DISSERTATION | DOCTORAL THESIS

Titel | Title

TRANSCRIPTOMICS OF THP-1 MACROPHAGES & BORDETELLA PERTUSSIS TOHAMA I INTERPLAY

verfasst von | submitted by

Mariam Rukhama Farman MSc

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Doctor of Philosophy (PhD)

Wien | Vienna, 2025

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 794 685 490
Dissertationsgebiet lt. Studienblatt | Field of study as it appears on the student record sheet:	Molekulare Biologie
Betreut von | Supervisor:	Univ.-Prof. Dipl.-Phys. Dr. Ivo Hofacker

To my late parents, who consistently motivated me to pursue my education and fostered in me the ambition to achieve success.

To my family, for all the time I was separated from them due to my academic pursuits.

ACKNOWLEDGEMENTS

This thesis signifies the culmination of my doctoral studies. Collaborating with the theoretical biochemistry department at the University of Vienna has provided a valuable learning opportunity. I am sincerely grateful to everyone who contributed in various ways to the achievement of this degree and made the journey memorable.

Dr. Ivo Hofacker – Ph.D. Supervisor I wish to sincerely thank you for the opportunity to pursue my Doctoral degree under your guidance, and for your patience over the past years. Your support and encouragement have been invaluable to my growth, both professionally and personally, whether I was dealing with complex technical issues, preparing talks and posters, drafting documents, or refining my English. It has truly been a pleasure working with you.

Dr. Fabian Amman – Ph.D. Advisor I will be forever thankful for your invaluable input, pieces of advice, mentorship, insight, and support throughout the Ph.D. He has been a terrific mentor academically, professionally, and personally.

Dr. Branislav Večerek - Collaborator I feel fortunate to have had the opportunity to become acquainted and work with you on a very fruitful collaboration during my Ph.D.

Farman & Bahar - Parents No one was more instrumental in helping me through this process than they were. I am blessed to have them. They have always been a constant source of love, support, and inspiration without any conditions. I will miss you forever.

Saad Altaf – Husband Thank you for your affection, understanding, and for sharing this dream with me.

The angels of mine - Halah and Isa I sincerely apologize for working long hours and weekends to complete my Ph.D., which may have taken away some of our precious moments together. I hope that someday you'll understand that pursuing a Ph.D. is a challenging journey, and recognize that while I am not perfect, I always tried my best. All of this is ultimately for your benefit. I want you to know that you are the source of my happiness, hope, and unconditional love.

Financial support

I am very grateful for the financial support the Austrian Science Fund provided under project I 2353-B22.

ABSTRACT

Infection is a dynamic process characterized by the invasion and proliferation of a pathogen within a host. RNA-seq is an effective method for quantifying the full spectrum of transcripts during infection. Instead of physically separating the interacting organisms, analyzing their relationship now requires less invasive experimental protocols. This is facilitated through dual RNA-seq, wherein the entire RNA of the host-pathogen system is extracted, and their respective transcriptomes are separated *in silico*, allowing for the concurrent examination of gene expression in these organisms as well as their direct interaction analysis during infection.

This study demonstrates the application of dual RNA-seq to investigate the processes involved in the infection of human THP-1 macrophages by Bordetella pertussis Tohama I (*B. pertussis*). Despite having both whole-cell and acellular vaccines, pertussis keeps re-emerging. The recent resurgence of pertussis underscores the imperative to broaden our comprehension of the molecular mechanisms that underlie the pathogenicity of *B. pertussis*. As part of my doctoral research, I conducted a comprehensive analysis of dual RNA-seq data using advanced bioinformatics techniques to elucidate the biological mechanisms of this obligate extracellular pathogen and to characterize the host immune response to infection. Our research identified: (i) despite being an extracellular pathogen, *B. pertussis* can survive inside macrophages; (ii) gene expression in THP-1 macrophages was extensively rewired; (iii) genes in THP-1 that encode cytokines, chemokines, and transcription regulators involved in initiating the M1 and M2 macrophage polarization pathways were highly expressed; (iv) THP-1 genes involved in controlling apoptosis and inflammation pathways frequently exploited by bacterial pathogens utilizing the Bvg- phase were overexpressed; (v) changes in expression of specific THP-1 gene subsets involved in signaling and metabolism pathways; (vi) a large number of genes involved in sulfur metabolism were significantly modulated within macrophages; (vii) and decreased expression of most virulence factors, along with increased activity of several transcriptional regulators of *B. pertussis*, during infection. These findings indicated that *B. pertussis* transitions from a Bvg+ (virulent) to a Bvg- (avirulent) phase to adapt to the intramacrophage environment and respond to antimicrobial activities triggered by THP-1 cells, a phenomenon not previously observed on such a scale.

To provide further insights into the complex interplay of THP-1 macrophages and *B. pertussis*, during the study, we developed a novel software tool, MSF using Java, to detect concertedly modulated sub-graphs in the THP-1 and *B. pertussis* to support our findings and identify genes that may strongly associate between the two species.

ZUSAMMENFASSUNG

Eine Infektion ist ein dynamischer Prozess, der durch das Eindringen und die Vermehrung eines Krankheitserregers in einem Wirt gekennzeichnet ist. RNA-seq ist eine effektive Methode zur Quantifizierung des gesamten Spektrums an Transkripten während einer Infektion. Anstatt die interagierenden Organismen physisch zu trennen, erfordert die Analyse ihrer Beziehung nun weniger invasive Versuchsprotokolle. Dies wird durch Dual-RNA-seq erleichtert, bei dem die gesamte RNA des Wirt-Erreger-Systems extrahiert und ihre jeweiligen Transkriptome *in silico* getrennt werden. Dieser Ansatz ermöglicht die gleichzeitige Profilerstellung der Genexpression in diesen Organismen und die Analyse ihrer direkten Interaktionen während der Infektion.

Diese Studie demonstriert die Anwendung von dualem RNA-Seq zur Untersuchung der Prozesse, die bei der Infektion menschlicher THP-1-Makrophagen durch Bordetella pertussis Tohama I (*B. pertussis*) eine Rolle spielen. Trotz der Verfügbarkeit von Ganzzell- und azellulären Impfstoffen tritt Pertussis immer wieder auf. Das jüngste Wiederauftreten von Pertussis unterstreicht die Notwendigkeit, unser Verständnis der molekularen Mechanismen, die der Pathogenität von *B. pertussis* zugrunde liegen, zu erweitern. Im Rahmen meiner Doktorarbeit habe ich eine umfassende Analyse von Dual-RNA-Seq-Daten unter Verwendung fortschrittlicher Bioinformatiktechniken durchgeführt, um die biologischen Mechanismen dieses obligat extrazellulären Erregers aufzuklären und die Immunantwort des Wirts auf die Infektion zu charakterisieren. Unsere Forschung ergab: (i) Obwohl es sich um einen extrazellulären Erreger handelt, kann *B. pertussis* in Makrophagen überleben; (ii) die Genexpression in THP-1-Makrophagen wurde umfassend umgestaltet; (iii) Gene in THP-1, die für Zytokine, Chemokine und Transkriptionsregulatoren kodieren, die an der Initiierung der M1- und M2-Makrophagen-Polarisationswege beteiligt sind, wurden stark exprimiert; (iv) THP-1-Gene, die an der Steuerung von Apoptose- und Entzündungswegen beteiligt sind, die häufig von bakteriellen Pathogenen unter Ausnutzung der Bvg-Phase genutzt werden, wurden überexprimiert; (v) Veränderungen in der Expression spezifischer THP-1-Genuntergruppen, die an Signal- und Stoffwechselwegen beteiligt sind; (vi) eine große Anzahl von Genen, die am Schwefelstoffwechsel beteiligt sind, wurden innerhalb der Makrophagen signifikant moduliert; (vii) und eine verminderte Expression der meisten Virulenzfaktoren, zusammen mit einer erhöhten Aktivität mehrerer Transkriptionsregulatoren von *B. pertussis* während der Infektion. Diese Ergebnisse deuteten darauf hin, dass *B. pertussis* von einer Bvg+ (virulenten) zu einer Bvg- (avirulenten) Phase übergeht, um sich an die intramakrophagale Umgebung anzupassen und auf die durch THP-1-Zellen ausgelösten antimikrobiellen

Aktivitäten zu reagieren, ein Phänomen, das bisher in diesem Ausmaß nicht beobachtet wurde.

Um weitere Einblicke in das komplexe Zusammenspiel von THP-1-Makrophagen und *B. pertussis* zu gewinnen, haben wir im Rahmen der Studie ein neuartiges Software-Tool namens MSF unter Verwendung von Java entwickelt, um gemeinsam modulierte Teilgraphen in THP-1 und *B. pertussis* zu erkennen, unsere Ergebnisse zu untermauern und Gene zu identifizieren, die möglicherweise in engem Zusammenhang zwischen den beiden Spezies stehen.

CONTENTS

1

PREAMBLE

Since the inception of my research trajectory, I have have been deeply fascinated by the intricate interplay between pathogens and their human hosts. The capacity of microbes to invade, adapt, and manipulate host biological systems, juxtaposed with the remarkable resilience and defense mechanisms of the human body, has consistently fueled my scientific curiosity. Upon engaging with the field of bioinformatics, I recognized that these host-pathogen interactions could be elucidated not solely through experimental laboratory methods but also via sophisticated computational techniques. This realization underscored the immense potential of integrative approaches, such as dual RNA sequencing, to dissect the molecular dialogue between host and pathogen with unprecedented resolution. This realization highlighted the immense potential of integrative strategies, such as dual RNA sequencing, to dissect the molecular dialogue between host and pathogen with unprecedented resolution, and it ultimately shaped the direction of this thesis.

As part of this project, we had two aims. The first was to conduct a dual RNA-seq experiment to investigate and gain a deeper understanding of the alterations in gene expression patterns of human THP-1 macrophages and *B. pertussis* during the progression of infection. Throughout my PhD research, I utilized a variety of bioinformatics methods and tools. I developed customized pipelines to analyze the dual RNA-seq data (produced in close collaboration with Dr. Branislav Večerek and Dr. Denisa Petráčková from the Czech Academy of Sciences, Prague, Czech Republic), which provided significant and valuable insights of the defense mechanism of the human THP-1 macrophages and the survival and infection mechanism of *B. pertussis* detailed in the pilot and comprehensive study. The project's second aim was to create coexpression networks between THP-1 and *B. pertussis* genes. We found that no existing tools were available for this purpose. Consequently, we developed a novel tool called Modulated Sub-graph Finder (MSF), which resulted in the first publication of my PhD. MSF uniquely identifies coordinated sub-graphs in an organism's overall cell signaling network. This tool was employed to generate the global signaling networks for humans. Subsequently, correlation methods were used to link these coexpression networks of THP-1 generated by MSF to *B. pertussis* genes.

The pilot investigation of dual RNA-seq preceded the comprehensive dual RNA-seq to assess the infection protocol and dual RNA-seq design for potential issues and to estimate the required time and resources. The pilot study was instrumental in shaping the comprehensive study by identifying critical parameters such as the multiplicity of infection (MOI), rRNA depletion, and

exclusion of RNA enrichment for *B. pertussis*, sequencing depth necessary to yield sufficient reads, the number of infection time points, specific hours post-infection to consider, the number of controls, and the required number of biological replicates. These details were integral in developing a successful dual RNA-seq data set for our comprehensive study.

This thesis is divided into five primary sections. **Chapter 1** presents an overview of essential background information regarding Bordetella Pertussis Tohama I. The chapter commences with a general introduction to *B. pertussis*, followed by detailed information on available vaccines against *B. pertussis*, its resurgence, clinical features of pertussis, and its pathogenesis. In **Chapter 2**, we explain next-generation sequencing and compare conventional RNA-seq with dual RNA-seq. We also describe the pipeline used to analyze our time-course dual RNA-seq data and the challenges encountered during the analysis. **Chapter 3** provides a brief overview of MSF software tool, detailing its development, implementation, and the packages used. In **Chapter 4**, the scientific contributions to the project are outlined, referencing three distinct publications. The first publication concerns the software tool MSF that we developed. The second publication presents the pilot study's findings, while the third details the comprehensive study conducted. These publications are arranged in chronological order according to their respective publication dates.

In conclusion, **Chapter 5** provides a comprehensive overview of the accomplishments attained within the scope of this thesis and delineates their implications for prospective research endeavors.

2

BORDETELLA PERTUSSIS TOHAMA I

Bordetella pertussis Tohama I (*B. pertussis*) is a Gram-negative bacterial agent causing highly contagious respiratory disease leading to pertussis or whooping cough in humans. The name 'whooping cough' was derived from the intense coughing fits that restrict airflow to the lungs, producing a distinctive 'whoop' sound. It was first isolated and implicated as the cause of the disease in 1906 by Bordet and Gengou [17].

Historically, pertussis has been a significant cause of illness and mortality worldwide, especially among infants. The widespread use of vaccinations has led to a substantial decrease in the number of cases among vaccinated populations. However, pertussis remains a leading cause of death from bacterial vaccine-preventable diseases. A recent assessment released by the European Centre for Disease Prevention and Control (ECDC) in May 2024 indicates a significant increase in pertussis cases within Europe. After a period of limited circulation, notably during the COVID-19 pandemic, the reported cases surpassed 25,000 in 2023 and exceeded 32,000 between January and March 2024. These figures are comparable to the incidences recorded in 2016 (41,026) and 2019 (34,468) [148].

2.1 BORDETELLA

2.1.1 *Genus of Bordetella*

To date, nine species of *Bordetella* have been identified, *Bordetella avium*, *Bordetella bronchiseptica*, *Bordetella hinzii*, *Bordetella holmesii*, *Bordetella parapertussis*, *Bordetella pertussis*, *Bordetella petrii*, *Bordetella pseudohinzii* and *Bordetella trematum*. Among these species, *Bordetella bronchiseptica*, *Bordetella parapertussis*, and *Bordetella pertussis* are the most well-known and studied due to their common ancestor and association with human disease [166]. Although it has evolved from a common ancestor shared with other Bordetella species, such as *Bordetella bronchiseptica*, which infects a broader range of hosts, primarily mammals. *Bordetella pertussis* is a pathogen that exclusively affects humans. One of the primary distinctions between these two species is that *B. pertussis* has undergone extensive gene loss as a result of its specialization for infecting humans [107].

2.1.2 *Genome of Bordetella Pertussis Tohama I*

Bordetella pertussis Tohama I genome was first published in 2003 [124]. It was isolated in Japan in the 1950s. It has a genome of approximately 4.1 Megabase pairs (Mpb), which consists of 3,981 annotated genes with average gene size of 978 base pairs (bp). The Guanine-Cytosine (GC) content is 67.72%. There are three ribosomal RNA (rRNA) operons, 51 transfer ribonucleic acid (tRNAs), and 261 insertion (IS) elements, of which 238 belong to the family IS481.

B. pertussis, the bacterium responsible for whooping cough, exhibits genome plasticity, allowing it to undergo various mechanisms such as deoxyribonucleic acid (DNA) inversion, rearrangements, deletion, duplication, and acquisition of additional genetic information. This genetic adaptability contributes to *B. pertussis*'s ability to evade the host immune response and develop resistance to antibiotics [37].

2.1.3 *Vaccines against B. pertussis*

Vaccination against *B. pertussis* plays a central role in controlling the spread of the disease by reducing the severity of symptoms and limiting bacterial transmission. The primary vaccines used to immunize against *B. pertussis* include whole-cell and acellular vaccines.

2.1.3.1 *Whole-cell Vaccine*

The first complete whole-cell vaccine targeting combined pertussis, diphtheria, and tetanus (DTP) was developed in the 1930s and subsequently adopted by most developed nations during the 1940s and 1950s [25]. These vaccines comprised inactivated bacterial cells, chemically killed for safety. Following the implementation of vaccination programs, a significant decline in disease incidence was observed shown in **Figure 1**, showing that whole-cell vaccination effectively protected infants and children from classic pertussis, commonly called whooping cough. During that era, this was the main and most well-known type of *B. pertussis* infection. [157].

However, whole-cell vaccinations were traditionally viewed as 'reactogenic', leading to side effects such as fussiness, soreness, and swelling at the injection site. More serious adverse events linked to the whole-cell pertussis (wP) vaccine include seizures, neurological issues, and, in rare cases, sudden death in infants [25].

2.1.3.2 *Acellular Vaccine*

An acellular pertussis vaccine (aP) was designed as DTaP (diphtheria, Tetanus, and pertussis) due to the adverse reactions associated with the whole-cell vaccine [112] and was introduced in 1996 and as tdap booster in 2005. Acellular

pertussis vaccines are made up of isolated antigens from *B. pertussis*. The current acellular pertussis vaccines contain up to five *B. pertussis* components: PT, FHA, FIM2 and FIM3, and Pertactin (PRN) [41]. The efficacy to prevent pertussis in aP vaccines appears to be slightly lower than that of whole-cell vaccines, but they have better safety profiles than a whole-cell vaccine in terms of causing adverse reactions [55].

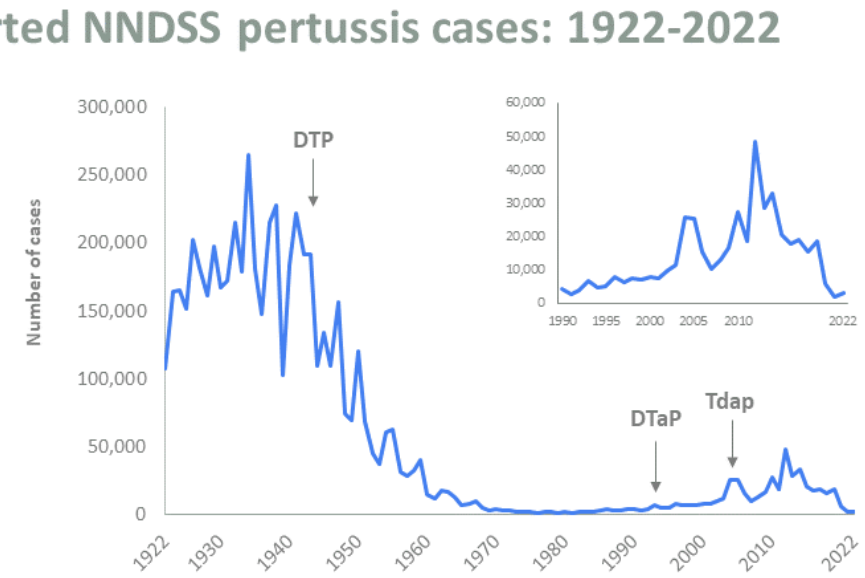


Figure 1: Introduction of DTP, DTap and Tdap vaccines in United States of America. The whole cell vaccine, labeled DTP here, was made widely available from 1948 and was replaced by the acellular vaccine, labeled DTaP, in 1996 and tdap in 2005. A sharp drop in pertussis cases is seen with the introduction of the DTP. The recent resurgence is seen in the inlay [157].

2.1.4 *Effectiveness of Pertussis Vaccines*

A distinction has been noted in the type of immunity triggered by acellular versus whole-cell vaccines. Nonetheless, both vaccine types stimulate T-cells, which are crucial for defending against *B. pertussis*. The whole-cell vaccine induces a T-helper 1 (Th1) response due to the secretion of IFN-γ. The acellular vaccine, on the other hand, has a mixed Th1 and T-helper 2 (Th2) response due to secretion of IFN-γ and IL-5 [161]. Th1 cells support cell-mediated immunity, which is essential for defending against intracellular pathogens such as viruses and certain bacteria. The cytokine IFN-γ activates macrophages and enhances the elimination of pathogens by infected cells. Conversely, Th2 cells drive humoral immunity by stimulating B cells to produce antibodies, particularly in response to extracellular pathogens like helminths. Additionally, IL-5 plays a role in activating and recruiting eosinophils, which are crucial in combatting parasitic infections. Compared to the acellular vaccine, the whole-cell vaccine

more closely resembled a naturally occurring infection. It was observed that the Th1 immune response played a vital role in stimulating the destruction of *B. pertussis*; At the same time, it was suggested that the Th2 response play a crucial role in triggering the production of antibodies by activating B cells. They achieve this by releasing cytokines such as IL-4, IL-5, and IL-13, which facilitate the differentiation of B cells into plasma cells. Subsequently, these plasma cells generate antibodies like IgG and IgE. These antibodies then coat the surface of bacteria through a process called opsonization. Once opsonized, phagocytes such as macrophages and neutrophils recognize the bound IgG through their Fc receptors, thereby aiding in the engulfment of the bacteria and leading to their destruction. The active toxin, Pertussis toxin (PT) and lipopolysaccharide of *B. pertussis* provoke Th1 response; these antigens are included in the whole-cell vaccine but are absent from the acellular vaccine [94].

The different responses of the two vaccines were tested, showing that individuals vaccinated with the whole-cell pertussis vaccine demonstrated greater protection against the disease compared to those who received the acellular vaccine [167]. The problem with the current vaccines is the short-lived immunity against the infection. The whole-cell vaccine offers effective protection against infection but can cause adverse side effects. In contrast, an acellular vaccine elicits a different immune response with fewer side effects, but declines faster than the whole-cell vaccine due to having fewer antigens [8]. Even though the whole-cell vaccine provides a more prolonged duration of immunity, unfortunately, the whole-cell vaccine and the acellular vaccine are unsuccessful in inducing sterilizing immunity [90, 160, 168].

2.1.5 *Resurgence*

The resurgence of pertussis has been noted worldwide despite the implementation of widespread vaccination programs (**Figure 2**). Following several years of limited circulation, particularly during the COVID-19 pandemic when measures such as lock downs and social distancing restricted the transmission of various infectious diseases, concerning trend has emerged. Between 2023 and April 2024, European countries collectively recorded nearly 60,000 cases of pertussis, representing more than a tenfold increase compared to both 2021 and 2022. This rise in cases is reminiscent of previous years, with 41,026 cases reported in 2016 and 34,468 in 2019 [127]. Notably, infants (those under one year of age) accounted for the highest incidence in 17 European countries. Conversely, in six countries, the highest incidence was reported among adolescents aged 10 to 19. In France, a significant outbreak occurred in 2024, marking the largest resurgence in 25 years. As of mid-2024, over 5,600 cases were reported, with a notable increase in severe cases, especially among unvaccinated infants, leading to 35 deaths by September 2024. Austria had the highest rates of cases per 100,000 population in infants, with 180.4 reported cases. Tragically, the ma-

jority of pertussis-related deaths occurred in infants. Across all age groups, the number of female notifications was higher than that of males [127].

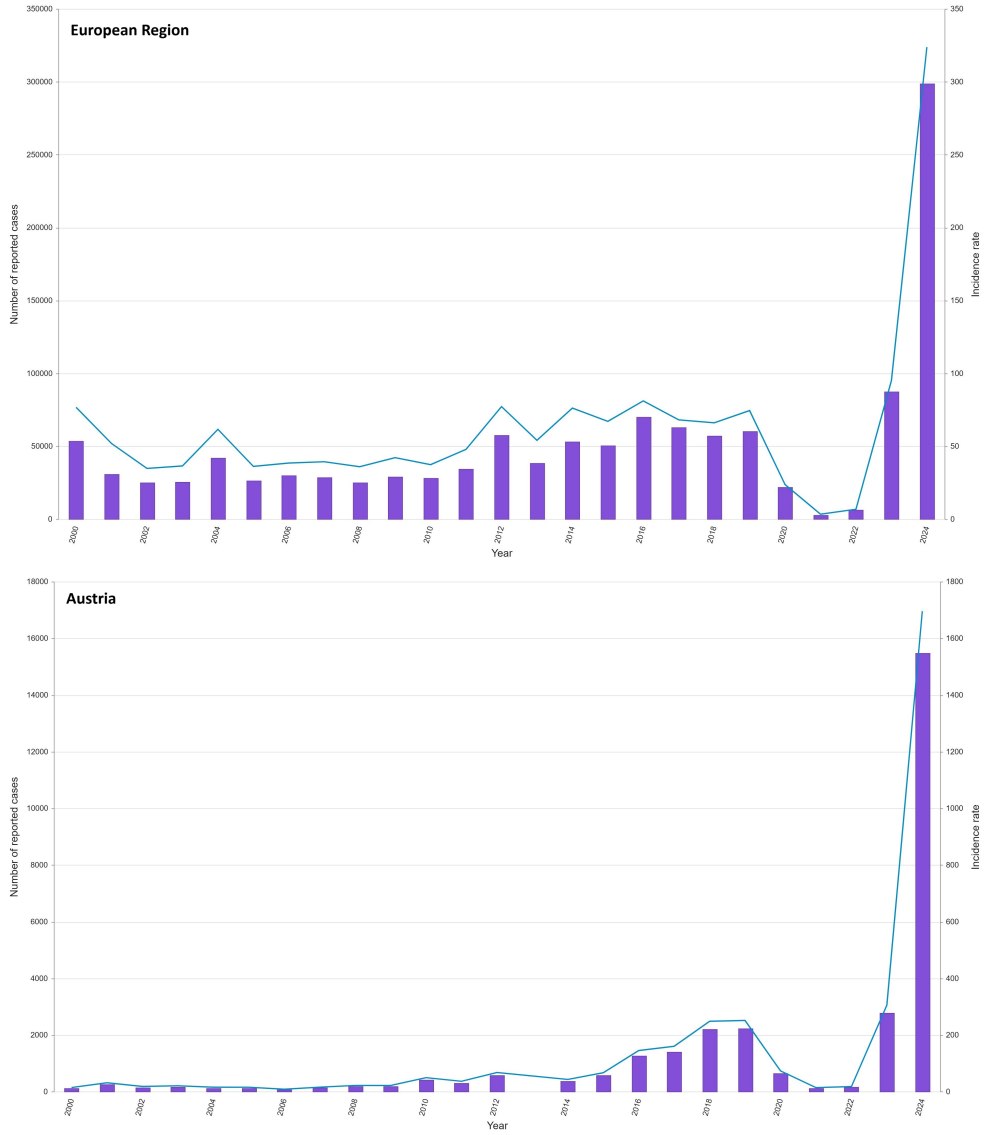


Figure 2: Reported pertussis cases from 2000 to 2024 are shown in European region, and in Austria. The top graph shows European region data, and the bottom on Austria [157]. Purple bars indicate case numbers, while blue lines show incidence rates per million

The resurgence of *B. pertussis* can be attributed to a variety of factors, including vaccine coverage gaps, immune responses, pathogen evolution, and waning immunity over time, among others. Understanding these contributing elements is crucial for developing effective strategies to control and prevent pertussis outbreaks.

2.1.5.1 *Waning Immunity*

Despite high vaccine coverage, cyclic epidemics of pertussis have been observed. Emerging reports suggests that the resurgence of pertussis could be because of replacement of the whole-cell vaccine with the acellular pertussis vaccine, though the exact reasons are still unclear. While the aP vaccine is less reactogenic, it may result in sub-optimal and waning immunity [92, 137].

Similar to immunization, contracting *B. pertussis* naturally does not necessarily guarantee lifelong immunity to pertussis [45]. Various studies have recorded instances of individuals experiencing a second episode of pertussis several years after the initial one. Currently, the length of protection provided by natural infection is estimated to be anywhere from 7 to 10 years [21] up to 20 years [83], though there are studies suggesting it can be as brief as 3.5 years [162].

Immunization with wP vaccine provides a similar or slightly reduced protection duration. In a 1962 study conducted by Lambert [84] in Kent County, Michigan, USA, it was observed that the frequency of pertussis cases among vaccinated people increased with the amount of time since their previous immunization with the whole-cell pertussis vaccine. Among the 210 individuals vaccinated against pertussis, the observed attack rates were 21% for those vaccinated within the past 4 years, 47% for those vaccinated 4 to 7 years ago, and 65% for those vaccinated 8 to 11 years prior. Jenkinson [72] carried out a research study in the UK spanning a decade, which found that the protection offered by the whole-cell pertussis vaccine remained effective in 85% of children four years post-immunization. However, this effectiveness decreased to 50% over the next three years. These studies indicate that wP vaccine does not confer permanent immunity and wanes over time. This underscores the challenge of achieving long-term immunity and underscores the significance of sustained immunization strategies in the management of pertussis outbreaks.

In comparison to wP vaccine, the immunity duration provided by aP vaccine seems to be shorter. Reports have also suggested that subjects who were fully vaccinated with aP vaccine had a higher risk of contracting pertussis earlier than those vaccinated with wP vaccine [27, 155]. Clark et al. [27] found in 2012 that children who received an acellular pertussis vaccine during infancy faced a higher risk of pertussis early in their school years, while those vaccinated with a whole-cell pertussis vaccine were more susceptible later, predominantly during adolescence. Similarly, Vickers et al. [155] observed that children vaccinated with an aP vaccine in infancy were at risk of pertussis within the first four years of life, whereas children who received the wP vaccine did not develop pertussis until ages 5 to 9.

2.1.5.2 *Immune Responses to Pertussis Vaccines*

Studies have shown that immune responses following natural *B. pertussis* infection differ from those induced by both whole-cell pertussis vaccines (wPVs) and acellular pertussis vaccines (aPVs). They reveal that the immune activation caused by aPVs is distinct from that resulting from natural infection and wPVs [43, 44]. The wP vaccine is typically administered via intramuscular injection as part of combination vaccines, such as DTP (diphtheria, tetanus, pertussis). This type of vaccine has been shown to elicit a robust mucosal immune response by stimulating both Th1 and Th17 pathways. These pathways are essential for producing secretory IgA, the primary antibody type present in mucosal surfaces, such as the respiratory tract. IgA plays a key role in neutralizing pathogens at mucosal surfaces before they can invade the body. In contrast, aP vaccine is also administered intramuscular, often in combination with other antigens like DTaP or Tdap, but it generally induces a Th2-dominant immune response, resulting in a strong systemic immune response but less effective mucosal immunity. This is primarily due to its induction of a systemic humoral response dominated by IgG rather than IgA [45].

2.1.5.3 *B. pertussis Genetic Modifications*

Over the past two decades, the incidence of the disease has shifted from infants to adolescents and adults [6, 63]. This might be due to the reason that *B. pertussis* is evolving in response to vaccination or as a consequence of *B. pertussis* adaption to aP vaccine use [20]. Alternatively, this could be attributed to advancements in screening programs and diagnostic tools.

Genes encoding antigens in aP vaccines evolved rapidly after introduction to pediatric immunization [141]. Studies have shown that the use of pertactin containing vaccines has led to the evolution of PRN-negative *B. pertussis* strains. Although there have been some exceptions, it has been demonstrated that the emergence of PRN-deficient strains is a direct result of the selective pressure induced by aP vaccination [10, 19]. No PRN-deficient isolates have been detected in Denmark, where an aP vaccine without PRN is used [171]. The data from Japan clearly indicates that the prevalence of PRN-negative strains has remained consistent over time, 41% from 2005 to 2007, 35% from 2008 to 2010 and 25% from 2011 to 2013, due to the administration of aPVs with PRN [110, 121, 171]. Another study that was carried out in the Netherlands showed variations in antigens between circulating strains and strains used in making the vaccines [112].

Recent studies have revealed that *B. pertussis* undergoes antigenic drift [109], which alters its infection mechanism and enables the bacterium to evade the host immune system. This particularly impacts vaccinated individuals, diminishing the effectiveness of the pertussis vaccine and leading to reduced immunity and increased susceptibility to pertussis. Consequently, there is an urgent

need for comprehensive research on the antigenic drift of *B. pertussis* and the resurgence of pertussis, ultimately paving the way for improved preventive and therapeutic strategies.

2.2 PERTUSSIS: THE DISEASE

2.2.1 *Clinical Features*

Clinical pertussis has a natural history consisting of three distinct stages. The first stage, known as Catarrhal, usually occurs when an individual contracts *B. pertussis* from infected droplets transmitted by other hosts. The bacteria have a strong attraction to the cilia found in the respiratory mucosa, which is where the infection primarily occurs [122]. The bacteria then multiply on the ciliated mucosal surface causing damage to the respiratory epithelium and resulting in ciliostasis, mucus release, and an inflammatory response in the respiratory tract. During the first stage of infection, mild symptoms, similar to those of a common cold, can be experienced such as sneezing, mild cough, and runny nose.

This stage is followed by the "Paroxysmal phase," where severe paroxysmal coughing, one of the classic symptoms of pertussis, can be experienced. During pertussis, individuals may experience frequent and intense coughing fits that can occur multiple times per hour for several weeks. These episodes can leave the person exhausted and struggling to catch their breath. The coughing often leads to a distinct "whoop" sound when inhaling. Despite ongoing research, the exact cause of these paroxysmal coughing fits remains unknown. It was once believed that pertussis only affected infants and children, but now it is understood that people of all ages are at risk of infection [136]. However, infants are particularly vulnerable to severe complications, such as seizures and choking, which can be life-threatening [98, 136]. Breathing difficulties known as apnoea are more common in newborn babies [136]. Unfortunately, this can result in pulmonary hypertension, leading to cardiac failure and death, particularly in previously healthy infants. This happens when pneumonia and apnoea work together to cause hypoxemia, which leads to pulmonary vasoconstriction. Severe pertussis is often characterized by pronounced leukocytosis, increasing vascular resistance. All these factors can result in pulmonary hypertension, cardiac failure, and even shock [122]. Older children and adolescents can also suffer from pertussis, but it is generally milder than infants. However, a systematic study by Cherry and Paddock [26] reveals that pertussis in adults and adolescents is not always like classic whooping cough. The burden of the disease in adults is unknown due to misdiagnosis [97].

As the patient recuperates during the final phase, coughing diminishes gradually, and the adaptive immune system helps eliminate bacteria from the respiratory system.

2.2.2 *Pathogenesis of B. pertussis*

Understanding the pathogenesis of *B. pertussis* is crucial for developing effective prevention strategies and vaccines. The pathogenesis of *B. pertussis* involves a series of steps that allow the bacteria to colonize and persist in the respiratory tract, evade the host immune response, and cause the characteristic symptoms of the disease. Before that, we should understand the different components of *B. pertussis* that it uses to get into the host and then prevail inside. A diagram illustrating the different components of *B. pertussis* involved in its infection is shown in **Figure 3**.

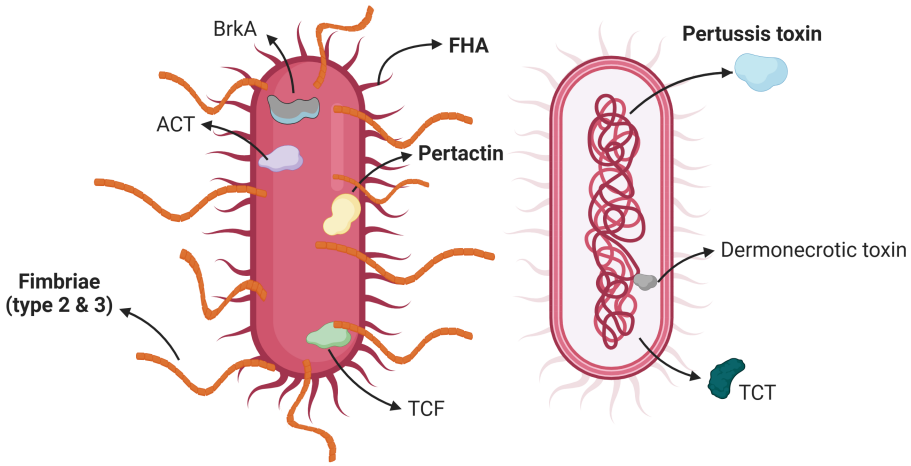


Figure 3: Diagram of *B. pertussis* organism. A schematic diagram illustrating *B. pertussis*, along with the antigens and structures crucial to its ability to cause whooping cough. The antigens included in acellular whooping cough vaccines are highlighted in bold.

2.2.2.1 *Adhesion Factors*

B. pertussis colonizes in the host's respiratory tract when the bacteria are inhaled [175]. Filamentous hemagglutinin (FHA), coded for by *fhaB*, is a protein produced by *B. pertussis*. It has multiple binding sites; an Arg-Gly-Asp triplet in the middle of the protein helps *B. pertussis* bind to macrophages, monocytes, ciliated respiratory cells, and nonciliated epithelial cells [60, 103]. FHA is a potent immunogen; due to its role in attachment and immunogenicity, it is one of the major virulence factors.

Fimbriae are filamentous structures composed of polymeric proteins found on the cell surface, encoded by the genes *fim2*, *fim3*, and *fimX*. These structures are responsible for binding to host cells [111]. A common component of fimbriae is *FimD*, which is located at the tip of the fimbriae [52]. These fimbriae are known to help *B. pertussis* in colonization, specifically binding to the monocytes and sugars in the epithelium [53, 61]. During the Bvg+ growth phase,

these genes are also active and, like other virulence-associated genes, have been found to have a significant role in the pathogenesis of *B. pertussis* and its ability to attach to host cells [165]. FIM proteins are also involved in bacterial agglutination [87].

Pertactin is a protein with the identical Arg-Gly-Asp triplet as FHA; it is an autotransporter. It is thought to be involved in attachment [42]. Antibodies are activated when PRN is detected. These antibodies prevent its attachment and encourage phagocytosis [65, 144]. Other autotransporters thought to play a role in colonization, adherence, and resistance include tracheal colonization factor (TcfA), BrkA, and Vag8 [47].

2.2.2.2 *BvgAS System*

The BvgAS system, a two-component regulatory mechanism in *Bordetella pertussis*, is responsible for numerous virulence factors. Activation of the sensor protein kinase BvgS and subsequent phosphorylation of the response regulator BvgA regulate the BvgAS genes. Upon entering the host, BvgS within the *bvgAS* locus is activated by an as-yet-unidentified signal, leading to the Bvg+ phase [60]. This phase initiates the transcription of genes encoding adhesins, toxins, and other virulence factors, collectively referred to as virulence-activated genes (vags) [31]. Understanding this system is paramount, as it significantly influences the pathogen's infectivity and disease-causing potential in the host.

2.2.2.3 *Pertussis Toxin*

The PT toxin is produced and released only by *B. pertussis*. The role of PT is adhering to ciliated respiratory cells [119]. Then, PT inhibits the chemotaxis of phagocytic cells to the site of inflammation [40]. PT binds to the glycoproteins on the cell surface of the host cell and then phagocytosis the whole holotoxin. The S1 subunit in the endoplasmic reticulum is released into the cytoplasm. Once in the cytoplasm, the S1 subunit is responsible for ADP-ribosylating the alpha subunit of heterotrimeric Gi proteins [24]. This process leads to increased cyclic adenosine monophosphate (cAMP) in the cytoplasm and dysregulation of the Gi protein-coupled signaling pathway. It is known that PT can specifically prevent the recruitment of macrophages and neutrophils to the site of infection [30]. Numerous studies have shown that neutralizing PT would result in vaccine-induced protection [18].

2.2.2.4 *Adenylate Cyclase Toxin*

Adenylate cyclase toxin (ACT) is a single-chain protein composed of five distinct regions: an N-terminal adenylate cyclase (AC) domain, a central hydrophobic segment, a modification region that facilitates the acylation of the protein,

a repeat-in-toxin (RTX) domain responsible for hemolysin binding, and a C-terminal secretion signal [40]. Type I secretion system exports ACT, which then passes through a channel formed by the CyaBDE protein complex [156]; the ACT is modified by acylation and then successfully secreted.

During secretion, ACT undergoes acylation, a modification that is essential for its activity. This acylation enhances its ability to bind to the complement receptor 3. In the host cell's cytoplasm, calmodulin increases the catalytic activity of ACT, leading to elevated cellular cAMP levels. These high cAMP concentrations can inhibit the activation of antigen-presenting cells (APCs) and promote apoptosis, contributing to the toxin's pathogenic effects [40].

2.2.2.5 *Other virulence factors*

Although *B. pertussis* possesses numerous genes associated with the type-III secretion system (T3SS), these genes do not cause cytotoxic effects in mammalian cells, unlike other *Bordetella* species, suggesting that *B. pertussis* may not utilize T3SS [99]. The conclusion was later rejected, showing that *B. pertussis* does use the T3SS, secreting effectors important for colonization and immune invasion. T3SS suppresses the innate immune response, the Th1 and Th17 response, and antibody responses. On the other hand, T3SS enhances the binding of the effectors to the macrophages, and the respiratory tract [46].

The lipopolysaccharides (LPS) produced by *B. pertussis* are toxic to host cells. These LPS lack an O-antigen in comparison to other *Bordetella* species. This lack of O-antigen helps *B. pertussis* to resist the killing by complement [23, 128].

2.3 THP-1 & BORDETELLA PERTUSSIS INTERPLAY

2.3.1 *Intracelullar Life of B. pertussis*

Sensing and responding to changes in the surroundings is crucial for organisms to adapt to different environments. *Bordetella* species have evolved mechanisms to adjust transcription to respond to such changes quickly. The *bvgAS* two-component system, which includes a sensor protein named BvgS, a transcriptional activator known as BvgA, and a transcriptional repressor called BvgR, primarily governs the activation of virulence in *Bordetella*. When growing in the presence of millimolar concentrations of nicotinate or magnesium sulfate, the BvgS sensor protein is inactive and unphosphorylated, and the bacteria are in the Bvg minus (Bvg-) phase, where the transcription of virulence genes is turned off and virulence repressed genes including genes that are required for flagella synthesis and motility are maximally expressed. When bacteria detect certain inducing signals, such as a rise in temperature to 37°C, similar to levels inside a mammalian host, the BvgS sensor protein undergoes autophosphorylation. This triggers a phosphorylation cascade, resulting in the transfer of a phosphate group to BvgA. Phosphorylated BvgA then binds to

the promoter regions of Bvg-activated genes, initiating the transcription of virulence-related genes in response to temperature changes. Consequently, this activates the production of virulence factors, including the type three secretion system (T3SS), PRN, FHA, ACT, PTX, and others [36, 66, 115].

BvgAS is not the only regulatory system involved in promoting bacterial persistence inside macrophages. Another two-component regulatory system was identified, named RisAS, which plays a crucial role in the persistence of Bordetella [75]. A study of a RisA mutant strain showed that the RisAS regulon is responsible for bolstering resistance to oxidative stress, facilitating acid phosphatase production, and promoting *in vivo* persistence. It was found that RisA is most effectively expressed when the bacteria are contained within eukaryotic phagocytes at a temperature of 37°C [75].

In vitro studies using a human alveolar epithelial cell line A549 have shown that a significant number of intracellular *B. pertussis* bacteria can avoid fusion with phagolysosomes. During this process, *B. pertussis* releases various proteins that facilitate its survival and persistence inside host cells. These proteins play roles in stress management, iron acquisition, and metabolic functions [85].

Another study on *B. bronchiseptica* showed that the avirulent Bvg- phase is essential for the bacterium's survival, persistence, and replication [150]. It is possible that, unlike other harmful bacteria that enhance their virulence under stress, *Bordetella* species rely on an age-old stress response to adapt to the intracellular environment they encounter. This indicates that the Bvg- phase plays a crucial role in regulating gene expression when these bacteria interact with host phagocytes. At the messenger RNA (mRNA) level, some studies have indicated that virulence is suppressed upon internalization in macrophages. However, when analyzing the *B. pertussis* proteome, it was observed that the levels of the virulence proteins PTX and ACT did not change in intracellular *B. pertussis* when compared to *in vitro* growth conditions under Bvg+ [85, 153].

While some research findings provide support for the existence of an intracellular phase in pertussis, the exact nature of its involvement in the pathogenesis of the disease and its implications for vaccine effectiveness are currently under intense scrutiny and debate within the scientific community.

2.3.2 *Host Immune Response to B. pertussis*

The immune response to *B. pertussis* encompasses both innate and adaptive elements. The innate response primarily functions to contain the bacteria through the activity of macrophages, neutrophils, and dendritic cells. Following the attachment of *B. pertussis* to ciliated cells within the trachea and bronchi of the infected host, resident alveolar macrophages and neutrophils phagocytose and eradicate bacteria, although their efficacy is diminished due to the impact of toxins. Incidences of pertussis surviving intracellularly within certain macrophages [48, 64] have been documented, indicating the potential for intra-

cellular survival, thereby extending the infectious period and facilitating disease transmission. Research has demonstrated that IFN-γ and IL-17 enhance macrophage killing of pertussis, with IFN-γ established as a critical cytokine, as evidenced by studies involving IFN-γ receptor knockout mouse models [9, 94]. IFN-γ prompts the production of Nitric Oxide (NO) by macrophages, whereas IL-17 does not. Additionally, TNF-α is identified as another cytokine that contributes to controlling *B. pertussis* infection. Notably, a TNF knockout mouse model exhibited elevated leukocyte levels in the lungs and a higher bacterial load [169].

Dendritic cells (DCs) have the primary role of presenting antigens to T-cells, thereby eliciting cytokine production and inducing T-cell differentiation into helper subsets. DCs situated in the respiratory tract are responsible for detecting pathogen-associated molecular patterns (PAMPS) and various virulence factors in pertussis. This detection triggers the activation of host T-cell responses, contributing to the adaptive immune response. The binding of LPS to TLR4 results in the secretion of IFN-γ and IL-12, both of which drive Th1 differentiation [67]. Meanwhile, ACT activates Nod-like receptor 3 (NLRP3), leading to the cleavage of pro IL-1b into its active form. This, in combination with IL-12, induces the generation of Th17 cells [38]. Notably, certain DC subtypes have been observed to play a pivotal role in the immune response to *B. pertussis*. Experimental depletion of CD103a+ DC cells resulted in reduced bacterial clearance and a diminished Th1 response [39].

