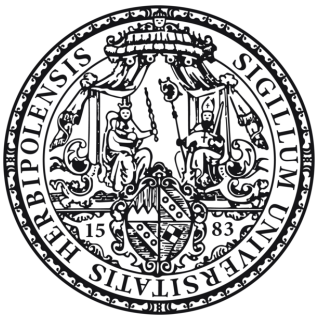

Implementation and application of bioinformatics methods to analyze and visualize single-cell RNA-sequencing data

Implementierung und Anwendung von bioinformatischen Methoden zur Analyse und Visualisierung von Einzelzell-Sequenzierungsdaten

Doctoral thesis for a doctoral degree
at the Graduate School of Life Sciences,
Julius-Maximilians-Universität Würzburg,
Section Infection and immunity

Submitted by
Ehsan Vafadarnejad

From
Gorgan, Iran

Würzburg, 2021

Submitted on: ……..

Office stamp

Members of the thesis committee:

Chairperson:	Prof. Dr. Christian Janzen
Primary supervisor:	Dr. Antoine-Emmanuel Saliba
Supervisor (second):	Prof. Dr. Thomas Dandekar
Supervisor (third):	Prof. Dr. Thomas Pietschmann
Supervisor (fourth):	Dr. Lars Barquist

Date of public defense: ……...

Date of receipt of certificate: ……...

Affidavit

I hereby confirm that my thesis entitled Implementation and application of bioinformatics methods to analyze and visualize single-cell RNA-sequencing data is the result of my own work. I did not receive any help or support from commercial consultants. All sources and / or materials applied are listed and specified in the thesis.

Furthermore, I confirm that this thesis has not yet been submitted as part of another examination process neither in identical nor in similar form.

Place, Date Signature

Eidesstattliche Erklärung

Hiermit erkläre ich an Eides statt, die Dissertation Implementierung und Anwendung von bioinformatischen Methoden zur Analyse und Visualisierung von Einzelzell-Sequenzierungsdaten eigenständig, d.h. insbesondere selbständig und ohne Hilfe eines kommerziellen Promotionsberaters, angefertigt und keine anderen als die von mir angegebenen Quellen und Hilfsmittel verwendet zu haben.

Ich erkläre außerdem, dass die Dissertation weder in gleicher noch in ähnlicher Form bereits in einem anderen Prüfungsverfahren vorgelegen hat.

Ort, Datum Unterschrift

Acknowledgement

I would particularly like to thank my mentor, Dr. Antoine-Emmanuel Saliba, for giving me the opportunity to perform this work in his group, for his instructive supervision and continuous support over the last years and for being always a source of motivation during my PhD. I am thankful to Dr. Clement Cochain and Dr. Lars Barquist for their precious feedback on my work during my PhD and for their valuable scientific support. Indeed, many of the achievements presented in this dissertation wouldn't be possible without their input. I thank Prof. Thomas Pietschmann and Prof. Thomas Dandekar for being a part of the thesis committee and for their supervision of my work over the last years. Above all, I express my gratitude to all members of the Saliba group who all played a major contribution in the work presented in this dissertation. I particularly thank Dr. Panagiota Arampatzi, Fabian Imdahl, Tobias Krammer, Oliver Dietrich and Christophe Toussaint.

Abstract

RNA sequencing (RNA-seq) has become a transformative method to profile genome-wide gene expression and whole transcriptome analysis over the last decade. In recent years, with the development of new technologies, it has become possible to study gene expression at single-cell level. This new advances in single-cell RNA-sequencing has revolutionized the way scientists study biological processes. Single-cell RNA-sequencing has been used in different areas to better understand the underlying mechanisms of biological processes.

In particular, single-RNA-sequencing is a suitable method to study infectious diseases. Infection is composed of heterogeneous mechanisms on either the host or pathogen side and the best way to understand the heterogeneity of these mechanisms and how they interact with each other is to study infectious diseases at the single-cell level. Studying infection processes at the single-cell level can reveal not only the heterogeneity but also the dynamics of infection and the interplay between the host and pathogen at the molecular level.

In this thesis, we implemented and applied different single-cell RNA-seq technologies to better understand infectious diseases. In the present work, we conducted four independent but related research works to shed light on different aspects of infection biology:

- We took advantage of this novel technology to study the consequences of RSV infection on primary human epithelial cells. The primary human epithelial cells were collected from six donors and cultured in air liquid interface (ALI) cell culture inoculated with **respiratory syncytial virus (RSV)**. In this project, we discovered ciliated cells as the susceptible cell types in RSV infection. We applied viral load as an indicator of infection progression and used it to reconstruct the dynamics of host response to RSV infection. Reconstruction of the dynamics of infection revealed many host genes and pathways that were

suppressed or induced as a result of RSV infection. Pathways related to innate immune response and interferon response were suppressed during the progression of infection and on the other hand pathways like protein targeting to endoplasmic reticulum and apoptosis were induced.

- We developed a new method which is capable of sequencing the **transcriptome of a bacterium** at the single-cell level and potentially can help us to characterize the bacterial heterogeneity during the course of infection. In this research project, bacteria were cultured in three different culture conditions namely Late stationary phase, Anaerobic shock and NaCl shock and we used a poly(A)-independent single-cell RNA-sequencing protocol to sequence bacteria at the single-cell level. In this work, we report the faithful capture of growth-dependent gene expression patterns in individual *Salmonella* and *Pseudomonas* bacteria. The results of our analysis showed that not only we could capture transcripts across different RNA classes but also our method is capable of discerning the transcriptome of bacteria across different culture conditions.
- We used single-cell RNA-sequencing technology to characterize the immune cells landscape over the course of atherosclerosis. Atherosclerosis is considered a cardiac disease which is highly related to infections and previous infections with bacteria or viruses is considered as a risk factor for atherosclerosis. We performed single-cell RNA sequencing of aortic $CD45^{+}$ cells extracted from healthy and atherosclerotic aorta of mice. We managed to find certain cell populations which were specifically present in atherosclerotic mice. One of the atheroschelorotic populations was previously undescribed **$TREM2^{high}$ macrophages** showing enrichment in *Trem2* gene expression. This population of macrophages seemed to be involved in functions like lipid metabolism and catabolism and lesion calcification. This work revealed the phenotypic heterogeneity and immune cells landscape of different immune cell populations at different stages of atherosclerosis. Our work paves the way to

better describe the relation between different infectious diseases and cardiovascular diseases.

- We developed a **web-based platform called Infection Atlas** to browse and visualize single-cell RNA-sequencing data. Infection Atlas platform provides a user-friendly interface to study different aspects of infectious diseases at the single-cell level and can potentially promote targeted approaches to intervene in infectious diseases. This platform which is available at infection-atlas.org in the short term provides a user-friendly interface to browse and visualize different aspects of infectious diseases and in the long-term is expected to be a comprehensive atlas of infection in human and mouse across different tissues and different pathogens.

Overall, in this thesis we provide a framework to study infectious diseases at the single cell level with providing novel data analysis methods and this thesis paves the way for future studies to study host-pathogen encounters at the single-cell level.

Zusammenfassung

RNA-Sequenzierung (RNA-Seq) ist in den letzten zehn Jahren zu einer revolutionären Technik für genomweite Genexpressionsanalysen, sowie für Gesamt-Transkriptom-Analysen geworden. In den letzten Jahren ist es mit der Entwicklung neuer Technologien möglich geworden die Genexpression auf Einzelzell-Niveau zu untersuchen. Diese Fortschritte in der Einzelzell-RNA-Sequenzierung haben die Art wie Wissenschaftler biologische Prozesse betrachten von Grund auf verändert. Einzelzell-Sequenzierung wird in unterschiedlichen Bereichen angewendet, um die grundlegenden Mechanismen biologischer Prozesse besser zu verstehen.

Besonders Einzelzell-Sequenzierung ist eine geeignete Methode, um Infektionskrankheiten zu untersuchen. Infektionen sind durch heterogene Mechanismen auf Wirts- und Erreger Seite gekennzeichnet. Der beste Weg die Heterogenität dieser Mechanismen zu verstehen und wie sie interagieren ist die Analyse von Infektionskrankheiten auf Einzelzell-Niveau.

Untersuchungen von Infektionsprozessen auf Einzelzell-Ebene können nicht nur die Heterogenität, sondern auch die Dynamik einer Infektion und das Wechselspiel zwischen Wirt und Pathogen auf molekularer Stufe aufzeigen.

In dieser Dissertation wurden unterschiedliche Einzelzell-RNA-Sequenzierung Technologien implementiert und angewandt um ein besseres Verständnis von Infektionskrankheiten zu erlangen. In der vorliegenden Arbeit haben wir vier unabhängige, aber verwandte Forschungsarbeiten durchgeführt, um unterschiedliche Aspekte von Infektionsbiologie näher zu betrachten.

- Wir nutzten die Vorteile dieser neuen Technologie, um die Konsequenzen einer RSV Infektion bei primären humanen Epithelzellen zu untersuchen. Die

primären humanen Epithelzellen stammten von sechs Spendern und wurden in Luft-Flüssigkeits-Grenzflächen (ALI) Zellkultur mit dem Respiratorischen Syncytial-Virus (kurz RS-Virus) infiziert. In diesem Projekt konnten wir ciliierte Zellen als anfällige Zelltypen einer RSV Infektion zeigen. Wir haben die Viruslast als Indikator für den Fortschritt der Infektion herangezogen, als auch für die Rekonstruktion der Wirtsantwort Dynamik gegenüber einer RSV Infektion. Die Rekonstruktion der Infektionsdynamik zeigte viele Wirtsgene und Signalwege, die durch die RSV Infektion unterdrückt oder induziert wurden. Signalwege, die mit der angeborenen Immunantwort und der Interferonantwort assoziiert waren, wurden durch die fortschreitende Infektion unterdrückt und andererseits waren Signalwege, wie die Zielsteuerung von Proteinen zum endoplasmatischen Retikulum und Apoptose induziert.

- Wir haben eine neue Methode entwickelt, die es ermöglicht das Transkriptom eines Bakteriums auf Einzelzell-Niveau zu sequenzieren und potenziell helfen könnte die bakterielle Heterogenität während des Verlaufs einer Infektion zu charakterisieren. In diesem Forschungsprojekt wurden Bakterien unter folgenden drei unterschiedlichen Konditionen angezogen: Späte stationäre Phase, anaerober Schock und Natriumchlorid Schock. Anschließend wendeten wir ein poly(A) unabhängiges Einzelzell-RNA Sequenzier-Protokoll an, um Bakterien auf Einzelzell-Niveau zu sequenzieren. In dieser Arbeit berichten wir die von wachstumsabhängigen Genexpressionsmustern in einzelnen Salmonellen und Pseudomonaden. Das Ergebnis unserer Analyse zeigte, dass wir nicht nur Transkripte unterschiedlicher RNA-Klassen, sondern auch das Transkriptom von Bakterien in unterschiedlichen Wachstumsbedingungen erfassen können.

- Wir haben Einzelzell-RNA Sequenzierungs-Technologien verwendet, um die Immunzellen Zusammensetzung während des Verlaufs der Athereosklerose zu betrachten. Die Atherosklerose wird als Herzkrankheit betrachtet, die eng mit Infektionen in Zusammenhang gebracht wird. Vorherige Infektionen mit Bakterien oder Viren werden als Risikofaktor für Atherosklerose angenommen. Wir haben für aortische CD45 Zellen von der gesunden und atherosklerotischen Aorta von Mäusen Einzelzell-RNA-Sequenzierungen durchgeführt. Hierbei konnten wir bestimmte Zellpopulationen identifizieren, die spezifisch in atherosklerotischen Mäusen vorkommen. Eine der athereosklerotischen Populationen war eine zuvor unbeschriebene TREM2high Makrophagen Population, die eine erhöhte Trem2 Genexpression zeigte. Diese Population von Makrophagen schien in Funktionen wie Lipid Metabolismus, Katabolismus, sowie Kalzifizierung von Verletzungen involviert zu sein. Diese Arbeit hat die phänotypische Heterogenität und das Feld unterschiedlicher Immunzellpopulationen in unterschiedlichen Stadien der Atherosklerose aufgezeigt. Unsere Arbeit bereitet den Weg, um die Beziehung zwischen unterschiedlichen Infektionskrankheiten und kardiovaskulären Krankheiten besser zu beschreiben.

- Wir haben eine webbasierte Plattform namens „Infektionsatlas“ entwickelt, um Einzelzell-RNA-Sequenzierungsdaten zu visualisieren und zu durchsuchen. Die „Infektionsatlas“ Plattform stellt eine nutzerfreundliche Oberfläche zur Untersuchung von unterschiedlichen Aspekten von Infektionskrankheiten auf Einzelzell-Niveau bereit und kann möglicherweise zielgerichtete Ansätze voranzutreiben, um Infektionskrankheiten zu verhindern. Diese Plattform, die unter „infection-atlas.org“ verfügbar ist, bietet im Moment eine nutzerfreundliche Oberfläche zum Durchsuchen und Darstellen unterschiedlicher Aspekte von Infektionskrankheiten. Langfristig soll es ein

umfangreicher Atlas für Infektionen in Maus und Mesch in unterschiedlichen Geweben und unterschiedlichen Pathogenen.

Insgesamt stellen wir in dieser Dissertation einen Rahmen zur Untersuchung von Infektionskrankheiten auf Einzelzell-Ebene mit neuen Methoden für die Datenanalyse zur Verfügung und bereiten den Weg für weitere Studien um Wirts-Pathogen Interaktionen auf Einzellzell-Niveau zu untersuchen.

Table of contents

Chapter 1: Introduction

From RNA-seq to single-cell RNA-seq

RNA sequencing

The RNA-seq method was developed more than a decade ago and ever since the method has become a ubiquitous tool to study the molecular mechanisms underlying many biological processes (Stark, Grzelak, and Hadfield 2019). The primary application of RNA-seq is to study the differential expression of genes across different tissues and treatment conditions. In an standard RNA-seq workflow, RNA is extracted and the mRNA is enriched via depletion of ribosomal RNA. Then, reverse transcription is applied to synthesize cDNA. Thereafter, adapters are added to the cDNA library and the library is sequenced on a sequencing platform. After sequencing, different bioinformatics methods are applied to detect genes that their expression changes across different conditions.

Although analysis of differentially expressed genes is the primary application of RNA-seq, several other methods are developed to study different aspects of RNA biology. For instance, nowadays, RNA-seq methods are used to study mRNA splicing and gene isoform expression, expression of non-coding RNAs and detection of structural genomic changes like gene fusion (Stark, Grzelak, and Hadfield 2019).

The most novel application of RNA-seq is to profile the transcriptome of tissues and cell types at the single cell level which became possible with technological advances over the last decade. Recent advances in application of single-cell RNA-seq transcriptomics brings up a totally new opportunity to not only study the heterogeneity but also the dynamics of different processes at the single cell level.

Single-cell RNA-sequencing technologies

The common single-cell RNA-sequencing technologies are composed of several steps (usually similar to common RNA-seq methods) including cell lysis, reverse transcription and synthesis of the first and the second cDNA strand followed by amplification of cDNA (Hwang, Lee, and Bang 2018). After cell lysis, poly dT primers are usually used to capture polyadenylated mRNA. Studies have shown that due to technical difficulties, only 10-20 % of the transcripts are reverse transcribed at this stage (Islam et al. 2014). Synthesis of the first strand of cDNA is usually carried out with reverse transcriptase and afterwards the second strand synthesis can be implemented by either poly-A tailing (Sasagawa et al. 2013) or template switching mechanism (Ramsköld et al. 2012). Template switching ensures uniform coverage of reads over the transcript length compared to poly-A tailing. After reverse transcription, cDNA is PCR amplified.

SMART-seq2, one of the most widely used methods for sequencing cells at the single-cell level, uses template switching to generate full-length transcripts and is suitable for the discovery of alternative splicing and allele-specific expression using single-nucleotide polymorphisms (Picelli et al. 2013). On the other hand, this method is limited by the sequencing costs and the number of cells that can be sequenced in a single run. This can be a major drawback for studies where an in depth profiling of several different cell types is required.

To overcome these limitations, researchers have developed methods that are focused on sequencing the 3’ or 5’ end of the transcripts. In these methods, usually unique

molecular identifiers (UMIs) which are 4-8 base pair barcodes associated with each transcript are incorporated during the reverse transcription step (Islam et al. 2014). Using UMIs, each read can be assigned to its original cell by limiting PCR bias and therefore improving accuracy. Overall, the main limitation of the current approaches that sequence the 5' or 3' end of the transcript is the fact that they are not capable of providing full-length gene coverage and therefore are not suitable to study allele specific gene expression or isoform usage.

Exponential increase in the number of single-cells profiled per study over the past decade

With the introduction of new technologies in the past few years, our ability to study cells at the single-cell level has exponentially improved. After just a decade since the first single-cell RNA-sequencing study was published, many sensitive and accurate single-cell RNA-sequencing protocols have been introduced. The recent advances not only improved the sensitivity and accuracy, but also the number of cells profiled per study has also dramatically changed (figure 1) (Svensson, Vento-Tormo, and Teichmann 2018). Nowadays, hundreds of thousands of single cells can be profiled in parallel in a single study which increases our ability to catalog different cell types and their corresponding signature. Modern advances in microfluidic technologies have made a substantial contribution in the improvement of different single cell sequencing technologies (Saliba et al. 2014). The commercial platform offered by 10x Genomics allows every laboratory to generate single-cell RNA-seq in a high-throughput manner (Zheng et al. 2017). Thanks to these advances in scRNA-seq technology, researchers are taking steps toward more fully understanding what people are made of, by reproducibly cataloging the diversity of cell types and gene expression patterns within them.

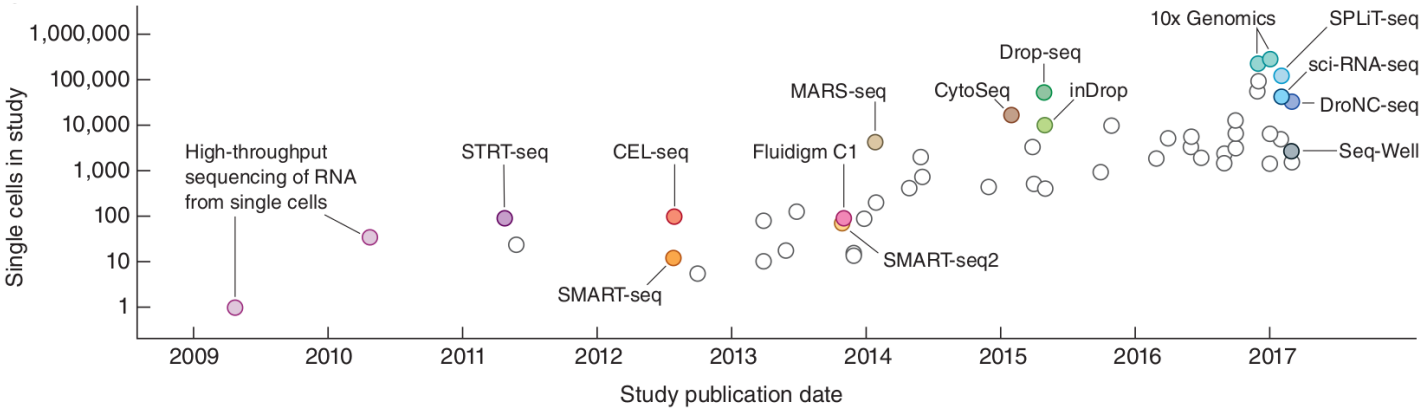

Figure 1. Exponential increase in the number of cells profiled per study as a result of technological improvements. Figure reprinted from Svensson, Vento-Tormo, and Teichmann 2018.

Single-cell transcriptomics and infection

Understanding the molecular mechanisms underpinning infection is crucial to rational design of therapeutics and to interfere with infectious disease progression. Conventional infection biology studies have relied on population-based approaches and important cell-to-cell variations which arise from both host and pathogen are usually neglected (Bumann 2015; Cadena, Fortune, and Flynn 2017). There are many different sources of heterogeneity in infection processes that could have an impact on the outcome of the disease. For example, in viral infections, diversity can arise from a mixture of mutated viral particles showing different infection abilities (Cristinelli and Ciuffi 2018) or in bacterial infections, heterogeneity could be an outcome of a population of cells showing different levels of resistance or tolerance to antibiotics (Gollan et al. 2019). On the other hand, host cells can also show variation in metabolism, cell cycle, activation status and infection history and a combination of these different elements are contributing factors that determine infection outcome (Cristinelli and Ciuffi 2018). Recent advances in single-cell transcriptomics methods provide an approach to characterize the diversity in host-pathogen encounters and the physiological outcomes of these diverse interactions (figure 2). Single-cell

transcriptomics methods can potentially help to deepen our understanding of heterogeneity in host and pathogen and it can contribute to clarify how the complex immune system reacts to infectious diseases.

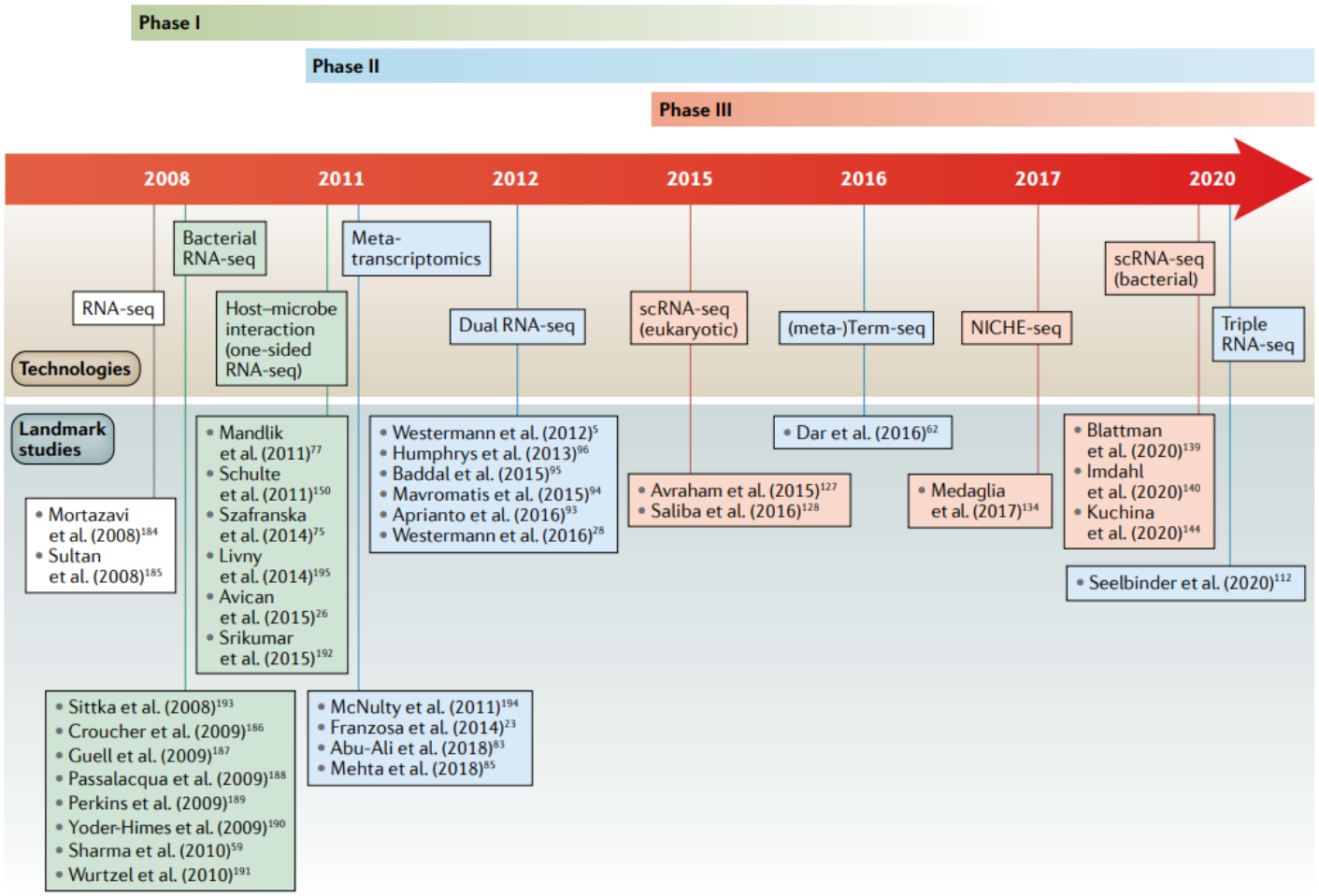

Figure 2. The history of RNA-seq-based infection research. Figure reprinted from Westermann and Vogel 2021.

Application of single-cell transcriptomics in infectious diseases

Single-cell transcriptomics has been already used in different experimental settings and so far, it has proved to be a valuable method to expand our understanding about infectious diseases at an unprecedented level. Using single-cell transcriptomics, researchers have managed to shed light on several aspects of infectious diseases and

host-pathogen interactions (Luo et al. 2020; Lin et al. 2020; Saliba, C Santos, and Vogel 2017). A brief overview of the different applications of single-cell transcriptomics in infection studies is detailed below.

Identification of susceptible cell types

Identification of susceptible cell types is crucial to understand the mechanism of pathogenesis of different pathogens. Single-cell sequencing is particularly powerful in identification of the cell types that are preferentially targeted by the pathogen. In many cases, the pathogen transcripts are polyadenylated. Therefore, pathogen transcripts could be detected by available single-cell RNA-sequencing methods and these transcripts can be associated to the cell types that are susceptible to the pathogen. For example, Angiotensin Converting Enzyme 2 (ACE2) has been characterized as one important receptor for SARS-CoV-2 (Hoffmann et al. 2020) and cell types with high expression of this receptor are more susceptible to SARS-CoV-2 infection (Shang et al. 2020; Wang et al. 2020). In a recent study, single-cell RNA-sequencing showed that *ACE2* is primarily expressed in several different cell types with highest expression in lung alveolar type II (AT2) cells (Qi et al. 2020; Muus et al. 2021). In another study, the atlas of influenza virus infection was studied in mice. In this study, $CD45^{+}$ and $CD45^{-}$ cells were isolated from the lung of mice infected with influenza virus and it was shown that 5 types of immune cells and 4 types of non-immune cells were targeted by influenza virus. Moreover, studies on these mouse models showed that the viral mRNA content was significantly higher in airway epithelial cells compared to other cell types (Steuerman et al. 2018). These examples demonstrate the ability of single-cell RNA-seq not only to identify the susceptible cell type but also to characterize key factors that are required for viral entry.

Characterizing infection dynamics

Understanding infection dynamics is crucially important because on the one hand, helps to understand how a pathogen proliferates and spreads over the course of

infection and on the other hand, shows how the host response changes over different steps of infection progression. However, understanding the dynamics is hindered by the heterogeneity in host-pathogen interaction (Bumann 2015) and this is the area in which single-cell RNA-sequencing is particularly powerful. The spread of pathogens between different cells and pathogen recognition by immune cells can be studied through single-cell RNA-sequencing. For instance, the dynamics of interaction between respiratory epithelial cells and influenza A virus (IAV) has been studied at the single-cell level and in this research, single-cell level patterns of expression of interferons and interferon stimulated genes provided a valuable resource of the contribution of the infected and bystander cells to the host innate immune response over the course of infection and highlighted the importance of early innate immune response in virus spread inhibition (Ramos et al. 2019). In another study, the dynamics of interaction between flavivirus and host cells was studied at the single-cell level and results clarified the dynamics of many host factors specifically related to flavivirus infection progression. For example pathways like membrane trafficking and signal peptide processing were contributing to infection progression (Zanini et al. 2018). Moreover, the dynamics of SARS-CoV-2 infection has also been recently studied via single-cell RNA-sequencing. The studies showed the composition of immune cells in patients at different stages of SARS-CoV-2 infection changes dramatically (Wen et al. 2020). Overall, these examples show the potential of single-cell RNA-sequencing in reconstructing the dynamics of host-pathogen interaction.

Profiling immune response to infection

Characterizing the immune cells landscape not only helps to study the differences in immune cell profile between healthy and infected persons but also provides a chance to study changes in immune cells profile at different stages of infection. The immune cells landscape provides key information to understand the pathogenic mechanisms of infectious diseases. For instance, a recent study used single-cell RNA-sequencing to characterize the immune cell profile of SARS-CoV-2 patients among 5 healthy donors

and 13 patients including moderate, severe and convalescent cases. This study showed that severe patients are marked with profound immune exhaustion, broad T cell expansion, and profound interferon response. Moreover, most cell types were characterized with interferon-α response and acute inflammatory response (J.-Y. Zhang et al. 2020). Overall, single-cell RNA sequencing can identify changes in immune cell landscape during infection and reveal the unique role of these populations to shed light on the mechanisms of pathogenesis.

Discovering novel immune cell subtypes

Diverse immune cell types are involved in different biological processes like, pathogen recognition, antigen presentation and pathogen clearance. For instance, there are several subtypes of tissue resident macrophages with significantly distinct transcriptomes. Single-cell RNA-sequencing not only can facilitate the identification of immune cell subgroups but also helps to characterize their kinetics and function. To demonstrate, it has been shown that $CD4^+$ T-cells infected with HIV could be categorized into two distinct cell types. Interestingly, the transcript level of HIV virus and the number of HIV genes in type I is significantly lower than type II cells. Moreover, type I cells were more difficult to activate (Golumbeanu et al. 2018). This example demonstrate that single-cell RNA-sequencing can significantly facilitate identification of new immune cell subtypes.

In this thesis, we tried to fully take advantage of different capabilities of single-cell RNA-sequencing to study infection:

- We took advantage of this novel technology to find the susceptible cell types in respiratory syncytial virus (RSV) infection and to characterize the dynamics of host cell response to RSV.
- We developed a new method which is capable of sequencing the transcriptome of a bacterium at the single-cell level and can potentially help us to characterize the bacterial heterogeneity during the course of infection.

- We used single-cell RNA-sequencing technology to characterize the immune cells landscape over the course of atherosclerosis. Characterizing the immune cell landscape in cardiovascular diseases can potentially help us to better understand their association with infectious diseases.
- We developed a web-based platform called Infection Atlas to browse and visualize single-cell RNA-sequencing data. Infection Atlas platform provides a user-friendly interface to study different aspects of infectious diseases at the single-cell level and can potentially promote targeted approaches to intervene infectious diseases.

In the following section, we provide an introduction to the different subjects mentioned above.

Respiratory Syncytial Virus (RSV)

Respiratory syncytial virus (RSV) is a leading cause of lower respiratory tract disease in young children and elderly people and it is responsible for roughly 60,000 deaths annually in children younger than 5 years of age. RSV is a pleomorphic enveloped virus (120-300 nm diameter) with a negative sense, single stranded RNA genome (15.2 kb) composed of 11 genes which 9 of them encode structural proteins and two genes encode non-structural proteins (figure 3). The first two genes in the RSV genome are NS1 and NS2 which code the non-structural proteins. These two proteins together are active in modulation of host innate immune response and they inhibit apoptosis and interferon response (Bitko, Shulyayeva, and Mazumder 2007; Spann, Tran, and Collins 2005).

The virion is covered by a lipid bilayer which is composed of F (fusion protein), G (attachment protein) and SH (a small hydrophobic protein) proteins. These three proteins are important for virus infectivity and the presence of all of them is necessary for efficient fusion of the virus. G protein mediates the attachment of the viral particle to the host cells by targeting airway ciliated cells. This further mediates

the fusion of infected cells to neighboring uninfected cells which ultimately leads to characteristic RSV synticia. The lipid bilayer is supported by two layers of M and M2-1 proteins. These two proteins are important for viral morphogenesis. Inside the viral envelope, the ribonucleoprotein is located which is composed of genomic RNA associated with nucleoprotein (N). RNA-dependent RNA-polymerase (RdRp) is another complex inside viral envelope that is composed of a large polymerase subunit (L), a phosphoprotein polymerase cofactor (P) and N (figure 3) (Nam and Ison 2019; Battles and McLellan 2019).

RSV mainly infects ciliated airway epithelial cells. It has been proposed that RSV's G protein binds to CX3CR1 which is present on the surface of the ciliated cells and especially on the cilia (Johnson et al. 2015). Other publications also proposed other cell surface receptors which mediate RSV attachment to the cell surface including annexin II, epidermal growth factor (EGF) receptor, Toll-like receptor 4 (TLR4), intercellular adhesion molecule 1 (ICAM-1), nucleolin, and heparan sulfate proteoglycans (HSPGs) (Griffiths, Drews, and Marchant 2017). After attachment of viral particles to the cell surface, the entry of the virus depends on F protein and it's fusogenic capacity. The virus is weakly cytopathic and causes slight cell lysis in human airway epithelial cells (Openshaw et al. 2017). However, the virus is transmitted from cell to cell and fuses neighboring cells to form syncytia. The formation of syncytia is mediated by F protein and a host GTPase called RhoA (Pastey, Crowe, and Graham 1999). It is worth mentioning that the ability of the virus to create syncytia is different between variants and different viral strains show a diverse capacity to form syncytia in epithelial airway.

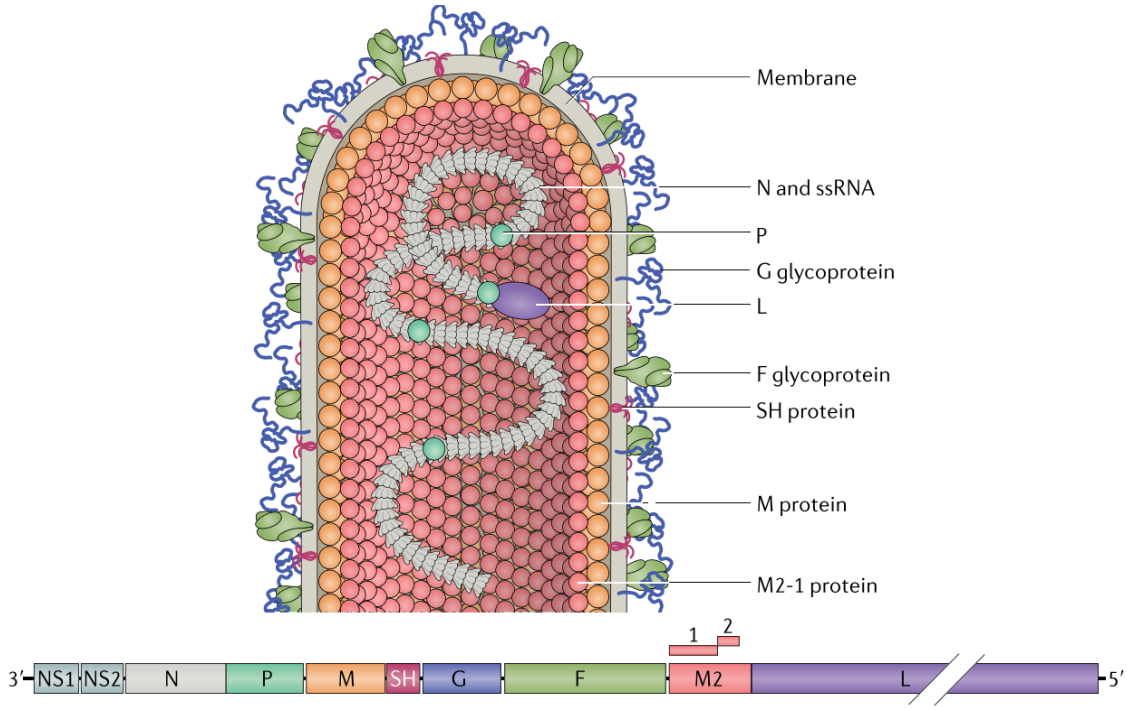

Figure 3. Schematic representation of Respiratory syncytial virus. Localization of different viral proteins on the viral particle (top) and the genomic position of different genes on RSV genome (bottom). Figure reprinted from Battles and McLellan 2019.

Host innate immune response to RSV

The innate immune response in the host cells is triggered after recognition of RSV surface proteins, RSV RNA and proteins by pattern recognition receptors. The TLR4/CD14 recognize extracellular RSV F protein and RIG-1, MDA5 and TLR3 receptors mainly recognize the intracellular transcripts and viral replication intermediates (Troy and Bosco 2016; Arruvito, Raiden, and Geffner 2015).

After recognition, a wide variety of transcription factors are activated which subsequently leads to activation of transcriptional programs that activate the host antiviral programs. Transcription factors like NF-kB, AP-1, JNK, JAK, p38 MAPK and STAT are among the first transcription factors that are activated following RSV recognition. The activation of these transcription factors is followed by expression of

a variety of chemokines and cytokines and type-I and type-III interferons (Rossi and Colin 2017).

Expression of TNF-α, ITAC/CXCL11 and CXCL8/IL-8 helps to recruit neutrophils and other granulocytes to the site of infection to eliminate the infected cells. Expression of type-I interferons results in decreased expression of host proteins and upregulation of pro-apoptotic mediators. Apoptosis following RSV infection is an efficient approach to control infection via decreasing viral replication and production of proinflammatory factors (Bueno et al. 2008; Troy and Bosco 2016).

Interferons have long been known to actively restrict viral infection via induction of expression of interferon stimulated genes (ISGs). Recognizing the viral antigens triggers the expression of different types of interferons. Interferon type-I induces the expression of gene programs which upregulates the expression of many ISGs in cells which subsequently leads to inflammatory response after RSV infection by activation of dendritic cells (DCs), natural killer (NK) cells, and T cells (Goritzka et al. 2015, 2014). Type-I interferon IFN-β also induces the production of B cell survival factor (BAFF) by the airway epithelium cells (McNamara et al. 2013). The production of interferons with epithelial cells is crucial for the course of RSV infection. Interferon Type-III also induces an antiviral program that limits RSV replication (Villenave et al. 2015).

RSV infection also leads to production of other chemokines that recruit monocytes and memory T-cells (Ioannidis et al. 2012). However, secretion of chemokines is not always protective. In RSV infection, secretion of a set of chemokines (RANTES/CCL5, eotaxin-1/CCL11, IL-6, TSLP and IL-17C.) results in recruitment of eosinophils, basophils and mast cells which have no protective role in RSV infection and exacerbate airway damage (Rossi and Colin 2017).

Severe cases of RSV infection have been associated with single nucleotide polymorphism (SNP) in genes related to host cell immune response. For instance airway epithelium cells carrying TLR4 SNPs fail to activate NF-kB signaling pathway

therefore decreased production of interferons leads to inefficient host response to RSV infection (Tal et al. 2004).

RSV modulates innate immune response

RSV takes advantage of different mechanisms to control and inhibit innate immune response. NS1, NS2 and G proteins are the main proteins used by RSV to control or to inhibit host immune mechanisms (Troy and Bosco 2016). Here, a brief overview of how viral proteins modulate the host cell environment in their advantage to evade innate immune response is provided.

RSV interferes with interferon response via NS1 and NS2 proteins

As mentioned before, the type-I interferon response can be inhibited by NS1 and NS2. These two proteins block interferon functions via inhibition of interferons or the signalling pathways in infected cells (Spann, Tran, and Collins 2005). Studies showed that *in vitro* infection with recombinant viruses which lack NS1 and NS2 leads to higher expression of IFN-β (Spann et al. 2004). Moreover, in vitro studies also showed that NS1 and NS2 enhance degradation of STAT2 protein (Whelan et al. 2016). These two proteins also disrupt binding of IRF3 to IFN-β promoter (Ren et al. 2011) which results in disruption in all the pathways controlled by type-I interferons. The effect of NS1 and NS2 on interferon response is not limited to the host cells. It has also been suggested that this inhibition might also have a diverse effect on maturation of monocyte derived dendritic cells (Munir et al. 2008). NS1 and NS2 also modulate TNF receptor-associated factor-3 and inhibitor-κB kinase ε which ultimately lead to decreased type-I interferon production (Swedan, Musiyenko, and Barik 2009). They also affect the innate immune response via interaction with cytoplasmic pattern recognition receptors like RIG-I and MDA5 (Lifland et al. 2012).

The role of G protein in immune modulation

G protein is considered as the most antigenic protein in RSV viral particles. This protein is highly glycosylated and shows an extensive level of variability. These features are important for the virus to be able to escape immune recognition and to neutralizing antibodies (Rossi and Colin 2015). This protein is also capable of inhibiting the host cell innate immune response. G protein performs it's immunomodulatory activity via interfering with normal functions of microRNAs (miRNAs). miRNAs are small single stranded RNA molecules with important functions in determining host cell response to RSV infection (Rossi, Silvestri, and Colin 2015; Bakre et al. 2012). Changes in miRNA synthesis has been linked to increased RSV infection severity, decrease in apoptotic signal and disruption of TLR4 function (Inchley et al. 2015; Liu et al. 2015). But modulation of miRNA activity is not the only way that G protein affects the innate immune response. G protein also works as a TLR antagonist and binds to RSV-specific antibodies to reduce their availability (Shingai et al. 2008).

Sequencing bacteria at the single-cell level

Whereas single-cell transcriptomics have proved to be a valuable method in infection studies and eukaryotic single-cell transcriptomics is revolutionizing the analysis of cell-to-cell variation in infection context, technical limitations have restricted its application to prokaryotes and due to these limitations most of the studies in microbiology are still relied on analysis of bacterial transcriptome at the bulk level. Currently, our understanding about the diversity inside microbial populations is limited and a robust method that can faithfully capture the transcriptome of a single bacterium can potentially bridge this gap (Imdahl and Saliba 2020).

Phenotypic heterogeneity in bacterial pathogens

Using fluorescent reporters, major advances have been made to comprehend bacterial phenotypic heterogeneity in the context of infection (Roche and Bumann 2021). Pathogens show an extraordinary level of variability within the same host tissue site (Gollan et al. 2019). This variability can be either a consequence of fluctuation in transcription factor expression and activity which leads to different gene expression patterns or due to the subtle differences in tissue microenvironments. As a result, pathogen and host encounters can potentially lead to different outcomes in individual cells and have a major impact on antibiotic treatment. For example, a subpopulation of non/slow-growing bacterial populations has a transient resistance to antibiotics that is not genetically encoded (Ackermann 2015).

Diversity in the bacterial population is beneficial for pathogen species. In case of sudden changes in environmental conditions, a subpopulation which has an advantageous phenotype is able to survive and reproduce in the new environmental condition. Bacterial pathogens are exposed to dramatic environmental changes during the course of infection and they have to be able to survive in diverse host niches. Therefore, they have adapted several mechanisms to generate and promote phenotypic diversity (Weigel and Dersch 2018). The molecular mechanisms underlying phenotypic diversity have been the subject of intense research over the past years and scientists have tried to explain the mechanistic molecular events which lead to pathogens phenotypic heterogeneity (Ackermann 2015).

Bacterial pathogens interact with different microenvironments in infection sites and the response of different individual pathogens to environmental conditions is slightly different and this leads to variable gene expression profile and phenotypic properties. Even when the spatial distance between individual cells is decreased as much as possible, they are still exposed to different gradients of concentration of metabolites, ROS/NOS and immune response activities (Bumann 2015). Demonstrated examples

of the effect of microenvironment on phenotype are *Salmonella* proliferating inside macrophage cells (Helaine et al. 2014) and *Yersinia* replication inside microcolonies nearby surrounding neutrophil cells (Davis, Mohammadi, and Isberg 2015).

Another source of diversity between individual pathogens is stochastic events and fluctuations which normally originate from the small number of molecules inside a single cell bacterium (Ackermann 2015; Ozbudak et al. 2002). Stochastic events can potentially lead to different transcriptional signatures and subsequent diverse phenotypes inside a genetically identical population even when cells are growing inside an identical microenvironment (Bódi et al. 2017; Ackermann 2015; Nuss et al. 2016). Specific types of genes are more likely to have stochastic gene expression. Studies showed genes involved in certain pathways like metabolic pathways and stress response pathways are more susceptible to have fluctuation in their expression. On the other hand, housekeeping genes and evolutionary conserved genes are less likely to show this type of gene expression pattern (Ackermann 2015; Raj and van Oudenaarden 2008). One of the best studied examples of fluctuation in gene expression that leads to distinct phenotype is the process of sporulation in *Bacillus subtilis* (Veening, Smits, and Kuipers 2008).

Although, Identification of the evolutionary benefits for pathogens phenotypic heterogeneity is experimentally challenging, scientists have attributed two main benefits to population heterogeneity named division of labor and bet-hedging.

In division of labor, bacteria cooperate with each other to increase the overall fitness of the population to the environmental condition. In this strategy, each subpopulation performs a particular function that is beneficial for the whole population and increases the compatibility of the bacterial population to the environmental condition. It is costly for a bacterium to perform the combination of these functions therefore the labor is distributed among the bacterial individuals inside the bacterial community (Z. Zhang, Claessen, and Rozen 2016). One example of division of labor that promotes pathogenicity of bacteria is the association between pathogen inside

micro colonies in infected tissue where different subpopulations perform specialized tasks like defense against surrounding neutrophils or activation of certain metabolic pathways and the combination of these distinct tasks increases the overall fitness of the pathogens (Davis, Mohammadi, and Isberg 2015).

Bet-hedging is when bacterial individuals express phenotypes that have no immediate benefit for them but this particular phenotype might be beneficial to the bacterium in the later stages of infection or under stress conditions. Bet-hedging increases the overall fitness of the population to sudden environmental changes. Examples of bet-hedging have been reported for bacterial pathogens like *Salmonella* and *Yersinia* which enter the host from the environment and during the process of infection have to cross several host barriers and are exposed to host immune cells. Therefore, bet-hedging allows a subpopulation of them to survive in different host environmental conditions (Nuss et al. 2016; Stewart and Cookson 2012).

