and comparability of the 9 fractionations by performing pair-wise alignments at the peptide-level traces and calculated the global correlation amongst all matching peptides (supplementary material). To be able to quantitatively compare the respective SEC runs, we then normalized the intensities using a cyclic loess method (supplementary material & supplementary figure). To increase the confidence downstream, we further filtered down the number of peptides using the CCprofiler framework by carrying over multiple pre-processing and filtering steps. These included filtering out peptides without 2 consecutive identifications in neighboring fractions, to only keep proteins supported

by more than one peptide, and that the remaining proteins are required to be supported by at least 2 high-correlating sibling peptides (ref materials and methods & supplementary figure). After the filtering steps, the number average number of identified proteins per SEC were reduced from 5,736 down to 4013. The resulting concatenated quantitative peptide map across all 9 SECs showed reliable elution features for 3335 proteins features at a 5% FDR cutoff, amongst which 27.32% were detected as monomers only whilst 2424 (72.68%) had at least one elution feature and 1389 had multiple suggesting their contributions to more than one complex (figure 2). To generate a comprehensive input hypothesis-set for the complex-centric analysis, we merged the CORUM database enriched for stable protein complexes supplemented with a partitioned version of the String database using the ClusterONE algorithm [24], originally created to detect potentially overlapping protein complexes from PPI datasets.

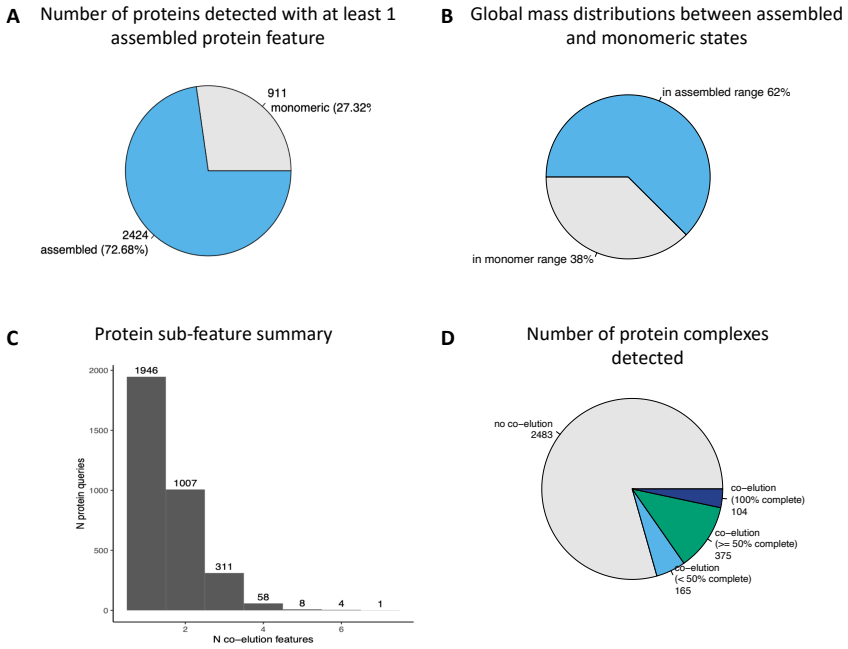


Figure 2 Panel A depict the proportion of protein having only a protein feature detected in the monomeric range compared to the proportion of proteins having at least one protein feature in the assembled range. Panel B shows the global proteome mass in assembly states compared to the proteome mass present in the monomeric range. Panel C shows the distribution of the number of protein sub-features for all the proteins detected. Panel D shows the number of detected protein complexes with 100% of the subunits detected, more than 50% or less than half of the subunits belonging to the complex present in the complex-feature.

Together this approach resulted into 3127 complex hypotheses, from which 104 were fully detected, 375 were detected with at least 50% of the subunits present and 165 were identified from less than 50% of the subunits present, all with a 5% FDR at the complex-detect level. As predicted, the three different quantitative CCprofiler modules demonstrated larger changes between the undifferentiated cells and the PMA-induced differentiated cells compared to the LPS-

stimulated cells, explained by a larger phenotypic change including the transformation into adherent cells and loss of proliferative properties leading to the cell-cycle arrest (figure 3).

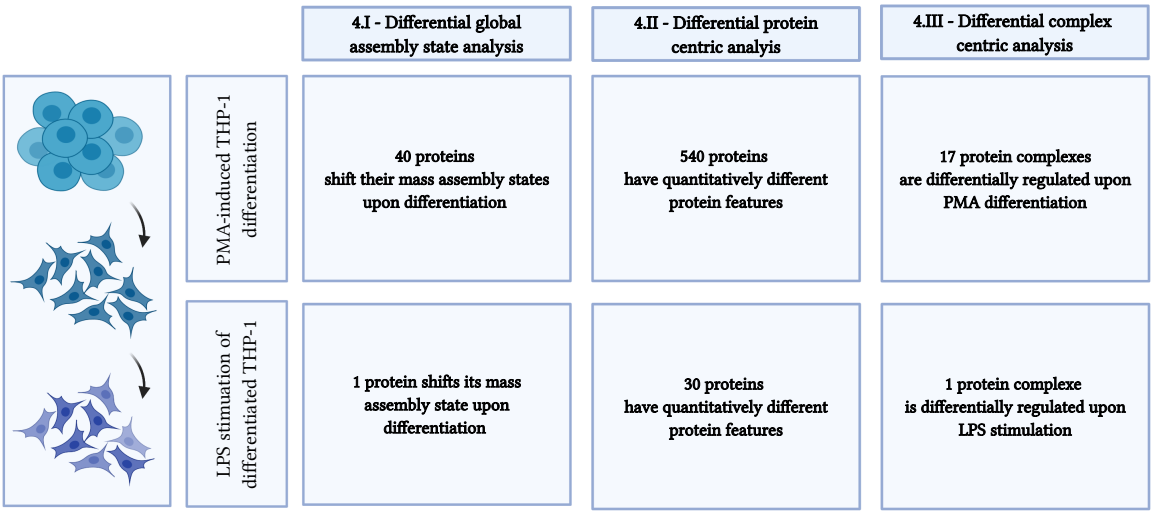


Figure 3 Here depicted are the number of (4.I) proteins undergoing shifts in their assembly states, (4.II) the number of proteins having different protein features or (4.III) the number of protein complexes that undergo qualitative or quantitative rearrangements upon the perturbations.

Over the entire dataset, 62% of the global proteome mass was estimated to be in assembled states and was consistent in all 9 SEC runs. 60 proteins were detected to undergo significant shifts in their Assembled Mass Fractions between all pair-wise comparisons, where the highest number of shifts in the AMF was retrieved when comparing undifferentiated to stimulated cells. From the 40 proteins having a significant different assembly state upon THP-1 differentiation, enriched for being members of the cell adhesion and extracellular GO Biological Process (Benjamini pval < 0.05), 35 shifted towards higher assembly states upon differentiation (figure 4). The LPS stimulation alone resulted in only 1 protein with a significant shift it is assembly states. However, when comparing undifferentiated to LPS stimulated cells, 42 proteins were retrieved and enriched in Type I interferon signaling pathway, negative regulation of viral genome replication, defense response to virus and innate immune responses, coherent with the stimulus (figure 4). Amongst the 60 proteins undergoing a shift in their assembly, the serine threonine kinase 4 was retrieved for shifting between undifferentiated to LPS-stimulated cells. Although the differentiated cells already started to mark a shift in the STK4 assembly state, it was only significant when comparing the undifferentiated to the stimulated cells. When summing up the peptide intensities belonging to STK4 across the SEC using the global protein-centric analysis approach of the protein-centric module (materials and methods), STK4 did not have a significant fold change in the apparent expression levels. Together this suggests that even though the STK4 expression levels are not affected by the differentiation and LPS stimulation, we could

hypothesize that STK4 changes its interactome and potentially its activity upon both stimulations by shifting towards higher assembly states.

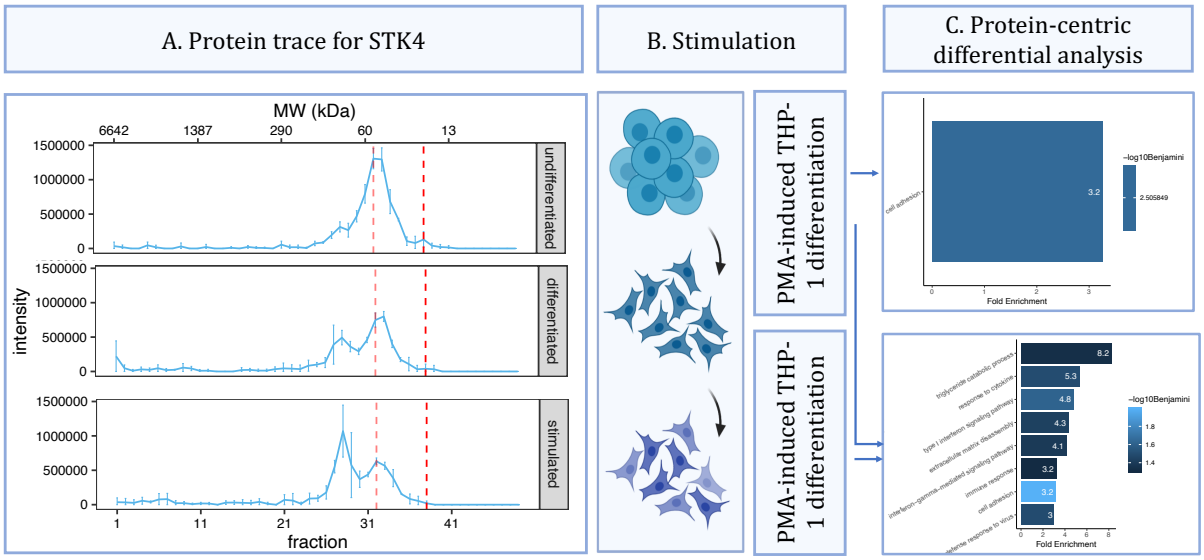


Figure 4 Panel A shows the protein trace belonging to STK4, the serine threonine 4 and how its assembled mass fraction is affected by the 2 cell perturbations. The red dotted line is placed at the expected monomeric Molecular Weight and the pink dotted line at the dimeric MW and holds the purpose of the cutoff to distinguish assembly states. The blue vertical bars represent the standard deviation of quantitation of all 3 biological replications. Panel C shows the gene ontology results for Biological Processes significantly enriched (Benjamini pvalue <0.05) in proteins with shifts in their assembly states upon the PMA-induced differentiation (upper) and upon differentiation and LPS stimulation (lower).

The protein-centric module encompasses two analyses. The first consist of calculating the global protein intensity by summing all the peptides belonging to each protein across the SEC and calculate fold (materials and methods) and can considered as a proxy for expression levels. The second analysis aims at finding quantitative or qualitative differences at the local protein-feature levels. As expected, global protein intensities are more sensitive, although potentially less informative about proteins' interactomes, in terms of differential changes as they are less specific than relative local protein-feature comparisons.

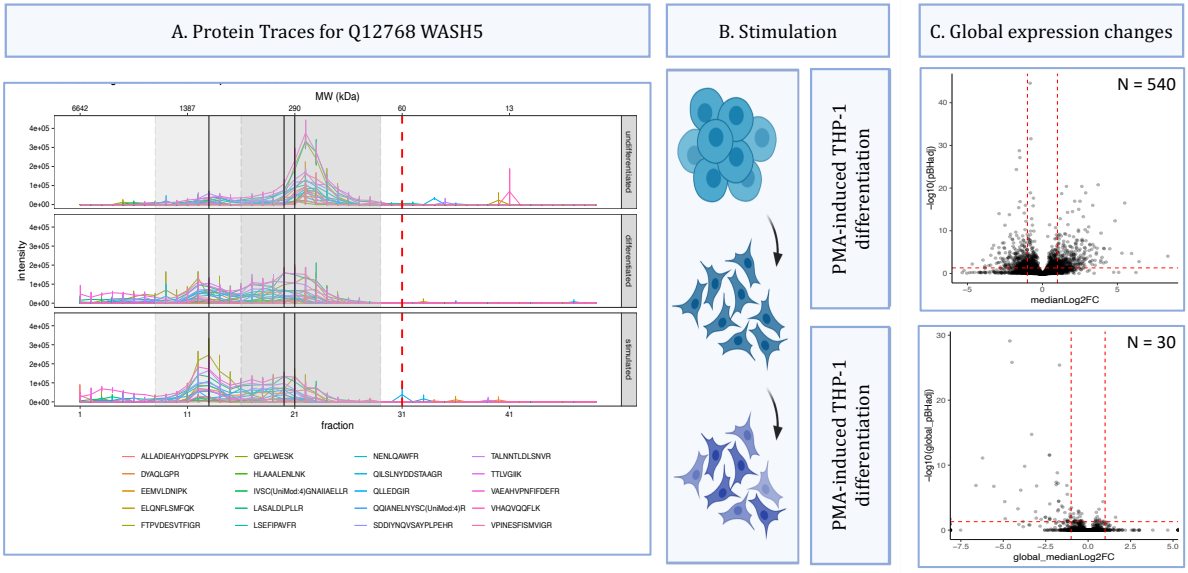


Figure 5 Panel A shows the protein trace belonging to WASH5, a member of the WASH core complex and how its relative local protein-features are affect upon both differentiations. Panel C entail volcano plots to visualize the number of proteins affected at their global quantitative levels upon PMA-induced differentiation (upper) and upon differentiation and LPS stimulation (lower) with the absolute fold change cutoff set at 1 and the Benjamini pvalue < 0.05. *N* equals the number of significantly affected proteins at

By applying our second module based on protein-centric analysis, 540 out of the 3813 proteins having protein features within the higher assembly ranges were identified to have significant fold changes (|FC| > 1, pvalue < 0.05) at their global intensity levels upon THP-1 cell differentiation and 30 in response to the LPS stimulation. Consistent with the biological perturbation, the LPS stimulations affected global protein intensities enriched in Type I interferon signaling and responses to viruses, whereas the PMA induced differentiation affected proteins involved cell adhesion and extracellular matrix organization (Benjamini pval < 0.05). The protein-centric analysis yielded a total of 156 proteins that had quantitatively altered local protein features. Notable examples include 3 WASH proteins (Figure 5 A) where they all shift from an assembly state of an estimated 290 kDA complex towards higher assembly of 1387 kDa.