Neutrophils typically migrate into the lungs approximately five days following infection, peaking around day 10 [101]. Recruitment of neutrophils is dependent on the presence of IL-6 and macrophage inflammatory protein 2 (MIP-2) [172]. Early neutrophil recruitment is hindered by pertussis toxin, while adenylate cyclase toxin suppresses their bactericidal functions by reducing phagocytosis, chemotaxis, and superoxide production [29]. Neutrophils exert intracellular killing mechanisms subsequent to bacterial uptake facilitated by opsonizing antibodies, crucial for infection control in immune mice.

Natural Killer cells (NK) traditionally associated with viral infections, have been observed to initiate IFN-γ release, aiding bacterial clearance through macrophage activation and Th1 differentiation and response [94]. Depletion of NK cells in pertussis has been shown to lead to bacterial dissemination, intensified Th2 response, and diminished Th1 response, underscoring their significance [108].

The complement system plays a critical role in the innate immune response, contributing to pathogen opsonization, the release of chemo attractants, and direct bactericidal activity. It serves as a crucial link between the innate and adaptive immune systems. *B. pertussis* exhibits resistance to the classical complement pathway through the Bordetella resistance to killing A (BrkA) protein, which impedes neutrophil phagocytosis by preventing complement component aggregation on the bacterial surface [11]. Moreover, pertussis inhibits the

recruitment of the C4b binding protein (C4BP) and the C1 esterase inhibitor, thereby influencing the function of complement components [12, 14].

2.3.2.1 *Adaptive Immune Response*

The adaptive immune system is essential in clearing *B. pertussis* infection and providing immunity against future infections. It starts with dendritic cells presenting antigens to naive T-cells and producing pro-inflammatory cytokines and chemokines. This leads to the generation of B and T-cells with antigen specificity for *B. pertussis*.

Helper T-cell responses are typically observed approximately three weeks after infection. It is noteworthy that Th2 cells are not significant responders to infections, as their induction following *B. pertussis* infection is minimal. Furthermore, studies have shown that IL-4 knockout mice exhibit the capability of bacterial clearance similar to that of wild type mice [102]. In various investigations, both Th1 and Th17 immunity have been identified as essential components of the adaptive response [3, 68]. The differentiation of Th1 cells is initiated by IL-12 signals stimulated by *B. pertussis* from innate immune cells. Notably, Th1 cells play a critical role in producing IFN-γ, which is vital for preventing lethal disseminated infection, as evidenced in receptor and cytokine knockout mice. The production of IFN-γ is essential for the NO responses by macrophages to effectively combat *B. pertussis* [95]. On the other hand, Th17 cells serve as producers of IL-6, TNF-α, and IL-17, which are pro-inflammatory cytokines that contribute to the eradication of extracellular pathogens. The differentiation of Th17 cells is driven by IL-1/23. These cells function to recruit neutrophils via chemokines and cytokines, while also activating innate immune cells, including macrophages and neutrophils, to mediate the elimination of *B. pertussis* [172]. It should be noted that Th17 responses have been associated with negative aspects, such as their link to conditions like asthma, cystic fibrosis, and other lung disorders.

3

TRANSCRIPTOMICS OF THP-1 & BORDETELLA PERTUSSIS

Over the last decades, our understanding of the building blocks of life DNA and RNA has advanced significantly, thanks to the combined efforts of physics, chemistry, biology, and computer science. This progress has generated an enormous amount of data, which data scientists analyze and interpret to uncover meaningful insights. The advent of next-generation sequencing (NGS) has further revolutionized the field by greatly increasing access to genetic information and producing genomic and transcriptomic data more efficiently in terms of cost and time. However, despite these advancements, managing and processing this vast wealth of data remains a major challenge. To effectively address research questions in this domain, it is essential to design appropriate protocols with optimal parameters tailored to each study, ensuring that sequencing efforts yield accurate and useful results.

The first Human Genome Project was launched in 1990 to identify the complete nucleotide base pair sequence that makes up the human DNA [1]. The project was completed, covering about 92% of the total human reference genome in 2003 with an accuracy of 99.99% [140] costing about $3 billion [138]. Unlike before, sequencing a complete human genome today would take only a few days and cost less than a thousand dollars with the emerging NGS technologies [173]. Due to the parallel, fast, and low-cost sequencing of whole genomes and transcriptomes, NGS technologies are now known as high-throughput sequencing technologies [54].

NGS is a vital tool for research across disciplines, from basic biology to clinical diagnostics [54]. It has revolutionized the field of genomics by enabling comprehensive genome sequencing, transcriptomics, epigenomics, meta-genomics, and other omics studies [86]. These approaches play an important role in elucidating the current physiological state and responses generated by external stimuli in a cell [117].

3.1 TRANSCRIPTOMICS: THEN AND NOW

A transcriptome is the complete set of all RNA molecules that are expressed in a cell, tissue, or organism at a specific time and under specific conditions. It reflects the genes that are actively being transcribed into RNA, providing a snapshot of gene activity and expression levels. The information gained from transcriptomics can provide a platform for researchers to understand better how individual genes or a group of genes are involved in biological processes. Traditionally, the cellular expression of transcripts was measured based on gene-specific DNA probes, detection by hybridization (Northern blot), or amplifi-

cation (polymerase chain reaction (PCR) or quantitative real-time PCR (qRT-PCR)). Nevertheless, these approaches could only analyze a small portion of predefined target genes. The era of traditional transcriptomics changed when microarrays were developed. Microarrays use glass slides or membranes with parallel immobilized DNA probes on them [139]. Fluorescence was used to label the complementary DNA (cDNA) before hybridizing it to the microchip; this assisted in inferring the abundance of a respective transcript depending on the intensity of the fluorescent signal. Although multiple genes could be studied simultaneously with microarrays, only known RNA molecules could be studied. To overcome the limitation of discovering novel transcripts, high-resolution tiling arrays were introduced [142]. Microarrays and tiling arrays helped in the revelation of the very first gene expression patterns of several bacteria such as Vibrio cholera [104], Chlamydia trachomatis [13] and C. pneumonia [114].

Sequencing technologies have revolutionized biology. The first NGS technologies introduced were Illumina/Solexa, Roche/454, or ABI/SOLiD. These technologies can sequence the entire genome in a couple of hours [105]. NGS technologies were used for genomics initially but later adapted to coding DNA sequence (CDS), now known as RNA sequencing (RNA-seq) [159]. RNA-seq has many advantages over micro-array; firstly, prior information on the genome or genomic features is unnecessary. This approach is appealing for non-model organisms that lack existing knowledge, as it is organism-independent and capable of identifying all cellular transcripts, both coding and non-coding. Secondly, RNA-seq has large dynamic ranges of expression levels, increasing the sensitivity for genes expressed at very low or very high levels. RNA-seq also shows very high levels of reproducibility for technical replicates. Lastly, RNA-seq requires only a small input amount; for single-cell RNA-seq, a few picograms (pg) suffice [117, 159].

The RNA-seq approach has a few drawbacks when compared to microarrays. One of the main disadvantages is the size of RNA-seq files, which are significantly larger than microarrays files. Furthermore, RNA-seq involves several processing steps, including library preparation, alignment to a reference genome, normalization, and addressing sequence-specific biases and mapping. These factors contribute to longer analysis times. Also RNA-seq tends to be more expensive than microarrays [50, 151]. Nonetheless, RNA-seq has become more prevalent due to decreasing sequencing costs and technological advancements.

3.2 RNA-SEQ

RNA sequencing is a widely utilized technique for comprehensive analysis of the transcriptome, encompassing the full array of RNA molecules within a cell or tissue at a given moment [132]. This includes messenger RNA (mRNA),

non-coding RNAs, and small RNAs. The process involves reverse transcription of RNA molecules into complementary DNA (cDNA), followed by high-throughput sequencing, commonly performed using platforms such as Illumina. RNA-seq provides valuable insights into gene expression quantification, alternative splicing events, RNA editing, and the profiling of non-coding RNAs. It is extensively employed in studies of gene expression regulation, notably in differential gene expression analysis (DGEA), facilitating the discovery of novel transcripts, untranslated regions (UTRs), processing sites, bacterial operon architectures, and splice isoforms in eukaryotic organisms. Due to its versatility and depth, RNA-seq has become the method of choice for transcriptomic investigations across a wide range of biological systems.

3.3 DUAL RNA-SEQ

Dual RNA-seq is an advanced extension of traditional RNA sequencing methodologies that enables the concurrent analysis of host and pathogen transcriptomes within a single mixed infection system. This technique involves the extraction and simultaneous sequencing of RNA molecules derived from both the host organism such as humans or plants and the infectious agent, including bacteria, viruses, or fungi. The application of dual RNA-seq facilitates a comprehensive examination of transcriptional responses during host-pathogen interactions, thereby illuminating mechanisms underpinning immune responses, microbial pathogenicity, and the temporal dynamics of infection [163]. Its utility is particularly pronounced in the study of infectious diseases, microbial ecology, and the molecular pathways influencing disease progression and host defense. In essence, while conventional RNA-seq provides insights into the transcriptome of a single organism, dual RNA-seq broadens this scope to encompass the simultaneous analysis of both host and pathogen transcriptomes, yielding a more integrated understanding of molecular interactions. Despite similarities between RNA-seq and dual RNA-seq in their capacity to analyze gene expression profiles, their applications diverge markedly, as illustrated in **Figure 4**.

Although dual RNA-seq presents promising opportunities, it also faces certain obstacles. Quantitatively and qualitatively, differences are evident: a mammalian cell contains roughly ~10-20 pg of total RNA, which exceeds the amount in a prokaryotic cell—about ~100 femtogram (fg) by more than a hundredfold. Practically, this implies that to detect gene expression changes in both organisms, the sample should have an appropriate ratio of bacterial to host cells. For dual RNA-seq to be practical, an estimated number of bacteria per infected host cell has been suggested, with the bacterial transcripts in a mixed host-pathogen RNA sample expected to constitute approximately 5% [163].

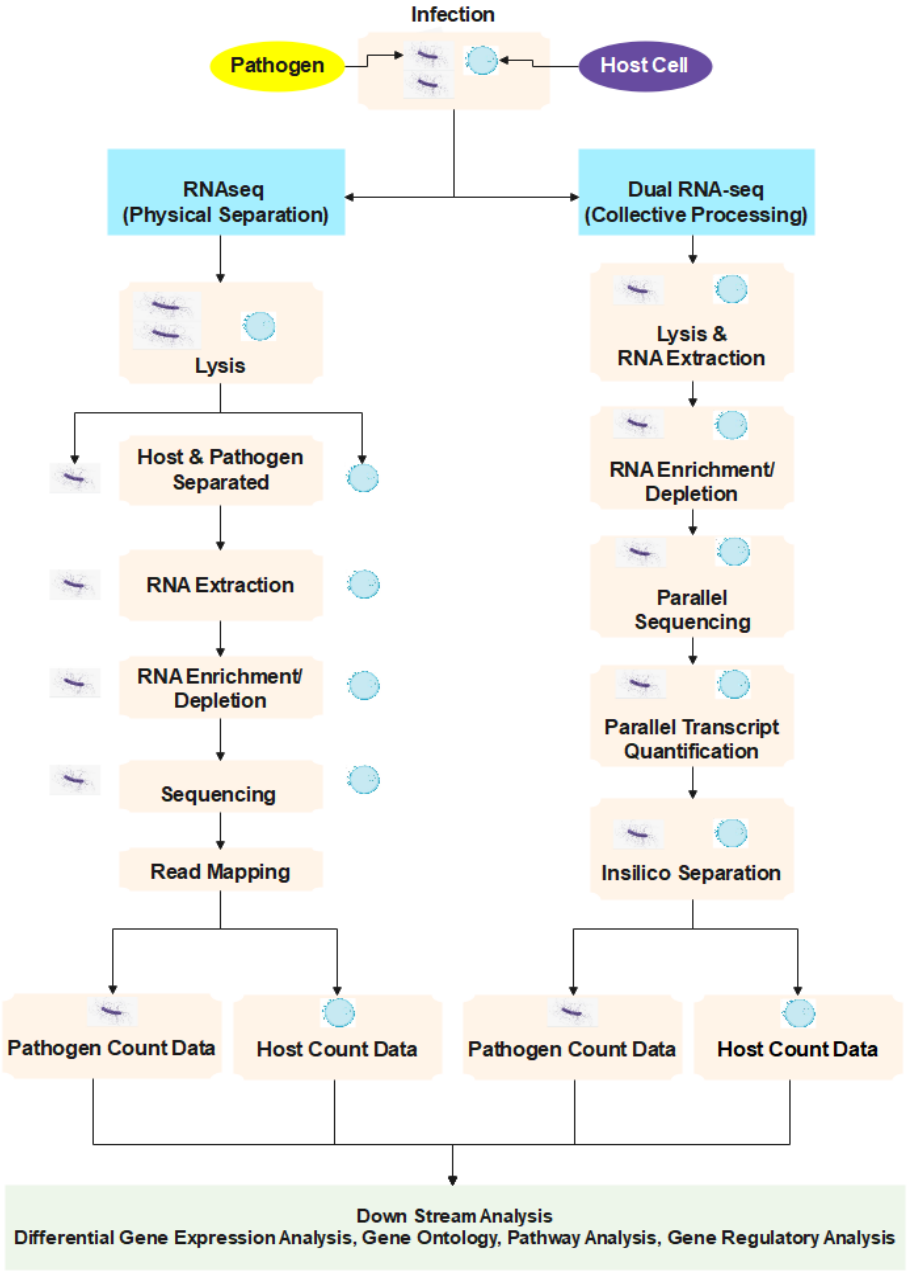


Figure 4: Comparison between RNA-seq and dual RNA-seq pipeline. The figure depicts two distinct methodologies employed in RNA-seq analysis. On the left, the traditional approach is illustrated, wherein total RNA is independently isolated from both host and pathogen cells within the infected sample. Subsequent processes include sequencing and read alignment, enabling the quantification of gene expression levels for each organism separately. On the right side, the dual RNA-seq technique is depicted, where total RNA is extracted from the infected cells and subjected to sequencing analysis before separating the mixed sequence reads for the original genomes.

3.4 DUAL RNA-SEQ EXPERIMENT DESIGN

The meticulous planning and coordination required for designing a dual RNA-seq experiment are essential to ensure the acquisition of high-quality data. This data, in turn, can yield valuable insights into host-pathogen interactions and the underlying molecular mechanisms of infection. Therefore, we decided to conduct a pilot study to address key questions pertinent to the dual RNA-seq design. These inquiries encompassed the assessment of the infection assay's efficacy, determination of the optimal number and timing of infection time points, identification of the suitable quantity of controls and replicates for each sample, and the establishment of the requisite sequencing depth for each infection time point. Before designing the dual RNA-seq setup, we carefully considered several factors, as listed below.

3.4.1 *Multiplicity of Infection*

In formulating the infection assay, our initial consideration was the multiplicity of infection (MOI). The pilot study employed an MOI of 100 bacteria per macrophage, which was subsequently reduced to 10 bacteria per macrophage in the comprehensive study. The adjustment of the MOI from 100 to 10 for the comprehensive study was necessitated by the utilization of a fresh batch of THP-1 cells which were more sensitive to infection. This alteration was implemented to maintain cell viability while replicating more realistic biological conditions, with the aim of ensuring that the experiment provides a more accurate reflection of natural *in vivo* conditions.

3.4.2 *Infection Time points*

The pilot study facilitated the selection of appropriate infection time points to ensure comprehensive coverage of gene expression changes during the infection process. Specifically, three time points were chosen: 2, 6, and 24 hours post-infection (hpi), with controls only at 6 hpi to validate temporal changes in gene expression profiles. The majority of infection dynamics observed between THP-1 cells and *B. pertussis* occurred during the earlier time points. To better capture these dynamics, the subsequent comprehensive investigation expanded the number of time points to five: 1, 2, 4, 8, and 12 hpi. Additionally, the decision to exclude the final time point of the pilot study (24 hpi) in the comprehensive investigation was based on the limited bacterial read abundance at that stage, which hindered the acquisition of meaningful insights. This expansion and adjustment enhanced the understanding of the infection dynamics and critical events. Notably, all infection time points in the comprehensive study were accompanied by controls.

3.4.3 *Sample Replicates*

It is crucial to thoughtfully determine the number of biological replicates in any study, as they greatly enhance the statistical power and reliability of the results while reducing the influence of variation on experimental outcomes. In both of our studies, each sample consisted of three biological replicates. This strategy ensured robust statistical assessments of gene expression variances, allowing us to differentiate genuine biological changes from random noise and technical artifacts.

3.4.4 *RNA Depletion*

RNA enrichment and depletion techniques are employed in RNA sequencing to selectively isolate or remove specific RNA molecules from a sample, allowing researchers to target specific RNA subsets or eliminate unwanted RNA to enhance the precision and accuracy of their analyses. Ribo-depletion methods are utilized to eliminate ribosomal RNA (rRNA) from the sample, as rRNA constitutes a substantial portion of total RNA and may not be pertinent to the conducted experiment. Following Ribo-depletion, other RNA species, such as mRNA or non-coding RNA, remain for further analysis. This approach serves to minimize the sequencing of undesired transcripts and optimize the sequencing of more relevant RNA classes. Depletion and enrichment can be performed at the RNA or cDNA levels. Commercially available kits, such as RiboMinus, Ribo-Zero, and MICROBExpress [146], utilize DNA oligonucleotides with a sequence complementary to rRNA to deplete rRNAs from RNA samples. Conversely, RNA enrichment for eukaryotic mRNA can be performed using kits that employ oligo(dT) to capture mRNAs. These techniques effectively reduce the required sequencing depth, thus lowering sequencing costs [33].

In our studies the enrichment step for bacterial RNA was skipped to avoid delaying of RNA isolation. Instead we focused on RNA depletion utilizing the Ribo-Zero Gold rRNA Removal Kit from Illumina to eradicate rRNA from total RNA samples.

3.4.5 *Library Preparation*

Different library preparation kits and protocols are offered by various manufacturers, each with distinct features and advantages tailored to the specific requirements of sequencing applications (e.g., DNA-seq, RNA-seq, ChIP-seq) and sequencing platforms (e.g., Illumina, Ion Torrent, PacBio). The NEBNext® Ultra™ II DNA Library Prep Kit was utilized for this investigation. This kit is designed to produce high-quality libraries suitable for sequencing on Illumina platforms. The sequencing for both the studies was performed on an Illumina HiSeq 2500 platform using HiSeqV4 chemistry.

3.4.6 *Transcript Length*

Both studies utilized single-end RNA-seq to balance cost and coverage. In the comprehensive study, the RNA-seq read length was increased from 50 base pairs (bp) in the pilot study to 100 bp, improving read specificity, mapping accuracy, and the resolution of transcriptional profiling. This adjustment in RNA-seq design allowed more precise detection of gene expression changes during *B. pertussis* infection, particularly for low-abundance transcripts, thereby enhancing the overall quality and interpretability of the RNA-seq data.

3.4.7 *Depth Requirements*

The initial RNA-seq research in mammals focused on the polyadenylated portion of cellular transcripts, with a sequencing depth of ~10-50 M reads [113, 146]. The first RNA-seq studies on prokaryotes yielded ~1-7 M reads after rRNA depletion [118, 126]. These depths were sufficient to deduce the differential mRNA expression, but more sequencing depth is required to provide information on low abundant RNAs, transcript boundaries, splice isoforms, and operon structures. The ideal depth suggested to deduce both quantitative and qualitative RNA-seq expression of bacteria and mammals is 1-10 M or ~100 M reads/library, respectively [57].

In our research, it was imperative to ensure the acquisition of a sufficient number of *B. pertussis* reads for each time point while also considering the sequencing cost. Our primary obstacle was to establish the sequencing depth required to obtain an ample number of reads for *B. pertussis*, given its reduction over the infection course in THP-1 cells. In the comprehensive study, we utilized the reads generated for *B. pertussis* at each time point in the pilot studies to design and calculate the sequencing depth and number of reads. We initiated with a low sequencing depth for the initial time points and incremented it for the later time points to ensure the acquisition of sufficient bacterial reads. For the initial time points, the sequencing depth was maintained at a lower level, and the anticipated reads to be retrieved for *B. pertussis* at 1 hpi and 2 hpi was 100k. For 4 hpi and 8 hpi, the sequencing depth was doubled, and the expected reads were estimated to be between 78k to 62k. For the final time point, 12 hpi, the sequencing depth was quadrupled, and the calculated reads to be retrieved amounted to 100k. Although the number of *B. pertussis* reads was lower than what the literature suggests, this was primarily due to the omission of enrichment steps, which resulted in fewer viable intracellular bacteria. However, the read counts at each time point were sufficient, even in the later stages of the study, and we successfully balanced the desired coverage with our budget constraints.

3.5 DUAL RNA-SEQ ANALYSIS PIPELINE

Analyzing dual RNA-seq data involves several key steps, and here's an overview of the downstream workflow used in our studies shown in **Figure 5**.

3.5.1 *Quality control*

High-throughput sequencers generate vast amounts of data in a single run; therefore, evaluating the library quality and sequencing performance is necessary before starting the analysis with the acquired sequences. Quality Control (QC) usually considers duplication rate, rRNA abundance, strand specificity, coverage continuity, the content of bases, the number of bases N, GC content, and adapter sequences.

For both of our studies, FastQC [4] was used to perform quality control checks on raw Illumina reads. Bedtools [129] was used to convert the raw bam files to fastq format. All the adapter sequences at the end of the reads were removed using the tool Cutadapt [96] before they can be further processed. The aim was to improve the quality and reliability of the raw sequencing data before downstream analysis as the read quality declines towards the 3′ end of the read.

3.5.2 *Transcript Quantification*

Transcript quantification is used to estimate the abundance or expression levels of individual transcripts (mRNAs or non-coding RNAs) in a biological sample, based on the number of reads that map to each transcript. This approach requires a gene annotation, e.g. in the form of a gene transfer format (GTF) file, which stores the genes and exons coordinates. This GTF file is used to map reads to transcripts. Tools commonly used to quantify transcript are HTseq-count [2], feature-count [88], and salmon [125].

Salmon employs the quasi-mapping algorithm to assign reads to transcripts by identifying a series of exact, unique matches between the read and the transcript that cannot be extended further in either direction. These matches are based on a distinctive sequence fragment that sets them apart from other matches, allowing reads that might otherwise align to multiple locations to be accurately assigned to a single transcript, thereby enabling precise quantification of transcripts [125]. Salmon quantifier was used for both our studies by utilizing a single transcriptome generated by combining the humans transcriptome from Ensembl GRCh38 and Bordetella pertussis Tohama I transcript from NCBI Assembly GCF000195715.1. The bordetella transcriptome was generated using the gffread tool, which converted the GFF (General Feature Format) file format to GTF (Gene Transfer Format).

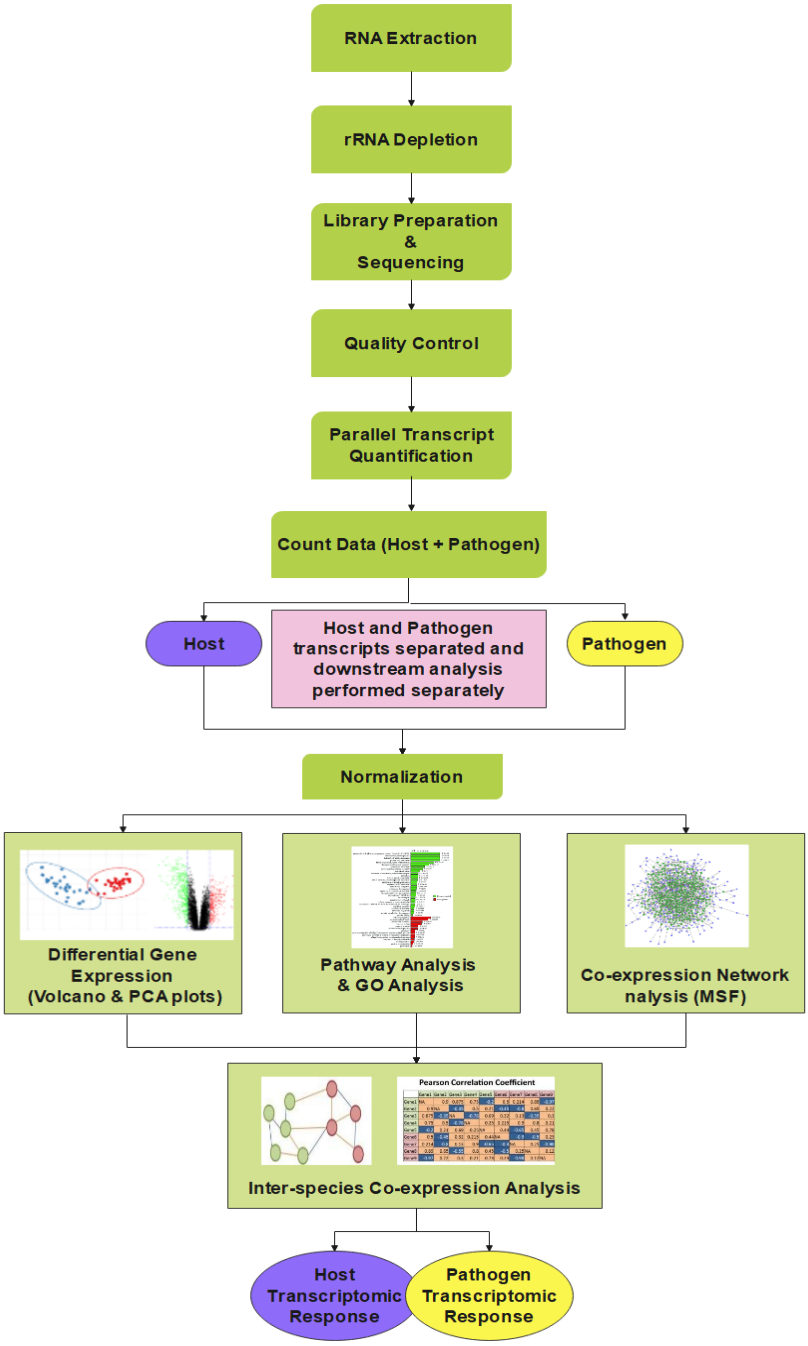


Figure 5: Pipeline for dual RNA-seq analysis. The figure illustrates the steps we used in our study for dual RNA-seq. Firstly, we lysed the infected cells and extracted the total RNA, followed by rRNA depletion. We then prepared libraries and performed sequencing on mixed RNA. The dual RNA-seq reads were filtered for quality before aligning to host and pathogen transcripts in parallel. Separated reads for host or pathogen were used for quantification and downstream bioinformatics analysis, such as differential expression, pathways, and network analysis.

After running the quantification process using Salmon, the human and *B. pertussis* reads were sorted to be used for further analysis of differential expression. All the genes with an Ensemble ID were found to belong to human macrophages, while the IDs that started with BP were assigned to *B. pertussis*.

3.5.3 *Principal component analysis*

Principal Component Analysis (PCA) is frequently applied in dual RNA-seq data analysis to explore patterns of gene expression variation across samples, detecting outliers, and visualizing relationships between samples [73].

In our studies, after obtaining the read counts per gene from salmon quantification, the regularized log transformation using the rlog function of DESeq2 [91] was applied for each time point to normalize the data. The PCA plots for visualizations were produced using the ggfortify package [149] and the ggplot2 package [164] in R [130]. The PCA plots from both studies offered valuable insights into how the samples clustered. The analysis showed that samples from uninfected macrophages were clearly separated from those taken from infected cells. Additionally, the samples from infected macrophages at different time points also formed distinct clusters, indicating differences in their gene expression profiles. The replicates of the same samples clustered closely together, demonstrating the consistency of the experimental results.

3.5.4 *Differential Gene Expression Analysis*

Differential Gene Expression Analysis highlights the genes that have changed significantly in abundance according to phenotype, disease, or experimental condition (e.g. disease and control). Effective use of gene expression analysis is time-series gene expression data [49], knowing that biological processes are often dynamic. Differentially expressed genes (DEG) are identified using a combination of expression level thresholds and expression score cut-offs, often determined by p-values obtained through statistical models. A gene's expression is quantified by the number of sequenced fragments that align with its transcript. The results of DGEA can be influenced by factors such as library preparation methods and quality, mapping settings, and transcript counting techniques [131].

The differential gene expression analysis, for both host and pathogen for both our studies was performed with edgeR [135] package in R using the read counts from salmon. Before conducting differential gene expression analysis, we removed any unwanted variation caused by batch effects or library preparation from the samples. This was achieved by using the RUVs correction method from the RUVseq [133] package in R. In edgeR, the normalization process employed involved the weighted trimmed mean of M-values (TMM) approach. This method adjusts for variations in sequencing depth across sam-

ples, thereby enhancing the accuracy of DGEA. TMM normalization is widely regarded for its robustness compared to alternative normalization techniques, especially when applied to datasets with large number of genes [158].

3.5.5 *Volcano Plots*

A volcano plot is a widely used graphical technique in the analysis of dual RNA-seq datasets, employed to concurrently represent the statistical significance and the fold change of gene expression changes between two experimental conditions or groups. It helps identify genes that are both significantly differentially expressed and exhibit large fold changes [100].

For our studies, volcano plots were generated between the control samples and treated samples for each time point and each species. The volcano plots were generated using the ggplot package in R. These plots helped our study in identifying and prioritizing significantly differentially expressed genes with high fold-change and very low p-value for further investigation.

3.5.6 *Gene Ontology*

After conducting DGEA, the first step is to perform Gene Ontology (GO) analysis. This helps uncover and identify key molecular functions, biological processes, and cellular components associated with differentially expressed genes under experimental conditions or treatments [7, 32]. GO annotations are associated with specific genes or gene products, and these annotations are based on experimental evidence, computational predictions, or manual curation from published literature. The annotations can be accessed through various databases and resources [7, 69], such as the Gene Ontology website [32], UniProt [93], and NCBI Gene [22].

For both our studies, pilot and comprehensive, the GO analysis for *B. pertussis* was performed using blast2go [28], we determined the gene ontology terms associated with more than one gene in the gene set for enrichment compared to the entire transcriptome using the Fisher Exact test. Revigo tool [147] was then used to condense, summarize, and visualize the GO term lists. The GO analysis helped observe that several important GO terms have been found to be dis-regulated and so we delved deeper into these GO terms. Heatmaps were generated for each of these GO terms using the ggplot2 package in the R.

For the host part the GO analysis was performed in R using the package GSEABase [134]. To perform this analysis, we first separated the complete list of significant differentially expressed genes based on their adjusted p-value, which was less than 0.05. These genes were then divided into two groups - one group with genes having a positive log fold change (log2FC > 0) and the other group with genes having a negative fold change (log2FC < 0).

3.5.7 *Pathway Analysis*

Pathway analysis is a methodological approach employed to elucidate the complex biological pathways and networks influenced by specific experimental conditions or genetic variations. In biological systems, genes, proteins, and metabolites frequently operate within interconnected pathways, collaborating to execute particular physiological functions or adapt responses to environmental stimuli [51, 116]. Pathway analysis aims to identify these sets of genes, proteins, or metabolites that are coordinately regulated and contribute to a particular biological process or disease. There are curated databases and tools that contain information about known biological pathways and their interactions for example, KEGG [76] Reactome [154] and the ingenuity pathway analysis (IPA) (QIAGEN Inc., https://www.qiagenbioinformatics.com).

In the pilot study, which was more focused on the host response upon infection, we performed GSEA to identify the enriched KEGG (Kyoto Encyclopedia of Genes and Genomes) terms in humans using the R package 'GSEABase'. To perform this analysis, we first separated the complete list of significant differentially expressed genes based on their adjusted p-value, which was less than 0.05. These genes were then divided into two groups - one group with genes having a positive log fold change (log2FC>0) and the other group with genes having a negative fold change (log2FC<0). The plots for GO and KEGG analysis were generated using the ggplot in R.

3.5.8 *Gene Regulatory Networks*

Gene regulatory networks (GRNs), are complex networks of interactions that govern the regulation of gene expression in a cell or an organism [59]. These networks play a fundamental role in controlling various cellular processes, including development, differentiation, response to environmental cues, and maintaining cellular homeostasis. GRN commonly shown as a Boolean network model. The model has nodes corresponding to genes and gene products (RNAs and proteins) with directed edges corresponding to the interaction type between the genes. The interaction of the nodes has different modes, including activation (upregulation), inhibition (downregulation), or regulation. Correct prediction of the GRNs would help elucidate the dynamics of a specific gene under a condition or disease. Hence, GRNs would assist in identifying dis-regulated genes, the dis-regulated signal transduction downstream of the gene, and develop new methods to test drugs for the genes [123]. A simple graphical representation of a gene regulatory network is illustrated in **Figure 6**.

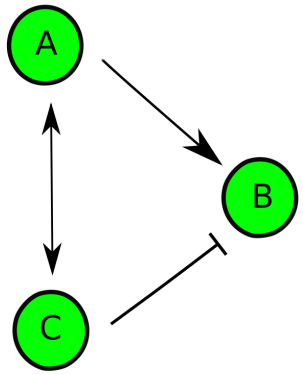


Figure 6: A representation of a simple GRN. With three nodes as genes and three edges as interactions between the genes. The graph shows gene A activating gene B, while gene A and gene C show an example of bi-regulation. On the contrary, gene C is inhibiting gene B.

We developed a novel tool called Modulated Sub-graph Finder (MSF) to discover networks of genes that function together in biological networks. Unlike other approaches that rely on pre-defined gene sets, MSF uses the network of interacting genes and individual gene p-values generated from differential gene expression analysis. The p-value indicates the probability that the expression of a specific gene has not changed in a statistical model. To detect significantly modulated sub-graphs, MSF combines the individual p-values of interacting genes into a single combined p-value of the network using the inverse of the standard normal distribution method. This method takes into account the dependent nature of the gene p-values. MSF can also identify possible trigger points or points of perturbation in these networks, which are the spots where the stimulus leads to downstream effects. The impact score of the trigger points is determined by the number of downstream genes within the corresponding network.

MSF was used to generate the gene regulatory networks for human genes in the pilot study. In pilot study, a significantly regulated gene network was identified by MSF. The identified network highlighted some critical role of cytokine and chemokine signaling in regulating immune cells and shaping the adaptive immune response. In our comprehensive study, we employed MSF to develop gene regulatory networks. However, due to the substantial size of our dataset and the incorporation of five time points of infection, the resultant gene networks became too extensive for effective visualization. Furthermore, interpreting these increasingly complex networks over the duration of the in-

fection was quite challenging. As a result, we opted not to include the MSF analysis in the final study.

Despite the potential benefits of analyzing the gene regulatory networks of *B. pertussis*, previous attempts to perform this analysis using MSF or any other tool have been unsuccessful. This is primarily due to the lack of an annotation file and gene interaction data available for *B. pertussis*.

3.5.9 *Inter-species Gene Coexpression Analysis*

Inter-species gene coexpression analysis is a computational approach commonly employed in bioinformatics to investigate and compare gene expression patterns across diverse species. This methodology facilitates a deeper understanding of the functional relationships among genes and elucidates the extent of their conservation or divergence throughout evolutionary processes [15, 145].

After completing all the necessary downstream analysis of dual RNA-seq, we planned to perform inter-species gene coexpression analysis between human and *B. pertussis* genes. To achieve this, we utilized the cor.test function in R, using both the Pearson and Spearman correlation methods. Additionally, we employed an online tool called LeMoNe [106] for correlation analysis. To ensure the accuracy of our findings, we further investigated the gene correlations that appeared in both the Pearson and Spearman correlation analysis in R, as well as the online tool analysis. Using the correlation analysis we identified a number of THP-1 gene showing strong correlation with *B. pertussis* genes.

4

MODULATED SUB-GRAPH FINDER (MSF)

4.1 OVERVIEW OF NETWORK MODELS

Biological processes consist of interconnected interactions among proteins, nucleic acids, metabolites, and various other ligands. Emerging high-throughput techniques can measure the gene expression of proteins, genes, regulatory RNAs, signaling agents, and other small ligands under different perturbations [77]. Studies in systems biology employ network models to determine protein interactions within a specific species [170]; identify conserved pathways across various species [145]; pinpoint protein groups implicated in diseases [71]; forecast protein functions [174]; explore the chemical mechanisms behind metabolic reactions [80]; and analyze topological features and other properties of biological networks [79, 82].

There are different classes of network models used to explore the available data, which vary in quantity, quality, and granularity:

Undirected graphs: These graphs feature nodes representing genes or proteins, with connections illustrating their interactions. Such gene-gene or protein-protein networks can aid in detecting potential functional groups, analyzing network structures, and pinpointing key nodes. Comparative studies of networks across different species can reveal conserved functional modules. Additionally, undirected graphs can depict various interactions among proteins, RNA, DNA, and other biological entities [143].

Undirected weighted graphs: In these graphs, the nodes represent genes, and the links represent the interaction. The interactions are given weight which models the similarity of expression patterns between genes. This type of graph is used in gene expression correlation networks to identify clusters of genes that show similar expression patterns. It can also be used for comparative analysis of networks from different tissues of the same species to distinguish gene coexpression patterns. Functional similarities between genes from different species could also be identified [145].

Directed graphs: The nodes represent genes, and directed links between nodes denote the regulatory interactions. Examples of good annotated directed graphs are pathway databases such as KEGG [78] and Reactome [35]. The direction in such graphs gives an insight into the regulatory events and regulatory mechanisms. Comparison of pathways can reveal common sub-graphs and putative evolutionary relations.

Undirected or directed multi-graphs: These graphs have each node, and each link is associated with a set of labels (e.g., nodes have GO term with it)

and weights. These weights can be further used in modular algorithms and system biology [145].

4.2 MSF

As part of our research objectives, we aimed to construct interspecies coexpression networks. To achieve this, multiple network models were systematically evaluated, including STRING App, SPIA, IntAct, and BioGRID. However, none of these tools fulfilled our specific requirements; consequently, we planned to develop an in-house tool to precisely perform the desired analyses. Ultimately, the **MSF** tool was implemented using a heuristic methodology in Java. This approach utilizes undirected, weighted graph constructs to identify co-modulated subgraphs within complex biological interaction networks. The degree of modulation within each subgraph is quantified by aggregating the individual *p*-values associated with nodes, employing a statistical technique for combining dependent *p*-values as described by Hartung [58] and elaborated below.

4.2.1 *Method of Combining Dependent Tests of Significance*

There has been a long tradition of combining several test statistics for clinical trials. These test statistics are often different, and the direct combination of the test statistics is impossible. In such cases, the p-values arising from test statistics are combined. This helps to test the overall effect of the outcome of several experiments in a meta-analysis. The traditional methods to combine p-values assume that the test statistics are independent of each other. Examples of such methods are by Birnbaum [16], Hedges and Olkin [62], Liptak (1958) [89], and Fisher [56].

Dependency allowed in the original test statistics, in the quasi means of the p-values characterized by Liptak (1958), is now termed the inverse normal method. The dependency, in this case, becomes equivalent to correlation. The modification of the inverse normal method suggested by Hartung with the incorporation of dependent test statistics is given below.

Let t_i be one-sided test statistics for testing null hypotheses $H_{i,0} : \vartheta_i = \vartheta_{i0}$ for $i = 1; \ldots ; n$; for the real-valued parameters ϑ_i, against the one-sided alternatives $H_{i,1} : \vartheta_i > \vartheta_{i0}$, where large values of t_i may lead to a rejection of $H_{i,0}$. The results from these test statistics are combined using a function of $T_1; ...; T_n$ to test the combined null hypothesis.

Considering T_i has a continuous distribution function F_{i0} under the null hypothesis. Then using the inverse normal cumulative distribution function Φ^{-1}, individual *p*-values are transformed to their corresponding normal score $t_i = \Phi^{-1}(F_{i0} - T_i)$ that is uniformly distributed on (0,1)

The dependency in the original test statistics $T_1; ...; T_n$ now becomes a dependency in the probits $t_1; ...; t_n$, which is equivalent to the correlation of the t_i's. Using the normal scores, the correlation between the test statistics is calculated $Cov(t_i, t_j) = \rho$ for $i \neq j$ and $i; j = 1; \ldots; n$.

Letting the weights be $\lambda_1; ...; \lambda_n$, with $\sum_{i=1}^{n} \lambda_i \neq 0$. Now applying the normal scores, correlations, and weights of the test statistics, the weighted inverse normal method to calculate the combined *p*-value $t(\rho)$ to test the combined null hypothesis is

$$t(\rho) = \frac{\sum_{i=1}^{n} \lambda_i t_i}{\sqrt{(1-\rho)\sum_{i=1}^{n} \lambda^2 + \rho(\sum_{i=1}^{n} \lambda_i)^2}}$$

4.2.2 *Implementation*

4.2.2.1 *Java*

In the era of structured programming, object-oriented languages like Java and C++ have become standard. Java, a class-based language built on C and C++, is compiled into machine language using a 'Compiler'. It runs in a virtual machine, with the core written in C and C++. [34]. As a result of being run on a virtual machine, Java can only access native platform functionality through the use of the Java Native Interface. Java Platform, Standard Edition (Java SE) refers to the basic version of Java, intended for developing and executing applications for personal use. Java Platform, Enterprise Edition (Java EE) provides an API and runtime environment tailored for enterprise application development and deployment.

In technical terms, the Java SE platform is also built upon two distinct levels:

1. Java Runtime Environment (JRE) consists of a Java Virtual Machine along with standard class libraries.

2. Java Development Kit (JDK) is a set of programming tools used for Java software development. It contains the Java compiler (javac), standard class libraries, examples, documentation, and the Java Runtime Environment (JRE).

4.2.2.2 *Prerequisites*

- Java Development Kit (JDK)

- Eclipse IDE

4.2.2.3 *Libraries*

The Java libraries utilized in the development of MSF were:

Commons CLI: The Apache Commons CLI [5]. The library provides an API designed to interpret command line options supplied to programs. Additionally, it can generate help messages that list the available options for a command-line interface.

4.2.2.4 *Special Classes*

A special package used in the development of MSF was java.math [120]. This package in Java provides classes for performing arbitrary-precision arithmetic operations. It is beneficial when dealing with huge numbers that might exceed the limits of primitive data types like 'int' or 'long', for example combining the *p-values* in our case, which generated exact and big decimal values.

Class BigIntegerMath: This class provides utility methods and mathematical operations on BigInteger values.

Class BigInteger: The BigInteger provides mathematical operations on integer numbers. This class handles very precisely very large and very small integer numbers.

Class BigDecimal: The BigDecimal provides mathematical operations on double numbers. This class handles very precisely very large and very small floating numbers.

Class BigSurd: The BigSurd class represents numbers of the type $r\sqrt{d}$, where r is a signed rational number and d is a non-negative rational number belonging to the Rational class.

Class BigSurdVec: The sums and differences of two BigSurd are represented as a BigSurdVec.

Class BigComplex: This class represents a complex number that uses BigDecimal for storing values with arbitrary precision. It not only supports accurate numeric calculations but also aids in DataType conversions with minimal data loss. Note that some methods may round values to a precision of 50 decimal places.

4.3 FUNCTIONALITY OF MSF

The functionality and application of **MSF** have been comprehensively examined in the initial publication.

5

RESULTS AND PUBLISHED WORK

Publication 1

"MSF: Modulated Sub-graph Finder"

Mariam R. Farman, Ivo L. Hofacker and Fabian Amman

Status: published in F1000, 2019
https://doi.org/10.12688/f1000research.16005.3

Author Contributions:

MARIAM R. FARMAN - Designed the methodological framework, conducted data collection, processing, and analysis, developed the software, and wrote the manuscript.

IVO HOFACKER - Supervised the project and contributed to the manuscript.

FABIAN AMMAN - Supervised the project and contributed to the manuscript.

SOFTWARE TOOL ARTICLE

REVISED MSF: Modulated Sub-graph Finder [version 3; peer review: 3 approved]

Mariam R. Farman [1], Ivo L. Hofacker[1], Fabian Amman [1,2]

[1]Institute for Theoretical Chemistry,Theoretical Biochemistry Group,, University of Vienna, Vienna, 1090, Austria
[2]Department of Chromosome Biology, Max F. Perutz Laboratories,, University of Vienna, Vienna, 1030, Austria

v3 **First published:** 29 Aug 2018, **7**:1346
https://doi.org/10.12688/f1000research.16005.1
Second version: 22 Mar 2019, **7**:1346
https://doi.org/10.12688/f1000research.16005.2
Latest published: 14 Apr 2019, **7**:1346
https://doi.org/10.12688/f1000research.16005.3

Abstract

High throughput techniques such as RNA-seq or microarray analysis have proven to be invaluable for the characterizing of global transcriptional gene activity changes due to external stimuli or diseases. Differential gene expression analysis (DGEA) is the first step in the course of data interpretation, typically producing lists of dozens to thousands of differentially expressed genes. To further guide the interpretation of these lists, different pathway analysis approaches have been developed. These tools typically rely on the classification of genes into sets of genes, such as pathways, based on the interactions between the genes and their function in a common biological process. Regardless of technical differences, these methods do not properly account for cross talk between different pathways and most of the methods rely on binary separation into differentially expressed gene and unaffected genes based on an arbitrarily set *p*-value cut-off.

To overcome this limitation, we developed a novel approach to identify concertedly modulated sub-graphs in the global cell signaling network, based on the DGEA results of all genes tested. To this end, expression patterns of genes are integrated according to the topology of their interactions and allow potentially to read the flow of information and identify the effectors. The described software, named Modulated Sub-graph Finder (MSF) is freely available at https://github.com/Modulated-Subgraph-Finder/MSF.