Challenges in studying RNA content of a bacterium

Bacterial heterogeneity has been mainly studied using fluorescence microscopy but using fluorescently tagged bacteria (Roche and Bumann 2021) but this method can only report the information for only few genes.. Furthermore, bacterial heterogeneity can not be captured via conventional bulk RNA-seq that average information over million of organisms. To better understand the heterogeneity inside a pathogen population and to potentially develop mechanisms to interfere with pathogens' ability to escape immune cells, one needs to study bacterial pathogens at the single-cell level.

Several technical difficulties need to be elevated to to perform single bacteria RNA-seq. First, one needs to first dissociate the cell wall via enzymatic digestion or mechanical dissociation and the diversity of cell wall among bacterial species means that different bacteria might require different cell wall dissociation methods and unfortunately, there is no universal method that can be applied to every bacterial

species. Secondly, a bacterium has very low RNA content. The size of a bacterium is roughly 1 µm whereas the size of a mammalian cell is around 10-30 µm. It is estimated that the total RNA content of a bacterium is around 10-100 fg whereas this number for a eukaryotic cell is around 10 pg and last but not least, the estimate for the average mRNA copies per bacterial cell is around 0.4 whereas this number is > 10 in a mammalian cell. Therefore very sensitive methods with a low detection limit are required to capture and amplify the RNA content of a single-cell bacterium. Third, bacterial mRNA molecules are not poly-adenylated making it very challenging to separate ribosomal RNAs which account for more than 90% of total RNA. Most of the single sequencing methods that have been developed so far are dependent on the poly-A tail to amplify mRNA and these methods are not applicable to sequence RNA content of bacterial cells (Imdahl and Saliba 2020).

State-of-the-art methods to sequence bacteria at the single-cell level

In 2020, a substantial effort has been devoted to develop new methods to sequence bacteria at the single cell level which overall lead to successful protocols that are capable of capturing the transcriptional diversity inside a bacterial population. In a research that was conducted in this thesis, using a poly(A)-independent single-cell RNA-sequencing protocol, we report the faithful capture of growth-dependent gene expression patterns in individual *Salmonella* and *Pseudomonas* bacteria across all RNA classes and genomic regions. In parallel efforts were made by other research groups to sequence bacteria at the single-cell level. Microbial split-pool ligation transcriptomics (microSPLiT) (Kuchina et al. 2019) and prokaryotic expression profiling by tagging RNA in situ and sequencing (PETRIseq) (Blattman et al. 2020) are two successful methods that have been published recently. These two methods are based on combinatorial indexing based single-cell RNA-sequencing methods (Rosenberg et al. 2018). Therefore, they are considered as high-throughput methods which are able to sequence the transcriptome of thousands of bacteria in a single run although the number of genes detected per cell in both of these methods is relatively low

compared to the work conducted in this thesis. In these methods, single-bacterium RNA-sequencing was performed across a wide range of microorganisms including both gram-negative and gram-positive bacterial pathogens.

Myocardial infarction and infection

The association between cardiovascular diseases and infection has long been a subject of studies. A large number of Infectious diseases like influenza, pneumonia and other chest infections have been associated with myocardial infarction (Kwong, Schwartz, and Campitelli 2018; Clayton, Thompson, and Meade 2008). A recent study revealed an increased risk of myocardial infarction during the week after infection with influenza virus, respiratory syncytial virus and other respiratory viruses compared to the risk in the first year before or after the onset of infection (Kwong, Schwartz, and Campitelli 2018). Another study showed that the risk of myocardial infarction is 8% higher in patients who were hospitalized due to pneumococcal pneumonia (Musher et al. 2007). Moreover, the link between myocardial infarction and pneumonia was further verified in patients with *Haemophilus influenzae* infection (Corrales-Medina et al. 2012) and studies showed that the risk of myocardial infarction associated with pneumonia is correlated to the severity of infection (Corrales-Medina et al. 2015). Respiratory infectious diseases are not the only type of infection that increase the risk of myocardial infarction. Other types of infections like urinary tract infection (Smeeth et al. 2004) and bacteremia (Dalager-Pedersen et al. 2014) are also strong risk factors for myocardial infarction. Moreover, the advent of the new SARS-CoV2

Potential mechanisms of association between myocardial infarction and infection

Researchers have proposed a causal relationship between infection and myocardial infarction (Musher, Abers, and Corrales-Medina 2019). As described before,

myocardial infarction is associated with a variety of pathogens (including bacterial and viral) and it is also linked to different infection sites. Moreover, more severe infections have higher associations with myocardial infarction. So far, several different mechanisms have been proposed to explain this association.

One possible mechanism which has been proposed is related to the existence of inflammatory cells in athrosclerotic plaque. In this scenario, infection in the body leads to generation of circulating cytokines like interleukin 1, 6 and 8 and tumor necrosis factor α which subsequently causes activation of inflammatory cells in athrosclerotic plaque (Brown et al. 2014). In this case, plaque inflammatory cells up-regulate expression of enzymes like peptidase and metalloproteinase which leads to oxidative burst. All of these factors subsequently contribute to destabilization of the plaque (Libby 2013).

Moreover, studies have attributed a second possible mechanism to explain association between myocardial infarction and infection. Myocardial infarction takes place when myocardial cells demand more oxygen than what blood can possibly supply to the cells. Infection can lead to inflammation and fever and subsequent increased metabolic activity of peripheral organs. The resulting increase in heart rate shortens the filling time during diastole, thereby compromising coronary perfusion. Also pneumonia can cause decreased levels of oxygen in blood and subsequently limits the myocardium oxygen supply. Moreover septic shock has a tremendous adverse effect on coronary blood supply (Musher, Abers, and Corrales-Medina 2019).

Lastly, the risk of other cardiac disorders like arrhythmias, stroke and heart failure also increase after different infectious diseases and therefore the underlying mechanisms that increase the risk of cardiac disease needs to be characterized. This is particularly important in the case of heart failure, because after pneumonia the risk of heart failure is even higher than the risk of myocardial infarction. A detailed understanding of the interaction between the cardiac disease and infection will help to reduce the risk of heart diseases after infection. In this thesis we tried to explore the

heterogeneity and dynamics of immune cells in cardiovascular diseases to pave the way for a better understanding of the mechanisms underlying the correlation between cardiac diseases and infection.

Aortic macrophages and atherosclerosis

Macrophages are the main immune cell type in atherosclerotic lesions and they play a crucial role during all stages of plaque development. Lesional macrophages perform a wide variety of different functions which controls or exacerbates lesion development (Clement Cochain and Zernecke 2017). Therefore, understanding the role of macrophages in vascular inflammation is required not only to develop new therapeutics for atherosclerosis but also to better explain the association between myocardial disorders and infectious diseases.

Diverse functions of macrophages in atherosclerosis

Several different roles have been attributed to aortic macrophages in the context of atherosclerosis. Here we provide an overview of the different functions which have been described for macrophages.

Production of inflammatory cytokines

Inside athrosclerotic lesions, macrophages contribute to the recruitment of immune cells by producing chemokines and also participate in tissue inflammation via secretion of proinflammatory cytokines. The role of athrosclerotic macrophages in inflammation has been described in many researches. Deficiency in production of IL-1α and IL-1β in bone marrow derived cells has been associated with reduced inflammation and plaque formation in *Apoe* $^{-/-}$ mice (Freigang et al. 2013). Moreover, CCL2 over production in bone marrow cells results in higher levels of macrophage recruitment to athrosclerotic plaque which suggests that chemokines produced by leukocytes can increase lesion inflammation (Aiello et al. 1999). Aortic macrophages

also produce anti-inflammatory cytokines and deficiency in IL-10 and IL-13 has been linked to increased plaque formation and inflammation (Cardilo-Reis et al. 2012).

Forming foam cells

Foam cells in the aorta are one of the main hallmarks of atherosclerosis. Various cell surface receptors in macrophages like CD36, scavenger receptor A1(SR-A1) and LDL receptor-related protein 1 (LRP-1) have been associated with the ability of macrophages to uptake LDL (Tabas and Bornfeldt 2016). Studies with mice deficient in expression of these receptors have proved the role of foam cell forming macrophages in atherosclerosis. Studies showed deficiency in LPR-1 receptors has a protective role in *Ldlr*$^{-/-}$ mice against atherosclerosis (Makowski et al. 2001) and studies showed that deficiencies in macrophage specific CD36 receptor leads to decreased foam cell formation in athrosclerotic plaque (M. Febbraio et al. 2000; Maria Febbraio, Guy, and Silverstein 2004).

Proliferation and senescence of macrophages in athroscelerotic lesion

The ability of plaque macrophages to proliferate is also considered an important feature of these cells. It has been proposed that proliferation of macrophages take place in more advanced athrosclerotic lesions. In addition, the athrosclerotic plaque also contains macrophages which have lost their proliferative ability and are known as senescent cells and these cells are considered to be pathogenic (Childs et al. 2015). Studies showed that clearance of senescent cells from plaque lead to decrease in lesion progression and results in stability of plaque (Childs et al. 2016). Overall, it is not clear how macrophage survival impacts athrosclerotic progression and contradictory results have been reported in different genetic mouse models (Clement Cochain and Zernecke 2017).

M1/M2 paradigm and its limitations to explain heterogeneity in athrosclerotic macrophages

A classification that is commonly used in literature to categorize macrophages phenotype is M1/M2 paradigm and its variants which is based on macrophages phenotype *in vitro*. M1 phenotype is characterized by increased expression of proinflammatory cytokines and chemokines, inducible nitric oxide synthase (iNOS) and production of reactive oxygen species (ROS). Stimulation of macrophages with IFNγ and LPS leads to M1 phenotype. On the other hand, stimulation of macrophages with Th2 cytokines IL-4 and IL-13 leads to anti-inflammatory M2 phenotype which is characterized with secretion of IL-10 and TGFβ. Using different stimulation methods, the M2 phenotype has been further categorized into M2a, M2b and M2c subtypes (Martinez and Gordon 2014).

Whereas the M1/M2 paradigm has been extensively used to define macrophage phenotype *in vivo* but one needs to bear in mind that M1/M2 polarization represents extremely artificial ends of the macrophage phenotype spectrum and could potentially be misleading in certain circumstances. For example, in a research on an obesity model adipose tissue, macrophages expressing the M2 marker CD301b were associated with glucose intolerance and weight gain whereas this finding contradicts with the beneficial role described for M2 macrophages in this setting (Knudsen and Lee 2016). Although the M1/M2 paradigm has been proved to be useful in certain *in vivo* circumstances like allergic reaction and parasite infection, it can not be readily transposed to different disease specific contexts where macrophages are exposed to more complex microenvironments (Sica and Mantovani 2012; Martinez and Gordon 2014). Due to this reason, it is very unlikely that the M1/M2 paradigm could be useful to describe the heterogeneity of the macrophage population in a complex microenvironment such as athrosclerotic plaque.

Researchers in the field of atherosclerosis have tried to elaborate on the M1/M2 paradigm by characterizing several different macrophage polarization states which are formed using different stimuli. For example, M4 induced by CXCL4, Mox induced by oxidized phospholipids and Mheme induced by Heme (Chinetti-Gbaguidi, Colin, and Staels 2015; Clément Cochain and Zernecke 2015). This type of categorization tries to take into account the plaque specific stimuli and their role in the control of macrophages function but they are not still sufficient to characterize the heterogeneity of macrophages in the plaque because in these experiments, macrophages from a single source have been studied with a limited number of stimuli and it can not potentially reflect the diversity of plaque macrophages with multiple origins and inside a complex microenvironment (Nahrendorf and Swirski 2016). On top of that, the presence of resident macrophages in arteries adds another layer of complexity to the study of macrophages in atherosclerosis because how they are functionally different from recruited macrophages and how they contribute to atherosclerosis is unknown (Ensan et al. 2016).

TREM2

Triggering Receptor Expressed on Myeloid cells-2 (TREM2) is a cell membrane receptor composed of an extracellular immunoglobulin domain followed by a transmembrane helix and a cytosolic tail. This receptor binds to a wide variety of ligands that many of them are potentially related to diverse pathological states and are markers of different types of tissue damage. These ligands include phospholipids, sulfatides, bacterial lipopolysaccharide (LPS) and DNA (Kober and Brett 2017). Surprisingly, TREM2 is inactive (or has severe restricted activity) in healthy tissues, however, upon tissue damage, the TREM2 pathway plays a pivotal role in detecting and restricting the spread of the damage (Deczkowska, Weiner, and Amit 2020). Therefore, the scientific community has paid extra attention to this protein in recent years due to its role in different disease contexts.

TREM2 triggers a downstream signalling pathway and activation of this pathway can lead to diverse cellular responses which are not necessarily similar in different cell types. In fact, the downstream pathways that are activated upon TREM2 activation are highly dependent on tissue context and cellular state and accessibility of different elements of signalling pathway determines the outcome of signal transmission. On top of that, binding of different ligands to the TREM2 receptor can lead to activation of diverse biochemical pathways with diverse consequences for the cell fate (Deczkowska, Weiner, and Amit 2020).

The importance of this signalling pathway has been identified in the recent years and due to that our current knowledge about this cell surface receptor is limited. Most studies that have been conducted to elucidate TREM2 function so far, are based on *in vitro* and classic biochemical studies and generally it has been pointed out that *in vivo* TREM2 signalling could be way more sophisticated.

Actually, the role of TREM2 in human health was first pointed out via genetic studies on a rare and fatal disease named Nasu-Hakola disease (NHD) which is characterized by progressive dementia and repeated pathological fractures during adolescence. Genetic studies on patients showed that a loss of function mutation on TREM2 and DAP12 (one of the proteins associated with TREM2 signalling pathway) leads to the symptoms of NHD (Klünemann et al. 2005). These early studies pointed out that TREM2 could potentially play a role in tissue development and maintenance whereas later studies on NHD patients demonstrated, lack of functional TREM2 and DAP12 causes inability in development and survival of myeloid cells (Otero et al. 2009).

Studies in mice show TREM2 is associated with two other transmembrane co-receptors, namely DAP10 and DAP12. Interaction of TREM2 with one of its ligands leads to phosphorylation of DAP10 and DAP12 which subsequently triggers recruitment of cellular signal transduction proteins. DAP12 activates a tyrosine kinase named Syk whereas DAP10 triggers a cascade of signal transduction events by activating PI3K (Ulland and Colonna 2018). As mentioned above, the downstream

signalling is dependent on the intracellular state and type of tissue. For example, in Macrophages, TREM2 has anti-inflammatory properties and limits inflammatory cytokines production (Turnbull et al. 2006) whereas in brain, TREM2 expression is restricted to microglia, and drives the acquisition of a protective disease-associated microglia (DAM) state during neurodegeneration (Keren-Shaul et al. 2017). Finally, cleavage of TREM2 receptor with two enzymes (ADAM17 and ADAM10) releases the extracellular domain of TREM2 and this process contributes in blocking TREM2 pathway (Schlepckow et al. 2017) (figure 4).

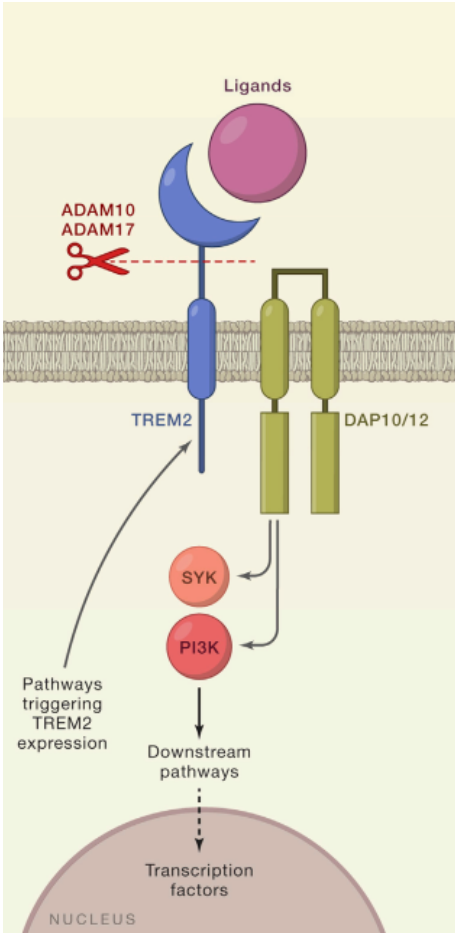

Figure 4. TREM2 signalling pathway, associated proteins and different mediators involved in transferring TREM2 signal. Figure reprinted from Deczkowska, Weiner, and Amit 2020.

Interestingly, studies on the expression of TREM2 in different human tissues using single-cell RNA-sequencing has shown that the expression of this gene is limited to a

small subset of the tissue specific macrophages. Based on these studies, TREM2 expression could be only found in microglia and macrophages in placenta, adrenal gland and adipose tissue (Han et al. 2020). Although, due to the limitation of these studies, it doesn't reflect the full spectrum of the functions TREM2 signalling involved in, but it potentially could give an overview of the sites in the human body that TREM2 could play a role.

In conclusion, TREM2 seems to be an important signalling mediator that is involved in many physiological roles and the cell response to the TREM2 activation is dependent on many contributing factors. The affinity of ligand to TREM2 receptor, the expression level of TREM2 and also the expression level of the downstream molecular machinery, the epigenetic state of the cell, the antagonism of other cell signalling pathway are just a few contributing factors to name and these factors coordinately determine the phenotypic outcome of the TREM2 activation.

Physiological role of TREM2 signalling pathway

It seems, TREM2 signalling is responsible for significant changes in the gene expression profile and function of the cells. This shows that TREM2 signalling is capable of inducing different cellular pathways in different tissue contexts which reflects several TREM2 dependent processes in the human body. Functions like activating phagocytosis, lipid metabolism, restricting inflammation and promoting cell survival have been attributed to the TREM2 signalling pathway (Deczkowska, Weiner, and Amit 2020).

TREM2 deletion has an immediate effect on the phagocytic capability in mice (Kleinberger et al. 2017; Takahashi, Rochford, and Neumann 2005) and interestingly, overexpression of TREM2 in cells that are not phagocytic under normal circumstances (like Chinese hamster ovary cells) can lead to phagocytosis of bacteria (N’Diaye et al. 2009). TREM2 seems to have a role in engulfment of the phagocytic particles as particles attached to TREM2 ligands are more likely to be cleared by cells expressing

this receptor (Kleinberger et al. 2017; Yeh et al. 2016). The mechanism which TREM2 contributes to phagocytosis is still poorly understood and needs to be further studied but this function is likely to be dependent on other proteins that need to be available in the cells. Whereas, TREM2 induces phagocytosis, it simultaneously turns off immune activation in myeloid cells. It has been shown that expression of several anti-inflammatory genes in mice macrophages are in correlation with TREM2 expression. Genes like interleukin 1 receptor antagonist (*Il1rn*), Galectin-1 (*Lgals1*), Galectin-3 (*Lgals3*), progranulin (*Gm*) and many others which are modulators of the inflammation are induced along with TREM2 (Jaitin et al. 2019).

Another function that has been attributed to TREM2 signalling pathway is promoting cell survival. This pathway promotes macrophages survival under stress (like tissue damage or inflammation) *in vivo* and also when the growth factor (CSF-1) has been depleted from the culture media *in vitro* (Kleinberger et al. 2017; Ulland et al. 2017). Also, it has been shown that TREM2 signalling pathway is required for osteoclasts maturation and mice with deficient TREM2 pathway have decreased number of osteoclasts (Otero et al. 2012). Another study propose that the differentiation program between the myeloid cells lacking TREM2 signalling pathway and cells with functional TREM2 pathway is fairly similar whereas, cells without deficiency in TREM2 are more likely to die shortly after differentiation which suggests that TREM2 could play a role in maintaining the myeloid cells viability (Deczkowska et al. 2018).

However, it is worth mentioning that the function of TREM2 signalling is constrained to specific niches in the body and the consequences of deficiency in this pathway can not be seen in many organs in the human body. Most importantly, The pathway becomes critically important under pathological conditions and the role of this mediator has to be studied and explained in this context.

TREM2 in diseases

Deficiencies in the function of TREM2 or the downstream signalling pathway have been implicated to different kinds of diseases. The role of this protein in disorders like neurodegeneration, fatty liver, obesity and cancer has been described and these discoveries have proposed the potential of manipulating TREM2 signalling pathway as a therapeutic measure. In the following section, we discuss some of the research conducted so far on TREM2's role in different diseases.

Evidence showing the role of TREM2 in cancer has been increased in recent years. A rather new study showed that TREM2 expression is substantially increased in peripheral blood monocytes and tumor-associated macrophages in patients suffering from lung cancer compared to healthy tissue and interestingly the expression of TREM2 in macrophages had a positive correlation with tumor progression. This shift in the signature of myeloid cells in lung cancer has been proved in both human patients and mouse models (Yao et al. 2016). Another research showed that the expression of TREM2 mRNA and protein is increased in gastric cancer tissue compared to the healthy tissue (X. Zhang et al. 2018). These studies are, unfortunately, based on bulk data analysis and they do not reveal which cell types are associated with higher levels of expression of TREM2 But recent single-cell RNA-seq studies show that this elevated expression is mainly within different subtypes of myeloid cells (Lavin et al. 2017) and surprisingly, expression of TREM2 in some subtypes of cancerous cells can promote their survival (Duggan et al. 2018). Overall, an increasing number of published papers are providing evidence about the role of this gene in cancer.

The TREM2 signalling pathway in macrophages is also related to some metabolic disorders. For example, $TREM2^{high}$ macrophages can be found in adipose tissue in mice treated with a high fat diet whereas this type of macrophage is absent in mice treated with chow diet. Interestingly, the $TREM2^{high}$ macrophages could also be found in obese human. Macrophages with a similar gene signature can also be found in fatty liver of

mice treated with a high fat diet and also other mice models with liver injury (Jaitin et al. 2019). This shows that TREM2 is also important in metabolism and metabolic disorders.

Implication of TREM2 signalling pathway to neurodegenerative diseases has also been extensively studied. Two independent studies in 2013 used whole exome and whole genome sequencing and showed that variants of the TREM2 gene could increase the risk of developing Alzheimer's Disease (Jonsson et al. 2013; Guerreiro et al. 2013). This finding prompted further studies on TREM2 deficient mice. These studies suggested that TREM2 is crucial in promoting microglial survival (Ulland et al. 2017). In fact, TREM2 deficiencies can lead to several outcomes including impaired microglial proliferation and inability of microglial cells to find lesions and migrate toward them. TREM2 deficiencies also lead to downregulation of homeostatic microglia markers (Ulland and Colonna 2018). Single-cell studies of TREM2 deficient mice demonstrates that a small subset of activation genes like *Dap12* and *ApoE* are upregulated whereas the cells failed to upregulate the majority of activation genes like *Spp1*, *Cst7*, *Axl* and *Lpl (Keren-Shaul et al. 2017)*. It seems that these deficiencies in activation are accumulated over the lifespan of individuals and they facilitate the development of Alzheimer’s disease via uncharacterized mechanisms.

Infection atlas

Infection Atlas (infection-atlas.org) is a web-based platform which provides a simple user interface to browse and visualize different aspects of single cell sequencing data. In the short term, infection atlas is meant to be used by researchers who are not familiar with bioinformatics tools to help them to access single cell sequencing data with just a few mouse clicks! However, in the long term, it is aimed to represent a comprehensive map of host-pathogen interaction at the single cell level in human and mouse. In the following section, we briefly review the current ongoing efforts to

create cell atlases based on single cell sequencing technologies and then we describe our perspective about infection atlas website.

Ongoing efforts to create cell atlases

Mouse was among the first organisms which was used to create a cell atlas and two cell atlases named Mouse Cell Atlas (http://bis.zju.edu.cn/MCA/) (Han et al. 2018) and Tabula Muris (https://tabula-muris.ds.czbiohub.org/) (Tabula Muris Consortium et al. 2018) were among the first attempts to create a partial global map of the cells in an organism. In these two atlases more than 500,000 cells from 40 different mouse tissues were sequenced. The Mouse Cell Atlas for the first time provided evidence for bipotent progenitor cells in mouse lungs and Tabula Muris for the first time revealed the functions of genes like *Chodl* in muscle and *Neurog3* and *Prss53* in pancreas.

However, the efforts to create a cell atlas has not been limited to mice. Malaria Cell Atlas (Howick et al. 2019) tried to create a temporal atlas of malaria parasite *Plasmodium berghei* across the complete life cycle. Currently, the function of around 40% of genes in *P. berghei* is unknown and Malaria Cell Atlas is hoped to be useful in elucidating the function of these unknown genes which could potentially lead to new therapeutics. Moreover, using this atlas the dynamics of gene expression of this parasite over the full life cycle can be studied in more detail.

Nematode worm *Caenorhabditis elegans* is also extensively profiled at different life stages and a report provides RNA expression profiles of more than 130,000 cells at different stages of development (Packer et al. 2019). In this atlas, 27 different cell types are identified including rare neuronal lineages and integrating data with ChIP-seq has revealed many transcription factors which are playing a role in different life stages.

Human Cell Atlas is probably the biggest ongoing effort to create a comprehensive atlas of human cells in different organs (Regev et al. 2018). In this initiative, hundreds of scientists from all over the world joint efforts to profile the heterogeneity of cells

in the human body. Samples from adults, pediatric samples and also developmental samples are going to be profiled in this multinational initiative and healthy human body along with some disease states (particularly in cancer) are going to be analyzed.

Moreover, many laboratories in the world are now devoted to creating atlases from individual organs in the human body. For instance, LungMap (Ardini-Poleske et al. 2017) and Human Lung Cell Atlas (Schiller et al. 2019) are jointly working to create a cell atlas of the human lungs at different developmental stages and in adults. The BRAIN initiative is working to provide a profile of the central nervous system in human.

Potential impact of cell atlases

One of the potential impacts of cell atlases is discovery of new cell types and cell subtypes. Single cell atlases have already made a huge contribution in the discovery of new cell types and their functional states. Pulmonary ionocytes with expression of cystic fibrosis gene *CFTR* (cystic fibrosis transmembrane conductance regulator) are already identified in respiratory airway (Vieira Braga et al. 2019). Rare progenitor populations of cells have been identified in different organs like kidney (Menon et al. 2018) and lung (Reyfman et al. 2019). These studies show that there are still many cell states and cell types that are not identified yet and cell atlases can facilitate their discovery.

Another important aspect of life science that could be addressed via cell atlases is transition of cells between different differentiation stages particularly during development. Single-cell studies have already made a contribution in reconstructing the developmental pathways of different tissues like kidney (Menon et al. 2018), heart (Cui et al. 2019) and brain (Polioudakis et al. 2019).

Elucidation of disease mechanisms is another area of impact of cell atlases. Reconstructing the developmental pathways that lead to a particular disease can be used to do comparison between healthy and diseased tissue which can potentially

help to infer the cell types in which the disease originates (Rajewsky et al. 2020). Cell atlases provide a detailed map of the cells in tissue which can potentially help scientists to understand the underlying mechanisms of disorders.

Why do we need an infection atlas?

Infection Atlas is a single cell based platform that in the long-term is expected to create a harmonized representation of the heterogeneity, dynamics and topology of the infection sites and their corresponding pathogens to facilitate biologists' understanding of the fundamental molecular features that lead to infection outcomes. This platform, in principle, sheds light on the intracellular and intercellular molecular circuits that are involved in host's response to a specific pathogen and the intriguing mechanisms used by the pathogen to evade host's response. Our focus will be particularly on reconstruction of viral and bacterial infection progression and characterizing changes in tissue and pathogen gene expression which ultimately leads to resistance to treatment or chronic diseases. Infection Atlas is a complement of other efforts to create a global map of human cells and is distinct in the respect that it focuses on elucidating pathogen specific molecular and cellular features and the particular immune responses that are activated upon pathogen encounter. Infection Atlas missions can be categorized into three main parts:

- Infection Atlas provides an interactive interface to reconstruct heterogeneity, spatial and temporal aspects of infection. Infection Atlas bridges the gap between experimental infection biologists and computational biologists by providing a user-friendly, interactive user interface that projects high-dimensional single cell sequencing data into simple and intuitive visualization graphs. Different features of infection and heterogeneity in pathogen encounters can be examined via this platform without any prior bioinformatics knowledge and high dimensional multi omics data can be browsed with a few mouse clicks!

- Infection Atlas promotes targeted approaches to intervene infectious diseases. To find novel therapeutic approaches and new targets to intervene infection, a global and comprehensive insight of underlying molecular features and the dynamics of host and pathogen genetic circuits is required. Infection Atlas promotes discovery of new therapeutics by in-depth study of different bacterial and viral pathogen encounters at genomics and transcriptomics level to unravel the ultimate molecular characteristics that define the outcome of infection. These molecular profiles can potentially contribute to development of new therapeutic measures.
- Infection Atlas provides a platform to study pathogen-specific immune response. Recognition and clearance of bacterial and viral pathogens is a result of the cross talk between a sophisticated and complex network of immune cells. Infection Atlas provides an in-depth knowledge of major mechanisms taken by different components of the immune surveillance system to recognize, evade and clear specific microorganisms. This platform investigates modulatory mechanisms that have a protective role against infectious diseases and also sheds light on the interaction between immune cells and tissue within infected organs.

Chapter 2: Material and Methods

In this chapter, different methods used in this thesis to analyze and visualize single-cell RNA-seq data are described. We outline different steps of data analysis from pre-processing of raw data to more advanced steps like data integration and gene regulatory network analysis.

Obtaining count matrix

During my PhD, we used illumina sequencing platform to sequence the sequencing libraries produced in the lab. After sequencing, the data was converted to a matrix of expression values where each column is a cell and each row is a gene. For 10X Genomics data, we used the CellRanger pipeline to create the count matrix. CellRanger uses STAR (Dobin et al. 2013) to align reads to the reference genome and then counts the number of unique molecular identifiers aligned to each gene.

For read-based protocols, in principle a same pipeline for processing bulk RNA-seq data was used. After demultiplexing, data quality was examined using FastQC (v.0.11.7). Sequencing adaptors were removed using cutadapt (v.2.8) and trimmed reads were mapped to the reference genome using STAR aligner (v.2.7.6a) with default settings. Depending on the project, different reference genomes including mouse (mm10), human (GRCh38), *Salmonella enterica* SL1344 (NCBI ASM21085v2) and

Pseudomonas aeruginosa strain PAO1 (ASM676v1) were used. Read counts for each gene were determined using the featureCounts (v.2.0.1) programme.

Quality control

Low quality libraries in single-cell RNA-seq experiments are quite common. These low quality librarie are usually due to the failure during tissue dissociation or library preparation and are marked by low number of expressed genes, high mitochondrial proportions and low number of reads (or UMIs) per cell.

In general, low quality libraries should be removed from single-cell RNA-seq data before performing data analysis. Otherwise, they would interfere with downstream data analysis. For instance

- Low quality libraries can form their own distinct cell clusters after data analysis which potentially makes data interpretation difficult. These low quality clusters are mainly driven by increase in the expression of mitochondrial genes which is a hallmark of dead cells.
- Low quality libraries can also distort the principal component analysis. The first few principal components only capture the differences that are arising from the quality of the data rather than the biology therefore the dimension reduction of the data can be also distorted by low quality libraries.
- Low quality libraries can also have a detrimental effect on normalization of the data. Due to the small library size of low quality libraries, small counts of the transcripts are transformed to large normalized expression values which can make the impression that these genes are upregulated in certain clusters.

To avoid these problems, we tried to select proper quality control measures before performing data analysis which will be explained in the following section.

Selection of quality control thresholds

In our data analysis pipeline, we selected different quality control metrics to identify and remove low quality libraries. These metrics are described below and are general between different projects that are conducted in this thesis.

- The library size is the total number of reads (or UMIs) uniquely aligned to different features in each library. Cells with low library size are considered as low quality libraries because the RNA of these cells has been probably lost during the different steps of library preparation. On the other hand, high library size can also be a sign of a library that is composed of two or more cells.
- The number of expressed genes in each library is also an important feature to assess the quality of a library. In this case, again the libraries with very low or very high number of expressed genes are considered as low quality cells due to the same reason. Low number of genes can show that the RNA of the cell was not properly captured and on the other hand, high number of genes is a sign of doublets.
- The proportion of reads mapped to mitochondrial genes is another criteria that can be used to assess the quality of a single-cell RNA-seq library. High proportions are indicative of low quality cells. The rationale behind is that modest damage to the cell membrane allows the transcripts to go out of the cells but mitochondria is too big compared to the transcripts and it is preserved inside the cells. Therefore, damaged cells have a high proportion of mitochondrial genes.

In the experiments conducted here, we selected fixed thresholds for each one of these quality control metrics and removed the cells that were outside of the selected thresholds. The selection of the thresholds were based on visualization of quality control metrics on diagnostic plots.

Data normalization

In single-cell RNA-sequencing data, differences in sequencing depth between different cells is quite common. These differences can potentially arise from different sources like differences in cDNA capture and amplification and it is always very difficult (or impossible) to achieve consistent libraries. Normalization of the data helps to remove these differences in a way that they won't interfere with downstream data analysis and comparison of the gene signature of different cell types. Proper data normalization ensures that the differences observed between different populations is driven by biology and not technical biases.

In the projects conducted in this thesis, scaling normalization was the main approach to normalize the data which is basically the most common method used to normalize single-cell RNA-sequencing data. In this normalization method, all counts for each cell is divided by a cell specific scaling factor named size factor. The size factor for each cell is an estimate of bias in that cell and the assumption is that dividing the counts by a size factor should remove this bias. The size factor for each cell is directly proportional to its library size where the proportionality constant is defined such that the mean size factor across all cells is equal to 1.

Log transformation

In the data analysis pipeline used in this thesis, we always log transformed the expression values after data normalization and a pseudo-count was added before log transformation. Log transformation is particularly important because in downstream data analysis, the contribution of genes with strong relative difference is promoted. For example, gene A with expression value of 1000 in cell 1 and 1100 in cell 2 has a higher contribution compared to gene B with expression value of 10 in cell 1 and 50 in cell 2.

Feature selection

To perform tasks like clustering and dimension reduction, it is important to select genes that contain information about the biology of the system. Clustering and dimension reduction compare cells based on their expression profile and they try to integrate differences between genes into a single metric. The genes that are selected to perform these calculations have a fundamental impact on the outcome of the analysis.

The goal of the feature selection is to preserve genes which contain biological information and meanwhile reduce the size of the data set to increase the computation efficiency and speed in downstream steps. The feature selection assumes that biologically informative genes have a higher degree of variation among cells compared to the baseline variation that might be due to the technical variation.

Several methods to select the highly variable genes (HVGs) are available. The most basic way to select highly variable genes is to calculate the variance of the log-normalized values for each gene across all cells and select the genes with the highest variation across the population for downstream analysis. The advantage of this method is that feature selection is performed on the same log transformed values that are used in the downstream data analysis. Using log-values guarantees that the quantitative definition of heterogeneity is consistent in the entire data analysis pipeline.

To select the highly variable features in this thesis, we modelled the mean-variance relationship. Here the motivation is the fact that the variance of a gene expression is driven by its abundance rather than its biological heterogeneity. To address this problem, we fitted the relationship between variance and gene abundance for every single gene. The assumption is that for most genes, the variance is dominated by technical variability and our fit actually models the technical noise as a function of abundance. Then, the total variability of a gene can be divided into technical and

biological variability and genes with highest biological variability are selected for downstream data analysis.

Dimension reduction

Single-cell RNA-sequencing data is high dimensional data composed of expression values for thousands of genes where each gene represents one dimension of the data. Dimension reduction goal is to reduce the number of separate dimensions in the single-cell RNA-sequencing data (Andrews and Hemberg 2018). In principle, many genes in the data are correlated with each other therefore there is no need to store the information of each gene separately and information related to multiple features can be stored in a single dimension. Reducing the dimensions of the data reduces the computational need in the downstream data analysis and enables us to visualize the heterogeneity in the data in 2 or 3 dimensional plots. In this thesis, we took advantage of different dimension reduction methods like principle component analysis (PCA), t-distributed stochastic neighbor embedding (t-SNE) and uniform manifold approximation and projection (UMAP) which are briefly explained below.

Principal component analysis (PCA)

Principal component analysis finds axes in the high dimensional space that represent the highest amount of variation in the data. In principal, if we draw a line in the high dimensional space and then move all the cells on that line with the shortest path, the variance across cells along that line is the variance captured by that particular axis. In principal component analysis, the lines which represent the highest amount of variation are discovered. In PCA, the first axis is selected in a way that it shows the highest amount of variation among cells and the second axis is selected in a way that is orthogonal to the first axis and also captures the highest remaining amount of variation.

In Practice, the top first axes in PCA are capable of representing the main elements of heterogeneity in the data and it is possible to restrict the downstream data analysis to the top PCs which makes computational data analysis even more efficient. When applying PCA to single-cell RNA-sequencing data, the assumption is that the first PCs contain most of the heterogeneity in the data and on the other hand the last PCs are dominated by technical noise in the data and are unlikely to capture biological variation. Due to this reason, in single-cell RNA-sequencing data, usually the first PCs are selected and used in the downstream data analysis pipeline. In the projects presented here, the PCA was either applied to log-normalized values or the scaled expression values. In either case, we always performed feature selection before PCA and used the highly variable genes as the input of our PCA.

Choosing the number of PCs

Choosing the number of PCs is one of the parameters that can potentially affect the outcome of the downstream analysis. Using more PCs would retain the biological signal at the cost of including more noise and it is always challenging to decide about the optimal number of PCs that should be included in the downstream analysis. One way to select the number of PCs is to use elbow plot. Elbow plot is a type of visualization which shows the percentage of variance explained by successive PCs in the data. As explained earlier, we assume that first PCs represent much more variance in the data compared to remaining PCs and in the elbow plot, always there is a sharp drop in the percentage of variance explained when we move towards last PCs. In this thesis, we tried to select an arbitrary but reasonable number of PCs after the sharp drop in the elbow plot which typically ranged from 10 to 30.

Visualization with PCA

The simplest approach to visualize single cell data is to plot top 2 PCs in the form of a scatter plot. This approach usually works well for the data sets with low heterogeneity where most of the variability in the data can be explained in the first two PCs but it is

not usually very efficient for complex data sets. The problem is that if the first PC is devoted to resolving the biggest difference between subpopulations, and the second PC is devoted to resolving the next biggest difference, then the remaining differences will not be visible in the plot. One solution is to plot different PCs against each other in a pairwise manner and create multiple scatter plots but this approach is not always desirable because these plots are usually difficult to interpret and some subpopulations might remain undiscovered. In this thesis, we applied PCA for visualization where the first two PCs were sufficient to explain the heterogeneity in the data and for more complex data sets we used more advanced visualization methods like t-stochastic neighbor embedding and uniform manifold approximation and projection.

Visualization with t-stochastic neighbor embedding (t-SNE)

One of the most popular methods to visualize single-cell RNA-sequencing-data is t-stochastic neighbor embedding (t-SNE). This method tries to find a two dimensional representation of the data while preserving the distance between the point in the high dimensional space. The algorithm is composed of two main steps. In the first step, a probability distribution is assigned to pairs of high dimensional points in a way that similar objects in the data have a higher probability and dissimilar objects have a lower probability. In the second step, the algorithm assigns the same probability distribution in the low dimensional space and then it minimizes the Kullback–Leibler divergence (KL divergence) between the two distributions with respect to the locations of the points in the map (van der Maaten and Hinton 2008).

In comparison to PCA, t-SNE is not limited to linear transformation and it doesn't have to precisely show the distance between the points. Therefore, it has more freedom to organize the cells in the low dimensional space and it is able to successfully separate the distinct clusters in the low dimensional space.

t-SNE requires fine tuning some of the parameters. In the projects conducted in this thesis, we always repeated the visualization via different parameters to ensure that the projection of the data to low dimensional space is always representative. We also set the seed in all cases to ensure that the results are reproducible. Furthemore, the perplexity parameter was also adjusted several times to ensure that the choice of perplexity doesn't affect the interpretation of the data.

Visualization with uniform manifold approximation and projection (UMAP)

Uniform manifold approximation and projection (UMAP) (Becht et al. 2018) is an alternative for t-SNE. It is more or less similar to a t-SNE and it also tries to find a two dimensional representation of the data while preserving the distance between the points in the high dimensional space but the two methods are based on different graph weighting equations consequently the visualization from the two methods are usually different. UMAP plots usually have more condensed clusters with more space between the clusters. UMAP is also much faster than t-SNE which makes it a method of choice for large single-cell RNA-seq data sets.

UMAP also requires fine-tuning of some parameters in order to have the best visualization outcome. The number of neighbors and the minimum distance between embedded points are among the most important parameters that need to be fine-tuned in order to find the best visualization. In this report, we always tried a range of different values for each one of these parameters to make sure that changing them doesn't make a substantial change in the overall visualization.

Clustering

The main clustering method used in the projects conducted in this thesis was graph-based clustering which is the default clustering algorithm used by the Seurat package. In nutshell, graph-based clustering builds a graph where each node is a cell and each cell is connected to its nearest neighbor in the high dimensional space via an

edge. In this method, edges are weighted based on the resemblance between the cells and higher weights are given to cells with higher similarity. Then an algorithm is applied that identifies cell communities that are more connected to the cells in the same community compared to the cells in other communities.

In this clustering method, each single cell has to be connected to a minimum number of neighboring cells which is useful because clusters with very low numbers of cells are not allowed. Therefore, clusters consisting of one or two outlier cells wont shape.

One of the most important parameters in this type of clustering is the number of nearest neighbors used to construct the graph. This parameter controls the resolution of clustering and changing this parameter can change the number of clusters identified by the algorithm. In this thesis, we always tested different resolutions and therefore different numbers of clusters and then we tried to inspect the data to select the optimal number of clusters that are meaningful from the biological point of view.

Differential expression analysis

After clustering the data, it is necessary to find the genes that drive separation of the clusters. Finding these genes enables us to biologically interpret the function and the identity of the clusters in the data. Several different methods for differential expression analysis are available. For single-bacterium RNA-seq, DESeq2 bioconductor package (Love, Huber, and Anders 2014) was used to perform differential expression analysis. For other datasets, we mainly used the default differential expression analysis method integrated in the Seurat package. The Seurat package uses “Wilcoxon rank sum test” as default method to perform differential expression analysis.

The statistics of Wilcoxon rank sum test corresponds to area under the curve (AUC) which is basically a probability of a random cell in one cluster having a higher expression of a certain gene than a random cell from another cluster. An AUC equal to

0.5 means that the difference is completely random whereas AUCs equal to 0 and 1 means that the two clusters show completely different expression distributions.

One of the main advantages of Wilcoxon rank sum test is that the test takes the size of the different groups of cells into account. This means that the results of the differential expression analysis is independent of the size of the groups being compared. On the other hand, a disadvantage of the test that should be taken into account is that this test slightly favors the genes with lower abundance and it is important to keep this in mind when interpreting the data (Lawlor et al. 2017). Another disadvantage of Wilcoxon rank sum test is that it is more computationally intensive and therefore slightly slower compared to other differential expression methods available for single-cell RNA-sequencing data.

Gene set enrichment analysis

Gene set enrichment analysis was performed to identify classes of genes that were over represented in a set of differentially expressed genes. Gene set enrichment was generally used to achieve a functional profile of the gene set corresponding to a specific cluster to better understand biological processes and underlying functional states of cells. To Perform gene set enrichment analysis on the marker genes, we mainly used clusterProfiler bioconductor package (Yu et al. 2012).

Data integration

In many projects that were conducted here, data was generated across multiple batches. In principle, preparing libraries in multiple batches is prone to technical differences. These differences lead to systematic differences in expression data which is known as batch effect. Batch effect is particularly important in the downstream data analysis because it can potentially be the main source of variability in the data and therefore it masks the true biological variability.

The batch effect should be computationally corrected before performing downstream data analysis. Different methods for batch effect correction in single-cell RNA-seq data have been developed so far and their cons and pros are extensively studied and corresponding benchmarks have been documented (Luecken et al. 2020). In this thesis, we mainly used Cellranger for batch correction between the data sets. The cellranger aggr command was used to integrate data sets with default settings. This method for integrating data sets, automatically equalizes the average read depth per cell between groups before merging. Using this approach, artifacts that might be introduced due to differences in sequencing depth are regressed out.

Single-cell regulatory network inference and clustering (SCENIC)

SCENIC is a computational method to infer gene regulatory networks and cell types from single-cell RNA-seq data (Aibar et al. 2017). The gene regulatory networks in this method are inferred based on co-expression and DNA motif analysis. In principle, SCENIC first identifies the transcription factors and genes that their expressions are correlated and considers them as candidate regulatory modules. Afterwards, SCENIC looks for enrichment of DNA motifs belonging to the candidate transcription factor nearby the genes inside the candidate regulatory module and in case the DNA motif is enriched, that regulatory module is considered as a hit and reported in the output of the program (figure 5). In this thesis, the SCENIC program was used to detect the gene regulatory networks in single-cell RNA-seq data.

Deconvolution of single-cell RNA-seq data based on genotyping

Deconvolution of single-cell RNA-seq data is used for samples containing a mixture of genotypes. By using variants detected in single-cell RNA-seq reads, it is possible to assign cells to their donor of origin. In this thesis, sourporcell (v.2.0) (Heaton et al. 2020) was used to cluster mixed-genotype single-cell RNA-seq experiments by individuals.