When applying the third CCprofiler module dedicated to the protein-complex level, we detected over the entire dataset 644 complexes. However, after removing protein complexes that were only supported by the presence of less than half of its annotated subunits and by collapsing the remaining protein complexes to retain non-redundant interacting proteins, we arrived at a total of 381 uniquely identified protein complexes. When applying the differential analysis, 17 complexes were differentially regulated (materials and methods) upon PMA-differentiation, 1

upon LPS-stimulation and 19 when comparing the monocytic precursor THP-1 with the LPS stimulated macrophages.

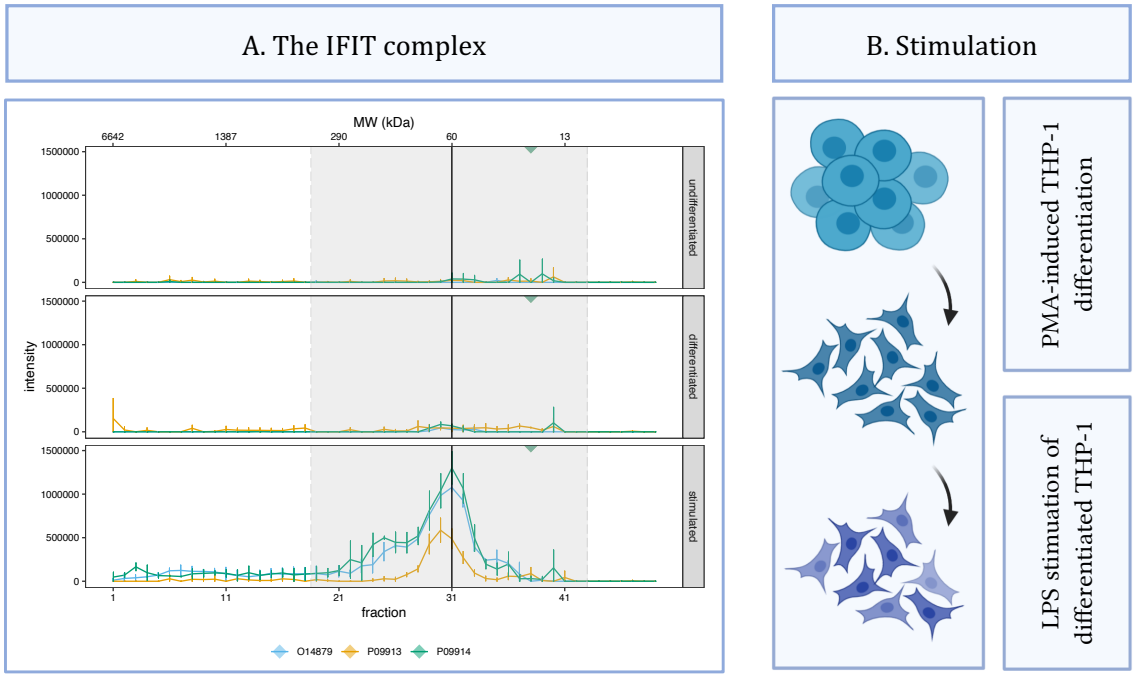


Figure 6 The IFIT complex composed of the IFIT1, IFIT2 and IFIT3 subunits forms upon LPS stimulation.

From the complex-centric analysis, one protein-complex was retrieved upon setting quantitative thresholds for their detection between conditions (materials and methods) to be highly up-regulated in LPS-stimulated macrophages. This protein complex, named the IFIT complex, is composed of the Interferon-induced protein with tetratricopeptide repeats 1-3 and is a known interferon-induced antiviral response [25]. Indeed, LPS stimulations in macrophages increases IFIT expression levels, in order to enhance the secretion of proinflammatory cytokines including TNF alpha and IL-6. From the protein traces, P09914 and O14879 (IFIT 1 and IFIT3 respectively) but not IFIT2 seem to interact with additional proteins and form a higher molecular weight complex. However, our complex hypothesis databases did not contain heavier protein complexes interacting with these proteins. In order to render the data easily accessible and viewed in depth, and enable manual query of community-based testing of novel putative interacting proteins supported by the presence of co-elution profiles, we have uploaded the data on the existing publicly available tool SECexplorer-cc [18] (https://sec-explorer.shinyapps.io/TestTestTest/). This online tool provides the opportunity to manually query the SEC profiles of our 3 biological THP-1 conditions by providing an interactive viewing. It enables the manual query for locally co-eluting proteins to potentially identify *de novo* interactions and to visualize the results from the 3 differential CCprofiler modules.

5.4 Discussion

In this study we presented a highly optimized workflow to enable the deep profiling of complexomes from multiple samples within a feasible time-frame. So far, the study of multi-conditions complexome profiling by CoFrac-MS was limited by 3 bottlenecks in the workflows. First, the sample handling and processing was highly time consuming and complex, potentially leading to high variability. Secondly the classical MS acquisition strategies relying on 60-90-minute-long gradients were requiring weeks of acquisition, highly limiting its feasibility as it would monopolize mass spectrometers for an extended period of time. Thirdly, there was a need to have a computational framework that would enable quantitative comparisons between multiple conditions in an automated fashion, while maintaining the error rates by using target-decoy strategy [7,9–13]. By combining FASP based protocols, avoiding manual reverse-phase based peptide cleanup steps [21], including 21-minute long liquid chromatographic gradients with SWATH-MS [22] and providing an extension to CCprofiler R package, we could reduce the processing and acquisition timeframes of 9 SEC runs of 64 fractions each, leading to a total of 576 MS runs from an estimated 40 days down to 10.5 days (including 9.5 days of measurement time).

By adding a clusterOne [24]partition of the String database [26], we could expand our hypothesis space beyond core complexes annotated in Corum [16] and thus allow a higher retrieval of protein complexes from our SEC profiles, exemplified by the inclusion of the IFIT complex originally absent from the Corum core complex annotation.

We then applied the 3 differential module analysis to study the complexomes rearrangements associated with PMA-induced THP-1 differentiation and subsequent LPS-stimulation. This enabled us to quantify the responses from both stimuli and to detect changes in the proteome organization otherwise invisible when using whole-lysate based proteomics. The first module, consisting of analyzing the assembled mass fraction, that most proteome rearrangement upon PMA differentiation entailed of proteins shifting towards higher assembly masses, so potentially functionally distinct, upon PMA induced differentiation. However, on the global proteome scale, none of the stimulations yielded shifts in the global assembly states. The global quantitative protein-centric module was the most sensitive in detecting stimulation-caused protein rearrangements, as it calculates global protein intensities and does not rely on fold changes between local protein-features. Relative changes in local protein-features, although less frequent, are potentially more informative as they reveal changes in the stochiometric assembly states of given proteins. Lastly, the complex-centric module is most likely the least sensitive as it requires

prior complex knowledge and feature-finding space, but retrieves valuable information as it calculates relative amounts of active protein complexes forms.

We predict that along with the rapid improvements in sensitivities and implementations from novel mass-spectrometry based approaches [27], and together with improvements in separation resolutions of SEC, that deep-profiling of complexomes will increase in resolution and enable a profound understanding on the proteome organization and highlight underlying the plasticity of biological samples in responses to perturbations.

5.5 Materials and methods

Cell culture

The human monocytic cell line THP-1 (LGC, ATCC-TIB-202) was cultured and expanded in RPMI 1640 media (Gibco, 61870-010) supplemented with 10% FCS (BioConcept, 2-01F00-I) and 1% Penicillin/Streptomycin (Gibco, 15140-122) and kept at a confluency between 0.5 - 1.2 x 10^6 cells per ml at 37°C in a 5% CO_2 incubator. 1.5 x 10^6 THP-1 cells were differentiated when supplemented with 50 ng/mL PMA (Sigma, P1585) for 48h and when stated, the differentiation treatment included a 24-hour stimulation with 100 ng/mL LPS (Sigma, L2630). The suspension cells or differentiated adherent cells were washed with PBS (Gibco, 10010-023) and were sedimented in a pellet by centrifugation at 300g kept at 4°C. The cell pellets were immediately snap-frozen in liquid nitrogen.

Total proteome cell lysates

The proteins were extracted from the frozen cell pellets by lysing the cells with 1% SDC (Sigma, D6750) in HNN Buffe pH 7.8 (50 mM HEPES, 150 mM NaCl, 50 mM NaF, 200 µM Na_3VO_4, 1 mM PMSF, 1x Protease Inhibitors (Sigma, P8215), 1x Benzonase (Sigma, E1014)), and incubated for 5 minutes at room temperature. The lysates were centrifugated at 13'000g for 10 minutes to remove insoluble materials. The extracted proteins were reduced at 5mM TCEP for 30 minutes at 37°C while shaking at 500 rpm and subsequently alkylated in 10 mM Iodoacetamide for 30 minutes at 37°C. The proteins were precipitated overnight in 100% Acetone at -20°C and pelleted by a 30-minute centrifugation step at 4°C. The protein pellets were then resuspended in 1% SDC, 8M Urea in 0.1 M Ammonium bicarbonate and sonicated for 10 minutes. The proteins were diluted to 0.1 M ammonium bicarbonate and digested overnight with Trypsin (Promega, V5113) at 37°C with a protein-to-enzyme ratio of 50:1. The digestions were stopped with 50% TFA and the SDC was removed by two centrifugation steps of 10 minutes each at 16'000g. The peptides were desalted and cleaned-up using C18 columns (The Nest Group, #SEM SS18V) and were resuspended in 5% acetonitrile, 0.1% formic acid with iRT peptides (Biognosys, Ki-3002).

Sample preparation for library Generation

For the spectral library generation, a fraction of all samples was pooled together, dried using a vacuum centrifugation at 45°C and resuspended in Buffer A (20 mM ammonium formate, 0.1% ammonia solution, pH 10). 200 µg of peptides were injected into an Agilent Infinity 1260 (HP Degasser, Vial Sampler, Cap Pump) and 1290 (Thermostat, FC-µS) system and separated on a 25 cm long C18 reverse-phase column (YMC Triart) with 3µm particle size and 12 nm of pore size. The peptides were separated at a flow rate of 12µL per minute by a linear 56-min gradient from 5% to 35% Buffer B (20 mM ammonium formate, 0.1% ammonia solution, 90% acetonitrile in water, pH 10) against Buffer A (20 mM ammonium formate, 0.1% ammonia solution, pH 10) followed by a linear 4-min gradient from 35% to 90% Buffer B against Buffer A and 6 min at 90% Buffer B. The resulting 36 fractions were pooled into 12 samples. The buffer of the pooled samples was evaporated using vacuum centrifugation at 45 °C and the resulting 12 samples were resuspended in 2% ACN, 0.1% FA with iRT peptides (Biognosys, Ki-3002).

SEC protein complex extraction and fractionation

Protein complexes fractionation was performed as previously described [2]. THP-1 cells were thawed and lysed in mild conditions by homogenization with a lysis buffer composed by 0.5% NP-40 detergent and protease and phosphatase inhibitors (50 mM HEPES pH 7.5, 150 mM NaCl, 0.5% NP-40). Cell debris and membranes were removed by 15 minutes of ultracentrifugation (55,000×g, 4 °C) and the detergent was removed by 30 kDa molecular weight cut-off membrane and exchanged with the SEC buffer (50 mM HEPES pH 7.5, 150 mM NaCl). The samples were concentrated for a final protein concentration between 7-12 µg/µl. After 5 min of centrifugation at 16,900 ×g at 4 °C, the supernatant was directly injected to a Yarra-SEC-4000 column (300 × 7.8 mm, pore size 500 Å, particle size 3 µm, Phenomenex, CA, USA). 0.8 mg of native proteome extract (estimated by Pierce™ BCA Protein Assay Kit, Thermo Fisher Scientific, MA, USA) was injected for each SEC run at 4 °C with a flow rate of 500 µl/min, for a total chromatographic time of 30 min. Fraction collection was performed in the retention time window from 10 to 26 min, at 0.25 min per fraction, for a total of 64 fractions collected.

The calibration curve for SEC fractionation was obtained by running in the column a protein standard mix (Column Performance Check Standard, Aqueous SEC 1, AL0-3042, Phenomenex, CA, USA) before each sample injection.

Sample preparation for Mass Spectrometry analysis

Sample processing for bottom-up analysis of SEC fractions was peformed on a 96-well plate MWCO filters (AcroPep Advance Filter Plates for Ultrafiltration 1mL Omega 10K MWCO; Pall

Corporation, USA) [21]. Prior the usage, the filters are washed twice with 200 µl of water that was successively removed by centrifugation at 1800 g for 30 min. 64 fractions for each sample (total fraction volume 125 µl) were loaded and concentrated on the filters through centrifugation, until the complete removal of the SEC buffer.

Protein denaturation and reduction was obtained incubating the samples at 37 °C for 30 min with 5 mM of TCEP in 8M Urea/20 mM ammonium bicarbonate (AMBIC) (pH 8.8). Alkylation of cysteine residues was performed adding a final concentration of 50 mM IAA/20 mM AMBIC and incubating in the dark and room temperature for 1 h. After the reaction, the plates were centrifuged for removing the Urea buffer and washed for three times with 20 mM AMBIC. Protein digestion was carried out at 37 °C for 16 h, adding to each well 1 µg of trypsin (Promega, Switzerland) and 0.3 µg of Lysyl Endopeptidase (Mass Spectrometry grade, FUJIFILM Wako Pure Chemical Industries, Japan). The resulting peptides were collected by centrifugation and the further wash of the filters with water solution.