Keywords

Differential gene expression analysis, pathway analysis, combining p-value, cell signalling network

Open Peer Review

Reviewer Status ✓ ✓ ✓

	Invited Reviewers		
	1	**2**	**3**
version 3 (revision) 14 Apr 2019	✓ report	✓ report	✓ report
version 2 (revision) 22 Mar 2019	? report		? report
version 1 29 Aug 2018	✗ report	✓ report	? report

1 **Guanming Wu**, Oregon Health and Science University (OHSU), Portland, USA

2 **Stefanie Widder**, Research Lab of Infection Biology, Waehringer Guertel 18-20, 1090 Vienna, Austria

3 **Haibo Liu**, Iowa State University, Ames, USA

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Ivo L. Hofacker (ivo@tbi.univie.ac.at)

Author roles: Farman MR: Data Curation, Investigation, Methodology, Software, Validation, Visualization, Writing – Original Draft Preparation; **Hofacker IL**: Conceptualization, Funding Acquisition, Project Administration, Resources, Supervision, Writing – Review & Editing; **Amman F**: Conceptualization, Project Administration, Supervision, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This work was funded by the FWF ("Fonds zur Förderung der wissenschaftlichen Forschung") within the project Internationalen Kooperationsprojektes - Intl cooperation Project (Joint Project - Lead Agency Verfahren) with the project number (I 2353-B22). The grant was assigned to ILH. FA was funded by the Austrian Science Fund (FWF) project SFB F43.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

How to cite this article: Farman MR, Hofacker IL and Amman F. **MSF: Modulated Sub-graph Finder [version 3; peer review: 3 approved]** F1000Research 2019, **7**:1346 https://doi.org/10.12688/f1000research.16005.3

First published: 29 Aug 2018, **7**:1346 https://doi.org/10.12688/f1000research.16005.1

REVISED **Amendments from Version 2**

Version 3 of our manuscript addresses the points from the peer-reviewers, whom we thank for their constructive feedback. We have updated the manuscript and published a new version on MSF (GitHub and Zenodo).

1. Human genes names are in upper-case for all output files of MSF.
2. For source reliability the *t*-test is performed on log transformed p-values.

Furthermore, we corrected typos, improved the documentation, and rephrased some sentences. The project number in the funding section has been corrected from 'I 1988-B22' to 'I 2353-B22'. Figure 2 nodes label size have been increased.

See referee reports

Introduction

High throughput sequencing techniques have been widely used to yield differentially expressed genes (DEG)[1]. The changes in transcript abundance are measured, e.g. by next generation sequencing techniques and interpreted as an indicator of differential expression of genes. DEGs can be used to gain insights into the mechanisms underlying differences between conditions of samples, such as healthy versus infected. Differential gene expression analysis (DGEA) informs about the magnitude of expression changes, which are often expressed as log-fold change. The sign of log-fold change and the confidence level of observing an authentic change, often expressed as *p*-value. The information from DEGs is further interpreted to extract meaningful biological insights. For example, genes that could be involved in the response to a particular stimulus. To this end, pathway-based analysis has become an important tool to further interpret the results of a DGEA and to acquire understandings of the perturbations in a biological system. These pathway-based methods use predefined pathways or networks which are sets of genes with their interactions forming a functional unit. DEGs help to identify pathways or networks that may be altered during an infection, providing important information about diseases and its treatment process[2]. The expression measurements of the genes obtained from DGEA in combination with statistical methods and the predefined pathways are used to identify specifically modulated pathways and processes[3].

Well established resources for pathway annotations are KEGG (Kyoto Encyclopedia of Genes and Genomes)[4] and Reactome[5]. KEGG pathways is a branch of KEGG database that hosts a collection of manually drawn pathway maps representing the molecular interaction, reaction and relation networks of cellular functions. Similarly, Reactome is an open-source, manually curated, peer-reviewed database for signaling and metabolic molecules with their interactions formed into different biological pathways[5]. Both provide predefined pathways which are sets of genes and their interactions categorized into functional units.

Existing pathway-based analysis approaches use different research designs, which can be categorized into ORA (Over-representation analysis), FCS (Functional class scoring) and pathway topology based methods. All of which aim to find a subset of genes, e.g., significantly differentially expressed genes, genes associated with a certain pathway more often than expected given the total set of examined genes, e.g. the whole genome background. ORA is the first and the most basic method of pathway analysis[3]. It uses a DEG list with user defined cut-off for the log-fold change and *p*-value (most commonly using absolute log-fold change ≥ 2 and *p*-value ≤ 0.05). Subsequently, sets of genes associated with annotated pathways are tested for being over-represented in the set of DEGs. To this end, hyper-geometric distribution, chi-square tests, binomial probability or the Fisher's exact test are used, whereas, the information of the topology of genes in the pathways is neglected[6]. Furthermore, ORA assumes that the biological pathways are independent of each other and ignores the fact that they cross-talk and overlap[2,3].

Unlike ORA, FCS has no artificial cut-off to define a DEG list. FCS works in three steps. First it calculates the gene-level statistics including correlation of molecular measurements using differential expression of individual genes, ANOVA, t-test and Z-test. In the second step the statistics of individual genes in a pathway are transformed to an individual pathway-level statistic commonly using Kolmogorov-Smirnov statistic, mean or median. Finally the statistical significance of the pathway-level statistics is assessed. Although FCS overcomes some of the limitations from ORA, it still ignores the topology of genes in a pathway, cross-talk and overlap of the pathways[2,3]. Pathway topology based methods are similar to FCS except that they consider the topology of each gene during the gene-level statistics but still lacks to link the different functional pathways[2].

From another perspective, network based approaches do not categorize sets of genes into functional pathways, but they consider all interactions to be equal. Thereby, they avoid distinguishing arbitrarily between interactions within a pathway and interactions between pathways (i.e., cross-talk). With this they aim to identify subnetworks that show modulation between two conditions or upon a stimuli[7]. To find these active modules heuristics solutions like simulated annealing (SA), greedy methods, genetic algorithms (GA), network propagation and co-clustering methods are used[7]. jActiveModules has been the first of this kind using simulated annealing to find modulated sub-networks[8]. The benefit of omitting pathways is brought by reduced interpretability of the results due to the lack of functional labels on the networks.

On these grounds we propose a novel approach to make use of the rich gene and protein interaction annotation resources available and combining it to functional pathway annotations to gain additional insights from basic DGEA. To this, we start with the presupposition that expression of neighboring genes within a functional pathway are not independent from each other. Rather, they are often regulating each other's expression or are part of the same regulon[9]. We understand that the categorization of links between genes into labeled pathways is often an arbitrary one, given the extensive cross talk between different pathways. Although these categories have proven to be useful in many situations, they force a certain perspective onto the interpretation of novel data. Based on these two principles, we aim to find sub-graphs of connected genes within cell signaling network, which exhibit as

a whole significant differential expression changes. This approach differs in two main aspects from common pathway analysis tools. First, it does not aim to identify functional pathways enriched in differentially expressed genes, but detects sub-graphs or branches in a network graph (potentially spanning more than one functionally grouped pathway) which is coherently modulated. Second, it aims to improve the DGEA on the gene level, by collecting the information of neighboring genes, which as a whole might exhibit prominent enough signal to be called significantly modulated. All of this can be helpful to understand the cause and effect of a stimulus and might inform about potential points of intervention.

As input, information on functional links between genes provided by e.g. KEGG or Reactome and information on the differential expression status of single genes resulting from a DGEA, are required. As a result the analysis returns sub-graphs and their joint confidence scores, reflecting how the perturbation migrates through the network. Furthermore, the entry points of perturbation in the networks and overlap with conventional pathway categories are returned. To facilitate prioritization of the perturbation entry points, to each an impact score and a measure of its reliability are assigned. The impact score expresses the fraction of the sub-module being downstream of the entry point. The reliability is measured using a t-test on log transformed *p*-values of the immediate upstream and downstream genes. The output is prepared in a directed adjacency matrix file, convenient for visualization, e.g., with StringApp[10], available as a Cytoscape plug-in[11].

The proposed algorithm is named `Modulated Sub-graph Finder` (abbreviated `MSF`). `MSF` can help transform the information obtained from DGEA into comprehensible knowledge of signal transduction of genes, hence being a valuable complement to existing pathway based methods. `MSF` is freely accessible on GitHub under the terms of the Creative Commons Attribution 4.0 International License.

Methods

`MSF` was implemented as a Java program. It is developed as a novel heuristic approach to find concertedly modulated sub-graphs in networks of biological interactions. `MSF` does not use predefined gene sets grouped into functional units, but rather relies purely on the network of interacting genes. The input network consists of nodes corresponding to genes and edges representing interactions. Furthermore it utilizes comprehensive results from a differential gene expression analysis to discover the sub-graphs, or modules, which are as a whole modulated.

`MSF` uses the individual gene's *p*-values generated from the DGEA. The *p*-value expresses the probability that the null hypothesis of unmodified gene expression can't be rejected for a given statistical model. To find significantly modulated sub-graphs individual *p*-values of the vicinal genes in the global network are combined into a single combined *p*-value, using a statistical method for combining dependent *p*-values described by Hartung[12]. Hartung's method uses the inverse of standard normal distribution function. Using the inverse normal cumulative distribution function Φ^{-1}, individual gene *p*-values T_i are transformed to their corresponding normal score $t_i = \Phi^{-1}(1 - T_i)$ that is uniformly distributed on (0,1). Then using these normal scores, the correlation between genes is calculated $Cov(t_i, t_j) = \rho$. The normal scores and correlation are applied to the weighted inverse normal function to calculate the combined *p*-value $t(\rho)$ for all genes examined, namely the examined sub-graph

$$t(\rho) = \frac{\sum_{i=1}^{n} \lambda_i t_i}{\sqrt{(1-\rho)\sum_{i=1}^{n} \lambda_i^2 + \rho(\sum_{i=1}^{n} \lambda_i)^2}}$$

Lambda λ be the weights for each gene, currently all genes have equal weight, i.e. 1. The combined *p*-value $t(\rho)$ of a sub-graph will express the significance of all genes in the sub-graph being modulated together. The information from the different genes are used as, although not independent, replicated measurement of the behavior of the whole sub-graph. This potentially increases the power to detect also significant sub-graphs consisting of genes which are not significant on there own.

Overview of our method

To reduce the complexity to score all possible connected sub-graphs `MSF` applies a four step heuristic as described in the following. The proceeding identification of modulated sub-graphs from a network by `MSF` is presented as a flowchart diagram (Figure 1).

Initializing modulated sub-graphs. `MSF` constructs the first sub-graph starting with the genes associated with the lowest (most significant) *p*-value deduced from the DGEA. From this seed it tries to extend the sub-graph by adding directly neighboring genes, starting with the next most significant one. A single combined *p*-value is calculated for the extended sub-graph. If the combined *p*-value is smaller than the *p*-value of the original sub-graph, the extended sub-graph is accepted. This step is iteratively repeated until no further extension is accepted. In this case the process starts over with all remaining genes not yet in the significantly modulated sub-graph. This step identifies all simple sub-graphs that are modulated in the whole network.

Extending modulated sub-graphs. In the next step, we check if any of the initial modulated sub-graphs could further be extended beyond the immediate neighboring genes. Instead of testing single genes and their compatibility to be added, groups of genes are considered. If the combined *p*-value of the initial modulated sub-graph and the extension genes is smaller than the *p*-value of the initial sub-graph the extension is accepted. All possible extension paths up to *3* (default 2) genes at all nodes in the sub-graph are tested. Again, this step is iteratively repeated until no further genes are added to the significant differentially expressed sub-graphs. This step bridges small gaps of genes without a clear differential signal in the DGEA.

Merging modulated sub-graphs. After detection and extension of the modulated sub-graphs, each pair of so far identified sub-graphs is tested if its combination scores better than each on its own. The merging of the sub-graphs is done by depth first search traversal from one sub-graph to the other sub-graph. If the two

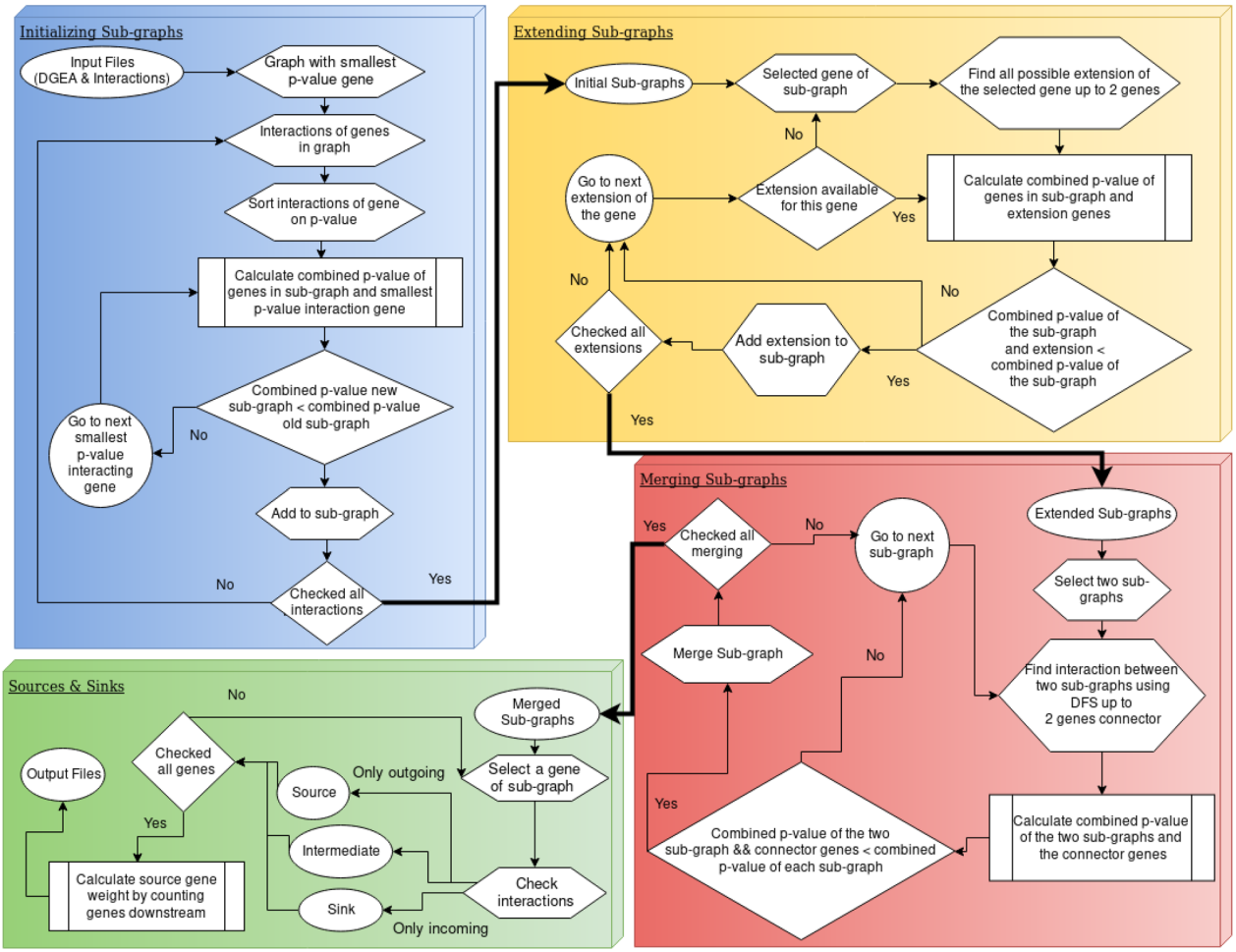


Figure 1. Graphical representation of the MSF heuristic approach to detect modulated sub-graphs in a global gene regulatory network.

sub-graphs merge with the connector of at most *3* genes (default 2 genes) and the combined *p*-value of the merged sub-graph including the bridging genes in between is less than the individual *p*-values of the two sub-graphs, the two sub-graphs are merged together to one bigger modulated sub-graph. This step is repeated iteratively until no sub-graphs can be merged anymore.

Finding sources & sinks. In a last post processing step MSF identifies the trigger points of the modulated sub-graphs. These trigger genes are the sources of the sub-graphs with only outgoing edges. These genes can be interpreted as the possible entry points of perturbation from where the stimulus causes downstream effects. Each individual source is given an impact score, expressing the relative number of downstream genes within the corresponding sub-graph directly connected by directed links. This score can be interpreted as an upper limit of how much of the sub-graphs' perturbation could have been introduced by the respective source, and thereby could be helpful to prioritize different identified sources for larger sub-graphs. In the same spirit as sources, the most downstream genes of the modulated sub-graphs are identified and defined as sinks. Due to loops not all sub-graphs are guaranteed to have sources or sinks. The reliability of each source is inspected using a t-test on log transformed *p*-values. The significant difference between the two groups, genes downstream the source and the genes upstream of the source is determined. This would help to assess if the source identified is reliable and indeed marks the border between two different regulation regimes.

MSF output. MSF generates a directed network file as an output, containing complete directed interactions of all modulated sub-graphs identified. This file could be imported into Cytoscape[11] for visualization. Additionally, a file containing details on all sources and sinks for all modulated sub-graphs is reported. Furthermore, for facilitated visualization in Cytoscape, a node attribute file is provided, containing the source weightage and the log-fold changes of all considered genes.

Operation. The only system requirements to run MSF are Java version 8 and JDK 1.8 or above. The few package dependencies are already been added to the release. The runtime of MSF is less than 10 minutes. To run MSF, the user must provide two text files, one containing the DGEA and the second one containing directed interactions in an adjacency matrix format file. Example files and a detailed tutorial to use MSF has been provided on GitHub https://github.com/Modulated-Subgraph-Finder/MSF.

Results

Case study

To demonstrate its usefulness, MSF is applied to RNA-seq data set of primary human monocyte derived macrophages (MDMs) infected with Ebola virus (GSE84188)[13]. Ebola Virus (EBOV) belongs to the Filoviridea family: filamentous, enveloped and single stranded RNA viruses. EBOV causes hemorrhagic fever in humans, inducing the host innate and adaptive immune response to be unable to control virus infection[14]. Currently, there are no approved antiviral drugs for the treatment of Ebola virus infection[15,16]. The initial targets of EBOV are the macrophages and dendritic immune cells[16,17]. EBOV inhibits the critical innate immune response of the host, which includes inhibiting the activation of alpha/beta interferon (IFN-α/β)[14,15,18]. It has been proposed that IFN-α/β should be tested against Ebola for its antiviral activity through clinical trials[15]. Ebola infection data was selected to test the approach because it has been well recognized for the last several decades, and vast literature is available for the pathogenesis of Ebola, hence facilitating the verification of the results of MSF with the vast literature present on Ebola infection. Especially, the detection of IFN-α/β as point of action for the virus could be considered as a basic indicator of the correctness and usefulness of the approach.

EBOV infection sequenced reads count data was downloaded from GEO (GSE84188), it describes the course of infection at three time-points 6, 24 and 48 hour post infection (hpi). Differential gene expression analysis was performed on the count data with edgeR package (version 3.4.2)[19] using an upper-quartile normalization. The DEG analysis results generated by edgeR were used as input for MSF. Cell signaling interactions were filtered from Reactome Functional interactions (FIs) Version 2016[20] for only direct interactions, which was used as a second input for MSF.

For the earliest time point at 6 hpi, three large modulated sub-graphs were identified with 42, 139, and 69 genes. The modulated sub-graphs consist predominantly of cytokines and chemokines (CXCL10, CCL8, CXCL9, CXCL11, CXCR4, CCR7, CCL4L1, CCL3L1, CCL4, CCL8, CCL20, CCL3, CCL19, IL6, IL27, IL23). IFNB1 and IFNA1 were both identified as two of the possible sources in the most significantly modulated sub-graph identified with 42 genes. IFNB1 has the highest impact score of 14.5. IFNA1 has an impact score of 8.7, in top 5 highest impact scores in the sub-graph it belongs to. Most of the sources identified by MSF were type I interferon induced genes (Supplement material). At 24 hpi seven modulated sub-graphs were identified with four main sub-graphs consisting of 61, 222, 130 and 242 genes, others being smaller than 6 genes. Again, IFNB1 and IFNA1 were identified as two sources out of the total sources with 3.9 and 1.6 impact score. IFNB1 was one of the top 5 impact score sources for the corresponding sub-graph. For the last time-point 48 hpi, six modulated sub-graphs were identified. Three of the sub-graphs were less than ten genes and main sub-graphs had 217, 224 and 276 genes. IFNB1 and IFNA1 were identified as sources in the most significantly modulated sub-graph with an impact score of 2.8 and 3.7, but not among the highest ranked sources (Supplement material).

As stated earlier IFN-α/β was reported to be one of the target genes of Ebola infection. We were able to successfully identify IFNA1 and IFNB1 as sources in all three Ebola infection time-points. Although IFNA1 and IFNB1 were already among the most significant genes in the DGEA during the later time points, MSF was also able to detect them as a source in the very early time-point when the genes were not significant based on the individual DGEA alone. Identifying the possible sources will reduce the search space for potential target genes and can help the biologist as the starting point of clinical testing for drugs and vaccines against an infection.

Table 1 compares the results of MSF, namely the number of detected sub-modules and their total genes numbers, to a simple analysis of mapping significantly modulated genes from the DGEA to the network and joining neighbors to modules. The numbers indicate that MSF detects less but larger and easier interpretable submodules, applying its statistical test. Furthermore, the

Table 1. Comparison of connected sub-graphs of modulated genes in the global network identified after the analysis with MSF and applying different *p*-value cut-offs from edgeR to genes in MSF identified modulated sub-graphs.

	Total number of genes in networks	Number of connected sub-graphs in network
6hpi		
edgeR + MSF	250	3
p-value ≤ 0.1	166	87
p-value ≤ 0.05	152	89
p-value ≤ 0.01	125	76
24hpi		
edgeR + MSF	656	7
p-value ≤ 0.1	457	183
p-value ≤ 0.05	418	198
p-value ≤ 0.01	332	216
48hpi		
edgeR + MSF	744	6
p-value ≤ 0.1	514	189
p-value ≤ 0.05	468	206
p-value ≤ 0.01	363	241

dependency of the results from the *p*-value cutoff choice is demonstrated for the DGEA, which is avoided for `MSF` altogether. It showcases how applying different cut-offs to the *p*-value of genes from edgeR to the sub-graphs identified by `MSF` breaks the larger sub-graphs to many smaller unconnected sub-graphs, many of which are single genes.

Modulated sub-graphs at 6 hpi

Three main modulated sub-graphs identified by `MSF` at 6 hpi are shown in Figure 2. The graphs represent the immediate output of the `MSF` analysis, visualized by StringApp[10] in Cytoscape[11]. Each node represents a gene part of a modulated sub-graph, whereby the associated colors code the functional annotation deduced from KEGG Pathways. The cross-talk between the pathways and also the multiple employment of many genes is evident. The flow of information between the sensors and effectors can be perceived given the directionality of each interaction, indicated by arrows. In more detail, sub-graph 1 (bottom) shows how the activation of Toll-like receptor, Cytokine, Chemokine activating Jak-STAT and MAPK genes, together with TNF leads into apoptosis. The next significant sub-graph (sub-graph 2: top right) reveals how information from the Extracellular matrix (ECM) receptor, which are reported to interact with Ebola glycoprotein (GP)[21], Chemokines, Cytokines, and Cytosolic DNA sensing are directly or indirectly controlling cell growth, differentiation, proliferation and apoptosis. It suggests that dysregulation of these pathways is responsible for modulation of apoptosis. Eventually, sub-graph 3 (top left) demonstrates how IFNA1 and IFNB1 modulates once more, via only a few intermediate steps, the apoptotic response of the cell. On the other side cAMP signaling genes activates platelet genes.

This display case might advertise with how little effort complex data can be processed and prepared for interpretation by the domain expert, to apprehend the dynamics of the underlying processes and suggest testable hypothesis and potential points of intervention.

Robustness

A potential concern is how noise in the gene expression measurements affects our analysis. To assess the robustness and stability of our method, we therefore added Poisson distributed noise to the read counts of the three time-points data set, used above. Then DGEA was carried out on the disturbed data with the same parameters as for the native data using edgeR, followed by analysis with `MSF`. This procedure was carried out 100 times. Every time the genes from the modulated sub-graphs identified from noisy data were compared to the genes of sub-graphs identified from the native data. The robustness of `MSF` analysis for the time-point 6, 24, and 48 hpi is shown in Figure 3. The procedure how data noise was modeled can be considered as rather stringent as `MSF` is sensitive to *p*-value changes across the

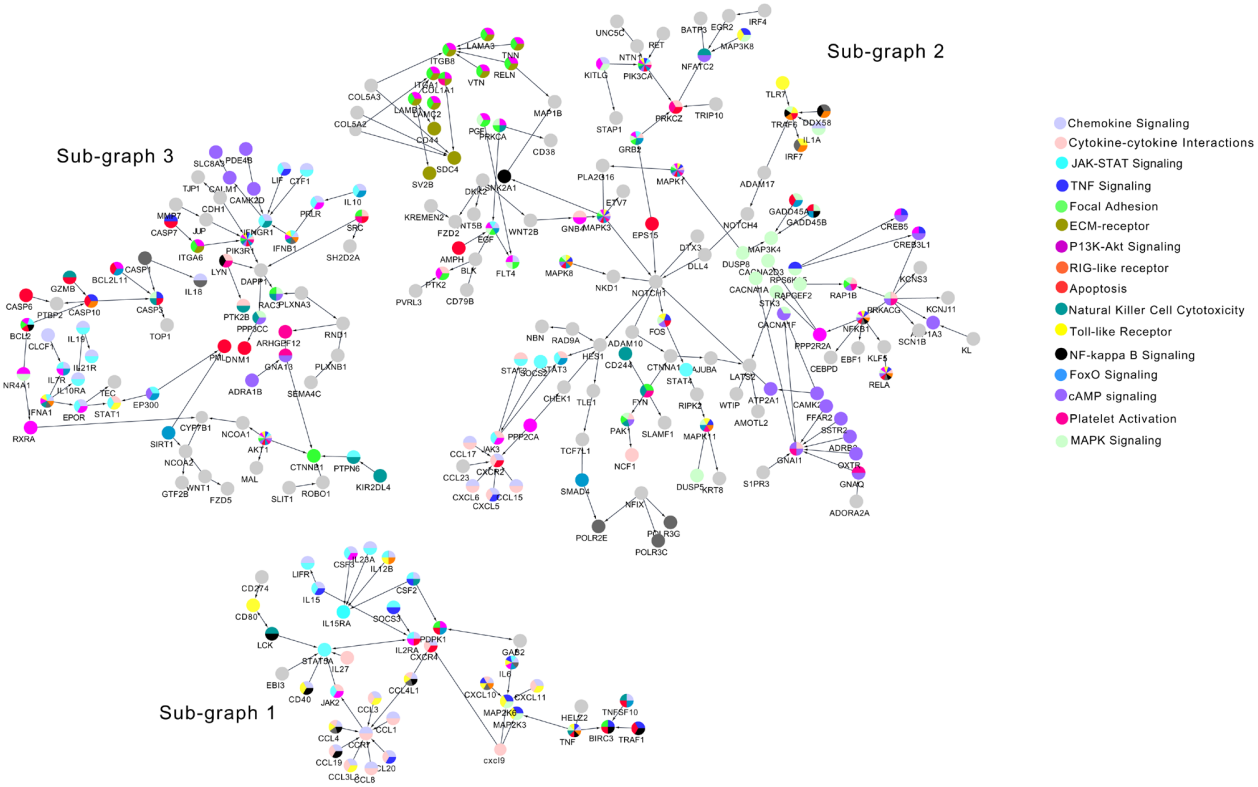


Figure 2. Visualization of the three modulated directed sub-graphs identified by `MSF` at 6 hpi. The node coloring is associated to KEGG pathways referring to the colors in the legend. The graph edges are from Reactome.

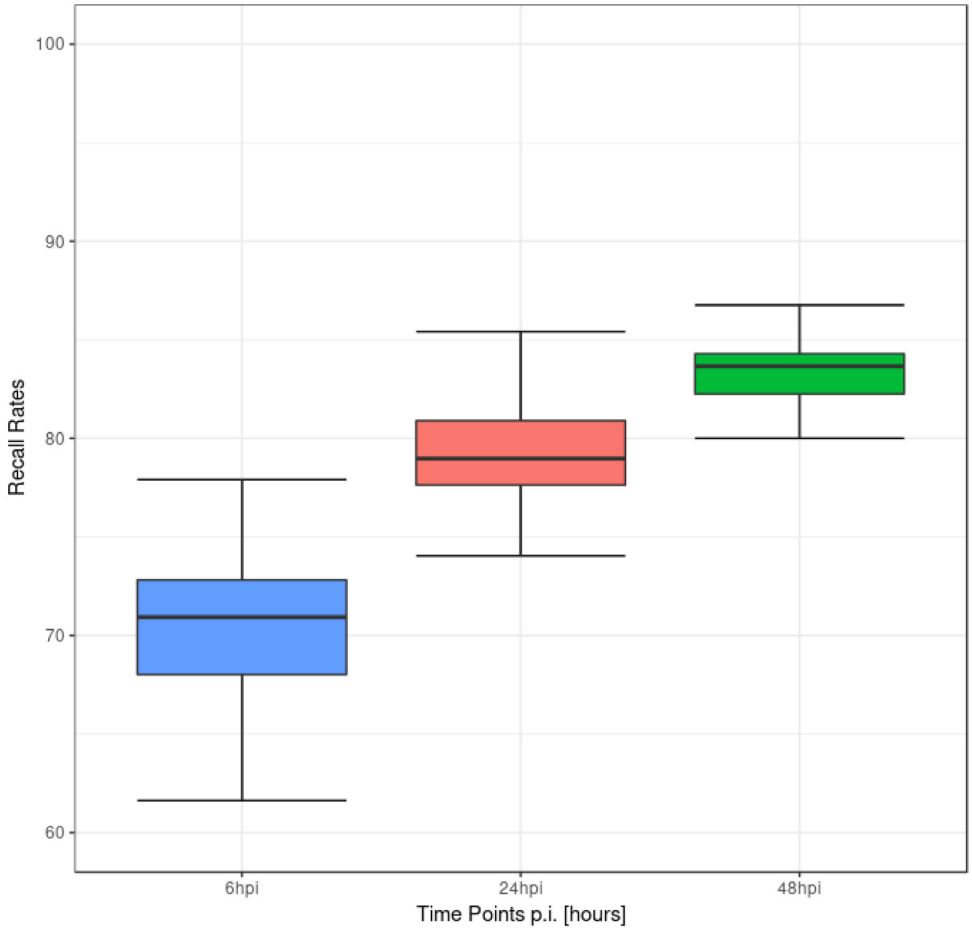


Figure 3. Recall rates for genes in `MSF` identified sub-graph for the three different time points of EBOV infection data for 100 simulations where Poisson distributed noise was added to the experimentally deduced reads per gene.

whole range of possible values. The observed median recall rates lay between 71 % (6 hpi) and 84 % (48 hpi).

Benchmark

The purpose of the comparisons to existing tools is to show the overall capabilities of `MSF`. `MSF` was compared to jActiveModules[8] since it uses similar approaches to find and score modules. For comparison to classical pathway enrichment analysis Reactome[5] and gene set enrichment analysis (GSEA)[22] was chosen since both are widely used and the latter does not rely on *p*-value cut-off.

jActiveModules. jActiveModules[8] is a plugin in Cytoscape that searches for molecular interaction network to find expression activated sub-networks. The method used to score the expression activated sub-networks is close to the method used in `MSF`. The difference is in how these sub-networks are identified. `MSF` starts building the sub-graphs from one gene, incorporating and combining the *p*-value of the next gene, with the check that the combined *p*-value of new sub-graph should be better than the original. On the other hand jActiveModules first transforms all the gene's *p*-values p_i to z-scores using $z_i = \Phi^{-1}(1-p_i)$, where Φ^{-1} is the inverse normal CDF and tries to find connected sets of genes with unexpectedly high levels of differential expression, in this case high z-scores. The overall score of the sub-network is calculated by combining the z-scores of the genes. Next using their extended simulated annealing method jActiveModules toggles multiple nodes to merge additional components.

The first time-point of Ebola infection data was analyzed using jActiveModules (Version 3.2.1) to compare the modulated sub-graphs identified by `MSF` and jActiveModules. The input files were same for both tools. From the modules identified by jActiveModules, the module with the highest pathScore was selected for comparison with `MSF` identified modulated sub-graphs. The module consists of a single graph with 314 genes. While `MSF` identified three directed modulated sub-graphs with 42, 69 and 139 genes. The overlap of the common genes identified between `MSF` and jActiveModules is shown in Figure 4. The sub-graphs identified by `MSF` are more fragmented than jActiveModules. Unfortunately there is no golden

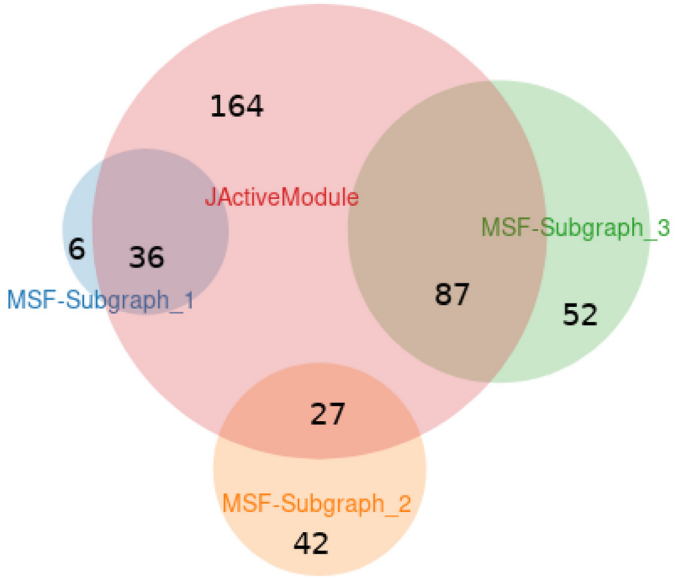


Figure 4. The Venn diagram shows the common genes identified as modulated from MSF identified sub-graphs and jActiveModules identified module.

standard example data that could help benchmark the method. MSF provides directionality, with the identification of possible perturbation sources of the sub-graphs. The predefined pathway labels could also be seen in MSF identified sub-graph with little effort using StringApp[10].

Reactome pathway analysis. Gene enrichment analysis was performed using Reactome analyze data tool[5] (version 67) on the different time-points of Ebola infection data. Reactome's over-representation analysis tool tests whether certain Reactome pathways are enriched for the lists of genes submitted to it. Genes from MSF identified sub-graphs for each time-point were analyzed for gene enrichment using this tool. For comparison the DEG results from edgeR for the three time-points were filtered using the cut-off of adjusted p-value < 0.05. These DEG lists were used for gene enrichment analysis. The compression of MSF identified sub-graphs gene lists and the DEG lists analysis is shown in Figure 5

All enriched pathways with a cut-off of p-value <0.05 for MSF and DEG lists for the three time-points were selected. The comparison shows most of the pathways known from literature to be dis-regulated by Ebola infection are enriched in both the enrichment analysis. EBOV glycoprotein (GP) interacts with the Toll-like recpetor signaling pathway, it triggers the activation of cytokines[13]. Toll-like receptor pathway is expected to be dis-regulated in the early stage of infection, this pathway was not identified as significantly dis-regulated when p-value cut off DEG lists were analyzed for enrichment. Nine Toll-like receptor cascades TLR10, TLR2, TLR3, TLR4, TLR5, TLR7/8, TLR9, TLR1:TLR2 and TLR6:TLR2 were identified as dis-regulated from gene enrichment analysis of MSF identified sub-graph genes, not a single one of these cascade was shown to be dis-regulated in pathway enrichment analysis from DEG cut-off lists. Since MSF considers the complete DEG results, even the weak signal at the earliest time-point was detected; for example Toll-like receptor signaling. While MSF is able to catch weak signals, it does not provide information about the functional relationships among genes like Reactome tool.

Gene set enrichment analysis. Gene set enrichment analysis (GSEA)[22] is a method to identify classes of genes or proteins that are overrepresented in a large set of genes or proteins. GSEA uses statistical approaches to identify significantly enriched or depleted groups of genes. The complete DEG list from DGEA of the first time-point 6 hpi was analyzed using bioconductor package GSEABase (version 1.44.0). GSEA was able to identify Toll-like receptor, chemokine signaling pathway, cytosolic DNA-sensing pathway, Jak-STAT signaling pathway, RIG-I-like receptor signaling pathway and apoptosis as the highest ranked pathways. Although GSEA identified the important pathways for Ebola infection, it did not show the topology of the genes in the different pathways identified and how they cross-talk. MSF and GSEA uses complete DEG list without any cut-off, that is why pathways important for Ebola infection showed up even with weak signal genes in it.

Discussion

Classic pathway analysis tools aim to detect in lists of significantly deregulated genes enriched associations with pathway genes categorized by their biological function and their interactions. Depending on the tool, the internal pathway topology is considered or neglected all together. Here presented tool, MSF, employs a different approach, by aiming to detect sub-graphs in whole gene regulatory networks which are significantly deregulated in a concerted manner. To this end, neighboring genes in the user provided network are tested for jointly common regulation. Exploiting that each gene's abundance, although not independent from its neighbors, is measured on its own, sensitivity can be increased by our applied p-value meta-analysis,

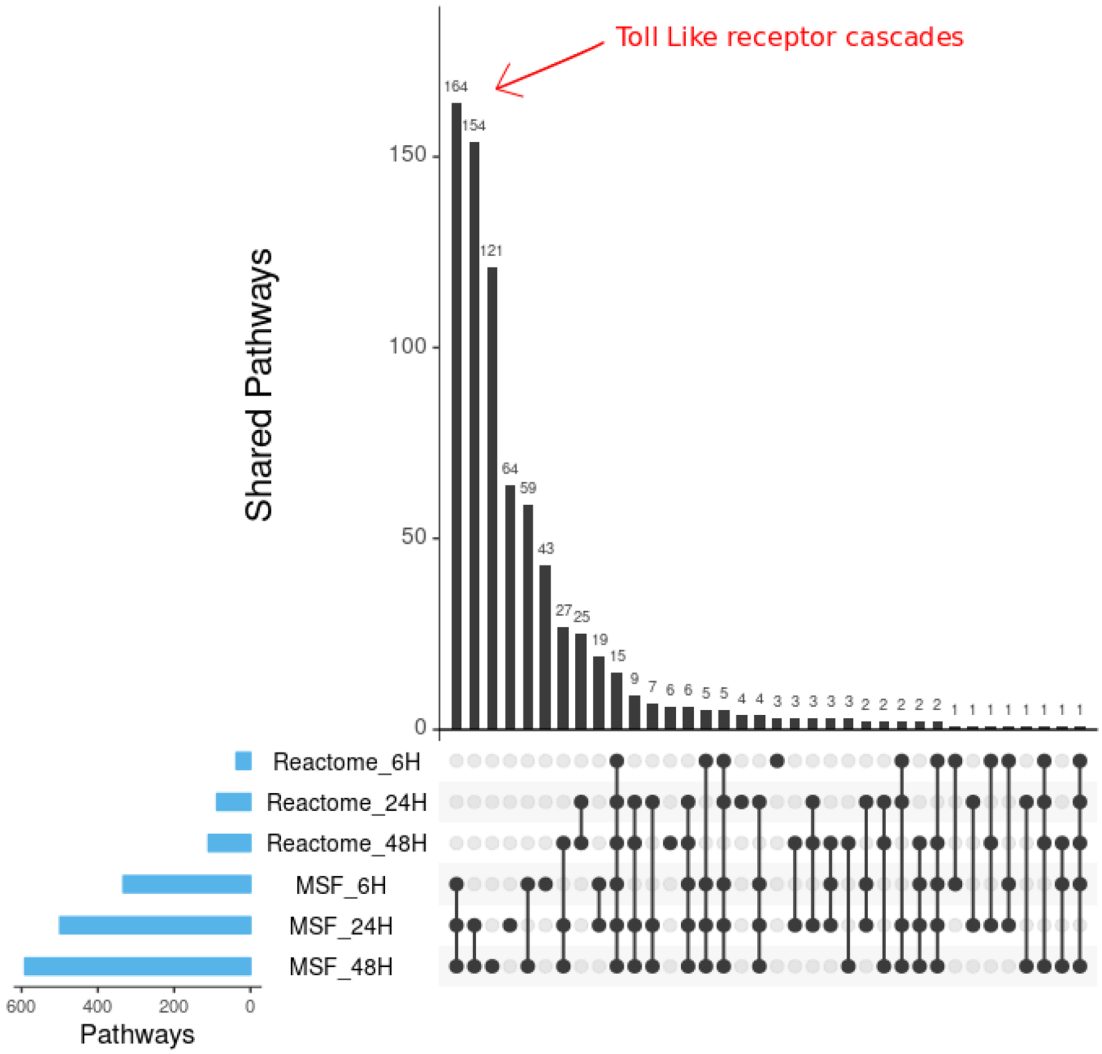


Figure 5. The Upset plot shows the number of shared pathways between `MSF` identified sub-graph gene list and DEG cut-off list for the three time-points. All 10 different Toll-like receptor cascades are in the set of 164 shared pathways only between `MSF` at different time-points.

namely Hartung's method. This potentially enables to call not just significant modulated genes based on the DGEA to be convincingly called to be part of a deregulated gene group. Furthermore, it allows to identify connected sub-graphs, representing the propagation of gene regulation perturbation in the input network. A better understanding of this propagation, especially the critical spots such as sensors, effectors, and hubs, facilitates the projection of potential intervention points, e.g., for drug development. Since `MSF` only uses interaction information in gene regulation network, but not the functional grouping of the genes into functional pathways, it is especially adapted to discover so called cross-talk between such pathways.

Conclusions

`MSF` is a fast and easy to use tool to find concertedly modulated sub-graphs in a given network. Its implementation in Java enables its use across many operating systems e.g. linux and windows. So far the raw output from edgeR[19] and DESeq2[23] are supported.

Data availability

The Ebola infection RNA-seq data set analyzed during the current study are available in the GEO repository (GSE84188)[13]. The cell signaling network file used is from Reactome Functional interactions (FIs) Version 2016[20]

Software availability

Source code is available from GitHub: https://github.com/Modulated-Subgraph-Finder/ MSF

Archived source code at time of publication: https://doi.org/10.5281/zenodo.2632973[24]

Software license: MIT license.

Grant information

This work was funded by the FWF ("Fonds zur Förderung der wissenschaftlichen Forschung") within the project Internationalen Kooperationsprojektes - Intl cooperation Project (Joint Project - Lead Agency Verfahren) with the project number (I 2353-B22). The grant was assigned to ILH. FA was funded by the Austrian Science Fund (FWF) project SFB F43.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Supplement material

Supplementary material is available form GitHub: https://github.com/Modulated-Subgraph-Finder/MSF

References

1. Malone JH, Oliver B: **Microarrays, deep sequencing and the true measure of the transcriptome.** *BMC Biol.* 2011; **9**(1): 34.
PubMed Abstract | Publisher Full Text | Free Full Text

2. Khatri P, Sirota M, Butte AJ: **Ten years of pathway analysis: current approaches and outstanding challenges.** *PLoS Comput Biol.* 2012; **8**(2): e1002375.
PubMed Abstract | Publisher Full Text | Free Full Text

3. García-Campos MA, Espinal-Enríquez J, Hernández-Lemus E: **Pathway Analysis: State of the Art.** *Front Physiol.* 2015; **6**: 383.
PubMed Abstract | Publisher Full Text | Free Full Text

4. Kanehisa M, Goto S: **KEGG: kyoto encyclopedia of genes and genomes.** *Nucleic Acids Res.* 2000; **28**(1): 27–30.
PubMed Abstract | Publisher Full Text | Free Full Text

5. Fabregat A, Jupe S, Matthews L, *et al.*: **The Reactome Pathway Knowledgebase.** *Nucleic Acids Res.* 2018; **46**(D1): D649–D655.
PubMed Abstract | Publisher Full Text | Free Full Text

6. Bayerlová M, Jung K, Kramer F, *et al.*: **Comparative study on gene set and pathway topology-based enrichment methods.** *BMC Bioinformatics.* 2015; **16**(1): 334.
PubMed Abstract | Publisher Full Text | Free Full Text

7. Mitra K, Carvunis AR, Ramesh SK, *et al.*: **Integrative approaches for finding modular structure in biological networks.** *Nat Rev Genet.* 2013; **14**(10): 719–732.
PubMed Abstract | Publisher Full Text | Free Full Text

8. Ideker T, Ozier O, Schwikowski B, *et al.*: **Discovering regulatory and signalling circuits in molecular interaction networks.** *Bioinformatics.* 2002; **18 Suppl 1**: S233–S240.
PubMed Abstract | Publisher Full Text

9. Michalak P: **Coexpression, coregulation, and cofunctionality of neighboring genes in eukaryotic genomes.** *Genomics.* 2008; **91**(3): 243–248.
PubMed Abstract | Publisher Full Text

10. Morris J, Jensen LJ, Doncheva NT: **stringApp 1.3.0**. [Online; accessed 19-Februar-2018]. 2018.
Reference Source

11. Shannon P, Markiel A, Ozier O, *et al.*: **Cytoscape: a software environment for integrated models of biomolecular interaction networks.** *Genome Res.* 2003; **13**(11): 2498–2504.
PubMed Abstract | Publisher Full Text | Free Full Text

12. Hartung J: **A note on combining dependent tests of significance**. Technical report, Technical Report, SFB 475: Komplexitätsreduktion in Multivariaten Datenstrukturen, Universität Dortmund, 1998.
Reference Source

13. Olejnik J, Forero A, Deflubé LR, *et al.*: **Ebolaviruses Associated with Differential Pathogenicity Induce Distinct Host Responses in Human Macrophages.** *J Virol.* 2017; **91**(11): pii: e00179-17.
PubMed Abstract | Publisher Full Text | Free Full Text

14. Prins KC, Cárdenas WB, Basler CF: **Ebola virus protein vp35 impairs the function of interferon regulatory factor-activating kinases IKKepsilon and TBK-1.** *J Virol.* 2009; **83**(7): 3069–3077.
PubMed Abstract | Publisher Full Text | Free Full Text

15. Konde MK, Baker DP, Traore FA, *et al.*: **Interferon β-1a for the treatment of Ebola virus disease: A historically controlled, single-arm proof-of-concept trial.** *PLoS One.* 2017; **12**(2): e0169255.
PubMed Abstract | Publisher Full Text | Free Full Text

16. Rhein BA, Powers LS, Rogers K, *et al.*: **Interferon-γ Inhibits Ebola Virus Infection.** *PLoS Pathog.* 2015; **11**(11): e1005263.
PubMed Abstract | Publisher Full Text | Free Full Text

17. Falasca L, Agrati C, Petrosillo N, *et al.*: **Molecular mechanisms of Ebola virus pathogenesis: focus on cell death.** *Cell Death Differ.* 2015; **22**(8): 1250–9.
PubMed Abstract | Publisher Full Text | Free Full Text

18. Cárdenas WB, Loo YM, Gale M Jr, *et al.*: **Ebola virus vp35 protein binds double-stranded RNA and inhibits alpha/beta interferon production induced by RIG-I signaling.** *J Virol.* 2006; **80**(11): 5168–5178.
PubMed Abstract | Publisher Full Text | Free Full Text

19. Robinson MD, McCarthy DJ, Smyth GK: **edger: a Bioconductor package for differential expression analysis of digital gene expression data.** *Bioinformatics.* 2010; **26**(1): 139–140.
PubMed Abstract | Publisher Full Text | Free Full Text

20. Wu G, Feng X, Stein L: **Reactome FIs**. [Online; Version 2016]. 2016.
Reference Source

21. Veljkovic V, Glisic S, Muller CP, *et al.*: ***In silico* analysis suggests interaction between Ebola virus and the extracellular matrix.** *Front Microbiol.* 2015; **6**: 135.
PubMed Abstract | Publisher Full Text | Free Full Text

22. Subramanian A, Tamayo P, Mootha VK, *et al.*: **Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.** *Proc Natl Acad Sci U S A.* 2005; **102**(43): 15545–50.
PubMed Abstract | Publisher Full Text | Free Full Text

23. Love MI, Huber W, Anders S: **Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.** *Genome Biol.* 2014; **15**(12): 550.
PubMed Abstract | Publisher Full Text | Free Full Text

24. Farman M: **Modulated-Subgraph-Finder/MSFv2.1 (Version v2.1).** *Zenodo.* 2019.
http://www.doi.org/10.5281/zenodo.2632973

Publication 2

"Transcriptional Profiling of Human Macrophages during Infection with Bordetella Pertussis"

Denisa Petráčkováa, **Mariam R. Farman**, Fabian Amman, Irena Linhartovád, Ana Dienstbiera, Dilip Kumara, Jakub Držmíšek, Ivo Hofacker, Maria Eugenia Rodriguezf and Branislav Večerek

Status: published in RNA Biology, 2020
https://doi.org/10.1080/15476286.2020.1727694

Author Contributions:

MARIAM R. FARMAN - Provided significant contributions to the planning and structuring of the experimental setup, including determination of time-points, number of replicates and sequencing depth. Developed the analytical pipeline, conducted comprehensive bioinformatics analyses, interpreted the results, and authored the manuscript.