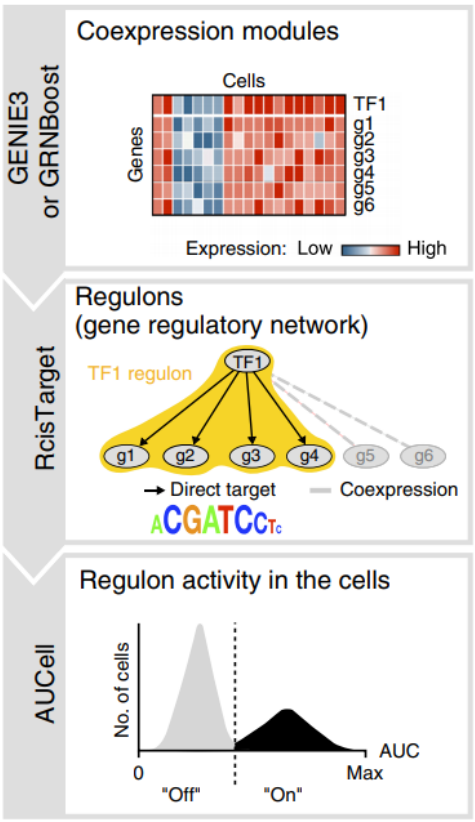

Figure 5. SCENIC workflow. In SCENIC, first genes and transcription factors that are correlated are selected as candidate regulons and then the regulons which the DNA motif corresponding to the transcription factor is enriched are detected as hits. Figure reprinted from Aibar et al. 2017.

Development of interactive data visualization

To develop interactive data visualization in Infection Atlas website, shiny (v.1.5.0) R package was used. Due to the difficulties in storing the full datasets in memory, we used HDF5 data format which offers efficient, on-disk storage, that is scalable to massive datasets. To develop HDF5 files, we used loompy (v.3.0) python package. Loompy develops an HDF5-based data structure to easily store single cell genomics datasets and corresponding metadata.

Data availability

The GEO accession number of the data sets used in this thesis can be found here:

Single-bacterium RNA-sequencing: GSE119888

Heterogeneity of aortic macrophages in murine atherosclerosis: GSE97310

Moreover the scripts that were used to analyse the single-bacterium RNA-seq and RSV infection are deposited on github in the following address:

https://github.com/saliba-lab

Chapter 3: The heterogeneity and dynamics of RSV infection in human respiratory tract

Experiment design

To study the heterogeneity and dynamics of RSV infection at the single-cell level, human airway epithelial cells were isolated from six donors. The isolated cells were cultured in air liquid interface (ALI) cell culture with two different treatment conditions. In one condition, cells were cultured without infection and in the second condition cell culture was inoculated with hRSV-GFP virus. This virus carries *gfp* gene and in principle, the infected cells would express GFP protein. This enables us to isolate the infected GFP^+ cells from GFP^- bystander cells. Cells were harvested at four different time points (1, 3, 5 and 8 days post infection) and fluorescent activated cell sorting (FACS) was used to isolate the GFP^+ infected cells from GFP^- bystanders. After that, cells from different time points were tagged with hashtag antibodies which enables us to backtrack the time point of origin of each single cell after sequencing. After tagging, cells were subjected to single-cell RNA-sequencing (figure 6). Due to the fact that the RSV transcripts are poly-adenylated, our sequencing protocol is capable of capturing and measuring the expression level of the viral transcripts along with the host transcriptional profile.

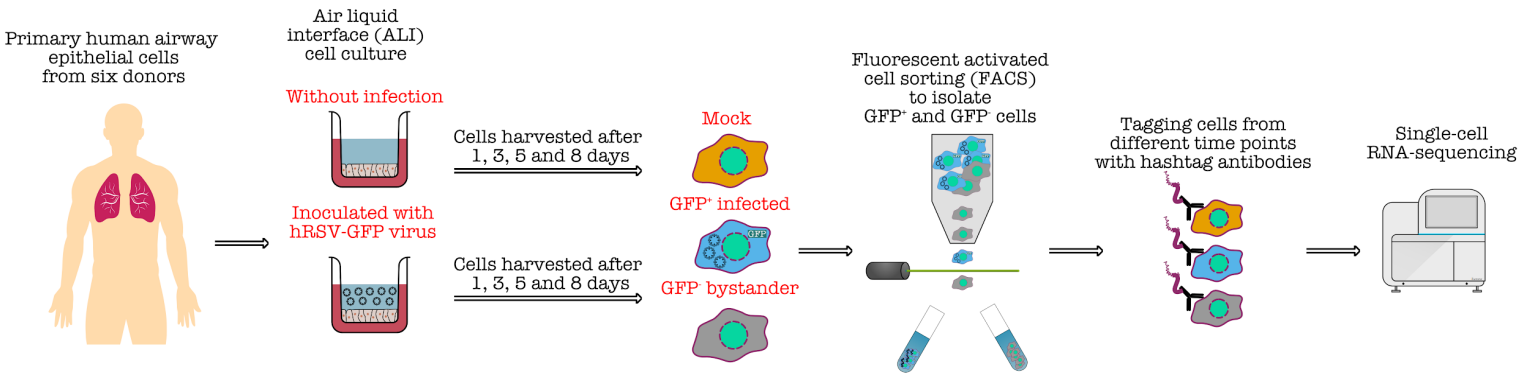

Figure 6. Schematic representation of experiment design.

Single-cell RNA-sequencing is able to discover all major cell types in human lung epithelium

After sequencing, the data from three treatment conditions (mock, bystander and infected) were integrated. Based on the expression of the hashtag antibodies, doublets and negative cells were excluded from the analysis. The common quality control metrics (number of UMIs, number of detected features and the percentage of mitochondrial genes) for each single cell were calculated and low quality cells were removed from the analysis. To remove low quality cells, the following thresholds were selected:

- 1500 < number of genes < 10,000
- 5000 < number of UMIs < 100,000
- Mitochondrial percentage < 30% (figure 7)

Moreover, in this analysis, we examined different treatment conditions for the number of viral UMIs in each single cell (figure 8) and the bystander cells with high number of viral UMIs (more than 50) and the infected cells with low number of viral UMIs (less than 100) were also removed from the data analysis.

After removal of low quality cells, overall 12967 cells (6837 mock, 4067 bystander, 2063 infected) passed our quality control criteria (figure 9).

Based on the expression of hashtag antibodies, cells were assigned to their time point of origin. The number and ratio of cells from different time points were calculated and the results showed that the ratio of cells from different time points was fairly similar between mock, bystander and infected condition (figure 9).

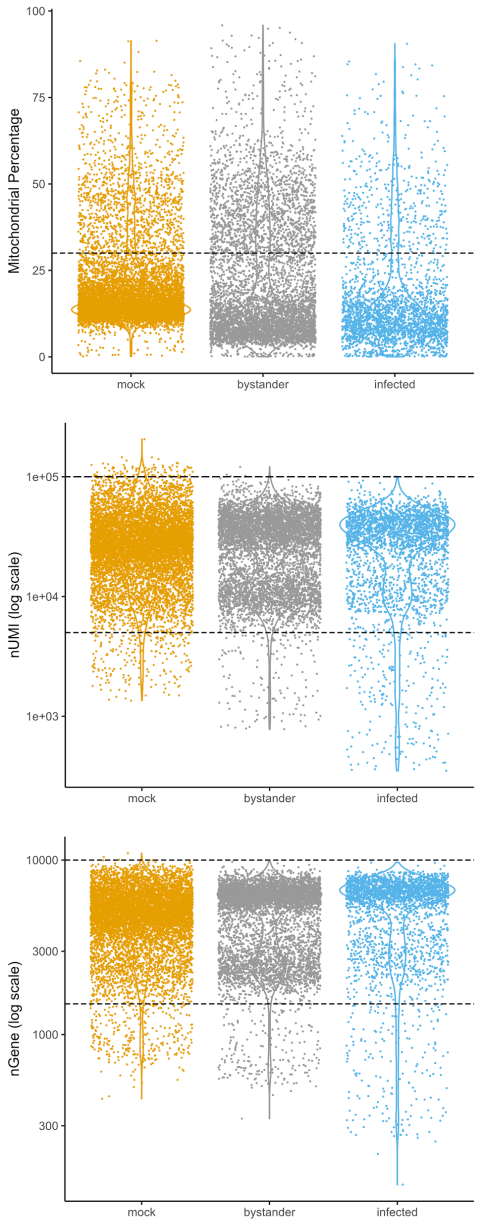

Figure 7. Common quality control metrics (mitochondrial percentage, number of UMIs and number of genes) in mock, bystander and infected cells. The selected thresholds are shown by dashed lines.

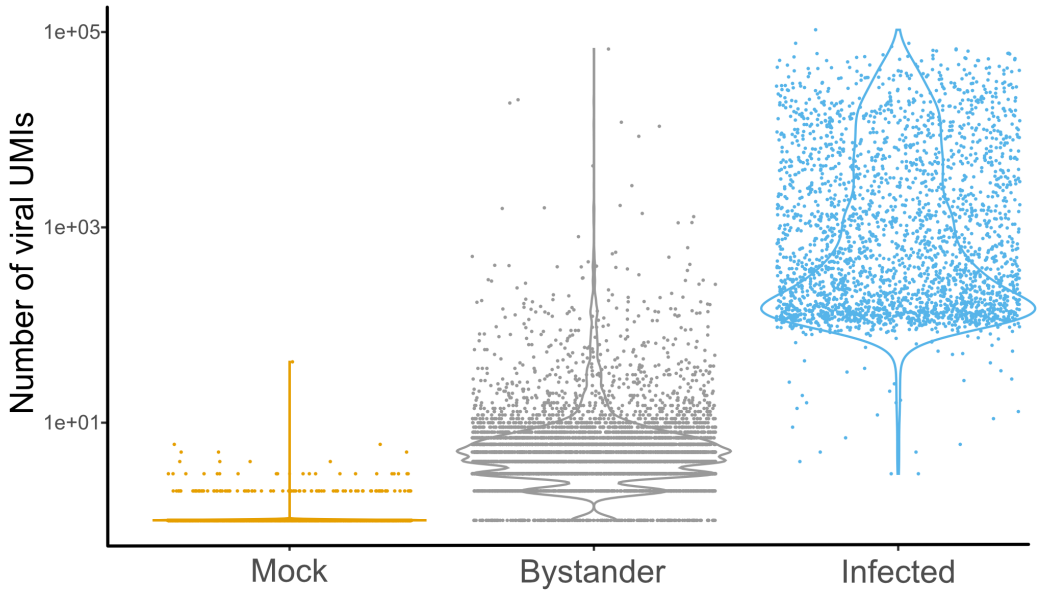

Figure 8. Number of viral UMIs detected per treatment condition. Bystander cells with more than 50 viral UMIs and infected cells with less than 100 viral UMIs were excluded from the downstream data analysis.

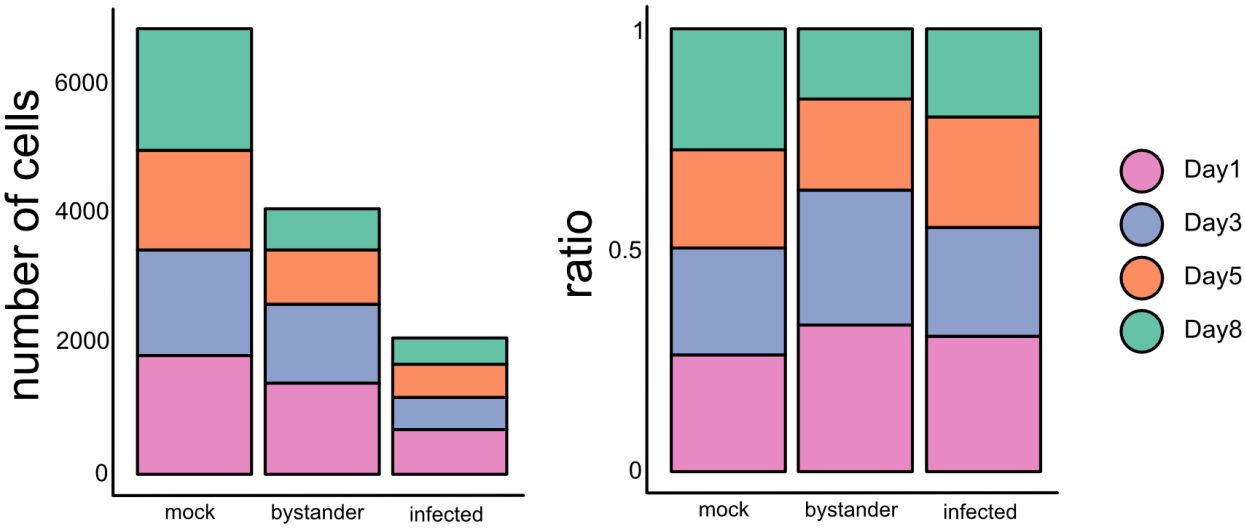

Figure 9. Number and ratio of cells assigned to each time point of origin in mock, bystander and infected.

The integrated data was subjected to dimension reduction and clustering and was projected on two dimensional space. Based on the expression of the canonical markers, all the major cell types in the lung epithelium including basal cells (*KRT5*, *TP63*), ionocytes (*CFTR*, *FOXI1*), ciliated cells (*FOXJ1*, *TP73*, *CCDC78*), mucous cells (*MUC5B*) and goblet and club cells (*SCGB3A1*, *SCGB1A1*, *SPDEF*) were identified. We also singled out a *NOTCH3*$^{+}$ population of cells which is probably an intermediate cell type which gives rise to other epithelium cell types from basal cells (Gomi et al. 2015). We also found two other unassigned populations which didn't express any of the known canonical marker genes (figure 10).

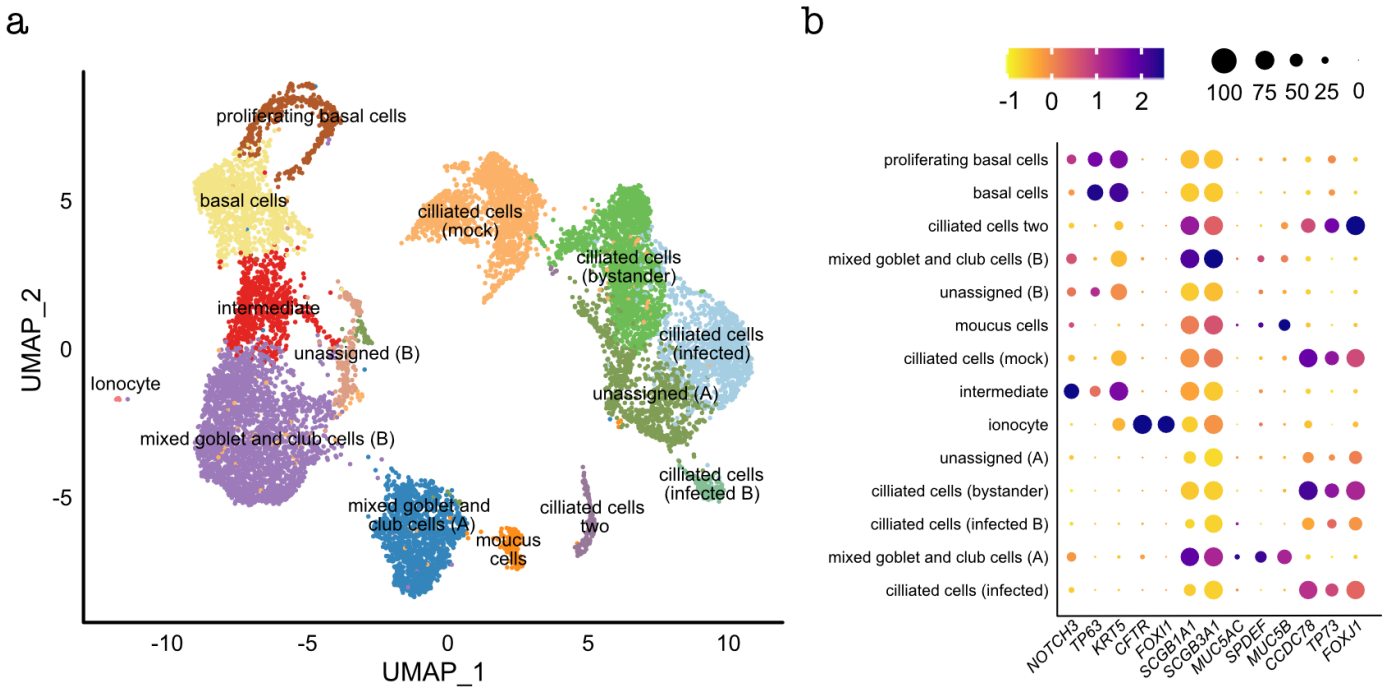

Figure 10. Single-cell RNA-sequencing captures all the major cell types in lung epithelium. a) UMAP visualization of the different clusters identified based on the expression of canonical marker genes. b) Dot plot showing the expression level of canonical marker genes in lung epithelium. Size of the dots corresponds to the number of cells expressing the marker gene.

RSV preferentially targets ciliated cells

To visualize the distribution of viral transcripts on different cell types, the normalized expression values of viral genes were pooled and log transformed. Then, the log transformed values were projected on the UMAP (figure 11). The distribution of viral transcripts on different cell types showed that ciliated cells are the main targets of RSV and the highest amount of viral load could be observed in this cell type. Other than that, the populations annotated as mixed goblet and club cells and one unassigned population were also, to a lower extent, infected with the virus.

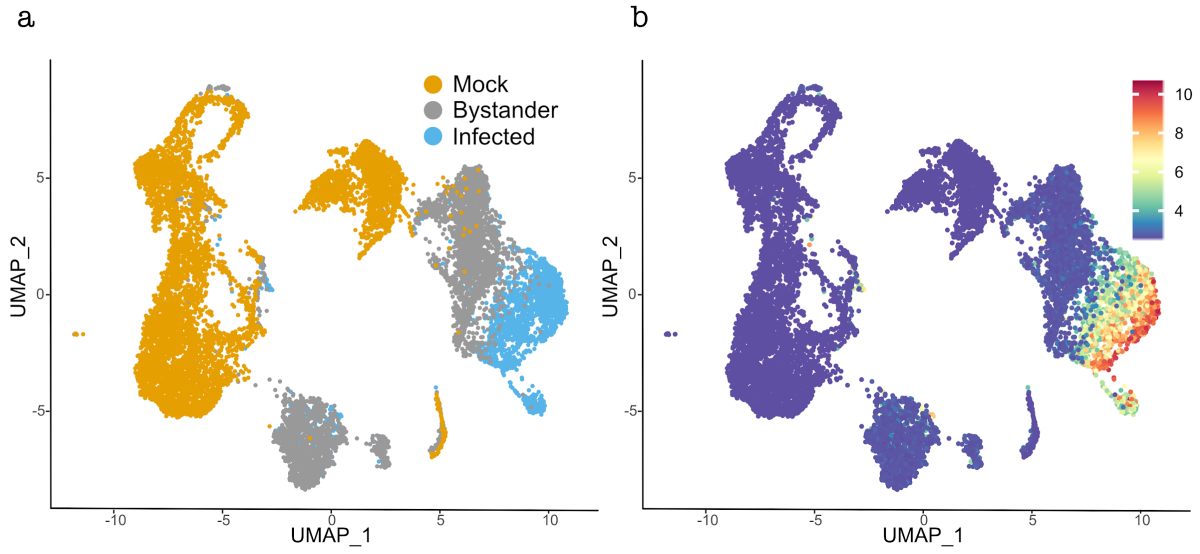

Figure 11. Ciliated cells are the main targets of RSV. a) color code showing different treatment conditions of cells. b) The viral load projected on the umap.

Cells from different time points showed relatively similar gene signature

To examine if the cells from different time points after RSV infection have a different expression profile, the time point of origin of each single cell was projected on the UMAP (figure 12). Overall, cells from different time points showed an even distribution on the UMAP plot. The localization of cells from different time points on

the UMAP suggests that the cells harvested from 1, 3, 5 and 8 days post RSV infection have relatively similar gene expression signatures.

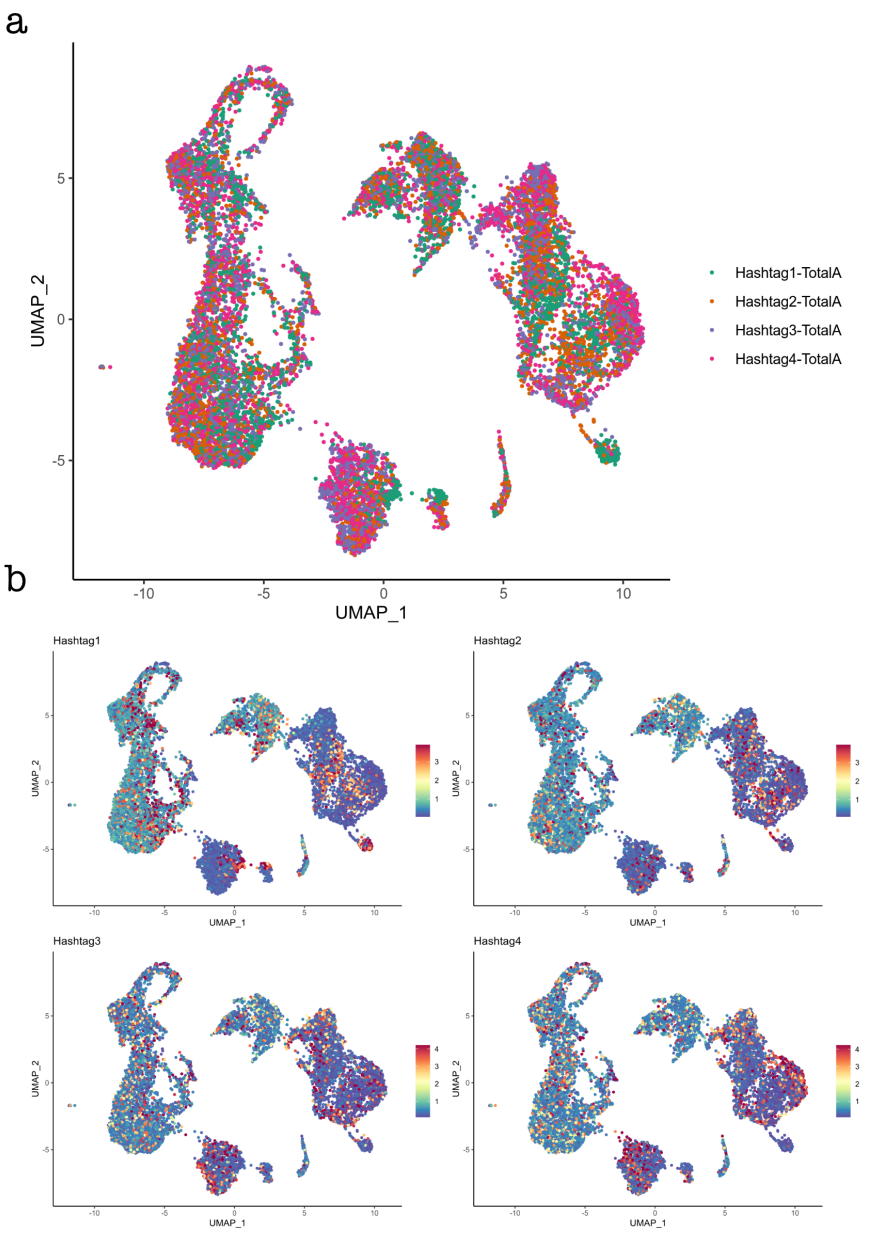

Figure 12. Cells from different time points don't show different gene expression profiles. a) color code presenting the time point of origin assigned to each cell based on the expression of hashtag antibodies. b) the expression value of each hashtag antibody projected on the UMAP. (Hashtag1 = day1, Hashtag2 = day3, Hashtag3 = day5 and Hashtag4 = day8)

Differential expression analysis between ciliated cells

To examine the genes that are up-regulated under different treatment conditions, we performed differential expression analysis between ciliated cells from mock, bystander and infected conditions. Top 20 differentially expressed genes between these cells are shown in figure 13.

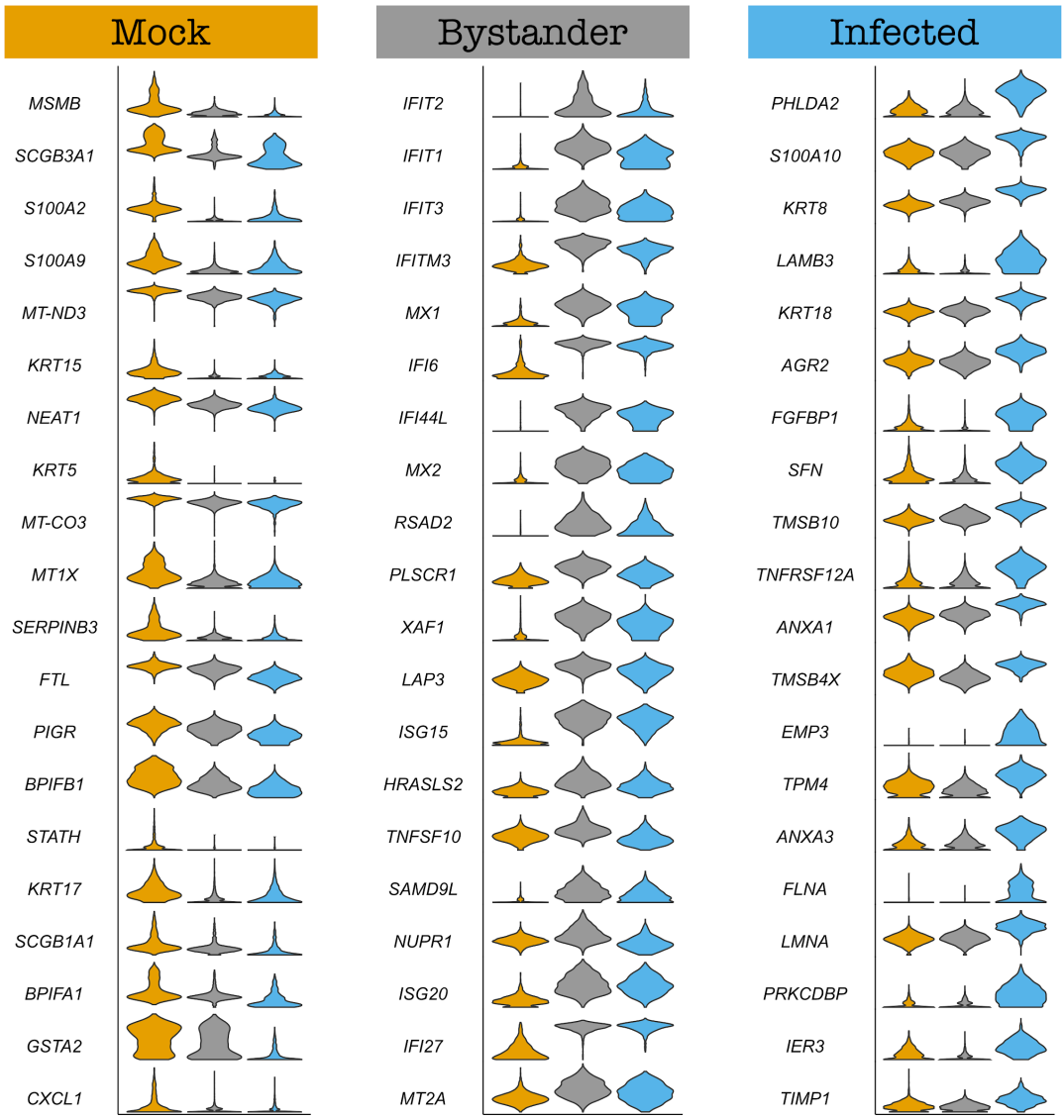

Figure 13. Violin plot showing the top 20 genes differentially expressed in mock, bystander and infected ciliated cells.

As expected, interferon stimulated genes (ISGs) like *IFIT1*, *IFIT2*, *IFIT3*, *MX1* and *MX2*, *ISG15* and *ISG20*, *IFI6* and *IFI27* were among the top differentially expressed genes in bystander cells. Although the expression of interferon stimulated genes was particularly prominent in bystander cells, their expression was also high (slightly lower than bystanders) in infected cells. On the other hand, genes like *TNFRSF12A* and *SFN* that are implicated in apoptosis and *IER3* that corresponds to cell stress response were particularly upregulated in infected ciliated cells.

Certain pathways were specifically induced in infected ciliated cells

We noticed a certain set of genes which were specifically expressed in the infected cells and were neither expressed in mock nor in bystander cells. These genes were extracted via defining a set of stringent thresholds. These thresholds select genes that their median is 2 times higher than bystander cells **AND** 2 times higher than mock cells (figure 14a) (for a full list of genes please see Appendix 1). A gene set enrichment analysis was performed to find the pathways that were enriched in these genes. Important pathways like regulation of calcineurin-NFAT signaling, ERBB signaling pathway and phosphatidylinositol 3-kinase signaling were among the enriched pathways (figure 14b).

Interferons are specifically produced by ciliated cells (B) population

The expression of different interferon genes were examined in our data set and overall, four type-I interferon genes (*IFNB1*, *IFNL1*, *IFNL2*, *IFNL3*) were expressed in our single-cell RNA-seq data. The expression of interferon genes were cluster specific and their expression was only detected in ciliated cells (B) population in our data (figure 15). Among the receptors of interferons, the expression of *IFNLR1*, *IFNGR1*, *IFNGR2*, *IFNAR1* and *IFNAR2* were detected with no cluster specific expression pattern.

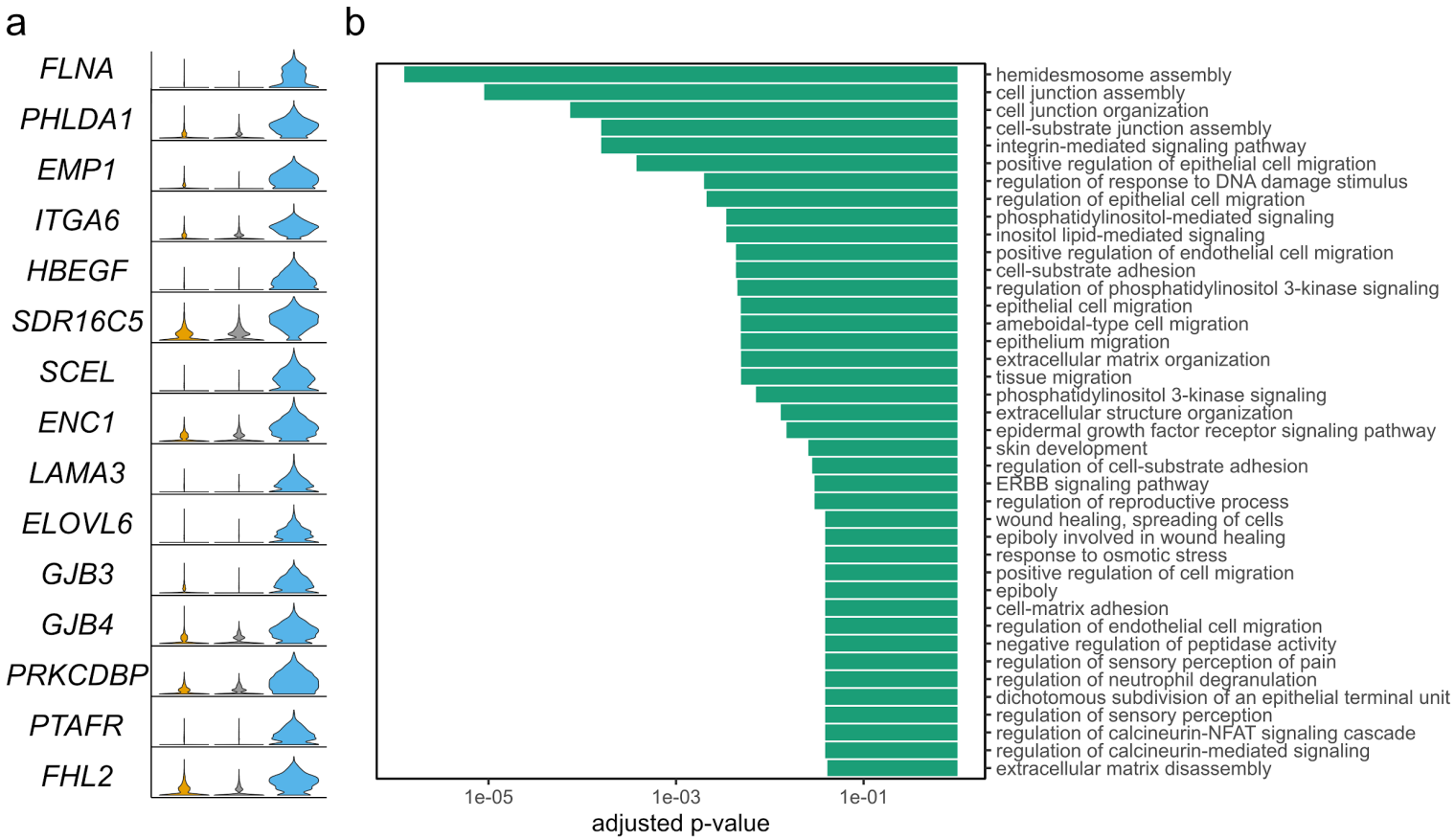

Figure 14. Genes and pathways specifically induced under the infected condition. a) Genes that were upregulated in the infected condition with their corresponding expression level in mock, bystander and infected conditions (for a full list of genes please see Appendix 1). b) Gene set enrichment analysis of the upregulated genes in infected condition.

Infected population

To better understand the heterogeneity and dynamics of viral infection, in the next steps, we focused on the infected population. The infected cells were extracted and dimension reduction and clustering was performed again on the infected data. The expression of canonical markers was used to assign the identity of each cluster. Inline with integrated data, three populations of ciliated cells (namely ciliated cell (A), ciliated cell (B) and ciliated cells (C)), a mixed club and goblet cell population and an

unassigned population were among the cells that were targeted by RSV virus (figure 16).

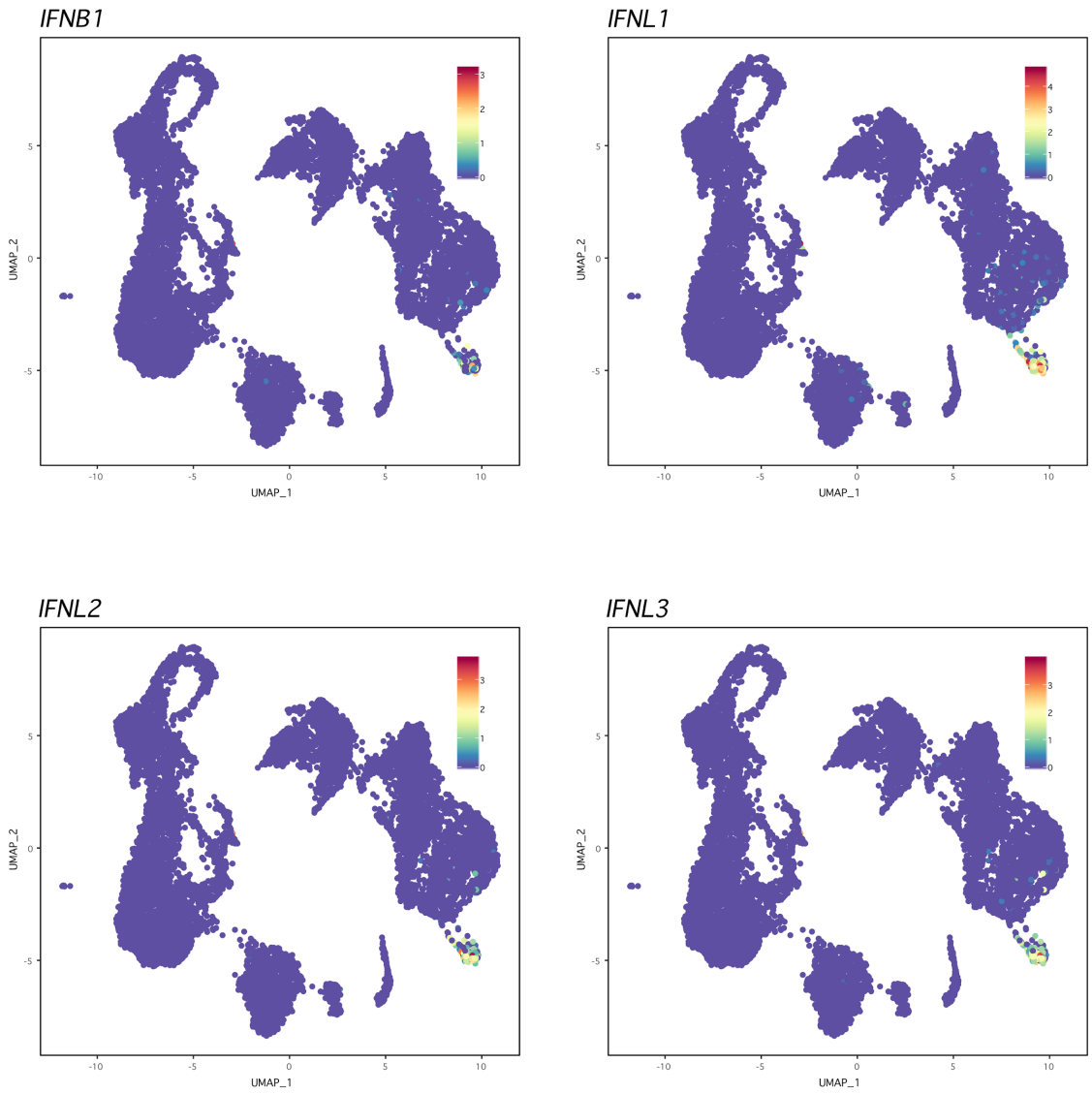

Figure 15. The expression of interferon genes is cluster specific. The expression of interferons was only detected in ciliated cells (B) population.

The infection load for each single cell was calculated and the corresponding load for each population was shown in the form of a box plot (figure 16c). Moreover, The infection load was also projected on the UMAP to have a better visualization of the

severely infected cells (figure 16d). Analysis of the viral load in each population showed again that the ciliated cell (A) was the main target of RSV (figure 16).

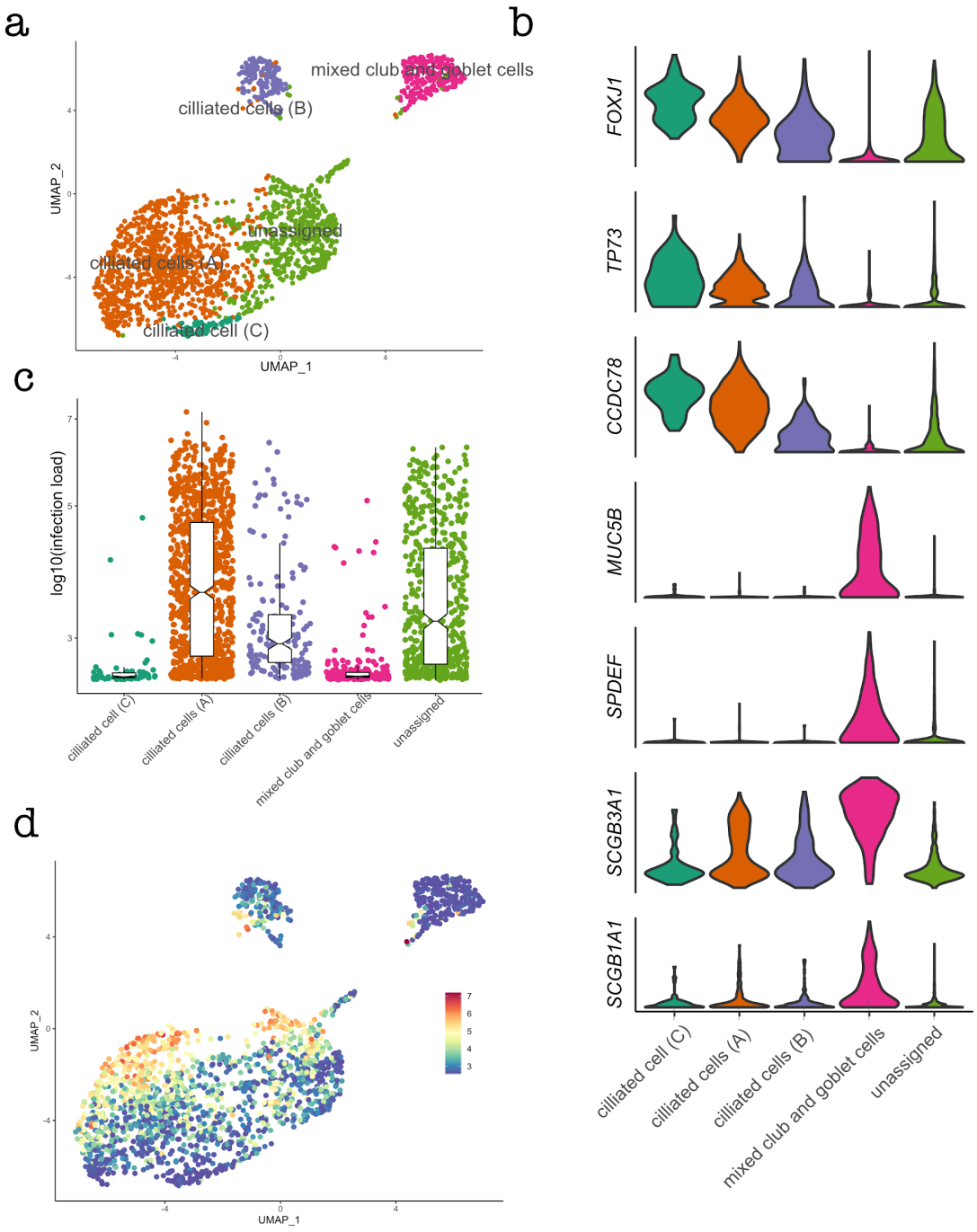

Figure 16. Heterogeneity in the infected population. a) UMAP visualization of different clusters identified in infected data. b) violin plot showing the expression of canonical marker genes. c) boxplot showing the infection load for each population. d) infection load of each cell projected on the UMAP.

Cell surface receptors

Receptors that viruses bind to trigger entry into the host cell are one of the candidates that can be used to develop therapeutics. Many candidate proteins have been introduced to have a role in RSV entry into the host cell, including annexin II, CX3 chemokine receptor 1 (CX3CR1), epidermal growth factor (EGF) receptor, Toll-like receptor 4 (TLR4), intercellular adhesion molecule 1 (ICAM-1), nucleolin, and heparan sulfate proteoglycans (HSPGs). We examined the coefficient of correlation between the corresponding gene of each one of the above mentioned receptors and infection load. Except for *ANXA2* which codes annexin II cell receptor, none of the genes showed a positive coefficient of correlation with RSV infection load. *ANXA2* showed a coefficient of correlation of 0.64 whereas this value for other suggested receptors was below 0.1 (figure 17).

Expression of viral genes gradually decrease from 3’ to 5’ end of viral genome

We also measured the relative abundance of different viral genes in the RSV genome. The abundance of each RSV gene in each single cell was calculated and was shown in the form of a box plot. Our results showed the expression of different viral genes is related to their localization on the viral genome and genes close to the 3’ end of the genome show the highest abundance whereas genes close to the 5’end have the lowest abundance. The analysis showed, the relative abundance of RSV genes gradually decreases from 3’ end to 5’ end with one exception in the abundance of the gene coding the G protein (figure 18).

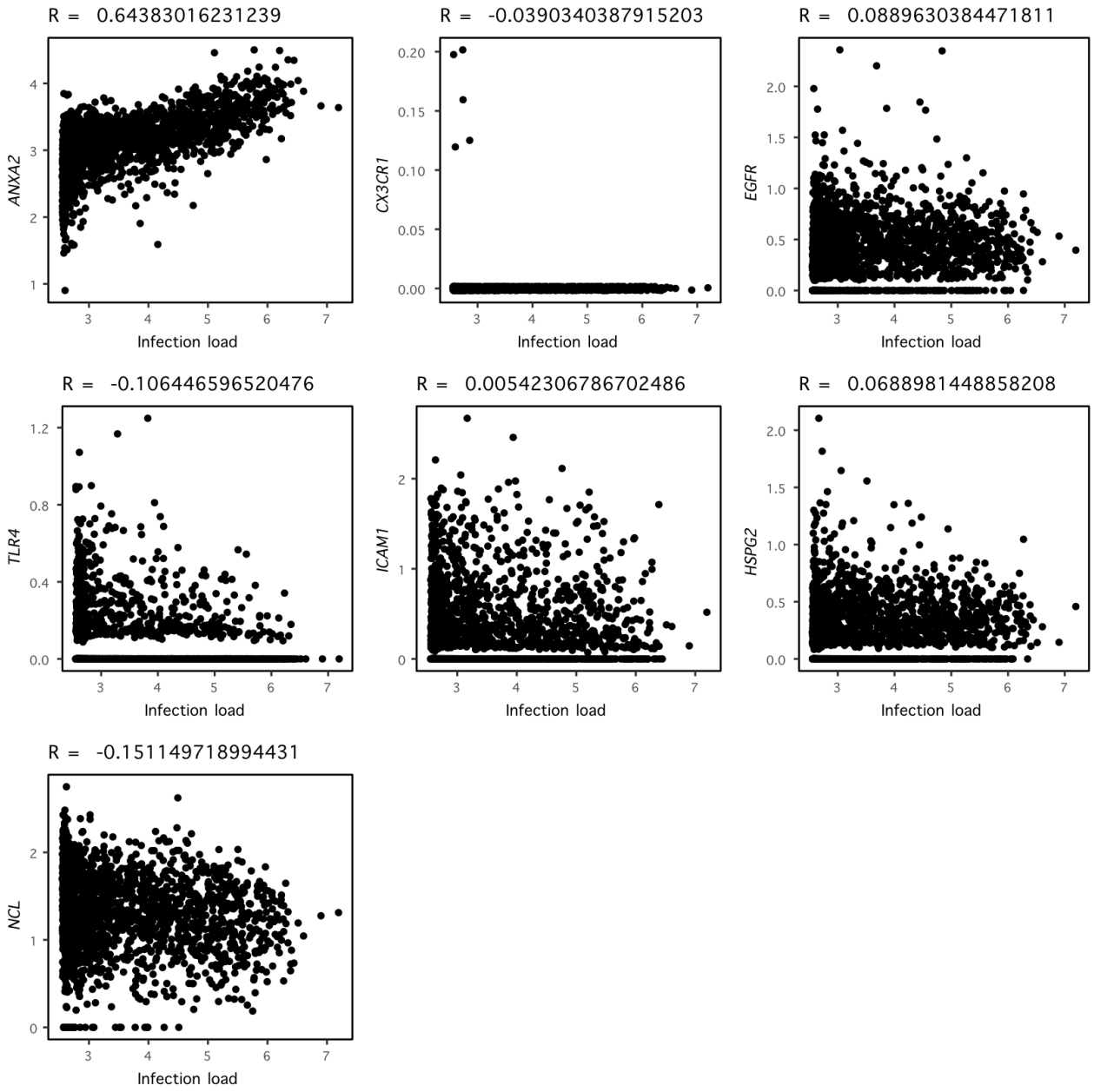

Figure 17. The correlation of different cell surface receptors with viral load.