LC-MS analysis

LC-SWATH MS analysis of the peptide fractions was performed on Evosep One system (Evosep Biosystems, Denmark) [22] coupled to an AB Sciex TripleTOF 6,600 instrument (Sciex, MA, USA) equipped with a NanoSpray III ion source (Sciex). Due to the heterogeneity of the SEC protein fractions, we normalized the MS injections according to the fraction volume and to the amount of protein fractionated, corresponding to a total amount of 600 µg of proteins injected in the SEC column.

MS analysis sample loading

The samples are loaded in Evotips (Evosep Biosystems, Denmark), after resuspension in solvent A (0.1% FA water solution, Fisher Scientific AG, Switzerland) and the addition of iRTs peptides (Biognosys AG, Switzerland) in a ratio 1:100 for the retention time alignment requested for SWATH acquisition.

For the loading, the C18 stage tips (Evotips) were soaked with 100 µl of 2-propanol during the activation and the conditioning steps. The activation step consisted in the washing with 20 µl of solvent B (0.1 % FA in ACN, Fisher Scientific AG, Switzerland), followed by the conditioning with 20 µl of solvent A. Prior the sample loading step, 10 µl of solvent A is added on top of the tips, ensuring that the tips remain wet during the loading step. For each steps, the Evotips were centrifuged for 1 min at a speed of 700 g for the elution of the solvents.

The last step (i. e. washing step) was performed using 100 µl of solvent A, and the loaded tips are added with 200 µl of solvent A for preserving the samples during the entire injection of the batch.

Evosep-ESI-DIA-SWATH

The separation of peptides was performed selecting the "60 samples per day" method, consisting in 24 minutes of total cycle time, for 21 minutes of gradient length, 3 minutes of overhead time at a flow rate of 1 µl/min. A partial gradient is applied (0-35% solvent B) in order to elute the peptides from the Evotip by two couples of low pressure pumps. The peptides were then pushed in a C-18 nanoConnect LC column (8 cm column, ID 100 µm packed with 3 µm Reprocil, PepSep, Denmark) using an high pressure pump and solvent A [22]. The ESI coupling was obtained using a Nano Source Emitter Stainless Steel Nano-bore 1/32 (Thermo Fisher Scientific).

The ESI tuning parameters were the following: spray voltage, 2800 V; ion source gas flow (GS1), 16; curtain gas flow (CUR), 35; interface heater temperature (IHT), 100°C and declustering potential, 100.

The Evosep system was controlled by the Axel Semrau Chronos software (Axel Semrau GmbH, Germany), while the mass spectrometer acquisition software was Analyst TF 1.7.1 (Sciex).

Data-independent acquisition (SWATH/DIA) mass spectrometry [14] was performed for the quantitative analysis of the 576 SEC fractions (64 fractions per sample) obtained from the 9 SEC experiment. SWATH scan performed using an updated scheme of 64 variably sized precursor co-isolation windows [15], covering similar precursor densities (in terms of number and intensity) within all SWATH windows. The SWATH windows cover the precursors ions in the range of 350-1500 m/z and 350-1500 in the MS^2 SWATH scans, the accumulation time was 100 ms for the MS1 and 20 ms for each SWATH window, resulting in a cycle time of 1.38 s. For fragmentation, it was applied a rolling collisional energy with a collisional energy spread of 15 eV.

DDA MS analysis for the library generation

The 12 high pH fractioned peptide samples were separated on an Eksigent nanoLC Ultra AS2 1D Plus and expert 400 autosampler system (Eksigent, Dublin, CA) coupled to a TripleTOF 5600 through a NanoSpray III ion source using a Data Dependent Acquisition scheme. The 20 cm long nanoLC column was packed in house using a 75 µm inner diameter PicoFrit emitter (New Objective, Woburn) with Magic C18 AQ 3 um, 200 Å particles (Bruker, Billerica). The separation was performed at room temperature with a flow rate of 300 nl/min. All the LC solvents were all of mass spectrometry grade. The LC solvent A was composed of 98% water, 2% acetonitrile and 0.1% formic acid, LC solvent B was 98% acetonitrile, 2% water and 0.1% formic acid. The peptides were eluted over 120 minutes, with a linear gradient from 5% to 35% LC solvent B. One MS1 scan with a m/z range of 360-1460 and an accumulation time of 250 ms was followed by 20 MS2 scans with m/z ranges of 50-2000 and accumulation times of 100 ms. The dynamic exclusion time was set to 20 seconds.

Cell lysates total proteome analyses

Total proteome measurements have been performed by means of Eksigent nanoLC Ultra 1D Plus and expert 400 autosampler system (Eksigent, Dublin, CA) coupled to a TripleTOF 6600 through a NanoSpray III ion source. The nanoLC column was packed in house using a 75 μm inner diameter PicoFrit emitter (New Objective, Woburn, MA) with Magic C18 AQ 3 um, 200 Å particles (Bruker, Billerica, MA) to a length of 30 cm. The separation was performed at room temperature with a flow rate of 300 nl/min. All the LC solvents were of mass spectrometry grade. The LC solvent A was composed of 98% water, 2% acetonitrile and 0.1% formic acid, LC solvent B was 98% acetonitrile, 2% water and 0.1% formic acid. 1ug of peptides were injected for the analysis was applied a linear gradient starting from 5 to 35% B in 90 min.

The ESI tuning parameters were the following: spray voltage, 2800 V; GS1, 16; CUR, 35; IHT, 750°C and declustering potential, 100.

SWATH/DIA scan was performed using a scheme of 100 variably sized precursor co-isolation windows. The SWATH windows cover the MS1 range of 50-2000 m/z and 50-2000 for the MS2 scans, the accumulation time was 200 ms for the MS1 and 30 ms for each SWATH window, for a 3.2 s cycle time of 1.38 s. For fragmentation, a rolling collisional energy with a collisional energy spread of 15 eV was applied.

Data processing

DDA data analysis for the library generation

DDA-MS data acquired from peptide fractionation of the full THP-1 cell lysates (see above) were processed for the SWATH library generation following the protocol previously described [28].

MS spectra were searched for peptide matches against the human UniProt/SwissProt reference database (reviewed, canonical entries, June 2017) using Comet 2018.01 rev. 0 MS/MS search engine. The search was carried out using trypsin cleavage, 30 ppm precursor and 0.05 Da fragment ion mass tolerance, carbamidomethyl (C) as static and oxidation (M) as variable modification and a maximum of 2 enzyme missed cleavages. The results from the search were statistically scored using Peptide Prophet (statistical validation of PSMs) and iProphet (peptide sequence validation) of the Trans-Proteomic Pipeline (TPP v5.0.0 POLAR VORTEX rev 0), filtering the results at 1% peptide FDR (0.815939 iprob) as determined using the tool Mayu [29]. A wider peptide-level FDR cut-off (5% FDR on protein level, compared to requiring 1% FDR) was chosen in order to increase sensitivity for the recovery of true positive peptide signals.

The resulting spectra were then gathered for the generation of the consensus spectra library using SpectraST including retention time calibration. The 6 most abundant fragment ion transitions per precursor from the b_n or y_n ion series were selected, with a m/z range of 350-2000

and aa fragment charge states 1-2. The final library contains query parameters for 506,717 precursors of 73,007 peptides mapping to 9375 protein groups. Moreover, to the spectra consensus library reverse decoy (506,581 decoys transitions) were generated for the FDR scoring provided by the SWATH/DIA data analysis workflow.

SWATH/DIA data analysis

The SWATH-MS data collected from the analysis of SEC fractions were analyzed through peptide-centric analysis, querying 506,717 fragment precursors from the sample-specific peptide library generated (see above) in the SWATH MS^2 spectra, using a modified OpenSWATH [30,31] PyProphet and TRIC workflow.

First, one global classifier was trained on a subsampled set of SEC fractions across the experiment using pyProphet-cli [32]. Peptides from all fractions were then quantified and scored using the pre-trained scoring function using OpenSWATH, pyProphet and TRIC.

CCProfiler

The first differential analysis module in CCprofiler is tailored towards detecting proteins that differ in their global assembly state, meaning that the relative distribution between monomeric and assembled protein mass is different across the conditions. Since this module depends on the assignment of the fractionation dimension into a monomeric and assembled range based on the monomeric molecular weight of each protein, the analysis is currently only available for SEC datasets and requires both a molecular weight calibration of the fractions and a monomeric molecular weight annotation of the measured proteins. The cutoff between the monomeric and assembled SEC range is set at the fraction corresponding to two times the expected monomeric molecular weight of a protein. Based on this initial division of the SEC dimension, the assembled mass fraction (AMF) of each protein can be estimated by the fraction of the detected MS signal in the assembled mass range relative to the total, globally detected signal:

$$\text{AMF} = \frac{\sum intensity\ assembled}{\sum intensity\ global}$$

A change in AMF is subsequently estimated by the difference in mean AMF across conditions:

$$diffAMF = mean(AMFA) - mean(AMFB)$$

Here, AMFA and AMFB denote the AMF values of two conditions A and B. Since AMF values are not normally distributed and bound by zero and one, a conventional t-test for significance estimation is not applicable. Instead, CCprofiler applies a beta-regression model and p-value estimation by a likelihood-ratio test to derive significance estimates (for details see below). Multiple testing correction is performed by Benjamini-Hochberg adjustment of the derived p-values [33]. Proteins with significant adjusted p-values and large AMF differences, are indicated to have a different proportion of individual proteins associated to higher order assemblies across the conditions. Notably, this information is derived independent from any feature (i.e. peak group) detection and does not require knowledge of the protein's exact interaction partners.

Differential analysis of distinct protein assembly states and detection of protein rewiring

To further gain insights into distinct protein assembly states, we have previously introduced the protein-centric analysis concept for CoFrac-MS data within a single condition [2]. Here, we extend the protein-centric analysis concept to enable the differential assessment of distinct protein assembly states. To achieve consistent protein feature (i.e. peptide co-elution peak group) detection across conditions and replicates, peptide-level traces are first integrated by summing the intensities across all samples in the provided tracesList. The integrated traces are subsequently used for protein-centric feature finding, applying random peptide assignments as decoy model for p- and q-value estimation [2]. Each protein can thereby be assigned to potentially multiple distinct assembly states, as indicated by the detection of multiple unique protein features. Following this initial protein feature detection, differential analysis is performed to compare the signal intensity within each protein feature across conditions.

Differential analysis is performed in 5 steps: (i) Peptide-level intensities are computed for each protein feature and sample. Missing values in single fractions, replicates or conditions are imputed by uniformly sampling values between zero and the minimum detected signal of a peptide. The peptide intensity of one feature is then calculated by summing the intensities of all fractions across the corresponding protein feature range. (ii) The mean intensity across all replicates within a condition (specified by the design matrix) is calculated. (iii) The log2-fold-change between conditions is calculated based on the mean feature intensities. (iv) If replicates are available, p-values are estimated by comparing the summed intensities across conditions by a non-paired t-test. If no replicates are available, p-values are estimated by comparing each fraction within a feature by a paired t-test across the conditions. (v) To subsequently derive protein-level information, the peptide-level tests are aggregated as follows: (1) protein log2-fold-changes are derived from the median log-2-fold change across all detected peptides of the protein,

and (2) protein p-values are estimated by determining the fold-change adjusted median p-value and applying a beta distribution as described by Teo et al. [34] and Suomi et al. [35] (for details see method section). (vi) Multiple testing correction is performed by Benjamini-Hochberg adjustment of the protein-level p-values [2].

In addition to the feature-specific differential analysis, global differential assessment is performed by comparing integrated intensities across the entire fractionation dimension instead of restricting the analysis to a feature-specific range. The same strategies as for feature-specific estimation of log2-fold-changes and p-values are performed. To assess whether the signal within a protein feature is changing because of a global change in the protein's expression or due to a rearrangement of the proteins relative distribution across different assembly states, an additional analysis step is available in CCprofiler. Here, the relative feature-specific mass fraction (FMF) is estimated by the fraction of the detected MS signal in the feature-specific mass range relative to the total detected signal:

$$FMF = \frac{\sum intensity\ feature}{\sum intensity\ global}$$

A change in FMF is subsequently estimated by the difference in mean FMF across conditions:

$$diffFMF = mean(FMFA) - mean(FMFB)$$

Here, FMFA and FMFB denote the FMF values of two conditions A and B. Similar to the concept introduced for comparing AMF values, CCprofiler applies a beta-regression model and p-value estimation by a likelihood ratio test [36] to derive significance estimates for the change in FMF across conditions (for details see methods section). Since the initial assessment of FMF values is performed on peptide-level data, protein-level information is derived by aggregation across all detected peptides as follows: (1) FMF differences are derived from the median diffFMF across all detected peptides of the protein, and (2) p-values are estimated by determining the difference adjusted median p-value and applying a beta distribution as described by Teo et al. [34] and Suomi et al. [35](for details see method section). Multiple testing correction is performed by a Benjamini-Hochberg adjustment of the p-values [33]. A significant change in the FMF across conditions indicates that the protein's relative contribution to different distinct assembly states has changed across the conditions, thus providing insights into protein rewiring which is not observable by global proteome analyses. In contrast to complex-centric analyses, described in the following section, protein-centric differential analysis enables the assessment of changes in distinct protein assembly states independent of actually knowing the protein's exact interaction partners.

Protein complex detection and differential analysis

The final analysis module in CCprofiler is focused on the complex-centric detection and differential assessment of protein complexes. We have previously introduced the basic concept of complex-centric analysis for CoFrac-MS data of a single condition [2]. In summary, prior protein connectivity information is used to query CoFrac-MS data directly for evidence of pre-defined complexes. By using random protein assignments as a decoy model for error rate estimation, complex-centric analysis enables the detection of hundreds of protein complexes at high sensitivity and under controlled FDR. Here, we expand the complex-centric analysis strategy to allow the quantitative comparison between complexes detected across different cellular conditions. Analogous to the protein-centric workflow described in the previous section, protein-level traces are first integrated by summing the intensities across all samples in the provided tracesList to ensure consistent signal detection across conditions and replicates. The integrated traces are subsequently used for complex-centric feature detection. Only the most complete complex feature (i.e. protein co-elution peak group) for each complex query is considered for scoring and FDR estimation. After filtering for q-values (e.g. 0.05), the complex features are appended by secondary features with high correlation values (peak correlation 0.7). These secondary features can for example entail potential sub-complexes or complex variants [2].