IVO HOFACKER - Supervised the project and contributed to the manuscript.

FABIAN AMMAN - Supervised the project and contributed to the manuscript.

COLLABORATORS - Conceived, designed, and performed the experiments.

RNA BIOLOGY
2020, VOL. 17, NO. 5, 731–742
https://doi.org/10.1080/15476286.2020.1727694

Taylor & Francis
Taylor & Francis Group

RESEARCH PAPER

OPEN ACCESS Check for updates

Transcriptional profiling of human macrophages during infection with *Bordetella pertussis*

Denisa Petráčková[a*], Mariam R. Farman [b*], Fabian Amman [b,c], Irena Linhartová[d], Ana Dienstbier[a], Dilip Kumar[a], Jakub Držmíšek [a], Ivo Hofacker [b,e], Maria Eugenia Rodriguez[f], and Branislav Večerek [a]

[a]Institute of Microbiology of the Czech Academy of Sciences, Laboratory of Post-transcriptional Control of Gene Expression, Prague, Czech Republic; [b]Institute for Theoretical Chemistry, University of Vienna, Vienna, Austria; [c]Division of Cell and Developmental Biology, Medical University of Vienna, Vienna, Austria; [d]Institute of Microbiology of the Czech Academy of Sciences, Laboratory of Molecular Biology of Bacterial Pathogens, Prague, Czech Republic; [e]Faculty of Computer Science, Research Group Bioinformatics and Computational Biology, University of Vienna, Vienna, Austria; [f]Facultad de Ciencias Exactas, Universidad Nacional de La Plata, CINDEFI (UNLP CONICET La Plata), La Plata, Argentina

ABSTRACT

Bordetella pertussis, a strictly human re-emerging pathogen and the causative agent of whooping cough, exploits a broad variety of virulence factors to establish efficient infection. Here, we used RNA sequencing to analyse the changes in gene expression profiles of human THP-1 macrophages resulting from *B. pertussis* infection. In parallel, we attempted to determine the changes in intracellular *B. pertussis*-specific transcriptomic profiles resulting from interaction with macrophages. Our analysis revealed that global gene expression profiles in THP-1 macrophages are extensively rewired 6 h post-infection. Among the highly expressed genes, we identified those encoding cytokines, chemokines, and transcription regulators involved in the induction of the M1 and M2 macrophage polarization programmes. Notably, several host genes involved in the control of apoptosis and inflammation which are known to be hijacked by intracellular bacterial pathogens were overexpressed upon infection. Furthermore, *in silico* analyses identified large temporal changes in expression of specific gene subsets involved in signalling and metabolic pathways. Despite limited numbers of the bacterial reads, we observed reduced expression of majority of virulence factors and upregulation of several transcriptional regulators during infection suggesting that intracellular *B. pertussis* cells switch from virulent to avirulent phase and actively adapt to intracellular environment, respectively.

ARTICLE HISTORY
Received 11 December 2019
Accepted 12 December 2019

KEYWORDS
Bordetella pertussis; infection; macrophage; intracellular survival; host-pathogen interaction

Introduction

Bordetella pertussis is a Gram-negative strictly human pathogen of the respiratory tract and the aetiological agent of whooping cough [1]. Despite vaccination programmes, pertussis remains one of the 10 most common causes of vaccine-preventable deaths [2]. Furthermore, pertussis incidence is currently on the rise in industrialized countries with highly vaccinated populations [3,4]. The re-emergence of pertussis strongly suggests that we need to widen our understanding of the molecular mechanisms underlying the pathogenicity of *B. pertussis* [5,6].

Upon infection, *B. pertussis* cells are challenged by the host innate immune system. While ciliated epithelial cells secrete mucus that mechanically eliminates the bacteria, resident as well as attracted phagocytic cells interact with the pathogen in order to clear the infection [7]. Successful infection, therefore, depends on activities of a great variety of virulence factors produced by *B. pertussis* cells [8]. In particular, adhesins such as filamentous haemagglutinin and fimbriae and toxins such as pertussis toxin and adenylate cyclase toxin contribute to efficient adhesion and colonization of the epithelium as well as to modulation of the immune system and thereby facilitate the survival of bacteria in the respiratory tract [9]. *B. pertussis* was historically described as an extracellular pathogen, which, in the presence of specific antibodies, can be internalized and killed by phagocytic cells [10]. Nevertheless, some early reports suggested that in the absence of opsonization this pathogen stimulates its own attachment to immune cells resulting in inefficient killing by phagocytes and increased intracellular survival [11–15]. Importantly, *B. pertussis* cells were found also within human, rabbit, and mouse alveolar macrophages *in vivo* [16–18]. Within the last decade, Rodriguez and her team presented data indicating that *B. pertussis* can survive in primary macrophages [19]. They have shown that the majority of *B. pertussis* cells are killed within acidic compartments; however, some residual cells evade this killing and survive and replicate inside a non-acidic early endosome-like compartment [20,21]. Indeed, *B. pertussis* uses several mechanisms to avoid phagocytic killing. For example, immune evasion mediated by both pertussis and adenylate cyclase toxins results in subversion of cellular signalling, repression of phagocytic activity and, eventually, in induction of apoptosis [22–25]. Collectively, these data led to speculations that *B. pertussis* could use macrophages as an intracellular niche and

CONTACT Fabian Amman fabian@tbi.univie.ac.at Institute for Theoretical Chemistry, University of Vienna, Währinger Straße 17, Vienna A-1090, Austria; Branislav Večerek vecerek@biomed.cas.cz Czech Academy of Sciences, Laboratory of Post-transcriptional Control of Gene Expression, Institute of Microbiology v.v.i., Prague 14220, Czech Republic*These authors contributed equally to this work
Supplemental data for this article can be accessed here

that the intramacrophage phase of infection could play a significant role in survival and persistence of bacteria within the host [12,17,18]. In support, gene expression profiling and proteomic analysis of infected monocyte-derived THP-1 macrophages revealed that intracellular *B. pertussis* cells modulate expression of several genes involved in response to the host immune system and reshape production of dozens of proteins implicated in virulence [20,21].

In this study, representing the first attempt to assess the global response of human phagocytes to *B. pertussis* infection, we aimed at RNA-seq analysis of transcriptomic profiles in monocyte-derived THP-1 macrophages infected with *B. pertussis*. In addition, we were tempted to decipher whether we can also monitor the pathogen gene expression profiles during the infection to gain insight into its response and adaptation to the intracellular environment.

Results

Global transcriptomic profile of THP-1 macrophages infected with *B. pertussis*

THP-1 monocyte-derived macrophages were infected with *B. pertussis* reference strain Tohama I. Samples of uninfected macrophages (C; control) were collected 6 h post-infection (pi) while infected macrophages were collected at three time points (T1-T3) corresponding to 2, 6 and 24 h pi with the objective of a) elucidation of the general macrophage response triggered by *B. pertussis* (6 h pi) and b) monitoring the time-dependent changes in gene expression profiles resulting from the host–pathogen interaction within the early phase of infection (T2 versus T1) and within the late phase of infection (T3 versus T2). We did not use the uninfected controls for T1 and T3 time points as the changes in gene expression profiles of THP-1 macrophages incubated for 24 h in RPMI medium are not substantial [26] and in particular, expression of several cytokines and chemokines remains unchanged during 24 h incubation in RPMI medium [27,28]. RNA-seq analysis of samples from infected cells and uninfected control yielded on average 16 million reads which were mapped to human and *B. pertussis* genomes (Fig S1). Principal component analysis (PCA) of RNA-seq data revealed that samples from uninfected macrophages clustered separately from infected cells and that also samples of infected macrophages collected at 2 h, 6 h and 24 h pi clustered apart from each other (Fig. 1A). Thus, PCA indicated temporal changes in gene expression profiles resulting from *B. pertussis* infection. To identify these alterations, differential expression (DE) analysis was performed. First, we compared global gene expression profiles in infected cells and control uninfected cells harvested 6 h pi (T2 versus C). For the host, DE analysis identified 679 macrophage genes as significantly modulated ($|\log_2 FC| \geq 1$; adjusted p-value < 0.05) upon infection (Table S1). Expression of hundreds of genes was strongly upregulated upon infection including those encoding cytokines, chemokines, cell receptors and transcription factors (Fig. 1B).

Many of the highly expressed genes encode important factors of the innate immunity such as IL-1A, IL-6, IL-23A, TNF superfamily proteins, CCL and CXCL chemokine families and NFκB and STAT4 regulators and belong to common expression pattern mounted by host cells upon bacterial infection [29]. Of note, among the significantly upregulated genes were also anti–inflammatory cytokines *IL-13* and *IL-10* and *ADORA2A* gene encoding the adenosine receptor A_{2A} which senses adenosine levels and

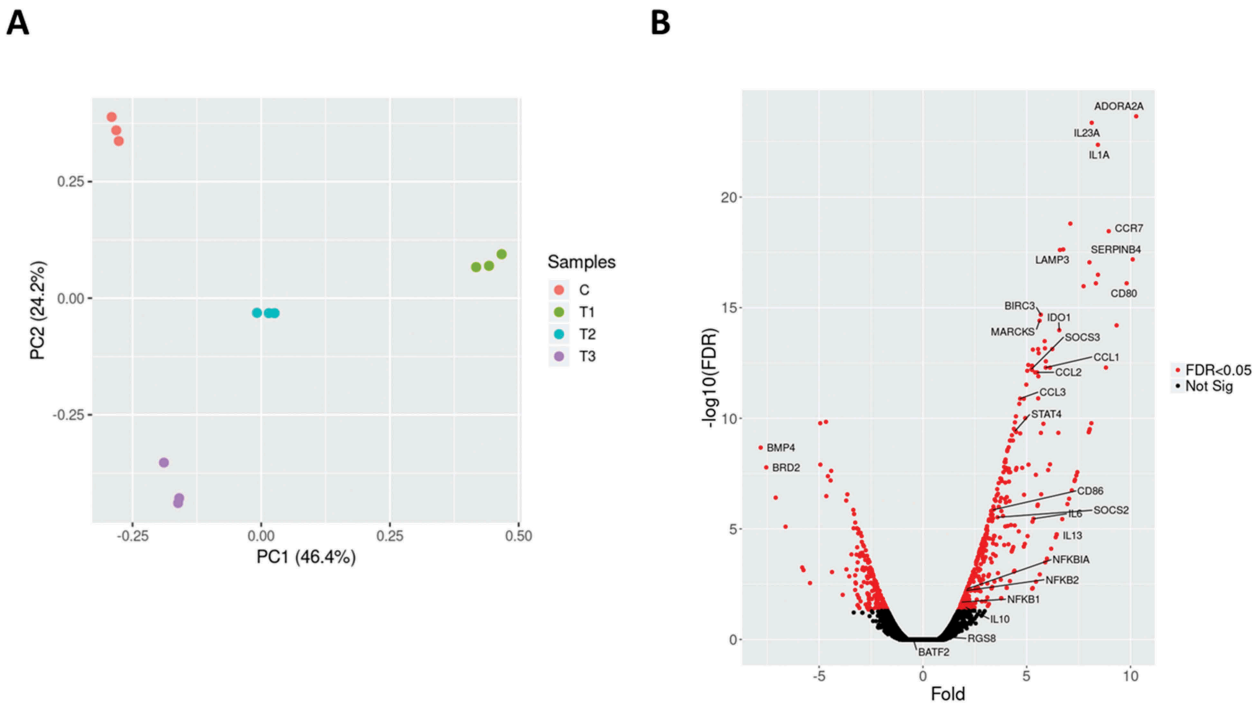


Figure 1. (A) Principal component analysis was applied to transcriptomic profiles of the uninfected THP-1 macrophages (C; red circles) and *B. pertussis*-infected THP-1 macrophages harvested 2 h (T1, green circles), 6 h (T2, blue circles) and 24 h (T3, cyan circles) pi. Each dot represents an independent biological replicate. (B) Volcano plot showing the global transcriptional changes in the infected THP-1 cells 6 h pi. The red dots represent significantly differentially expressed genes, labelled genes are discussed in this work. The black dots represent nonsignificantly modulated genes.

triggers the anti–inflammatory signalling [30]. On the other side, among the highly downregulated genes were *BMP4* and *BRD2*. BMP4 belongs to the TGF-β superfamily of proteins and was reported to induce M2 polarization in macrophages [31], while transcriptional regulator BRD2 interacts with chromatin and plays important role in inflammatory response in mice [32]. Collectively, these results suggest that both pro-inflammatory and anti–inflammatory programmes that presumably result from host-pathogen interplay are active in infected cells.

To obtain a deeper insight into the functional clustering of up- and downregulated DE genes, we performed GO term and KEGG pathway enrichment analyses. As shown in Fig. 2, the Gene Set Enrichment Analysis (GSEA) revealed that genes upregulated upon infection were significantly and selectively enriched within GO terms such as 'Response to external stimulus', 'Immune response', 'Leukocyte activation', 'Locomotion' and 'Regulation of immune system process', while downregulated genes were enriched predominantly within 'Respiratory chain' and 'Mitochondrial membrane part' terms. KEGG pathway analysis presented in Fig. 3 identified several regulatory circuits to be selectively enriched for upregulated genes such as 'NOD-like receptor signalling pathway', 'Chemokine signalling pathway'or 'Antigen processing and presentation'. On the other side, 'mTOR signalling pathway' and several metabolic pathways were enriched for downregulated genes.

Finally, to identify coherently modulated genes within the global cell signalling network we used the Modulated Subgraph Finder (MSF) tool. MSF combines the whole cell signalling network with data from DE analysis to identify the modulated networks. The networks identified by MSF help to reveal the flow and cross-talk between different KEGG pathways. As presented in Fig. 4, the selected network shows the signalling between cytokine and chemokines affecting Th1, Th2 and Th17 profiling and apoptosis in immune cells. For example, the graph depicts activation of STAT4 by IL-23 and IL-18R1 or activation of CXCR4 receptor by members of CCL and CXCL chemokine families.

Temporal changes in global transcriptomic profiles of infected THP-1 macrophages

Next, we focused on time-dependent changes in transcriptomic profiles of infected macrophages. We compared the gene expression profiles in cells harvested 6 h and 2 h pi (T2 versus T1) and in cells harvested 24 h and 6 h pi (T3 versus T2). DE analysis identified 561 and 553 of significantly ($|\log_2 FC| \geq 1$; adjusted p-value < 0.05) modulated genes between T2-T1 and T3-T2 time points, respectively (Table S2). Significantly up- and downregulated genes were used for GSEA to identify enriched GO terms between T2-T1 and T3-T2 time points (Fig. S2). Apparently, most of the GO terms were not selectively enriched for either up- or downregulated genes. Nevertheless, as shown in Fig. S3a, when we compared DE genes modulated between T2 and T1 time points, KEGG pathway analysis unravelled several pathways consistently enriched for upregulated (e.g. 'Phagosome', 'Cell adhesion molecules' and 'Antigen processing and presentation') or downregulated (e.g. 'DNA replication', 'Ubiquitin mediated proteolysis' and 'Insulin signalling pathway') genes. Of note, several pathways such as 'MAPK signalling pathway',

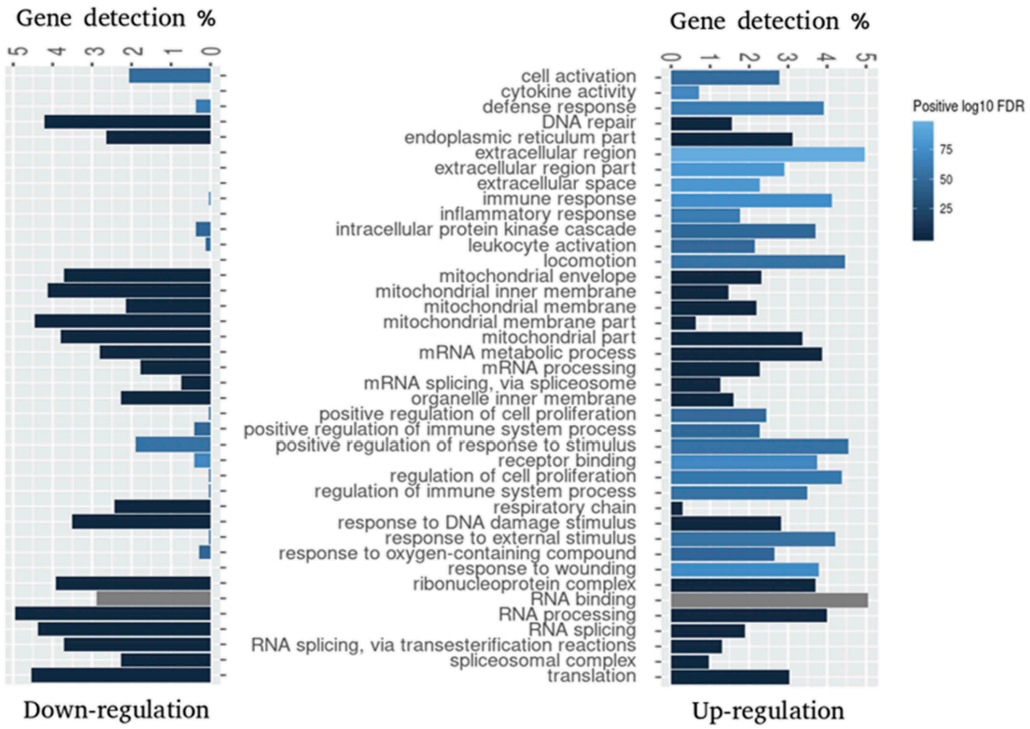


Figure 2. Gene Set Enrichment Analysis (GSEA) of GO terms enriched for differentially expressed genes 6 h pi. GSEA was applied to identify enriched GO terms within either downregulated ($\log_2 FC < 0$) or upregulated ($\log_2 FC > 0$) gene sets. The bars depict the percentage of DE genes associated with the specific GO term and shades of blue indicate positive $\log_{10}$ of FDR range. Grey bars indicate that this GO term was not significantly enriched within the corresponding gene set. Top 20 enriched gene sets were selected for visualization.

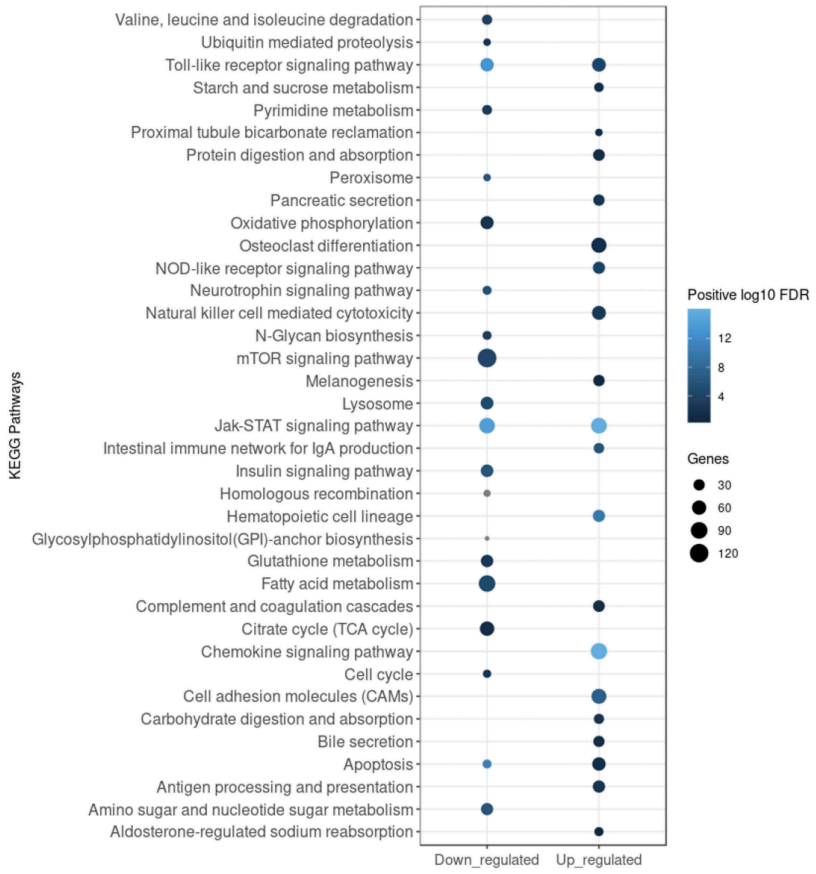


Figure 3. Enrichment analysis of KEGG pathways enriched for DE genes 6 h pi. Analysis was applied to identify pathways enriched for either downregulated ($log_2FC < 0$) or upregulated ($log_2FC > 0$) DE genes. Top 20 enriched KEGG pathways for each subset of genes were selected for visualization. The legend shows the positive log_{10} of FDR range (shades of blue) and the number of DE genes in each KEGG pathway expressed by the size of the dots.

'Toll-like receptor pathway' and 'Hedgehog signalling pathway' were highly enriched for genes whose expression was upregulated between 6 h and 24 h pi (Fig. S3b).

Validation of the RNA-seq data by quantitative PCR

To validate the gene expression profiles obtained from RNA-seq analysis, the expression of selected THP-1 genes was assayed by RT-qPCR method. Comparative RT-qPCR analysis of infected and uninfected cells performed at all three time points allowed us also to confirm temporal changes in gene expression profiles deduced from RNA-seq results. As shown in Fig. 5A, results of quantitative PCR analysis of eight selected genes of different functional categories were in very good agreement with RNA-seq data. Moreover, temporal changes in the gene expression deduced from RNA-seq results are also concordant with the outcome of quantitative PCR analysis. We could, for example, verify that expression of *IL10* or *SOCS3* genes is reduced during the infection or that expression of *STAT4, LAMP3, ADORA2A* and *BIRC3* genes peaks 6 h pi.

Changes in gene expression profiles of intracellular *B. pertussis* cells

Additionally, as an important part of this project, we tested the suitability of our infection protocol for parallel monitoring of the pathogen gene expression responses. Therefore, at this stage we did not fully focus on DE analysis of intracellular *B. pertussis* cells, nevertheless, we quantified abundant bacterial transcripts in samples of infected macrophages to monitor their expression profiles in the course of infection. Expectedly, more than 99% of the reads could be mapped to the human genome while the number of *Bordetella*-specific informative reads (those mapping to protein-coding genes) in samples harvested 6 h (≈ 10 000 reads/sample) and 24 h pi (≈ 6 500 reads/sample) decreased when compared to samples harvested 2 h pi (≈ 38 000 reads/sample; Fig. S1) and thereby mirrored reduced numbers of viable intracellular *Bordetella* cells recovered from macrophages in the course of infection (from 2.1 x 10^6/well at 2 h pi to 1.0 x 10^6/well at 6 h pi and 1.5 x 10^5/well at 24 h pi). In spite of low sequencing yield (see Table S3 for numbers of mapped reads), we asked whether we could compare the gene expression profiles of intracellular bacteria between time points T2 and T1 and between time points T3 and T2. This analysis identified 281 and 55 significantly ($|log_2FC| \geq 1$; adjusted p-value < 0.1) differentially expressed *B. pertussis* genes, respectively (Table S4). Among the genes strongly induced during infection were ferrous iron transporter *ftrA*, several genes implicated in sulphur metabolism including transcriptional factor *iscR*, cysteine desulfurase *iscS*, and cysteine dioxygenase *BP2871* and genes involved in vitamin B6 metabolism and homoeostasis (*BP1920* and *BP1921*) (Table 1). Of note, large number of genes encoding

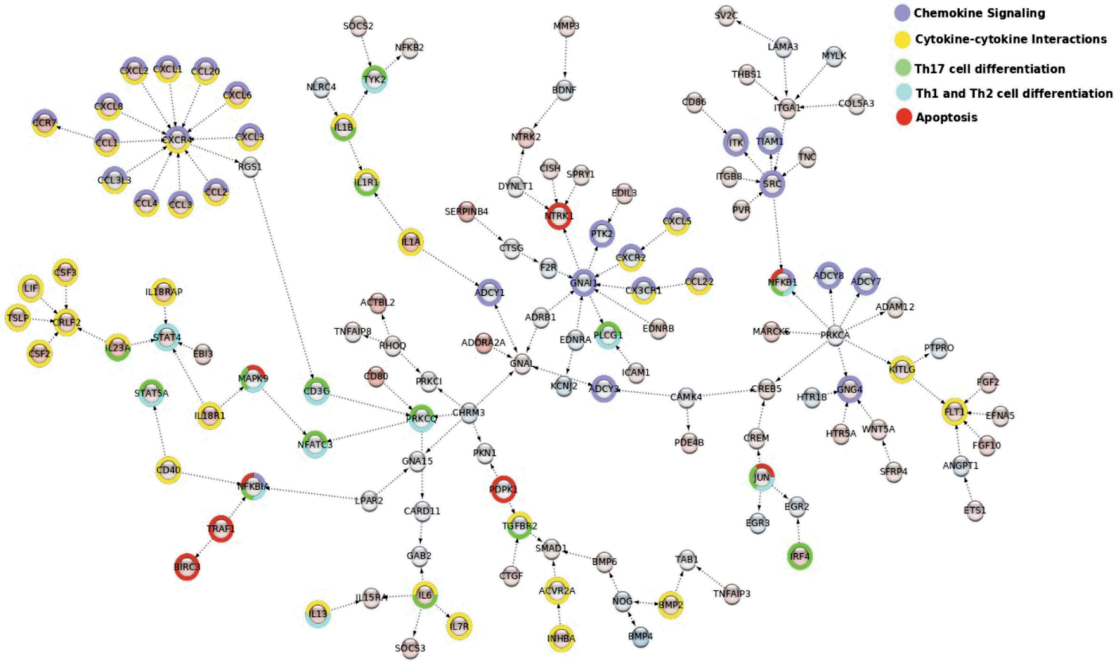


Figure 4. MSF identification of modulated interacting networks. Networks were generated using data from DE analysis. The network file created by MSF was used to visualize the modulated networks in Cytoscape. The nodes represent the genes and the edges show the direction of interaction. The enrichment within different KEGG pathways (see legend) is shown by the colours around the nodes. The colouring inside the node depicts the range of log fold change for upregulated (shades of red) and downregulated genes (shades of blue).

transcriptional regulators including *BP0135, BP0823, BP1319, BP2216*, sigma factors *rpoE* and *rpoH* and RNA chaperone *hfq* were overexpressed during the infection.

On the other side, expression of several virulence genes such as *cyaDE, tcfA, prn, vag8, brkA* and genes within *fim, ptx/ptl* and T3SS-specific *bsc/btr* operons was significantly reduced in the course of infection (Table 2).

In support, the expression of *bvgR* regulatory gene which is required for expression of virulence genes was also reduced upon infection while the expression of virulence repressed gene *vrg6* was increased in the course of infection. Importantly, the outcome of quantitative PCR analysis of *vag8, prn, vrg6* and *BP2871* genes was fully in agreement with DE analysis of *B. pertussis*-specific reads determined in samples of infected macrophages (Fig. 5B). Indeed, as inferred from RNA-seq data (Tables 1 and 2), the expression of *vag8* and *prn* virulence genes was decreased in the course of infection while the expression of *vrg6* and *BP2871* genes was increased and these trends were clearly confirmed by RT-qPCR analysis. Notably, when compared to unexposed control, the expression of cysteine dioxygenase gene *BP2871* was increased more than 70-fold already 2 hours pi.

GO term enrichment analysis of genes significantly modulated between time points T2 and T1 revealed several affected cellular processes. As shown in Fig. 6A, within the set of genes which were significantly upregulated, processes such as 'Cell redox homoeostasis', 'Transmembrane transport', 'Phosphorylation' and 'Oxidation-reduction process' were highly enriched. On the other side, genes belonging to 'Ribosome biogenesis', 'Dephosphorylation', 'Oxidation-reduction process', 'Pathogenesis' and 'ATP synthesis coupled proton transport' terms were enriched among the transcripts which were significantly downregulated (Fig. 6B). Although only very limited number of genes displayed differential expression levels between time points T3 and T2, biological processes such as 'Phosphorelay signal transduction' and 'Regulation of transcription' were enriched for upregulated genes while 'Pathogenesis', 'Proteolysis' and'Cell adhesion' terms were among the processes enriched for downregulated genes (data not shown).

Discussion

This study represents the first RNA-seq-based attempt to assess the global transcriptomic response of human phagocytes triggered by infection with *B. pertussis* cells. So far the transcriptional profiling of *B. pertussis* interactions with host cells was studied predominantly in mouse model and was focused either on the host [33,34] or the pathogen [35,36]. Nevertheless, mouse is not a natural host of *B. pertussis* and in our study we were interested in specific interaction with human macrophages rather than with complex mouse immune system. To avoid donor-to-donor variation observed in experiments with primary macrophages, we used the THP-1 monocyte-derived cells, a popular cell model for immune modulations studies [37] which has also been shown to be a good surrogate for primary cells in experiments with *B. pertussis* [21]. Profiling of human respiratory epithelial cells [38] and THP-1 macrophages [21] following *B. pertussis* infection was studied using DNA microarrays and quantitative PCR, respectively, however, global RNA-seq

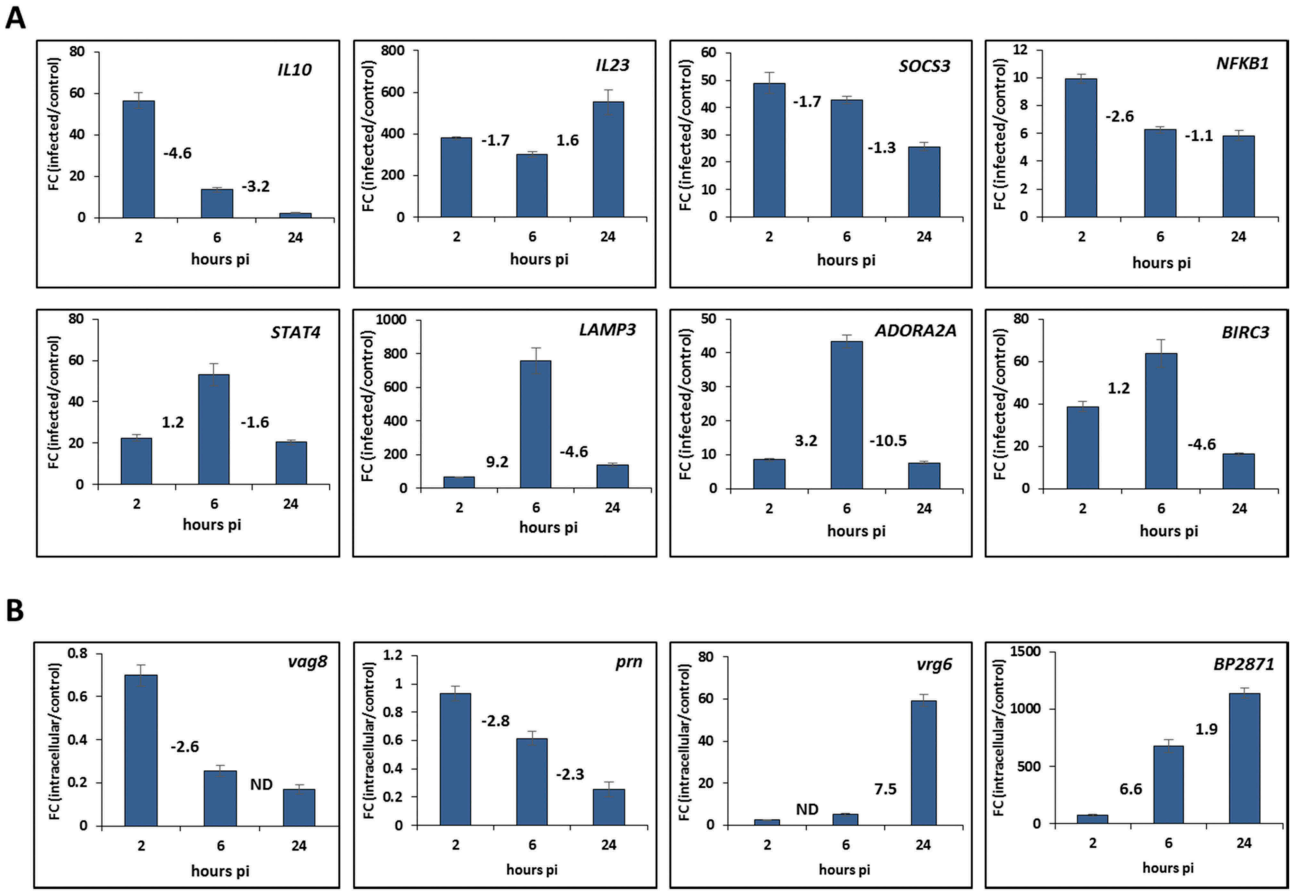


Figure 5. Validation of the RNA-seq results with quantitative PCR. (A) RT-qPCR analysis was performed to assay the relative expression profiles of *IL10, IL23, SOCS3, NFKB1, STAT4, LAMP3, ADORA2A* and *BIRC3* genes in infected THP-1 macrophages. Relative gene expression was compared between infected and uninfected macrophages harvested 2, 6 and 24 h post-infection (pi). (B) RT-qPCR analysis was performed to assay the relative expression profiles of *vag8, prn, vrg6* and *BP2871* genes in intracellular *B. pertussis* cells. Relative gene expression was compared between intracellular and unexposed bacteria 2, 6 and 24 h post-infection (pi). Fold change (FC) values are means (bars) ± standard deviations (error bars) from three biological replicate experiments. Values depicted between the bars denote the fold changes in expression of the specific gene between corresponding time points deduced from RNA-seq results determined 2, 6 and 24 h pi. ND, not determined in the corresponding analysis.

Table 1. List of selected *B. pertussis* genes upregulated in the course of infection[a].

Gene ID	log_2FC T2/T1	log_2FC T3/T2	Name	Function
BP0135	ND	3.68	*BP0135*	LysR transcriptional regulator
BP0726	4.02	*0.95*	*BP0726*	ABC transporter
BP0823	ND	3.34	*BP0823*	transcriptional regulator
BP1152	2.09	4.01	*ftrA*	ferrous iron transport
BP1319	2.75	ND	*BP1319*	GntR transcriptional regulator
BP1320	2.67	1.76	*BP1320*	pyridoxal homoeostasis protein
BP1321	2.05	1.66	*pdxK*	pyridoxine kinase
BP1736	ND	2.55	*BP1736*	OsmY-like protein
BP1798	*1.18*	2.96	*iscR*	Fe-S transcriptional regulator
BP1799	1.69	2.73	*iscS*	cysteine desulfurase
BP2193	1.13	1.35	*hfq*	RNA chaperon
BP2216	3.12	1.80	*BP2216*	MarR transcriptional regulator
BP2437	*1.16*	1.56	*sigE*	sigma factor RpoE
BP2468	ND	2.90	*vrg6*	virulence repressed gene
BP2871	2.72	*0.90*	*BP2871*	cysteine dioxygenase
BP2782	ND	2.41	*BP2782*	lipoprotein
BP3213	1.49	1.46	*gloB*	glyoxalase II
BP3748	2.33	1.71	*BP3748*	sigma factor RpoH

Values which did not pass the statistical significance but showed similar trend are shown in italics. ND, not determined in the respective analysis.

analysis of transcriptional responses of human phagocytic cells to *B. pertussis* cells has not been performed as yet. We have applied the dual RNA-seq approach allowing for analysis of transcriptomic profiles in both host and pathogen [39]; however, the major focus was set on the host side. To minimize the possible bias coming from additional treatment of samples delaying the RNA isolation (enrichment of the sample for bacterial RNA, time-consuming separation of infected and uninfected macrophages by flow cytometry), the macrophages were instantly lysed and used for isolation of the total RNA. This approach, however, brought some possible limitation to our study. First, as expected from previously published data [19] and our preliminary results, the numbers of viable intracellular bacteria dropped during infection and consequently, reduced relative amounts of isolated bacterial RNA combined with only a moderate sequencing depth limited the statistical power to call significantly differentially expressed genes for all but highly expressed bacterial genes. Second, we were analysing rather heterogeneous population of phagocytes composed of infected cells and cells which were either not infected or were infected but already cleared the pathogen. We accepted these limitations in order to develop an easy to control and reproducible model of infection. Indeed, obtained results clearly indicate that macrophages underwent substantial reprogramming of their transcriptome upon infection. Macrophages, like other immune cells, react to bacterial

Table 2. List of selected *B. pertussis* genes downregulated in the course of infection[a].

Gene ID	log_2FC T2/T1	log_2FC T3/T2	Name	Function
BP0015	−1.77	−1.68	*rpoB*	sigma factor RpoB
BP0016	−1.45	−2.10	*rpoC*	sigma factor RpoC
BP0762	−1.62	ND	*cyaD*	transport protein CyaD
BP0763	−1.65	ND	*cyaE*	transport protein CyaE
BP1054	−1.48	−1.18	*prn*	pertactin
BP1119	−2.46	ND	*fim2*	fimbriae Fim2
BP1201	*−0.75*	−3.27	*tcfA*	tracheal colonization factor TcfA
BP1251	−1.75	ND	*BP1251*	aerolysin
BP1876	ND	−1.72	*bvgR*	regulator of virulence BvgR
BP2241	−5.99	ND	*bscC*	type III secretion protein BscC
BP2244	−5.41	ND	*bscF*	type III secretion protein
BP2248	−5.35	ND	*bscI*	type III secretion protein
BP2256	−5.79	ND	*bscU*	putative secreted protein Bsp22
BP2262	−1.81	ND	*bopD*	type III secretion protein
BP2315	−1.39	ND	*vag8*	autotransporter Vag8
BP3282	−7.08	ND	*atpB*	ATP synthase subunit A
BP3494	−1.17	−1.44	*brkA*	autotransporter BrkA
BP3785	ND	−3.28	*ptxD*	pertussis toxin subunit 4 precursor
BP3786	ND	−3.28	*ptxE*	pertussis toxin subunit 5 precursor
BP3787	ND	−3.53	*ptxC*	pertussis toxin subunit 3 precursor
BP3790	ND	−2.71	*ptlC*	pertussis toxin transport protein PtlC
BP3791	ND	−1.85	*ptlD*	pertussis toxin transport protein PtlD

Values which did not pass the statistical significance but showed similar trend are shown in italics. ND, not determined in the respective analysis.

infection with common gene expression programmes [29,40]. One of the important roles of immune cells encountering the pathogen is to attract and activate other immune cells by secretion of cytokines and chemokines. Interestingly, we observed induction of both pro- and anti-inflammatory cytokines which is in line with previous reports demonstrating that TLR4-dependent recognition of *B. pertussis*-associated products triggers mixed immune response in the host [41,42]. For example, the highly expressed *IL-23, IL-1A, IL-6, CCR7, CCL2, CCL3, CXCL8, CD80* and *CD86* genes belong to set of genes characteristic for induction of M1 macrophage polarization and promoting pro-inflammatory Th1 and Th17 immune responses [43,44]. On the other side, increased expression of *IL-13* and *IL-10* cytokine genes inducing M2 programme in macrophages suggests that Th2 response is also mounted in infected macrophages. Notably, such a mixed immune response was also detected in mouse lungs after primary *B. pertussis* infection [34] and also in primary macrophages infected with several other pathogens [45–47]. Nevertheless, while the expression of *IL-23* remained highly increased in infected cells during the infection, expression of *IL-10* culminated at 2 h pi and then it gradually dropped. Importantly, these expression profiles of *IL23, IL10* and several other genes were clearly confirmed by quantitative PCR analysis. Of note, our results are in line with already documented observation that *B. pertussis*-infected monocyte-derived dendritic cells promote synthesis of IL-23 but not of IL-12 [48]. IL-23 is an IL-12-like heterodimer sharing the p40 subunit with IL-12 and was shown to play different roles in shaping of the immune response [49]. Besides its role in mounting of Th1 response, it is one of the essential factors required for the expansion of a memory T cell population, which is characterized by the production of IL-17 (Th17) [44]. Importantly, both Th1 and Th17 cells contribute to natural immunity induced by infection with *B. pertussis* in mice [50]. Also baboons, previously infected with *B. pertussis*, possess strong specific Th1 and Th17 memory [51]. Interestingly, IL-23/IL-17 axis plays a substantial role also in the development of chronic inflammations causing a variety of allergic and autoimmune diseases such as rheumatoid arthrosis, inflammatory bowel disease and asthma [52]. Based on environmental observations, Rubin and Glazer proposed a hypothesis that subclinical *B. pertussis* colonization is an important cause of asthma and diseases of allergic sensitization [53]. Therefore, it is tempting to speculate whether the paroxysmal cough, a hallmark of pertussis disease, could also result from strong pro-inflammatory IL-23 production and expansion of Th17 cells.

To prevent phagocytic killing, extracellular *B. pertussis* cells produce a very potent adenylate cyclase toxin which can disrupt cell signalling in macrophages and eventually induce apoptosis [24,25,54]. Nevertheless, our data suggest that upon internalization of *B. pertussis* cells, the expression of several

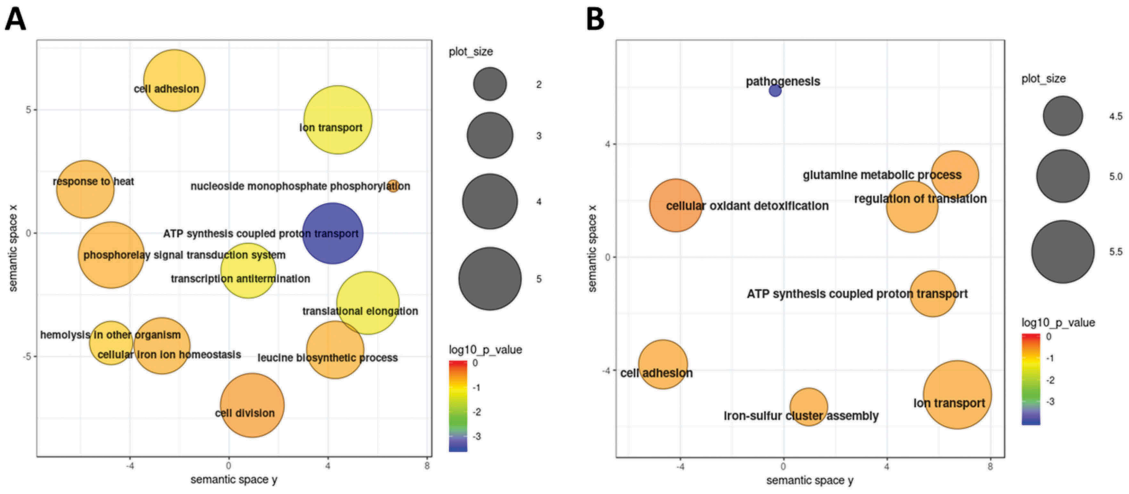


Figure 6. GO term enrichment analysis of *B. pertussis* genes modulated during infection. Significantly enriched terms from the domain 'Biological processes' were identified within gene sets either upregulated (A) or downregulated (B) between the time points T2 and T1. Results were summarized and visualized by REVIGO as a scatter plot. Circle size encodes number of genes associated with respective category, colours encode the significance level of the enrichment.

genes involved in apoptotic programme is modulated which may result in prevention of cell death and permitting the pathogen to persist within the immune cells. Interestingly, antiapoptotic signalling was observed also in human respiratory epithelial cells infected with *B. pertussis* and seems to be a molecular signature of *B. pertussis* infections [38]. Indeed, majority of serpins, a group of serine protease inhibitors which are known to control an array of biological processes including inflammation [55], was significantly upregulated upon infection. Among them, SERPINB3 and SERPINB4 were reported to inhibit apoptosis in THP-1 macrophages infected with *T. gondii* [56]. Likewise, the expression of *BIRC3* (*cIAP2*) gene was highly increased in infected cells. BIRC3 is a member of the inhibitor of apoptosis family of proteins that inhibit cell death by interfering with the activation of caspases [57] and, of note, its expression was found to be increased upon *B. pertussis* infection of human bronchial cells [38].