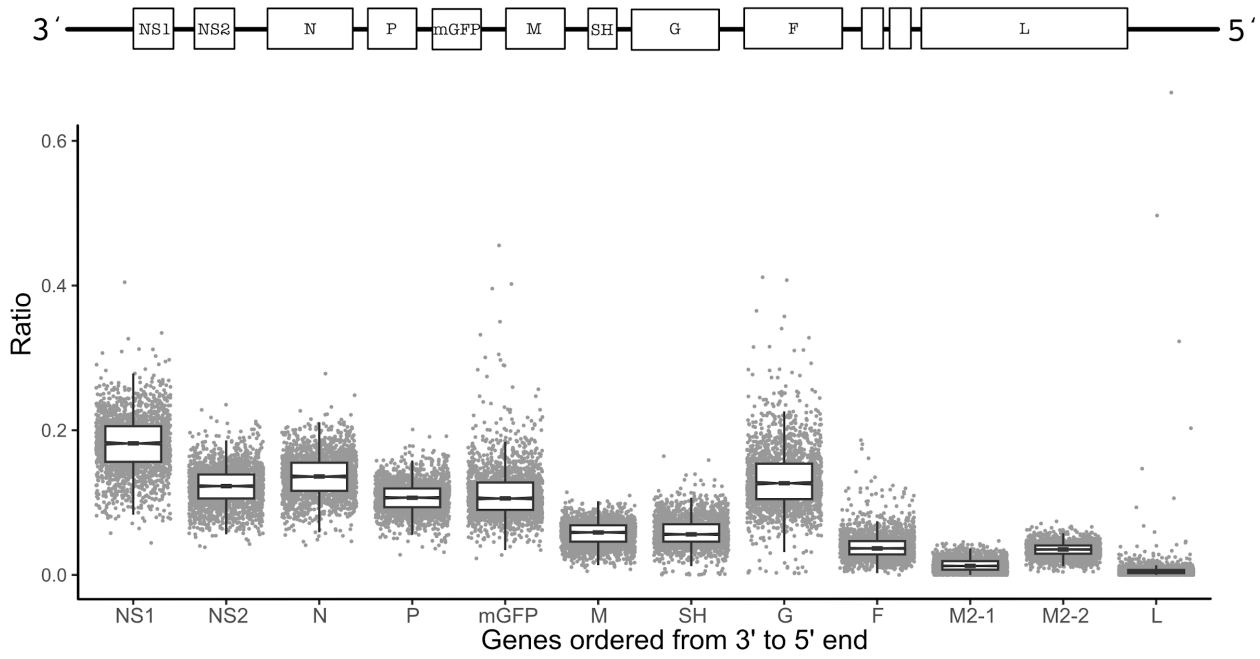

Figure 18. The relative abundance of different viral genes. Schematic presentation of RSV genome (top) and the relative abundance of different genes in our data (bottom).

Infected cell clusters are not patient specific

As mentioned before, the airway epithelial cells were isolated from six different donors. To examine whether different individuals show a specific response to infection, we assigned each single cell in our data to an individual based on their genotype (Heaton et al. 2020) (also see material and methods). The cell assignment showed that the expression profile of infected cells was not patient specific and different patients in this analysis showed relatively similar responses to infection. Overall, we didn't identify any patient specific cluster of cells in our data (figure 19).

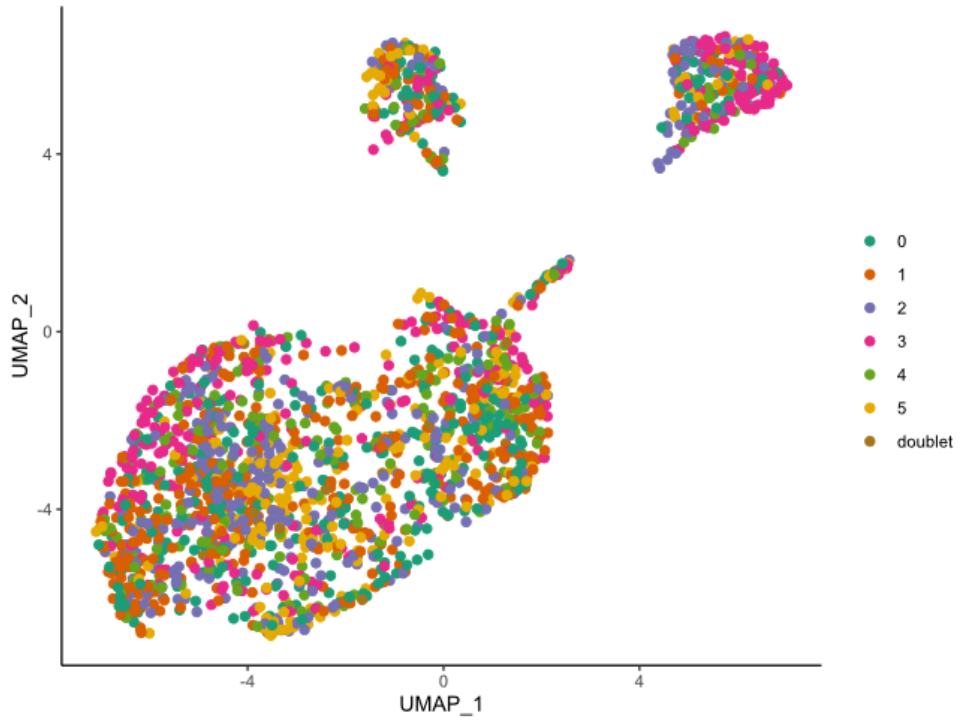

Figure 19. Response to RSV is not patient specific. Number 0 to 5 corresponds to the six donors that were used in this experiment. Doublet corresponds to the droplets that were identified as doublets based on the genotyping.

Infection load as an indicator of infection progression

To see if viral load can be a measure of infection progression, the coefficient of correlation of host genes with viral load was calculated. Certain host genes showed a particularly strong correlation (positive or negative) with the viral load (figure 20). These strongly correlated genes can be an indicator of the genes that their expression increases or decreases over the course of infection.

The expression of highly correlated genes was visualized in a heatmap where the cells were ordered based on their infection load (figure 21). Interestingly, typical interferon stimulated genes like *MX1*, *MX2*, interferon induced *IFIT* family genes and *STAT1* were among the genes which their expression decreased as a function of infection load. On the other hand, genes like *HSPA8* and *HSP90AA1* which are chaperons required for

protein folding and many ribosomal proteins such as *RPL3*, *RPS6*, *RPL9* which might be implicated in viral mRNA translation showed a positive correlation with viral load.

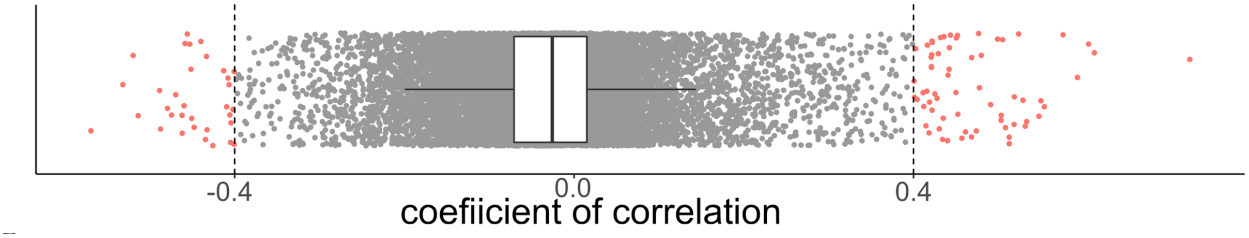

Figure 20. The coefficient of correlation between host genes and viral load. Genes with strong coefficients of correlation (positive or negative) are highlighted in red.

Gene set enrichment analysis reveals the early and late response of host to RSV infection

To see the functional pathways that were enriched at the two ends of the spectrum of viral load, the genes with an absolute coefficient of correlation higher than 0.3 were selected and used for gene ontology enrichment analysis (figure 22).

As expected, the ontologies related to host innate immune response like response to type-I interferon (GO:0034340), type-I interferon signaling pathway (GO:0060337) and negative regulation of viral life cycle (GO:1903901) were among the gene ontologies enriched at low viral load.

On the other hand, a different set of gene ontologies were enriched at high viral load. Ontologies such as protein targeting to ER (GO:0045047), protein refolding (GO:0042026), cellular heat acclimation (GO:0070370), establishment of protein localization to membrane (GO:0090150) were among the most important enriched gene ontologies.

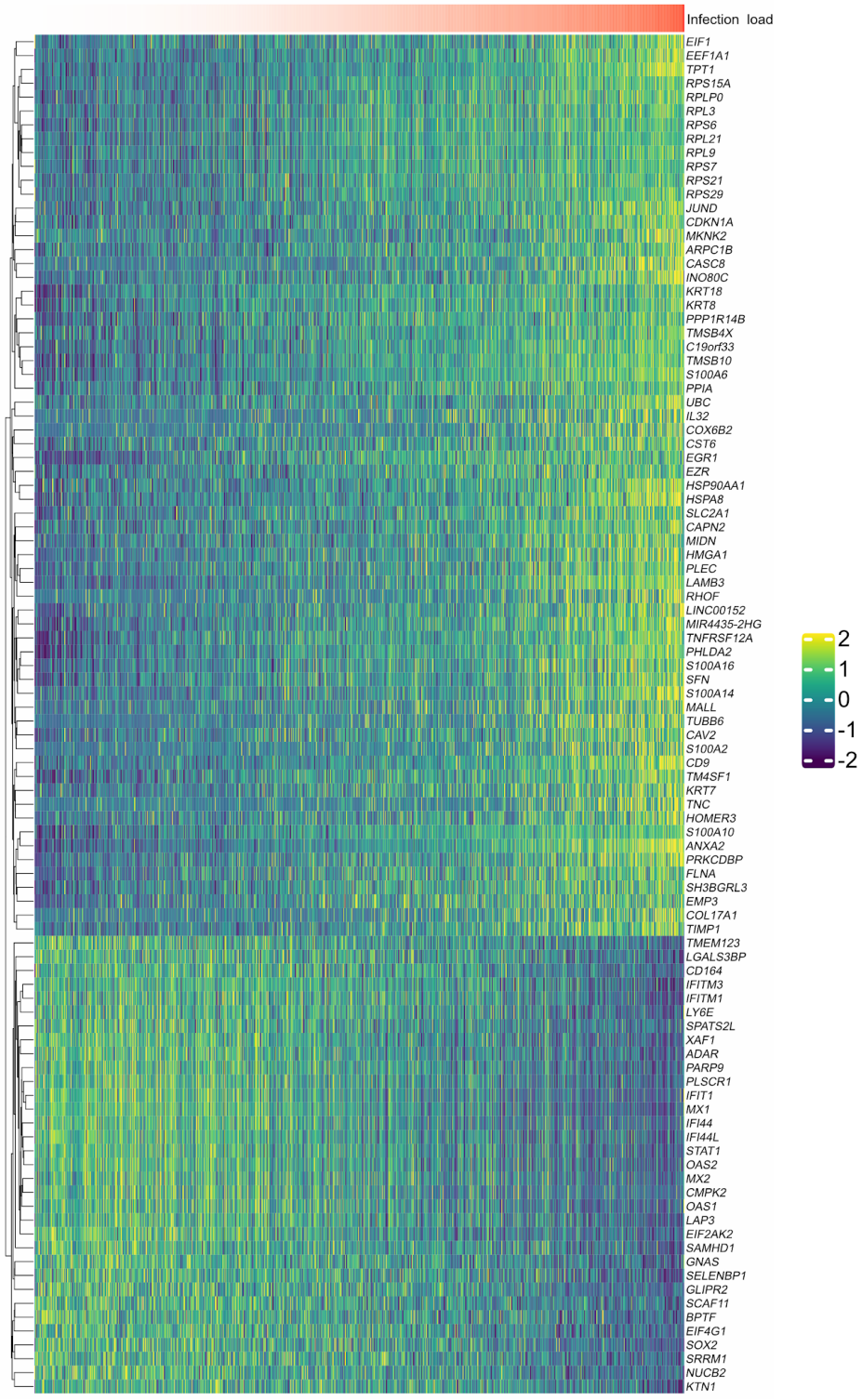

Figure 21. heatmap showing the scaled expression values of strongly correlated genes. The heatmap is ordered according to infection load (annotation bar).

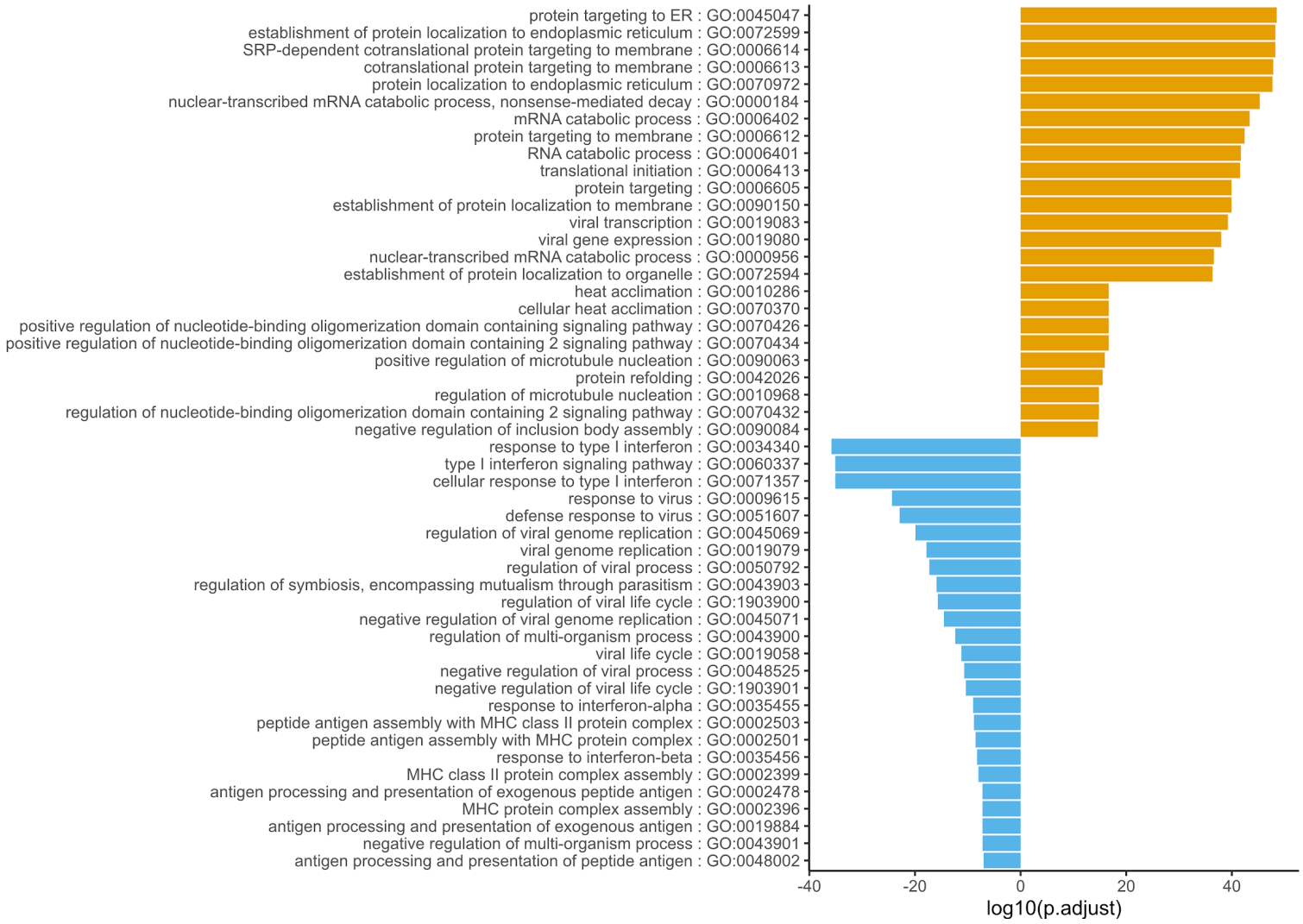

Figure 22. Gene ontology enrichment analysis of strongly correlated genes with viral load. The blue bars show the gene ontologies enriched in negatively correlated genes and brown color represents the pathways enriched in positively correlated genes.

Pseudo-bulk analysis finds the intermediate host responses

As mentioned earlier, analysis of the correlation of host genes with infection load can give us information about the potential early and late responses of the host to RSV infection but doesn't provide any information about the intermediate pathways that

might be activated during the course of infection. To detect the intermediate pathways, we performed pseud-bulk data analysis. In this analysis, the single cells were ordered according to their infection load and then the cells were divided into 20 parts with equal number of cells (~50 cells) on the infection axis. In the next step, cells in each part were pooled to create 20 pseudo-bulk samples. Subsequently, we performed principal component and differential expression analysis on pseudo-bulk samples. A schematic presentation of the workflow is shown in figure 23.

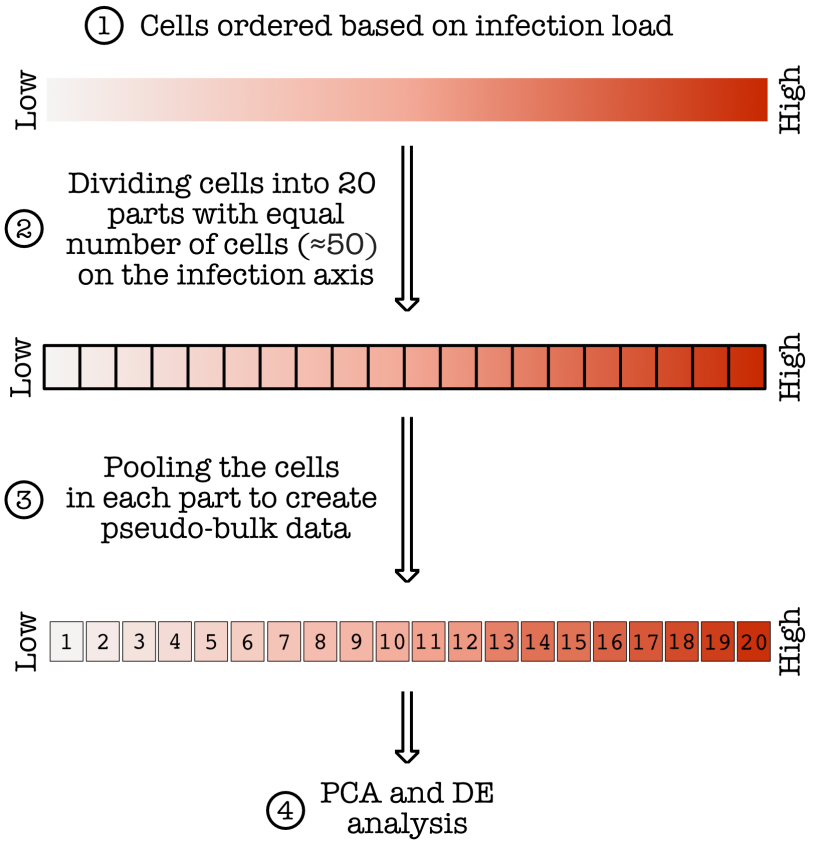

Figure 23. Schematic framework of pseudo-bulk sample preparation and analysis.

After preparing pseudo-bulk samples, principal component analysis was performed on the data. The results of pseudo bulk analysis revealed a trajectory of pseudo-bulk samples where the trajectory starts with pseudo-bulk data with low viral load and it ends with pseudo-bulk samples with high viral load and the samples with relatively close viral load were localized close to each other on the PCA plot (figure 24).

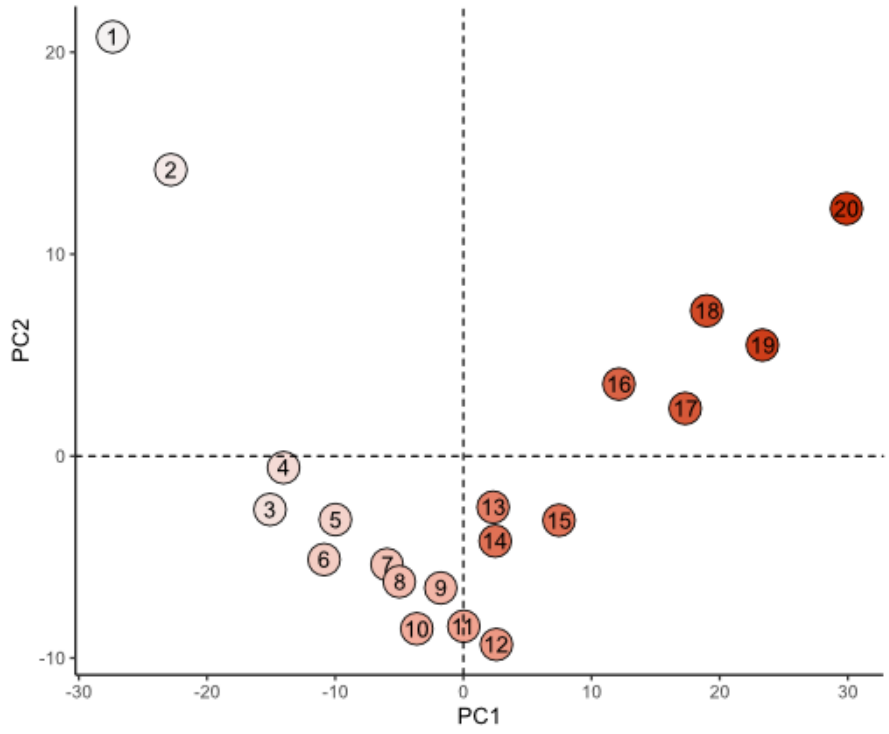

Figure 24. Principal component analysis on pseudo-bulk samples

Principal component analysis suggested a gradual change in the gene expression profile of pseudo-bulk samples over the course of infection. To address this gradual change in gene expression profile, the pseudo-bulk samples were divided into six different groups composed of pseudo-bulk samples that were close to each other on the PCA plot (two groups composed of the samples at the two ends of the trajectory and 4 intermediate groups). Then iterative differential expression analysis was performed between the intermediate groups versus the two end point groups. The differentially expressed genes from each comparison were pooled and used to create a heatmap of pseudo-bulk samples (figure 25). This heatmap is ordered based on the infection load. The heatmap showed five waves of transcription (wave A-E) that were activated sequentially one after each other over the course of RSV infection. To see a complete list of genes in each wave of transcription, please see Appendix 3.

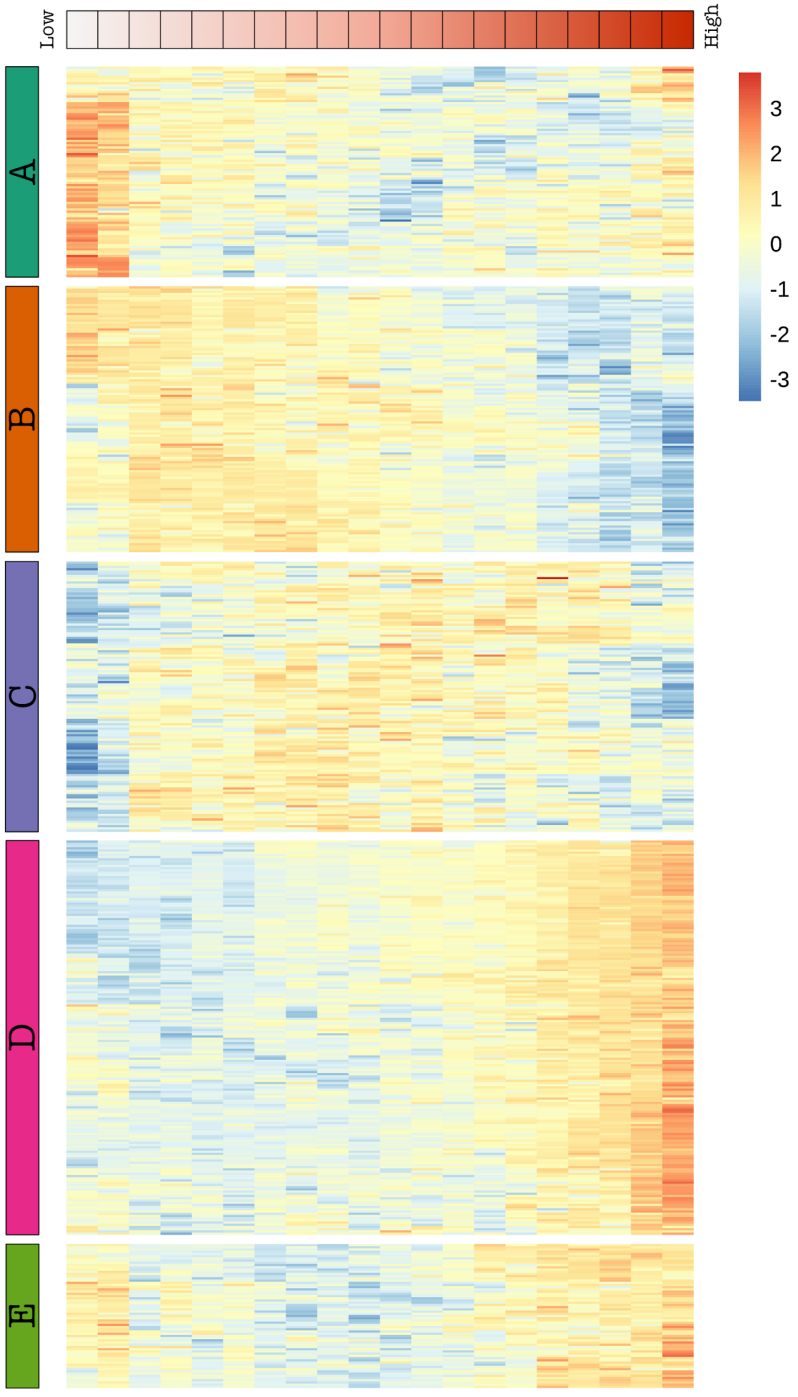

Figure 25. Heatmap showing the dynamics of host response to RSV infection. The heatmap is ordered based on viral load from low to high (annotation bar). Six waves of transcription are labeled with a capital letter (A-F). The heatmap shows the scaled expression values in each pseudo-bulk sample.

Gene regulatory networks over the course of RSV infection

We then tried to identify transcriptional regulators which modulate gene expression in ciliated cells and underlie their temporal heterogeneity. Single-cell regulatory network inference and clustering (SCENIC) analysis (Aibar et al. 2017) (also see methods) was used to find potential regulators of host response to RSV infection. SCENIC infers activity of gene regulatory networks based on co-expression of transcription factors and their putative target genes and also looks for enrichment of corresponding DNA motifs in proximity of genes inside a regulatory network. SCENIC identified several gene regulatory networks.

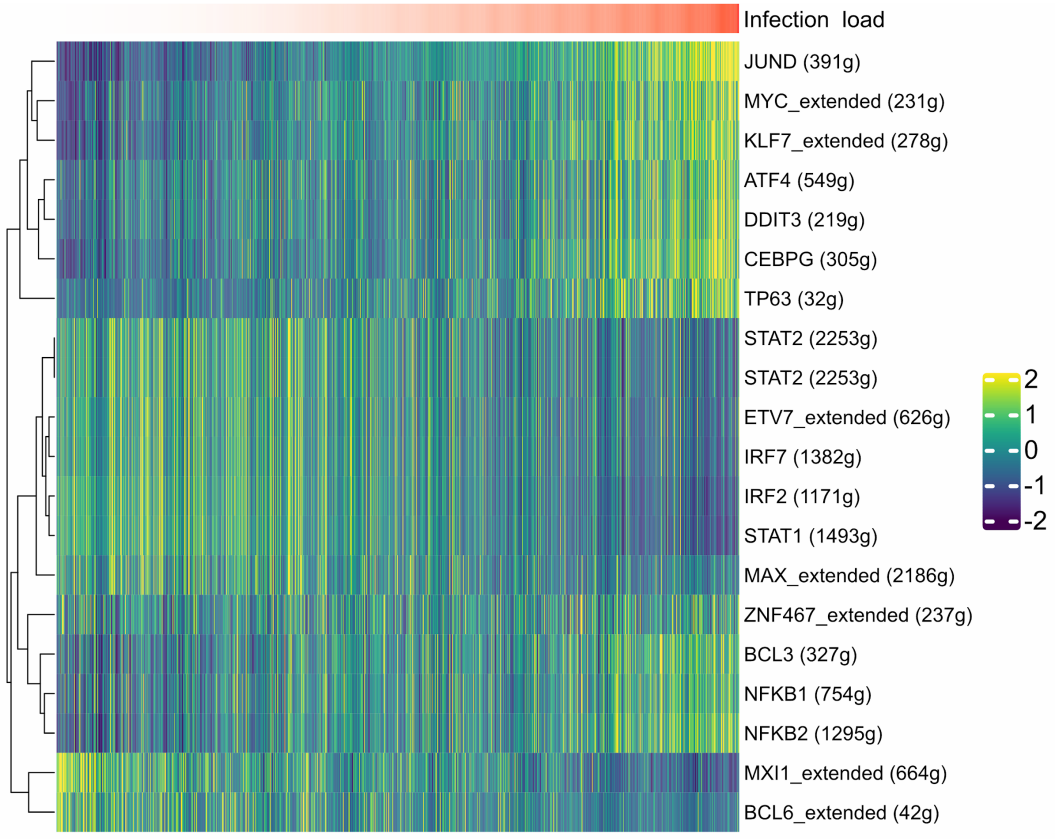

Figure 26. Gene regulatory networks detected in ciliated cells. The cells are ordered based on infection load (annotation bar). The heatmap shows the scaled AUC values calculated by SCENIC.

As expected, transcription factors controlling the innate immune response and interferon response were dominated at low infection loads. For instance, transcription factors like *STAT1*, *STAT2*, *IRF2* and *IRF7* control the early response of infected ciliated cells to infection. On the other hand, transcription factors such as *MYC*, *NFKB1* and *NFKB2* which control cell stress response and apoptosis were dominant at high viral loads (figure 26).

Discussion

Using single-cell RNA-sequencing to unbiasedly investigate the heterogeneity and dynamics of RSV infection, we profiled the gene expression signature of primary human airway epithelial cells and uncovered all the major cell types that are targeted by RSV. The interaction between the host and virus was investigated and we managed to reconstruct the dynamics of host gene expression profile in response to RSV infection. We further established a new framework to study infectious diseases at the single-cell level where infection load can be an indicative of infection progression and the host response can be analysed in the context of infection load.

Our results showed that ciliated cells were the main target of RSV infection and the infection load in this population of cells was prominently higher than any other cell type. This is inline with previous studies which showed RSV replicates almost exclusively in the highly differentiated ciliated cells in human airway (L. Zhang et al. 2002). We also observed the replication of RSV in club and goblet cells and in an unassigned population of cells although the infection load in these cell types was significantly lower compared to ciliated cells.

Overall, three subtypes of ciliated cells were present in our data set named ciliated cell A, B and C. Whereas all the ciliated cells were targeted by RSV, the three populations seemed to show different responses to the infection. Whereas ciliated cell A was the main target of RSV infection, ciliated cell B seemed to be involved in the

expression of interferons and the expression of interferon genes was specific to this population. On the other hand, ciliated cell C showed a strong innate immune response and in the global data, they clustered with bystander cells which suggests that their response to infection is very much similar to the bystander cells and upregulation of interferon stimulated genes and genes related to host innate immune response was particularly strong in ciliated cell C. Moreover the infection load in ciliated cell C was significantly lower than other ciliated cell subtypes which suggests that probably this population is more successful in taking RSV infection under control.

Differential expression analysis between mock, bystander and infected ciliated cells revealed several interesting genes and pathways that were induced upon RSV infection. As expected, in bystander cells, interferon stimulated genes (ISGs) like *IFIT1*, *IFIT2*, *IFIT3*, *MX1* and *MX2*, *ISG15* and *ISG20*, *IFI6* and *IFI27* were among the top differentially expressed genes. The expression of ISGs was also relatively high in infected cells (although it was lower than bystanders). On the other hand, genes like *TNFRSF12A* and *SFN* that are implicated in apoptosis and *IER3* that corresponds to cell stress response were particularly upregulated in infected ciliated cells. Strikingly, we observed a set of genes that their expression was specific to the infected cells and they were neither induced in mock nor in bystander cells. We argued that these genes are part of the programs which are induced by the virus to proceed the infection process and they can potentially be targets for therapeutics (for a full list of the infected specific genes please check Appendix 2). A gene set enrichment analysis was performed on these genes and pathways like regulation of calcineurin-NFAT signaling, ERBB signaling pathway and phosphatidylinositol 3-kinase signaling were among the enriched pathways. For instance calcineurin-NFAT signaling has already been implicated in the context of different diseases (Park et al. 2020) and several inhibitors of this pathway are already accessible (Martínez-Martínez and Redondo 2004) which based on this study they might also be potentially useful to control RSV infection. Further studies are required to show if inhibition of these pathways can be an approach to control RSV infection.

Different cell surface proteins have been suggested to be involved in RSV entry into the host cell, including annexin II (Malhotra et al. 2003), CX3 chemokine receptor 1 (CX3CR1) (Battles and McLellan 2019), epidermal growth factor (EGF) receptor (Currier et al. 2016), Toll-like receptor 4 (TLR4) (Marchant et al. 2010), intercellular adhesion molecule 1 (ICAM-1) (Behera et al. 2001), nucleolin (Tayyari et al. 2011), and heparan sulfate proteoglycans (HSPGs) (Krusat and Streckert 1997). We examined the expression of the corresponding gene related to each one of the above mentioned receptors at the RNA level. Results showed that among these genes, only *ANXA2* showed a positive correlation with viral load and none of the other receptors were correlated with RSV infection at the mRNA level. This suggests that ANXA2 might be an important cell surface receptor which mediates the entry of RSV into the host cells. Further experiments can indicate the potential role of this cell receptor and whether it can be a candidate target protein for development of therapeutics which inhibit RSV entry into the host cell.

Host response to infection might be affected by several factors including genetic diversity, age, sex etc. (Jones, Le Sage, and Lakdawala 2020) and based on these factors individuals might respond to viral infection differently. In the case of RSV, variations in genes like *SFPA/D*, *VDR*, *IL8*, *IL4*, *IL13*, *IL10* and *IL4RA* have been associated with different host responses to infection (Kenney et al. 2017). To assess whether we have an individual specific response to RSV infection, the cells were assigned to different individuals based on their genotype (Heaton et al. 2020). We couldn't identify any individual specific clusters in our data although the number of participants in the experiment was low and a higher number of participants are required to detect this kind of individual specific responses.

Next we tried to understand the dynamics of host response to RSV infection via analysis of the correlation of host genes with viral load. We selected a set of genes with strong correlation (positive or negative) and examined their expression over the course of infection. Interestingly, typical interferon stimulated genes like *MX1*, *MX2*,

interferon induced *IFIT* family genes and *STAT1* were among the genes which their expression decreased as a function of infection load. This is inline with previous findindings which showed after RSV recognition in the host cell, a wide variety of transcription factors including STAT, are activated which subsequently leads to activation of host antiviral programs (Rossi and Colin 2017). On the other hand, RSV has developed mechanisms to modulate the host innate immune response. NS1, NS2 and G proteins are the proteins used by RSV to control or to inhibit host immune mechanisms (Troy and Bosco 2016). Therefore, the expression interferon stimulated genes decreases as RSV infection progresses. On the other hand, genes like *HSPA8* and *HSP90AA1* which are chaperons required for protein folding and many ribosomal proteins such as *RPL3*, *RPS6*, *RPL9* which might be implicated in viral mRNA translation showed a positive correlation with viral load. This implies the ability of RSV to hijack transcription and translation machinery of the host for production of viral components (Battles and McLellan 2019).

We reconstructed the dynamics of host response to RVS infection via pseudo-bulk analysis. Pseudo-bulk analysis revealed five transcriptional waves (wave A-E) sequentially activated over the course of RSV infection. The details of the genes in each one of these transcriptional waves can be found in Appendix 2. Overall this analysis reveals the dynamics of host response to RSV infection at an unprecedented level.

Analysis of gene regulatory networks dynamics in ciliated cells supported our previous findings that interferon stimulated programs are activated at the beginning of infection and their activity gradually decreases as infection proceeds. For instance, gene programs controlled by transcription factors like STAT1, STAT2, IRF2 and IRF7 were active at the beginning of infection and their activity decreased as a function of infection progression. On the other hand, transcription factors such as MYC, NFKB1 and NFKB2 which control cell stress response and apoptosis were dominant at high viral loads.

Chapter 4: Single-bacterium RNA-sequencing

Experiment design

As discussed in the introduction, the sequencing of bacterial transcriptome at the single-cell level is hindered by several difficulties such as low copy number of bacterial transcripts, lack of poly-A tail and low RNA content of a bacterial cell compared to eukaryotic cells. To overcome these barriers, we used a highly sensitive protocol with low detection limit called multiple annealing and dC-tailing-based quantitative scRNA-seq (MATQ-seq) amenable to capture non polyadenylated transcripts (Sheng et al. 2017). MATQ-seq was reported to have very low detection limits and we reasoned that it would be able to perform single-bacterium RNA-sequencing. In our experiment, we analysed *Salmonella enterica* serovar Typhimurium (hereafter, *Salmonella*) which is a well studied model pathogen with a wealth of resources on the transcriptional response to different environmental stimuli (Kröger et al. 2013). To examine if our single-cell RNA-sequencing method is capable to correctly capture the transcriptional response of a single bacterium to different stimuli, *Salmonella* was grown in three different culture conditions:

1. Late stationary phase reflecting the transcriptional profile of resting cells.

2. Anaerobic shock which resembles the transcriptional status of bacteria in intestine
3. NaCl shock which shows the transcriptional response of bacteria to the increased concentration of NaCl in growth medium

Samples of either ten (10-pooled) or single bacteria were sorted using fluorescence-activated flow cytometry (FACS) and after cell lysis, reverse transcription and amplification, libraries were sequenced (figure 27). The libraries were sequenced to a depth of 62.4±20.9 million reads per library. Whereas this depth of sequencing is relatively high for an RNA-sequencing experiment, it guarantees that after removal of reads aligned to rRNA and tRNA, sufficient reads are available for downstream data analysis. Overall, 60 samples of 10-pooled and 71 samples of single bacteria were sequenced in this experiment.

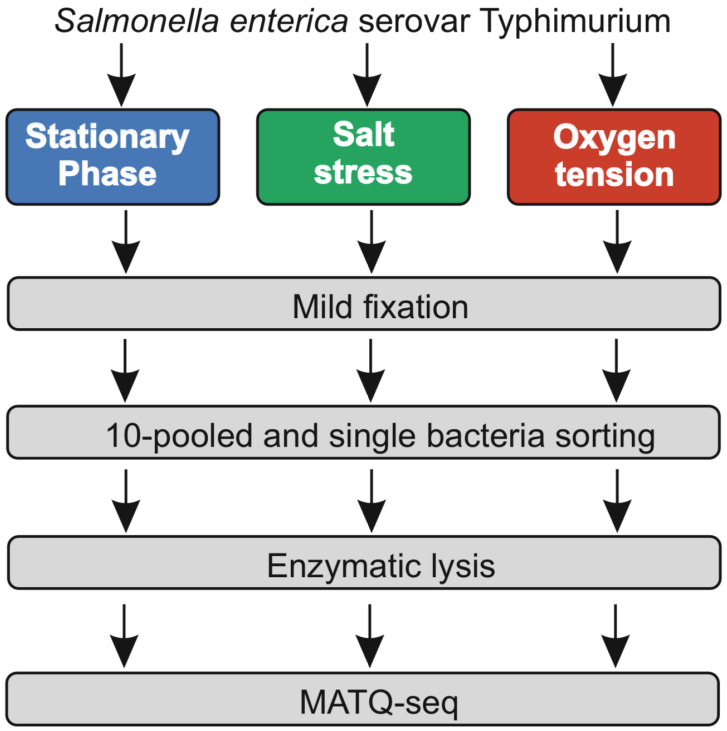

Figure 27. Schematic representation of experiment design

Quality assessment of sequencing data

To examine the quality of the sequenced libraries, a series of different quality control assessments were performed. In the following section, different aspects of quality of 10-pooled and single-bacterium RNA-sequencing data have been described and where relevant, it has been compared to bulk RNA-seq libraries.

Libraries were dominated by rRNA and tRNA

The number of reads aligned to different RNA species was counted in order to examine the percentage of reads associated with each RNA type. As expected for a random hexamer priming protocol, the cDNA libraries were dominated by rRNA and tRNA which on average accounted for ~93% of all mapped reads. On the other hand, mRNA and small non-coding RNA genes were represented by ~5 and ~1.2% of mapped reads and there was no significant difference between single and 10-pooled libraries (figure 28). We also compared the proportion of different RNA classes in our single and 10-pooled libraries to a previously published bulk RNA-seq library (Westermann et al. 2016) and noticed that the proportion of different RNA classes are fairly similar between the two experiments (figure 28).

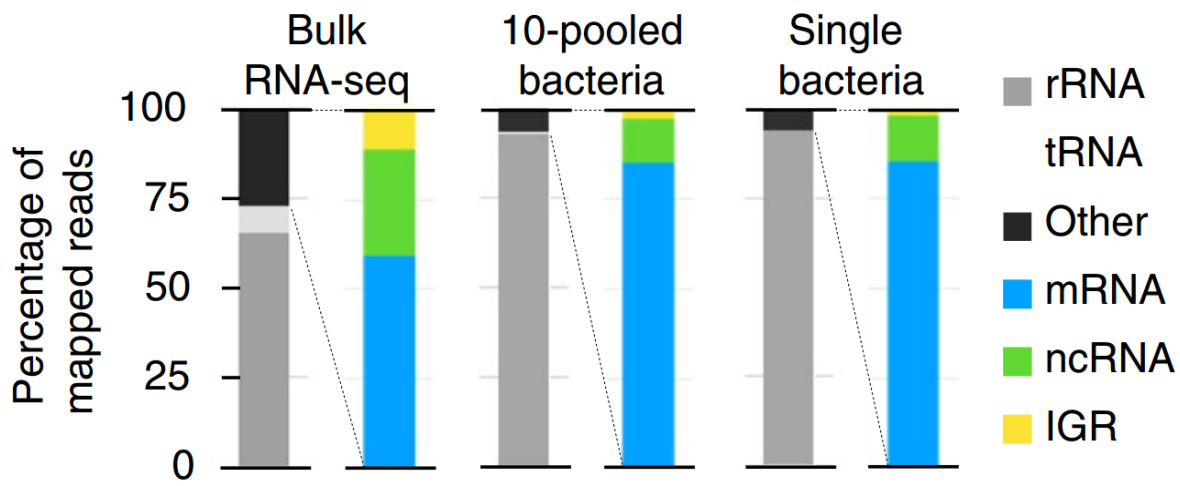

Figure 28. The percentage of different RNA species in single-cell and 10-pooled libraries along with the percentage of RNA species in bulk RNA-seq data (Westermann et al. 2016) as a benchmark.

Library size and number of genes detected per library

The library size of each one of the single-cell and 10-pooled samples were calculated by counting the sum of uniquely aligned reads in each library. We observed discrepancy regarding library size between different culture conditions where libraries cultured in late stationary phase showed lower library size compared to NaCl shock and Anaerobic shock (figure 29). Overall, on average, we had ~1,500,000 (for 10-pooled) and ~800,000 (for single bacteria) uniquely mapped reads per library. This number of uniquely aligned reads is commonly used in single-cell RNA-sequencing experiments to study the identity and functional state of a eukaryotic cell (Picelli et al. 2013) therefore is sufficient to study a bacterial cell which basically expresses less number of genes and has lower RNA content.