Following this initial protein complex feature detection, a differential analysis step can be performed to compare the signal intensity within each complex feature across different conditions. The analysis concept is analogous to the differential analysis strategy implemented on the level of protein features (see previous section). The initial differential testing is performed on peptide level, while results are subsequently aggregated on the protein level. For complex-centric analysis, the protein-level results are additionally aggregated to the complex level, again following the same strategy as compared to aggregation from peptide to protein level. Finally, multiple testing correction is performed by a Benjamini-Hochberg adjustment of the p-values [33].

P-value estimation for AMF and FMF differences

P-value estimation for AMF and FMF differences was performed by first transforming the AMF and FMF (y) to values between zero and one, while excluding the extremes (0 and 1) [37,38]:

$$y' = \frac{(y * (n-1) + 0.5)}{n}$$

Here, n denotes the sample size, which was six for the presented dataset. The resulting y' values were used for fitting a beta-regression model with the betareg R package with default parameters [37,38]. The lrtest function of the lmtest R package [36] was subsequently used for p-value estimation by a likelihood-ratio test with default parameters. Multiple testing correction was performed by the p.adjust function of the stats base package, using the "fdr" method corresponding to correction by Benjamini-Hochberg [33].

P-value estimation for aggregating peptide-level tests to the protein and complex level

Peptide-level p-values were aggregated to the protein-level by applying the strategy presented by Teo at al. [34] and Suomi et al. [35]. First the median of peptide-level p-values is used as a score for each protein taking the direction of change into account. The protein-level significance of the detection is subsequently calculated using a beta distribution [35]. The same strategy is applied to aggregate protein-level p-values to the complex level. Multiple testing correction is performed by the p.adjust function of the stats base package, using the "fdr" method corresponding to correction by Benjamini-Hochberg [33].

CCprofiler analysis workflow and parameters

All R-scripts for the CCprofiler analysis are openly available on github. The following provides a summary of the most important processing steps and the selected parameters for the presented analysis.

Due to the very low molecular weight of later SEC factions, the data was limited do fractions 1 to 49 for CCprofiler analysis. Missing peptide intensity values (for which both the previous and following fraction contained measured intensity values) were imputed by a spline fit across the SEC dimension. After missing value imputation, peptide intensity values were normalized across conditions and replicates by applying a cyclic loess normalization [39–41]. Low-confidence peptides were subsequently removed, keeping only peptides with (1) at least three consecutive detections across any replicate, (2) at least one high correlating sibling peptide (maximum correlation >= 0.5), and (3) a good average sibling peptide correlation (>= 0.2). Protein quantification was performed by summing the top two most intense peptides consistently across all replicates.

To determine proteins with a significant change in their assembly state across conditions, a mean difference in AMF of >= 30% and a Benjamini-Hochberg adjusted p-value <= 0.05 were required.

Protein-centric analysis was performed with following parameters: corr_cutoff=0.9, window_size=7, rt_height=1, smoothing_length=7, perturb_cutoff="5%" and collapse_method="apex_only". Only protein features passing the 5% FDR threshold were further

considered. For the differential analysis, a minimum log2 fold-change of one and a Benjamini-Hochberg corrected p-value of 0.05 were required for significance in all pairwise analyses. To determine protein features with a significant change in their relative abundance in comparison to the total protein intensity across conditions, a mean difference in FMF of >= 30% and a Benjamini-Hochberg adjusted p-value <= 0.05 were required.

For complex-centric analysis, we first defined a set of target protein complex queries. This was achieved by combining queries derived from CORUM [16] and StringDB [17]. We derived protein complex queries from StringDB v10 (9606.protein.links.v10.txt). Protein identifiers were mapped to Uniprot accessions via BioMart. The interactions were filtered for a minimal combined_score of 980. We applied the ClusterONE algorithm [24] for PPI network partitioning with following parameters: d=0.95. Weights were set to the combined_score divided by 1000. CORUM derived protein complex queries were taken directly from within the CCprofiler package [2]. The complex queries were combined and decoys were generated randomly by requiring a minimum edge distance of 3. Complex-centric analysis was performed with following parameters: corr_cutoff=0.9, window_size=7, rt_height=1, smoothing_length=7, perturb_cutoff="5%" and collapse_method="apex_network". Only complex features with a molecular weight higher than two times the largest monomeric molecular weight of any of its participating subunits were considered. For each protein complex query, the complex feature with the highest number of participating subunits was selected for FDR estimation, filtering for a maximum FDR of 5%. Secondary features were appended to the final results based on a minimum peak correlation threshold of 0.7. To reduce redundancy across the detected complex features between different queries, features were collapsed with following parameters: rt_height = 0 and distance_cutoff = 1.25. For the differential analysis, a minimum log2 fold-change of one and a Benjamini-Hochberg corrected p-value of 0.05 were required for significance in all pairwise analyses.

5.6 References

1. Hartwell LH, Hopfield JJ, Leibler S, Murray AW: **From molecular to modular cell biology. [Internet]**. *Nature* 1999, **402**:C47-52.
2. Heusel M, Bludau I, Rosenberger G, Hafen R, Frank M, Banaei-Esfahani A, Drogen A, Collins BC, Gstaiger M, Aebersold R: **Complex-centric proteome profiling by <scp>SEC</scp> - <scp>SWATH</scp> - <scp>MS</scp> [Internet]**. *Mol. Syst. Biol.* 2019, **15**.
3. Gingras A-C, Gstaiger M, Raught B, Aebersold R: **Analysis of protein complexes using mass spectrometry [Internet]**. *Nat. Rev. Mol. Cell Biol.* 2007, **8**:645–654.
4. Majorek KA, Kuhn ML, Chruszcz M, Anderson WF, Minor W: **Double trouble - Buffer selection and his-tag presence may be responsible for nonreproducibility of biomedical experiments**. *Protein Sci.* 2014, **23**:1359–1368.
5. Meyer K, Selbach M: **Quantitative affinity purification mass spectrometry: a versatile technology to study protein–protein interactions [Internet]**. *Front. Genet.* 2015, **6**:237.
6. Kristensen AR, Gsponer J, Foster LJ: **A high-throughput approach for measuring temporal changes in the interactome [Internet]**. *Nat. Methods* 2012, **9**:907–909.
7. Havugimana PC, Hart GT, Nepusz T, Yang H, Turinsky AL, Li Z, Wang PI, Boutz DR, Fong V, Phanse

S, et al.: **A Census of Human Soluble Protein Complexes [Internet]**. *Cell* 2012, **150**:1068–1081.
8. Wan C, Borgeson B, Phanse S, Tu F, Drew K, Clark G, Xiong X, Kagan O, Kwan J, Bezginov A, et al.: **Panorama of ancient metazoan macromolecular complexes**. *Nature* 2015, **525**:339–344.
9. Kristensen AR, Foster LJ: **Protein correlation profiling-SILAC to study protein-protein interactions**. *Methods Mol. Biol.* 2014, **1188**:263–270.
10. Kirkwood KJ, Ahmad Y, Larance M, Lamond AI: **Characterization of native protein complexes and protein isoform variation using size-fractionation-based quantitative proteomics. [Internet]**. *Mol. Cell. Proteomics* 2013, **12**:3851–73.
11. Dong M, Lee ⊥, Yang L, Williams K, Fisher SJ, Hall SC, Biggin MD, Jin J, Witkowska HE: **A " Tagless " Strategy for Identification of Stable Protein Complexes Genome-wide by Multidimensional Orthogonal Chromatographic Separation and iTRAQ Reagent Tracking [Internet]**. [date unknown], doi:10.1021/pr700624e.
12. Hu LZM, Goebels F, Tan JH, Wolf E, Kuzmanov U, Wan C, Phanse S, Xu C, Schertzberg M, Fraser AG, et al.: **EPIC: software toolkit for elution profile-based inference of protein complexes**. *Nat. Methods* 2019, **16**:737–742.
13. Stacey RG, Skinnider MA, Scott NE, Foster LJ: **A rapid and accurate approach for prediction of interactomes from co-elution data (PrInCE) [Internet]**. *BMC Bioinformatics* 2017, **18**:457.
14. Gillet LC, Navarro P, Tate S, Röst H, Selevsek N, Reiter L, Bonner R, Aebersold R: **Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. [Internet]**. *Mol. Cell. Proteomics* 2012, **11**:O111.016717.
15. Collins BC, Hunter CL, Liu Y, Schilling B, Rosenberger G, Bader SL, Chan DW, Gibson BW, Gingras A-C, Held JM, et al.: **Multi-laboratory assessment of reproducibility, qualitative and quantitative performance of SWATH-mass spectrometry [Internet]**. *Nat. Commun.* 2017, **8**:291.
16. Ruepp A, Waegele B, Lechner M, Brauner B, Dunger-Kaltenbach I, Fobo G, Frishman G, Montrone C, Mewes HW: **CORUM: The comprehensive resource of mammalian protein complexes-2009 [Internet]**. *Nucleic Acids Res.* 2009, **38**:D497-501.
17. Snel B: **STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene**. *Nucleic Acids Res.* 2000, **28**:3442–3444.
18. Heusel M, Frank M, Köhler M, Amon S, Frommelt F, Rosenberger G, Bludau I, Aulakh S, Linder MI, Liu Y, et al.: **A Global Screen for Assembly State Changes of the Mitotic Proteome by SEC-SWATH-MS [Internet]**. *Cell Syst.* 2020, **0**.
19. Chanput W, Mes JJ, Wichers HJ: **THP-1 cell line: An in vitro cell model for immune modulation approach**. *Int. Immunopharmacol.* 2014, **23**:37–45.
20. Hambleton J, Weinsteint SL, Lemt L, Defrancots AL: *Activation of c-Jun N-terminal kinase in bacterial lipopolysaccharide-stimulated macrophages (c-Jun N-terminal kinase/endotoxin/macrophages)*. 1996.
21. Potriquet J, Laohaviroj M, Bethony JM, Mulvenna J: **A modified FASP protocol for high-throughput preparation of protein samples for mass spectrometry [Internet]**. *PLoS One* 2017, **12**:e0175967.
22. Bache N, Geyer PE, Bekker-Jensen DB, Hoerning O, Falkenby L, Treit P V., Doll S, Paron I, Müller JB, Meier F, et al.: **A novel LC system embeds analytes in pre-formed gradients for rapid, ultra-robust proteomics**. *Mol. Cell. Proteomics* 2018, **17**:2284–2296.
23. Schwende H, Fitzke E, Ambs P, Dieter P: **Differences in the state of differentiation of THP-1 cells induced by phorbol ester and 1,25-dihydroxyvitamin D3**. *J. Leukoc. Biol.* 1996, **59**:555–561.
24. Nepusz T, Yu H, Paccanaro A: **Detecting overlapping protein complexes in protein-protein interaction networks**. *Nat. Methods* 2012, **9**:471–472.
25. Autenrieth IB, Bohn Hrabe De Angelis E, Weber F, Hornef MW, Gailus-Durner V, Fuchs H, Kalinke MU, Hofer MJ, Hatesuer B, Deuschle KE, et al.: **Secretion and LPS-Induced Endotoxin Shock α Lipopolysaccharide (LPS)-Induced TNF-Mediated Amplification of – IFIT2 Is an Effector Protein of Type I IFN [Internet]**. 2013, doi:10.4049/jimmunol.1203305.
26. **The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible [Internet]**. [date unknown], [no volume].
27. Meier F, Brunner A-D, Frank M, Ha A, Voytik E, Kaspar-Schoenefeld S, Lubeck M, Raether O, Aebersold R, Collins BC, et al.: **Parallel accumulation – serial fragmentation combined with data-independent acquisition (diaPASEF): Bottom-up proteomics with near optimal ion usage [Internet]**. *bioRxiv* 2019, doi:10.1101/656207.