Remarkably, within the genes strongly induced upon infection we also identified set of factors which were reported to be hijacked by several intracellular pathogens to escape from host immunity. For example, suppressors of cytokine signalling (SOCSs) proteins SOCS3 and SOCS2 belonging to important family of inflammatory response regulators [58] were shown to augment the infection with *M. tuberculosis* and *S. enterica* [59,60]. Highly induced expression of *SOCS3* in infected macrophages is also in line with previous experiments, showing increased expression of *SOCS3* gene in *B. pertussis*-infected THP-1 macrophages [21]. Similarly, increased expression of *IDO1, LAMP3* and *MARCKS* genes in infected cells may correlate with attenuation of innate immune response and increased intracellular survival as it was shown for *Mycobacterium leprae* infection of monocytes [61], *S. typhimurium* infection of HeLa cells [62] and *Burkholderia thailandensis* infection of THP-1 macrophages and epithelial A549 cells [63], respectively.

GO term and KEGG pathway enrichment analyses were applied to identify molecular signatures of macrophage response to *B. pertussis* infection. Our results showed that several signalling and metabolic pathways were significantly modulated during the infection. Challenge with *B. pertussis* (6 h pi) led among others to downregulation of mTOR and insulin signalling pathways, fatty acid metabolism, TCA cycle and oxidative phosphorylation while NOD-like receptor and chemokine signalling pathways were induced. Of note, downregulation of insulin signalling pathway may correlate with reported *B. pertussis*-induced hyperinsulinemia in human and mice [64]. Interestingly, apoptosis, Toll-like receptor and Jak-STAT signalling pathways were concomitantly up- and downregulated. These data suggest that besides modulation of signalling pathways controlling the immune response, *B. pertussis* has also profound effect on host metabolism [65].

One of the important outcomes of our study was the finding that a significant number of bacterial reads can be obtained from infected macrophage cells without enrichment steps. Yet, we are aware that our data must be interpreted with caution and corroborated in larger study yielding substantially increased number of mapped reads. Nevertheless, in this study, we could identify several *Bordetella* genes which were differentially expressed between 2 and 6 h pi and/or between 6 and 24 h pi and validate observed effects for selected genes by quantitative PCR. Among those induced during infection were genes involved in iron transport and iron-sulphur cluster assembly indicating that intracellular *B. pertussis* cells face iron scarcity and oxidative stress. These results are in line with previous reports showing that iron acquisition and transport systems and iron-sulphur cluster sensors contribute to fitness of *B. pertussis* cells during *in vivo* infections [35,36,66,67]. Within the genes displaying reduced expression during infection were also those involved in pathogenesis. The expression of *bvgR* regulator of virulence was reduced during the infection and, consequently, the expression of several virulence factors was also downregulated. This is in contrast with *in vivo* data obtained previously in mouse model of respiratory tract infection where expression of majority of virulence factors was increased during the infection of lungs [35,36,68]. Our observations indicate that opposed to *B. pertussis* cells isolated from mouse organs and facing systemic immune response, the intracellular *B. pertussis* cells switch from Bvg^+ to Bvg^- phase during the intramacrophage phase. We assume that the Bvg^+ phase plays a less significant role in the intracellular survival of *B. pertussis* and that avirulent Bvg^- phase can be beneficial to the pathogen as it was shown for closely related *B. bronchiseptica* [69,70]. Interestingly, Bvg^- mutants have been shown to accumulate among persistent *B. pertussis* cells within the upper respiratory tract of experimentally infected rhesus monkeys [71]. Furthermore, multiple metabolic pathways are upregulated in the Bvg^- phase and potentially help to overcome the environmental and nutritional stress [72]. Recently, we have shown that Bvg-repressed small regulatory RNA RgtA is involved in control of transport of glutamate, a key metabolite in *B. pertussis* physiology [73]. In line with these observations, the expression of virulence repressed gene *vrg6* was induced during the infection. Of note, increased expression of *vrg6* gene was observed also in *B. pertussis* cells recovered from lungs of intranasally infected mice [74]. While this gene was one of the first described virulence repressed genes in *B. pertussis* [75], its cellular function remained unclear. We suggest that Vrg6 protein belongs to family of PXP(V) proteins forming potassium channels [76,77] as it contains five PXP(V) motifs and two transmembrane domains (Fig. S4). These proteins control the activity of the pores in response to different stimuli such as changes in pH or membrane potential [78]. Considering the possible effects of harsh macrophage environment on bacterial cell wall integrity, the increased expression of the *vrg6* gene seems to be biologically relevant.

Collectively, observed changes in the gene expression profiles suggest that intracellular *B. pertussis* cells actively adapt to intramacrophage environment and respond to bactericidal activities triggered by THP-1 cells. As mentioned above, large number of host genes involved in macrophage polarization, signalling, immunosuppression and apoptosis were overexpressed upon infection. It seems that interplay between macrophages and intracellular *B. pertussis* cells may result in alternative polarization of macrophages and induction of the processes that augment survival of the bacteria within

phagocytic cells. Nevertheless, strongly modulated expression of dozens of cytokines and chemokines suggests that some of the observed effects result from cell-to-cell signalling within the population of infected macrophages. Importantly, this study indicates that *in vitro* expressions profiles of several genes and pathways in both human host cells and pathogen are in line with *in vivo* data obtained with mouse model of infection [34–36,66], although our data suggest higher importance of avirulent phase in intramacrophage infection than previously thought. Nevertheless, we are aware of certain limitations of our study and it is evident that we need to perform additional experiments to delineate the *B. pertussis*-specific activities modulating the host immune response and macrophage-specific responses induced by *B. pertussis* infection in order to clarify the importance of the intracellular phase for the infectious cycle of *B. pertussis*.

Materials and methods

Growth and differentiation of THP-1 cells

Monocytic THP-1 cells (obtained from ATCC) were grown in suspension in 75 cm^2 flasks in Roswell Park Memorial Institute (RPMI) 1640 medium (Sigma) supplemented with 10% heat-inactivated foetal bovine serum (FBS, Sigma) at 37°C in a humidified incubator (5% CO_2). For differentiation into macrophages, the THP-1 cells were seeded in 6-well plates (4 x 10^6 per well) and stimulated by the addition of 10 nM phorbol 12-myristate 13-acetate (PMA, Sigma) for 72 h followed by resting for 24 h in RPMI medium without PMA.

Cultivation of the *B. pertussis* cells

Bordetella pertussis reference strain Tohama I was grown on Bordet-Gengou (BG) agar plates (Difco) supplemented with 15% sheep blood for 3 to 4 days at 37°C. For liquid cultures, bacteria were grown in Stainer-Scholte (SS) medium supplemented with 0.1% cyclodextrin (Sigma) and 0.5% casamino acids (Difco) at 37°C. For infection assays, the *B. pertussis* cells were grown overnight in SS medium to mid-exponential phase of growth (OD_{600} ≈ 1.0) and diluted in RPMI medium to appropriate cell density.

Infection of THP-1-derived macrophages

Differentiated THP-1 macrophages were infected with *B. pertussis* cells diluted in RPMI medium at multiplicity of infection of 100 bacteria per macrophage (4 x 10^8 per well). The plates were then centrifuged for 3 min at 600 g to facilitate the interaction of bacteria with macrophages. After 1 h of incubation (37°C; 5% CO_2), nonadherent bacteria were removed by washing and remaining extracellular bacteria were killed by incubation in RPMI medium containing 100 µg/ml of polymyxin B sulphate (Sigma) for 1 h (37°C; 5% CO_2). No viable bacteria were detected by direct plating of cell culture supernatants on BG agar. From this time point (T1), corresponding to 2 hours pi (1 h of co-incubation of *B. pertussis* cells with macrophages and 1 h of polymyxin B treatment), the concentration of the antibiotic was kept at 20 µg/ml which was sufficient to prevent the repopulation of RPMI medium with released intracellular bacteria (data not shown). At indicated times (2, 6 and 24 h pi; time points T1-T3), the infected cells were intensely washed with prewarmed phosphate-buffered saline (PBS) and directly lysed. In parallel, samples of uninfected cells were treated in the same way and served as controls. In addition, at each time point, the infected macrophages from separate well were lysed with sterile water containing 0.01% Triton X-100 and the lysates were plated onto BG agars to assess the numbers of viable intracellular bacteria. At each time point, three biological replicates of infected and uninfected macrophages as well as of unexposed bacteria cultivated in RPMI medium were harvested for transcriptomic and quantitative PCR analyses.

RNA isolation

Macrophages were lysed with 0.5 ml of TRI reagent (Sigma) and total RNA was isolated from lysates according to the manufacturer´s protocol. Removal of DNA was achieved by treatment of samples with TURBO DNA-free kit (Thermo Fisher Scientific). RNA quality and quantity were determined by agarose gel electrophoresis and using the Nanodrop 2000 device (Thermo Fisher Scientific). Furthermore, the RNA quality was assessed at sequencing facility (Vienna Biocenter Core Facility, NGS unit) on an Agilent 2100 Bioanalyzer. All samples displayed RNA integrity numbers higher than 9.

Library preparation, sequencing and read mapping

Ribosomal RNAs of the host and pathogen were simultaneously depleted with the Ribo-Zero Gold rRNA Removal Kit from Illumina. Libraries were prepared using NEBNext® Ultra™ II DNA Library Prep Kit for Illumina and sequenced on an Illumina HiSeq 2500 platform using HiSeqV4 chemistry with single-end 50-base-pair reads. FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc; version 0.11.8) was used to perform quality control checks on raw Illumina reads in FASTQ format. A single transcriptome generated by combining human transcriptome (Ensembl GRCh38) and *B. pertussis* Tohama I transcriptome was used to map and quantify the reads using Salmon quantifier (version 0.11.3) [79]. *B. pertussis* transcriptome was produced from NCBI Assembly GCF 000195715.1 using gffread (https://github.com/gpertea/gffread). After quantification with Salmon, human and bacterial read counts per gene were separated for subsequent differential expression analysis. RNA-seq data from the sequencing runs were deposited at the European Nucleotide Archive under project accession number PRJEB33395.

Differential expression (DE) analysis

Prior to DE analysis, unwanted variation caused by batch effects or library preparation was removed from the samples using RUVs correction method from RUVseq (version 1.16.0) [80]. DE was assessed using gene read counts as inputs and analysis was performed by EdgeR package (version 3.24.1) [81]. Human genes with at least five uniquely

mapped reads in at least two replicates were considered with default k = 1 using the upper-quartile normalization. Bacterial genes with at least 10 uniquely mapped reads (average of the triplicate) at least at one time point were considered with default k = 1, using a weighted trimmed mean of M-values (TMM) normalization. DE analysis for samples collected at time point T2 versus control was evaluated only for human genes, DE analysis of samples collected at time points T2 versus T1 and T3 versus T2 was determined for both human and *B. pertussis* genes.

MSF analysis

The DE analysis results generated by edgeR were used to identify modulated sub-graphs using Modulated Sub-graph Finder (MSF). Cell signalling interactions were filtered from Reactome functional interactions version 2016 (https://reactome.org/download-data) for only direct interactions, which was used as a second input for MSF tool [82]. The parameters used for MSF were extension 2 and merging 2. The network file containing the identified interactions produced by MSF was imported and visualized by Cytoscape (version 3.6.1). KEGG pathway enrichment was generated using StringApp (version 1.4.2).

GSEA analysis

The Gene Set Enrichment Analysis was done for GO and KEGG term enrichment using R package 'GSEABase' (version 1.44.0). For each comparison, the complete list of significant DE genes (adjusted p-value < 0.05) was separated into two groups, one composed of genes with positive log fold change ($log_2FC > 0$) and the second composed of genes with negative fold change ($log_2FC < 0$). Top 20 enriched 'Upregulated' and 'Downregulated' gene sets were selected for the visualization.

GO term enrichment for *B. pertussis*

The gene ontology (GO) terms per gene were deduced using blast2go [83]. Each GO term which was associated with more than one gene in the gene set was tested for enrichment compared to the whole transcriptome by applying the Fisher Exact test. Afterwards, determined p-values were corrected for multiple testing by the method of Benjamini and Hochberg [84]. The GO term lists were condensed, summarized and visualized using the Revigo tool [85].

Quantitative PCR

To validate gene expression profiles of infected THP-1 macrophages and intracellular *B. pertussis* cells, total RNA isolated for RNA-seq analysis from triplicate samples of infected macrophages 2, 6 and 24 h pi was used. As controls, total RNA was isolated from triplicates of uninfected macrophages and unexposed bacteria cultivated in parallel in RPMI medium. RT-qPCR analysis of both host and pathogen transcript levels was performed as described earlier [86]. Briefly, duplicates of 1 µg of total isolated RNA treated with TURBO DNA-free kit (Thermo Fisher Scientific) were reverse transcribed into cDNA following the manufacturer's instructions in a 25-µl reaction using the MMLV Reverse Transcriptase (Promega). Quantitative PCR (RT-qPCR) was performed on Bio-Rad CFX384 Touch™ instrument using HOT FIREPol® Eva®Green qPCR Mix Plus™ (Solis Biodyne), 200 nmoles of each primer and 40 ng of reverse transcribed RNA in a 12-µl qPCR reaction with an initial step at 95°C for 12 min, followed by 40 cycles of 95°C for 15 s, 60°C for 25 s, 72°C for 30 s. For THP-1 macrophages, the *ACTB* gene was used as the reference gene to assess relative gene expression levels in infected and uninfected cells. For *B. pertussis*, the *rbfA* gene was used as the reference gene to assay the relative gene expression levels in intracellular and unexposed bacteria. Primers used to quantify THP-1 and *B. pertussis* genes (Table S5) were analysed for secondary structures and for cross-dimers in primer pairs. Relative expression levels were quantified using amplification efficiency values [87]. To validate RNA-seq data, relative expression levels of selected genes were compared between infected and uninfected THP-1 cells and between intracellular and unexposed bacteria at each time point.

Disclosure of potential conflicts of interest

No potential conflict of interest was reported by the authors.

Funding

This research was funded by the Austrian FWF project I 2353-B22 (I.H.), by the grants [16-34825L] and [19-12338S] from Czech Science Foundation (www.gacr.cz) (B.V.), by the grant [16-30782A] from Czech Health Research Council (www.azvcr.cz/) (B.V.) by project International mobility of researchers of the Institute of Microbiology of the CAS, v. v. i., (CZ.02.2.69/0.0/0.0/16_027/0007990) (DK) and by funding from RVO61388971.

ORCID

Mariam R. Farman http://orcid.org/0000-0002-4419-8421
Fabian Amman http://orcid.org/0000-0002-8646-859X
Jakub Držmíšek http://orcid.org/0000-0002-8585-3189
Ivo Hofacker http://orcid.org/0000-0001-7132-0800
Branislav Večerek http://orcid.org/0000-0001-9520-6324

References

[1] Mattoo S, Cherry JD. Molecular pathogenesis, epidemiology, and clinical manifestations of respiratory infections due to Bordetella pertussis and other Bordetella subspecies. Clin Microbiol Rev. 2005;18:326–382.
[2] WHO. Vaccine preventable deaths and the global immunization vision and strategy, 2006–2015. MMWR Morb Mortal Wkly Rep. 2006;55:511–515.
[3] Cherry JD. The present and future control of pertussis. Clin Infect Dis. 2010;51:663–667.
[4] Raguckas SE, VandenBussche HL, Jacobs C, et al. Pertussis resurgence: diagnosis, treatment, prevention, and beyond. Pharmacotherapy. 2007;27:41–52.
[5] Consortium P. PERISCOPE: road towards effective control of pertussis. Lancet Infect Dis. 2018. DOI:10.1016/S1473-3099(18)30646-7

[6] Melvin JA, Scheller EV, Miller JF, et al. Bordetella pertussis pathogenesis: current and future challenges. Nat Rev Microbiol. 2014;12:274–288.
[7] Higgs R, Higgins SC, Ross PJ, et al. Immunity to the respiratory pathogen Bordetella pertussis. Mucosal Immunol. 2012;5:485–500.
[8] Locht C. Molecular aspects of Bordetella pertussis pathogenesis. Int Microbiol. 1999;2:137–144.
[9] de Gouw D, Diavatopoulos DA, Bootsma HJ, et al. Pertussis: a matter of immune modulation. FEMS Microbiol Rev. 2011;35:441–474.
[10] Lenz DH, Weingart CL, Weiss AA. Phagocytosed Bordetella pertussis fails to survive in human neutrophils. Infect Immun. 2000;68:956–959.
[11] Ishibashi Y, Claus S, Relman DA. Bordetella pertussis filamentous hemagglutinin interacts with a leukocyte signal transduction complex and stimulates bacterial adherence to monocyte CR3 (CD11b/CD18). J Exp Med. 1994;180:1225–1233.
[12] Saukkonen K, Cabellos C, Burroughs M, et al. Integrin-mediated localization of Bordetella pertussis within macrophages: role in pulmonary colonization. J Exp Med. 1991;173:1143–1149.
[13] Hazenbos WL, van den Berg BM, van Furth R. Very late antigen-5 and complement receptor type 3 cooperatively mediate the interaction between Bordetella pertussis and human monocytes. J Immunol. 1993;151:6274–6282.
[14] Rodriguez ME, Hellwig SM, Hozbor DF, et al. Fc receptor-mediated immunity against Bordetella pertussis. J Immunol. 2001;167:6545–6551.
[15] Friedman RL, Nordensson K, Wilson L, et al. Uptake and intracellular survival of Bordetella pertussis in human macrophages. Infect Immun. 1992;60:4578–4585.
[16] Paddock CD, Sanden GN, Cherry JD, et al. Pathology and pathogenesis of fatal Bordetella pertussis infection in infants. Clin Infect Dis. 2008;47:328–338.
[17] Hellwig SM, Hazenbos WL, van de Winkel JG, et al. Evidence for an intracellular niche for Bordetella pertussis in broncho-alveolar lavage cells of mice. FEMS Immunol Med Microbiol. 1999;26:203–207.
[18] Vandebriel RJ, Hellwig SM, Vermeulen JP, et al. Association of Bordetella pertussis with host immune cells in the mouse lung. Microb Pathog. 2003;35:19–29.
[19] Lamberti YA, Hayes JA, Perez Vidakovics ML, et al. Intracellular trafficking of Bordetella pertussis in human macrophages. Infect Immun. 2010;78:907–913.
[20] Lamberti Y, Cafiero JH, Surmann K, et al. Proteome analysis of Bordetella pertussis isolated from human macrophages. J Proteomics. 2016;136:55–67.
[21] Valdez HA, Oviedo JM, Gorgojo JP, et al. Bordetella pertussis modulates human macrophage defense gene expression. Pathog Dis. 2016;74. DOI:10.1093/femspd/ftw073.
[22] Carbonetti NH, Artamonova GV, Van Rooijen N, et al. Pertussis toxin targets airway macrophages to promote Bordetella pertussis infection of the respiratory tract. Infect Immun. 2007;75:1713–1720.
[23] Kamanova J, Kofronova O, Masin J, et al. Adenylate cyclase toxin subverts phagocyte function by RhoA inhibition and unproductive ruffling. J Immunol. 2008;181:5587–5597.
[24] Cerny O, Kamanova J, Masin J, et al. Bordetella pertussis adenylate cyclase toxin blocks induction of bactericidal nitric oxide in macrophages through cAMP-dependent activation of the SHP-1 phosphatase. J Immunol. 2015;194:4901–4913.
[25] Gueirard P, Druilhe A, Pretolani M, et al. Role of adenylate cyclase-hemolysin in alveolar macrophage apoptosis during Bordetella pertussis infection in vivo. Infect Immun. 1998;66:1718–1725.
[26] Aulicino A, Dinan AM, Miranda-CasoLuengo AA, et al. High-throughput transcriptomics reveals common and strain-specific responses of human macrophages to infection with mycobacterium abscessus smooth and rough variants. BMC Genomics. 2015;16:1046.
[27] Mizuno K, Fukami T, Toyoda Y, et al. Terbinafine stimulates the pro-inflammatory responses in human monocytic THP-1 cells through an ERK signaling pathway. Life Sci. 2010;87:537–544.
[28] Huang H, Fletcher A, Niu Y, et al. Characterization of lipopolysaccharide-stimulated cytokine expression in macrophages and monocytes. Inflamm Res. 2012;61:1329–1338.
[29] Jenner RG, Young RA. Insights into host responses against pathogens from transcriptional profiling. Nat Rev Microbiol. 2005;3:281–294.
[30] Idzko M, Ferrari D, Eltzschig HK. Nucleotide signalling during inflammation. Nature. 2014;509:310–317.
[31] Martinez VG, Rubio C, Martinez-Fernandez M, et al. BMP4 induces M2 macrophage polarization and favors tumor progression in bladder cancer. Clin Cancer Res. 2017;23:7388–7399.
[32] Belkina AC, Nikolajczyk BS, Denis GV. BET protein function is required for inflammation: brd2 genetic disruption and BET inhibitor JQ1 impair mouse macrophage inflammatory responses. J Immunol. 2013;190:3670–3678.
[33] Banus S, Pennings J, Vandebriel R, et al. Lung response to Bordetella pertussis infection in mice identified by gene-expression profiling. Immunogenetics. 2007;59:555–564.
[34] Raeven RH, Brummelman J, Pennings JL, et al. Molecular signatures of the evolving immune response in mice following a Bordetella pertussis infection. PLoS One. 2014;9:e104548.
[35] van Beek LF, de Gouw D, Eleveld MJ, et al. Adaptation of Bordetella pertussis to the respiratory tract. J Infect Dis. 2018;217:1987–1996.
[36] Wong TY, Hall JM, Nowak ES, et al. Analysis of the in vivo transcriptome of Bordetella pertussis during infection of mice. mSphere. 2019;4. DOI:10.1128/mSphereDirect.00154-19.
[37] 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.
[38] Belcher CE, Drenkow J, Kehoe B, et al. The transcriptional responses of respiratory epithelial cells to Bordetella pertussis reveal host defensive and pathogen counter-defensive strategies. Proc Natl Acad Sci U S A. 2000;97:13847–13852.
[39] Westermann AJ, Forstner KU, Amman F, et al. Dual RNA-seq unveils noncoding RNA functions in host-pathogen interactions. Nature. 2016;529:496–501.
[40] Boldrick JC, Alizadeh AA, Diehn M, et al. Stereotyped and specific gene expression programs in human innate immune responses to bacteria. Proc Natl Acad Sci U S A. 2002;99:972–977.
[41] Ross PJ, Lavelle EC, Mills KH, et al. Adenylate cyclase toxin from Bordetella pertussis synergizes with lipopolysaccharide to promote innate interleukin-10 production and enhances the induction of Th2 and regulatory T cells. Infect Immun. 2004;72:1568–1579.
[42] Higgins SC, Jarnicki AG, Lavelle EC, et al. TLR4 mediates vaccine-induced protective cellular immunity to Bordetella pertussis: role of IL-17-producing T cells. J Immunol. 2006;177:7980–7989.
[43] Benoit M, Desnues B, Mege JL. Macrophage polarization in bacterial infections. J Immunol. 2008;181:3733–3739.
[44] Aggarwal S, Ghilardi N, Xie MH, et al. Interleukin-23 promotes a distinct CD4 T cell activation state characterized by the production of interleukin-17. J Biol Chem. 2003;278:1910–1914.
[45] Li W, Katz BP, Spinola SM. Haemophilus ducreyi-induced interleukin-10 promotes a mixed M1 and M2 activation program in human macrophages. Infect Immun. 2012;80:4426–4434.
[46] Benoit M, Barbarat B, Bernard A, et al. Coxiella burnetii, the agent of Q fever, stimulates an atypical M2 activation program in human macrophages. Eur J Immunol. 2008;38:1065–1070.
[47] Smith MW, Schmidt JE, Rehg JE, et al. Induction of pro- and anti-inflammatory molecules in a mouse model of pneumococcal pneumonia after influenza. Comp Med. 2007;57:82–89.
[48] Fedele G, Stefanelli P, Spensieri F, et al. Bordetella pertussis-infected human monocyte-derived dendritic cells undergo maturation and induce Th1 polarization and interleukin-23 expression. Infect Immun. 2005;73:1590–1597.

[49] Bastos KR, Marinho CR, Barboza R, et al. What kind of message does IL-12/IL-23 bring to macrophages and dendritic cells?. Microbes Infect. 2004;6:630–636. DOI:10.1016/j.micinf.2004.02.012
[50] Ross PJ, Sutton CE, Higgins S, et al. Relative contribution of Th1 and Th17 cells in adaptive immunity to Bordetella pertussis: towards the rational design of an improved acellular pertussis vaccine. PLoS Pathog. 2013;9:e1003264.
[51] Warfel JM, Zimmerman LI, Merkel TJ. Acellular pertussis vaccines protect against disease but fail to prevent infection and transmission in a nonhuman primate model. Proc Natl Acad Sci U S A. 2014;111:787–792.
[52] Iwakura Y, Ishigame H. The IL-23/IL-17 axis in inflammation. J Clin Invest. 2006;116:1218–1222.
[53] Rubin K, Glazer S. The pertussis hypothesis: bordetella pertussis colonization in the etiology of asthma and diseases of allergic sensitization. Med Hypotheses. 2018;120:101–115.
[54] Khelef N, Zychlinsky A, Guiso N. Bordetella pertussis induces apoptosis in macrophages: role of adenylate cyclase-hemolysin. Infect Immun. 1993;61:4064–4071.
[55] Law RH, Zhang Q, McGowan S, et al. An overview of the serpin superfamily. Genome Biol. 2006;7:216.
[56] Song KJ, Ahn HJ, Nam HW. Anti-apoptotic effects of SERPIN B3 and B4 via STAT6 activation in macrophages after infection with Toxoplasma gondii. Korean J Parasitol. 2012;50:1–6.
[57] Roy N, Deveraux QL, Takahashi R, et al. The c-IAP-1 and c-IAP-2 proteins are direct inhibitors of specific caspases. Embo J. 1997;16:6914–6925.
[58] Delgado-Ortega M, Marc D, Dupont J, et al. SOCS proteins in infectious diseases of mammals. Vet Immunol Immunopathol. 2013;151:1–19.
[59] Schmok E, Abad Dar M, Behrends J, et al. Suppressor of cytokine signaling 3 in macrophages prevents exacerbated interleukin-6-dependent arginase-1 activity and early permissiveness to experimental tuberculosis. Front Immunol. 2017;8:1537.
[60] Uchiya K, Nikai T. Salmonella pathogenicity island 2-dependent expression of suppressor of cytokine signaling 3 in macrophages. Infect Immun. 2005;73:5587–5594.
[61] de Souza Sales J, Lara FA, Amadeu TP, et al. The role of indoleamine 2, 3-dioxygenase in lepromatous leprosy immunosuppression. Clin Exp Immunol. 2011;165:251–263.
[62] Lee EJ, Park KS, Jeon IS, et al. LAMP-3 (lysosome-associated membrane protein 3) promotes the intracellular proliferation of salmonella typhimurium. Mol Cells. 2016;39:566–572.
[63] Micheva-Viteva SN, Shou Y, Ganguly K, et al. PKC-eta-MARCKS signaling promotes intracellular survival of unopsonized Burkholderia thailandensis. Front Cell Infect Microbiol. 2017;7:231.
[64] Furman BL, Sidey FM, Wardlaw AC. Role of insulin in the hypoglycaemic effect of sublethal Bordetella pertussis infection in mice. Br J Exp Pathol. 1986;67:305–312.
[65] Freyberg Z, Harvill ET. Pathogen manipulation of host metabolism: A common strategy for immune evasion. PLoS Pathog. 2017;13:e1006669.
[66] Brickman TJ, Vanderpool CK, Armstrong SK. Heme transport contributes to in vivo fitness of Bordetella pertussis during primary infection in mice. Infect Immun. 2006;74:1741–1744.
[67] Miller HK, Auerbuch V. Bacterial iron-sulfur cluster sensors in mammalian pathogens. Metallomics. 2015;7:943–956.
[68] Bibova I, Hot D, Keidel K, et al. Transcriptional profiling of Bordetella pertussis reveals requirement of RNA chaperone Hfq for Type III secretion system functionality. RNA Biol. 2015;12:175–185.
[69] Jungnitz H, West NP, Walker MJ, et al. A second two-component regulatory system of Bordetella bronchiseptica required for bacterial resistance to oxidative stress, production of acid phosphatase, and in vivo persistence. Infect Immun. 1998;66:4640–4650.
[70] Zimna K, Medina E, Jungnitz H, et al. Role played by the response regulator Ris in Bordetella bronchiseptica resistance to macrophage killing. FEMS Microbiol Lett. 2001;201:177–180.
[71] Karataev GI, Sinyashina LN, Medkova AY, et al. [Insertional inactivation of virulence operon in population of persistent Bordetella pertussis bacteria]. Genetika. 2016;52:422–430.
[72] Moon K, Bonocora RP, Kim DD, et al. The BvgAS regulon of Bordetella pertussis. MBio. 2017;8. DOI:10.1128/mBio.01526-17.
[73] Keidel K, Amman F, Bibova I, et al. Signal transduction-dependent small regulatory RNA is involved in glutamate metabolism of the human pathogen Bordetella pertussis. RNA. 2018;24:1530–1541.
[74] Hoo R, Lam JH, Huot L, et al. Evidence for a role of the polysaccharide capsule transport proteins in pertussis pathogenesis. PLoS One. 2014;9:e115243.
[75] Croinin TO, Grippe VK, Merkel TJ. Activation of the vrg6 promoter of Bordetella pertussis by RisA. J Bacteriol. 2005;187:1648–1658.
[76] Del Camino D, Holmgren M, Liu Y, et al. Blocker protection in the pore of a voltage-gated K+ channel and its structural implications. Nature. 2000;403:321–325.
[77] Tieleman DP, Shrivastava IH, Ulmschneider MR, et al. Proline-induced hinges in transmembrane helices: possible roles in ion channel gating. Proteins. 2001;44:63–72.
[78] Labro AJ, Snyders DJ. Being flexible: the voltage-controllable activation gate of kv channels. Front Pharmacol. 2012;3:168.
[79] Patro R, Duggal G, Love MI, et al. Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods. 2017;14:417–419.
[80] Risso D, Ngai J, Speed TP, et al. Normalization of RNA-seq data using factor analysis of control genes or samples. Nat Biotechnol. 2014;32:896–902.
[81] Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26:139–140.
[82] Farman MR, Hofacker IL, Amman F. MSF: modulated sub-graph finder. F1000Res. 2018;7:1346.
[83] Conesa A, Gotz S, Garcia-Gomez JM, et al. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics. 2005;21:3674–3676.
[84] Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B. 1995;57:289–300.
[85] Supek F, Bosnjak M, Skunca N, et al. REVIGO summarizes and visualizes long lists of gene ontology terms. PLoS One. 2011;6:e21800.
[86] Bibova I, Skopova K, Masin J, et al. The RNA chaperone Hfq is required for virulence of Bordetella pertussis. Infect Immun. 2013;81:4081–4090.
[87] Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res. 2001;29:e45.

Publication 3

"Avirulent phenotype promotes Bordetella pertussis adaptation to the intramacrophage environment"

Mariam R. Farman, **Denisa Petráčková**, Dilip Kumar, Jakub Držmíšek, Argha Saha, Ivana Čurnová, Jan Čapek, Václava Hejnarová, Fabian Amman, Ivo Hofacker and Branislav Večerek

Status: published in Emerging Microbes & Infections, 2023
https://doi.org/10.1080/22221751.2022.2146536

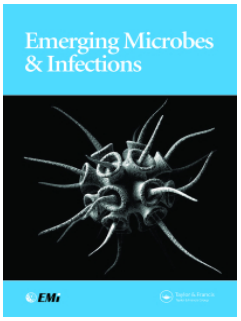
Author Contributions:

MARIAM R. FARMAN - Provided significant contributions to the planning and structuring of the experimental setup, including determination of time-points, number of replicates and sequencing depth. Developed the analytical pipeline, conducted comprehensive bioinformatics analyses, interpreted the results, and authored the manuscript.

IVO HOFACKER - Supervised the project and contributed to the manuscript.

FABIAN AMMAN - Supervised the project and contributed to the manuscript.

COLLABORATORS - Conceived, designed, and performed the experiments.



Emerging Microbes & Infections

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/temi20

Avirulent phenotype promotes *Bordetella pertussis* adaptation to the intramacrophage environment

Mariam R. Farman, Denisa Petráčková, Dilip Kumar, Jakub Držmíšek, Argha Saha, Ivana Čurnová, Jan Čapek, Václava Hejnarová, Fabian Amman, Ivo Hofacker & Branislav Večerek

To cite this article: Mariam R. Farman, Denisa Petráčková, Dilip Kumar, Jakub Držmíšek, Argha Saha, Ivana Čurnová, Jan Čapek, Václava Hejnarová, Fabian Amman, Ivo Hofacker & Branislav Večerek (2023) Avirulent phenotype promotes *Bordetella pertussis* adaptation to the intramacrophage environment, Emerging Microbes & Infections, 12:1, e2146536, DOI: 10.1080/22221751.2022.2146536

To link to this article: https://doi.org/10.1080/22221751.2022.2146536

Emerging Microbes & Infections
2023, VOL. 12, e2146536 (16 pages)
https://doi.org/10.1080/22221751.2022.2146536

RESEARCH ARTICLE

OPEN ACCESS Check for updates

Avirulent phenotype promotes *Bordetella pertussis* adaptation to the intramacrophage environment

Mariam R. Farman [a*†], Denisa Petráčková [b*], Dilip Kumar [b], Jakub Držmíšek [b], Argha Saha [b], Ivana Čurnová [b], Jan Čapek [b], Václava Hejnarová [b], Fabian Amman [a‡], Ivo Hofacker [a] and Branislav Večerek [b]

[a]Institute for Theoretical Chemistry, University of Vienna, Vienna, Austria; [b]Czech Academy of Sciences, Laboratory of Post-transcriptional Control of Gene Expression, Institute of Microbiology, Prague, Czech Republic

ABSTRACT

Bordetella pertussis, the causative agent of whooping cough, is an extracellular, strictly human pathogen. However, it has been shown that *B. pertussis* cells can escape phagocytic killing and survive in macrophages upon internalization. Our time-resolved RNA-seq data suggest that *B. pertussis* efficiently adapts to the intramacrophage environment and responds to host bactericidal activities. We show that this adaptive response is multifaceted and, surprisingly, related to the BvgAS two-component system, a master regulator of virulence. Our results show that the expression of this regulatory circuit is downregulated upon internalization. Moreover, we demonstrate that the switch to the avirulent Bvg^- phase augments a very complex process based on the adjustment of central and energy metabolism, cell wall reinforcement, maintenance of appropriate redox and metal homeostasis, and repair of damaged macromolecules. Nevertheless, not all observed effects could be simply attributed to the transition to Bvg^- phase, suggesting that additional regulators are involved in the adaptation to the intramacrophage environment. Interestingly, a large number of genes required for the metabolism of sulphur were strongly modulated within macrophages. In particular, the mutant lacking two genes encoding cysteine dioxygenases displayed strongly attenuated cytotoxicity toward THP-1 cells. Collectively, our results suggest that intracellular *B. pertussis* cells have adopted the Bvg^- mode to acclimate to the intramacrophage environment and respond to antimicrobial activities elicited by THP-1 cells. Therefore, we hypothesize that the avirulent phase represents an authentic phenotype of internalized *B. pertussis* cells.

ARTICLE HISTORY Received 10 July 2022; Revised 12 October 2022; Accepted 8 November 2022

KEYWORDS *Bordetella pertussis*; BvgAS; intramacrophage environment; adaptation to stress; cysteine toxicity; avirulent phase

Introduction

Bordetella pertussis is a strictly human reemergent pathogen of the respiratory tract and the etiologic agent of whooping cough [1]. This highly contagious disease is particularly severe in infants and remains a major cause of infant mortality worldwide, especially in developing countries [2].

B. pertussis has been traditionally described and viewed as an extracellular pathogen that colonizes the epithelium of the upper respiratory tract (recently reviewed in [3]). While ciliated epithelial cells secrete mucus that mechanically eliminates the bacteria, both resident and attracted phagocytic cells interact with the pathogen to clear the infection [4]. Successful infection, therefore, depends on the activities of a variety of virulence factors produced by *B. pertussis* cells [3]. In particular, adhesins such as filamentous haemagglutinin and fimbriae contribute to efficient adhesion and colonization of the epithelium, while toxins such as pertussis toxin and adenylate cyclase toxin disrupt the epithelium and modulate the activity of the immune system, thereby facilitating the survival of bacteria in the respiratory tract [3,5,6]. The expression of most virulence factors is controlled by the two-component system BvgAS, which consists of the sensor kinase BvgS and the response regulator BvgA, which upon phosphorylation with BvgS promotes expression of virulence-activated genes [7]. Basically, three different phenotypes can be distinguished based on the BvgA phosphorylation status: Bvg^+ (virulent), Bvg^- (avirulent) and Bvg^i (intermediate). However, the exact role of the avirulent phase in the infection cycle of *B. pertussis* remains enigmatic [8-10].

In the presence of specific antibodies, opsonized *B. pertussis* cells are internalized and killed by phagocytic cells [11,12]. However, some early reports suggested that in the absence of opsonizing antibodies, this pathogen stimulates its own attachment to immune cells, resulting in inefficient killing by

CONTACT Branislav Večerek vecerek@biomed.cas.cz Czech Academy of Sciences, Laboratory of Post-transcriptional Control of Gene Expression, Institute of Microbiology, 14220 Prague, Czech Republic
*These authors contributed equally.
†Present address: Johannes Kepler University Linz, Department of Paediatrics and Adolescent Medicine, Krankenhausstrasse 26-30, 4020 Linz, Austria
‡Present address: CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, Lazarettgasse 14, 1090 Vienna, Austria
Supplemental data for this article can be accessed online at https://doi.org/10.1080/22221751.2022.2146536.

phagocytes and increased intracellular survival [12-16]. Importantly, *B. pertussis* cells have also been found *in vivo* in human, rabbit and mouse alveolar macrophages [17-19]. In the last decade, Rodriguez and her team have presented data showing that *B. pertussis* can survive in primary macrophages [20] as well as in monocyte-derived THP-1 macrophages [21,22]. They have shown that most *B. pertussis* cells are killed within acidic compartments, but some residual cells escape this killing and proliferate in a nonacidic, early endosomal compartment. These data led to speculation that *B. pertussis* may utilize macrophages as an intracellular niche and that the intramacrophage phase of infection may play a significant role in the *B. pertussis* infection cycle [14,18,19]. In support of this hypothesis, gene expression profiling and proteomic analysis of infected monocyte-derived THP-1 macrophages revealed that intracellular *B. pertussis* cells modulate the expression of several genes involved in the host immune response and reshape the production of dozens of proteins implicated in virulence [21-23]. Furthermore, Harvill and colleagues have demonstrated that the closely related pathogens *B. bronchiseptica* and *B. parapertussis* survive intracellularly in amoebae [24] and primary macrophages [25], respectively. These observations suggest that the ability for intracellular survival has an ancestral origin and is widespread among pathogenic *Bordetella* species [26,27].

To date, transcriptional profiling of *B. pertussis* interactions with host cells has been studied primarily in mouse models and has focused on either the host [28,29] or the pathogen [30,31]. Moreover, these studies did not address the importance of the intracellular phase in the infection cycle of *B. pertussis*. Therefore, we recently engaged the dual RNA-seq method to analyze parallel transcriptomic profiles of human macrophages infected with *B. pertussis* to elucidate the nature of the interactions between the pathogen and the immune cell [23]. To avoid the donor-to-donor variation observed in experiments with primary macrophages, we used THP-1 monocyte-derived cells, a popular cell model for immune modulations studies [32] that has been shown to be a good surrogate for primary cells in experiments with *B. pertussis* [22]. In the pilot experiment, which focused primarily on host cells, we demonstrated that in response to *B. pertussis* infection, the global gene expression profiles of THP-1 macrophages were extensively rewired during the course of infection [23]. Importantly, we have also shown that significant numbers of bacterial reads can be obtained from infected macrophage cells, provided we focus on the early stages of infection and substantially increase the sequencing depth.

In this manuscript, we focused on determining temporal changes in gene expression profiles of intracellular *B. pertussis* cells using time-resolved dual RNA-seq analysis. Our results indicate that intracellular *B. pertussis* cells switch from a virulent to an avirulent phenotype, which appears to be beneficial for intramacrophage survival and persistence, and support the emerging hypothesis that the avirulent phase represents an authentic phenotype of intracellular *B. pertussis* cells.

Materials and methods

Cultivation of B. pertussis *cells*

The *Bordetella pertussis* reference strain Tohama I and its derivatives (Table S3) were grown on Bordet-Gengou (BG) agar plates (Difco) supplemented with 15% sheep blood for 3 days at 37°C. For liquid cultures, cultures were inoculated from agar plates and grown in Stainer-Scholte (SS) medium [33] supplemented with 0.1% cyclodextrin (Sigma) and 0.5% casamino acids (Difco) at 37°C. For infection assays, the *B. pertussis* cells were grown overnight in SS medium to the mid-exponential phase of growth ($OD_{600} \approx 1$) and diluted to the appropriate cell density in RPMI medium.

Construction of B. pertussis *mutants*

The deletions were introduced into the chromosome of the *B. pertussis* Tohama I strain as described previously [34]. To construct the Δ*BP2871* deletion mutant, two DNA fragments of approximately 750 bp flanking the corresponding gene were amplified from the upstream region (which ends with the ATG codon of the Δ*BP2871* gene and the *Spe*I site) and the downstream region (which begins with the *Spe*I site and the TGA stop codon of the Δ*BP2871* gene). The resulting PCR products were ligated via a *Spe*I site, and the ligation mixture was used as a template to generate a PCR product of approximately 1.5 kb containing the intended deletion. In the resulting product, the start and stop codons of the corresponding gene separated by the *Spe*I restriction site formed a markerless in-frame deletion. The Δ*BP3011* deletion was constructed in the same manner, but the *Nhe*I site was used instead of *Spe*I to ligate both DNA fragments. In all cases, the final PCR products were ligated into the allelic exchange plasmid pSS4245 and cloned into the *E. coli* strain XL1-Blue. Finally, the resulting recombinant plasmid was transformed into the *E. coli* SM10 strain (donor strain) and transferred to *B. pertussis* Tohama I (recipient strain) by conjugation, as described elsewhere [35]. After two recombination events, the strain with the desired deletion was obtained. To create Δ*BP2871*Δ*BP3011* double mutant, the Δ*BP2871* mutant was conjugated with pSS4245 derivative carrying the Δ*BP3011* allele. All

mutations were confirmed by sequencing the amplified PCR product covering the regions adjacent to the deletion. The plasmids and primers used in this study are listed in Table S2 and Table 3, respectively.

Growth and differentiation of THP-1 cells

Monocytic THP-1 cells (ATCC; TIB-202) were grown in suspension in tissue culture-treated dishes (100 mm; Corning) in Roswell Park Memorial Institute (RPMI) 1640 medium (Sigma, R8758) supplemented with 10% heat-inactivated fetal bovine serum (Sigma) at 37°C in a humidified incubator (5% CO_2). For differentiation into macrophages, THP-1 monocytes were seeded in 6-well plates (4×10^6 cells per well) and stimulated by the addition of 100 nM phorbol 12-myristate 13-acetate (Sigma) for 72 h, followed by a 24-h resting period in plain RPMI medium.