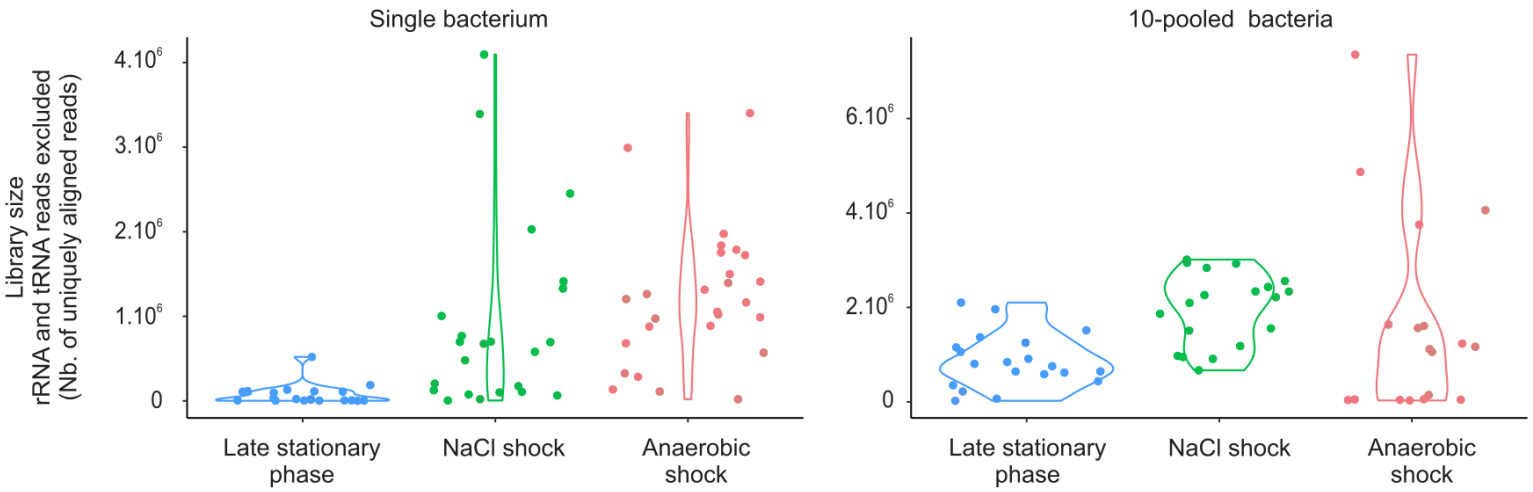

Figure 29. The library size of single-cell (left) and 10-pooled (right) libraries cultured under three different growth conditions.

Overall, our sequencing method on average detected 413±237 and 170±81 genes per library in 10-pooled and single-cell libraries respectively (genes with more than 10 unique reads aligned to them were considered as detected genes). The difference between the culture conditions was also observed regarding the number of detected genes and cells cultured under late stationary phase expressed lower number of genes and the number of genes detected in NaCl shock and anaerobic shock was relatively higher (figure 30).

The relation between the library size and the number of detected genes was also visualized in the form of scatter plot and in general the number of detected genes and the library size showed a positive correlation (figure 31).

Saturation analysis

To ensure that our libraries were sequenced to a sufficient sequencing depth, we performed saturation analysis. The count table related to each library was randomly down sampled and the number of detected genes for each round of down sampling was calculated. The analysis showed that the libraries were saturated approximately

at a sequencing depth of 200,000 reads per library and increasing the sequencing depth beyond this value only marginally increases the number of detected genes (figure 32).

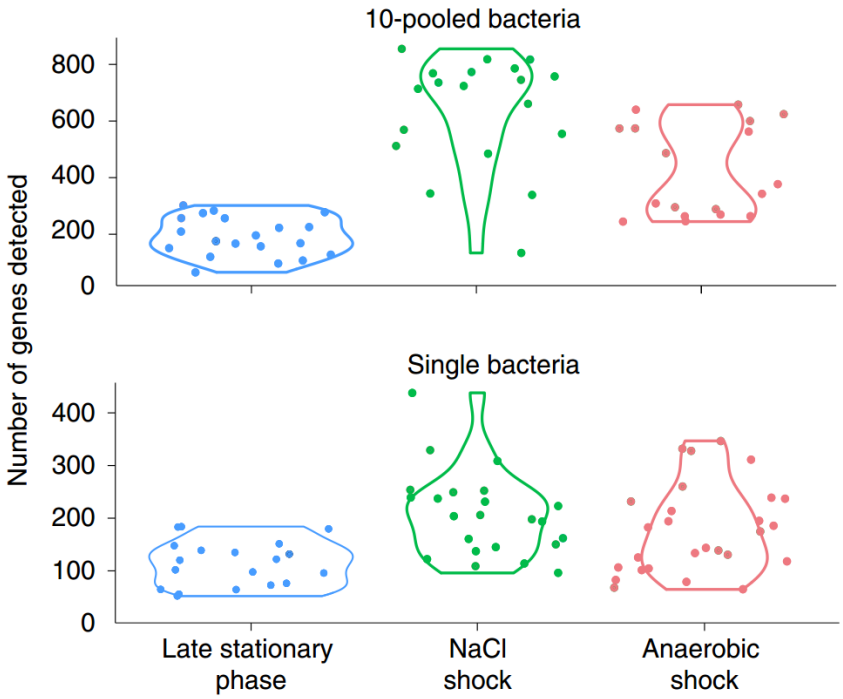

Figure 30. Number of detected genes in 10-pooled (top) and single-cell (bottom) libraries.

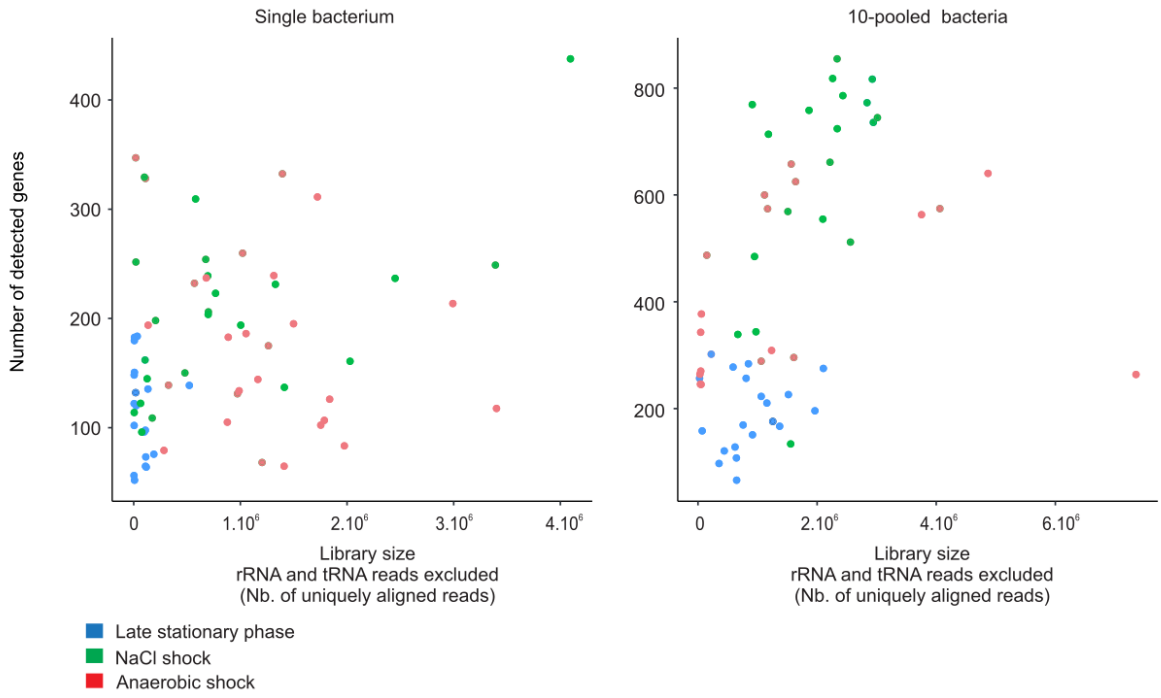

Figure 31. The relation between the number of detected genes and library size in single-cell (left) and 10-pooled (right) libraries

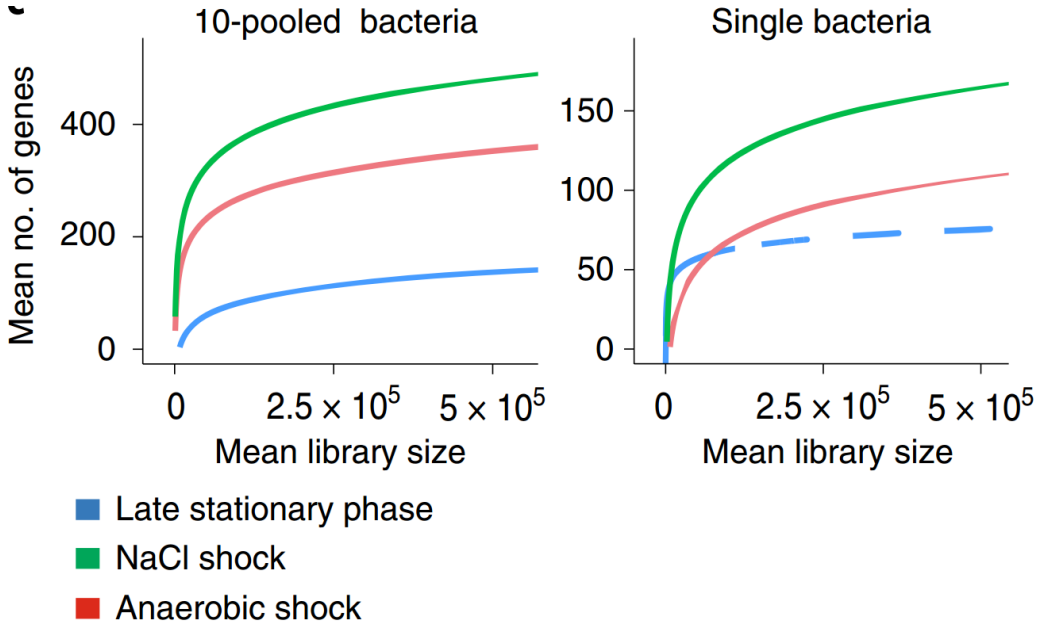

Figure 32. Saturation analysis of 10-pooled (left) and single-cell (right) libraries.

Reads coverage on the reference genome

To examine the distribution of reads on the reference genome, coverage plots for highly expressed genes and also genes with differential expression between the culture conditions were created. The read densities on the coverage plots showed that the uniquely aligned reads only covered the transcribed regions of the genes and were hardly detectable in intergenic regions. This shows that we sequenced cDNA derived from RNA and not genomic DNA. For instance, the coverage plots related to *ssrA*, *fliC*, 23S ribosomal RNA and *rrsH* 16SrRNA which were among the highly expressed genes and also *sodA*, *yadF*, *ompD* and *ansB* which were differentially expressed in 10-pooled bacteria are shown in figure 33.

Measuring variability in 10-pooled and single-cell data

To have an estimate of technical and biological variability in the data, the coefficient of variation (CV) of 10-pooled and single-cell libraries were calculated. In variability analysis, the assumption is that the majority of the genes show a basic level of technical variability and the technical variability can be modelled by fitting the variability among all the genes. On the other hand, we also assume that a small minority of the genes in the data show the interesting biological variability. Therefore, the biological variability can be perceived from the certain genes that their variability deviates from the fit. For each gene we plotted the coefficient of variation against the average expression level in each culture condition and obtained a profile similar to that for conventional single-cell eukaryotic transcriptomes. The resemblance between the profiles in our single-cell and 10-pooled libraries and eukaryotic cells profile is another layer of evidence that shows the data obtained from prokaryotic cells have acceptable quality. As expected, we observed that the technical variability is significantly lower for genes with high expression and on the other hand, genes with low expression show a relatively high technical variability (figure 34).

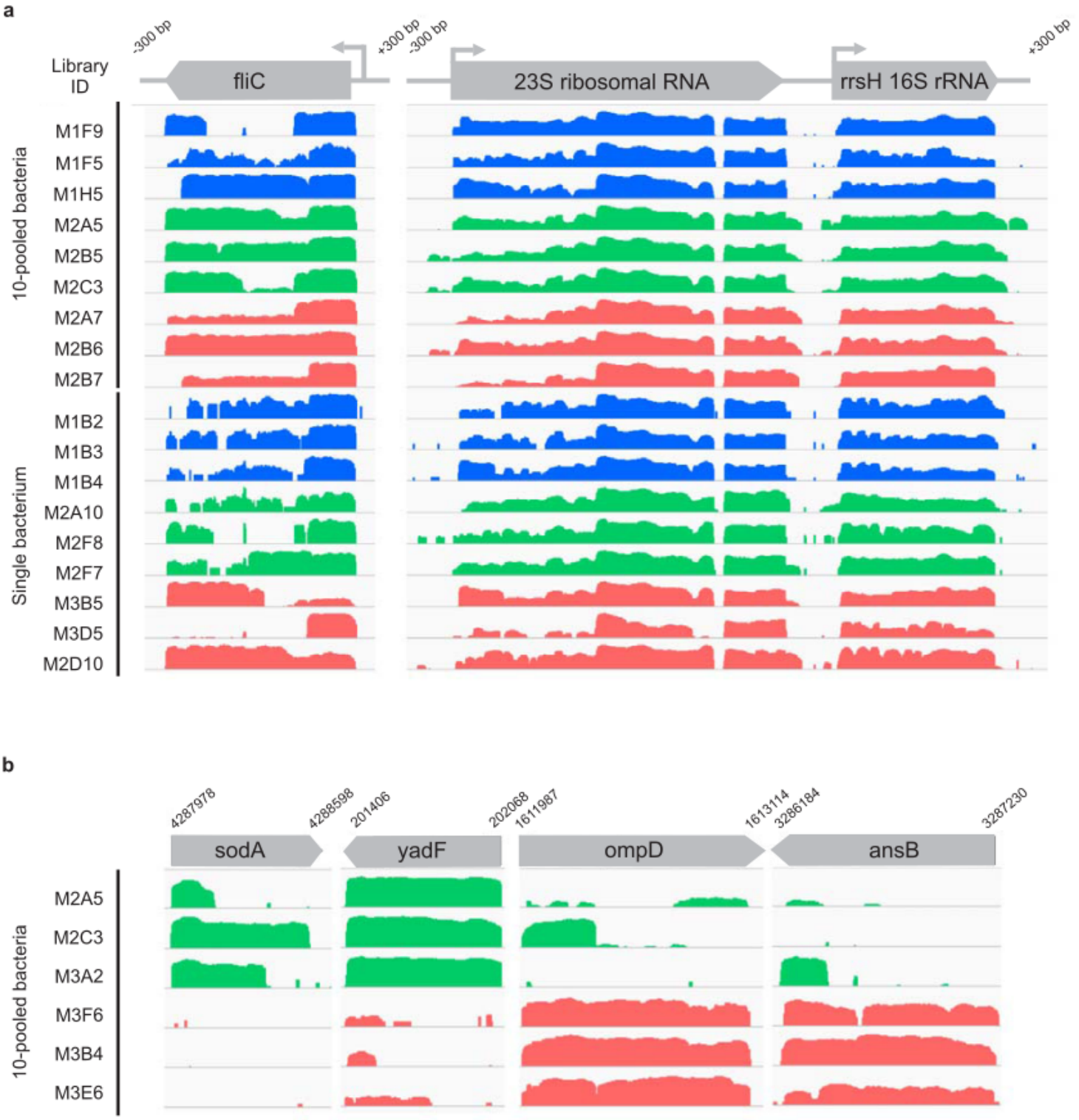

Figure 33. Coverage plots showing read densities on highly expressed genes (a) and differentially expressed genes (b). The library ID indicates the arbitrary names given to different libraries.

10-pooled and single-cell libraries are highly correlated

To measure the (dis) similarities between the 10-pooled and single-cell libraries, the correlation between the data from the same culture condition was calculated. Ten-pooled and single-cell libraries showed Spearman's correlation ρ = 0.5 in anaerobic condition and NaCl shock and ρ = 0.42 in late stationary phase. Also, a high-level of zero value could be observed in single-cell libraries compared to the 10-pooled libraries (figure 35).

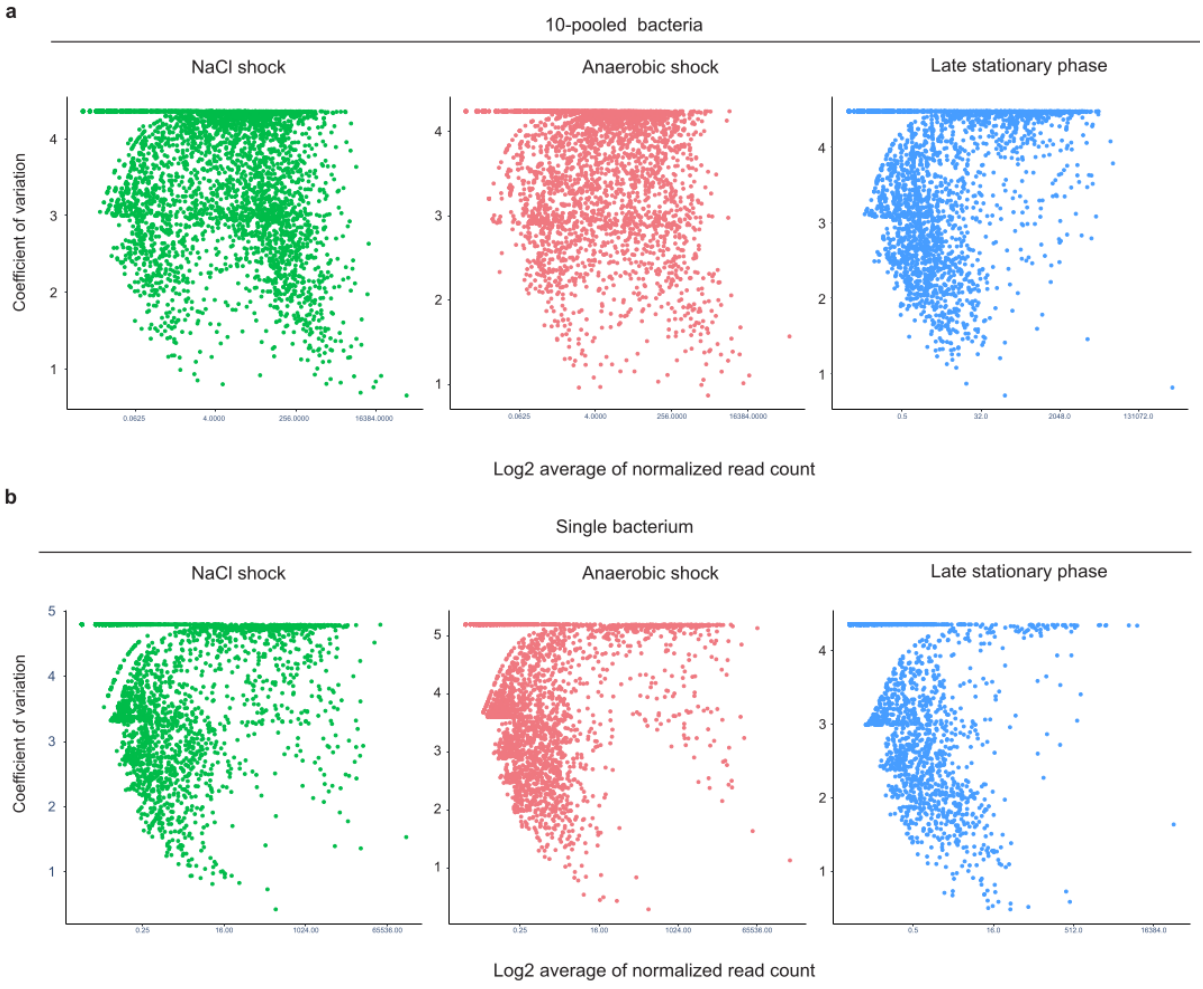

Figure 34. Coefficient of variation as a function of the level of gene expression in 10-pooled bacteria (a) and single bacterium (b)

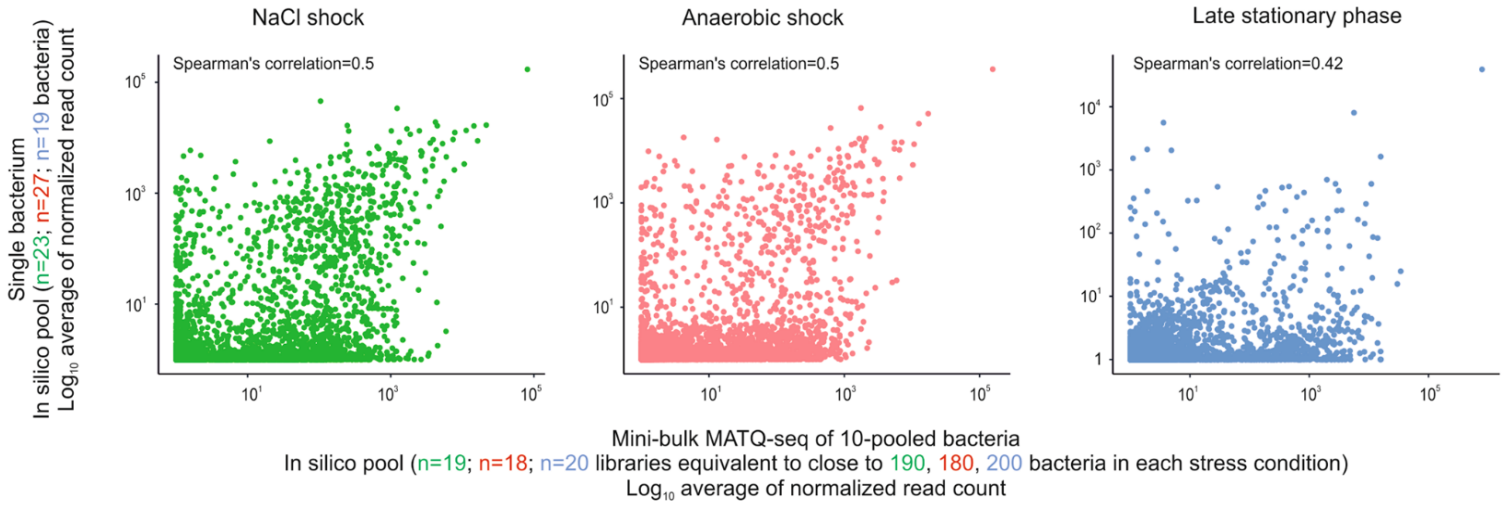

Figure 35. The correlation between the expression level of 10-pooled and single-cell libraries in three different culture conditions.

Principal component analysis (PCA)

To determine if the transcription profile of libraries faithfully reflects their culture condition, principal component analysis was performed on 10-pooled and single-cell libraries. Our PCA analysis showed that the three culture conditions were successfully delineated on the PCA plot and the first and the second principal components could successfully distinguish between libraries cultured under late stationary phase, anaerobic shock and NaCl shock in both 10-pooled and single-cell libraries (figure 36). Overall, in 10-pooled libraries, PC1 and PC2 explained 13.8% and 7.3% of variance respectively and in single-cell libraries, PC1 represented 12.9% and PC2 showed 8.5% of variance (figure 36).

To prove that the principal component analysis is not driven by the differences in the library size and the number of genes, we projected these variables on the PCA plot. Moreover, to show the main genes driving the PCA analysis, the top 15 genes with the highest contribution to principal component analysis were also selected and their related loading vectors showing how the original variables contribute to creating the principal component were shown on the PCA plot (figure 37).

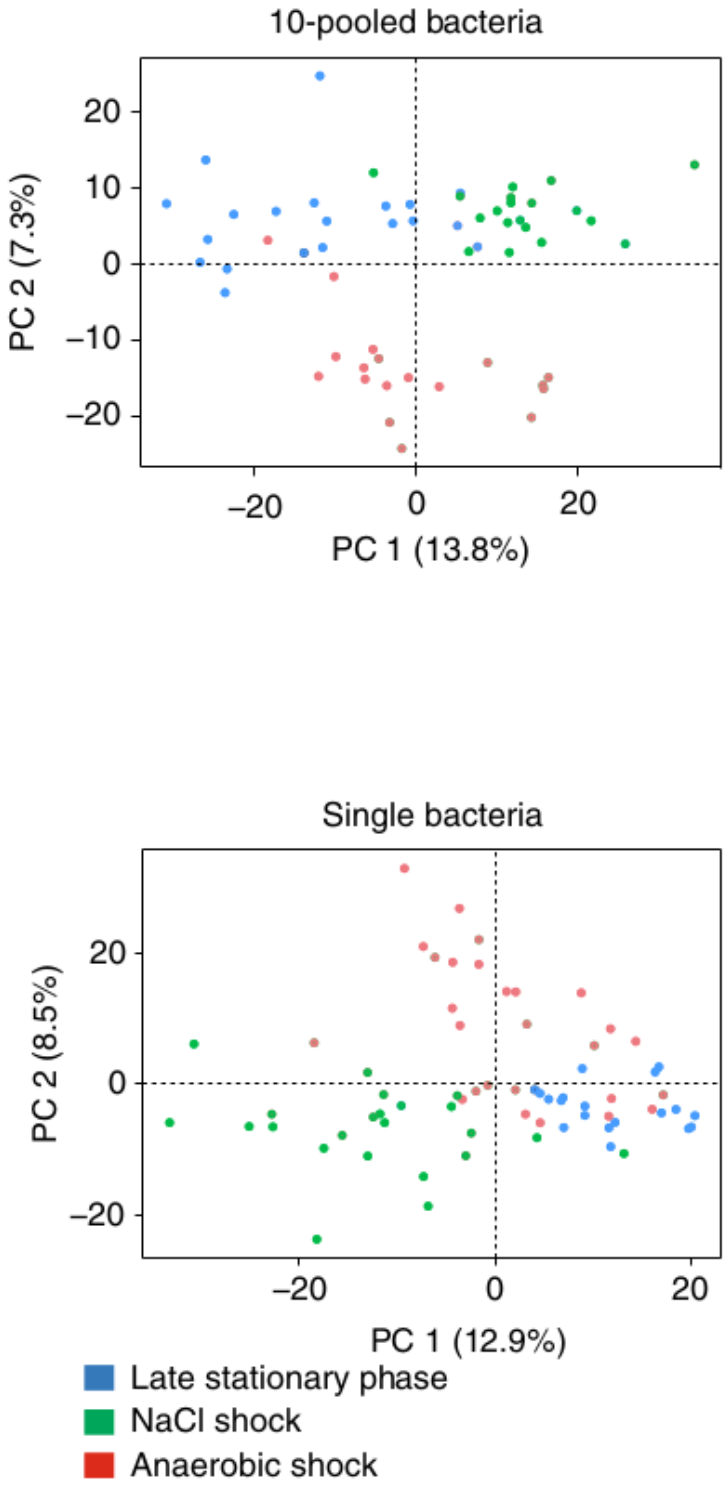

Figure 36. Principal component analysis faithfully delineates the libraries from different culture conditions in both 10-pooled (top) and single-bacteria (bottom) libraries.

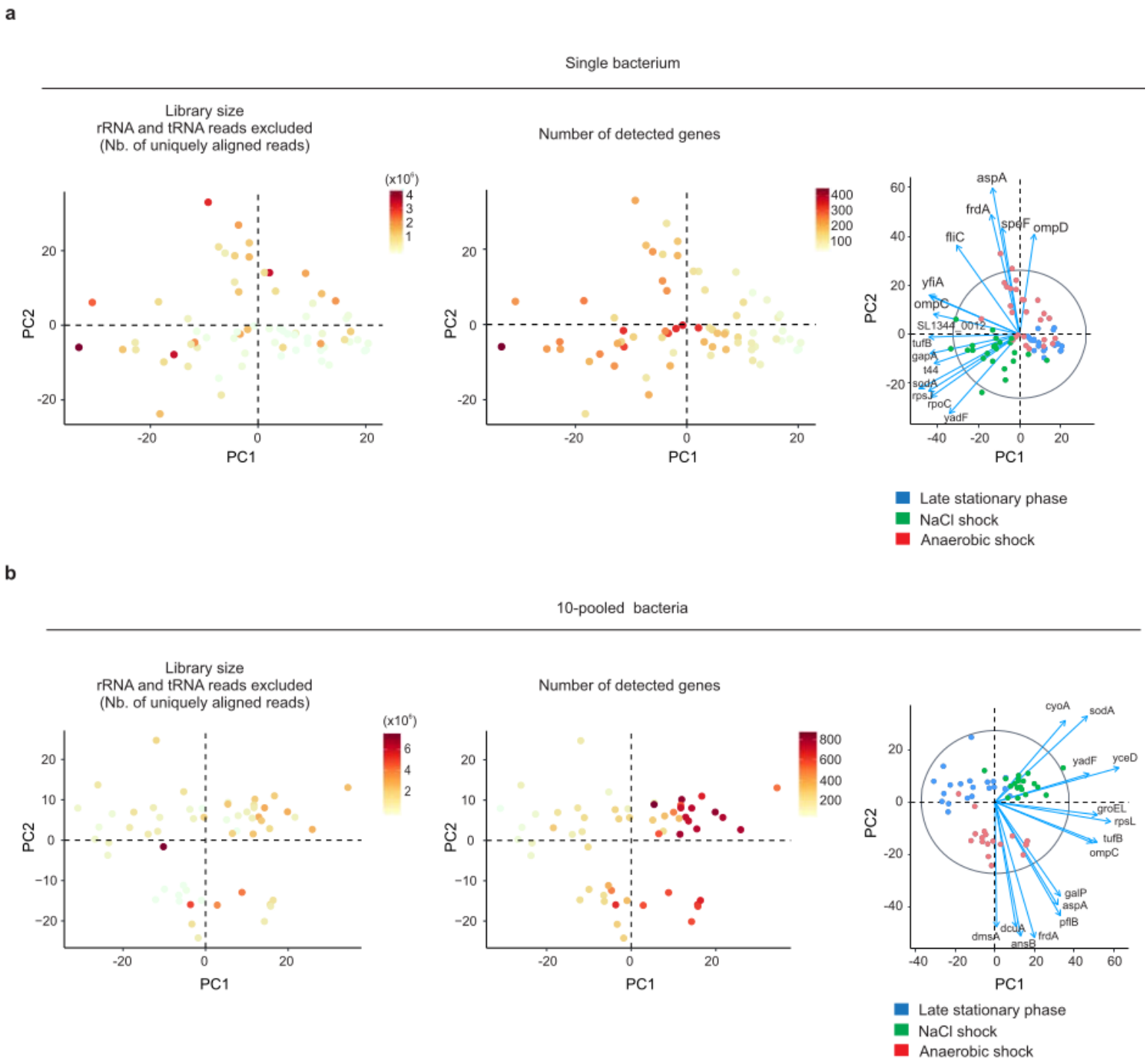

Figure 37. Library size and number of detected genes projected on PCA plot along with the top 15 genes with the highest contribution to principal components in single-cell libraries (a) and 10 pooled libraries (b).

Differential expression analysis (10-pooled bacteria)

Differential expression analysis between different culture conditions was performed to find out if our method is capable of finding genes that their expression changes under each treatment condition. First, we performed differential expression analysis in 10-pooled libraries using DESeq2 (Love, Huber, and Anders 2014). Overall, we identified 274 upregulated genes in NaCl shock and 101 upregulated genes in anaerobic shock. Differential expression analysis revealed important upregulated genes like *ompD*, *glpA*, *tdcC*, *ansB* and *ompD* in anaerobic condition and *rpsJ*, *groEL*, *sodA*, *ydgI*, *yadF*, *sopA*, *hilC*, *iacP* in NaCl shock treatment condition (figure 38). For the full list of differentially expressed genes, please see Appendix 3.

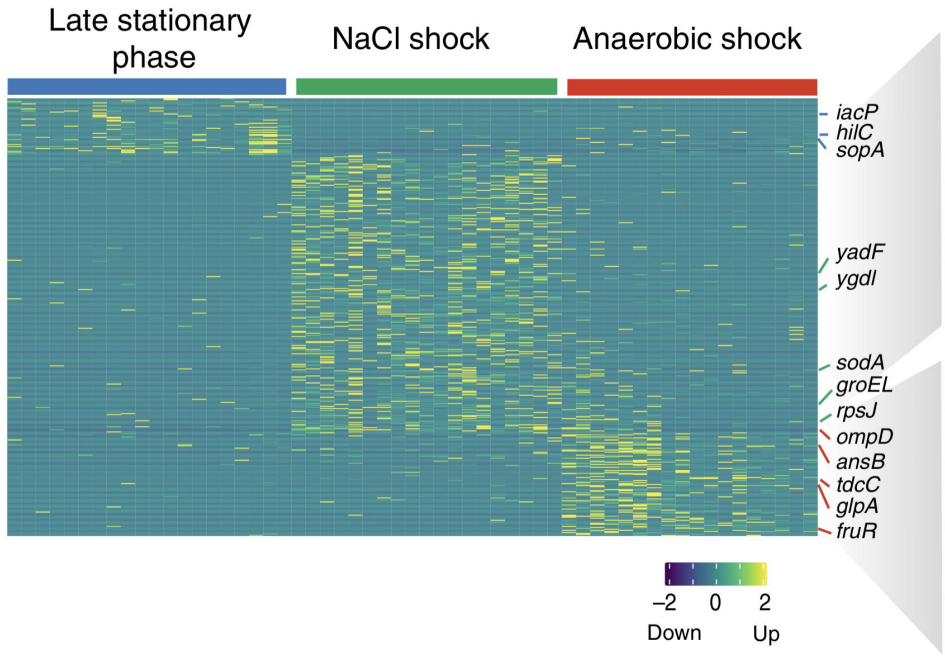

Figure 38. Heatmap showing the expression of differentially expressed genes in 10-pooled libraries in different culture conditions.

We benchmarked the list of differentially expressed genes based on a published bulk RNA-seq data (Kröger et al. 2013). For each gene found to be differentially expressed between NaCl and anaerobic shock, the log10-transformed ratio of expression values

measured in bulk RNA-seq study was calculated and is represented in histograms and boxplots (figure 39). The results of the benchmark showed that roughly 75% of the genes upregulated in our 10 pooled libraries were also up-regulated in bulk RNA-seq data.

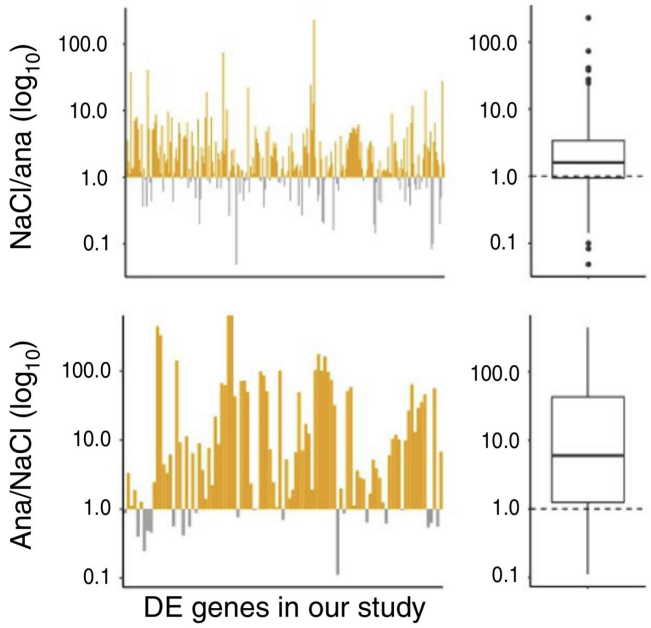

Figure 39. Benchmarking differentially expressed genes in 10-pooled libraries. The histogram shows the log transformed ratio of anaerobic and NaCl shock expression values in the benchmark study (colour coding represents whether the log-transformed ratio is >1 or <1). The boxplot shows the distribution of the log transformed ratios.

Differential expression analysis (single-cell bacteria)

Then we examined the expression of differentially expressed genes in the single-cell libraries. Single-cell libraries were also capable of identifying the differential expression of 63 and 131 genes in anaerobic and NaCl shock respectively. In the single-cell data again a set of biologically relevant genes for each culture condition were identified. In anaerobic shock genes like *ompD*, *tdcC*, *uvrD*, *glpA* and *frdA* were upregulated whereas in NaCl shock, genes like *groEL*, *rpsJ*, *sodA*, *yadF*, *ygdL* and *sopA* were upregulated (figure 40) (for a full list of genes, please see Appendix 4).

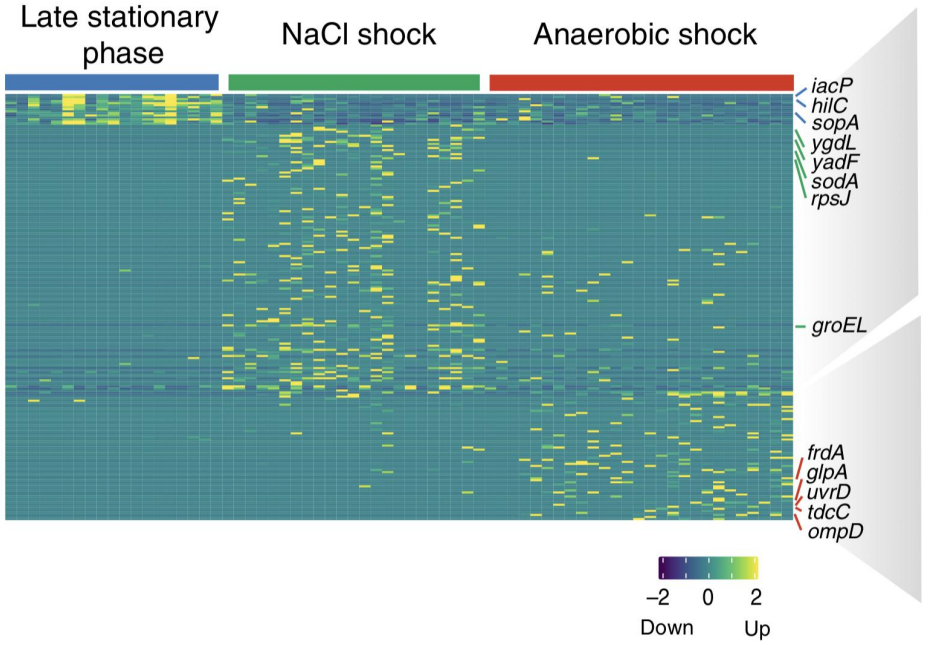

Figure 40. Heatmap showing the expression of differentially expressed genes in single-cell libraries in different culture conditions.

We benchmarked the list of differentially expressed genes in single-cell libraries based on a published bulk RNA-seq data (Kröger et al. 2013). The same approach with 10 pooled libraries was also used here. For each gene found to be differentially expressed between NaCl and anaerobic shock, the log10-transformed ratio of expression values measured in bulk RNA-seq study was calculated and was represented in histograms and boxplots (figure 41). The results of the benchmark showed that roughly 75% of the genes upregulated in our 10 pooled libraries were also up-regulated in bulk RNA-seq data.

Overall, the fact that a big fraction of upregulated genes in either 10-pooled or single-cell libraries were also upregulated in bulk RNA-seq experiments is an indicative of the robustness of our approach and shows it is capable of recapitulating biologically relevant features.

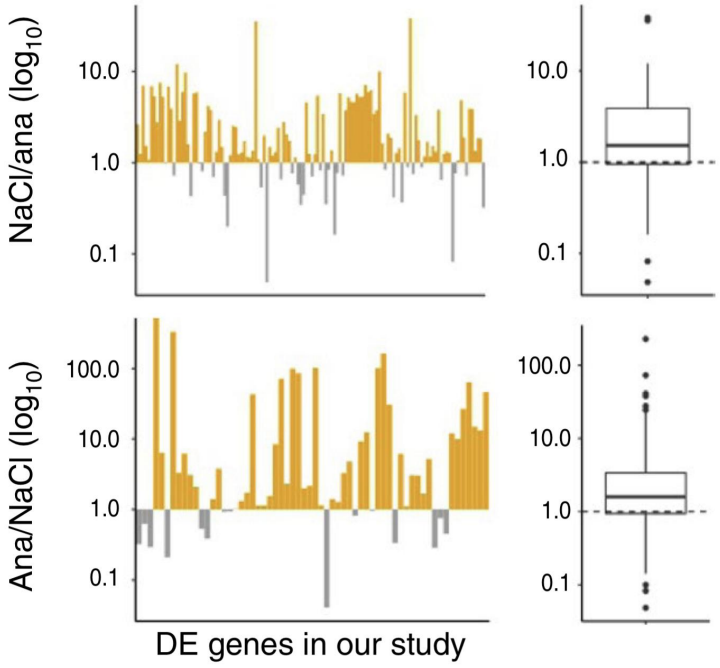

Figure 41. Benchmarking differentially expressed genes in single-cell libraries. The histogram shows the log transformed ratio of anaerobic and NaCl shock expression values in the benchmark study (colour coding represents whether the log-transformed ratio is >1 or <1). The boxplot shows the distribution of the log transformed ratios.

Pseudomonas aeruginosa

To show that our method can be generalized to other bacterial species, libraries of the important pathogen *Pseudomonas aeruginosa*, whose genome has a much higher GC content than that of *Salmonella* (>67% versus ~50%) was prepared and the same data analysis procedure was conducted on the libraries. We prepared 10-pooled and single-cell libraries and examined the library size and the number of genes expressed in each library. In *Pseudomonas* we detected on average 284 genes in 10-pooled libraries and 102 genes at the single cells level which is comparable to the number of genes detected in *salmonella* libraries. As expected the *Pseudomonas* data was also dominated by rRNA and tRNA. Around 97% of the reads were aligned to these RNA species. Moreover, all other RNA species were also detected in this data. We also examined the distribution of the reads on the reference genome and noticed that

most of the uniquely aligned reads were aligned to the genes whereas the reads aligned to intergenic regions were relatively rare (figure 42).

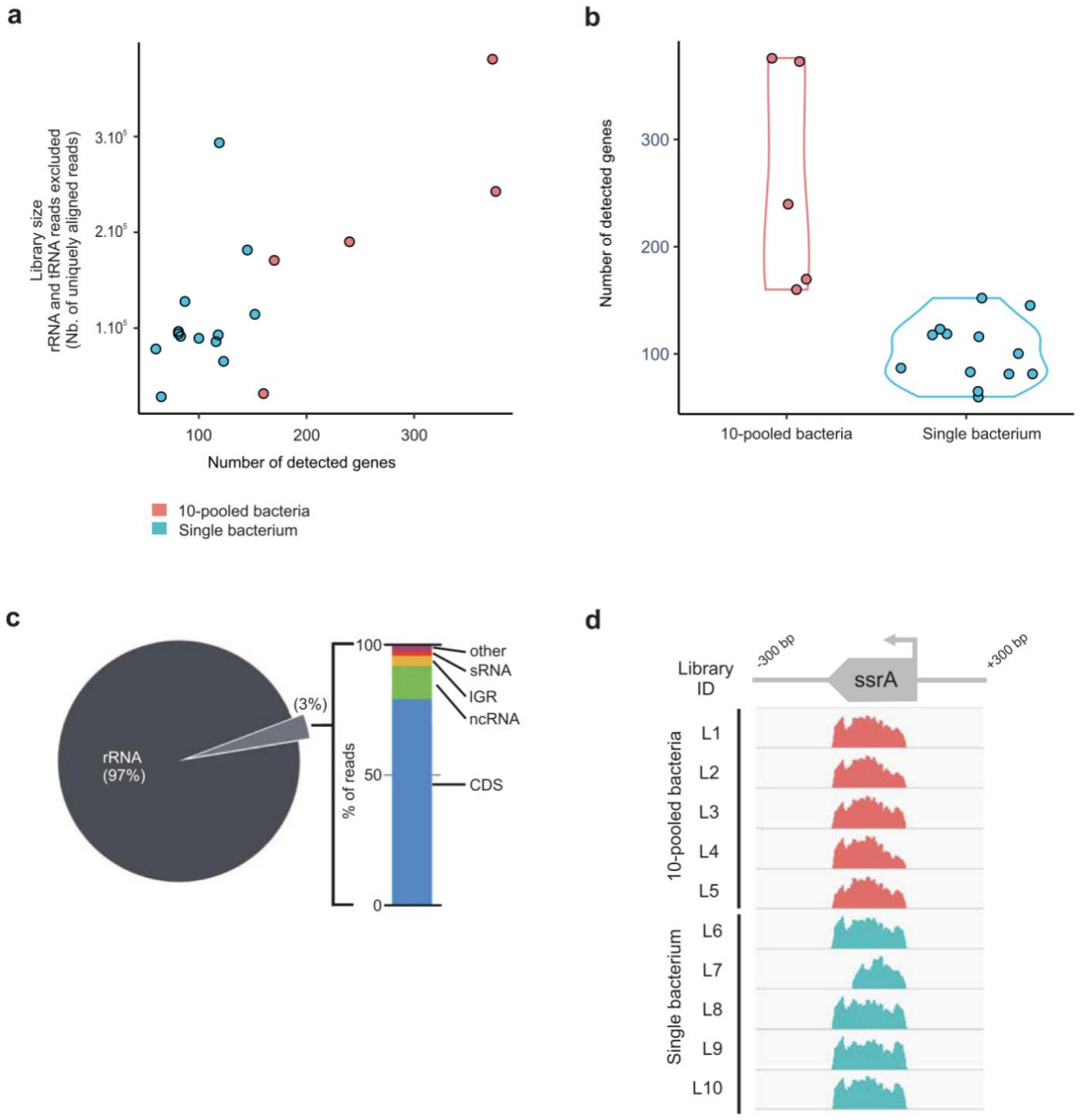

Figure 42. MATQ-seq captures the transcriptome of Pseudomonas aeruginosa. a) the relation between the library size and number of detected genes in Pseudomonas data. b) violin plot showing the number of detected genes in 10-pooled and single-bacterium libraries. c) the percentage of reads aligned to different RNA species. d) coverage plot showing the reads densities of the uniquely aligned reads mapped to ssrA gene.