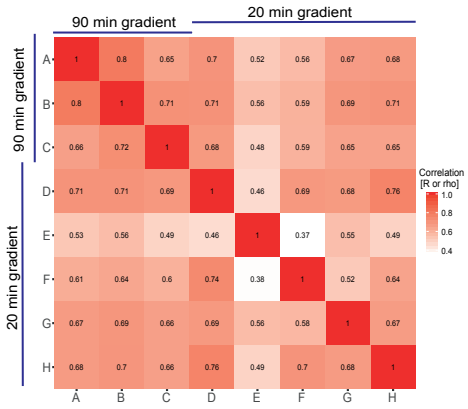
28. Schubert OT, Gillet LC, Collins BC, Navarro P, Rosenberger G, Wolski WE, Lam H, Amodei D, Mallick P, Maclean B, et al.: **Building high-quality assay libraries for targeted analysis of SWATH MS data [Internet]**. *Nat. Protoc.* 2015, **10**:426–441.
29. Reiter L, Claassen M, Schrimpf SP, Jovanovic M, Schmidt A, Buhmann JM, Hengartner MO, Aebersold R: **Protein identification false discovery rates for very large proteomics data sets generated by tandem mass spectrometry**. *Mol. Cell. Proteomics* 2009, **8**:2405–2417.
30. Röst HL, Rosenberger G, Navarro P, Gillet L, Miladinović SM, Schubert OT, Wolski W, Collins BC, Malmström J, Malmström L, et al.: **OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data [Internet]**. *Nat. Biotechnol.* 2014, **32**:219–223.
31. Rosenberger G, Liu Y, Röst HL, Ludwig C, Buil A, Bensimon A, Soste M, Spector TD, Dermitzakis ET, Collins BC, et al.: **Inference and quantification of peptidoforms in large sample cohorts by SWATH-MS [Internet]**. *Nat. Biotechnol.* 2017, **35**:781–788.
32. Rosenberger G, Bludau I, Schmitt U, Heusel M, Hunter CL, Liu Y, Maccoss MJ, Maclean BX, Nesvizhskii AI, Pedrioli PGA, et al.: **Statistical control of peptide and protein error rates in large-scale targeted data-independent acquisition analyses**. *Nat. Methods* 2017, **14**:921–927.
33. Benjamini Y, Hochberg Y: *Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple*. 1995.
34. Teo G, Vogel C, Ghosh D, Kim S, Choi H: **PECA: A novel statistical tool for deconvoluting time-dependent gene expression regulation**. *J. Proteome Res.* 2014, **13**:29–37.
35. Suomi T, Elo LL: **Enhanced differential expression statistics for data-independent acquisition proteomics**. *Sci. Rep.* 2017, **7**:1–8.
36. Zeileis A, Hothorn T: *Diagnostic Checking in Regression Relationships*. [date unknown].
37. Cribari-Neto F, Zeileis A: **Beta regression in R**. *J. Stat. Softw.* 2010, **34**:1–24.
38. Grün B, Kosmidis I, Zeileis A: **Extended beta regression in R: Shaken, Stirred, Mixed, and partitioned**. *J. Stat. Softw.* 2012, **48**:1–25.
39. Rosenberger G, Heusel M, Bludau I, Collins B, Martelli C, Williams E, Xue P, Liu Y, Aebersold R, Califano A: **SECAT: Quantifying differential protein-protein interaction states by network-centric analysis [Internet]**. *bioRxiv* 2019, doi:10.1101/819755.
40. Bolstad BM, Irizarry RA, Åstrand M, Speed TP: **A comparison of normalization methods for high density oligonucleotide array data based on variance and bias**. *Bioinformatics* 2003, **19**:185–193.
41. Ballman K V, Grill DE, Oberg AL, Therneau TM: **Faster cyclic loess: normalizing RNA arrays via linear models. [Internet]**. *Bioinformatics* 2004, **20**:2778–86.

5.7 Supplementary Figures

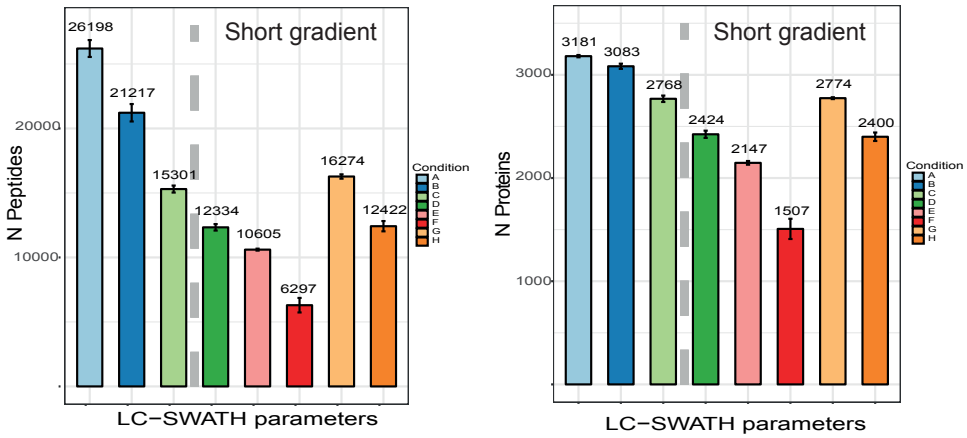
A

Condition	Column length (cm)	Flow rate (nl/min)	Gradient lenght (min)	N variable windows	MS² scan time (msec)	MS² mass range	Sample load (µgs)
A) Standard DIA/SWATH method	20	300	90	64	50	50-2000	1
B) Shorten column	10	300	90	64	50	50-2000	1
C) Increase flowrate	10	1000	90	64	50	50-2000	1
D) Short gradient-Method 1	10	1000	20	64	25	50-2000	1
E) Short gradient-Method 2	10	1000	20	32	50	50-2000	1
F) Short gradient-Method 3	10	1000	20	100	15	50-2000	1
G) Short gradient-Method 4	10	1000	20	64	25	350-1500	1
H) Short gradient-Method 5	10	1000	20	64	25	50-2000	2

B



C



Supplementary Figure 1.

Effect of the column length, flow rate, gradient length and number of DIA/SWATH variable windows in the DIA/SWATH performances (sample correlation, number of peptides and proteins identification). All the runs were acquired with Eksigent nanoLC Ultra AS2 1D Plus and expert 400 autosampler system (Eksigent, Dublin, CA) coupled to a TripleTOF 5600 (Sciex) through a NanoSpray III ion source (Sciex). The 20 cm and 10 cm nanoLC columns were packed in house using a 75 µm inner diameter PicoFrit emitter (New Objective, Woburn) and ProntoSIL 3 um, 200 Å particles (Bischoff Chromatography, Germany).

For the chromatographic separation an aqueous solution with 2% ACN/0.1 FA % was used as buffer A, while a 98 % ACN%/0.1% FA solution as buffer B.

Panel A. Short scheme of the different conditions tested.

A) Standard DIA/SWATH method using a scheme of 64 variably sized precursor co-isolation windows. The SWATH windows cover the precursor ions in the range of 360-1460 m/z and 50-2000 in the MS^2 SWATH scans, the accumulation time was 250 ms for the MS1 and 50 ms for each SWATH window, resulting in a cycle time of 3.5 s. For fragmentation, it was applied a rolling collisional energy with a collisional energy spread of 15 eV. 1 µg of peptides were loaded in the sample loop. The separation was carried out with a 20-cm capillary column at 300 nl/min flow for a 90-min gradient from 5 to 30% of buffer B.

B) Shorten column using a scheme of 64 variably sized precursor co-isolation windows. The SWATH windows cover the precursors ions in the range of 360-1460 m/z and 50-2000 in the MS2 SWATH scans, the accumulation time was 250 ms for the MS1 and 50 ms for each SWATH window, resulting in a cycle time of 3.5 s. For fragmentation, it was applied a rolling collisional energy with a collisional energy spread of 15 eV. 1 µg of peptides were loaded in the sample loop. The separation was carried out with a 10-cm capillary column at 300 nl/min flow for a 90-min gradient from 5 to 30% of buffer B.

C) Increased separation flowrate using a scheme of 64 variably sized precursor co-isolation windows. The SWATH windows cover the precursors ions in the range of 360-1460 m/z and 50-2000 in the MS^2 SWATH scans, the accumulation time was 250 ms for the MS1 and 50 ms for each SWATH window, resulting in a cycle time of 3.5 s. For fragmentation, it was applied a rolling collisional energy with a collisional energy spread of 15 eV. 1 µg of peptides were loaded in the sample loop. The separation was carried out with a 10-cm capillary column at 1000 nl/min flow for a 90-min gradient from 5 to 30% of buffer B.

D) Short gradient-Method 1. 64 variably sized precursor co-isolation windows scheme. The SWATH windows cover the precursors ions in the range of 360-1460 m/z and 50-2000 in the MS^2 SWATH scans, the accumulation time was 250 ms for the MS1 and 25 ms for each SWATH window, resulting in a cycle time of 1.9 s. For fragmentation, it was applied a rolling collisional energy with a collisional energy spread of 15 eV. 1 µg of peptides were loaded in the sample loop. The separation was carried out with a 120-cm capillary column at 1000 nl/min flow for a 20-min gradient from 5 to 30% of buffer B.

E) Short gradient-Method 2. 32 variably sized precursor co-isolation windows scheme. The SWATH windows cover the precursors ions in the range of 360-1460 m/z and 50-2000 in the MS^2 SWATH scans, the accumulation time was 250 ms for the MS1 and 50 ms for each SWATH window, resulting in a cycle time of 1.9 s. For fragmentation, it was applied a rolling collisional energy with a collisional energy spread of 15 eV. 1 µg of peptides were loaded in the sample loop. The separation was carried out with a 10-cm capillary column at 1000 nl/min flow for a 20-min gradient from 5 to 30% of buffer B.

F) Short gradient-Method 3. 100 variably sized precursor co-isolation windows scheme. The SWATH windows cover the precursors ions in the range of 360-1460 m/z and 50-2000 in the MS^2 SWATH scans, the accumulation time was 250 ms for the MS^1 and 15 ms for each SWATH window, resulting in a cycle time of 1.8 s. For fragmentation, it was applied a rolling collisional energy with a collisional energy spread of 15 eV. 1 µg of peptides were loaded in the sample loop. The separation was carried out with a 10-cm capillary column at 1000 nl/min flow for a 20-min gradient from 5 to 30% of buffer B.

G) Short gradient-Method 4. 64 variably sized precursor co-isolation windows scheme. The SWATH windows cover the precursors ions in the range of 360-1460 m/z and 350-1500 in the MS^2 SWATH scans, the accumulation time was 250 ms for the MS^1 and 25 ms for each SWATH window, resulting in a cycle time of 1.9 s. For fragmentation, it was applied a rolling collisional energy with a collisional energy spread of 15 eV. 1 µg of peptides were loaded in the sample loop. The separation was carried out with a 120-cm capillary column at 1000 nl/min flow for a 20-min gradient from 5 to 30% of buffer B.

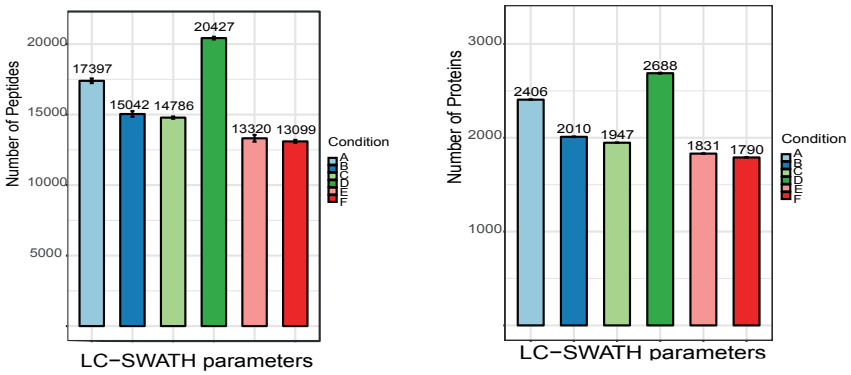
F) Short gradient-Method 5. 64 variably sized precursor co-isolation windows scheme. The SWATH windows cover the precursors ions in the range of 360-1460 m/z and 50-2000 in the MS^2 SWATH scans, the accumulation time was 250 ms for the MS^1 and 50 ms for each SWATH window, resulting in a cycle time of 3.5 s. For fragmentation, it was applied a rolling collisional energy with a collisional energy spread of 15 eV. 2 µg of peptides were loaded in the sample loop. The separation was carried out with a 10-cm capillary column at 1000 nl/min flow for a 90-min gradient from 5 to 30% of buffer B.

Panel B. Pearson (upper triangle) and Spearman correlation (lower triangle) of fragment ions intensities across the tested conditions (90 min gradient vs 20 min gradient).
Effect of the column length, flow rate, gradient length and number of DIA/SWATH variable windows in terms of the correlation of the total fragment intensities. The SWATH-MS data were analyzed using OpenSWATH, PyProphet and TRIC workflow (see Methods section). No further normalization was applied. From the plot results clear as the SWATH acquisition method [i.e. SWATH number of windows; Condition E (32 SWATH windows) vs Condition D (64 SWATH windows) and Condition F (100 SWATH SWATH windows)] influences most the performances of the analysis with respect that the column length (i.e. Condition A vs Condition B) and the flow rate (i.e. Conditions A/B vs Condition C).
Panel C. Number of peptides (left panel) and number of proteins (right panel) identified across the tested conditions.
Effect of the column length, flow rate, gradient length and number of DIA/SWATH variable windows in terms of total number of peptides and proteins identified.
The short gradient decrease of the 53% the number of peptides identified (Condition A vs Condition D), whereas of 24% the total proteins. As shown for the correlation analysis, the major effect observed for short gradients is determined by the selection of the number of SWATH windows and the MS^2 scan range. The 100 SWATH window method (Condition F) showed 76% and 53% of peptides and protein loss, respectively, with respect to the standard conditions (Condition A). The smaller MS^2 scan range (Condition H) improved the performances of 24% and 13% of peptides and proteins characterized, respectively, in comparison with the broader MS^2 scan range (Condition D).