Infection of THP-1-derived macrophages

Differentiated THP-1 macrophages were infected with *B. pertussis* cells diluted in RPMI medium at a multiplicity of infection of 10 bacteria per macrophage (4×10^7 per well). The plates were then centrifuged at 600 g for 3 min to facilitate the interaction of bacteria with macrophages. After 1 h of incubation (37°C; 5% CO_2), extracellular bacteria were removed by washing with RPMI medium, and the remaining bacteria were killed by incubation in RPMI medium containing 100 µg/ml of polymyxin B sulphate (Sigma) for 1 h (37°C; 5% CO_2). No viable bacteria were detected by direct plating of cell culture supernatants on BG agar (data not shown). From this time point (T2), corresponding to 2 h pi (1 h of co-incubation of macrophages with *B. pertussis* cells and 1 h of polymyxin B treatment), the concentration of antibiotic was maintained at 25 µg/ml which was sufficient to prevent repopulation of the RPMI medium with released intracellular bacteria. To obtain samples of infected macrophages 1 h pi (T1), both co-incubation of *B. pertussis* with THP-1 cells and treatment with polymyxin B were performed for 30 min each. Again, no viable bacteria were detected by direct plating of cell culture supernatants on BG agar. At the indicated time points (1, 2, 4, 8 and 12 h pi; corresponding to time points T1–T5), infected cells were washed intensely with prewarmed phosphate-buffered saline (PBS) and directly lysed. In parallel, samples of uninfected macrophages treated in the same manner and samples of intact bacteria incubated in RPMI at 37°C served as controls. In addition, at each time point infected macrophages from a separate well were lysed with sterile water containing 0.01% Triton X-100, and the lysates were plated onto BG agars to determine the numbers of viable intracellular bacteria. At each time point, three biological replicates of infected and uninfected macrophages and unexposed bacteria were harvested for RNA isolation.

RNA isolation

Samples of biological triplicates of infected and uninfected macrophages were lysed at each time point with 0.5 ml of TRI reagent (Sigma), and the lysates were stored at −80°C. Total RNA was isolated from the lysates according to the manufacturer's protocol. Cultures of *B. pertussis* cells grown in biological triplicates in RPMI medium (controls) were mixed 4:1 with stop solution (95% ethanol, 5% phenol), pelleted, and stored at −80°C. Bacterial pellets were suspended in TE buffer (10 mM Tris, 1 mM EDTA; pH 8.0) containing 1 mg/ml lysozyme (Sigma) and total RNA was isolated from lysed cells using TRI reagent (Sigma) according to the manufacturer's protocol. DNA was removed by treatment of all samples with the TURBO DNA-free kit (Thermo Fisher Scientific). Quality and quantity of RNA were determined by agarose gel electrophoresis and using the Nanodrop 2000 instrument (Thermo Fisher Scientific). In addition, RNA quality was assessed at the sequencing facility (Vienna Biocenter Core Facility, NGS unit) on an Agilent 2100 Bioanalyzer. All samples displayed RNA integrity numbers greater than 9.

Library preparation, sequencing, read mapping and DE analysis

Host and pathogen ribosomal RNAs were simultaneously depleted using the Ribo-Zero Gold rRNA Removal Kit (Illumina). Libraries were prepared using the NEBNext® Ultra™ II DNA Library Prep Kit for Illumina and sequenced on an Illumina HiSeq 2500 platform using HiSeqV4 chemistry with single-end 100-base-pair reads. FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc; version 0.11.9) was used to perform quality control checks on raw Illumina reads in FASTQ format. A single reference transcriptome generated by combining the human transcriptome (Ensembl GRCh38) and *B. pertussis* Tohama I transcriptome was used to map and quantify reads with Salmon quantifier (version 1.1.0) [36]. The *B. pertussis* reference transcriptome sequence file was generated using the gffread tool (https://github.com/gpertea/gffread), the Tohama I reference sequence (GCF 000195715.1) from the NCBI assembly, and the corresponding Tohama I gene annotation in GFF file format (GCF_000195715.1_ASM19571v1_genomic.gff). After quantification performed with Salmon and the reference transcriptome file, the human and bacterial read counts per gene were separated for subsequent differential expression analysis. Unwanted variations caused by batch effects or library preparation

were removed from the samples prior to DE analysis using the RUVs correction method of RUVseq (version 1.24.0) [37] in R (version 4.0.5). DE analysis was performed with the EdgeR package (version 3.32.1) [38] using the gene read counts as inputs. To normalize for sequencing depth, we calculated library size normalization factors for each library using the weighted Trimmed Mean of M-values method implemented in the EdgeR by means of calcNormFactors function.

Bioinformatics analyzes

To perform principal component analysis (PCA) for THP-1 macrophages and *B. pertussis* samples, read counts per gene were first normalized by regularized log transformation implemented in the rlog function of DESeq2 [39] for each time point. The plots were generated using the ggfortify (https://github.com/sinhrks/ggfortify; version 0.4.14) and ggplot2 (version 3.3.5) [40] packages in R. GO term enrichment analysis for *B. pertussis* genes was deduced using blast2go [41]. The log fold changes obtained from DE analysis were used for each gene and sample for the corresponding term. Each GO term associated with more than one gene was tested for enrichment compared to the whole transcriptome using Fisher's exact test. Subsequently, the determined *p*-values were corrected for multiple testing by the method of Benjamini and Hochberg [42]. The 20 most enriched "Upregulated" and "Downregulated" gene sets for each time point were selected for visualization. The package ggplot2 was used in R to generate heatmaps for the selected terms and corresponding genes for the five infection time points.

Validation of the RNA-seq data by quantitative PCR

To validate the gene expression profiles obtained from RNA-seq analysis, the expression of selected *B. pertussis* genes was examined by RT-qPCR method using the same RNA samples previously used for RNA-seq analysis. Isolated RNA (600 ng) was reverse transcribed into cDNA using random hexamers and M-MLV reverse transcriptase (Promega) at 37°C in a 25-µl reaction according to the manufacturer's instructions. RT-qPCR reactions were performed in at least three technical replicates per sample on the Bio-Rad CFX96 instrument using SYBR® Green JumpStart™ Taq ReadyMix™ (Sigma), 4 pmoles of each primer, and 2 µl of fiftyfold diluted cDNA in a 20-µl reaction. Each reaction consisted of an initial step at 95°C for 2 min, 40 cycles of 95°C for 15 s, 65°C for 30 s, and 72°C for 30 s, followed by a melting curve recording. The sequences of primers used for gene expression analysis of THP-1 and *B. pertussis* have been published previously [23,34,43] or are listed in Table S3. The *rbfA* gene was used as a reference gene to determine relative gene expression in intracellular and unexposed bacteria. Relative gene expression was determined using the delta-delta C_t method [44].

THP-1 cells viability assay

The viability of infected THP-1 macrophages was determined spectrophotometrically as the capacity of mitochondrial dehydrogenase to induce cleavage of the tetrazolium salt to formazan using the WST-1 assay kit (Roche) according to the manufacturer's protocol. THP-1 monocytes were seeded in 48-well plates (4.6×10^5 cells per well) and differentiated into macrophages. Macrophages were infected with *B. pertussis* cells at different MOIs as described above. Briefly, after 1 h of co-incubation, nonadherent bacteria were removed by washing, and remaining extracellular bacteria were killed by incubation in RPMI medium containing 100 µg/ml of polymyxin B sulphate (Sigma) for 1 h. At this time point, corresponding to 2 h pi, infected cells were washed intensely with prewarmed RPMI medium and then incubated in 200 µl of RPMI and 20 µl of WST-1 substrate for 40 min at 37°C. In parallel, uninfected cells were treated in the same manner and served as controls (100% viability). The absorbance of formazan dye measured at 450 nm correlates directly with the number of viable cells and was measured with a scanning multi-well spectrophotometer Epoch (BioTek).

Cytotoxicity toward THP-1 cells was determined using the CellTox™ Green Cytotoxicity Assay (Promega), which measures changes in membrane integrity that occur as a result of cell death. The fluorescent signal generated by binding of the dye to the dead-cell DNA is proportional to cytotoxicity. THP-1 monocytes grown in colorless RPMI medium (R7509, Sigma) supplemented with L-glutamine (0.03%) were seeded in 96-well plates (1×10^5 cells per well) and differentiated into macrophages. The differentiated macrophages were infected by adding 100 µl of a *B. pertussis* cell suspension (containing 5×10^6 cells; MOI 50). After 30 min of co-incubation, nonadherent bacteria were removed by washing, and the remaining extracellular bacteria were killed by incubation in RPMI medium containing 100 µg/ml of polymyxin B sulphate (Sigma) for 30 min. At this time, the infected cells were washed intensely with prewarmed RPMI medium, and the assay was started by adding of CellTox™ reagent according to the manufacturer's protocol. Uninfected cells were treated in the same manner and served as controls. Next, the macrophages were incubated in a Tecan Spark multimode microplate reader (37°C, 5% CO2), and the fluorescent signal was measured at Ex/Em = 490 ± 10/520 ± 10 nm for 12 h after addition of reagent.

Results

Parallel transcriptomic profiling of infected macrophages

Monocyte-derived THP-1 macrophages were infected in triplicate with 4.0×10^7 cells per well of the *B. pertussis* reference strain Tohama I. Samples of infected macrophages were collected at 1, 2, 4, 8, and 12 h (time points T1–T5) post-infection (pi). Control samples (C1–C5) of uninfected macrophages and intact bacteria cultivated in triplicate in RPMI medium were incubated and harvested simultaneously. In parallel with RNA isolation, plating of lysed infected macrophages revealed that the macrophages contained approximately 4.1×10^6, 4.8×10^6, 4.0×10^6, 4.4×10^6, and 3.3×10^6 viable bacteria per well at time points T1, T2, T3, T4, and T5, respectively.

RNA-seq analysis of samples isolated from infected cells yielded between 50 and 120 million reads, whereas control samples (uninfected macrophages and bacteria incubated in RPMI medium) yielded an average of 20 million reads (Figure S1). Reads were mapped to human and B. *pertussis* genomes and all samples were clustered by principal component analysis (PCA). PCA revealed that samples of infected macrophages clustered separately from uninfected cells (Figure 1(a)). Similarly, samples of intracellular bacteria clustered apart of control samples along principal component 1 (PC1), which accounted for nearly 42% of the observed variance (Figure 1(b)). The control samples also exhibited some degree of variability along PC2 ($\approx$ 12%), although the samples of intracellular bacteria displayed a similar trend toward PC2, but to a lesser extent. Therefore, we cannot exclude the possibility that some variance in the control samples was due to RPMI medium-specific effects. Importantly, samples of both infected macrophages and intracellular bacteria collected at each time point clustered separately, indicating significant time-dependent changes in gene expression profiles during infection in both the host and the pathogen as early as 1 h pi.

Global changes in transcriptional profiles of internalized B. pertussis cells

Next, we focused on samples of *B. pertussis* to monitor temporal changes in gene expression profiles of intracellular bacteria during the course of infection. To this end, we performed pairwise differential expression (DE) analysis to compare the transcriptomic profiles of intracellular and control bacteria at each of the five time points. The DE analysis identified 455, 708, 1112, 1255, and 1243 significantly modulated ($|\log_2 FC| \geq 1$; adjusted p-value < 0.05) genes at time points T1 to T5, respectively (Table S1). Similarly, 86, 137, 160, 130, and 110 non-coding RNAs, including rRNAs, tRNAs, and previously identified candidate non-coding RNAs [45], were significantly modulated at the corresponding time points. Detailed analysis of protein-coding genes revealed that 1368 genes were modulated at least at two time points and 148 genes were modulated at all five time points (Figure S2). To validate the RNA-seq data, we analyzed the expression profiles of randomly selected significantly modulated genes by quantitative PCR. This experiment revealed that, in agreement with the RNA-seq data, the expression of the virulence genes *fhaB*, *prn*, *vag8*, *bsp22*, *brkA*, and *tcfA* was decreased in internalized cells and the expression of *BP2871*, *vrg6* and *vrg24* genes was increased (Figure S3).

To gain deeper insight into the functional clustering of upregulated and downregulated genes at each time point, we performed gene ontology (GO) term enrichment analyzes. This analysis focused on biological processes and revealed the GO terms enriched for genes that displayed either increased or decreased expression during the intramacrophage phase (Figure 2). Among the biological processes enriched in upregulated genes, we identified those related to the "Phosphorelay signal transduction system," "Regulation of transcription," "Cell redox homeostasis," and "Protein repair" (Figure 2(a)). On the other hand, processes such as "Pathogenesis," "Siderophore transport," "Intracellular protein membrane transport" and several processes linked to the energy metabolism and biosynthesis of amino acids were significantly enriched in downregulated genes (Figure 2(b)).

Adaptation of B. pertussis to the intramacrophage environment

GO enrichment analysis revealed that several important biological processes are enriched in significantly modulated genes. Therefore, we examined these processes in more detail (Figure 3).

Pathogenicity

Consistent with our previous report [23], the expression of majority of virulence factors was significantly reduced within macrophages. As can be seen from Figure 3(a), this effect occurred mainly after 2 h pi and was due to reduced expression of the *bvgA* (*BP1877*) and *bvgS* (*BP1878*) genes, the main regulators of virulence. Among the downregulated genes, we identified those encoding adenylate cyclase toxin (*BP0760*), dermonecrotic toxin (*BP3439*), adhesins such as filamentous haemagglutinin (*BP1879*), serotype 2 fimbriae (*BP1119*), and pertactin (*BP1054*), autotransporters Vag8 (*BP2315*) and BrkA (*BP3494*), tracheal colonization factor A (*BP1201*), and the nearly complete *bsc* locus encoding the type 3 secretion system (*BP2225–BP2264*).

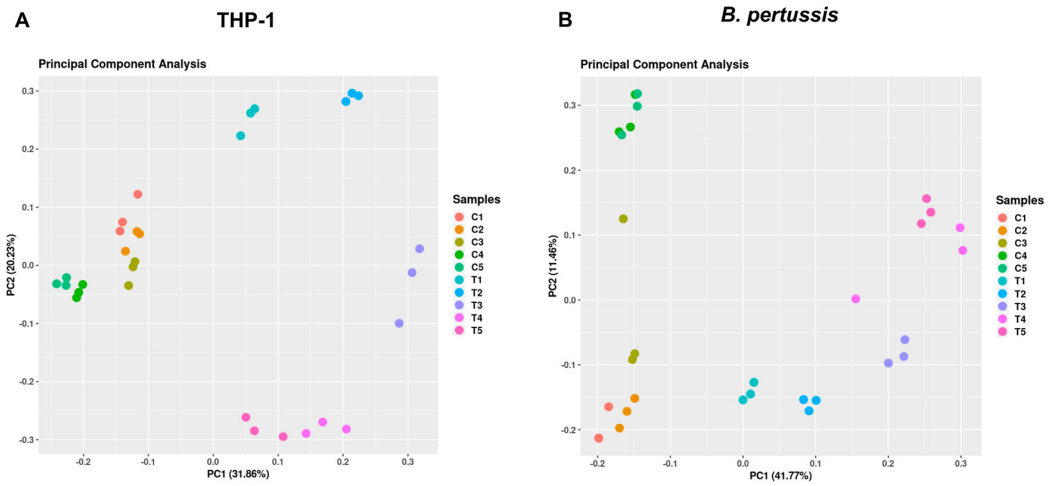


Figure 1. (a) Principal component analysis was performed with transcriptomic profiles of uninfected (C1-C5) and infected (T1–T5) THP-1 macrophages harvested 1 h (T1), 2 h (T2), 4 h (T3), 8 h (T4), and 12 h (T5) pi. (b) Principal component analysis was performed with transcriptomic profiles of control *B. pertussis* cells incubated in RPMI medium (C1–C5) and intracellular *B. pertussis* cells (T1–T5) harvested 1 h (T1), 2 h (T2), 4 h (T3), 8 h (T4), and 12 h (T5) pi. Each dot in both panels represents an independent biological replicate.

Membrane composition and envelope stress

Internalized bacteria are subjected to various stresses resulting from harsh conditions within macrophages. Low pH, and the activity of reactive oxide species and antimicrobial peptides exert strong effects on cell wall integrity. Our results indicate that intracellular *B. pertussis* cells actively adapted their membranes to macrophage-induced stress, as the expression of several genes and operons involved in fortifying the cell wall was increased, especially at later time points (Figure 3(b)). For example, the expression of the *BP0396–0399* cluster, which affects LPS composition and lipid A remodelling, was strongly upregulated. Similarly, the *BP0769* gene, encoding the D-amino acid transferase required for the synthesis of the peptidoglycan components D-alanine and D-glutamate, and the *dra* locus (*BP2987–BP2991*), required for the incorporation of D-alanine into outer membrane, were incessantly upregulated during infection. In addition, the locus *BP2434–BP2437*, which contains the *mucD*, *mucB*, *rseA*, and *rpoE* genes that encode factors involved in the response to outer membrane stress, was highly overexpressed. The expression of the lipid A palmitoyltransferase gene *pagP* (*BP3006*)

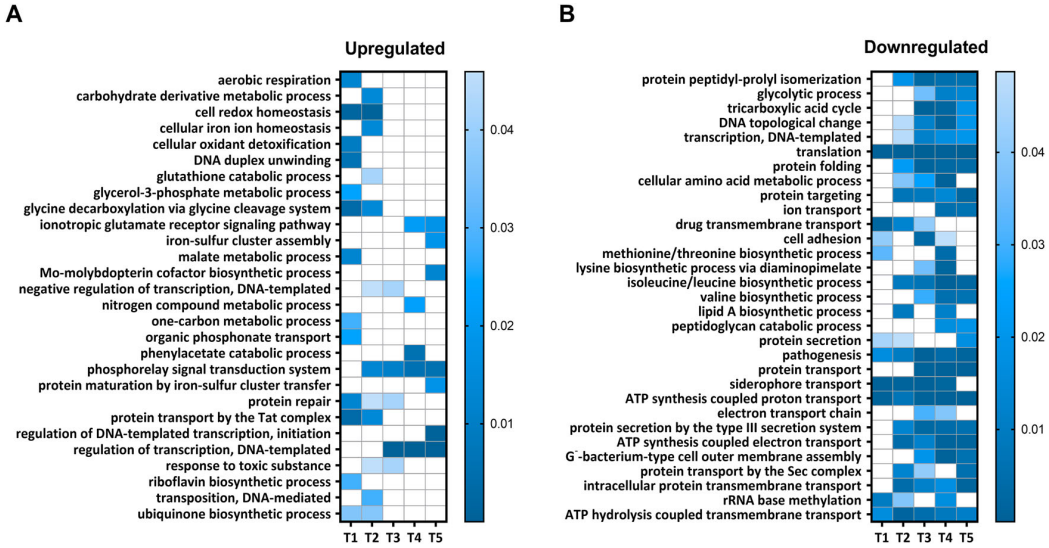


Figure 2. GO enrichment analysis of processes modulated in intracellular *B. pertussis* cells. Analysis was applied to identify GO terms from the domain "Biological processes" enriched for either upregulated (panel a; $\log_2FC > 0$) or downregulated (panel b; $\log_2FC < 0$) DE genes at all experimental time points (T1–T5). The processes that were significantly enriched at least at one time point were selected for visualization. The vertical bars show the adjusted *p*-values (shades of blue) for both sets of genes.

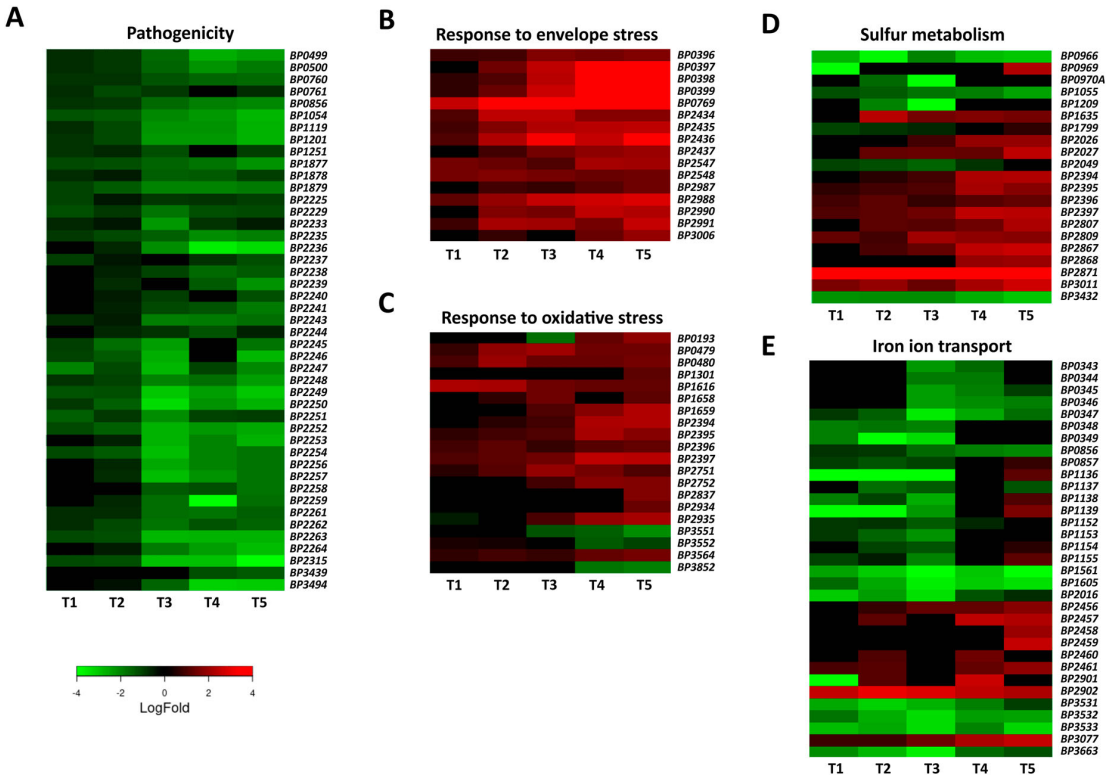


Figure 3. Adaptation of *B. pertussis* to the intramacrophage environment. Heat maps depicting $\log_2$ fold changes in the expression of *B. pertussis* genes (intracellular/intact bacteria) during infection (T1–T5). Shown are genes involved in pathogenesis (a), response to membrane stress (b), response to oxidative stress (c), sulphur metabolism (d), and iron transport (e).

was increased at later time points. On the other hand, the expression of porin genes such as *BP0840*, *BP0943* (*ompA)*, *BP3405* (*ompQ*), and *BP3755* (*ompW*), which enable the transport of small molecules including antimicrobial peptides, was downregulated (Table S1). The expression of *lpx* (*BP1429–BP1432*) and *bpl* (*BP0089–BP0093*) operons, which are responsible for lipooligosaccharide biosynthesis, was also significantly decreased (Table S1).

Oxidative stress and redox homeostasis

During infection, the production of reactive oxygen species resulting from bacterial metabolism and from the activity of immune cells can substantially damage bacterial macromolecules. Our data suggest that intracellular *B. pertussis* cells responded to oxidative stress (Figure 3(c)). Indeed, the expression of the *dps* (*BP1616*) gene encoding the DNA-binding protein that confers resistance to hydrogen peroxide by absorbing the oxidative damage in place of the DNA [46] was increased throughout the experiment. The expression of the redox-sensitive regulator gene *soxR* (*BP2837*) and the Mn-dependent superoxide dismutase *sodA* (*BP0193*) was significantly increased, though only at later time points. In addition, the expression of paraquat-inducible genes *pqiA* (*BP2751*) and *pqiB* (*BP2752*) and glutathione peroxidase gene (*BP1301*) was also increased during infection. Moreover, our analysis revealed several highly expressed genes involved in the repair of damaged proteins. Indeed, the sulphur-containing amino acids methionine and cysteine are particularly sensitive to oxidative stress because of their electron-rich sulphur atom. During infection, genes encoding cytosolic MsrB (*BP3564*) and periplasmic MsrP (*BP0479*) and MsrQ (*BP0480*) methionine reductases were significantly upregulated. Our data also showed that the locus encoding the putative ATP-binding cassette transporter (*BP2394–BP2397*) with specificity for di- and tripeptides (*e.g.* glutathione) was strongly overexpressed during the course of infection. In contrast, the entire *pha* locus (*BP1691–BP1696*), which encodes the potassium efflux system involved in pH adjustment, was downregulated (Table S1).

Sulphur metabolism

Transcriptomic analysis revealed that the expression of a variety of genes controlling sulphur metabolism was strongly modulated (Figure 3(d)). The reductive sulphate assimilation pathway, which is required for sulphate transport and cysteine biosynthesis, was attenuated during infection. In particular, the *sbp* (*BP0966*), *cysD* (*BP0970*), *cysG* (*BP1055*), *cysJ* (*BP1209*), sulphate permease *BP2049*, *cysI* (*BP3432*), and adenosylhomocysteinase *BP3068* genes were significantly downregulated. The *BP2816–BP2818* locus,

which encodes the putative methionine importer, was also downregulated. On the other hand, the expression of genes involved in cysteine export and catabolism was strongly increased. For example, the genes *BP2026* and *BP2027* genes, encoding a probable cysteine exporter and a sulphite exporter, respectively, and the two putative cysteine dioxygenase genes *BP2871* and *BP3011* were strongly upregulated.

Metal homeostasis

During infection, bacterial pathogens enter a hostile and variable environment that contains limited amounts of essential nutrients, including metals. Therefore, we were surprised to see that several iron transport genes and systems were downregulated during infection, predominantly within the first 4 h pi (Figure 3(e)). For example, the expression of the *fecI* and *fecR* genes, which control the transport of ferric citrate, and the *ftrABCD* operon (*BP1152–BP1155*), which facilitates the transport of ferrous iron, was strongly downregulated within the first four hours after infection. Similarly, expression of the haem transport system encoded by the *bhuRSTUV* locus (*BP0343–BP0347*) and adjacent regulatory genes *hurR* (*BP0348*) and *hurI* (*BP0349*) was significantly downregulated during infection. The expression of *BP3531–BP3533* cluster, which consists of the *tonB*, *exbB*, and *exbD* genes involved in iron binding and transport, was also strongly reduced. In contrast, the expression of the *alc* operon (*BP2456–BP2461*), which is responsible for the biosynthesis of alcaligin siderophores, increased significantly 8 h after infection. Genes responsible for the binding (*BP2901*) and release (*BP2902*) of enterobactin siderophores were also overexpressed.

The expression of several genes involved in the homeostasis of other essential metals was modulated as well. For example, the magnesium importer gene *mgtC* (*BP0414*), the putative manganese importer locus *BP3080–82*, and the putative nickel/cobalt transporter gene *BP3078* were strongly upregulated (Table S1).

Central and energy metabolism

Our data indicate that intracellular cells extensively rewired their metabolic activity and energy production upon entry into host macrophages (Table S1). Most strikingly, the large *nuo*, *sdh* and *atp* operons, encoding the electron transfer chain complexes of NADH:ubiquinone oxidoreductase, succinate dehydrogenase, and ATP synthase, respectively, were strongly downregulated from the beginning of infection. Furthermore, expression of the *cyo* and *cyd* operons encoding terminal ubiquinol oxidases and genes encoding terminal cytochrome cbb3 oxidase were also significantly downregulated.

In addition, several important metabolic processes were substantially modulated in internalized bacteria (Table S1). Among others, several genes involved in glycolysis, such as *gap*, *pgk*, and *ldh*, were downregulated. On the other hand, the expression of *glcDEF* genes involved in glycolate/glyoxylate metabolism (*BP2903–BP2905*) and the glyoxalase gene *BP2863* was strongly upregulated. Similarly, the expression of genes responsible for the biosynthesis (acetolactate synthase *BP0467*) and transport (ABC transporters *BP0619–BP0622* and *BP3573–BP3577*) of branched-chain amino acids, required for the metabolism of fatty acids (acetyl-CoA synthetase *BP0661*, enoyl-CoA hydratases *BP0627* and *BP0662*, and long-chain fatty acid ligase *fadD*), and for phenylacetic acid catabolic pathway (*BP2675–BP2684*), was significantly increased.

Large part of the BvgAS regulon is modulated in internalized cells

The greatly reduced expression of virulence genes suggested that intramacrophage environment induces an avirulent phenotype in internalized *B. pertussis* cells. We noticed that in addition to a large number of virulence genes, a great variety of other genes belonging to the BvgAS regulon were identified among the DE genes. Therefore, we determined the extent to which the BvgAS regulon overlaps with the set of genes found in our DE analysis which we named "intracellular adaptome". In this analysis, we considered all *bvg*-dependent genes published by Moon et al. [47]. The rationale behind this was that when internalized cells develop the Bvg$^-$ phase, then all genes that are downregulated by the BvgAS system (virulence-repressed genes) should be upregulated and *vice versa*, genes that are activated by the BvgAS system (virulence-activated genes), should be downregulated upon phagocytosis. Interestingly, 207 of 322 virulence-repressed genes (64.2%) were significantly upregulated at least at one time point in internalized cells and 172 of 237 virulence-activated genes (72.5%) were significantly downregulated in intracellular bacteria (Figure 4(a)). A large proportion of the DE genes do not overlap with the BvgAS regulon.

Expression of at least some of the virulence-repressed genes in *B. pertussis* is induced by the RisAK two-component system [48]. Therefore, we compared the intracellular adaptome with the list of genes shown to be transcriptionally regulated by the RisA factor (Figure 4(b)). Analogously to BvgA, we determined whether genes activated by RisA were upregulated and genes repressed by RisA were downregulated in intracellular bacteria. This analysis showed that among the genes activated by RisA, 8 of 21 (38.0%) were upregulated, including *vrg6* and several genes encoding putative lipoproteins and small

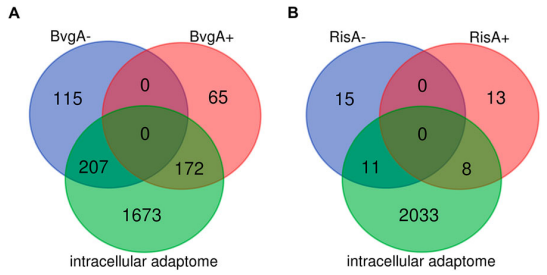


Figure 4. Comparison of the intracellular adaptome with BvgAS and RisAK regulons. (a) Venn diagrams showing the number of genes differentially expressed in intracellular *B. pertussis* cells at least at one time point (intracellular adaptome), the number of *B. pertussis* genes repressed (BvgA−) or activated (BvgA+) by the BvgAS system (reference [9] in the manuscript), and the overlap between each dataset. (b) Venn diagrams showing the number of genes differentially expressed in intracellular *B. pertussis* cells at least at one time point (intracellular adaptome), the number of *B. pertussis* genes repressed (RisA−) or activated (RisA+) by the RisAK system (reference [48] in the manuscript), and the overlap between each dataset.

membrane proteins. Of the genes repressed by RisA, 11 of 26 (42.3%) were downregulated in intracellular bacteria, including several flagellar genes. Similar to BvgA, a large part of the DE genes do not overlap with the RisA regulon.

Cysteine dioxygenases impact on pathogenesis of B. pertussis

DE analysis revealed that a large number of genes responsible for the transport and metabolism of cysteine were modulated in intracellular *B. pertussis* cells. Among them, two putative cysteine dioxygenases (CDO) genes, *BP2871* and *BP3011*, were highly upregulated upon internalization, and *BP2871* in particular was among the most deregulated genes in the DE analysis ($\log_2$FC > 7). Consistent with their annotation, alignment of the two enzyme sequences with the Uniprot database identified several domains characteristic of the CDO protein family (Figure 5(a)). Given the extent of upregulation, we wondered whether the *cdo* genes play a significant role in intracellular *B. pertussis* cells. Therefore, macrophages were infected at three different multiplicity of infection (MOI) doses of 10, 20, and 50 bacteria per macrophage and viability was tested 2 h pi (Figure 5(b)). First, we tested a single Δ*BP2871* mutant and determined its cytotoxicity, *i.e.* its capacity to reduce the viability of THP-1 macrophages, however, we found no significant difference compared to wt strain. The viability of macrophages infected with the Δ*BP3011* mutant lacking the other *cdo* gene was slightly but significantly increased at all MOIs, however, the viability of infected cells was still substantially reduced compared with uninfected macrophages, especially at the highest MOI. We hypothesized that the two *cdo* genes might complement each other and therefore, we constructed a double Δ*BP2871*Δ*BP3011* mutant lacking both *cdo* genes. The viability of THP-1 cells infected with the double mutant was significantly higher than that of cells infected with the wt strain and both single *cdo* mutants at all MOIs tested (Figure 5(b)). To confirm these results, we examined the cytotoxic effects of all strains toward macrophages in real time using an assay that measures changes in membrane integrity resulting from cell death (Figure 5(c)). Consistent with the cell viability assays, the cytotoxic effects toward THP-1 cells caused by the double mutant were strongly reduced compared with the wt strain and the two single *cdo* mutants (Figure 5(c)).

Discussion

Although *B. pertussis* is classified as an extracellular pathogen, it has been reported that these bacteria can also survive and persist within host cells, including both professional and nonprofessional phagocytes [20,49]. It has been hypothesized that the ability for intracellular persistence is conserved among mammalian pathogens of *Bordetella* species and that the intracellular phase contributes to infection [26,27,50]. In this regard, our data indicate that after the initial elimination of extracellular bacteria, viable intracellular *B. pertussis* cells repopulate the antibiotic-free RPMI medium during infection. Approximately 10^4 *B. pertussis* cells per well (≈ 1% of intracellular bacteria) are found in the extramacrophage environment (data not shown), a number very similar to that of *B. pertussis* cells released from human epithelial A549 cells [49]. This observation suggests that bacteria that survive in macrophages may contribute to disease transmission.

The conservation of intracellular persistence would imply that *Bordetella* species have evolved a specific expression programme to adapt to the intramacrophage environment. Therefore, we have employed time-resolved RNA-seq approach to specifically monitor and describe the adaptation of *B. pertussis* to the intramacrophage environment to elucidate the capacity of this reemerging pathogen to persist within this niche. Our data clearly demonstrate that *B. pertussis* cells actively adapt to the harsh environment within macrophages early after internalization. The changes in global gene expression profiles were enormous, with more than 30% of genes differentially expressed at least at two time points during the course of infection. Nevertheless, we are aware that part of the observed effect could be due to RPMI-specific effects on the expression of control bacteria, especially at the last two time points. In our pilot experiment we identified 310 bacterial genes that were induced 6 h

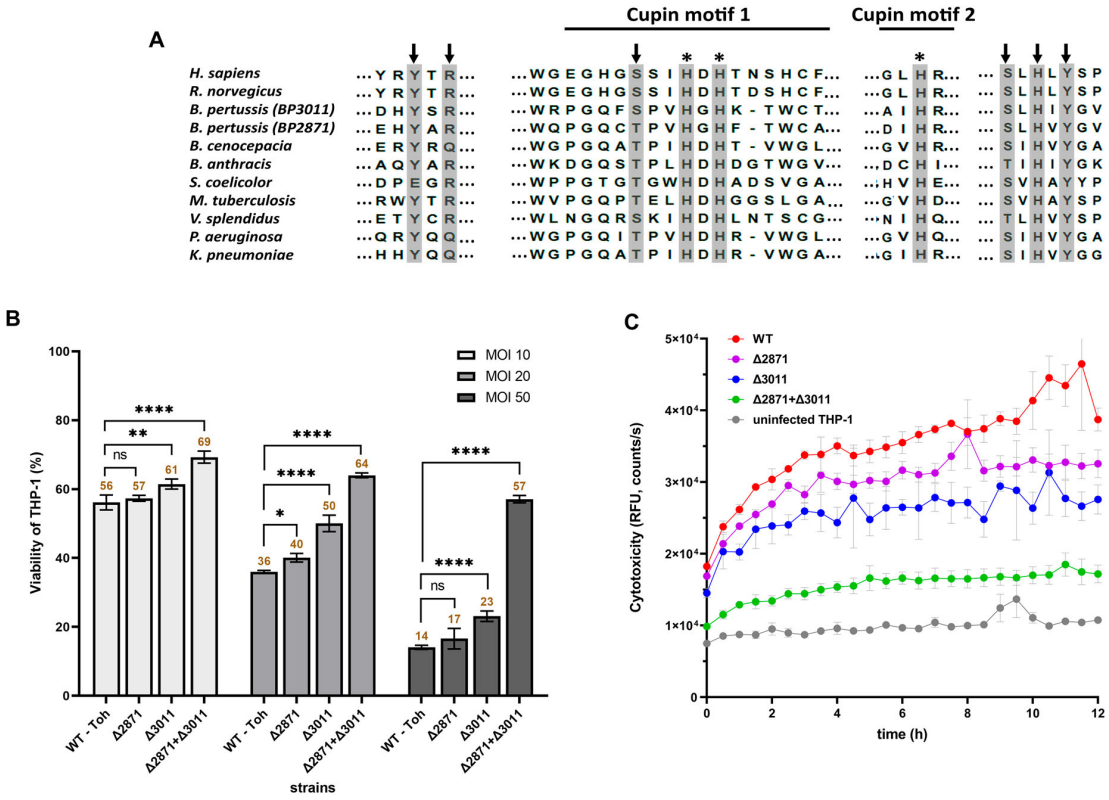


Figure 5. Impact of *cdo* genes on *B. pertussis* pathogenicity. (a) Conservation of functional residues in CDO enzymes between *B. pertussis* and other organisms. The grey rectangles show important residues that are either part of the CDO active site (arrows) or involved in iron binding (asterisks). The extent of cupin motifs 1 and 2 is depicted with solid lines. Uniprot query numbers are as follows: *Homo sapiens*, (Q16878); *Rattus norvegicus*, (P21816); *Bordetella pertussis BP3011*, (Q7VUR5); *Bordetella pertussis BP2871*, (Q7VV35); *Burkholderia cenocepacia*, (A0A142PCQ1); *Bacillus anthracis*, (A0A1T3V1L5); *Streptomyces coelicolor*, (A0A6M9XRZ5); *Mycobacterium tuberculosis*, (A0A045IN01); *Vibrio splendidus*, (A0A0P6ZFX5); *Pseudomonas aeruginosa*, (A0A0F7QLY4); *Klebsiella pneumoniae* (A0A086IHW3). (b) Viability of THP-1 macrophages infected with *cdo* mutants of *B. pertussis*. Macrophages were infected in triplicate with the *B. pertussis* wt strain and Δ*BP2871*, Δ*BP3011*, and Δ*BP2871*Δ*BP3011* mutants at MOIs of 10, 20, and 50 bacteria per macrophage and incubated for 2 h at 37°C. Next, macrophages were washed with fresh RPMI and finally incubated in 200 µl RPMI and 20 µl WST-1 reagent for 40 min. After incubation, the absorbance of the samples, which is proportional to cell viability, was measured at 450 nm using a multi-well spectrophotometer. The absorbance of uninfected cells treated in parallel in the same manner was arbitrarily set to 100%. The bars represent mean values ± standard deviation, the labels above the bars indicate the mean values of cell viability (%). Statistical analysis was performed using a two-way ANOVA test for multiple comparisons (Dunnett's test); ns, *p*-value > 0.05, *, *p*-value < 0.05, **, *p*-value < 0.005, ****, *p*-value < 0.00005. The result is representative of three independent experiments. (c) Cytotoxicity of *cdo* mutants of *B. pertussis* toward THP-1 macrophages. Macrophages were infected in triplicate with the *B. pertussis* wt strain and the Δ*BP2871*, Δ*BP3011*, and Δ*BP2871*Δ*BP3011* mutants (MOI of 50). Uninfected cells served as controls. Immediately after addition of the fluorescent dye, THP-1 cells were incubated for 12 h (37°C, 5% CO_2) in the microplate reader. During incubation, the fluorescence of the samples, which is proportional to cytotoxicity, was measured every 30 min. The graph shows mean values and the standard errors of the means. The result is representative of two independent experiments.

post-infection [23]. Compared to here presented study, 79.3% of these genes displayed identical expression patterns (data not shown). In addition, we compared the results of our study with proteomic analysis of intracellular *B. pertussis* cells isolated from THP-1 macrophages. This study differed in some important details (*e.g.* MOI) that could affect the dynamics of infection and gene/protein expression, nevertheless, of the 203 proteins modulated 3 h pi we found 82 genes (40%) that displayed similar trends in expression/production.

Remarkably, this global rewiring of gene expression profiles in intracellular cells was closely associated with the switch from the Bvg^+ to Bvg^- phase. Yet, the magnitude of the "intracellular adaptome" exceeded more than twice the extent reported for the BvgAS regulon [9], suggesting that the adaptation of *B. pertussis* to the intramacrophage environment is orchestrated also by additional regulators. Of note, 100% of *B. pertussis* colonies obtained by direct plating of lysed macrophages (containing viable intracellular bacteria) produced adenylate cyclase toxin, as demonstrated by clear zones on BG agar supplemented with blood (data not shown). This result suggests that the BvgAS system remained functional and was only temporarily silenced in macrophages.

In agreement with previous observations [23,26], our data indicated that the transcriptomic profiles of intracellular *Bordetella* cells differ substantially from those of bacteria isolated from the respiratory tract of mice [30,31,43]. Hallmark of this difference, similar to *B. bronchiseptica* [26], was the expression of virulence factors, which was strongly downregulated in intracellular cells. After host infection, extracellular bacteria express virulence genes to successfully colonize epithelial layers and manipulate immune cells to evade phagocytic clearance. Most likely, these "contact weapons" such as adhesins, toxins, autotransporters, and T3SS are required for colonization and immune system evasion in the host tissues, but are not required after internalization. We presume that intracellular *B. pertussis* cells have evolved an alternative strategy which appears to be based in part on a rapid shift from virulent to avirulent mode resulting from downregulation of the *bvgAS* two-component system, a key regulator of *Bordetella* virulence. Our data indicate that activation of the Bvg^- phase augments a very complex adaptive process based on adaptation of central and energy metabolism, fortification of the cell wall, maintenance of appropriate redox and metal homeostasis, and repair of damaged macromolecules.

Internalized *B. pertussis* cells are exposed to various stress factors such as reactive oxidative species, nutrient limitation, antimicrobial peptides, and low pH. Our results indicate that the adaptation of intracellular *B. pertussis* cells is multifaceted to integrate the response to multiple stressors. Maintenance of the cell envelope is one of these crucial processes, and our results suggest that *B. pertussis* copes with outer and inner membrane stresses induced by a hostile environment. For example, several genes that have been shown to be involved in membrane integrity protection (*pqiA* and *pqiB*) [51] or resistance to hydrogen peroxide (*dps*, glutathione peroxidase gene) [46,52] were upregulated. We also noticed that as a common theme, the mechanisms controlling cell envelope modifications were strongly upregulated, whereas the processes leading to *de novo* synthesis of the cell wall components were attenuated. Indeed, several factors involved in the modification of already transported lipooligosaccharide and peptidoglycan required for cell wall integrity of bacterial pathogens were strongly overexpressed. For example, lipid A modifications by glucosamine and palmitate [53,54] and peptidoglycan or outer membrane modification by D-alanine [55,56] have been shown to be required for resistance to antimicrobial peptides. In contrast, genes involved in the biosynthesis of lipid A and lipooligosaccharide were attenuated. We also observed downregulation of the *nuo*, *cyo* and *atp* operons, which encode large respiratory protein complexes in the cytoplasmic membrane. Remarkably, downregulation of these complexes has also been observed in bacteria responding to membrane stress by activation of the two-component regulatory system CpxAR [57]. Indeed, stressors such as low pH and the presence of misfolded proteins trigger the Cpx stress response, which mediates adaptation to cell envelope stresses by downregulation of nonessential protein trafficking between the cytosol and the inner membrane [58]. Although we did not identify plausible homologs of *cpxAR* genes in *B. pertussis*, we speculate that the observed reduced *de novo* biosynthesis of membrane components and decreased expression of large membrane complexes in intracellular *B. pertussis* cells also contributes to the maintenance of envelope integrity. Of note, the expression of several genes encoding outer membrane proteins was decreased. A similar pattern was observed in *Salmonella* and reduced outer membrane permeability was found to provide protection against toxic molecules [59]. Furthermore, to escape the antimicrobial effects of acidic pH, *B. pertussis* cells modulated the expression of several genes involved in the maintenance of pH homeostasis. For example, in agreement with a previous report [60], the expression of the magnesium importer *mgtC* homolog (*BP0414*) was increased during infection. MgtC has been shown to be an important adaptive factor for intramacrophage survival of bacterial pathogens [61], as it inhibits ATP synthase activity and consequently, hinders the proton influx and acidification of the environment [62]. Also, the strong repression of the *pha* locus, which encodes the potassium efflux pump, may have contributed to the neutralization of the intramacrophage milieu as the accumulation of potassium has been shown to be essential for pH homeostasis under acidic conditions [63].

The phagosome represents a hostile, nutrient-poor environment, and therefore, intracellular pathogens must adjust their metabolic activities. Our study revealed that internalized *B. pertussis* cells have extensively remodelled their metabolism to take advantage of the nutrients available within the phagosome. Importantly, most of the upregulated metabolic processes were facilitated by genes previously shown to be repressed by the BvgAS system [9]. This metabolic shift was highlighted by the induction of genes involved in the β-oxidation of fatty acids and the glycolate/glyoxylate pathway, which allow cells to utilize fatty acids-derived metabolites and convert them into simple carbon molecules [64]. These genes have also been upregulated in intracellular *B. bronchiseptica* cells [26], and glyoxylate production has been shown to be important for intramacrophage survival of *M. tuberculosis* [65]. In addition, genes associated with the tricarboxylic acid cycle, lipid metabolism, and branched amino acid transport were strongly upregulated. Of note, gene encoding glyoxalase, an enzyme that detoxifies methylglyoxal,

was also intensely induced. Methylglyoxal is a reactive electrophilic byproduct of metabolism and its removal has been reported to contribute to the survival of the intracellular pathogen *Listeria monocytogenes* [66]. The remodelling of metabolic fluxes in intracellular cells was also illustrated by extensive changes in the expression of *Bordetella* uptake genes (bugs). This group of genes encodes periplasmic solute-binding proteins that bind and deliver their specific cargo, such as amino acids and citrate, to their cognate membrane components [67,68]. *B. pertussis* contains more than 70 Bug proteins [69], and we found that about 73% of them were significantly modulated at least at one time point during infection.