Overall, In this research, we managed to adopt the MATQ-seq method to sequence low abundant transcripts down to single bacteria. More importantly, we managed to

infer the corresponding growth condition according to the transcriptional profile. This method is capable of detection of hundreds of genes per single bacteria which allows to study the physiological state of single cell bacteria. This work can potentially pave the way for the study of more complicated microbial communities like microbiome and sophisticated microbial interactions like host-pathogen interactions. In principal developing methods to deplete ribosomal RNA can considerably decrease the cost of running such experiments therefore making the analysis of single bacteria available for further studies.

Chapter 5: Heterogeneity of aortic macrophages in murine atherosclerosis

To study the diversity and transcriptional landscape of aortic macrophages, we sorted total viable $CD45^+$ leukocytes from the aortas of *Ldlr*$^{-/-}$ mice (LDL receptor deficient mice). To obtain $CD45^+$ cells, we conducted two separate experiments. In the first experiment, two groups of mice were treated with either a high fat diet (HFD) or a chow diet for 11 weeks. Treatment with chow diet was considered as a control condition in this experiment and HFD treatment was representative of the intermediate stages of aortic plaque development. In this experiment, data from two treatment conditions (chow diet and HFD) were aligned and analysed together to compare the diversity of macrophages in healthy and athrosclerotic aorta. In the second experiment, a group of mice were treated with HFD for 20 weeks and were considered as representative of advanced stages of atherosclerosis and the data obtained from this experiment was analysed separately to understand the diversity of macrophages at more advanced stages of atherosclerosis (figure 43). Overall, in this experiment design, not only the heterogeneity of $CD45^+$ cells can be studied in healthy aorta versus intermediate and advanced stages of atherosclerosis but the dynamics of cells in different stages of atherosclerosis can also be examined.

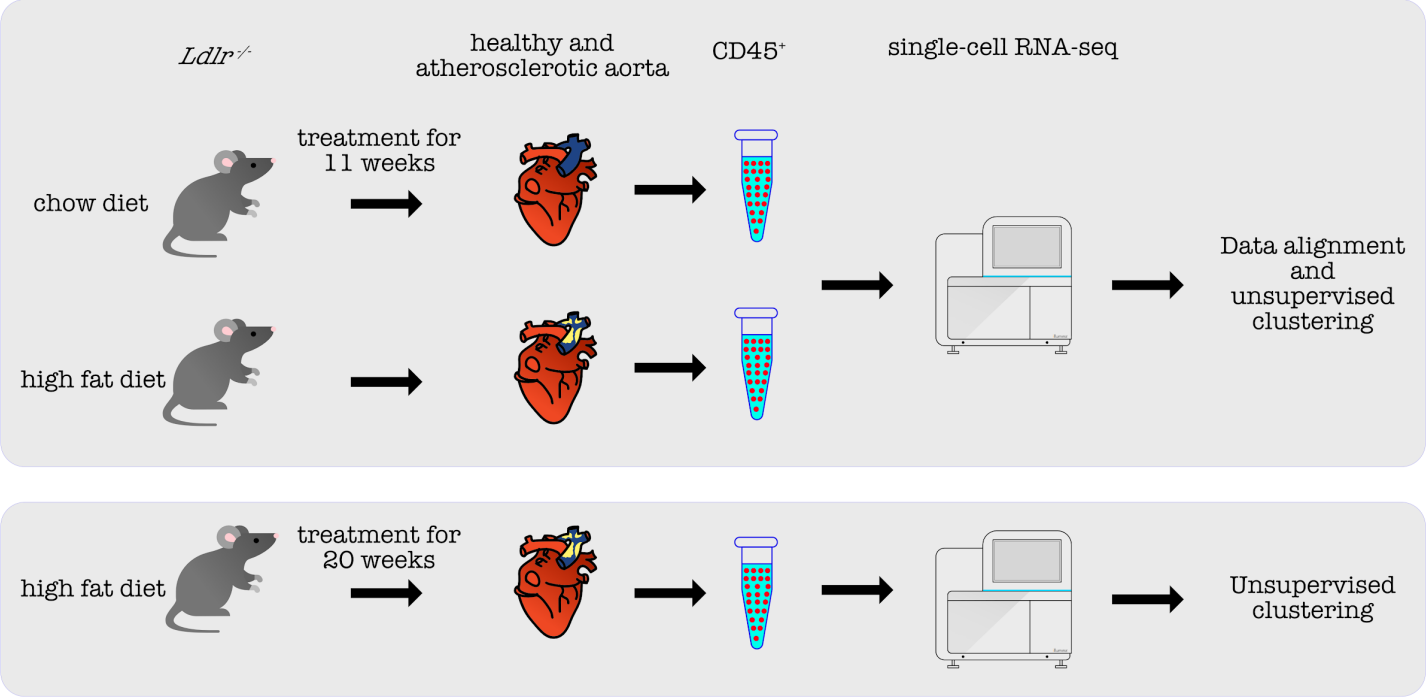

Figure 43. Schematic diagram of experimental design.

Heterogeneity of immune cells in healthy and diseased aorta

First we tried to compare the diversity of immune cells in healthy and atherosclerotic aorta after 11 weeks of treatment. After sequencing, the data from two treatment conditions were integrated. The quality of the data was examined based on the common quality control metrics and after removal of low quality cells, 372 cells in the control condition and 854 cells in diseased aorta were included for further data analysis. The gene expression of the integrated data was subjected to dimension reduction and clustering and was projected on two dimensional space. Based on the alignment of the two data sets and expression of the canonical markers, atherosclerosis-associated immune cells were identified. Overall, we identified 13 transcriptionally distinct clusters in the data which some of them were only atherosclerosis specific (figure 44).

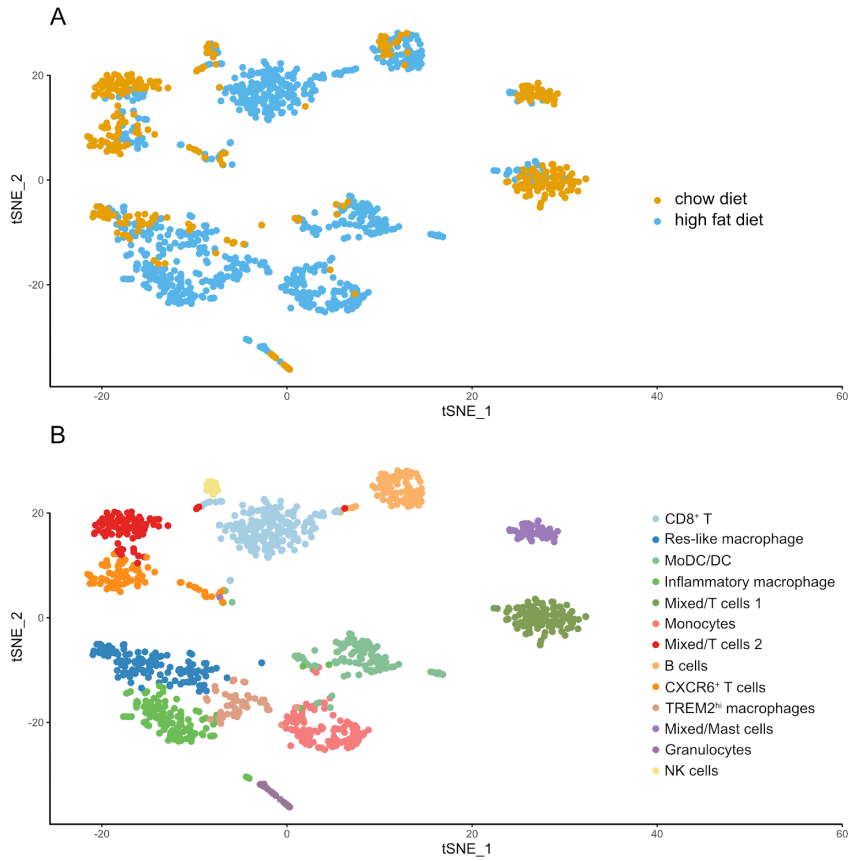

Figure 44. t-SNE visualization of aggregated data after 11 weeks of chow diet (n = 372) or high fat diet (n=854) treatment A. cells color represents the treatment condition. B. cells colored according to the 13 distinct clusters which were identified based on the expression of canonical markers.

Non macrophage populations were present in both healthy and diseased libraries. These populations were composed of B cells (*Cd79a*, *Cd79b*, *Ly6D* and *Mzb1*), CXCR6^{+} T cells (*Cxcr6*, *Icos*, *Cd3g* and *Il7r*), mixed cells / mast cells (*Calca*, *Furin* and *Il1rl1*), granulocytes (*S100A8*, *S100A9*, *Ngp* and *Camp*), natural killer cells (*Klrb1c*, *Ncr1*, *Klra8* and *Klrc1*) and 2 clusters with a T cell phenotype.

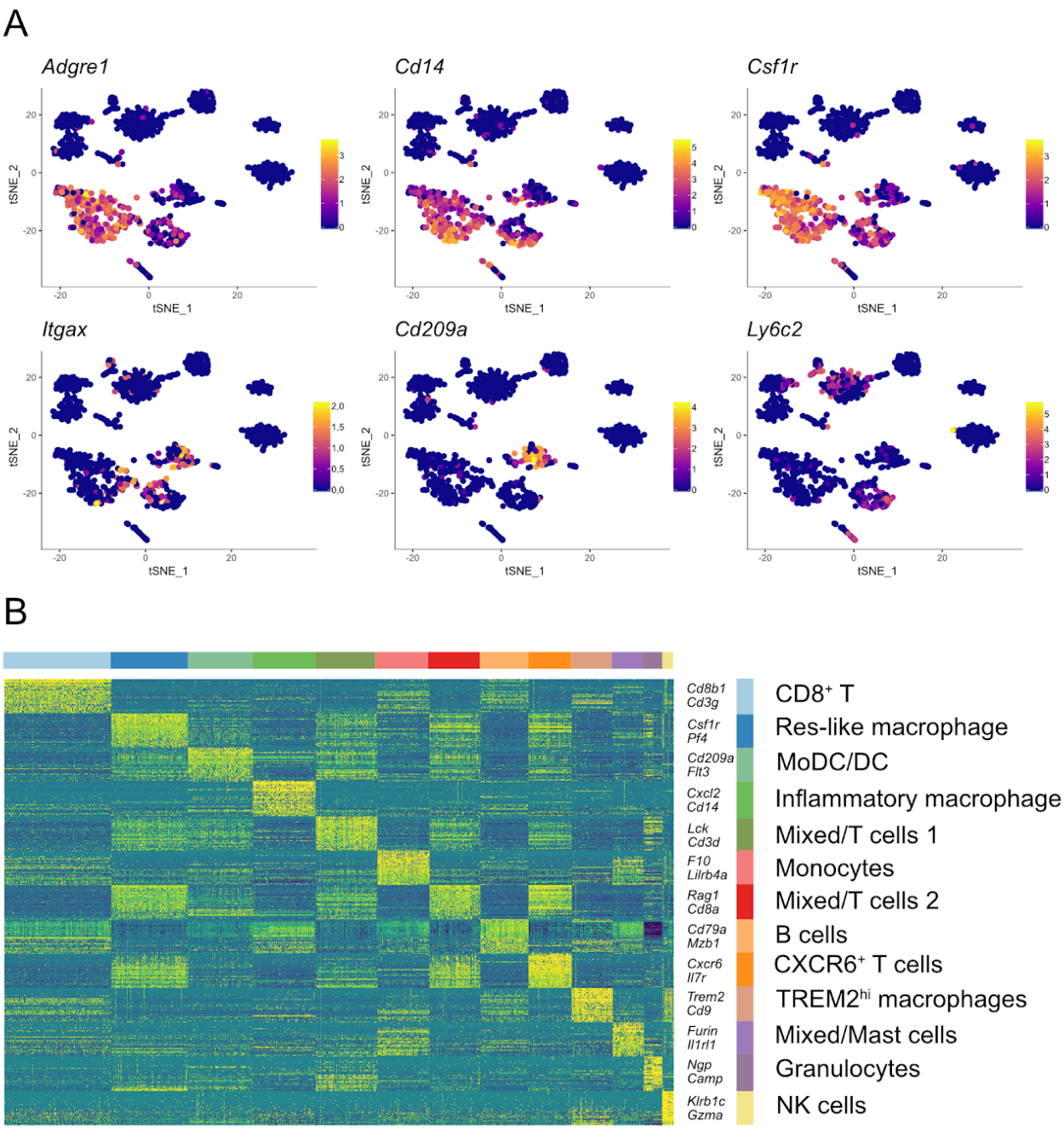

Figure 45. Expression pattern of canonical marker genes. A. expression of Adgre1, Cd14, Csf1r, Itgax, Cd209a, Ly6c2 projected on two dimensional space. B. Heatmap showing the top 20 differentially expressed genes in each cluster. Canonical marker gene in each cluster is highlighted on the right.

In addition, one single cluster of macrophages (*Adgre1*, *Csf1r*, *Fcgr1* and *Cd68*) was identified that was composed of both healthy and diseased libraries. This population

showed the gene signature of resident macrophages (*F13a1*, *Lyve1* and *Gas6*). We tried to find the difference in gene expression signature of the cells from healthy and diseased libraries in this particular population. The results of differential expression analysis pointed out that cells from healthy aorta had higher expression of *Lyve1* whereas cells from diseased aorta showed higher expression of *Ccr2*. Overall, the results of differential expression analysis suggests that atherosclerotic aortas contain both resident macrophages and recruited cells that adopt a similar gene signature when infiltrated in the aorta. Therefore, this macrophage population is hereafter referred to as Res-like macrophages.

Five clusters could only be identified in atherosclerotic aorta and represented the atherosclerosis-associated immune cell populations. These five clusters consist of $CD8^+$ T cells (*Cd8a*, *Cd8b1* and *Nkg7*), MoDCs / dendritic cells (*Cd209a*, *Cd74*, *Flt3* and *H2-Eb1*), monocytes (*Ly6c2*, *Ccr2* and *Csf1r*) and two population of atherosclerosis-associated macrophages.

Heterogeneity of macrophage populations in atherosclerotic aorta

Overall, we detected three populations of macrophages in atherosclerotic aorta namely Res-like, Inflammatory and $TREM2^{hi}$ macrophages (figure 46). Population of macrophages had the highest proportion of cells in the atherosclerotic aorta. In total, 28.8% of cells in the diseased condition were macrophages which was composed of Inflammatory macrophages with 47%, Res-like with 34.4% and $TREM2^{hi}$ with 18.6% of cells (figure 46).

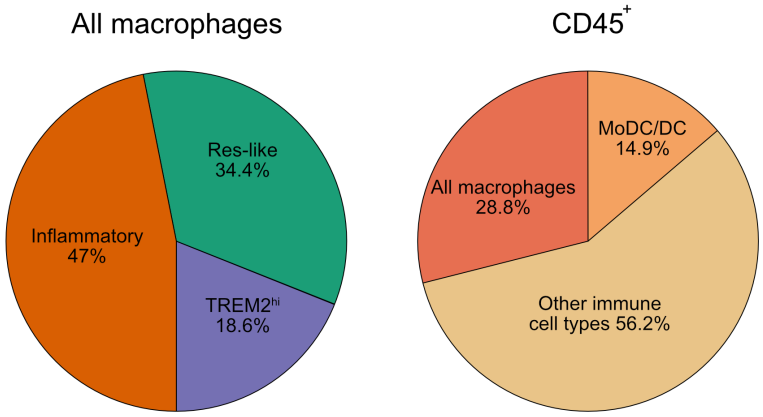

Figure 46. The ratio of different cell types in athrosclerotic aorta. The ratios are shown for total CD45+ cells (right) and all macrophage populations (left).

Next we tried to better understand the gene signature specific for each one of these groups of macrophages. To do that, we performed differential expression analysis between the groups of macrophages to see what are the genetic programs that discern these groups of cells from each other. We used single-cell differential expression (SCDE) analyses in each group versus the 2 other groups. Our analysis showed Inflammatory macrophages have higher expression of genes which are associated with proatherogenic roles (*Ccl3*, *Il1b*, *Il1a*, *Nlrp3*, *Cebpb*, *Egr1*, and *Phlda1*). Res-like macrophages expressed genes related to aortic resident macrophages (*F13a1* and *Lyve1*) and genes associated with M2-like macrophages (*Folr2*, *Cbr2*, and *Mrc1*). TREM2^{hi} macrophages, not only expressed *Trem2* at a higher level but also genes like *Cd9*, *Spp1* (encoding osteopontin), *Hvcn1*, and several cathepsins (*Ctsd*, *Ctsb*, and *Ctsz*) were enriched in this group of cells (figure 47a).

To see the functional state of each one of these groups of cells, we performed gene set enrichment analysis. Res-like and inflammatory macrophages shared several gene ontologies like processes related to inflammation and myeloid leukocyte activation. On the other hand, TREM2^{hi} cells had a unique profile associated with organic substance and cellular catabolic processes, lipid metabolic processes, regulation of cholesterol efflux and oxidative stress (figure 47b).

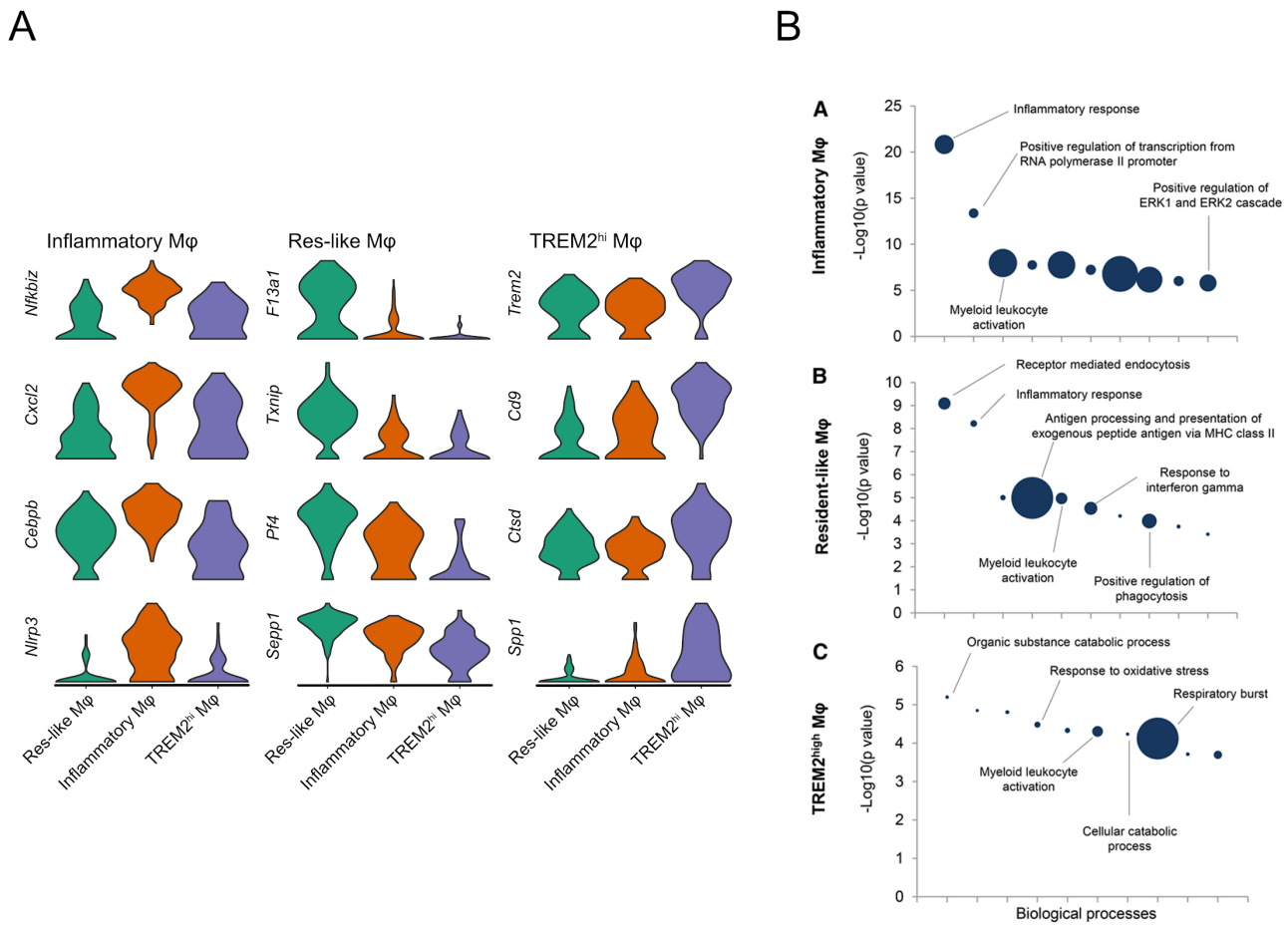

Figure 47. The gene profile of three macrophage populations. A. Violin plots showing the expression level of selected genes differentially expressed in inflammatory Mφ (left), Resident-like Mφ (center) and TREM2hi Mφ. B. Gene ontology enrichment analysis on three macrophage populations.

Gene signature of Aortic MoDC/DC

It has been assumed that Dendritic cells (DCs) and macrophages have many functions in common in atherosclerotic aorta and how to discern dendritic cells from macrophages in the plaque and what are the phenotypic markers that can be used to differentiate between these two population of cells is a long-standing question in the field of atherosclerosis research (Zernecke 2015). In our data, we could identify a major population of MoDC/DC which in total 14.9% of the cells in the data set belongs to this population (figure 48a).

We performed gene ontology enrichment of the genes expressed in this group of cells and the results showed that expected ontologies like antigen processing and presentation through MHCII or positive regulation of T cell activation were enriched in this group of cells. Interestingly, some of the gene ontologies were overlapping with inflammatory and Res-like macrophages (figure 48b).

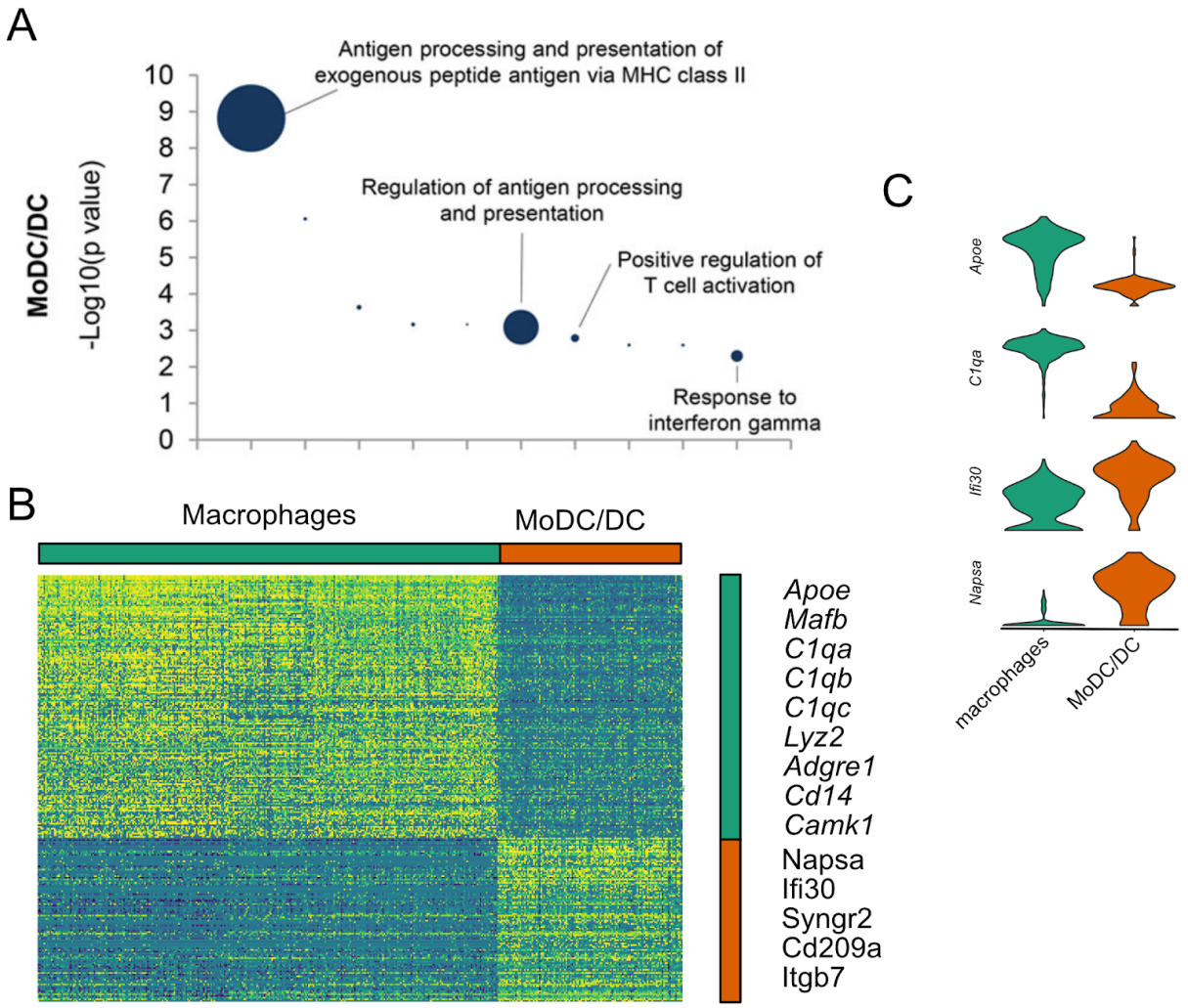

Figure 48. Gene signature of monocyte derived dendritic cells / dendritic cells in atherosclerotic aorta. A. Gene ontology enrichment analysis showing biological processes activated in MoDC/DC. Size of the circles shows the fold enrichment. B. Heatmap of differentially expressed genes between macrophages and MoDC/DC. C. Violin plots showing the expression level of selected differentially expressed genes in macrophages and MoDC/DC populations.

To further characterize the gene expression signature that differentiates between macrophages and MoDC/DC, we performed a differential expression analysis between the MoDC/DC population and total macrophages (including Res-like, inflammatory and TREM2^{hi} macrophages). Using differential expression analysis, we managed to find several genes that were exclusively expressed in one or another population. Overall, 531 genes were upregulated in MoDC/DC compared to 845 genes which were upregulated in macrophages. Genes such as *Apoe* and genes encoding complement C1q chains (*C1qa*, *C1qb*, and *C1qc*) were expressed in macrophages and on the other hand, genes like *Flt3*, *Ifi30* (γ-interferon-inducible lysosomal thiol reductase), *Napsa* (Napsin A), *Itgb7* (integrin β7), *Syngr2* (synaptogyrin-2) were expressed in M0DC/DC population (figure 48c).

MoDC/DC and macrophage cells populate aorta after 20 weeks of high fat diet treatment

To study if three populations of macrophages and MoDC/DC also populate the aorta plaque at more advanced stages of atherosclerosis, we repeated our experiment after 20 weeks of high fat diet treatment. We sorted CD45^{+} cells isolated from *Ldlr*$^{-/-}$ mice aorta after 20 weeks and performed single-cell RNA-seq on them. After performing common data analysis procedure, dimension reduction and clustering, we could again identify the three populations of macrophages (Res-like, inflammatory and TREM2^{hi}) and a group of cells with a transcription profile similar to MoDC/DC. Our data showed that the proportion of macrophages among the CD45^{+} cells increased after 20 weeks of high fat diet treatment and macrophages represented 49.6% of leukocytes. The macrophage population was composed of 46.3% Res-like macrophages, 13.4% TREM2^{hi} macrophages and 40.3% inflammatory macrophages (figure 49). Comparison of the proportion of the macrophages after 11 and 20 weeks of treatment showed that the level of TREM2^{hi} macrophages changed slightly (from 5.4% to 6.6%) whereas level of inflammatory macrophages (from 13.6% to 20%) and Res-like macrophages

(from 9.9% to 23%) experienced a substantial increase. Overall, this data shows that the three populations of macrophages also populate the lesion in aorta in more advanced stages of the disease.

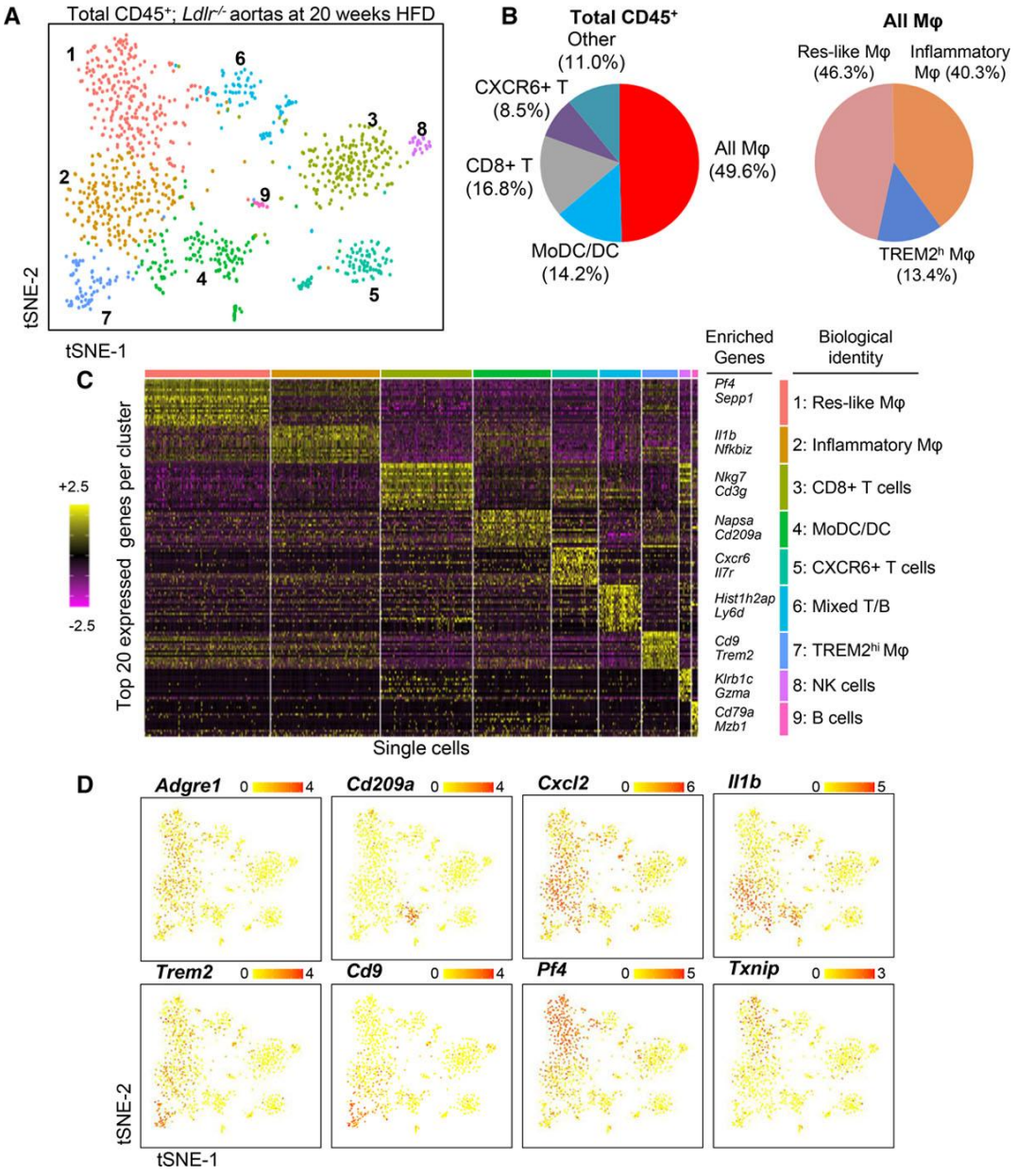

Figure 49. Immune cells population in advanced atherosclerosis. A. t-SNE visualization of 9 distinct clusters of immune cells. B. proportion of different cell types within CD45+ cells (left) and proportion of different macrophage subpopulations (right). C. Heatmap showing the expression of top 20 differentially expressed genes between identified clusters. D. Expression of Adgre1, Cd209a, Cxcl2, Il1b, Trem2, Cd9, Pf4 and Txnip genes projected on t-SNE.

Discussion

We applied single-cell RNA-sequencing to measure the heterogeneity and landscape of the leukocytes infiltrating the athrosclerotic aorta and we managed to discover three populations of macrophages associated with atherosclerosis. One population with resident-like gene expression profile and two other populations of macrophages named inflammatory macrophages and TREM2^{hi} macrophages. We also managed to discriminate the expression profile of MoDC/DC from macrophages and showed the differences in the expression profile of these populations. In our analysis, we showed that three populations of macrophages and MODC/DC cells were populating the athrosclerotic aorta at different stages of atherosclerosis progression and they were present in the intermediate and advanced stages of atherosclerosis.

All of the three types of macrophages discovered in this analysis showed high expression of *Apoe* and complement *C1q* genes. These genes are involved in pathways like reverse cholesterol transport and restricting inflammation which can be associated with the protective role of macrophages (Bhatia et al. 2007; Shi et al. 2004; Zanotti et al. 2011). Another gene with common expression between the three populations of macrophages was *Mafb* which controls the survival of macrophages in plaque (Hasegawa et al. 2016; Hamada et al. 2014).

According to the results of this analysis, the relative proportion of macrophages in CD45^{+} increased in advanced stages of atherosclerosis. This trend is inline with previous reports (Galkina et al. 2006). We showed that this increase is mainly due to the increase in the number of Res-like and inflammatory macrophages and suggests that a shift in macrophage population may occur in advanced stages of atherosclerosis.

The role of MoDC/DC in plaque formation is unclear. This is to some extent because dissecting MoDC/DC from macrophages and monocytes via flow cytometry or

immunohistochemistry is difficult (Zernecke 2015). The power and resolution of single-cell RNA-sequencing enabled us to compare the transcriptional profile and function of MoDC/DC and macrophages. Interestingly, the enriched pathways were quite similar between the two populations but they showed a clearly different gene expression profile. Further experiments are needed to characterize the functional consequences of the difference in gene expression to elucidate the precise role of these cell subsets in atherosclerosis.

Inflammatory macrophages showed a gene signature which is prototypical of proinflammatory plaque macrophages. Expression of genes like *Il1b* and *Nlrp3* shows that these subsets of cells are capable of expression of IL-1β. It has been shown that IL-1β cytokine has a crucial proatherogenic role in aorta (Kamari et al. 2011). This subset of cells also expressed chemokines and cytokines like *Cxcl2* and *Tnf*. Expression of NF-κB (nuclear factor κB) inhibitors like *Nfkbia*, *Nfkbiz*, *Nfkbid* was also a prominent feature of these cells. Inflammatory macrophages also expressed *Zpf36* which is shown to be able to limit atherosclerosis (Kang et al. 2011).

Res-like macrophages showed a gene expression profile that is similar to M2 macrophages. These cells expressed genes such as *Folr2*, *Mrc1*. On the other hand the expression of *Tnf* and *Il1b* was downregulated. This expression profile suggests an antiatherogenic phenotype among Res-like macrophages. However, Res-like macrophages also expressed genes that were associated with proatherogenic activity. For example, expression of *Txnip* and *Pf4* which were highly upregulated in Res-like macrophages, has been previously linked to promotion of atherosclerosis (Byon et al. 2015).

In healthy aorta, macrophages with expression of *Lyve1* and *F13a1* can also be found. In our data analysis, some of the macrophages from the healthy aorta are also clustered as Res-like macrophages which means that these cells share a similar gene expression profile. Although genes like *CCr2* and *Lyve1* were differentially expressed

between healthy and diseased conditions. Further studies and development of lineage tracing models are required to accurately bactrack these populations.

The third population of atherosclerotic macrophages were a novel population which we called TREM2^{hi} macrophages. This population of cells showed very distinct functional features based on the gene ontology enrichment analysis and they seemed to be involved in tasks like lipid handling and catabolic pathways. The gene signature of these macrophages is surprisingly similar to other macrophage lineages like osteoclasts and disease associated microglia. Differentiation of macrophages into osteoclasts in chronic inflammatory conditions has already been reported (for example in rheumatoid arthritis (Adamopoulos and Mellins 2015)) and it might also take place in the case of atherosclerosis. In advanced atherosclerosis, calcification occurs in lesions (Nakahara et al. 2017) and the osteoclast like phenotype of TREM2^{hi} macrophages suggest that this population might play a role in atherosclerotic plaque calcification.

The gene signature of TREM2^{hi} macrophages also resembles TREM2-dependent disease-associated microglia in Alzheimer's disease. TREM2 was shown to be required to maintain microglia and macrophage metabolic fitness and to preserve their function in Alzheimer's disease, which could also be relevant for macrophages in the atherosclerotic plaque microenvironment (Ulland et al. 2017).

Whereas the origin of TREM2^{hi} macrophages could not be determined based on this experiment and requires further investigation but the origin of disease associated microglia might give some ideas about the origin of TREM2^{hi} cells. Disease associated microglia are proposed to differentiate locally from microglia (Keren-Shaul et al. 2017). Microglia are self-renewing tissue resident myeloid cells which originate from yolk sac (Prinz, Erny, and Hagemeyer 2017).

It has already been suggested that Alzheimer's disease and atherosclerosis share important underlying mechanisms (Lathe, Sapronova, and Kotelevtsev 2014). Therefore, regardless of their origin, TREM2^{hi} myeloid cells might arise in response to

microenvironments that are shared between atherosclerosis and Alzeihmer's disease. Microenvironments like local inflammation, unfolded protein, metabolic stress and altered lipid metabolism might be the contributing factors which lead to this phenotype.

The concept of polarized M1/M2 macrophages seemed to be insufficient to describe the heterogeneity of macrophages in our *in vivo* model. Although one could argue that Res-like macrophages and inflammatory macrophages are similar to the polarized M1/M2 phenotype, we noticed that the expression of markers that are commonly used to describe M1/M2 states are overlapping between the two populations. On top of that, the phenotype of TREM2^{hi} macrophages could not be associated with any of the polarized phenotypes that are described *in vitro*. These findings once again show that the previously defined nomenclature to describe macrophages *in vitro* is insufficient to characterize the diversity of these cell types in an *in vivo* model.

In conclusion, based on an unbiased single-cell RNA-sequencing, we here established the transcriptional profile of 3 major macrophage populations and MoDC in mice atherosclerotic aorta. We demonstrated the existence of a novel atherosclerosis associated, TREM2^{hi} macrophage population with distinct gene profile and highly specific functions.

Chapter 6: Infection Atlas

Infection Atlas is a web based platform to browse and visualize single-cell RNA-seq data across different technologies with a focus on infection studies. The platform was developed based on the shiny R package and generates a number of diagnostic plots for visual inspection of single-cell sequencing data. To date, five different data sets have been published via this platform that can be accessed via the links below:

- https://infection-atlas.org/6970782209/

  Description: Blood and heart monocyte/macrophage and neutrophils post myocardial infarction

  Chemistry: Cell Hashing + CITE-seq + 10X Single Cell 3' v3

  Reference Genome: Mus musculus - mm10

- https://infection-atlas.org/4099491356/

  Description: Heart monocyte/macrophages, days 0, 1, 3, 5 and 7 post myocardial infarction

  Chemistry: 10X Single Cell 3' v2

  Reference Genome: Mus musculus - mm10

- https://infection-atlas.org/1760719889/

  Description: Heart neutrophils, days 1, 3 and 5 post myocardial infarction

  Chemistry: Cell Hashing + CITE-seq + 10X Single Cell 3' v3

Reference Genome: Mus musculus - mm10

- https://infection-atlas.org/4629609923/

  Description: Heart CD11b+ cells, days 1, 3 and 5 post myocardial infarction

  Chemistry: Cell Hashing + CITE-seq + 10X Single Cell 3' v3

  Reference Genome: Mus musculus - mm10

- https://infection-atlas.org/9074526738/

  Description: Heart monocyte/macrophages, days 1, 3 and 5 post myocardial infarction

  Chemistry: Cell Hashing + CITE-seq + 10X Single Cell 3' v3

  Reference Genome: Mus musculus - mm10

A set of data visualization plots were implemented in the website. In principle, these plots can be expanded or modified in case new visualizations are required for a specific data set. Overall, the platform is composed of five different tabs. Each tab provides a set of data visualizations which covers a certain aspect of the data. Basic information, dynamics, gene expression, differentially expressed genes and SCENIC are the five tabs that were implemented in the platform and they can also be expanded if required. In the following section, a description of the visualizations available in each tab along with the related screenshots are provided.

Basic information

In the basic information tab, a brief description of the data like the type of cells under investigation, the chemistry of sequencing run and the reference genome used in the study were provided. A set of drop down menus, allow the user to simultaneously visualize different aspects of the data. For instance, the UMAP plots can be color coded via these drop down menus to see the cells corresponding to different cell populations or time points. The expression of hashtag antibodies used in the experiment can also be projected on the UMAP and drop down menus allow users to switch between the expression level of different hashtag antibodies. Moreover, the

common quality control metrics are also available for visualization in the basic information tab (figure 50).

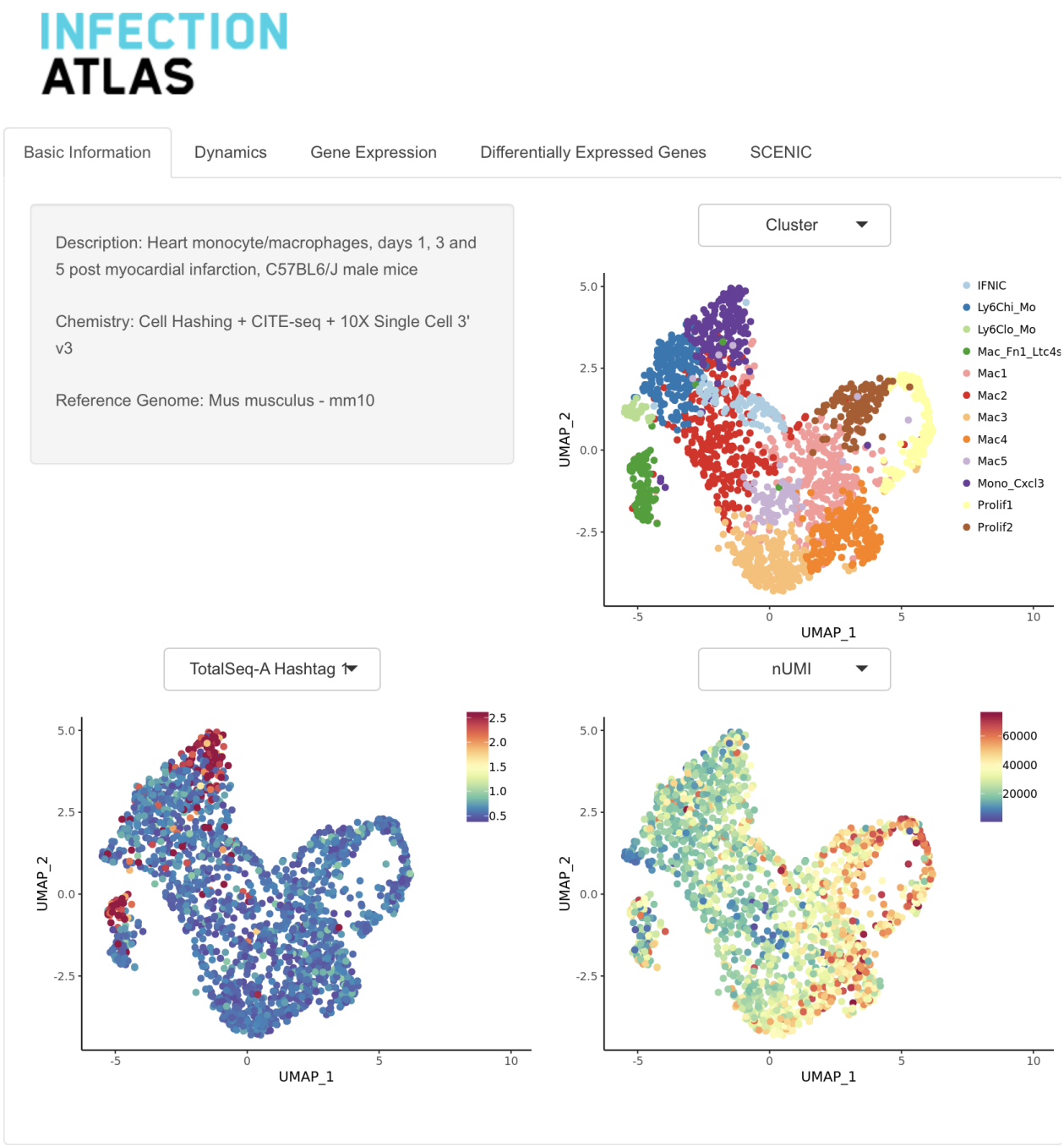

Figure 50. Screenshot of basic information tab.

Dynamics

The dynamics tab is devoted to visualize the temporal aspects of the data. A histogram is provided which shows the ratio of each cell type in each time point that study was conducted. In principal, increase or decrease in the abundance of different cell types over time can be investigated via the histogram. Moreover, a cell specific line graph which shows the change in abundance of cells over time was also implemented in the dynamics tab. The user can use the drop down menu in this tab to switch between cell types on the line graph (figure 51).