A

Condition	N variable windows	MS² scan time (msec)	MS² mass range	Cycle time (msec)
A	32	40	300-1500	1430
B	32	40	50-1500	1430
C	32	40	50-2000	1430
D	64	20	300-1500	1430
E	64	20	50-1500	1430
F	64	20	50-2000	1430

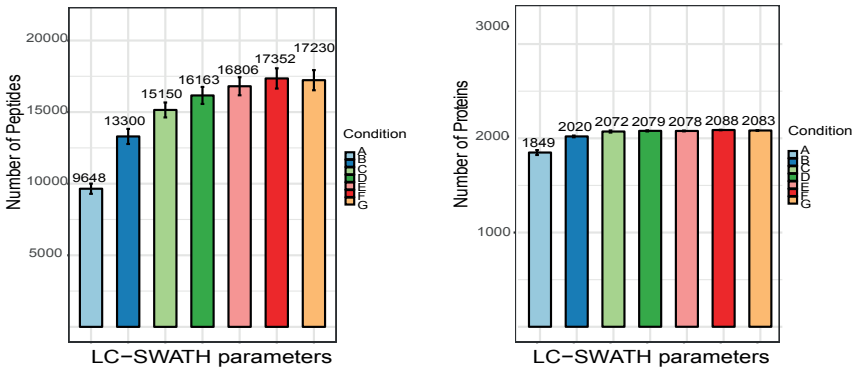
B



C

Condition	N variable windows	MS² scan time (msec)	MS² mass range	Cycle time (msec)
A	16	80	300-1500	1430
B	24	54	300-1500	1446
C	32	40	300-1500	1430
D	40	32	300-1500	1430
E	48	27	300-1500	1446
F	56	23	300-1500	1438
G	64	20	300-1500	1430

D



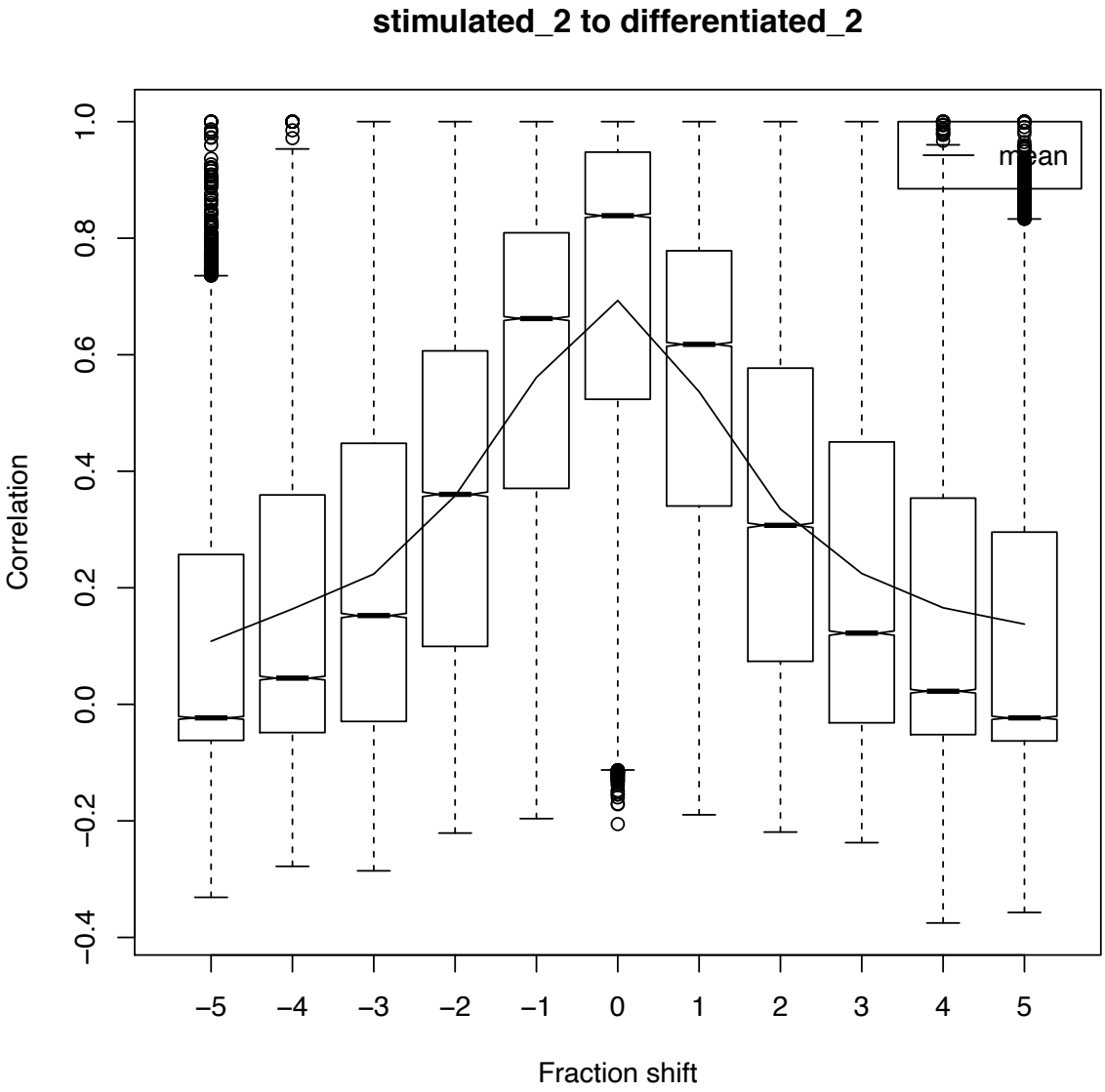
Supplementary Figure 2.

Effect of the MS² scan range (Panels A and B) and number of SWATH windows (Panels C and D) in the DIA/SWATH performances (number of peptides and proteins identified). All the runs were performed on Evosep One system (Evosep Biosystems) couple to an TripleTOF 5,600 instrument (Sciex) equipped with a NanoSpray III ion source (Sciex) selecting the "60 samples per day" method and using the nanoConnect LC column (8 cm column, ID 100 µm packed with 3 µm Reprocil, PepSep) (see Methods section).

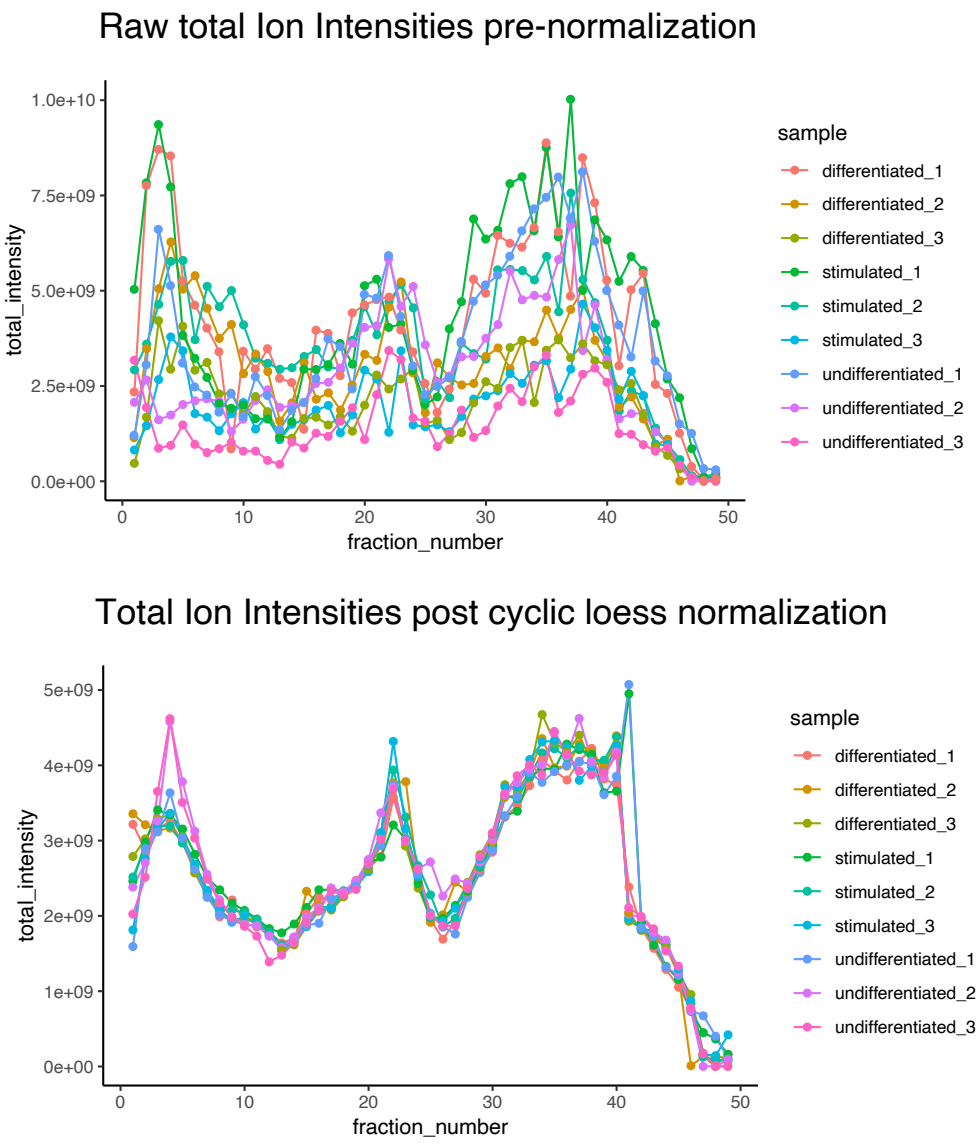
Panel A. Short scheme of the different conditions tested for the MS^2 scan range.
A) 32 variably sized precursor co-isolation windows. The SWATH windows cover the precursors ions in the range of 350-1460 m/z and 350-1500 in the MS^2 SWATH scans, the accumulation time was 100 ms for the MS1 and 40 ms for each SWATH window, resulting in a cycle time of 1.430 s. For fragmentation, it was applied a rolling collisional energy with a collisional energy spread of 15 eV. 1 µg of peptides were loaded in the sample loop.
B) 32 variably sized precursor co-isolation windows with the MS^2 SWATH scans range of 50-1500. The other parameters were the same of Condition A.
C) 32 variably sized precursor co-isolation windows with the MS^2 SWATH scans range of 50-2000. The other parameters were the same of Condition A.
D) 64 variably sized precursor co-isolation windows. The SWATH windows cover the precursor ions in the range of 350-1460 m/z and 350-1500 in the MS^2 SWATH scans, the accumulation time was 100 ms for the MS^1 and 20 ms for each SWATH window, resulting in a cycle time of 1.430 s. For fragmentation, it was applied a rolling collisional energy with a collisional energy spread of 15 eV. 1 µg of peptides were loaded in the sample loop.
E) 64 variably sized precursor co-isolation windows with the MS2 SWATH scans range of 50-1500. The other parameters were the same of Condition D.
F) 64 variably sized precursor co-isolation windows with the MS2 SWATH scans range of 50-2000. The other parameters were the same of Condition D.
Panel B. Number of peptides (left panel) and number of proteins (right panel) identified across the conditions tested for the MS^2 scan range.
Effect of the MS^2 scan range in the DIA/SWATH performances in terms of number of peptides and proteins identified.
The smaller MS^2 scan range improves the numbers of peptides and proteins identified both with 32 and 64 SWATH windows methods. For the 32 SWATH windows method, the 350-1500 MS^2 scan range (Condition A) improves of ~14% and ~16% the number of peptides and proteins identifications (with respect of conditions B and C). For the 64 SWATH windows method, the 350-1500 MS^2 scan range (ConditionD) improves of ~35% and ~32% the number of peptides and proteins identifications (with respect of conditions E and F). As shown here, the influence of the MS^2 scan range in the mass spectrometry acquisition increases with an increased number of SWATH windows; a smaller mass range improves the sensitivity of the detection and allowing a higher number of events for the same full cycle time (1.430 secs).
Panel C. Short scheme of the different conditions tested for the number of SWATH windows.
A) 16 variably sized precursor co-isolation windows. The SWATH windows cover the precursor ions in the range of 350-1460 m/z and 350-1500 in the MS^2 SWATH scans, the accumulation time was 100 ms for the MS^1 and 80 ms for each SWATH window, resulting in a cycle time of 1.430 s. For fragmentation, it was applied a rolling collisional energy with a collisional energy spread of 15 eV. 1 µg of peptides were loaded in the sample loop.
B) 24 variably sized precursor co-isolation windows. The accumulation time for each SWATH window is 54 ms, for a total cycle time of 1.446 s. The other parameters were the same of Condition A.

C) 32 variably sized precursor co-isolation windows. The accumulation time for each SWATH window is 40 ms, for a total cycle time of 1.430 s. The other parameters were the same of Condition A.
D) 40 variably sized precursor co-isolation windows. The accumulation time for each SWATH window is 32 ms, for a total cycle time of 1.430 s. The other parameters were the same of Condition A.
C) 48 variably sized precursor co-isolation windows. The accumulation time for each SWATH window is 27 ms, for a total cycle time of 1.446 s. The other parameters were the same of Condition A.
C) 56 variably sized precursor co-isolation windows. The accumulation time for each SWATH window is 23 ms, for a total cycle time of 1.438 s. The other parameters were the same of Condition A.
C) 64 variably sized precursor co-isolation windows. The accumulation time for each SWATH window is 20 ms, for a total cycle time of 1.436 s. The other parameters were the same of Condition A.