One of the most interesting observations in our study was the large number of strongly and significantly modulated genes required for sulphur metabolism. Among the significantly downregulated genes were those involved in the sulphate assimilation pathway. Intriguingly, speciation of *B. pertussis* was accompanied by the loss of function of several genes required for sulphate uptake, such as *cysT*, *cysW*, and *cysH* [70]. Furthermore, our data indicate that most of the functional genes within the *cys* operon and genes involved in the biosynthesis and transport of sulphur-containing compounds such as cysteine and methionine were downregulated. Apparently, intracellular *B. pertussis* cells diminished the biosynthesis and uptake of cysteine and other sulphur-containing compounds in several steps. Cysteine is an important amino acid and nutrient, however, due to its highly reactive thiol group, elevated cytosolic cysteine levels are toxic [71]. Cysteine homeostasis is therefore maintained by several mechanisms including metabolic redistribution to non-reactive thiols (*e.g.* glutathione), cysteine efflux, and cysteine degradation, all of which ensure the removal of excess cysteine [72,73]. This is likely also true for *B. pertussis*, as several genes required for cysteine catabolism and export have been strongly upregulated including the cysteine exporter and two CDOs. The biology and function of CDO enzymes are best studied in eukaryotic cells [74], yet, CDOs have also been identified in bacterial cells [73]. These enzymes, which belong to the cupin family of proteins, play an important role in cysteine homeostasis by initiating the conversion of cysteine into beneficial products such as pyruvate [73,74]. The human pathogenic yeast *Candida albicans* lacking the CDO gene induced delayed mortality in a mouse model [75]. Our experiments revealed that the cytotoxicity of the *B. pertussis* double CDO mutant toward human macrophages was strongly reduced and thereby highlighted the importance of cysteine homeostasis for *B. pertussis* adaptation and survival. Though we provide only preliminary evidence, we speculate that CDO activity plays an important role in the physiological fitness of other bacterial pathogens.

The intramacrophage environment is deprived of essential metals, and therefore pathogens need to scavenge metals using high-affinity acquisition systems. After infection, the expression of iron utilization systems of *B. pertussis* was differentially phased within macrophages. A similar phenomenon was observed in *B. pertussis* cells during colonization of the mouse respiratory tract [76]. Especially within the first 4 h pi, several iron transport systems and iron starvation-inducible genes were downregulated, suggesting that intracellular *B. pertussis* cells were not limited by iron during the early phase of infection. On the other hand, expression of loci involved in the utilization of enterobactin and alcaligin siderophore systems was increased at later time points. These results suggest that siderophores became an essential and preferential system for iron acquisition in intracellular bacteria during the course of infection. We hypothesize that *B. pertussis* cells adapt to the iron resources available within the intracellular niche (lack of free iron) and use the siderophores to acquire iron from host iron-loaded molecules such as ferritin.

Remarkably, expression of the *BP3077–BP3082* operon increased strongly and steadily during infection. This operon encodes a TonB-dependent receptor (*BP3077*), a putative nickel/cobalt transporter (*BP3078*), a DUF1007 family protein (*BP3079*), and an ABC transporter (*BP3080–BP3082*) that shares a substantial homology with the manganese importer SitABCD of *Salmonella enterica* subsp. *Typhimurium* [77]. We have previously shown that, as an evolutionary adaptation to "nutritional immunity" evolved by immune cells [78], *B. pertussis* has decayed its manganese exporter to accumulate this important metal within the cytosol, thereby enhancing its survival under oxidative stress [79]. Furthermore, a neighbouring *BP3083* gene transcribed opposite to this operon encodes a Fur family transcriptional regulator. In contrast to adjacent operon, the expression of the regulatory gene was not affected upon internalization. It will be of utmost importance to decipher the role of this locus in the adaptation of *B. pertussis* to the intramacrophage environment.

Transcriptomic analysis also revealed several significantly modulated non-coding RNAs previously identified in *B. pertussis* [45]. Among the strongly downregulated candidate transcripts, we identified CT_433 and CT_233, whereas the expression of CT_347, CT_481, CT_483, and CT_531 was significantly increased during infection (see Table S1). Of particular note, the CT_433 transcript was recently identified also by bioinformatics (referred to as the S17 transcript) and subsequent analyzes showed that this sRNA represents a Bvg-activated gene [47]. In support, expression of the CT _433 sRNA gene was

among the most downregulated genes in our analysis ($\log_2$FC < −10), providing further evidence that the intramacrophage environment is inducible for Bvg$^-$ mode.

Importantly, our results may well explain the preservation of the seemingly superfluous Bvg$^-$ phase in the infection cycle of *B. pertussis*. The Bvg system in the closely related mammalian pathogen *B. bronchiseptica* has been shown to contribute to survival in nutrient-poor environments outside the host, as well as to intracellular survival and replication within amoebae [24,80]. *B. pertussis* represents a strictly human pathogen that does not survive outside its host, and thus it has been hypothesized that the Bvg$^-$ phase represents an evolutionary remnant that is otherwise useless in the infection cycle of this pathogen [8,10]. Alternatively, it has been speculated that the avirulent phase may play a role in either the transmission of the disease (aerosol route) associated with the temperature drop [9,81] or the survival and persistence of the bacterium in the host [27]. Regarding the first alternative, we have shown that, unlike *B. bronchiseptica*, the activity of the *bvg* system in *B. pertussis* is only partially reduced at low temperatures, allowing the production of virulence factors at temperatures below 25°C [82]. We hypothesized that *B. pertussis* cells do not switch to Bvg$^-$ mode during host-to-host transmission, and maintain production of virulence factors required for reinfection. Thus, we speculate that similar to *B. bronchiseptica*, the Bvg$^-$ mode serves *B. pertussis* cells to adapt to and survive in a hostile and nutrient-poor niche such as the intramacrophage environment. This concept is also congruent with other observations showing that Bvg$^-$ mutants accumulate among persistent *B. pertussis* cells within the upper respiratory tract of experimentally infected rhesus monkeys [83] and that a spontaneous mutant of *B. pertussis* producing highly reduced levels of virulence factors was not attenuated in survival in primary human macrophages [84]. Moreover, several Bvg-activated genes were downregulated and, *vice versa*, several Bvg-repressed genes were induced during *B. pertussis* infection of the respiratory tract of mice [30,31].

Nevertheless, not all effects observed in intracellular cells can be simply explained by the shift to Bvg$^-$ phase. First, we identified a large number of genes that were strongly modulated in internalized cells but were not reported as Bvg-dependent. Among these, we found genes encoding electron transport chain complexes, transport systems, ribosomal proteins, and various transcriptional regulators. Second, for some previously described Bvg-dependent genes, we observed either negligible change in expression or even opposite changes than expected. For example, the *btrS* gene encodes a transcription factor responsible for expression of the T3SS locus, which is highly upregulated in Bvg$^+$ cells [9]. However, the expression of *btrS* was not affected at all in internalized cells, whereas the expression of almost the entire T3SS locus was strongly downregulated, as expected for Bvg-activated genes. Most strikingly, the operon required for lipid A remodelling (*BP0396–0399*) consists of Bvg-activated genes [9] and therefore, should be downregulated in the Bvg$^-$ mode. Nevertheless, the expression of this locus was greatly increased in intracellular *B. pertussis* cells. Altogether, these observations indicate that other regulators besides the BvgAS system are involved in adaptation to the intramacrophage environment.

Overall, the data presented here demonstrate that *B. pertussis* cells actively adapt to the intramacrophage environment; suggest that the intracellularity of *B. pertussis* is in the context of its lifestyle; and expose the avirulent phase as an authentic phenotype of internalized *B. pertussis* cells.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by grant 23-05634S (to B.V.) from the Czech Science Foundation (www.gacr.cz), GAUK grant 301221 from the Grant Agency of Charles University in Prague (to A.S.), by funding from RVO61388971 and from the project National Institute of Virology and Bacteriology (Programme EXCELES, ID Project No. LX22NPO5103) funded by the European Union – Next Generation EU.

Data availability statement

RNA-seq data from the sequencing runs were deposited in the European Nucleotide Archive under project accession number PRJEB33395 and will be available immediately from the date of publication.

ORCID

Mariam R. Farman http://orcid.org/0000-0002-4419-8421
Denisa Petráčková http://orcid.org/0000-0002-6207-2099
Dilip Kumar http://orcid.org/0000-0003-4566-1446
Jakub Držmíšek http://orcid.org/0000-0002-8585-3189
Argha Saha http://orcid.org/0000-0002-4031-3939
Ivana Čurnová http://orcid.org/0000-0002-5372-6337
Jan Čapek http://orcid.org/0000-0001-5094-6812
Václava Hejnarová http://orcid.org/0000-0003-3476-8125
Fabian Amman http://orcid.org/0000-0002-8646-859X
Ivo Hofacker http://orcid.org/0000-0001-7132-0800
Branislav Večerek http://orcid.org/0000-0001-9520-6324

References

[1] Mattoo S, Cherry JD. Molecular pathogenesis, epidemiology, and clinical manifestations of respiratory infections due to Bordetella pertussis and other Bordetella subspecies. Clin Microbiol Rev. 2005 Apr;18(2):326–382.

[2] Black RE, Cousens S, Johnson HL, et al. Global, regional, and national causes of child mortality in 2008: a systematic analysis. Lancet. 2010 Jun 5;375 (9730):1969–1987.
[3] Belcher T, Dubois V, Rivera-Millot A, et al. Pathogenicity and virulence of Bordetella pertussis and its adaptation to its strictly human host. Virulence. 2021 Dec;12(1):2608–2632.
[4] Higgs R, Higgins SC, Ross PJ, et al. Immunity to the respiratory pathogen Bordetella pertussis. Mucosal Immunol. 2012 Sep;5(5):485–500.
[5] de Gouw D, Diavatopoulos DA, Bootsma HJ, et al. Pertussis: a matter of immune modulation. FEMS Microbiol Rev. 2011 May;35(3):441–474.
[6] Novak J, Cerny O, Osickova A, et al. Structure-Function relationships underlying the capacity of Bordetella adenylate cyclase toxin to disarm host phagocytes. Toxins (Basel). 2017 Sep 24;9(10): toxins9100300.
[7] Cotter PA, Jones AM. Phosphorelay control of virulence gene expression in Bordetella. Trends Microbiol. 2003 Aug;11(8):367–373.
[8] Merkel TJ, Stibitz S, Keith JM, et al. Contribution of regulation by the bvg locus to respiratory infection of mice by Bordetella pertussis. Infect Immun. 1998 Sep;66(9):4367–4373.
[9] Moon K, Bonocora RP, Kim DD, et al. The BvgAS regulon of Bordetella pertussis. MBio. 2017 Oct 10;8(5): e01526-17.
[10] de Tejada G M, Cotter PA, Heininger U, et al. Neither the Bvg- phase nor the vrg6 locus of Bordetella pertussis is required for respiratory infection in mice. Infect Immun. 1998 Jun;66(6):2762–2768.
[11] Lenz DH, Weingart CL, Weiss AA. Phagocytosed Bordetella pertussis fails to survive in human neutrophils. Infect Immun. 2000 Feb;68(2):956–959.
[12] Rodriguez ME, Hellwig SM, Hozbor DF, et al. Fc receptor-mediated immunity against Bordetella pertussis. J Immunol. 2001 Dec 1;167(11):6545–6551.
[13] Ishibashi Y, Claus S, Relman DA. Bordetella pertussis filamentous hemagglutinin interacts with a leukocyte signal transduction complex and stimulates bacterial adherence to monocyte CR3 (CD11b/CD18). J Exp Med. 1994 Oct 1;180(4):1225–1233.
[14] Saukkonen K, Cabellos C, Burroughs M, et al. Integrin-mediated localization of Bordetella pertussis within macrophages: role in pulmonary colonization. J Exp Med. 1991 May 1;173(5):1143–1149.
[15] Hazenbos WL, van den Berg BM, van Furth R. Very late antigen-5 and complement receptor type 3 cooperatively mediate the interaction between Bordetella pertussis and human monocytes. J Immunol. 1993 Dec 1;151(11):6274–6282.
[16] Friedman RL, Nordensson K, Wilson L, et al. Uptake and intracellular survival of Bordetella pertussis in human macrophages. Infect Immun. 1992 Nov;60 (11):4578–4585.
[17] Paddock CD, Sanden GN, Cherry JD, et al. Pathology and pathogenesis of fatal Bordetella pertussis infection in infants. Clin Infect Dis. 2008 Aug 1;47 (3):328–338.
[18] Hellwig SM, Hazenbos WL, van de Winkel JG, et al. Evidence for an intracellular niche for Bordetella pertussis in broncho-alveolar lavage cells of mice. FEMS Immunol Med Microbiol. 1999 Dec;26(3-4):203–207.
[19] Vandebriel RJ, Hellwig SM, Vermeulen JP, et al. Association of Bordetella pertussis with host immune cells in the mouse lung. Microb Pathog. 2003 Jul;35 (1):19–29.
[20] Lamberti YA, Hayes JA, Perez Vidakovics ML, et al. Intracellular trafficking of Bordetella pertussis in human macrophages. Infect Immun. 2010 Mar;78 (3):907–913.
[21] Lamberti Y, Cafiero JH, Surmann K, et al. Proteome analysis of Bordetella pertussis isolated from human macrophages. J Proteomics. 2016 Mar 16;136:55–67.
[22] Valdez HA, Oviedo JM, Gorgojo JP, et al. Bordetella pertussis modulates human macrophage defense gene expression. Pathog Dis. 2016 Aug;74(6):ftw073.
[23] Petrackova D, Farman MR, Amman F, et al. Transcriptional profiling of human macrophages during infection with Bordetella pertussis. RNA Biol. 2020 May;17(5):731–742.
[24] Taylor-Mulneix DL, Bendor L, Linz B, et al. Bordetella bronchiseptica exploits the complex life cycle of Dictyostelium discoideum as an amplifying transmission vector. PLoS Biol. 2017 Apr;15(4): e2000420.
[25] Gorgojo J, Harvill ET, Rodriguez ME. Bordetella parapertussis survives inside human macrophages in lipid raft-enriched phagosomes. Infect Immun. 2014 Dec;82(12):5175–5184.
[26] Rivera I, Linz B, Dewan KK, et al. Conservation of ancient genetic pathways for intracellular persistence among animal pathogenic Bordetellae. Front Microbiol. 2019;10:2839.
[27] Rivera I, Linz B, Harvill ET. Evolution and conservation of Bordetella intracellular survival in eukaryotic host cells. Front Microbiol. 2020;11:557819.
[28] Banus S, Pennings J, Vandebriel R, et al. Lung response to Bordetella pertussis infection in mice identified by gene-expression profiling. Immunogenetics. 2007 Jul;59(7):555–564.
[29] Raeven RH, Brummelman J, Pennings JL, et al. Molecular signatures of the evolving immune response in mice following a Bordetella pertussis infection. PLoS One. 2014;9(8):e104548.
[30] van Beek LF, de Gouw D, Eleveld MJ, et al. Adaptation of Bordetella pertussis to the respiratory tract. J Infect Dis. 2018 May 25;217(12):1987–1996.
[31] Wong TY, Hall JM, Nowak ES, et al. Analysis of the in vivo transcriptome of Bordetella pertussis during infection of mice. mSphere. 2019 Apr 17;4(2):e00154-19.
[32] Chanput W, Mes JJ, Wichers HJ. THP-1 cell line: an in vitro cell model for immune modulation approach. Int Immunopharmacol. 2014 Nov;23(1):37–45.
[33] Stainer DW, Scholte MJ. A simple chemically defined medium for the production of phase I Bordetella pertussis. J Gen Microbiol. 1970 Oct;63(2):211–220.
[34] Bibova I, Skopova K, Masin J, et al. The RNA chaperone Hfq is required for virulence of Bordetella pertussis. Infect Immun. 2013 Nov;81(11):4081–4090.
[35] Inatsuka CS, Xu Q, Vujkovic-Cvijin I, et al. Pertactin is required for Bordetella species to resist neutrophil-mediated clearance. Infect Immun. 2010 Jul;78 (7):2901–2909.
[36] Patro R, Duggal G, Love MI, et al. Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods. 2017 Apr;14(4):417–419.
[37] Risso D, Ngai J, Speed TP, et al. Normalization of RNA-seq data using factor analysis of control genes or samples. Nat Biotechnol. 2014 Sep;32 (9):896–902.

[38] Robinson MD, McCarthy DJ, Smyth GK. Edger: a bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010 Jan 1;26(1):139–140.
[39] Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.
[40] Wickham H. Ggplot2: elegant graphics for data analysis. New York (NY): Springer; 2009.
[41] Conesa A, Gotz S, Garcia-Gomez JM, et al. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics. 2005 Sep 15;21(18):3674–3676.
[42] Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B. 1995;57:289–300.
[43] Bibova I, Hot D, Keidel K, et al. Transcriptional profiling of Bordetella pertussis reveals requirement of RNA chaperone Hfq for type III secretion system functionality. RNA Biol. 2015;12(2):175–185.
[44] Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res. 2001 May 1;29(9):e45.
[45] Amman F, D'Halluin A, Antoine R, et al. Primary transcriptome analysis reveals importance of IS elements for the shaping of the transcriptional landscape of Bordetella pertussis. RNA Biol. 2018 Apr 23;15(7):967–975.
[46] Grant RA, Filman DJ, Finkel SE, et al. The crystal structure of Dps, a ferritin homolog that binds and protects DNA. Nat Struct Biol. 1998 Apr;5(4):294–303.
[47] Moon K, Sim M, Tai CH, et al. Identification of BvgA-dependent and BvgA-independent small RNAs (sRNAs) in Bordetella pertussis using the prokaryotic sRNA prediction toolkit ANNOgesic. Microbiol Spectr. 2021 Oct 31;9(2):e0004421.
[48] Coutte L, Huot L, Antoine R, et al. The multifaceted RisA regulon of Bordetella pertussis. Sci Rep. 2016 Sep 13;6:32774.
[49] Lamberti Y, Gorgojo J, Massillo C, et al. Bordetella pertussis entry into respiratory epithelial cells and intracellular survival. Pathog Dis. 2013 Dec;69 (3):194–204.
[50] Lamberti Y, Surmann K. The intracellular phase of extracellular respiratory tract bacterial pathogens and its role on pathogen-host interactions during infection. Curr Opin Infect Dis. 2021 Jun 1;34(3):197–205.
[51] Nakayama T, Zhang-Akiyama QM. pqiABC and yebST, putative mce operons of escherichia coli, encode transport pathways and contribute to membrane integrity. J Bacteriol. 2017 Jan 1;199(1): e00606-16.
[52] Brenot A, King KY, Janowiak B, et al. Contribution of glutathione peroxidase to the virulence of Streptococcus pyogenes. Infect Immun. 2004 Jan;72 (1):408–413.
[53] Guo L, Lim KB, Poduje CM, et al. Lipid A acylation and bacterial resistance against vertebrate antimicrobial peptides. Cell. 1998 Oct 16;95(2):189–198.
[54] Shah NR, Hancock RE, Fernandez RC. Bordetella pertussis lipid A glucosamine modification confers resistance to cationic antimicrobial peptides and increases resistance to outer membrane perturbation. Antimicrob Agents Chemother. 2014 Aug;58 (8):4931–4934.
[55] Mortuza R, Aung HL, Taiaroa G, et al. Overexpression of a newly identified d-amino acid transaminase in mycobacterium smegmatis complements glutamate racemase deletion. Mol Microbiol. 2018 Jan;107 (2):198–213.
[56] Taneja NK, Ganguly T, Bakaletz LO, et al. D-alanine modification of a protease-susceptible outer membrane component by the Bordetella pertussis dra locus promotes resistance to antimicrobial peptides and polymorphonuclear leukocyte-mediated killing. J Bacteriol. 2013 Nov;195(22):5102–5111.
[57] Guest RL, Wang J, Wong JL, et al. A bacterial stress response regulates respiratory protein complexes To control envelope stress adaptation. J Bacteriol. 2017 Oct 15;199(20):e00433-17.
[58] Raivio TL, Silhavy TJ. Transduction of envelope stress in Escherichia coli by the Cpx two-component system. J Bacteriol. 1997 Dec;179(24):7724–7733.
[59] van der Heijden J, Reynolds LA, Deng W, et al. Salmonella rapidly regulates membrane permeability To survive oxidative stress. mBio. 2016 Aug 9;7(4): e01238-16.
[60] Cafiero JH, Lamberti YA, Surmann K, et al. A Bordetella pertussis MgtC homolog plays a role in the intracellular survival. PLoS One. 2018;13(8): e0203204.
[61] Alix E, Blanc-Potard AB. Mgtc: a key player in intramacrophage survival. Trends Microbiol. 2007 Jun;15 (6):252–256.
[62] Lee EJ, Pontes MH, Groisman EA. A bacterial virulence protein promotes pathogenicity by inhibiting the bacterium's own F1Fo ATP synthase. Cell. 2013 Jul 3;154(1):146–156.
[63] Ochrombel I, Ott L, Kramer R, et al. Impact of improved potassium accumulation on pH homeostasis, membrane potential adjustment and survival of Corynebacterium glutamicum. Biochim Biophys Acta. 2011 Apr;1807(4):444–450.
[64] Schnappinger D, Ehrt S, Voskuil MI, et al. Transcriptional adaptation of mycobacterium tuberculosis within macrophages: insights into the phagosomal environment. J Exp Med. 2003 Sep 1;198 (5):693–704.
[65] McKinney JD, Honer zu Bentrup K, Munoz-Elias EJ, et al. Persistence of Mycobacterium tuberculosis in macrophages and mice requires the glyoxylate shunt enzyme isocitrate lyase. Nature. 2000 Aug 17;406 (6797):735–738.
[66] Anaya-Sanchez A, Feng Y, Berude JC, et al. Detoxification of methylglyoxal by the glyoxalase system is required for glutathione availability and virulence activation in Listeria monocytogenes. PLoS Pathog. 2021 Aug;17(8):e1009819.
[67] Antoine R, Huvent I, Chemlal K, et al. The periplasmic binding protein of a tripartite tricarboxylate transporter is involved in signal transduction. J Mol Biol. 2005 Aug 26;351(4):799–809.
[68] Huvent I, Belrhali H, Antoine R, et al. Crystal structure of Bordetella pertussis BugD solute receptor unveils the basis of ligand binding in a new family of periplasmic binding proteins. J Mol Biol. 2006 Mar 3;356(4):1014–1026.
[69] Antoine R, Jacob-Dubuisson F, Drobecq H, et al. Overrepresentation of a gene family encoding extracytoplasmic solute receptors in Bordetella. J Bacteriol. 2003 Feb;185(4):1470–1474.

[70] Parkhill J, Sebaihia M, Preston A, et al. Comparative analysis of the genome sequences of Bordetella pertussis, Bordetella parapertussis and Bordetella bronchiseptica. Nat Genet. 2003 Sep;35(1):32–40.
[71] Park S, Imlay JA. High levels of intracellular cysteine promote oxidative DNA damage by driving the fenton reaction. J Bacteriol. 2003 Mar;185(6):1942–1950.
[72] Deshpande AA, Bhatia M, Laxman S, et al. Thiol trapping and metabolic redistribution of sulfur metabolites enable cells to overcome cysteine overload. Microb Cell. 2017 Mar 27;4(4):112–126.
[73] Dominy JE J, Simmons CR, Karplus PA, et al. Identification and characterization of bacterial cysteine dioxygenases: a new route of cysteine degradation for eubacteria. J Bacteriol. 2006 Aug;188 (15):5561–5569.
[74] Stipanuk MH, Dominy JE Jr., Lee JI, et al. Mammalian cysteine metabolism: new insights into regulation of cysteine metabolism. J Nutr. 2006 Jun;136(6 Suppl):1652S–1659S.
[75] Hennicke F, Grumbt M, Lermann U, et al. Factors supporting cysteine tolerance and sulfite production in Candida albicans. Eukaryot Cell. 2013 Apr;12 (4):604–613.
[76] Brickman TJ, Hanawa T, Anderson MT, et al. Differential expression of Bordetella pertussis iron transport system genes during infection. Mol Microbiol. 2008 Oct;70(1):3–14.
[77] Zhou D, Hardt WD, Galan JE. Salmonella typhimurium encodes a putative iron transport system within the centisome 63 pathogenicity island. Infect Immun. 1999 Apr;67(4):1974–1981.
[78] Kehl-Fie TE, Skaar EP. Nutritional immunity beyond iron: a role for manganese and zinc. Curr Opin Chem Biol. 2010 Apr;14(2):218–224.
[79] Capek J, Prochazkova I, Matousek T, et al. A unique reverse adaptation mechanism assists Bordetella pertussis in resistance to both scarcity and toxicity of manganese. mBio. 2021 Oct 26;12(5):e0190221.
[80] Porter JF, Wardlaw AC. Long-term survival of Bordetella bronchiseptica in lakewater and in buffered saline without added nutrients. FEMS Microbiol Lett. 1993 Jun 1;110(1):33–36.
[81] Trainor EA, Nicholson TL, Merkel TJ. Bordetella pertussis transmission. Pathog Dis. 2015 Nov 73;8, ftv068.
[82] Seydlova G, Beranova J, Bibova I, et al. The extent of the temperature-induced membrane remodeling in two closely related Bordetella species reflects their adaptation to diverse environmental niches. J Biol Chem. 2017 May 12;292(19):8048–8058.
[83] Karataev GI, Sinyashina LN, Medkova AY, et al. Insertional inactivation of virulence operon in population of persistent Bordetella pertussis bacteria. Genetika. 2016 Apr;52(4):422–430.
[84] Novak J, Jurnecka D, Linhartova I, et al. A mutation upstream of the rplN-rpsD ribosomal operon downregulates Bordetella pertussis virulence factor production without compromising bacterial survival within human macrophages. mSystems. 2020 Dec 8;5(6):e00612-20.

6

SUMMARY AND OUTLOOK

Pertussis is reemerging as a significant public health concern in many regions worldwide despite the widespread use of vaccines. While considerable research has been conducted on this pathogen, and much is understood, the resurgence of pertussis highlights existing gaps in our knowledge regarding the fundamental aspects of *B. pertussis* virulence and infection, host immune responses, and the pathogenesis of pertussis disease. Recent studies on *B. pertussis* have predominantly aimed to elucidate the complex interactions between the pathogen and the host during infection. This involves characterizing bacterial surface structures, determining host specificity, and analyzing host immune responses throughout the disease progression. For this, we need to study the interaction between hosts and pathogens by analyzing their transcriptomes simultaneously, which has been a challenging task in infection biology. This was mainly due to the limitations of array-based techniques, which cannot simultaneously capture and analyze prokaryotic and eukaryotic transcripts from model systems due to the finite dynamic range. However, the development of RNA-seq methodology in recent years has enabled the simultaneous profiling of RNA from various host-pathogen systems. This has led to new insights into our understanding of infection processes [70, 74].

Our research is the first to utilize dual RNA-seq analysis to create detailed transcriptome profiles of *B. pertussis* inside human macrophages. This simultaneous profiling helped us understand how each organism responds to the presence of the other in real-time, providing valuable insights into the dynamic interaction between THP-1 and *B. pertussis*. Our study reveals the ability to detect THP-1 and *B. pertussis* transcriptomes simultaneously, allowing us to profile gene expression in both organisms accurately. We obtained a significant number of bacterial reads from infected macrophage cells without using enrichment steps and without incurring excessive sequencing costs. The infection protocol and the dual RNA-seq pipeline are easy to control and highly reproducible, making them a valuable baseline for other research groups.

We aimed to elucidate the nature of the interactions of *B. pertussis* within its natural host. To date, studies investigating the transcriptional profiling of *B. pertussis* interactions with host cells have primarily utilized murine models, which do not accurately represent the natural host of *B. pertussis*. While these studies have provided valuable insights into the interaction of *B. pertussis* with the respiratory mucosa, the use of animal models may be limited due to inter-species disparities [81]. This issue holds particular significance as animal models may not faithfully replicate the human immune response and disease progression. Hence, it was imperative to examine *B. pertussis* within

its natural host to gain insights into the dynamics of natural infection and the potential evolution or adaptation of *B. pertussis* in response to human immune pressures. To determine whether the intracellular phase of *B. pertussis* was a lab-induced artifact; our objective was to investigate the infection process of *B. pertussis* in its natural host, with a particular emphasis on its intracellular phase. Rodriguez and team proposed that *B. pertussis* uses macrophages as a cell-internal environment, suggesting that the intracellular stage of infection could be essential for the bacteria's survival and enduring presence in the host. [152, 153].

To fully understand how the reemerging pathogen *B. pertussis* can survive in this particular environment; we used a time-resolved dual RNA-seq method to monitor its activity. In our initial investigation, we observed significant changes in the gene expression of human macrophages after infection. The results indicated that the induction of pro- and anti-inflammatory cytokines, along with heightened expression of genes associated with the polarization of M1 macrophages, promotion of pro-inflammatory Th1 and Th17 immune responses, i.e., IL-23, IL-1A, IL6, CCR7, CCL2, CCL3, CXCL8, CD80, and CD86, as well as the M2 program in macrophages and the corresponding Th2 response including genes IL-13 and IL-10. Furthermore, the study identified a notable upregulation of suppressors of cytokine signaling proteins such as SOCS3 and SOCS2 by *B. pertussis* as an evasion strategy against the host's immune system. While the investigation focused on a specific time point (6hpi), it revealed reduced expression of several pathogenesis-associated genes during infection. Notably, the study highlighted the manipulation of apoptosis-related genes by *B. pertussis* within macrophages, inhibiting cell death and facilitating pathogen survival. This distinctive mechanism was also observed in human respiratory epithelial cells infected with *B. pertussis* [85], underscoring its potential as a characteristic molecular response to such infections. Notably, the expression of serpins, proteins that regulate various biological processes, including inflammation, increased substantially upon infection. Specifically, the upregulation of SERPINB3, SERPINB4, and BIRC3 (cIAP2) was identified, indicating their role in inhibiting apoptosis and cell death. It was also seen that various metabolic and signaling pathways are significantly impaired during infection with *B. pertussis* i.e., the decrease in mTOR and insulin signaling pathways, fatty acid metabolism, TCA cycle, and oxidative phosphorylation. On the other hand, NOD-like receptor and chemokine signaling pathways were stimulated. In addition, our studies showed that the apoptosis, Toll-like receptor, and Jak-STAT signaling pathways were simultaneously upregulated and downregulated. These results suggest that *B. pertussis* not only alters signaling pathways that control the immune response but also has profound effects on host metabolism. The observations above sparked speculation that *B. pertussis* utilizes macrophages as an intracellular niche for its survival and persistence within the host achieved through a transition from the Bvg+ to the Bvg- phase. Consequently, we recognized the need for further experiments encompassing

a broader spectrum of infection time points and enhanced sequencing depth to elucidate the specific activities of *B. pertussis* that influence the host immune response and the responses of macrophages to *B. pertussis* infection. Thus, a comprehensive investigation was undertaken to elucidate the significance of the intracellular phase in the infectious cycle of *B. pertussis*.

To further investigate *B. pertussis*'s adaptation to the intramacrophage environment, we conducted a comprehensive dual RNA-seq experiment focusing on the intracellular escape and survival of the bacterium. Although Bordetella pertussis is typically categorized as an extracellular pathogen, our investigation has demonstrated the ability of intracellular *B. pertussis* to survive and repopulate in an antibiotic-free medium during infection. The intracellular persistence suggests that the evolutionary development of a specific gene expression mechanism within Bordetella species has enabled adaptation to the environment within macrophages. We speculated that this adaptation is associated with the transition from the Bvg+ to the Bvg- phase, where the genes that are typically downregulated by the BvgAS system (Bvg-) are expected to be upregulated, while conversely, genes activated by the BvgAS system (Bvg+) should be downregulated. This phenomenon formed the focus of our inquiry. Interestingly, we found that from the 322 of the Bvg- genes, 207 (64.2%) exhibited significant upregulation in at least one time point within internalized cells. In contrast, 172 out of 237 Bvg+ genes (72.5%) demonstrated significant downregulation within intracellular bacteria. We observed consistent downregulation of PTX components used by *B. pertussis* in the Bvg+ phase, specifically *ptxA*, *ptxB*, *ptxC*, *ptxD*, and *ptxE*, throughout the differential gene expression analysis at various infection time points. Similarly, the gene encoding adenylyl cyclase toxin (ACT), cyaA, crucial for toxin activity, decreased during infection, as indicated by the DGEA. This signifies the prompt adaptability of *B. pertussis* cells post-internalization and suggests that upon internalization within a host cell, *B. pertussis* undergoes a transition to the avirulent Bvg- phase.

The Bvg- phase is essential for the bacterium's survival, persistence, and replication [150]. Once inside the challenging macrophage environment, *B. pertussis* adapts a strategic survival mechanism. Our observations indicated that RisA, essential for *B. pertussis* persistence significantly influences gene expression. Research revealed that 38% of the genes activated by RisA exhibited increased expression levels, while 42% of the genes repressed by RisA showed decreased expression levels. Although our data did not show RisA as being dysregulated, it is reasonable to assume that its phosphorylation status, which affects its activity, cannot be directly inferred from RNA-seq data, as RNA-seq measures transcript levels rather than post-translational modifications such as phosphorylation. Additionally, genes related to oxidative stress, iron ion transport, membrane stress, and sulfur metabolism were mostly upregulated, while genes associated with pH homeostasis were suppressed by *B. pertussis*. This indicates that the Bvg+ phase plays a limited role in *B. pertussis*'s intracellular survival, while the avirulent Bvg- phase seems advantageous.

Our research findings, as detailed in the two studies, provide compelling evidence that *B. pertussis* cells have the ability to actively modify their behavior in response to the environment within macrophages, characterized by a switch from a virulent to an avirulent phase, which is accompanied by significant changes in the cell's metabolic and energy production processes. These intracellular survival strategies may facilitate the persistence of *B. pertussis* in vaccinated populations, pertussis enables the bacterium to adapt its behavior to evade immune clearance, thus contributing to its resurgence. In summary, the study suggests the ongoing evolution of *B. pertussis* cells could have a significant impact on the management of pertussis and may require the development of new strategies to control the disease.

In this specific study, one of the objectives was to construct inter-species gene coexpression networks between the host THP-1 and *B. pertussis* genes utilizing the novel MSF tool developed during my doctoral studies. MSF uses the approach of using p-values of individual genes obtained from the DGEA. These individual gene p-values are combined to form a combined p-value for the sub-graph. We successfully constructed coexpression networks for host THP1 genes using the MSF software and biological pathways and interactions from Reactome. The next step involved creating a correlation matrix using the read counts of THP1 and *B. pertussis* genes. This correlation matrix would connect the dis-regulated sub-graphs from THP-1 and *B. pertussis*. Establishing coexpression networks between different species proved to be quite challenging. Initially, we acknowledged the absence of pathway annotations for *B. pertussis*, which hindered our ability to generate modulated sub-graphs for this organism. Furthermore, we constructed a correlation matrix for THP1 and *B. pertussis* genes, we did not identify any significant or confirmed targets. This may be attributed to the complexity of pertussis as a disease, which is influenced by multiple factors rather than a single gene or target. We found that several human genes strongly correlate with *B. pertussis* genes. However, we faced a considerable challenge: many *B. pertussis* genes lacked functional annotations and were classified as hypothetical proteins.

Our comprehensive study found that one of the bacterioferritin genes in *B. pertussis BP0546*, exhibited a strong positive correlation with several innate immune response cytokines and chemokines, specifically IL23A, IL15, and IL11. This observation indicates that these cytokines also increase as bacterioferritin levels rise to acquire iron. Macrophages are vital for defending the host and regulating iron levels by phagocytosing aging red blood cells and recycling their iron content. Hepcidin (HAMP) serves as the main hormone managing iron balance, reducing iron absorption from the diet and controlling its release from macrophages. However, we noted that HAMP was downregulated by less than onefold at the first two and last two time points, although it showed a slight upregulation at time point 3. Additionally, we observed that several Transposase elements associated with IS481, including *BP3311*, *BP3203*, *BP2087*, and *BP0871*, demonstrated a strong positive correlation with IL23A

and IL6. Unfortunately, we were unable to derive any significant conclusions from this correlation.

Furthermore, *BP1178*, a drug/metabolite transporter, revealed a strong negative correlation with CAMK1G, IL1A, CXCL2, CXCL3, and IL10. Regrettably, we were unable to further test these correlations in the lab further due to time and resource limitations. Our collaborator compiled the Bordetella annotation file used in our analysis through a literature review, which helps to explain why many of the dysregulated genes could not be verified and why their function remained unknown.

Another key function of MSF is its ability to identify critical components within biological pathways, such as sensors, effectors, and hubs, as well as potential points for intervention. MSF pinpointed several candidate genes from the human biological system; these candidate genes hold promise for further investigation. This helpful functionality of MSF has the potential to benefit researchers significantly by identifying possible target areas for potential therapeutic intervention within the biological pathway. Doing so paves the way for more focused testing and development of targeted pharmaceutical interventions.

We applied the dual RNAseq methodology to investigate the interaction between THP-1 macrophages and *B. pertussis*, which has yielded favorable outcomes. Our approach's meticulous planning facilitated the extraction of RNA from both the host organism and the bacterium while minimizing sequencing expenses. As a result, we identified subtle and substantial alterations in gene expression within THP-1 macrophages and *B. pertussis* throughout the infection process. Using the dual RNAseq method, this research could provide a framework for other investigators seeking to study their respective bacteria within the host.

BIBLIOGRAPHY

[1] M. Adams et al. "Complementary DNA sequencing: expressed sequence tags and human genome project." In: *Science* 252.5013 (1991), 1651–1656. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.2047873.

[2] S. Anders, P. T. Pyl, and W. Huber. "HTSeq–a Python framework to work with high-throughput sequencing data." In: *Bioinformatics* 31.2 (2014), pp. 166–169. DOI: 10.1093/bioinformatics/btu638.

[3] Charlotte Andreasen, Daniel A. Powell, and Nicholas H. Carbonetti. "Pertussis toxin stimulates IL-17 production in response to bordetella pertussis infection in mice." In: *PLoS ONE* 4.9 (2009). DOI: 10.1371/journal.pone.0007079.

[4] Simon Andrews, Felix Krueger, Anne Segonds-Pichon, Laura Biggins, Christel Krueger, and Steven Wingett. *FastQC*. Babraham Institute. Babraham, UK, Jan. 2012.

[5] Apache. *Apache Commons CLI*. https://commons.apache.org/proper/commons-cli/. accessed 2023.

[6] Laura Arnal, Tom Grunert, Natalia Cattelan, Daan de Gouw, María I. Villalba, Diego O. Serra, Frits R. Mooi, Monika Ehling-Schulz, and Osvaldo M. Yantorno. "Bordetella pertussis Isolates from Argentinean Whooping Cough Patients Display Enhanced Biofilm Formation Capacity Compared to Tohama I Reference Strain." In: *Frontiers in Microbiology* 6 (2015), p. 1352. ISSN: 1664-302X. DOI: 10.3389/fmicb.2015.01352. URL: https://www.frontiersin.org/article/10.3389/fmicb.2015.01352.

[7] Michael Ashburner, Catherine A. Ball, Judith A. Blake, David Botstein, Heather Butler, J. Michael Cherry, Allan P. Davis, Kara Dolinski, Selina S. Dwight, and Janan T. et al. Eppig. "Gene Ontology: tool for the unification of biology." In: *Nature Genetics* 25.1 (2000), pp. 25–29. DOI: 10.1038/75556.

[8] Clara Maria Ausiello and Antonio Cassone. "Acellular Pertussis Vaccines and Pertussis Resurgence: Revise or Replace?" In: *mBio* 5.3 (2014), e01339–14. ISSN: 2150-7511. DOI: 10.1128/mBio.01339-14.

[9] J Barbic, M F Leef, D L Burns, and R D Shahin. "Role of gamma interferon in natural clearance of Bordetella pertussis infection." In: *Infection and immunity* 65.12 (1997), 4904–4908. ISSN: 0019-9567. DOI: 10.1128/IAI.65.12.4904-4908.1997.

[10] Alex-Mikael Barkoff, Jussi Mertsola, Sophie Guillot, Nicole Guiso, Guy Berbers, and Qiushui He. "Appearance of bordetella pertussis strains not expressing the vaccine antigen pertactin in Finland." In: *Clinical and Vaccine Immunology* 19.10 (2012), 1703–1704. DOI: 10.1128/cvi.00367-12.

[11] Michael G. Barnes and Alison A. Weiss. "Brka protein ofbordetella pertussisinhibits the classical pathway of complement after C1 deposition." In: *Infection and Immunity* 69.5 (2001), 3067–3072. DOI: 10.1128/iai.69.5.3067-3072.2001.

[12] Michael G Barnes and Alison A Weiss. "Activation of the complement cascade bybordetella pertussis." In: *FEMS Microbiology Letters* 220.2 (2003), 271–275. DOI: 10.1016/s0378-1097(03)00132-0.

[13] Robert J. Belland, Guangming Zhong, Deborah D. Crane, Daniel Hogan, Daniel Sturdevant, Jyotika Sharma, Wandy L. Beatty, and Harlan D. Caldwell. "Genomic transcriptional profiling of the developmental cycle of Chlamydia trachomatis." In: *Proceedings of the National Academy of Sciences* 100.14 (2003), 8478–8483. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1331135100.

[14] K Berggård, E Johnsson, F R Mooi, and G Lindahl. "Bordetella pertussis binds the human complement regulator C4BP: Role of filamentous hemagglutinin." In: *Infection and Immunity* 65.9 (1997), 3638–3643. DOI: 10.1128/iai.65.9.3638-3643.1997.

[15] Sven Bergmann, Jan Ihmels, and Naama Barkai. "Similarities and differences in genome-wide expression data of six organisms." In: *PLoS Biology* 2.1 (2003). DOI: 10.1371/journal.pbio.0020009.

[16] Allan Birnbaum. "Combining Independent Tests of Significance." In: *Journal of the American Statistical Association* 49.267 (1954), p. 559. ISSN: 01621459. DOI: 10.2307/2281130.

[17] O. Bordet J. Gengou. *Le microbe de la coqueluche*. Annales de I'lnstitut Pasteur, 1906. ISBN: https://wellcomecollection.org/works/xb56g32v.

[18] D. Bottero et al. "Characterization of the immune response induced by pertussis OMVs-based vaccine." In: *Vaccine* 34.28 (2016), 3303–3309. ISSN: 0264410X. DOI: 10.1016/j.vaccine.2016.04.079.

[19] V. Bouchez, D. Brun, T. Cantinelli, G. Dore, E. Njamkepo, and N. Guiso. "First report and detailed characterization of B. Pertussis isolates not expressing pertussis toxin or pertactin." In: *Vaccine* 27.43 (2009), 6034–6041. DOI: 10.1016/j.vaccine.2009.07.074.

[20] Lucy Breakwell, Patsy Kelso, Christine Finley, Susan Schoenfeld, Brant Goode, Lara K. Misegades, Stacey W. Martin, and Anna M. Acosta. "Pertussis vaccine effectiveness in the setting of pertactin-deficient pertussis." In: *Pediatrics* 137.5 (2016). DOI: 10.1542/peds.2015-3973.

[21] Hélène Broutin, François Simondon, Pejman Rohani, Jean-François Guégan, and Bryan T. Grenfell. "Loss of immunity to pertussis in a rural community in Senegal." In: *Vaccine* 22.5–6 (2004), 594–596. DOI: 10.1016/j.vaccine.2003.07.018.

[22] Garth R. Brown et al. "Gene: A gene-centered information resource at NCBI." In: *Nucleic Acids Research* 43.D1 (2014). DOI: 10.1093/nar/gku1055.

[23] Valorie C. Burns, Elizabeth J. Pishko, Andrew Preston, Duncan J. Maskell, and Eric T. Harvill. "Role of Bordetella O Antigen in Respiratory Tract Infection." In: *Infection and Immunity* 71.1 (2003), 86–94. ISSN: 0019-9567, 1098-5522. DOI: 10.1128/IAI.71.1.86-94.2003.

[24] Nicholas H. Carbonetti, Galina V. Artamonova, Charlotte Andreasen, and Nicholas Bushar. "Pertussis Toxin and Adenylate Cyclase Toxin Provide a One-Two Punch for Establishment of Bordetella pertussis Infection of the Respiratory Tract." In: *Infection and Immunity* 73.5 (2005), 2698–2703. ISSN: 0019-9567, 1098-5522. DOI: 10.1128/IAI.73.5.2698-2703.2005.

[25] James D. Cherry. "Pertussis: the trials and tribulations of old and new pertussis vaccines." In: *Vaccine* 10.14 (1992), 1033–1038. ISSN: 0264410X. DOI: 10.1016/0264-410X(92)90113-X.

[26] James D Cherry and Christopher D Paddock. "Pathogenesis and histopathology of pertussis: Implications for immunization." In: *Expert Review of Vaccines* 13.9 (2014), 1115–1123. DOI: 10.1586/14760584.2014.935766.

[27] Thomas A. Clark, Nancy E. Messonnier, and Stephen C. Hadler. "Pertussis control: Time for something new?" In: *Trends in Microbiology* 20.5 (2012), 211–213. DOI: 10.1016/j.tim.2012.03.003.

[28] Ana Conesa, Stefan Götz, Juan Miguel García-Gómez, Javier Terol, Manuel Talón, and Montserrat Robles. "Blast2GO: A universal tool for annotation, visualization and analysis in Functional Genomics Research." In: *Bioinformatics* 21.18 (2005), 3674–3676. DOI: 10.1093/bioinformatics/bti610.

[29] Dennis L. Confer and John W. Eaton. "Phagocyte impotence caused by an invasive bacterial adenylate cyclase." In: *Science* 217.4563 (1982), 948–950. DOI: 10.1126/science.6287574.