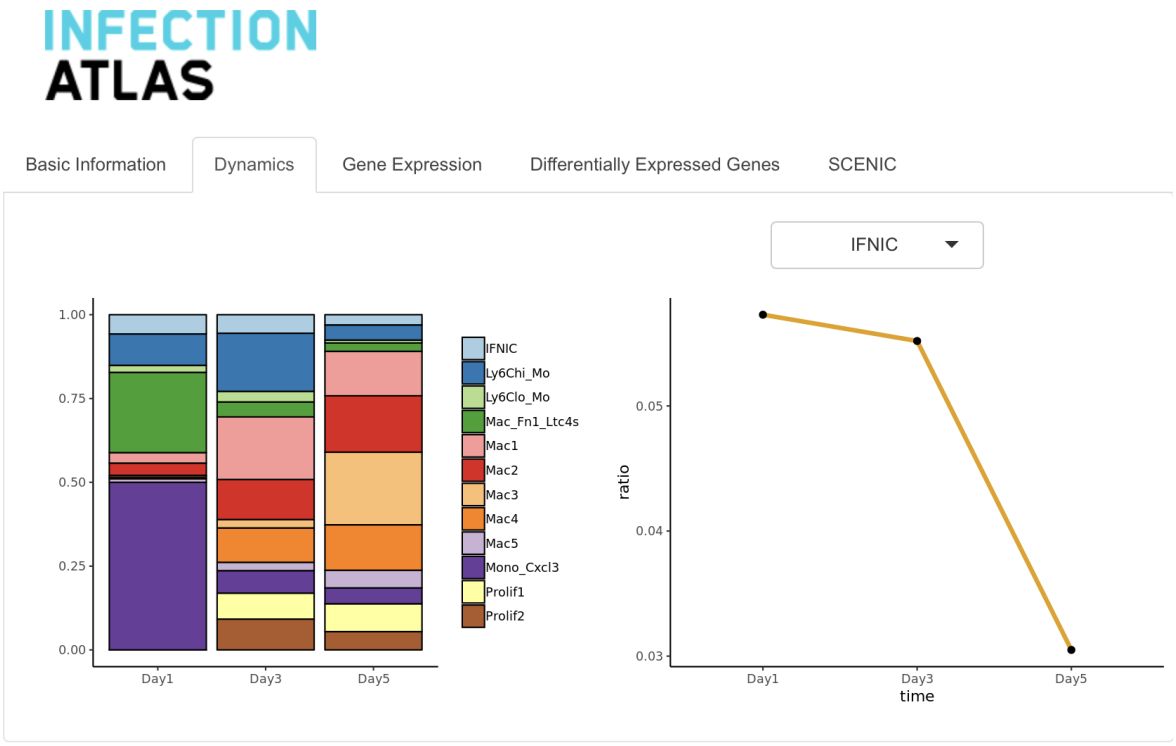

Figure 51. Screenshot of dynamics tab.

Gene expression

In the gene expression tab, different graphs to show the expression level of a gene in different cell subsets is provided. The drop down menu can be used to search for a

gene name. After selecting a gene, the graphs in this tab are updated to show the expression level of the gene in cell subsets. The expression level of the gene of interest is projected on the UMAP or it can be visualized via ridges plot or a violin plot (figure 52).

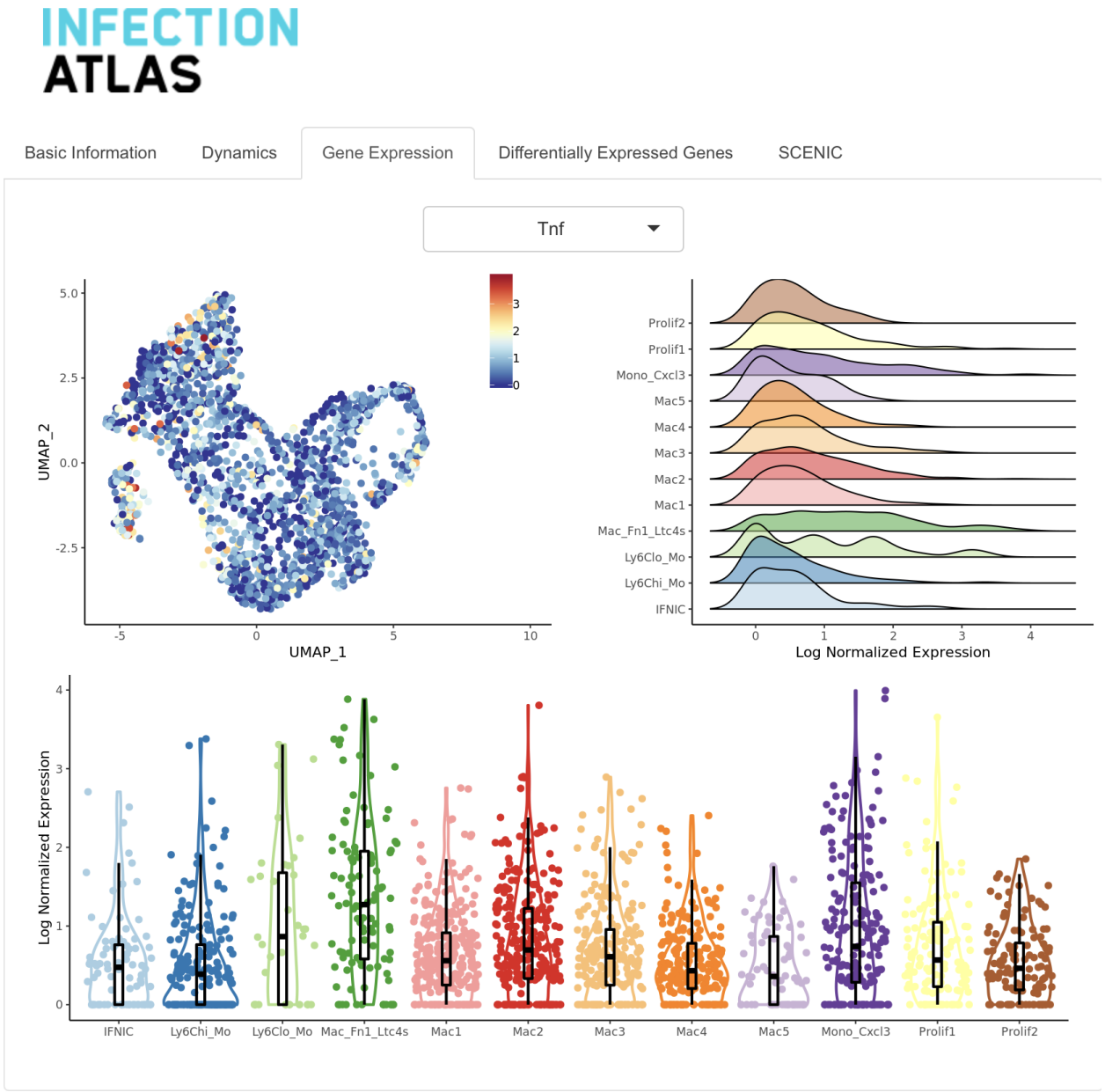

Figure 52. Screenshot of gene expression tab.

Differentially expressed genes

This tab provides a list of upregulated genes in each cell subset along with corresponding statistics. A radio button on the left side of the page was implemented that can be used to select a certain subset of cells. After selecting a cell subset, the list of upregulated genes is updated to show the differentially expressed genes in that subset. Moreover, a search button is provided that can be used to find the gene of interest in the list (figure 53).

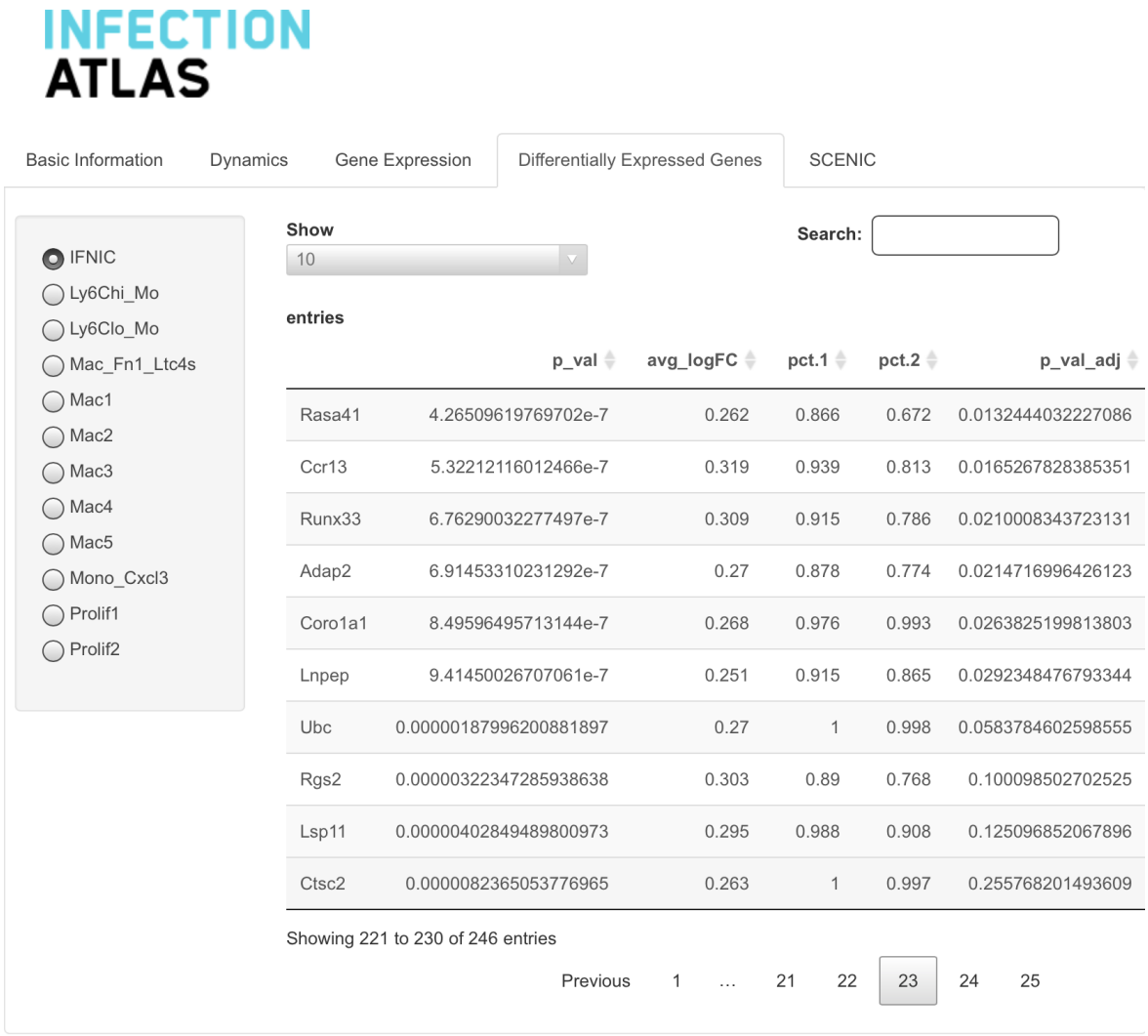

Figure 53. Screenshot of differentially expressed genes tab.

SCENIC

This tab is designed to interactively show the results of single-cell regulatory network inference and clustering (SCENIC) (Aibar et al. 2017). It provides a drop down menu where different regulons detected by SCENIC can be selected. By selection of regulons on the drop down menu, all the figures in this tab are updated to show the information related to that particular regulon. Overall four different figures are shown in this tab and each one of them show different information related to the chosen regulon. “AUC histogram” shows the overall distribution of AUC values in the data and frequency of each value can be visualized in the form of a histogram. A dashed vertical line on the histogram shows the value of the selected AUC threshold on the histogram. This threshold can be modified via the slider button on the top of the page and according to the selected threshold, the cells with AUC values higher than the threshold are considered ON for that regulon and all the other cells are considered OFF. On the UMAP in the upper right corner of the page, the ON and OFF assignments of different cells can also be visualized which helps the user to understand which clusters in the data are ON regarding that gene program. In this UMAP the cells that are ON based on the threshold are in blue whereas the other cells are in grey color. This UMAP is also updated while the value on the slider button changes. In the “Gene set activity (AUC)” plot, the AUC values corresponding to each single cell is projected on the UMAP and the “Expression of transcription factor” plot simply shows the RNA expression level of the transcription factor related to the selected regulon. Overall this tab provides an easy to use interface to browse and visualize SCENIC results which otherwise would be challenging for biologists with no bioinformatics skills to access (figure 54).

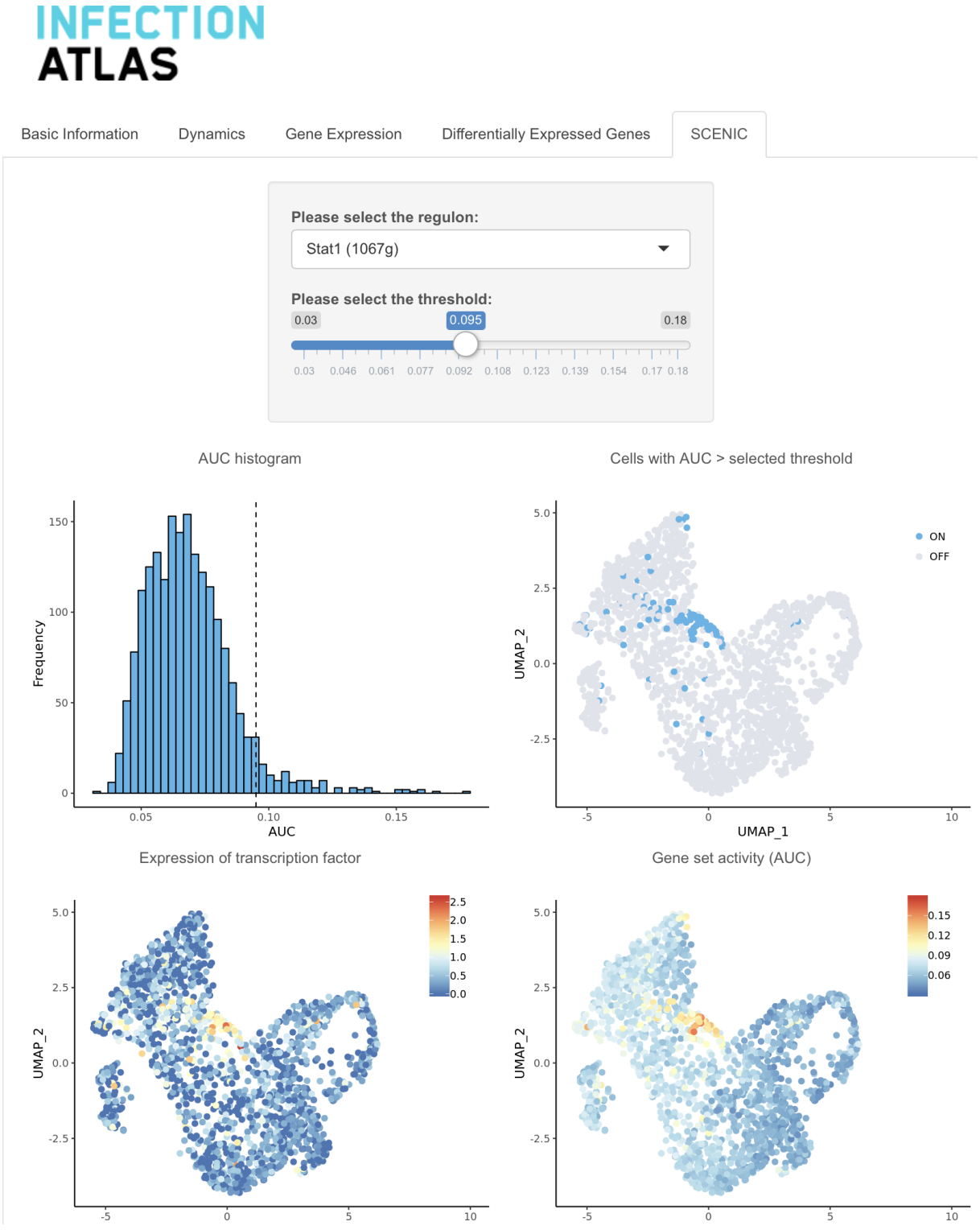

Figure 54. Screenshot of SCENIC tab.

Discussion

Technological advances have made single-cell sequencing affordable and a routine task in many laboratories. This technique has proved to be a valuable tool to study infection. Different aspects of infectious diseases like heterogeneity, dynamics and topology of infection can be studied via single-cell sequencing methods. Here, we provided a user-friendly interface that can be used to browse and visualize different aspects of single-cell sequencing data.

Whereas the current version of Infection Atlas website provides an interface to visualize single-cell RNA-seq data, it is not limited to this type of data and in principle any kind of -omics data can be loaded and visualized via this platform. Also the visualization methods can be expanded if required for a particular data. The corresponding scripts automates loading the data on the website and makes it easy to administer the web server and to add new data sets to the website. Once the data sets have been annotated by experts, they can be easily loaded on the server and used by biologists to further explore the infection process.

The data on the website is implemented and structured on a HDF5 file format. This file format offers efficient, on-disk storage of the data. Therefore, deploying and accessing data on the server is not computationally demanding and in principle, many datasets can be added to the web server without increasing the computational capacity of the machine.

Infection Atlas will expand in the future to include a greater number of normal and diseased cell types. Platforms to visualize sequencing data other than single-cell RNA-sequencing data will be added to the website to provide a global and comprehensive map of host-pathogen encounters and the outcomes of this interaction.

Infection Atlas is a complement of other efforts to create a global map of human cells and is distinct in the respect that it focuses on elucidating pathogen specific molecular and cellular features and the particular immune responses that are activated upon pathogen encounter.

Appendix 1: List of genes specific to infected ciliated cells

After performing differential expression analysis between mock, bystander and infected ciliated cells, upregulated genes in infected condition were selected. Then, a set of stringent thresholds were defined to select the genes that are specific to infected ciliated cells (genes that were neither highly expressed in bystander nor in mock). These thresholds select genes that their median is 2 times higher than bystander cells **AND** 2 times higher than mock cells. With these thresholds, we try to find genes that are not a part of the host innate immune response therefore are solely induced by RSV virus. Overall, 89 genes were identified that passed these thresholds. Here, the full list of genes is provided:

	Median infected	Median bystander	Median mock
HLDA2	"2,009"	"0,476"	"0,515"
LAMB3	"1,222"	"0,163"	"0,301"
FGFBP1	"1,231"	"0,152"	"0,36"
SFN	"1,548"	"0,357"	"0,624"
TNFRSF12A	"1,231"	"0,355"	"0,334"
EMP3	"0,71"	"0,026"	"0,03"
TPM4	"1,353"	"0,469"	"0,661"
ANXA3	"1,1"	"0,377"	"0,35"
FLNA	"0,642"	"0,02"	"0,06"
PHLDA1	"0,737"	"0,117"	"0,123"
EMP1	"0,653"	"0,056"	"0,095"
ITGA6	"0,728"	"0,127"	"0,111"
HBEGF	"0,611"	"0,06"	"0,086"
SDR16C5	"0,821"	"0,208"	"0,2"
SCEL	"0,489"	"0,023"	"0,043"
ENC1	"0,568"	"0,129"	"0,116"
LAMA3	"0,327"	"0,04"	"0,051"
ELOVL6	"0,339"	"0,051"	"0,056"
GJB3	"0,377"	"0,051"	"0,074"
GJB4	"0,505"	"0,121"	"0,105"
PRKCDBP	"0,817"	"0,171"	"0,165"
PTAFR	"0,379"	"0,078"	"0,061"
FHL2	"0,543"	"0,107"	"0,167"
SEPT11	"0,797"	"0,317"	"0,313"

ECE1	"0,584"	"0,184"	"0,172"
RHOD	"0,691"	"0,255"	"0,247"
HMOX2	"0,771"	"0,337"	"0,291"
IER3	"1,048"	"0,207"	"0,473"
PDLIM7	"0,441"	"0,087"	"0,093"
SDCBP2	"0,591"	"0,131"	"0,169"
BIRC3	"0,765"	"0,261"	"0,166"
AQP5	"0,652"	"0,082"	"0,205"
TIMP1	"1,278"	"0,54"	"0,604"
SRPX2	"0,607"	"0,237"	"0,198"
UXS1	"0,435"	"0,151"	"0,118"
DUSP6	"0,44"	"0,093"	"0,109"
HDAC9	"0,601"	"0,247"	"0,159"
LAMC2	"0,777"	"0,307"	"0,293"
CTSV	"0,278"	"0,045"	"0,039"
IL1RN	"0,343"	"0,067"	"0,083"
CD44	"0,305"	"0,016"	"0,095"
TUBA1C	"0,885"	"0,44"	"0,431"
HOMER3	"0,411"	"0,056"	"0,132"
ITGA3	"0,765"	"0,34"	"0,313"
TM4SF1	"1,081"	"0,409"	"0,482"
SLC9A1	"0,327"	"0,093"	"0,096"
PKP3	"0,619"	"0,28"	"0,271"
PLK3	"0,388"	"0,1"	"0,111"
PLEC	"0,823"	"0,39"	"0,348"
SLC12A2	"0,508"	"0,125"	"0,207"
RP11-445L6.3	"0,405"	"0,14"	"0,091"
PPL	"0,726"	"0,324"	"0,318"
ELL2	"0,537"	"0,191"	"0,237"
CAV2	"0,569"	"0,124"	"0,248"
BCR	"0,467"	"0,182"	"0,182"
SERPINB8	"0,393"	"0,132"	"0,107"
VCL	"0,57"	"0,209"	"0,254"
LRRC8A	"0,493"	"0,117"	"0,194"
EHD1	"0,362"	"0,119"	"0,094"
CKAP4	"0,408"	"0,104"	"0,163"
TIMP2	"0,352"	"0,089"	"0,113"
BID	"0,619"	"0,26"	"0,275"
MALL	"0,314"	"0,068"	"0,082"
GLUD1	"0,526"	"0,234"	"0,234"
HMGA1	"0,657"	"0,2"	"0,312"
HEXIM1	"0,589"	"0,278"	"0,247"
CTNNBIP1	"0,41"	"0,156"	"0,172"
TNIP1	"0,492"	"0,213"	"0,208"
TMEM63A	"0,457"	"0,215"	"0,184"

PLAUR	"0,38"	"0,059"	"0,15"
EHBP1L1	"0,375"	"0,126"	"0,147"
TMEM61	"0,364"	"0,159"	"0,106"
BCAR1	"0,393"	"0,147"	"0,168"
IL32	"0,423"	"0,07"	"0,129"
TUBB6	"0,229"	"0,014"	"0,052"
AREG	"0,619"	"0,19"	"0,294"
MGLL	"0,448"	"0,22"	"0,144"
SEMA3A	"0,457"	"0,196"	"0,205"
KRT80	"0,417"	"0,176"	"0,175"
VASP	"0,441"	"0,213"	"0,205"
NEDD9	"0,414"	"0,153"	"0,173"
CAMK2N1	"0,259"	"0,019"	"0,056"
INTS6	"0,453"	"0,198"	"0,192"
MYL9	"0,455"	"0,183"	"0,135"
LINC00152	"0,457"	"0,195"	"0,219"
C8orf4	"0,313"	"0,074"	"0,094"
PDE4C	"0,38"	"0,158"	"0,15"
LINC00326	"0,395"	"0,197"	"0,156"
G0S2	"0,202"	"0,024"	"0,039"

Appendix 2: The list of genes in each wave of transcription

The list of genes in each wave of transcription upon RSV infection is provided below.

Wave A:

SCGB1A1, CDC20, NEK2, CCNO, STC1, CENPW, CDC20B, SGOL2, C2orf74, MT1G, HIST1H2BJ, ANLN, AKR7A2, PTTG1, HIST1H2BC, HIST1H2AC, TMEM106C, CKS2, HTATIP2, PRR32, HIST1H1C, CDK1, TOP2A, SPAG5, MAD2L1, CROCC2, SMC4, RIBC2, TACC3, YPEL1, IL5RA, BBOX1, GJB7, RP3-467N11.1, GAS6-AS1, PWWP2B, RP11-872J21.3, CKS1B, CCDC60, HAGH, RNF32, BPIFB1, HLA-DRB5, MARCKSL1, VMO1, RP11-60L3.6, USP13, BRCA2, LINC00240, REEP1, MROH9, ISYNA1, NOS3, WBP5, LINC01091, HIST1H2BD, AC195454.1, WIPI1, MYD88, PBXIP1, C16orf71, LPCAT1, C2orf76, ZBTB21, FANCI, OASL, MT1E, PSIP1, SLF2, CEACAM1, BISPR, AC009133.12, ACSL1, BTG1, MEIG1, TECR, YPEL2, ANKRD65, METRNL, CDKN1B, TGIF1, SHISA8, CEP78, RP11-297N6.4, CALM3, STIL, TXLNB, CCDC18, WFDC6, REC8, AURKA, RP11-166P13.3, KIAA1841, OSBPL1A, STAU2

Wave B:

GSTA2, GSTA1, CEL, ALDH3A1, AKR1B10, HRASLS2, ISG15, GPX2, ADH7, ALDH1A1, ISG20, IFI35, APOL1, IRF7, PPM1K, LAMP3, LAP3, GBP1, IFITM2, SLC15A3, UBE2L6, SHISA2, TFRC, FAM3D, CYP4B1, FAM46A, PLSCR1, AKR1C2, CRYM, CTGF, RABAC1, CCDC88A, IQGAP2, CNKSR3, ENDOD1, CYP2J2, HAS3, DMRTA2, HLA-E, FAM184A, LYPD1, RP11-902B17.1, LGALS9, LIFR, PALLD, LGMN, RASSF6, HEXIM1, SLFN12, FMN1, GDAP1, F2RL1, C21orf91, BTC, TP53INP2, CTTNBP2NL, ATP13A3, ZNF22, PIK3R1, GLTP, PLS1, BDP1, MRPL17, DTX3L, MUCL1, LY6E, RDH10, ITPRIPL2, ADAM28, RAI14, DDT, EML4, TM7SF2, ZNF37A, IFIT2, RSAD2, HERC5, DHX58, MUC13, IFIT1, IFIT3, TNFSF10, CNTD1, MT1M, LINC01207, OAS1, IFIH1, EPSTI1, USP18, GBP6, TCHP, SOX4, SECTM1, PTPRT, SLC7A2, USP30-AS1, KDELC2, RNF213, SPRY1, CASP1, TRIM38, TCEAL8, SYNPO2,

GALM, JMY, SMIM1, NUDT14, ZNF292, ANKRD13A, TMEM37, PRR15L, CBX1, BARD1, SMIM19, SCLT1, LRMP, FAM111A, ANKIB1, RIN2, GRAMD1A

Wave C:

PCDH11Y, MMP1, HBB, PSCA, ATP12A, C2orf40, CX3CL1, GABRB3, NQO1, CKB, ID4, CCL15, MCTP1, KCTD12, PLEKHO1, SLC12A2, GSDMB, EGLN3, SLC16A9, ARHGAP29, INHBB, CAPS, TRNP1, CD55, FXYD3, ZBTB38, GATA6, PSMB10, PPP3CA, VCL, SLC30A1, GBP3, UBA6, SCD5, FABP5, TRIM31, XIST, CES1, AQP5, SPON2, MYL9, BIRC3, CITED4, THBS1, SCPEP1, ARL4C, CEBPA, FERMT1, PSTPIP2, PTMS, CCDC181, ZBED2, C11orf85, RAB30, MAP2, ENC1, RARRES3, SLC16A5, SLC16A2, GLUL, RP5-1185I7.1, ATP2B1, TMEM61, GCNT1, PAQR8, PDP1, PPP1R3B, TNFRSF10A, CLDN1, A4GALT, ADGRF1, ATP5G1, AGO2, MFSD2A, TAP2, GPATCH4, LPCAT4, WNT7B, FBLIM1, STK17B, PTPN12, RAB31, COX7A1, PEG10, LY6D, UNC5B-AS1, MTRNR2L8, STATH, RASGRP3, CENPM, TRIM7, RIN1, CCDC173, ANKRD66, SRM, LINC01513, SLC16A1, NOP56, HSPA12A, GNL3, RBM41, CKLF, LGALS9B, NOP14, RP11-766F14.2, PSMB9, HAUS2, MZF1-AS1, CYP2S1, PALMD, CCDC190, SLC25A5, ARHGAP18, RP11-489E7.4, RRS1, MT-CO3, DIAPH2, TMEM41A, ZP3, SIPA1L2, AP3M2, PLEKHG3

Wave D:

G0S2, S100A2" "SCGB3A1, KRT6A, SPRR1B, CXCL10, IGFBP6, CCL5, HSPA1A, IL32, IFNL3, LGALS7B, LGALS7, DDIT4, KLK7, CAMK2N1, DDIT3, KRT13, LAMB3, PLAU, RHOF, SERPINB4, C11orf96, KRT7, HSPA1B, S100A14, TUBB6, IL23A, RBP1, PRKCDBP, TIMP1, C3, TRIB3, UPK1B, IL11, MUC5AC, KLF2, KLK6, CCNA1, RRAD, RP13-463N16.6, MT-ATP8, PTRF, RPS4Y1, CRABP2, RHOB, HMGA1, TNC, TRIM55, IL20RB, CASC8, BCL2A1, FXYD5, WNT4, MT-ND4L, COL17A1, CDK2AP2, JUN, HES1, HOMER3, MT-ND5, MKNK2, RP11-21B23.2, LMTK3, MALL, KCP, NT5E, BBC3, KLK11, LY6G6C, PYCARD, ATF3, CORO1A, HERPUD1, NEAT1, AP000769.1, CHAC1, FASN, FAM83A, CAV2, RP4-666F24.3,

RFX8, LINC00342, AKAP14, COX6B2, TMEM132A, LAMA3, ZFP36, RP11-284F21.10, RP11-263K19.4, CLDN9, RPS27L, KYNU, VIM-AS1, S100A13, AHNAK, LYPD3, SEMA3B, WNT7A, HOMER2, LETM2, ZFAS1, ZNF487, KCTD11, ADGRF4, NGEF, PORCN, ITGAE, OVOL1, TNFRSF10D, S100A6, SLC31A2, PIM3, CSRP2, BID, TMSB4X, CALR" ADAM8, PRR34-AS1, ERCC3, NRP1, ITGB1BP1, MMP28, CAPG, UBE2V2, MFGE8, TNS3, C15orf52, PPID, CRELD2, LINC00704, PPP1R14C, FBN2, IFT20, RHEBL1, PNRC1, MRVI1-AS1, CD9, PIEZO1, RAC2, ANXA2, CFAP126, NFIL3, SNX3, DUSP8, JUND, RP5-875H18.9, BAD, NCK1, HSP90AA1, LTBP3, RP11-465N4.4, ARL16, TMEM14B, TAGLN2, P4HA2, RP11-660L16.2, NECAP2, TMA7, MAPK8IP1, AKNA, CCDC167, ADTRP, PSAT1, SERPINA1, GPX8, LLNLR-245B6.1, LINC00920, XBP1, CEP19, EIF5, ERICH2, PPP1R1C, MAFG, TTC29, RRAGD, CBX4, TTLL10, FRMPD2

Wave E:

MUC12, SYT5, HES6, RP11-620J15.3, TPPP3, CDKN2D, MYLK, PARD6G-AS1, PPP1R7, ZMYND10, CBY1, AC007405.6, IFT22, DTNA, EFCAB1, FAM46B, TEKT1, LRRC10B, HIST1H4H, DYNC2H1, TMEM99, ZDHHC1, PSMC3, FAM227A, BPIFA1, LYPD2, RP11-356K23.1, ZC3HAV1, CCDC74B, CCDC74A, RP11-60L3.1, SRGAP3-AS2, FAM92B, FAM216B, FAM229B, ERCC1, FHL1, LINC01571, MRLN, CARS, CFAP36, CTXN1, FAM183A, ATG9B, FAM104B, PPP1R42, C1orf141, PCAT6, ARL3, C11orf74, RP11-1223D19.1, DENND6B, IRF2BPL, PNMA1, PMM1, DGUOK-AS1, CFAP45, KNDC1, GTPBP2, SHFM1, CCDC114, TCTN1, C14orf1, EXOC5, ST8SIA1

Appendix 3: list of differentially expressed genes in 10-pooled bacteria

Differentially expressed genes for 10-pooled bacteria and comparison of our study with the reference bulk RNA-seq dataset from Kröger et al. 2013. Gene ID and gene name of differentially expressed genes of 10-pooled bacteria in the three growth conditions are listed in column 1 and 2, respectively (color code in Column 2: blue: 'Late Stationary Phase'; green: 'Salt (NaCl) shock'; red: 'Anaerobic shock'). For each gene differentially expressed between Salt (NaCl) and Anaerobic shocks, column 3 and 4 recapitulate the gene expression extracted from Kröger et al. under Salt (NaCl) shock and Anaerobic shock, respectively. The ratio and log10 transformed ratio of (NaCl/Anaerobic) of the latter expression values from bulk-RNA-seq from Kröger et al. 2013 are calculated in columns 5 and 6, respectively.

Gene ID	Gene name	Expression values Salt (NaCl) Shock (Kröger et al) (TPM)	Expression values Anaerobic Shock (Kröger et al.) (TPM)	Ratio Salt(NaCl) /Anaerobic	Ratio Salt(NaCl) /Anaerobic (Log10)
SL1344_1619	pspB				
SL1344_2376	mntH				
SL1344_1618	pspC				
SL1344_0462	rpmE2				
SL1344_3761	ilvN				
SL1344_2773	csiD				
SL1344_P1_0022	traC				
SL1344_1620	pspA				
SL1344_P2_0074	traJ				
SL1344_0649					
SL1344_1063	putP				
SL1344_P2_0081	shfB				
SL1344_2756	fljB				
SL1344_3300	arcB				
SL1344_P2_0082	shfB2				
SL1344_1223					
SL1344_0899	lolA				
SL1344_1006	rmf				
SL1344_2287	nuoL				
SL1344_3936	fadB				
SL1344_0936	ompF				
EBG00001133906	Bacteria_small_SRP				
SL1344_1409	pntA				
SL1344_4056	katG				
SL1344_4487	yjjG				
SL1344_1763	yebN				
SL1344_4180	pspG				
SL1344_2131	mrp				
SL1344_2860	iacP				

SL1344_1148	ndh				
SL1344_2059	rfbP				
SL1344_3301	yhcC				
SL1344_4216	nrfD				
SL1344_0573	fepA				
SL1344_2239	rcsB				
SL1344_3106					
SL1344_4465					
EBG00000241440	lsrM				
SL1344_1616	pspE				
SL1344_2731					
SL1344_3746					
SL1344_1415	asr				
SL1344_2847	hilC				
SL1344_2838	hypE				
SL1344_1828	ruvB				
SL1344_4499	stjB				
SL1344_2043	sopA				
EBG00001133739	tmRNA				
SL1344_4346					
EBG00001133849	RNaseP_bact_a				
SL1344_4217	nrfE				
SL1344_1508	narZ				
SL1344_1696	ychP				
SL1344_4507	nadR				
SL1344_0348					
SL1344_4520	sthB				
SL1344_2786	ygaM	367	73.3	5.0068	0.6996
SL1344_1915	gcpA	7.9	6.9	1.1449	0.0588
SL1344_4224	phnO	42.7	13.5	3.163	0.5001
SL1344_3030	gcvH	107.1	524.4	0.2042	-0.6899
SL1344_0818	ybiV(1)	139.7	21.1	6.6209	0.8209
EBG00001133861	P26	NA	NA	NA	NA
SL1344_2297	nuoA	228	109.9	2.0746	0.3169
SL1344_P1_0080	ccdA	NA	NA	NA	NA
SL1344_0081		444.8	11.7	38.0171	1.58
SL1344_2792	nrdE	29.6	2.3	12.8696	1.1096
SL1344_4228	basS	18	25.9	0.695	-0.158
SL1344_0354		690	16.8	41.0714	1.6135
SL1344_1914	mngB	4.6	2.3	2	0.301
SL1344_0431		23.7	4.6	5.1522	0.712
SL1344_2304	yfbU	75	102.1	0.7346	-0.134
SL1344_1019	yccV	596.4	80.8	7.3812	0.8681
SL1344_0426	phnX	1	2.3	0.4348	-0.3617
SL1344_3471	yhgF	32	10.1	3.1683	0.5008
SL1344_1288	ydiQ	0	0.6	0	-Inf
SL1344_3397	rplN	1856.6	405.5	4.5785	0.6607
SL1344_1202		1183.3	62.5	18.9328	1.2772
SL1344_2370	yfdZ	56	51.7	1.0832	0.0347
SL1344_2796	proX	762	3.3	230.9091	2.3634
SL1344_2625	pheA	54.3	11.9	4.563	0.6593
SL1344_2436	tal	38.3	8.4	4.5595	0.6589
SL1344_0433	cyoE	232	43.9	5.2847	0.723
SL1344_0459	ybaY	266.1	44.8	5.9397	0.7738
SL1344_4369	cybC	539.4	225	2.3973	0.3797
SL1344_0461	ylaB	23.5	38.7	0.6072	-0.2166
SL1344_1732	ycgB	42.3	28.5	1.4842	0.1715
SL1344_1304	sufB	18.2	5.3	3.434	0.5358
SL1344_2660	smpB	240.3	106.4	2.2585	0.3538
SL1344_0868	ybjP	59.8	45.6	1.3114	0.1177

SL1344_2182	fruK	3	8.5	0.3529	-0.4523
SL1344_3472	feoA	52.3	15	3.4867	0.5424
SL1344_0212	yaeH	363.7	1003.5	0.3624	-0.4408
SL1344_3202	yqjE	159.7	200	0.7985	-0.0977
SL1344_1132	fabG	358.4	225.1	1.5922	0.202
SL1344_2648	rimM	523	175.2	2.9852	0.475
SL1344_3492	glpE	242.8	82.5	2.943	0.4688
SL1344_0386	yaiE	185.4	223.4	0.8299	-0.081
SL1344_0763	ybhC	17.5	16.8	1.0417	0.0177
SL1344_2183	fruB	2.3	3.5	0.6571	-0.1823
SL1344_0737	aroG	67.3	71.6	0.9399	-0.0269
SL1344_2108	baeR	8.5	9.3	0.914	-0.0391
SL1344_2486	yfgA	111.2	71	1.5662	0.1948
SL1344_3237	yraM	28.9	46	0.6283	-0.2019
SL1344_4456	hsdM	7.4	36.8	0.2011	-0.6966
SL1344_4450		13.3	8.8	1.5114	0.1794
SL1344_3955	typA	75.3	40	1.8825	0.2747
SL1344_3282	yrbC	89.5	58.6	1.5273	0.1839
SL1344_1245	nadE	139.7	81.4	1.7162	0.2346
SL1344_2464	ppk	70.5	56.8	1.2412	0.0938
SL1344_3396	rplX	3303.7	608.5	5.4293	0.7347
SL1344_4503	lplA	22.3	15.7	1.4204	0.1524
SL1344_4227	proP	1160.2	58	20.0034	1.3011
SL1344_0498	ybbO	24.8	21.7	1.1429	0.058
SL1344_3394	rpsN	1616.8	309	5.2324	0.7187
SL1344_2747		24.7	34.9	0.7077	-0.1501
SL1344_2118	fbaB	24	12.2	1.9672	0.2939
SL1344_2811	mltB	49.2	25.1	1.9602	0.2923
SL1344_2938	rumA	61.3	113.3	0.541	-0.2668
EBG00000241426	STnc700	NA	NA	NA	NA
SL1344_1774	prc	59.7	51.1	1.1683	0.0676
SL1344_P1_0055	parA	NA	NA	NA	NA
SL1344_4230	yjdB	65.6	23.2	2.8276	0.4514
SL1344_3838	atpB	208.8	211.6	0.9868	-0.0058
SL1344_P2_0010	yafB	NA	NA	NA	NA
SL1344_0871		13.3	15.5	0.8581	-0.0665
SL1344_1496	sfcA	20.8	23.6	0.8814	-0.0548
SL1344_0866	artI	32.1	46	0.6978	-0.1563
SL1344_3452	damX	63.6	72.6	0.876	-0.0575
SL1344_3505	glgB	38.5	45.7	0.8425	-0.0745
SL1344_3494		4.4	22.7	0.1938	-0.7126
SL1344_4076	yijC	158.9	144.3	1.1012	0.0419
SL1344_0089	apaH	51.8	38.7	1.3385	0.1266
SL1344_1537		33.1	125.7	0.2633	-0.5795
SL1344_0047	ileS	69.4	62.4	1.1122	0.0462
SL1344_1243		84.5	10.6	7.9717	0.9016
SL1344_3041	yggE	106.1	50.8	2.0886	0.3199
SL1344_3034	pepP	59.2	64.4	0.9193	-0.0366
SL1344_2280	elaB	254.5	147.8	1.7219	0.236
SL1344_3890	hemY	59.8	67.1	0.8912	-0.05
SL1344_2607	srmB	42	27.9	1.5054	0.1776
SL1344_3201	yqjC	788.3	229	3.4424	0.5369
SL1344_3708	spoT	86.3	74.3	1.1615	0.065
SL1344_1666	ispZ	49.9	44.7	1.1163	0.0478
SL1344_3393	rpsH	1802.1	318.1	5.6652	0.7532
SL1344_3523	ugpB	39.3	29.6	1.3277	0.1231
SL1344_1767	yobF	2068.5	2360.9	0.8761	-0.0574
SL1344_0044	rpsT	3440.4	967.7	3.5552	0.5509
SL1344_1493	osmC	260.7	25.4	10.2638	1.0113
SL1344_3029	gcvP	24.8	114.2	0.2172	-0.6632
SL1344_2290	nuoI	85.7	88.6	0.9673	-0.0145

SL1344_3875	trxA	1663.3	477	3.487	0.5425
SL1344_0442	clpP	509.3	218.4	2.332	0.3677
SL1344_4013	cdh-a	134.8	23.6	5.7119	0.7568
SL1344_0045	yaaY	86.8	50.9	1.7053	0.2318
SL1344_4008	cpxR	56.3	46.8	1.203	0.0803
SL1344_1237	xthA	21.1	26	0.8115	-0.0907
SL1344_0187	dksA	412.6	225	1.8338	0.2633
SL1344_1887	fliB	35.5	24.7	1.4372	0.1575
SL1344_0658	phoL	223	82.1	2.7162	0.434
SL1344_2273	pmrD	1077.3	329.7	3.2675	0.5142
SL1344_0627	mrdB	22.8	19.5	1.1692	0.0679
SL1344_3462	igaA	72.4	56.5	1.2814	0.1077
SL1344_3419	yheO	114.1	85.4	1.3361	0.1258
SL1344_0937	asnCa	80.5	87.9	0.9158	-0.0382
SL1344_0192	fhuA	49.5	7.9	6.2658	0.797
SL1344_2649	rps16	779.5	143.9	5.417	0.7338
SL1344_1196		52	6.7	7.7612	0.8899
SL1344_2267	pmrF	49	17.7	2.7684	0.4422
SL1344_0225	yaeT	181.2	134.8	1.3442	0.1285
SL1344_3184	dnaG	113.2	46.3	2.4449	0.3883
SL1344_1646	topA	87.8	32.6	2.6933	0.4303
SL1344_1633	rnb	36.9	24.8	1.4879	0.1726
SL1344_0436	cyoB	185.5	21.4	8.6682	0.9379
SL1344_3667	secB	451.1	348.6	1.294	0.1119
SL1344_2203	yejK	20.3	19.8	1.0253	0.0108
SL1344_3410	bfr	420.1	228.3	1.8401	0.2648
SL1344_0434	cyoD	333.1	49.7	6.7022	0.8262
SL1344_0172	yadF	468.8	89.8	5.2205	0.7177
SL1344_2483	yfgM	76	51	1.4902	0.1732
SL1344_2245	ubiG	96.5	49.1	1.9654	0.2934
SL1344_1577	ldhA	108.1	42.7	2.5316	0.4034
SL1344_4311	yjfN	20.7	249.9	0.0828	-1.0818
SL1344_3501	glgP	13.9	97.9	0.142	-0.8478
SL1344_3382	pez	994	310.2	3.2044	0.5057
SL1344_1755	sdaA	20.7	46.6	0.4442	-0.3524
SL1344_3938	yigZ	93.4	67.1	1.392	0.1436
SL1344_3862	ilvM	137.1	43.2	3.1736	0.5016
SL1344_1571	hrpA	24.5	20	1.225	0.0881
SL1344_3947	polA	31.9	48.6	0.6564	-0.1828
SL1344_2289	nuoJ	45.1	53.2	0.8477	-0.0717
SL1344_4325	priB	496.6	145.9	3.4037	0.532
SL1344_0191	mrcB	45.3	38.5	1.1766	0.0706
SL1344_4489	osmY	2451.9	88.2	27.7993	1.444
SL1344_2937	relA	34.8	79.9	0.4355	-0.361
SL1344_0656	ybeX	129.7	67.7	1.9158	0.2824
SL1344_2963	ygdI	1685.4	492.6	3.4214	0.5342
SL1344_1396	ydgA	44.6	17.8	2.5056	0.3989
SL1344_0119	mraZ	405	288.2	1.4053	0.1478
SL1344_1220	yeaG	25.4	28.6	0.8881	-0.0515
SL1344_0444	lon	358	118.8	3.0135	0.4791
SL1344_3391	rl18	1798.9	388.7	4.628	0.6654
SL1344_1226	yeaA	290.2	98.7	2.9402	0.4684
SL1344_0788	ybhO	10.2	2.6	3.9231	0.5936
SL1344_3277	ispB	116.2	92.8	1.2522	0.0977
SL1344_0170	gcd	60.9	7.5	8.12	0.9096
SL1344_4181	qor	31.5	36.7	0.8583	-0.0664
SL1344_1422		162.3	2.2	73.7727	1.8679
SL1344_2482	yfgL	154.9	114.4	1.354	0.1316
SL1344_1863	otsB	130.6	5.9	22.1356	1.3451
SL1344_2296	nuoB	156.3	117.5	1.3302	0.1239
SL1344_3463	yrfG	118.2	43.3	2.7298	0.4361