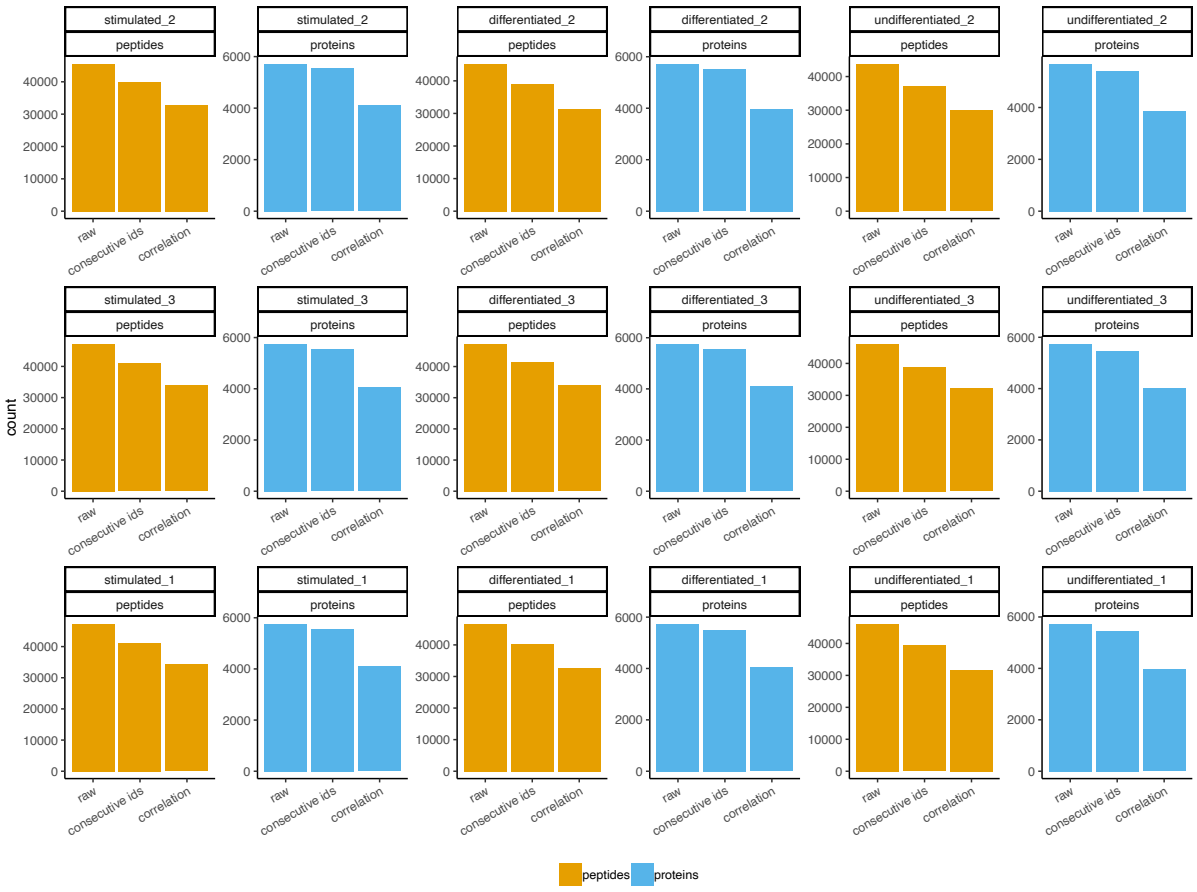
Panel D. Number of peptides (left panel) and number of proteins (right panel) identified across the conditions tested for the number of SWATH windows.
Evaluation of the performances of different DIA/SWATH methods in function of the number of variable SWATH windows. The shorter gradient
The smaller MS^2 scan range improves the numbers of peptides and proteins identified both with 32 and 64 SWATH windows methods. For the 32 SWATH windows method, the 350-1500 MS^2 scan range (Condition A) improves of ~14% and ~16% the number of peptides and proteins identifications (with respect of conditions B and C). For the 64 SWATH windows method, the 350-1500 MS^2 scan range (ConditionD) improves of ~35% and ~32% the number of peptides and proteins identifications (with respect of conditions E and F). As shown here, the influence of the MS^2 scan range in the mass spectrometry acquisition increases with an increased number of SWATH windows; a smaller mass range improves the sensitivity of the detection and allowing a higher number of events for the same full cycle time (1.430 secs).



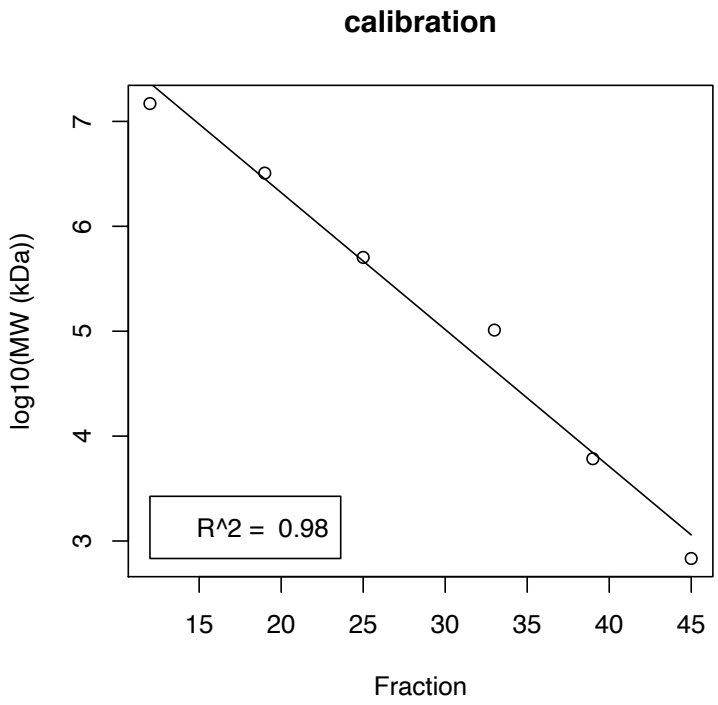
Supplementary figure 1 Example of a pair-wise alignment between the peptide-level SEC traces to illustrate different fraction shift offsets.



Supplementary figure 2 Collective total ion intensities for all 9 SEC runs before and after cyclic loess normalization.



Supplementary figure 3 Number of peptides (orange) or proteins (in blue) before and after the 3 peptide-filtering steps. The filtering steps are based on the identification of peptides in consecutive fractions and based on the correlation of sibling peptides mapping to the same proteins.



Supplementary figure 4 Example of a molecular weight calibration based on the log-linear relationship of the calibrant's molecular weights and the elution fraction.

Chapter 6

Conclusions and discussion

Parts of this chapter was published in:

Nicod C, Banaei-Esfahani A, Collins BC
Elucidation of host–pathogen protein–protein interactions to uncover mechanisms of host cell rewiring.
Curr. Opin. Microbiol. 2017, **39**:7–15.

The rest of the chapter was exclusively written by Charlotte Nicod.

6.1 Key findings

In this thesis, we aimed at unraveling host-pathogen interactions at the proteome level upon *Mycobacterium tuberculosis* infections using quantitative mass spectrometry. *Mycobacterium tuberculosis* is the causative agent of tuberculosis and infects primarily alveolar macrophages. Virulent *Mtb* strains having an intact RD1 locus encoding the ESX-1 protein secretion system successfully immunomodulate their host macrophages to favour their intracellular survival and replications by interfering with host cellular processes. Therefore, in order to gain a better understanding in the underlying mechanisms of *Mtb* pathogenicity, we aimed at mapping the indirect and direct physical interactions between the virulence-associated Region of Difference 1 genomic locus and host macrophages. To achieve this goal, we separated our research efforts into two main categories.

The first approach aimed at profiling at an unprecedented depth and coverage the host proteomic response of primary macrophages infected with WT and ΔRD1 *Mtb* upon dynamic time-course infections. By measuring the samples in a data independent acquisition fashion coupled to parallel accumulation serial fragmentation, we could monitor nearly 6,000 host proteins in a data-complete, robust and quantitatively accurate manner [1–3]. The first finding was that host macrophage responses were more impacted by the virulence of the strain used for the infection rather than the time of the infection time itself. Secondly, despite a nearly 5-fold increase of the intracellular bacterial burden in WT *Mtb* infections at the late-infection stage, WT strains dampen the proteomic dysregulation of host protein levels by a factor of 2 thus suppressing adequate immune responses. When we look in closer detail at the dysregulation of host proteins, the RD1 region specifically represses the cell adhesion molecules, proteins associated in the herpes simplex virus 1 infection and the complement and coagulation cascades including the Complement Receptors 1,3 and 5. By performing fuzzy c-means clustering to identify dynamic patterns in the host responses, amongst others we identify a gradual dysregulation of Rab-related signalling proteins. Although known Rab molecules known to be affected by *Mtb* infections were intact at their intensity levels (namely Rab5 and Rab7), we identified altered expression patterns for 5 distinct Rab molecules also involved in phagosomal maturation and intracellular recycling. By further overlapping our host proteomic profiling with siRNA screens aimed at identifying host restrictive or supportive factors for intracellular bacterial replication, we found an enrichment in the overlap with host proteins whose intensity levels are significantly altered upon WT infections. Indeed, *Mycobacterium tuberculosis* infections tend to increase the expression levels of host supportive factors, or conversely suppress host restrictive factors.

Our second approach was targeted towards mapping the exact physical protein-protein interactions between bacterial secreted proteins and the human host macrophages' proteins. To do so, we employed the robust and sensitive method based on affinity-purifications coupled to mass spectrometry, already widely applied in the field of host-pathogen interaction studies [4–17]. We first selected bacterial secreted proteins, including those secreted by the ESX-1 secretion system encoded in the RD1 locus, and identified 207 novel *Mtb*-human protein-protein interactions between 142 host proteins and 14 *Mtb* bacterial baits. We then overlapped our results with previously published *Mtb*-human maps [9] to evaluate the overlap and carry over a comprehensive analysis on host targeted proteins. Amongst the results, we found that only 48% of tested bacterial baits had high confidence host interactors. Secondly, we found that 8.2% of the host counterparts were targeted by more than one bacterial bait, suggesting a redundancy in the *Mtb*-human interactions to ensure the robustness of the intervention. Thirdly, we identified the 5 Rab proteins known prevent in the phagosomal maturation arrest of bacilli-containing phagosomes. Interestingly, these physical *Mtb* interactors were not dysregulated at the global proteome level, suggesting that *Mtb* employs 2 strategies to modulate host proteins: (i) at the overall protein intensity level and (ii) by physically interacting with them thus leading to a spatial reorganisation.

6.2 Outlook

Host-pathogen protein-protein interactions are essential mediators of any infection. In the case of *Mycobacterium tuberculosis* infection, this is corroborated by the fact that upon loss of the virulence-associated ESX-1 protein secretion apparatus, intervention in the host pathways is no longer carried out and consequently *Mtb* loses in virulence. Thus, mapping interspecies protein interactions can shed light on the molecular mechanisms of virulence-associated proteins, and teach us how pathogens may utilize host resources for their own benefit.

Most of the studies, including ours, have employed the high throughput PPI mapping method called affinity-purification coupled to mass spectrometry, or AP-MS in short. AP-MS consists of fusing an affinity tag to a bait protein, which are then overexpressed in the desired cellular environment. The tagged bait protein along with its direct and indirect interactors are co-purified and identified via bottom-up proteomics. Although highly robust and sensitive, AP-MS presents a few considerable general drawbacks [18]. First, the addition of tag fused to the carboxyl or amine end of the bait protein might impair certain of its interactions [19]. Additionally, it provides no or little stochiometric information on the retrieved proteins and provides no protein complex boundaries. More specifically to the field of bacterial-host interactions, it brings additional

challenges. The first is to identify all secreted proteins upon infection, where *in silico* predictions and experimental findings don't always corroborate [20]. Secondly, due to their increased genomic complexity compared to viruses, the generation of transgenic cell lines to ectopically express each putative secreted protein would be highly time-consuming. Thirdly, bacterial systems generally lack adequate genetic tools preventing endogenous tagging of their secreted proteins. Fourthly, certain protein-protein interactions might differ from overexpression systems to the physiological context of a real infection, due to altered expression profiles and spatial organisation of the proteome responses [5,9,21].

Thusly, we hypothesize that more global approaches for bacterial - host PPI detection may be useful. In the last years, numerous groups have been working towards developing methods which do not require genetically engineered cells to systematically identify in an unbiased manner endogenous protein complexes in physiological samples. Those rely on correlating protein elution profiles from co-fractionated cell lysates (CoFrac-MS) across various biochemical separations or chromatographic techniques [22,23]. Not only does this mass-spectrometry based approach yields lists of putative protein complexes, but it also reports stoichiometric and quantitative information for the identified components. Unfortunately, despite tremendous improvements in the field, the sensitivity remains the limiting factor. It is especially problematic in infectious diseases, where the dynamic ranges in terms of protein abundances from pathogenic organisms are generally several orders of magnitude lower than those of the host proteome [24]. Additionally, CoFrac-MS studies are highly limited in feasibility due to lengthy and cumbersome sample preparation workflows, high time-consuming MS acquisition requirements and the lack of available software to detect quantitative changes in the proteome organisation in an automated manner. To address these limitations, in the 5th chapter of this thesis we improved and optimized a workflow that enables deep and rapid analysis of the complexomes across multiple conditions. This workflow is based on the native extraction of protein complexes from a mild lysis of cells which are subsequently separated by high-resolution Size Exclusion Chromatography (SEC) according to their hydrodynamic radius. The proteins contained in each fractions are then identified and quantified via SWATH-MS [1], a Data Independent Acquisition (DIA) scheme enabling a highly reproducible, robust and sensitive quantification of peptides across multiple samples [25]. As a proof-of-concept, we applied the workflow to study the THP-1 differentiation from a monocytic precursor cell-line towards a macrophage-like cell line and subsequent LPS-stimulation to mimic a bacterial infection.

Because our results show that SEC-SWATH/MS could capture stable protein complexes from macrophages, including complexes targeted by *Mycobacterium tuberculosis* uncovered in chapter

5 such as the endosomal-sorting WASH complex, the ribosome, CCT or the spliceosome complex, we hypothesize that performing SEC-SWATH/MS on infected macrophages with *Mtb* could provide for the first time physiologically relevant *Mtb*-human interactions at the proteome-wide scale. In order to generate the host-pathogen complex hypothesis, we could supplement Corum [26] and String [27] based complex hypothesis with results obtained from the existing AP-MS protein interactions. However, if the aim would be to uncover novel host-pathogen interactions, software solutions relying on *de novo* identification of protein-protein based on the protein co-elution profiles should be employed or manual query of the data using a visualization software [21] could be used. By mapping physiologically relevant host-pathogen protein-protein interaction upon a real infection, this could lead to the identification of pathogen-mediated immuno-modulation and thus to the discovery of potential host-directed therapeutic intervention targets.