[30] M. S. Conover, C. J. Redfern, T. Ganguly, N. Sukumar, G. Sloan, M. Mishra, and R. Deora. "BpsR Modulates Bordetella Biofilm Formation by Negatively Regulating the Expression of the Bps Polysaccharide." In: *Journal of Bacteriology* 194.2 (2012), 233–242. ISSN: 0021-9193. DOI: 10.1128/JB.06020-11.

[31] Matt S. Conover, Gina Parise Sloan, Cheraton F. Love, Neelima Sukumar, and Rajendar Deora. "The Bps polysaccharide of Bordetella pertussis promotes colonization and biofilm formation in the nose by functioning as an adhesin: Bps polysaccharide of Bordetella pertussis." In: *Molecular Microbiology* 77.6 (2010), 1439–1455. ISSN: 0950382X. DOI: 10.1111/j.1365-2958.2010.07297.x.

[32] Gene Ontology Consortium. "The Gene Ontology (GO) database and informatics resource." In: *Nucleic Acids Research* 32.90001 (2004). DOI: 10.1093/nar/gkh036.

[33] SEQC/MAQC-III Consortium. "A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium." In: *Nature Biotechnology* 32.9 (2014), 903–914. ISSN: 1087-0156, 1546-1696. DOI: 10.1038/nbt.2957.

[34] Cowell. *Essential java fast*. Springer London, 1997, 24–42.

[35] D. Croft et al. "Reactome: a database of reactions, pathways and biological processes." In: *Nucleic Acids Research* 39.Database (2011), D691–D697. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gkq1018.

[36] C. A. Cummings, H. J. Bootsma, D. A. Relman, and J. F. Miller. "Species- and strain-specific control of a complex, flexible regulon by bordetella bvgas." In: *Journal of Bacteriology* 188.5 (2006), 1775–1785. DOI: 10.1128/jb.188.5.1775-1785.2006.

[37] U Dobrindt. "Whole genome plasticity in pathogenic bacteria." In: *Current Opinion in Microbiology* 4.5 (2001), 550–557. ISSN: 13695274. DOI: 10.1016/S1369-5274(00)00250-2.

[38] Aisling Dunne, Pádraig J. Ross, Eva Pospisilova, Jiri Masin, Aoife Meaney, Caroline E. Sutton, Yoichiro Iwakura, Jurg Tschopp, Peter Sebo, and Kingston H. Mills. "Inflammasome activation by adenylate cyclase toxin directs th17 responses and protection against bordetella pertussis." In: *The Journal of Immunology* 185.3 (2010), 1711–1719. DOI: 10.4049/jimmunol.1000105.

[39] Paédraic J. Dunne, Barry Moran, Robert C. Cummins, and Kingston H. Mills. "CD11C + CD8a + dendritic cells promote protective immunity to respiratory infection with bordetella pertussis." In: *The Journal of Immunology* 183.1 (2009), 400–410. DOI: 10.4049/jimmunol.0900169.

[40] Joshua C. Eby, Casandra L. Hoffman, Laura A. Gonyar, and Erik L. Hewlett. "Review of the neutrophil response to Bordetella pertussis infection." In: *Pathogens and Disease* 73.9 (2015). DOI: 10.1093/femspd/ftv081.

[41] Kathryn M. Edwards et al. "Comparison of 13 acellular pertussis vaccines: Overview and serologic response." In: *Pediatrics* 96.3 (1995), 548–557. DOI: 10.1542/peds.96.3.548.

[42] Paul Emsley, Gerry McDermott, Ian G. Charles, Neil F. Fairweather, and Neil W. Isaacs. "Crystallographic Characterization of Pertactin, a Membrane-associated Protein from Bordetella pertussis." In: *Journal of Molecular Biology* 235.2 (1994), 772–773. ISSN: 00222836. DOI: 10.1006/jmbi.1994.1029.

[43] Susanna Esposito, Tiziano Agliardi, Anna Giammanco, Giacomo Faldella, Antonio Cascio, Samantha Bosis, Ornella Friscia, Mario Clerici, and Nicola Principi. "Long-term pertussis-specific immunity after primary vaccination with a combined diphtheria, tetanus, tricomponent acellular pertussis, and hepatitis B vaccine in comparison with that after natural infection." In: *Infection and Immunity* 69.7 (2001), 4516–4520. DOI: 10.1128/iai.69.7.4516-4520.2001.

[44] Susanna Esposito and Nicola Principi. "Prevention of pertussis: An unresolved problem." In: *Human Vaccines & Immunotherapeutics* 14.10 (2018), 2452–2459. DOI: 10.1080/21645515.2018.1480298.

[45] Susanna Esposito et al. "Pertussis prevention: Reasons for resurgence, and differences in the current acellular pertussis vaccines." In: *Frontiers in Immunology* 10 (2019). DOI: 10.3389/fimmu.2019.01344.

[46] Neil K. Fennelly, Federico Sisti, Sarah C. Higgins, Paédraig J. Ross, Han van der Heide, Frits R. Mooi, Aoife Boyd, and Kingston H. G. Mills. "Bordetella pertussis Expresses a Functional Type III Secretion System That Subverts Protective Innate and Adaptive Immune Responses." In: *Infection and Immunity* 76.3 (2008), 1257–1266. ISSN: 0019-9567, 1098-5522. DOI: 10.1128/IAI.00836-07.

[47] Theresa M. Finn and Lisa A. Stevens. "Tracheal colonization factor: a Bordetella pertussis secreted virulence determinant." In: *Molecular Microbiology* 16.4 (1995), 625–634. ISSN: 0950-382X, 1365-2958. DOI: 10.1111/j.1365-2958.1995.tb02425.x.

[48] R L Friedman, K Nordensson, L Wilson, E T Akporiaye, and D E Yocum. *Uptake and intracellular survival of bordetella pertussis in human macrophages*. 1992. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC258205/.

[49] André Fujita, Patricia Severino, Kaname Kojima, João Sato, Alexandre Patriota, and Satoru Miyano. "Functional clustering of time series gene expression data by Granger causality." In: *BMC Systems Biology* 6.1 (2012), p. 137. DOI: 10.1186/1752-0509-6-137.

[50] Manuel Garber, Manfred G Grabherr, Mitchell Guttman, and Cole Trapnell. "Computational methods for transcriptome annotation and quantification using RNA-seq." In: *Nature Methods* 8.6 (2011), 469–477. DOI: 10.1038/nmeth.1613.

[51] Miguel A. García-Campos, Jesús Espinal-Enríquez, and Enrique Hernández-Lemus. "Pathway analysis: State of the art." In: *Frontiers in Physiology* 6 (2015). DOI: 10.3389/fphys.2015.00383.

[52] C A Geuijen, R J Willems, M Bongaerts, J Top, H Gielen, and F R Mooi. "Role of the Bordetella pertussis minor fimbrial subunit, FimD, in colonization of the mouse respiratory tract." In: *Infection and immunity* 65.10 (1997), 4222–4228. ISSN: 0019-9567. DOI: 10.1128/IAI.65.10.4222-4228.1997.

[53] C A Geuijen, R J Willems, and F R Mooi. "The major fimbrial subunit of Bordetella pertussis binds to sulfated sugars." In: *Infection and immunity* 64.7 (1996), 2657–2665. ISSN: 0019-9567. DOI: 10.1128/IAI.64.7.2657-2665.1996.

[54] Sara Goodwin, John D. McPherson, and W. Richard McCombie. "Coming of age: ten years of next-generation sequencing technologies." In: *Nature Reviews Genetics* 17.6 (2016), 333–351. ISSN: 1471-0056, 1471-0064. DOI: 10.1038/nrg.2016.49.

[55] Roy Gross et al. "The missing link: Bordetella petrii is endowed with both the metabolic versatility of environmental bacteria and virulence traits of pathogenic Bordetellae." In: *BMC Genomics* 9.1 (2008), p. 449. DOI: 10.1186/1471-2164-9-449.

[56] F. Z. H. "The Design of Experiments: By B. A. Fisher. Edinburgh: Oliver and Boyd. IX + 252 pages, illus. 12/6 net. 1935." In: *Agronomy Journal* 27.12 (1935), 1004–1005. ISSN: 0002-1962, 1435-0645. DOI: 10.2134/agronj1935.00021962002700120010x.

[57] Brian J Haas, Melissa Chin, Chad Nusbaum, Bruce W Birren, and Jonathan Livny. "How deep is deep enough for RNA-Seq profiling of bacterial transcriptomes?" In: *BMC Genomics* 13.1 (2012), p. 734. ISSN: 1471-2164. DOI: 10.1186/1471-2164-13-734.

[58] Joachim Hartung. *A note on combining dependent tests of significance*. Tech. rep. Technical Report, SFB 475: Komplexitätsreduktion in Multivariaten Datenstrukturen, Universität Dortmund, 1998.

[59] Jeff Hasty, David McMillen, Farren Isaacs, and James J. Collins. "Computational studies of gene regulatory networks: in numero molecular biology." In: *Nature Reviews Genetics* 2.4 (2001), pp. 268–279. DOI: 10.1038/35066056.

[60] David L. Hava, Julianna LeMieux, and Andrew Camilli. "From nose to lung: the regulation behind Streptococcus pneumoniae virulence factors: Virulence gene regulation in S. pneumoniae." In: *Molecular Microbiology* 50.4 (2003), 1103–1110. ISSN: 0950382X, 13652958. DOI: 10.1046/j.1365-2958.2003.03764.x.

[61] W. L. W. Hazenbos, C. A. W. Geuijen, B. M. van den Berg, F. R. Mooi, and R. van Furth. "Bordetella pertussis Fimbriae Bind to Human Monocytes via the Minor Fimbria) Subunit FimD." In: *Journal of Infectious Diseases* 171.4 (1995), 924–929. ISSN: 0022-1899, 1537-6613. DOI: 10.1093/infdis/171.4.924.

[62] Larry V. Hedges and Ingram Olkin. *Statistical methods for meta-analysis*. Academic Press, 1985. ISBN: 9780123363800.

[63] Eriikka Heikkinen, Teemu Kallonen, Lilli Saarinen, Rolf Sara, Audrey J. King, Frits R. Mooi, Juhani T. Soini, Jussi Mertsola, and Qiushui He. "Comparative Genomics of Bordetella pertussis Reveals Progressive Gene Loss in Finnish Strains." In: *PLOS ONE* 2.9 (Sept. 2007), pp. 1–8. DOI: 10.1371/journal.pone.0000904. URL: https://doi.org/10.1371/journal.pone.0000904.

[64] S Hellwig. "Evidence for an intracellular niche for bordetella pertussis in broncho-alveolar lavage cells of mice." In: *FEMS Immunology and Medical Microbiology* 26.3–4 (1999), 203–207. DOI: 10.1016/s0928-8244(99)00146-7.

[65] Sandra M. M. Hellwig, Maria E. Rodriguez, Guy A. M. Berbers, Jan G. J. van de Winkel, and Frits R. Mooi. "Crucial Role of Antibodies to Pertactin in Bordetella pertussis Immunity." In: *The Journal of Infectious Diseases* 188.5 (2003), 738–742. ISSN: 0022-1899, 1537-6613. DOI: 10.1086/377283.

[66] Sara E. Hester, Minghsun Lui, Tracy Nicholson, Daryl Nowacki, and Eric T. Harvill. "Identification of a CO2 responsive regulon in bordetella." In: *PLoS ONE* 7.10 (2012). DOI: 10.1371/journal.pone.0047635.

[67] Sarah C. Higgins, Ed C. Lavelle, Chantelle McCann, Brian Keogh, Edel McNeela, Patricia Byrne, Brian O'Gorman, Andrew Jarnicki, Peter McGuirk, and Kingston H. Mills. "Toll-like receptor 4-mediated innate IL-10 activates antigen-specific regulatory T cells and confers resistance to bordetella pertussis by inhibiting inflammatory pathology." In: *The Journal of Immunology* 171.6 (2003), 3119–3127. DOI: 10.4049/jimmunol.171.6.3119.

[68] R Higgs, S C Higgins, P J Ross, and K H G Mills. "Immunity to the respiratory pathogen Bordetella pertussis." In: *Mucosal Immunology* 5.5 (2012), 485–500. ISSN: 1933-0219, 1935-3456. DOI: 10.1038/mi.2012.54.

[69] David P Hill, Barry Smith, Monica S McAndrews-Hill, and Judith A Blake. "Gene ontology annotations: What they mean and where they come from." In: *BMC Bioinformatics* 9.S5 (2008). DOI: 10.1186/1471-2105-9-s5-s2.

[70] Michael S. Humphrys et al. "Simultaneous transcriptional profiling of bacteria and their host cells." In: *PLoS ONE* 8.12 (2013). DOI: 10.1371/journal.pone.0080597.

[71] T. Ideker and R. Sharan. "Protein networks in disease." In: *Genome Research* 18.4 (2008), 644–652. ISSN: 1088-9051. DOI: 10.1101/gr.071852.107.

[72] D. Jenkinson. "Duration of effectiveness of pertussis vaccine: Evidence from a 10 year community study." In: *BMJ* 296.6622 (1988), 612–614. DOI: 10.1136/bmj.296.6622.612.

[73] Ian T. Jolliffe and Jorge Cadima. "Principal component analysis: A review and recent developments." In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374.2065 (2016), p. 20150202. DOI: 10.1098/rsta.2015.0202.

[74] Meleri Jones et al. "RNA-seq analysis of host and viral gene expression highlights interaction between varicella zoster virus and keratinocyte differentiation." In: *PLoS Pathogens* 10.1 (2014). DOI: 10.1371/journal.ppat.1003896.

[75] Heidrun Jungnitz, Nicholas P. West, Mark J. Walker, Gursharan S. Chhatwal, and Carlos A. Guzman. "A second two-component regulatory system of bordetella bronchiseptica required for bacterial resistance to oxidative stress, production of acid phosphatase, and in vivo persistence." In: *Infection and Immunity* 66.10 (1998), 4640–4650. DOI: 10.1128/iai.66.10.4640-4650.1998.

[76] M. Kanehisa. "KEGG: Kyoto Encyclopedia of Genes and Genomes." In: *Nucleic Acids Research* 28.1 (2000), pp. 27–30. DOI: 10.1093/nar/28.1.27.

[77] M. Kanehisa et al. "KEGG for linking genomes to life and the environment." In: *Nucleic Acids Research* 36.Database (2007), D480–D484. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gkm882.

[78] M. Kanehisa et al. "KEGG for linking genomes to life and the environment." In: *Nucleic Acids Research* 36.Database (2007), D480–D484. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gkm882.

[79] N. Kashtan, S. Itzkovitz, R. Milo, and U. Alon. "Topological generalizations of network motifs." In: *Physical Review E* 70.3 (2004), p. 031909. ISSN: 1539-3755, 1550-2376. DOI: 10.1103/PhysRevE.70.031909.

[80] Peter Kharchenko, George M Church, and Dennis Vitkup. "Expression dynamics of a cellular metabolic network." In: *Molecular Systems Biology* 1.1 (2005). ISSN: 1744-4292, 1744-4292. DOI: 10.1038/msb4100023. URL: https://onlinelibrary.wiley.com/doi/abs/10.1038/msb4100023.

[81] Tadele G. Kiros, Benoît Levast, Gaël Auray, Stacy Strom, Jill van Kessel, and Volker Gerdts. "The importance of animal models in the development of vaccines." In: *Innovation in Vaccinology* (2012), 251–264. DOI: 10.1007/978-94-007-4543-8_11.

[82] K. Klemm and S. Bornholdt. "Topology of biological networks and reliability of information processing." In: *Proceedings of the National Academy of Sciences* 102.51 (2005), 18414–18419. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.0509132102.

[83] C.H Wirsing von König, S Postels-Multan, H.J Schmitt, and H.L Bock. "Pertussis in adults: Frequency of transmission after household exposure." In: *The Lancet* 346.8986 (1995), 1326–1329. DOI: 10.1016/s0140-6736(95)92343-8.

[84] Harold J. Lambert. "Epidemiology of a small pertussis outbreak in Kent County, Michigan." In: *Public Health Reports (1896-1970)* 80.4 (1965), p. 365. DOI: 10.2307/4592424.

[85] Yanina Lamberti, Juan Gorgojo, Cintia Massillo, and Maria E. Rodriguez. "bordetella pertussisentry into respiratory epithelial cells and intracellular survival." In: *Pathogens and Disease* 69.3 (2013), 194–204. DOI: 10.1111/2049-632x.12072.

[86] Shawn E. Levy and Richard M. Myers. "Advancements in next-generation sequencing." In: *Annual Review of Genomics and Human Genetics* 17.1 (2016), 95–115. DOI: 10.1146/annurev-genom-083115-022413.

[87] Z M Li, M J Brennan, J L David, P H Carter, J L Cowell, and C R Manclark. "Comparison of type 2 and type 6 fimbriae of Bordetella pertussis by using agglutinating monoclonal antibodies." In: *Infection and Immunity* 56.12 (1988), 3184–3188. ISSN: 0019-9567, 1098-5522. DOI: 10.1128/IAI.56.12.3184-3188.1988.

[88] Y. Liao, G. K. Smyth, and W. Shi. "featureCounts: an efficient general purpose program for assigning sequence reads to genomic features." In: *Bioinformatics* 30.7 (2013), pp. 923–930. DOI: 10.1093/bioinformatics/btt656.

[89] Tamás Lipták. "On the combination of independent tests." In: *Magyar Tud Akad Mat Kutato Int Kozl* 3 (1958), pp. 171–197.

[90] C Locht and J. Keith. "Pertussis toxin gene: nucleotide sequence and genetic organization." In: *Science* 232.4755 (1986), 1258–1264. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.3704651.

[91] Michael I Love, Wolfgang Huber, and Simon Anders. "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2." In: *Genome Biology* 15.12 (2014). DOI: 10.1186/s13059-014-0550-8.

[92] Denis Macina and Keith E. Evans. "Bordetella pertussis in school-age children, adolescents and adults: A systematic review of Epidemiology and mortality in Europe." In: *Infectious Diseases and Therapy* 10.4 (2021), 2071–2118. DOI: 10.1007/s40121-021-00520-9.

[93] M. Magrane and U. Consortium. "Uniprot Knowledgebase: A hub of Integrated Protein Data." In: *Database* 2011.0 (2011). DOI: 10.1093/database/bar009.

[94] B P Mahon, M S Ryan, F Griffin, and K H Mills. "Interleukin-12 is produced by macrophages in response to live or killed Bordetella pertussis and enhances the efficacy of an acellular pertussis vaccine by promoting induction of Th1 cells." In: *Infection and immunity* 64.12 (1996), 5295–5301. DOI: 10.1128/iai.64.12.5295-5301.1996.

[95] Bernard P. Mahon, Brian J. Sheahan, Fiona Griffin, Geraldine Murphy, and Kingston H.G. Mills. "Atypical disease after bordetella pertussis respiratory infection of mice with targeted disruptions of interferon-gamma receptor or immunoglobulin-M chain genes." In: *The Journal of Experimental Medicine* 186.11 (1997), 1843–1851. DOI: 10.1084/jem.186.11.1843.

[96] Marcel Martin. "Cutadapt removes adapter sequences from high-throughput sequencing reads." In: *EMBnet.journal* 17.1 (2011), pp. 10–12. ISSN: 2226-6089. DOI: 10.14806/ej.17.1.200. URL: https://journal.embnet.org/index.php/embnetjournal/article/view/200.

[97] Cristina Masseria and Girishanthy Krishnarajah. "The estimated incidence of pertussis in people aged 50 years old in the United States, 2006–2010." In: *BMC Infectious Diseases* 15.1 (2015). DOI: 10.1186/s12879-015-1269-1.

[98] Seema Mattoo and James D. Cherry. "Molecular pathogenesis, epidemiology, and clinical manifestations of respiratory infections due tobordetella pertussisand otherbordetellasubspecies." In: *Clinical Microbiology Reviews* 18.2 (2005), 326–382. DOI: 10.1128/cmr.18.2.326-382.2005.

[99] Seema Mattoo, Ming H. Yuk, Lisa L. Huang, and Jeff F. Miller. "Regulation of type III secretion in Bordetella: Bordetella type III secretion." In: *Molecular Microbiology* 52.4 (2004), 1201–1214. ISSN: 0950382X, 13652958. DOI: 10.1111/j.1365-2958.2004.04053.x.

[100] Adam McDermaid, Brandon Monier, Jing Zhao, Bingqiang Liu, and Qin Ma. "Interpretation of differential gene expression results of RNA-Seq Data: Review and Integration." In: *Briefings in Bioinformatics* 20.6 (2018), 2044–2054. DOI: 10.1093/bib/bby067.

[101] Peter McGuirk, Bernhard P. Mahon, Fiona Griffin, and Kingston H. Mills. "Compartmentalization of T cell responses following respiratory infection with bordetella pertussis: Hyporesponsiveness of lung T cells is associated with modulated expression of the co-stimulatory molecule CD28." In: *European Journal of Immunology* 28.01 (1998), 153–163. DOI: 10.1002/(sici)1521-4141(199801)28:01lt;153::aid-immu153gt;3.3.co;2-r.

[102] Peter McGuirk and Kingston H. Mills. "A regulatory role for interleukin 4 in differential inflammatory responses in the lung following infection of mice primed with th1- or th2-inducing pertussis vaccines." In: *Infec-*

tion and Immunity 68.3 (2000), 1383–1390. DOI: 10.1128/iai.68.3.1383-1390.2000.

[103] Jeffrey A. Melvin, Erich V. Scheller, Christopher R. Noël, and Peggy A. Cotter. "New Insight into Filamentous Hemagglutinin Secretion Reveals a Role for Full-Length FhaB in Bordetella Virulence." In: *mBio* 6.4 (2015). Ed. by Jeff F. Miller, e01189–15. ISSN: 2150-7511. DOI: 10.1128/mBio.01189-15.

[104] D. Scott Merrell, Susan M. Butler, Firdausi Qadri, Nadia A. Dolganov, Ahsfaqul Alam, Mitchell B. Cohen, Stephen B. Calderwood, Gary K. Schoolnik, and Andrew Camilli. "Host-induced epidemic spread of the cholera bacterium." In: *Nature* 417.6889 (2002), 642–645. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature00778.

[105] Michael L. Metzker. "Sequencing technologies — the next generation." In: *Nature Reviews Genetics* 11.1 (2010), 31–46. ISSN: 1471-0056, 1471-0064. DOI: 10.1038/nrg2626.

[106] Tom Michoel et al. "Validating module network learning algorithms using simulated data." In: *BMC Bioinformatics* 8.S2 (2007). DOI: 10.1186/1471-2105-8-s2-s5.

[107] Beatriz Miguelena Chamorro, Karelle De Luca, Gokul Swaminathan, Stéphanie Longet, Egbert Mundt, and Stéphane Paul. "Bordetella bronchiseptica and bordetella pertussis: Similarities and differences in infection, immuno-modulation, and vaccine considerations." In: *Clinical Microbiology Reviews* 36.3 (2023). DOI: 10.1128/cmr.00164-22.

[108] K H Mills, A Barnard, J Watkins, and K Redhead. "Cell-mediated immunity to Bordetella pertussis: role of Th1 cells in bacterial clearance in a murine respiratory infection model." In: *Infection and Immunity* 61.2 (1993), 399–410. ISSN: 0019-9567, 1098-5522. DOI: 10.1128/IAI.61.2.399-410.1993.

[109] Alba Mir-Cros et al. "Population Dynamics and antigenic drift of bordetella pertussis following whole cell vaccine replacement, Barcelona, Spain, 1986–2015." In: *Emerging Microbes & Infections* 8.1 (2019), 1711–1720. DOI: 10.1080/22221751.2019.1694395.

[110] Yusuke Miyaji, Nao Otsuka, Hiromi Toyoizumi-Ajisaka, Keigo Shibayama, and Kazunari Kamachi. "Genetic analysis of bordetella pertussis isolates from the 2008–2010 pertussis epidemic in Japan." In: *PLoS ONE* 8.10 (2013). DOI: 10.1371/journal.pone.0077165.

[111] Frits R. Mooi, Han G.J. van der Heide, Anja R. ter Avest, Karen G. Welinder, Ian Livey, Ben A.M. van der Zeijst, and Wim Gaastra. "Characterization of fimbrial subunits from Bordetella species." In: *Microbial Pathogenesis* 2.6 (1987), 473–484. ISSN: 08824010. DOI: 10.1016/0882-4010(87)90054-4.

[112] Frits R. Mooi, Inge H. M. Van Loo, and Audrey King. "Adaptation of-Bordetella pertussisto Vaccination: A Cause for Its Reemergence?" In: *Emerging Infectious Diseases* 7.7 (2001), 526–528. DOI: 10.3201/eid0707.017708.

[113] Ali Mortazavi, Brian A Williams, Kenneth McCue, Lorian Schaeffer, and Barbara Wold. "Mapping and quantifying mammalian transcriptomes by RNA-Seq." In: *Nature Methods* 5.7 (2008), 621–628. ISSN: 1548-7091, 1548-7105. DOI: 10.1038/nmeth.1226.

[114] André P Mäurer, Adrian Mehlitz, Hans J Mollenkopf, and Thomas F Meyer. "Gene Expression Profiles of Chlamydophila pneumoniae during the Developmental Cycle and Iron Depletion–Mediated Persistence." In: *PLoS Pathogens* 3.6 (2007). Ed. by Joanne N Engel, e83. ISSN: 1553-7374. DOI: 10.1371/journal.ppat.0030083.

[115] Mari M. Nakamura, Sin-Yee Liew, Craig A. Cummings, Mary M. Brinig, Christine Dieterich, and David A. Relman. "Growth phase- and nutrient limitation-associated transcript abundance regulation in bordetella pertussis." In: *Infection and Immunity* 74.10 (2006), 5537–5548. DOI: 10.1128/iai.00781-06.

[116] Tuan-Minh Nguyen, Adib Shafi, Tin Nguyen, and Sorin Draghici. "Identifying significantly impacted pathways: A Comprehensive Review and assessment." In: *Genome Biology* 20.1 (2019). DOI: 10.1186/s13059-019-1790-4.

[117] Hiroyuki Ohashi, Mai Hasegawa, Kentaro Wakimoto, and Etsuko Miyamoto-Sato. "Next-generation technologies for multiomics approaches including Interactome sequencing." In: *BioMed Research International* 2015 (2015), 1–9. DOI: 10.1155/2015/104209.

[118] Haley F Oliver, Renato H Orsi, Lalit Ponnala, Uri Keich, Wei Wang, Qi Sun, Samuel W Cartinhour, Melanie J Filiatrault, Martin Wiedmann, and Kathryn J Boor. "Deep RNA sequencing of L. monocytogenes reveals overlapping and extensive stationary phase and sigma B-dependent transcriptomes, including multiple highly transcribed noncoding RNAs." In: *BMC Genomics* 10.1 (2009), p. 641. ISSN: 1471-2164. DOI: 10.1186/1471-2164-10-641.

[119] Shao-En Ong and Matthias Mann. "Mass spectrometry–based proteomics turns quantitative." In: *Nature Chemical Biology* 1.5 (2005), 252–262. ISSN: 1552-4450, 1552-4469. DOI: 10.1038/nchembio736.

[120] Oracle. *Java Platform SE 8*. https://docs.oracle.com/javase/8/docs/api/java/math/package-summary.html. accessed 2023.

[121] Nao Otsuka, Hyun-Ja Han, Hiromi Toyoizumi-Ajisaka, Yukitsugu Nakamura, Yoshichika Arakawa, Keigo Shibayama, and Kazunari Kamachi. "Prevalence and genetic characterization of pertactin-deficient bordetella pertussis in Japan." In: *PLoS ONE* 7.2 (2012). DOI: 10.1371/journal.pone.0031985.

[122] Christopher;D. Paddock et al. "Pathology and pathogenesis of fatal bordetella pertussis infection in infants." In: *Clinical Infectious Diseases* 47.3 (2008), 328–338. DOI: 10.1086/589753.

[123] Chanda Panse and Manali Kshirsagar. "Survey on Modelling Methods Applicable to Gene Regulatory Network." In: *International Journal on Bioinformatics & Biosciences* 3.3 (2013), pp. 13–23. DOI: 10.5121/ijbb.2013.3302.

[124] Julian Parkhill et al. "Comparative analysis of the genome sequences of Bordetella pertussis, Bordetella parapertussis and Bordetella bronchiseptica." In: *Nature Genetics* 35.1 (2003), 32–40. ISSN: 1061-4036, 1546-1718. DOI: 10.1038/ng1227.

[125] Rob Patro, Geet Duggal, Michael I Love, Rafael A Irizarry, and Carl Kingsford. "Salmon provides fast and bias-aware quantification of transcript expression." In: *Nature Methods* 14.4 (2017), 417–419. DOI: 10.1038/nmeth.4197.

[126] Timothy T. Perkins et al. "A Strand-Specific RNA–Seq Analysis of the Transcriptome of the Typhoid Bacillus Salmonella Typhi." In: *PLoS Genetics* 5.7 (2009). Ed. by Josep Casadesús, e1000569. ISSN: 1553-7404. DOI: 10.1371/journal.pgen.1000569.

[127] *Pertussis In ECDC: Annual epidemiological report for 2018*. 2020. URL: https://www.ecdc.europa.eu/en/publications-data/pertussis-annual-epidemiological-report-2018.

[128] Elizabeth J. Pishko, David J. Betting, Christina S. Hutter, and Eric T. Harvill. "Bordetella pertussis Acquires Resistance to Complement-Mediated Killing In Vivo." In: *Infection and Immunity* 71.9 (2003), 4936–4942. ISSN: 0019-9567, 1098-5522. DOI: 10.1128/IAI.71.9.4936-4942.2003.

[129] Aaron R. Quinlan and Ira M. Hall. "BEDTools: a flexible suite of utilities for comparing genomic features." In: *Bioinformatics* 26.6 (Jan. 2010), pp. 841–842. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btq033. eprint: https://academic.oup.com/bioinformatics/article-pdf/26/6/841/48854754/bioinformatics_26_6_841.pdf. URL: https://doi.org/10.1093/bioinformatics/btq033.

[130] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2021. URL: https://www.R-project.org/.

[131] Franck Rapaport, Raya Khanin, Yupu Liang, Mono Pirun, Azra Krek, Paul Zumbo, Christopher E Mason, Nicholas D Socci, and Doron Betel. "Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data." In: *Genome Biology* 14.9 (2013), R95. ISSN: 1465-6906. DOI: 10.1186/gb-2013-14-9-r95.

[132] R. Rappuoli. "Pushing the limits of cellular microbiology: Microarrays to study bacteria-host cell intimate contacts." In: *Proceedings of the National Academy of Sciences* 97.25 (2000), 13467–13469. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.010505497.

[133] Davide Risso, John Ngai, Terence P Speed, and Sandrine Dudoit. "Normalization of RNA-seq data using factor analysis of control genes or samples." In: *Nature Biotechnology* 32.9 (2014), 896–902. DOI: 10.1038/nbt.2931.

[134] Robert Gentleman, Vincent Carey, Martin Morgan, and others. *GSEAbase: Gene set enrichment data structures and methods*. 2023. URL: https://bioconductor.org/packages/GSEAbase.

[135] M. D. Robinson, D. J. McCarthy, and G. K. Smyth. "edgeR: a Bioconductor package for differential expression analysis of digital gene expression data." In: *Bioinformatics* 26.1 (2009), pp. 139–140. DOI: 10.1093/bioinformatics/btp616.

[136] MJ Saffar, A Hashemi, GR Ghorbani, MS Rezai, S Shahmohammadi, and J Shojaei. "Clinical and laboratory features of pertussis in hospitalized infants with confirmed versus probable pertussis cases." In: *Annals of Medical and Health Sciences Research* 4.6 (2014), p. 910. DOI: 10.4103/2141-9248.144911.

[137] Anja Saso, Beate Kampmann, and Sophie Roetynck. "Vaccine-induced cellular immunity against bordetella pertussis: Harnessing Lessons from animal and human studies to improve design and testing of novel pertussis vaccines." In: *Vaccines* 9.8 (2021), p. 877. DOI: 10.3390/vaccines9080877.

[138] Andrea Sboner, Xinmeng Mu, Dov Greenbaum, Raymond K Auerbach, and Mark B Gerstein. "The real cost of sequencing: Higher than you think!" In: *Genome Biology* 12.8 (2011), p. 125. DOI: 10.1186/gb-2011-12-8-125.

[139] M. Schena, D. Shalon, R. W. Davis, and P. O. Brown. "Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray." In: *Science* 270.5235 (1995), 467–470. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.270.5235.467.

[140] Jeremy Schmutz et al. "Quality assessment of the human genome sequence." In: *Nature* 429.6990 (2004), 365–368. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature02390.

[141] Katie L. Sealey, Simon R. Harris, Norman K. Fry, Laurence D. Hurst, Andrew R. Gorringe, Julian Parkhill, and Andrew Preston. "Genomic analysis of isolates from the United Kingdom 2012 pertussis outbreak reveals that vaccine antigen genes are unusually fast evolving." In: *Journal of Infectious Diseases* 212.2 (2014), 294–301. DOI: 10.1093/infdis/jiu665.

[142] D. W. Selinger. "Global RNA Half-Life Analysis in Escherichia coli Reveals Positional Patterns of Transcript Degradation." In: *Genome Research* 13.2 (2003), 216–223. ISSN: 10889051. DOI: 10.1101/gr.912603.

[143] Ulrich Stelzl et al. "A human protein-protein interaction network: A resource for annotating the proteome." In: *Cell* 122.6 (2005), 957–968. DOI: 10.1016/j.cell.2005.08.029.

[144] Jann Storsaeter, Hans O Hallander, Lennart Gustafsson, and Patrick Olin. "Levels of anti-pertussis antibodies related to protection after household exposure to Bordetella pertussis." In: *Vaccine* 16.20 (1998), 1907–1916. ISSN: 0264410X. DOI: 10.1016/S0264-410X(98)00227-8.

[145] J. M. Stuart. "A Gene-Coexpression Network for Global Discovery of Conserved Genetic Modules." In: *Science* 302.5643 (2003), 249–255. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1087447.

[146] Marc Sultan, Vyacheslav Amstislavskiy, Thomas Risch, Moritz Schuette, Simon Dökel, Meryem Ralser, Daniela Balzereit, Hans Lehrach, and Marie-Laure Yaspo. "Influence of RNA extraction methods and library selection schemes on RNA-seq data." In: *BMC Genomics* 15.1 (2014), p. 675. ISSN: 1471-2164. DOI: 10.1186/1471-2164-15-675.

[147] Fran Supek, Matko Bošnjak, Nives Škunca, and Tomislav Šmuc. "Revigo summarizes and visualizes long lists of gene ontology terms." In: *PLoS ONE* 6.7 (2011). DOI: 10.1371/journal.pone.0021800.

[148] *Surveillance atlas of infectious diseases*. URL: https://atlas.ecdc.europa.eu/public/index.aspx?Dataset.

[149] Yuan Tang, Masaaki Horikoshi, and Wenxuan Li. "GGFORTIFY: Unified interface to visualize statistical results of popular R packages." In: *The R Journal* 8.2 (2016), p. 474. DOI: 10.32614/rj-2016-060.

[150] Dawn L. Taylor-Mulneix et al. "Bordetella bronchiseptica exploits the complex life cycle of dictyostelium discoideum as an amplifying transmission vector." In: *PLOS Biology* 15.4 (2017). DOI: 10.1371/journal.pbio.2000420.

[151] Cole Trapnell, Adam Roberts, Loyal Goff, Geo Pertea, Daehwan Kim, David R Kelley, Harold Pimentel, Steven L Salzberg, John L Rinn, and Lior Pachter. "Erratum: Corrigendum: Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and cufflinks."

In: *Nature Protocols* 9.10 (2014), 2513–2513. DOI: 10.1038/nprot1014-2513a.

[152] Hugo Alberto Valdez, Juan Marcos Oviedo, Juan Pablo Gorgojo, Yanina Lamberti, and Maria Eugenia Rodriguez. "Bordetella pertussis modulates human macrophage defense gene expression." In: *Pathogens and Disease* 74.6 (2016). Ed. by Nicholas Carbonetti, ftw073. ISSN: 2049-632X, 2049-632X. DOI: 10.1093/femspd/ftw073.

[153] Hugo Alberto Valdez, Juan Marcos Oviedo, Juan Pablo Gorgojo, Yanina Lamberti, and Maria Eugenia Rodriguez. "bordetella pertussismodulates human macrophage defense gene expression." In: *Pathogens and Disease* 74.6 (2016). DOI: 10.1093/femspd/ftw073.

[154] Imre Vastrik et al. "Reactome: a knowledge base of biologic pathways and processes." In: *Genome Biology* 8.3 (2007), R39. ISSN: 1474-760X. DOI: 10.1186/gb-2007-8-3-r39. URL: https://doi.org/10.1186/gb-2007-8-3-r39.

[155] D. Vickers, A. G. Ross, R. C. Mainar-Jaime, C. Neudorf, and S. Shah. "Whole-cell and acellular pertussis vaccination programs and rates of pertussis among infants and young children." In: *Canadian Medical Association Journal* 175.10 (2006), 1213–1217. DOI: 10.1503/cmaj.051637.

[156] Jana Vojtova, Jana Kamanova, and Peter Sebo. "Bordetella adenylate cyclase toxin: a swift saboteur of host defense." In: *Current Opinion in Microbiology* 9.1 (2006), 69–75. ISSN: 13695274. DOI: 10.1016/j.mib.2005.12.011.

[157] *WHO: Pertussis reported cases and incidence*. 2023. URL: https://immunizationdata.who.int/pages/incidence/pertussis.html.

[158] Beibei Wang, Fengzhu Sun, and Yihui Luan. "Comparison of the effectiveness of different normalization methods for metagenomic cross-study phenotype prediction under heterogeneity." In: *Scientific Reports* 14.1 (2024). DOI: 10.1038/s41598-024-57670-2.

[159] Zhong Wang, Mark Gerstein, and Michael Snyder. "RNA-Seq: a revolutionary tool for transcriptomics." In: *Nature Reviews Genetics* 10.1 (2009), 57–63. ISSN: 1471-0056, 1471-0064. DOI: 10.1038/nrg2484.

[160] Jason M. Warfel, Lindsey I. Zimmerman, and Tod J. Merkel. "Acellular pertussis vaccines protect against disease but fail to prevent infection and transmission in a nonhuman primate model." In: *Proceedings of the National Academy of Sciences* 111.2 (2013), 787–792. DOI: 10.1073/pnas.1314688110.

[161] Christine L. Weingart and Alison A. Weiss. "Bordetella pertussis Virulence Factors Affect Phagocytosis by Human Neutrophils." In: *Infection and Immunity* 68.3 (2000). Ed. by D. L. Burns, 1735–1739. ISSN: 1098-5522, 0019-9567. DOI: 10.1128/IAI.68.3.1735-1739.2000.

[162] Aaron M. Wendelboe, Annelies Van Rie, Stefania Salmaso, and Janet A. Englund. "Duration of immunity against pertussis after natural infection or vaccination." In: *Pediatric Infectious Disease Journal* 24.5 (2005). DOI: 10.1097/01.inf.0000160914.59160.41.

[163] Alexander J. Westermann, Stanislaw A. Gorski, and Jörg Vogel. "Dual RNA-seq of pathogen and host." In: *Nature Reviews Microbiology* 10.9 (2012), 618–630. ISSN: 1740-1526, 1740-1534. DOI: 10.1038/nrmicro2852.

[164] Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016. ISBN: 978-3-319-24277-4. URL: https://ggplot2.tidyverse.org.

[165] R. Willems, A. Paul, H. G. van der Heide, A. R. ter Avest, and F. R. Mooi. "Fimbrial phase variation in Bordetella pertussis: a novel mechanism for transcriptional regulation." In: *The EMBO Journal* 9.9 (1990), 2803–2809. ISSN: 02614189. DOI: 10.1002/j.1460-2075.1990.tb07468.x.

[166] F. von Wintzingerode, G. Gerlach, B. Schneider, and R. Gross. "Phylogenetic relationships and virulence evolution in the genus bordetella." In: *Pathogenicity Islands and the Evolution of Pathogenic Microbes* (2002), 177–199. DOI: 10.1007/978-3-642-56031-6_10.

[167] Maxwell A. Witt, Larry Arias, Paul H. Katz, Elizabeth T. Truong, and David J. Witt. "Reduced Risk of Pertussis Among Persons Ever Vaccinated With Whole Cell Pertussis Vaccine Compared to Recipients of Acellular Pertussis Vaccines in a Large US Cohort." In: *Clinical Infectious Diseases* 56.9 (2013), 1248–1254. DOI: 10.1093/cid/cit046.

[168] Maxwell A. Witt, Paul H. Katz, and David J. Witt. "Unexpectedly limited durability of immunity following acellular pertussis vaccination in preadolescents in a North American outbreak." In: *Clinical Infectious Diseases* 54.12 (2012), 1730–1735. DOI: 10.1093/cid/cis287.

[169] Daniel N. Wolfe, Paul B. Mann, Anne M. Buboltz, and Eric T. Harvill. "Delayed role of tumor necrosis factor–a in overcoming the effects of pertussis toxin." In: *The Journal of Infectious Diseases* 196.8 (2007), 1228–1236. DOI: 10.1086/521303.

[170] Haiyuan Yu, Nicholas M. Luscombe, Hao Xin Lu, Xiaowei Zhu, Yu Xia, Jing-Dong J. Han, Nicolas Bertin, Sambath Chung, Marc Vidal, and Mark Gerstein. "Annotation transfer between genomes: Protein–protein Interologs and protein–DNA regulogs." In: *Genome Research* 14.6 (2004), 1107–1118. DOI: 10.1101/gr.1774904.

[171] A Zeddeman et al. "Investigations into the emergence of pertactin-deficient bordetella pertussis isolates in six European countries, 1996 to 2012." In: *Eurosurveillance* 19.33 (2014). DOI: 10.2807/1560-7917.es2014.19.33.20881.

[172] Xuqing Zhang, Tania Goel, Laura L. Goodfield, Sarah J. Muse, and Eric T. Harvill. "Decreased leukocyte accumulation and delayed bordetella pertussis clearance in IL-6-/-mice." In: *The Journal of Immunology* 186.8 (2011), 4895–4904. DOI: 10.4049/jimmunol.1000594.

[173] Yiming Zhong, Feng Xu, Jinhua Wu, Jeffrey Schubert, and Marilyn M. Li. "Application of next generation sequencing in laboratory medicine." In: *Annals of Laboratory Medicine* 41.1 (2021), 25–43. DOI: 10.3343/alm.2021.41.1.25.

[174] Xianghong Zhou, Ming-Chih J. Kao, and Wing Hung Wong. "Transitive functional annotation by shortest-path analysis of gene expression data." In: *Proceedings of the National Academy of Sciences* 99.20 (2002), 12783–12788. DOI: 10.1073/pnas.192159399.

[175] María Eugenia Zurita, Mieszko M. Wilk, Francisco Carriquiriborde, Erika Bartel, Griselda Moreno, Alicja Misiak, Kingston H. G. Mills, and Daniela Hozbor. "A Pertussis Outer Membrane Vesicle-Based Vaccine Induces Lung-Resident Memory CD4 T Cells and Protection Against Bordetella pertussis, Including Pertactin Deficient Strains." In: *Frontiers in Cellular and Infection Microbiology* 9 (2019), p. 125. ISSN: 2235-2988. DOI: 10.3389/fcimb.2019.00125.

DECLARATION

I confirm that this dissertation is entirely original, and I have not submitted any portion of this for consideration towards any other degree or qualification at this university or elsewhere, unless explicitly stated. I have given proper credit to the works of others wherever necessary. I have given proper credit to all sources used in the text. I have also adhered to the principles of 'Good Scientific Practice' in both this dissertation and any related publications.

Wien, 2025

Mariam Rukhama Farman,
BSc MSc

Mbp	Megabase pairs
bp	Base pairs
GC	Guanine-Cytosine
rRNA	ribosomal Ribonucleic Acid
tRNA	transfer Ribonucleic Acid
IS	Insertion
DNA	Deoxyribonucleic Acid
Fim	Fimbriae
FHA	Filamentous Haemagglutinin
wP	Whole-cell Pertussis
aP	Acellular Pertussis
Th1	T-helper 1
Th2	T-helper 2
vags	Virulence-activated genes
PT	Pertussis Toxin
T4SS	Type-IV Secretory System
G-proteins	Guanine nucleotide-binding proteins
ADP	Adenosine Diphosphate
cAMP	Cyclic Adenosine Monophosphate
ACT	Adenylyl Cyclase Toxin
AC	Adenylate Cyclase
RTX	Repeat in Toxin
APC	Antigen-presenting Cells
T3SS	Type-III Secretion System
LPS	Lipopolysaccharide
PAMP	Pathogen-associated Molecular Patterns
DC	Dendritic Cells

DTP	Diphtheria, Tetanus, Pertussis
Prn	Pertactin
TcfA	Tracheal colonization factor A
NK	Natural Killer Cells
BrkA	Bordetella resistance to killing A
NGS	Next-generation Sequencing
PCR	Polymerase Chain Reaction
qRT-PCR	Quantitative Real-time PCR
cDNA	Complementary DNA
CDA	Coding DNA Sequence
RNA-seq	RNA sequencing
UTR	Untranslated Region
mRNA	Messenger RNA
NO	Nitric Oxide
pg	Picograms
fg	Femtograms
QC	Quality Control
poly-A	Polyadenylic acid
GTF	Gene Transfer File
DGEA	Differential Gene Expression Analysis
GO	Gene Ontology
IPA	Ingenuity Pathway Analysis
GRN	Gene Regulatory Network
p-values	Probability Values
hpi	Hours Post Infection
MOI	Multiplicity of Infection
TMM	Trimmed Mean of M-values
MSF	Modulated Sub-graph Finder