6.3 References

1. Gillet LC, Navarro P, Tate S, Röst H, Selevsek N, Reiter L, Bonner R, Aebersold R: **Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. [Internet]**. *Mol. Cell. Proteomics* 2012, **11**:O111.016717.
2. Röst HL, Rosenberger G, Navarro P, Gillet L, Miladinović SM, Schubert OT, Wolski W, Collins BC, Malmström J, Malmström L, et al.: **OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data [Internet]**. *Nat. Biotechnol.* 2014, **32**:219–223.
3. Meier F, Brunner A-D, Frank M, Ha A, Voytik E, Kaspar-Schoenefeld S, Lubeck M, Raether O, Aebersold R, Collins BC, et al.: **Parallel accumulation – serial fragmentation combined with data-independent acquisition (diaPASEF): Bottom-up proteomics with near optimal ion usage [Internet]**. *bioRxiv* 2019, doi:10.1101/656207.
4. Ramage HR, Kumar GR, Verschueren E, Johnson JR, Von Dollen J, Johnson T, Newton B, Shah P, Horner J, Krogan NJ, et al.: **A Combined Proteomics/Genomics Approach Links Hepatitis C Virus Infection with Nonsense-Mediated mRNA Decay [Internet]**. *Mol. Cell* 2015, **57**:329–340.
5. Mirrashidi KM, Elwell CA, Verschueren E, Johnson JR, Frando A, Von Dollen J, Rosenberg O, Gulbahce N, Jang G, Johnson T, et al.: **Global Mapping of the Inc-Human Interactome Reveals that Retromer Restricts Chlamydia Infection [Internet]**. *Cell Host Microbe* 2015, **18**:109–121.
6. Auweter SD, Bhavsar AP, de Hoog CL, Li Y, Chan YA, van der Heijden J, Lowden MJ, Coombes BK, Rogers LD, Stoynov N, et al.: **Quantitative mass spectrometry catalogues Salmonella pathogenicity island-2 effectors and identifies their cognate host binding partners. [Internet]**. *J. Biol. Chem.* 2011, **286**:24023–35.
7. Kane JR, Stanley DJ, Hultquist JF, Johnson JR, Mietrach N, Binning JM, Jónsson SR, Barelier S, Newton BW, Johnson TL, et al.: **Lineage-Specific Viral Hijacking of Non-canonical E3 Ubiquitin Ligase Cofactors in the Evolution of Vif Anti-APOBEC3 Activity [Internet]**. *Cell Rep.* 2015, **11**:1236–1250.
8. Colpitts TM, Cox J, Nguyen A, Feitosa F, Krishnan MN, Fikrig E: **Use of a tandem affinity purification assay to detect interactions between West Nile and dengue viral proteins and proteins of the mosquito vector [Internet]**. *Virology* 2011, **417**:179–187.
9. Penn BH, Netter Z, Johnson JR, Von Dollen J, Jang GM, Johnson T, Ohol YM, Maher C, Bell SL, Geiger K, et al.: **An Mtb-Human Protein-Protein Interaction Map Identifies a Switch between Host Antiviral and Antibacterial Responses [Internet]**. *Mol. Cell* 2018, **71**:637-648.e5.
10. Jäger S, Cimermancic P, Gulbahce N, Johnson JR, McGovern KE, Clarke SC, Shales M, Mercenne G, Pache L, Li K, et al.: **Global landscape of HIV–human protein complexes [Internet]**. *Nature*

2011, **481**:365.
11. Watanabe T, Kawakami E, Shoemaker JE, Lopes TJS, Matsuoka Y, Tomita Y, Kozuka-Hata H, Gorai T, Kuwahara T, Takeda E, et al.: **Influenza Virus-Host Interactome Screen as a Platform for Antiviral Drug Development [Internet]**. *Cell Host Microbe* 2014, **16**:795–805.
12. Davis ZH, Verschueren E, Jang GM, Kleffman K, Johnson JR, Park J, Von Dollen J, Maher MC, Johnson T, Newton W, et al.: **Global Mapping of Herpesvirus-Host Protein Complexes Reveals a Transcription Strategy for Late Genes [Internet]**. *Mol. Cell* 2015, **57**:349–360.
13. Pichlmair A, Kandasamy K, Alvisi G, Mulhern O, Sacco R, Habjan M, Binder M, Stefanovic A, Eberle C-A, Goncalves A, et al.: **Viral immune modulators perturb the human molecular network by common and unique strategies [Internet]**. *Nature* 2012, **487**:486–490.
14. Germain M-A, Chatel-Chaix L, Gagné B, Bonneil É, Thibault P, Pradezynski F, de Chassey B, Meyniel-Schicklin L, Lotteau V, Baril M, et al.: **Elucidating novel hepatitis C virus-host interactions using combined mass spectrometry and functional genomics approaches. [Internet]**. *Mol. Cell. Proteomics* 2014, **13**:184–203.
15. White EA, Kramer RE, Tan MJA, Hayes SD, Harper JW, Howley PM: **Comprehensive analysis of host cellular interactions with human papillomavirus E6 proteins identifies new E6 binding partners and reflects viral diversity. [Internet]**. *J. Virol.* 2012, **86**:13174–86.
16. Rozenblatt-Rosen O, Deo RC, Padi M, Adelmant G, Calderwood MA, Rolland T, Grace M, Dricot A, Askenazi M, Tavares M, et al.: **Interpreting cancer genomes using systematic host network perturbations by tumour virus proteins. [Internet]**. *Nature* 2012, **487**:491–5.
17. White EA, Sowa ME, Tan MJA, Jeudy S, Hayes SD, Santha S, Münger K, Harper JW, Howley PM: **Systematic identification of interactions between host cell proteins and E7 oncoproteins from diverse human papillomaviruses. [Internet]**. *Proc. Natl. Acad. Sci. U. S. A.* 2012, **109**:E260-7.
18. Morris JH, Knudsen GM, Verschueren E, Johnson JR, Cimermancic P, Greninger AL, Pico AR: **Affinity purification-mass spectrometry and network analysis to understand protein-protein interactions**. *Nat. Protoc.* 2014, **9**:2539–2554.
19. Majorek KA, Kuhn ML, Chruszcz M, Anderson WF, Minor W: **Double trouble - Buffer selection and his-tag presence may be responsible for nonreproducibility of biomedical experiments**. *Protein Sci.* 2014, **23**:1359–1368.
20. Wang G, Xia Y, Song X, Ai L: **Common Non-classically Secreted Bacterial Proteins with Experimental Evidence [Internet]**. *Curr. Microbiol.* 2016, **72**:102–111.
21. Heusel M, Frank M, Köhler M, Amon S, Frommelt F, Rosenberger G, Bludau I, Aulakh S, Linder MI, Liu Y, et al.: **A Global Screen for Assembly State Changes of the Mitotic Proteome by SEC-SWATH-MS [Internet]**. *Cell Syst.* 2020, **0**.
22. Havugimana PC, Hart GT, Nepusz TS, Yang H, Turinsky AL, Li Z, Wang PI, Boutz DR, Fong V, Phanse S, et al.: **A Census of Human Soluble Protein Complexes [Internet]**. *Cell* 2012, **150**:1068–1081.
23. Kristensen AR, Gsponer J, Foster LJ: **A high-throughput approach for measuring temporal changes in the interactome [Internet]**. *Nat. Methods* 2012, **9**:907–909.
24. Schmidt F, Völker U: **Proteome analysis of host-pathogen interactions: Investigation of pathogen responses to the host cell environment [Internet]**. *Proteomics* 2011, **11**:3203–3211.
25. Collins BC, Hunter CL, Liu Y, Schilling B, Rosenberger G, Bader SL, Chan DW, Gibson BW, Gingras A-C, Held JM, et al.: **Multi-laboratory assessment of reproducibility, qualitative and quantitative performance of SWATH-mass spectrometry [Internet]**. *Nat. Commun.* 2017, **8**:291.
26. Ruepp A, Waegele B, Lechner M, Brauner B, Dunger-Kaltenbach I, Fobo G, Frishman G, Montrone C, Mewes HW: **CORUM: The comprehensive resource of mammalian protein complexes-2009 [Internet]**. *Nucleic Acids Res.* 2009, **38**:D497-501.
27. **The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible [Internet]**. [date unknown], [no volume].

Chapter 7

Acknowledgments

This thesis marks the end of four incredible years in the Aebersold group. This achievement could not have been possible without the help, guidance and support of countless people whom I was lucky to meet over the years. I would thus like to take a few lines to acknowledge them and express my gratitude.

First and foremost, I would like to thank Ben Collins for trusting me and giving me the opportunity to become the first PhD student in his group. Over the years, I highly enjoyed his guidance and support. He gave me the freedom to pursue my scientific interests, but was always present when help was needed. Ben's knowledge and technical skills in proteomics are an inexhaustible source of help for all of us around him. He has been a great teacher.

Secondly but just as importantly, I would like to thank Prof. Aebersold for giving me the opportunity to work in his group. Over the years, he has been an inspiring mentor. His critical thinking, positivity and ability to boil down complex subjects to their essential is always an inspiration. He provided an outstanding environment, filled with incredible scientists across a wide array of discipline.

I am very grateful to Prof. Gagneux, who gave me the opportunity to work in his lab at the Swiss Tropical Health Institute, where I discovered the world of *Mycobacterium tuberculosis* research in the BSL3 environment. It has been a real pleasure working in his group.

I would like to thank my fourth committee member, Prof. Hardt. His knowledge about intracellular bacterial pathogens and years of expertise and research in this field has yielded many useful advices during my committee meetings.

Further, I would like to thank Audrey van Drogen. Theoretically, she was our lab manager but in reality, she played a much larger role than this. I thank her for all the scientific help she provided,

but more importantly I would like thank her for having become an incredible friend and a great source of laughter and happiness in and outside of the lab.

I would like to thank all the past and current co-workers I have had the pleasure to meet over the years. I would like to thank first my fellow PhD buddies which whom we shared lots of great moments and discussions within and outside the PhD meetings: Moritz Heusel, Isabell Bludau, Amir Banaei-Esfahani, Chris Sarnowski, Fabian Frommelt, Andrea Fossati, Federico Uliana, Betty-Friedrich-Grube and Jelena Cuklina. Additionally, I would like to thank my co-workers who were always helpful when dealing with the mass spectrometers and spent many hours making sure we could all measure our samples, Ludovic Gillet, Alexander Leitner, Martin Mehnert and Matej Vizovisek. I would like thank Claudia Martelli, Peng Xue, Sabine Amon, and Chen Li for having been great office-mates. I would like to thank Evan Williams, Patrick Pedrioli, Mathias Gstaiger, Tatjana Sajic, and Emanuela Milani for all the scientific and non-scientific discussions.

I am indebted and grateful to all the team at the Swiss Tropical Health Institute who trained me and who helped me navigate in the BSL3 environment. Thus, I would like to thank Sonia Borell, Miriam Reinhard, Ainhoa Arbues Arribas and Damian Portevin specifically. I would also like to thank the all other co-workers, including Rhastin Casto, Chloé Loiseau, Selma Mode and Andrej Trauner and Sebastian Gygli, who were very welcoming and helpful, and provided a great working environment.

I would like to especially acknowledge my collaborator Johannes Nemeth, with whom I worked on various projects. Working with him was always a source of energy and motivation. I would like to thank him for all the efforts and experiments he did as well as for all the discussions.

I would like to thank Pascal Kägi, Martin Pestalozzi, Edita Schröppel and Silas Krämer for always being so helpful and having their doors always open.

I would like to thank Sarah Cherkaoui, Silas Krämer, Yannick Severin, Klara Kropivsek, Pontrelli Samuel, Mariëlle van Kooten, Tomek Diederen and Thijs Wildschut for the countless laughers in the corridors, at the happy hours or wherever we met.

The last four years in Zürich would have been incredibly different if I hadn't been surrounded by the most cheering and wonderful group of friends. I would like to thank Laura Gabella, who has made Zürich a much brighter place and the past 20 years filled of laughter. I am forever thankful to live with Laurent Magnet and Gaetan Petit, who make everyday life after work so much fun. I

would like to thank my Zürich friends from the group “la belle”, namely Arthur, Ben, Gaetan, Gilles, Guillaume, Laura, Laurent, Leo, Lucas, Nicolas and Rachel who, during the past years which were in many ways an emotional rollercoaster, have definitely contributed to most highs.

Last but not least, I would like to express my deepest gratitude to my family to whom I owe everything. I would like to thank my parents Alain and Isabelle, for their unconditional support especially during this nearly decade-long academic journey. I would like to thank François, my big brother, for being in many ways a guide in our lives. Finally, I would like to thank Pips for being a source of inspiration and with whom everything is better when shared.

Charlotte Nicod
April 2020

Abbreviations

aLFQ	absolute label free quantification
AMF	assembled mass fraction
AP	affinity-purification
AP-MS	affinity-purification coupled to mass spectrometry
APEX	ascorbic acid peroxidase
BCG	Bacille Calmette-Guérin
BDQ	bedaquiline
BP	biological processes
CoF-MS	co-fractionation coupled to mass spectrometry
DDA	data dependent acquisition
DEG	differential gene expressions
DIA	data independent acquisition
DNA	deoxyribonucleic acid
DRP	differentially regulated proteins
EMB	ethambutol
ESAT-6	early secreted antigen target of 6 kDa
ESCRT	endosomal sorting complex required for transport
ESI	electrospray ionization
FASP	filter-aided sample preparations
FDR	false discovery rates
FMF	feature specific mass fraction
GFP	green fluorescent protein
GO	gene ontology
HDT	host directed therapies
HP	host-pathogen
HP-PPI	host-pathogen protein-protein interaction
IL	interleukin
INH	isoniazid
IP	immuno-purification
LC	liquid chromatography
LC-MS/MS	liquid chromatography coupled to tandem mass spectrometry
LPS	lipopolysaccharides
m/z	mass over charge ratio
mBMDM	mouse bone marrow derived macrophages
MDR	multi-drug resistance
MRC	membrane-associated oxidoreductase complex
mRNA	messenger deoxyribonucleic acid
MS	mass spectrometry
MS/MS	tandem mass spectrometry
Mtb	*Mycobacterium tuberculosis*
PASEF	parallel accumulation - serial fragmentation
PE	proline - glutamic acid
PMA	phorbol 12-myristate 13-acetate

PPE	proline-proline-glutamic acid
PPI	protein-protein interaction
PZA	pyrazinamide
RD1	region of difference 1
RIF	rifampin
RNA	ribonucleic acid
RP-HPLC	reverse phase high pressure liquid chromatography
SEC	Size exclusion chromatography
SWATH-MS	sequential window acquisition of all theoretical mass spectra
TA	toxin-antitoxin
Tat	twin arginine translocation
TB	tuberculosis
TIMS	trapped ion mobility spectrometer
TLR	Toll-like receptor
TNF	tumor necrosis factor
WT	wild type
WXG	tryptophan-X-glycine
XDR	extended-drug resistance
XL-MS	chemical cross-linking coupled to mass spectrometry
Y2H	yeast two Hybrid