SL1344_1876	uvrC	26.7	44.9	0.5947	-0.2257
SL1344_1253	katE	23.5	20.3	1.1576	0.0636
SL1344_1982		15	2.6	5.7692	0.7611
SL1344_1753	pabB	32	24.7	1.2955	0.1125
SL1344_0443	clpX	611	318.7	1.9172	0.2827
SL1344_2068	rfbG	311.5	129.1	2.4129	0.3825
SL1344_3587	yhjS	88	49.2	1.7886	0.2525
SL1344_0808	ompX	2073.6	547.7	3.786	0.5782
SL1344_0997	pepN	28.6	59.2	0.4831	-0.316
SL1344_3676	rfaD	167.1	137.4	1.2162	0.085
SL1344_3668	grxC	362.5	285.4	1.2701	0.1039
SL1344_1394	ydgJ	22.8	33.7	0.6766	-0.1697
SL1344_4312	yjfO	100.1	999.5	0.1002	-0.9993
SL1344_2232	eco	214.5	37.2	5.7661	0.7609
SL1344_2393	cysK	184.8	64	2.8875	0.4605
SL1344_0468	acrB	69.6	42.3	1.6454	0.2163
SL1344_3199	yqjA	69.7	73.6	0.947	-0.0236
SL1344_P1_0072		NA	NA	NA	NA
SL1344_2190	yeiU	118.6	24.2	4.9008	0.6903
SL1344_4060	ptsA	18	23.3	0.7725	-0.1121
SL1344_4005	yiiM	22.7	52.7	0.4307	-0.3658
SL1344_3614	yiaG	218.4	180	1.2133	0.084
SL1344_2048	hisG	18.6	4.9	3.7959	0.5793
SL1344_2381	gltX	42.5	41.6	1.0216	0.0093
SL1344_3112		164	134.2	1.2221	0.0871
SL1344_3696	dfp	75.5	56.8	1.3292	0.1236
SL1344_3191	oat	7.3	4.6	1.587	0.2006
SL1344_2932	pyrG	177.8	188.8	0.9417	-0.0261
SL1344_1737		1090	871.1	1.2513	0.0974
SL1344_4040	hslU	485.8	74.1	6.556	0.8166
SL1344_1131	fabD	109.8	52.3	2.0994	0.3221
SL1344_0659	miaB	44.5	25.7	1.7315	0.2384
SL1344_0714	sdhC	413.3	62.6	6.6022	0.8197
SL1344_0384	yaiA	564.5	104.1	5.4227	0.7342
SL1344_1489		5.3	2.4	2.2083	0.3441
SL1344_1683	galU	112.1	70.3	1.5946	0.2027
SL1344_3196	ygjR	7.7	47.7	0.1614	-0.792
SL1344_0469	acrA	97.6	52.8	1.8485	0.2668
SL1344_1057	yccJ	153.2	770	0.199	-0.7012
SL1344_3707	rpoZ	450.1	157.2	2.8632	0.4569
SL1344_2437	tktB	22.5	9.5	2.3684	0.3745
SL1344_4041	hslV	670.4	57.5	11.6591	1.0667
SL1344_0467	ybaJ	901.6	399.2	2.2585	0.3538
SL1344_2105	yegO	3	3.4	0.8824	-0.0544
SL1344_4326	rpsR	562.8	86.4	6.5139	0.8138
SL1344_0529	ppiB	575.3	169.3	3.3981	0.5312
SL1344_1733	fadR	80.9	141.1	0.5734	-0.2416
SL1344_3453	aroB	74	72.5	1.0207	0.0089
SL1344_0616	pagP	593.3	74.7	7.9424	0.9
SL1344_2594		926.6	729	1.2711	0.1042
SL1344_2391	zipA	381.2	307.6	1.2393	0.0932
SL1344_3276	rplU	1855.4	443.3	4.1854	0.6217
SL1344_1062	putA	9.9	3.5	2.8286	0.4516
SL1344_2515	csiE	16.3	43.4	0.3756	-0.4253
SL1344_4029		15.6	21.8	0.7156	-0.1453
SL1344_1377	slyB	628.1	272	2.3092	0.3635
SL1344_1372	sodCb	39.8	58	0.6862	-0.1635
SL1344_P2_0012	cib	NA	NA	NA	NA
SL1344_3053	yggG	29	28.8	1.0069	0.003
SL1344_3161	ygiB	152.4	107.3	1.4203	0.1524
SL1344_2241	gyrA	127.6	60.4	2.1126	0.3248

SL1344_4190	uvrA	35.2	18.5	1.9027	0.2794
SL1344_1820	zwf	34.4	38.3	0.8982	-0.0466
SL1344_1058	wrbA	101.3	216.8	0.4673	-0.3304
SL1344_4004	sodA	468.1	123.2	3.7995	0.5797
SL1344_3554	pitA	19.8	43.4	0.4562	-0.3408
SL1344_0480	htpG	1057.3	109.6	9.6469	0.9844
SL1344_3454	aroK	171.1	165.1	1.0363	0.0155
SL1344_0160		783.7	104.5	7.4995	0.875
SL1344_3353	accC	94	71.7	1.311	0.1176
SL1344_1490		8.5	2.3	3.6957	0.5677
SL1344_3407	rplC	1142	208	5.4904	0.7396
SL1344_2615	yfiQ	14.4	53.4	0.2697	-0.5692
SL1344_2292	nuoG	86.1	104.3	0.8255	-0.0833
SL1344_2628	aroF	75.4	26.9	2.803	0.4476
SL1344_1384	tppB	49.2	26.6	1.8496	0.2671
SL1344_3346	mreB	76.9	81.5	0.9436	-0.0252
SL1344_3255	pnp	80.4	79.7	1.0088	0.0038
SL1344_3392	rplF	1327.5	290.3	4.5729	0.6602
SL1344_0617	cspE	2444.1	5682.8	0.4301	-0.3664
SL1344_0879	ybjX	306.5	63.5	4.8268	0.6837
SL1344_4496	deoB	150.1	89.9	1.6696	0.2226
SL1344_1905	fliM	32.4	30.4	1.0658	0.0277
SL1344_3596	dppA	40.8	98.4	0.4146	-0.3823
SL1344_0715	sdhD	299.7	64.8	4.625	0.6651
SL1344_0446	cypD	63.7	56.4	1.1294	0.0529
SL1344_4463	cstAb	41.4	83.5	0.4958	-0.3047
SL1344_0312	pepD	46.8	127.8	0.3662	-0.4363
SL1344_2266	yfbE	43.1	12.9	3.3411	0.5239
SL1344_2785	ygaC	289.8	11.8	24.5593	1.3902
SL1344_0235	ldcC	23.7	25.6	0.9258	-0.0335
SL1344_4177	zur	60.4	57.5	1.0504	0.0214
SL1344_3473	feoB	24	14.7	1.6327	0.2129
SL1344_0152	aceE	283.1	41.1	6.8881	0.8381
SL1344_3383	rpsD	1077.6	414.4	2.6004	0.415
SL1344_4260	dipZ	27.4	11.3	2.4248	0.3847
SL1344_1002	uup	22.9	13.3	1.7218	0.236
SL1344_4044	priA	42.9	26.9	1.5948	0.2027
SL1344_4267	groEL	1361.7	282.1	4.827	0.6837
SL1344_3015	idi	88.9	47.1	1.8875	0.2759
SL1344_3720	rhuM	28.2	24.7	1.1417	0.0576
SL1344_3385	rpsM	1974.7	530.9	3.7195	0.5705
SL1344_0802	ybiO	14.7	3.5	4.2	0.6232
SL1344_2237	ompC	739	835	0.885	-0.053
SL1344_4030	tpiA	247.6	647.8	0.3822	-0.4177
SL1344_3806	rpmH	7727.5	849.3	9.0987	0.959
SL1344_3876	rho	276.8	158.2	1.7497	0.243
SL1344_4405	valS	57.7	32.3	1.7864	0.252
SL1344_4157	pgi	69.2	105.9	0.6534	-0.1848
SL1344_4069	ppc	43.4	66.7	0.6507	-0.1866
SL1344_0918	rpsA	698.5	237.6	2.9398	0.4683
SL1344_2067	rfbH	421.7	130.1	3.2414	0.5107
SL1344_3984	fdoG	86	65.4	1.315	0.1189
SL1344_4090	rplJ	1267.1	576.3	2.1987	0.3422
SL1344_3408	rpsJ	1883.8	302.8	6.2213	0.7939
SL1344_1680	adh	21.2	435.8	0.0486	-1.313
SL1344_2395	ptsI	170.9	386.3	0.4424	-0.3542
SL1344_3412	tufA	441.9	129.7	3.4071	0.5324
SL1344_0005	yaaA	62.3	50.7	1.2288	0.0895
SL1344_3675	kbl	52.8	143.3	2.7140	0.4336
SL1344_0910	pflB	47.5	439.7	9.2568	0.9665

SL1344_1429	dmsA1	4.1	15.1	3.6829	0.5662
SL1344_1503	ompD	1394.7	1950.4	1.3984	0.1456
SL1344_4348	fbp	126.9	69.3	0.5461	-0.2627
SL1344_3310	nanT	4	234.3	58.5750	1.7677
SL1344_0596	ahpC	341.4	166.5	0.4877	-0.3118
SL1344_0136	secA	66.8	73.7	1.1033	0.0427
SL1344_3850	rbsC	13.8	22.9	1.6594	0.2200
SL1344_4169	malM	17.3	176	10.1734	1.0075
SL1344_4405	valS	57.7	32.3	0.5598	-0.2520
SL1344_4166	malE	16.9	175.1	10.3609	1.0154
SL1344_4168	lamB	15	179.9	11.9933	1.0789
SL1344_4457	hsdR	11.1	75.1	6.7658	0.8303
SL1344_0752	galT	9	55.6	6.1778	0.7908
SL1344_3265	glmM	28.3	55.5	1.9611	0.2925
SL1344_3848	rbsD	46.3	29.6	0.6393	-0.1943
SL1344_4263	aspA	188.7	2469.5	13.0869	1.1168
SL1344_3081	ansB	21	1041.3	49.5857	1.6954
SL1344_P2_0049	nikB	NA	NA	NA	NA
SL1344_1025	yccA	84.5	35	0.4142	-0.3828
SL1344_3450	rpe	18.1	20.2	1.1160	0.0477
SL1344_4108	yjaG	163.2	100.5	0.6158	-0.2106
SL1344_3217	tdcA	20.9	1556.3	74.4641	1.8719
SL1344_0666	nagB	35.1	86.1	2.4530	0.3897
SL1344_0487	ushA	44.6	56.7	1.2713	0.1042
SL1344_1592	ydaA	110	243.6	2.2145	0.3453
SL1344_1224A		9.4	61	6.4894	0.8122
SL1344_3851	rbsB	71.1	368.1	5.1772	0.7141
SL1344_2306	ackA	149.2	143.4	0.9611	-0.0172
SL1344_4280	frdA	10	460.4	46.0400	1.6631
SL1344_0309	dbh	37.6	15	0.3989	-0.3991
SL1344_3256	rpsO	4654.6	518	0.1113	-0.9536
SL1344_0751	galK	24.2	80.1	3.3099	0.5198
SL1344_3110	uxuB	6.4	45.6	7.1250	0.8528
SL1344_2458	yfgD	41.2	44	1.0680	0.0286
SL1344_1822	pykA	23.3	202.3	8.6824	0.9386
SL1344_2041	phsA	3.4	146	42.9412	1.6329
SL1344_1056	agp	9.5	106.6	11.2211	1.0500
SL1344_2950	sdaC	43.9	292	6.6515	0.8229
SL1344_4221	SC4B5,11c	7	69.2	9.8857	0.9950
SL1344_4186	yjbQ	65.8	64	0.9726	-0.0120
SL1344_2003	cbiH	1.2	81.2	67.6667	1.8304
SL1344_2255	glpC	11.6	579.4	49.9483	1.6985
SL1344_1582	ynaF	142.2	1095.5	7.7039	0.8867
SL1344_2347	fabB	129.9	316.5	2.4365	0.3868
SL1344_0557	ybdG	89.2	22.2	0.2489	-0.6040
SL1344_2521	cadA	7.7	782.2	101.5844	2.0068
SL1344_2098	dcd	29.6	22.6	0.7635	-0.1172
SL1344_P2_0006	yadA	NA	NA	NA	NA
SL1344_4164	malF	3.1	18.6	6.0000	0.7782
SL1344_3215	tdcC	7.8	1264.6	162.1282	2.2099
SL1344_4361	iolC	4.6	2.9	0.6304	-0.2004
SL1344_4079	btuB	44.4	55.2	1.2432	0.0946
SL1344_2663		0.7	1	1.4286	0.1549
SL1344_2254	glpB	6	435.9	72.6500	1.8612
SL1344_2253	glpA	7.4	526.2	71.1081	1.8519
SL1344_3852	rbsK	17.4	67.4	3.8736	0.5881
SL1344_4279	frdB	16.2	570.7	35.2284	1.5469
SL1344_0162	kdgT	48.2	90.8	1.8838	0.2750
SL1344_0683	speF	1.9	626.6	329.7895	2.5182
SL1344_3309	nanE	3.2	165.7	51.7813	1.7142
SL1344_1073	ycdX	61	34	0.5574	-0.2539

SL1344_2644		31.4	164.4	5.2357	0.7190
SL1344_2314		11.5	84.7	7.3652	0.8672
SL1344_0625	dacA	87.3	39.7	0.4548	-0.3422
SL1344_2307	pta	68	69.6	1.0235	0.0101
SL1344_1368	nemA	31.3	27.7	0.8850	-0.0531
SL1344_3652	mtlR	11.8	42.7	3.6186	0.5585
SL1344_2014	pduA	0	296.5	inf	inf
SL1344_2311		6.5	557.3	85.7385	1.9332
SL1344_4278	frdC	16.4	475.7	29.0061	1.4625
SL1344_0902	dmsA	1.7	241.1	141.8235	2.1517
SL1344_3859	yifE	293.6	836.1	2.8478	0.4545
SL1344_1428	dmsA2	0.8	7.2	9.0000	0.9542
SL1344_2831	hycC	1.1	2.1	1.9091	0.2808
SL1344_2011	cbiA	1.7	105.9	62.2941	1.7944
SL1344_3212	tdcG	5.5	571.4	103.8909	2.0166
SL1344_2310		4.1	406.2	99.0732	1.9960
SL1344_3299	yhbL	101.6	87.9	0.8652	-0.0629
SL1344_2312		8.7	444.9	51.1379	1.7087
SL1344_4382	nrdD	4.8	270.2	56.2917	1.7504
SL1344_2016	dhaB	0	156.3	inf	inf
SL1344_0019		3.1	2.7	0.8710	-0.0600
SL1344_3222	garL	6.9	221.2	32.0580	1.5059
SL1344_4237	fumB	6.9	183.7	26.6232	1.4253
SL1344_3208	yhaK	12.9	24.6	1.9070	0.2803
SL1344_4238	dcuB	3.9	249	63.8462	1.8051
SL1344_3124	hypO	13.1	162.1	12.3740	1.0925
SL1344_0750	galM	30.1	133.5	4.4352	0.6469
SL1344_3213	tdcE	4.5	792.4	176.0889	2.2457
SL1344_3120	hybD	11.7	200.4	17.1282	1.2337
SL1344_3674	tdh	98.2	280.9	2.8605	0.4564
SL1344_1761		22.4	489.5	21.8527	1.3395
SL1344_2260	yfaW	5.5	12.9	2.3455	0.3702
SL1344_3216	tdcB	16.7	1611.5	96.4970	1.9845
SL1344_3214	tdcD	9.7	974.6	100.4742	2.0021
SL1344_0805	glnH	154.2	86.6	0.5616	-0.2506
SL1344_2529	yfhD	12.6	8.8	0.6984	-0.1559
SL1344_0682	potE	1.7	751.9	442.2941	2.6457
SL1344_0118	fruR	38.3	128.1	3.3446	0.5244

Appendix 4: list of differentially expressed genes in single-cell bacteria

Differentially expressed genes for single bacteria and comparison of our study with reference bulk RNA-seq dataset from Kröger et al. 2013. Gene ID and gene name of differentially expressed genes of single bacteria in the three growth conditions are listed in column 1 and 2, respectively (color code in Column 2: blue: 'Late Stationary Phase' ; green: 'Salt (NaCl) shock'; red: 'Anaerobic shock'). For each gene differentially expressed between Salt (NaCl) and Anaerobic shocks, column 3 and 4 recapitulate the gene expression extracted from Kröger et al. under Salt (NaCl) shock and Anaerobic shock, respectively. Missing values in the benchmark dataset (Kröger et al. 2013) are indicated with 'NA'. The ratio and log10 transformed ratio of (NaCl/Anaerobic) of the latter expression values from bulk-RNA-seq from Kröger et al. 2013 are calculated in columns 5 and 6, respectively.

Gene ID	Gene name	Expression values Salt (NaCl) Shock (Kröger et al) (TPM)	Expression values Anaerobic Shock (Kröger et al.) (TPM)	Ratio Salt(NaCl) /Anaerobic	Ratio Salt(NaCl) /Anaerobic (Log10)
SL1344_3301	yhcC				
SL1344_3356	prmA				
SL1344_2860	iacP				
SL1344_2847	hilC				
SL1344_1508	narZ				
SL1344_2650	ffh				
EBG00000241440	lsrM				
SL1344_4123	metH				
SL1344_3746					
SL1344_1948					
SL1344_2043	sopA				
SL1344_3106					
SL1344_2795	proW				
SL1344_4520	sthB				
SL1344_4217	nrfE				
SL1344_3709	spoU	44.2	49.6	0.8911	-0.0501
SL1344_3574	yhjE	13.5	6.5	2.0769	0.3174
SL1344_2963	ygdI	1685.4	492.6	3.4214	0.5342
SL1344_0002	thrA	34.6	13.2	2.6212	0.4185
SL1344_3843	asnCb	102.5	30.8	3.3279	0.5222
SL1344_3390	rpsE	1034.9	199.1	5.1979	0.7158
SL1344_0172	yadF	468.8	89.8	5.2205	0.7177
SL1344_4004	sodA	468.1	123.2	3.7995	0.5797
SL1344_3408	rpsJ	1883.8	302.8	6.2213	0.7939
SL1344_4093	rpoC	91	71	1.2817	0.1078
EBG00001133793	t44	NA	NA	NA	NA

SL1344_0718	sucA	73.2	75.1	0.9747	-0.0111
SL1344_3894	cyaA	155.3	92.2	1.6844	0.2264
SL1344_3890	hemY	59.8	67.1	0.8912	-0.05
SL1344_0217	rpsB	1213	178.7	6.7879	0.8317
SL1344_4172	plsB	56.2	73.3	0.7667	-0.1154
SL1344_2295	nuoC	120.4	106.2	1.1337	0.0545
SL1344_1581	nifJ	13.4	6.8	1.9706	0.2946
SL1344_4348	fbp	126.9	69.3	1.8312	0.2627
SL1344_1170	purB	17.2	7	2.4571	0.3904
SL1344_4285	psd	53.5	28.4	1.8838	0.275
SL1344_1442	dcp	10.3	19.2	0.5365	-0.2705
SL1344_2280	elaB	254.5	147.8	1.7219	0.236
SL1344_0134	lpxC	528.9	347	1.5242	0.183
SL1344_3668	grxC	362.5	285.4	1.2701	0.1039
SL1344_3833	atpG	131.9	175.4	0.752	-0.1238
SL1344_1441	ydfG	135.3	124.2	1.0894	0.0372
SL1344_0417	ispA	47.9	66.3	0.7225	-0.1412
SL1344_3932	yigC	136.3	89.6	1.5212	0.1822
SL1344_2293	nuoF	71.1	93.1	0.7637	-0.1171
SL1344_3581	yhjL	35.1	18.9	1.8571	0.2688
SL1344_4177	zur	60.4	57.5	1.0504	0.0214
SL1344_0972		1.7	3.9	0.4359	-0.3606
SL1344_0437	cyoA	448.4	37.3	12.0214	1.08
SL1344_4301	rnr	59.1	82.1	0.7199	-0.1428
SL1344_3669	yibN	280.5	195.9	1.4319	0.1559
SL1344_1323	orf319	37.6	28.1	1.3381	0.1265
SL1344_3393	rpsH	1802.1	318.1	5.6652	0.7532
SL1344_1856	cheA	228	174.7	1.3051	0.1156
SL1344_1121	flgL	272.4	227.3	1.1984	0.0786
SL1344_3392	rplF	1327.5	290.3	4.5729	0.6602
SL1344_2068	rfbG	311.5	129.1	2.4129	0.3825
SL1344_3401	rpsC	1473	206.8	7.1228	0.8527
SL1344_2267	pmrF	49	17.7	2.7684	0.4422
SL1344_4092	rpoB	94	71.9	1.3074	0.1164
SL1344_0012	dnaK	1100.5	157.9	6.9696	0.8432
SL1344_4343	ytfN	77.9	58.3	1.3362	0.1259
SL1344_0441	tig	328.9	115	2.86	0.4564
SL1344_3160	tolC	130.9	96.8	1.3523	0.1311
SL1344_3350		11	15.2	0.7237	-0.1405
SL1344_0154	lpdA	471.4	88.5	5.3266	0.7264
SL1344_0566	nfnB	75.8	47.8	1.5858	0.2002
SL1344_3196	ygjR	7.7	47.7	0.1614	-0.792
SL1344_3317	rplM	1220.7	212.6	5.7418	0.759
SL1344_0160		783.7	104.5	7.4995	0.875
SL1344_0152	aceE	283.1	41.1	6.8881	0.8381
SL1344_3405	rplW	1435.2	242.6	5.9159	0.772
SL1344_P2_0096	pilK	NA	NA	NA	NA
SL1344_2563	gipA	370.8	525.9	0.7051	-0.1518
SL1344_1057	yccJ	153.2	770	0.199	-0.7012
SL1344_2649	rps16	779.5	143.9	5.417	0.7338
SL1344_P1_0072		NA	NA	NA	NA
SL1344_P1_0055	parA	NA	NA	NA	NA
EBG00001133868	StyR-44	NA	NA	NA	NA
SL1344_4167	malK	5.2	63.4	0.082	-1.0861
EBG00000241426	STnc700	NA	NA	NA	NA
SL1344_3394	rpsN	1616.8	309	5.2324	0.7187
SL1344_2395	ptsI	170.9	386.3	0.4424	-0.3542
SL1344_3891	hemX	103.7	89.3	1.1613	0.0649
SL1344_3596	dppA	40.8	98.4	0.4146	-0.3823
SL1344_3069	yggJ	38.5	45.7	0.8425	-0.0745
SL1344_3505	glgB	38.5	45.7	0.8425	-0.0745

SL1344_1240	astD	2.2	1.7	1.2941	0.112
SL1344_3413	fusA	767.4	204.3	3.7562	0.5748
SL1344_3395	rplE	3185.3	609.7	5.2244	0.718
SL1344_0214	dapD	186.5	190.1	0.9811	-0.0083
SL1344_3045	pgk	113.8	324.6	0.3506	-0.4552
SL1344_4085	tufB	388.3	310.6	1.2502	0.097
SL1344_4347	ppa	964	526	1.8327	0.2631
SL1344_4324	rpsF	330.1	84.7	3.8973	0.5908
SL1344_3694	rpmB	3495.1	595.6	5.8682	0.7685
SL1344_1680	adh	21.2	435.8	0.0486	-1.313
SL1344_2464	ppk	70.5	56.8	1.2412	0.0938
SL1344_1168	phoQ	143.1	57	2.5105	0.3998
SL1344_1732	ycgB	42.3	28.5	1.4842	0.1715
SL1344_0802	ybiO	14.7	3.5	4.2	0.6232
SL1344_0480	htpG	1057.3	109.6	9.6469	0.9844
SL1344_3984	fdoG	86	65.4	1.315	0.1189
SL1344_3926	aarF	101.2	86.7	1.1672	0.0672
SL1344_3387	prlA	1823	486.1	3.7503	0.5741
SL1344_0868	ybjP	59.8	45.6	1.3114	0.1177
SL1344_3391	rl18	1798.9	388.7	4.628	0.6654
SL1344_1234	gdhA	48.2	39.4	1.2234	0.0876
SL1344_0866	artI	32.1	46	0.6978	-0.1563
SL1344_1245	nadE	139.7	81.4	1.7162	0.2346
SL1344_2183	fruB	2.3	3.5	0.6571	-0.1823
SL1344_1253	katE	23.5	20.3	1.1576	0.0636
SL1344_2698		511.9	620.4	0.8251	-0.0835
SL1344_0653	gltJ	22.3	3.9	5.7179	0.7572
SL1344_0712	gltA	638.9	108.2	5.9048	0.7712
SL1344_1813	opdB	21.3	17.8	1.1966	0.078
SL1344_0808	ompX	2073.6	547.7	3.786	0.5782
SL1344_P1_0084	repA2	NA	NA	NA	NA
SL1344_3876	rho	276.8	158.2	1.7497	0.243
SL1344_0617	cspE	2444.1	5682.8	0.4301	-0.3664
SL1344_0918	rpsA	698.5	237.6	2.9398	0.4683
SL1344_0459	ybaY	266.1	44.8	5.9397	0.7738
SL1344_3493	glpD	1131.1	698.3	1.6198	0.2095
SL1344_2300	yfbQ	9.4	16.4	0.5732	-0.2417
SL1344_3478	yhgI	435.5	43.6	9.9885	0.9995
SL1344_0929	mukF	43.4	29.3	1.4812	0.1706
SL1344_4267	groEL	1361.7	282.1	4.827	0.6837
SL1344_2268	yfbG	25.4	12.5	2.032	0.3079
SL1344_1328	ssaC	7.1	0.2	35.5	1.5502
SL1344_4339	ytfK	6191.4	1596.9	3.8771	0.5885
SL1344_2436	tal	38.3	8.4	4.5595	0.6589
SL1344_4526	yjjY	8.5	26.3	0.3232	-0.4905
SL1344_1267	thrS	222.2	198.1	1.1217	0.0499
SL1344_3312	yhcK	43.5	56.2	0.774	-0.1112
SL1344_0792	ybhS	69.5	32.2	2.1584	0.3341
SL1344_0159	acnB	190.3	69.2	2.75	0.4393
SL1344_0336		99.8	25.4	3.9291	0.5943
SL1344_3824	pstS	481.4	12.6	38.2063	1.5821
SL1344_4069	ppc	43.4	66.7	0.6507	-0.1866
SL1344_0735	ybgR	18.5	22.8	0.8114	-0.0908
SL1344_3412	tufA	441.9	129.7	3.4071	0.5324
SL1344_2364	pgtA	5.2	15.1	0.3444	-0.463
SL1344_0005	yaaA	62.3	50.7	1.2288	0.0895
SL1344_0149		4.6	4.3	1.0698	0.0293
SL1344_2641		1.1	0.9	1.2222	0.0872
SL1344_3675	kbl	52.8	143.3	0.3685	-0.4336
EBG00001133739	tmRNA	NA	NA	NA	NA

SL1344_1888	fliC	2532.9	3334	1.3163	0.1193
SL1344_3556	uspA	1048.7	1147	1.0937	0.0389
SL1344_2394	ptsH	200.1	399.4	1.9960	0.3002
SL1344_2756	fljB	29.9	34.1	1.1405	0.0571
SL1344_4168	lamB	15	179.9	11.9933	1.0789
SL1344_2237	ompC	739	835	1.1299	0.0530
SL1344_2785	ygaC	289.8	11.8	0.0407	-1.3902
SL1344_2246	nrdA	66.5	102.6	1.5429	0.1883
SL1344_0753	galE	23.8	73.4	3.0840	0.4891
SL1344_3875	trxA	1663.3	477	0.2868	-0.5425
SL1344_3850	rbsC	13.8	22.9	1.6594	0.2200
SL1344_3223	garD	1.4	42.7	30.5000	1.4843
SL1344_3625	xylB	3.9	11.8	3.0256	0.4808
SL1344_2283		230.3	527.6	2.2909	0.3600
SL1344_2517	glyA	72.3	156.7	2.1674	0.3359
SL1344_0997	pepN	28.6	59.2	2.0699	0.3160
SL1344_3023	ygfY	99.6	126.2	1.2671	0.1028
SL1344_3403	rpsS	2990.4	1010.8	0.3380	-0.4711
SL1344_2521	cadA	7.7	782.2	101.5844	2.0068
SL1344_0752	galT	9	55.6	6.1778	0.7908
SL1344_1113	flgD	52.4	27.8	0.5305	-0.2753
SL1344_2252	glpT	63.7	535.1	8.4003	0.9243
SL1344_1854	cheM	264.7	267.8	1.0117	0.0051
SL1344_2105	yegO	3	3.4	1.1333	0.0544
SL1344_0654	gltI	59.4	12.5	0.2104	-0.6769
SL1344_4090	rplJ	1267.1	576.3	0.4548	-0.3422
SL1344_1707	ipk	69.2	64.7	0.9350	-0.0292
SL1344_0209	dgt	34.5	21.8	0.6319	-0.1994
SL1344_4221	SC4B5,11c	7	69.2	9.8857	0.9950
SL1344_2310		4.1	406.2	99.0732	1.9960
SL1344_0529	ppiB	575.3	169.3	0.2943	-0.5312
SL1344_0751	galK	24.2	80.1	3.3099	0.5198
SL1344_4237	fumB	6.9	183.7	26.6232	1.4253
SL1344_2931	eno	215.7	300.7	1.3941	0.1443
SL1344_3467	pckA	77.6	476.5	6.1405	0.7882
SL1344_3169	yqiC	167.6	163.2	0.9737	-0.0116
SL1344_4238	dcuB	3.9	249	63.8462	1.8051
SL1344_0193	fhuC	11.1	3.6	0.3243	-0.4890
SL1344_4263	aspA	188.7	2469.5	13.0869	1.1168
SL1344_3031	gcvT	53.4	174.7	3.2715	0.5148
SL1344_3112		164	134.2	0.8183	-0.0871
SL1344_1902	fliJ	13.4	23.1	1.7239	0.2365
SL1344_3066	galP	35.8	170.6	4.7654	0.6781
SL1344_4280	frdA	10	460.4	46.0400	1.6631
SL1344_1114	flgE	72.7	28.1	0.3865	-0.4128
SL1344_4262	dcuA	23.7	353	14.8945	1.1730
SL1344_3577	kdgK	39.2	120.3	3.0689	0.4870
SL1344_1537		33.1	125.7	3.7976	0.5795
SL1344_1708	prs	50.9	48.2	0.9470	-0.0237
SL1344_2311		6.5	557.3	85.7385	1.9332
SL1344_0683	speF	1.9	626.6	329.7895	2.5182
SL1344_0612	citC	0.3	1.9	6.3333	0.8016
SL1344_2253	glpA	7.4	526.2	71.1081	1.8519
SL1344_3851	rbsB	71.1	368.1	5.1772	0.7141
SL1344_3905	uvrD	41.8	31.6	0.7560	-0.1215
SL1344_3215	tdcC	7.8	1264.6	162.1282	2.2099
SL1344_2041	phsA	3.4	146	42.9412	1.6329
SL1344_3124	hypO	13.1	162.1	12.3740	1.0925
SL1344_3121	hybC	14.3	131	9.1608	0.9619
SL1344_1503	ompD	1394.7	1950.4	1.3984	0.1456
SL1344_0610	citE	0	1.1	inf	inf

SL1344_3214	tdcD	9.7	974.6	100.4742	2.0021

Curriculum vitae

Personal details

Born: 16th December 1987

Nationality: Iranian

Education

PhD candidate| Helmholtz Institute for RNA Based Infection Research (HIRI) | Würzburg, Germany | Since 2017

- **Thesis title:** Implementation and application of bioinformatics methods to analyze and visualize single-cell RNA-seq data
- **Mentor:** Dr. Antoine-Emmanuel Saliba

Master's Degree | University of Tehran (the most competitive university in Iran) | Tehran, Iran | 2011 - 2014

- **Thesis title:** Construction of *rad2* antisense RNA coding vector and analysis of the effect of *rad2* knock down on DNA repair mechanisms of *Halobacterium salinarum*.
- **Mentor:** Dr. Mohammad Ali Amoozegar

Bachelor's Degree | University of Zabol | Zabol, Iran | 2008 - 2011

List of publications

- Imdahl, Fabian, **Ehsan Vafadarnejad**, Christina Homberger, Antoine-Emmanuel Saliba, and Jörg Vogel. "Single-cell RNA-sequencing reports growth-condition-specific global transcriptomes of individual bacteria." Nature Microbiology 5, no. 10 (2020): 1202-1206.
- **Ehsan Vafadarnejad**, Giuseppe Rizzo, Laura Krampert, Panagiota Arampatzi, Anahi-Paula Arias-Loza, Yara Nazzal, Anna Rizakou et al. "Dynamics of cardiac neutrophil diversity in murine myocardial infarction." Circulation research 127, no. 9 (2020): e232-e249.
- Cochain, Clément, **Ehsan Vafadarnejad**, Panagiota Arampatzi, Jaroslav Pelisek, Holger Winkels, Klaus Ley, Dennis Wolf, Antoine-Emmanuel Saliba, and Alma Zernecke. "Single-cell RNA-seq reveals the transcriptional landscape and heterogeneity of aortic macrophages in murine atherosclerosis." Circulation research 122, no. 12 (2018): 1661-1674.

- Fuhr, Viktoria, **Ehsan Vafadarnejad**, Oliver Dietrich, Panagiota Arampatzi, Angela Riedel, Antoine-Emmanuel Saliba, Andreas Rosenwald, and Hilka Rauert-Wunderlich. "Time-Resolved scRNA-Seq Tracks the Adaptation of a Sensitive MCL Cell Line to Ibrutinib Treatment." International journal of molecular sciences 22, no. 5 (2021): 2276.
- Tegtmeyer, B., Vieyres, G., Todt, D., Lauber, C., Ginkel, C., Engelmann, M., Herrmann, M., Pfaller, C.K., Vondran, F.W., Broering, R. and **Ehsan Vafadarnejad**, 2021. Initial Hepatitis C Virus Infection of Adult Hepatocytes Triggers a Temporally Structured Transcriptional Program Containing Diverse Pro-and Antiviral Elements. Journal of Virology, 95(10), pp.e00245-21.
- Pezoldt, Joern, Maria Pasztoi, Mangge Zou, Carolin Wiechers, Michael Beckstette, Guilhem R. Thierry, **Ehsan Vafadarnejad** et al. "Neonatally imprinted stromal cell subsets induce tolerogenic dendritic cells in mesenteric lymph nodes." Nature communications 9, no. 1 (2018): 1-14.
- Winkels, H., Ehinger, E., Vassallo, M., Buscher, K., Dinh, H.Q., Kobiyama, K., Hamers, A.A., Cochain, C., **Ehsan Vafadarnejad**, Saliba, A.E. and Zernecke, A., 2018. Atlas of the immune cell repertoire in mouse atherosclerosis defined by single-cell RNA-sequencing and mass cytometry. Circulation research, 122(12), pp.1675-1688.

References

Ackermann, Martin. 2015. “A Functional Perspective on Phenotypic Heterogeneity in Microorganisms.” *Nature Reviews. Microbiology* 13 (8): 497–508.

Adamopoulos, Iannis E., and Elizabeth D. Mellins. 2015. “Alternative Pathways of Osteoclastogenesis in Inflammatory Arthritis.” *Nature Reviews. Rheumatology* 11 (3): 189.

Aibar, Sara, Carmen Bravo González-Blas, Thomas Moerman, Vân Anh Huynh-Thu, Hana Imrichova, Gert Hulselmans, Florian Rambow, et al. 2017. “SCENIC: Single-Cell Regulatory Network Inference and Clustering.” *Nature Methods* 14 (11): 1083–86.

Aiello, R. J., P. A. Bourassa, S. Lindsey, W. Weng, E. Natoli, B. J. Rollins, and P. M. Milos. 1999. “Monocyte Chemoattractant Protein-1 Accelerates Atherosclerosis in Apolipoprotein E-Deficient Mice.” *Arteriosclerosis, Thrombosis, and Vascular Biology* 19 (6): 1518–25.

Andrews, Tallulah S., and Martin Hemberg. 2018. “Identifying Cell Populations with scRNASeq.” *Molecular Aspects of Medicine* 59 (February): 114–22.

Ardini-Poleske, Maryanne E., Robert F. Clark, Charles Ansong, James P. Carson, Richard A. Corley, Gail H. Deutsch, James S. Hagood, et al. 2017. “LungMAP: The Molecular Atlas of Lung Development Program.” *American Journal of Physiology. Lung Cellular and Molecular Physiology* 313 (5): L733–40.

Arruvito, Lourdes, Silvina Raiden, and Jorge Geffner. 2015. “Host Response to Respiratory Syncytial Virus Infection.” *Current Opinion in Infectious Diseases* 28 (3): 259–66.

Bakre, Abhijeet, Patricia Mitchell, Jonathan K. Coleman, Les P. Jones, Geraldine Saavedra, Michael Teng, S. Mark Tompkins, and Ralph A. Tripp. 2012. “Respiratory Syncytial Virus Modifies microRNAs Regulating Host Genes That Affect Virus Replication.” *The Journal of General Virology* 93 (Pt 11): 2346–56.

Battles, Michael B., and Jason S. McLellan. 2019. “Respiratory Syncytial Virus Entry and How to Block It.” *Nature Reviews. Microbiology* 17 (4): 233–45.

Becht, Etienne, Leland McInnes, John Healy, Charles-Antoine Dutertre, Immanuel W. H. Kwok, Lai Guan Ng, Florent Ginhoux, and Evan W. Newell. 2018. “Dimensionality Reduction for Visualizing Single-Cell Data Using UMAP.” *Nature Biotechnology* 37 (1): 38–44.

Behera, A. K., H. Matsuse, M. Kumar, X. Kong, R. F. Lockey, and S. S. Mohapatra. 2001. “Blocking Intercellular Adhesion Molecule-1 on Human Epithelial Cells Decreases Respiratory Syncytial Virus Infection.” *Biochemical and Biophysical Research Communications* 280 (1): 188–95.

Bhatia, Vinay K., Sheng Yun, Viola Leung, David C. Grimsditch, G. Martin Benson, Marina B. Botto, Joseph J. Boyle, and Dorian O. Haskard. 2007. “Complement C1q Reduces Early Atherosclerosis in Low-Density Lipoprotein Receptor-Deficient Mice.” *The American Journal of Pathology* 170 (1): 416–26.

Bitko, V., O. Shulyayeva, and B. Mazumder. 2007. “Nonstructural Proteins of Respiratory Syncytial Virus Suppress Premature Apoptosis by an NF-κB-Dependent, Interferon-Independent Mechanism and Facilitate Virus ….” *Journal of*. https://jvi.asm.org/content/81/4/1786.short.

Blattman, Sydney B., Wenyan Jiang, Panos Oikonomou, and Saeed Tavazoie. 2020. “Prokaryotic Single-Cell RNA Sequencing by in Situ Combinatorial Indexing.” *Nature Microbiology* 5 (10): 1192–1201.

Bódi, Zoltán, Zoltán Farkas, Dmitry Nevozhay, Dorottya Kalapis, Viktória Lázár, Bálint Csörgő,

Ákos Nyerges, et al. 2017. “Phenotypic Heterogeneity Promotes Adaptive Evolution.” *PLoS Biology* 15 (5): e2000644.
Brown, Armand O., Beth Mann, Geli Gao, Jane S. Hankins, Jessica Humann, Jonathan Giardina, Paola Faverio, et al. 2014. “Streptococcus Pneumoniae Translocates into the Myocardium and Forms Unique Microlesions That Disrupt Cardiac Function.” *PLoS Pathogens* 10 (9): e1004383.
Bueno, Susan M., Pablo A. González, Rodrigo Pacheco, Eduardo D. Leiva, Kelly M. Cautivo, Hugo E. Tobar, Jorge E. Mora, et al. 2008. “Host Immunity during RSV Pathogenesis.” *International Immunopharmacology* 8 (10): 1320–29.
Bumann, Dirk. 2015. “Heterogeneous Host-Pathogen Encounters: Act Locally, Think Globally.” *Cell Host & Microbe* 17 (1): 13–19.
Byon, Chang Hyun, Tieyan Han, Judy Wu, and Simon T. Hui. 2015. “Txnip Ablation Reduces Vascular Smooth Muscle Cell Inflammation and Ameliorates Atherosclerosis in Apolipoprotein E Knockout Mice.” *Atherosclerosis* 241 (2): 313–21.
Cadena, Anthony M., Sarah M. Fortune, and Joanne L. Flynn. 2017. “Heterogeneity in Tuberculosis.” *Nature Reviews. Immunology* 17 (11): 691–702.
Cardilo-Reis, Larissa, Sabrina Gruber, Sabine M. Schreier, Maik Drechsler, Nikolina Papac-Milicevic, Christian Weber, Oswald Wagner, Herbert Stangl, Oliver Soehnlein, and Christoph J. Binder. 2012. “Interleukin-13 Protects from Atherosclerosis and Modulates Plaque Composition by Skewing the Macrophage Phenotype.” *EMBO Molecular Medicine* 4 (10): 1072–86.
Childs, Bennett G., Darren J. Baker, Tobias Wijshake, Cheryl A. Conover, Judith Campisi, and Jan M. van Deursen. 2016. “Senescent Intimal Foam Cells Are Deleterious at All Stages of Atherosclerosis.” *Science* 354 (6311): 472–77.
Childs, Bennett G., Matej Durik, Darren J. Baker, and Jan M. van Deursen. 2015. “Cellular Senescence in Aging and Age-Related Disease: From Mechanisms to Therapy.” *Nature Medicine* 21 (12): 1424–35.
Chinetti-Gbaguidi, Giulia, Sophie Colin, and Bart Staels. 2015. “Macrophage Subsets in Atherosclerosis.” *Nature Reviews. Cardiology* 12 (1): 10–17.
Clayton, Tim C., Mary Thompson, and Tom W. Meade. 2008. “Recent Respiratory Infection and Risk of Cardiovascular Disease: Case-Control Study through a General Practice Database.” *European Heart Journal* 29 (1): 96–103.
Cochain, Clément, and Alma Zernecke. 2015. “Macrophages and Immune Cells in Atherosclerosis: Recent Advances and Novel Concepts.” *Basic Research in Cardiology* 110 (4): 34.
Cochain, Clement, and Alma Zernecke. 2017. “Macrophages in Vascular Inflammation and Atherosclerosis.” *Pflugers Archiv: European Journal of Physiology* 469 (3-4): 485–99.
Corrales-Medina, Vicente F., Karina N. Alvarez, Lisa A. Weissfeld, Derek C. Angus, Julio A. Chirinos, Chung-Chou H. Chang, Anne Newman, et al. 2015. “Association between Hospitalization for Pneumonia and Subsequent Risk of Cardiovascular Disease.” *JAMA: The Journal of the American Medical Association* 313 (3): 264–74.
Corrales-Medina, Vicente F., Daniel M. Musher, George A. Wells, Julio A. Chirinos, Li Chen, and Michael J. Fine. 2012. “Cardiac Complications in Patients with Community-Acquired Pneumonia: Incidence, Timing, Risk Factors, and Association with Short-Term Mortality.” *Circulation* 125 (6): 773–81.
Cristinelli, Sara, and Angela Ciuffi. 2018. “The Use of Single-Cell RNA-Seq to Understand Virus–host Interactions.” *Current Opinion in Virology* 29 (April): 39–50.
Cui, Yueli, Yuxuan Zheng, Xixi Liu, Liying Yan, Xiaoying Fan, Jun Yong, Yuqiong Hu, et al. 2019.

"Single-Cell Transcriptome Analysis Maps the Developmental Track of the Human Heart." *Cell Reports* 26 (7): 1934–50.e5.
Currier, Michael G., Sujin Lee, Christopher C. Stobart, Anne L. Hotard, Remi Villenave, Jia Meng, Carla D. Pretto, et al. 2016. "EGFR Interacts with the Fusion Protein of Respiratory Syncytial Virus Strain 2-20 and Mediates Infection and Mucin Expression." *PLoS Pathogens* 12 (5): e1005622.
Dalager-Pedersen, Michael, Mette Søgaard, Henrik Carl Schønheyder, Henrik Nielsen, and Reimar Wernich Thomsen. 2014. "Risk for Myocardial Infarction and Stroke after Community-Acquired Bacteremia: A 20-Year Population-Based Cohort Study." *Circulation* 129 (13): 1387–96.
Davis, Kimberly M., Sina Mohammadi, and Ralph R. Isberg. 2015. "Community Behavior and Spatial Regulation within a Bacterial Microcolony in Deep Tissue Sites Serves to Protect against Host Attack." *Cell Host & Microbe* 17 (1): 21–31.
Deczkowska, Aleksandra, Hadas Keren-Shaul, Assaf Weiner, Marco Colonna, Michal Schwartz, and Ido Amit. 2018. "Disease-Associated Microglia: A Universal Immune Sensor of Neurodegeneration." *Cell* 173 (5): 1073–81.
Deczkowska, Aleksandra, Assaf Weiner, and Ido Amit. 2020. "The Physiology, Pathology, and Potential Therapeutic Applications of the TREM2 Signaling Pathway." *Cell* 181 (6): 1207–17.
Dobin, Alexander, Carrie A. Davis, Felix Schlesinger, Jorg Drenkow, Chris Zaleski, Sonali Jha, Philippe Batut, Mark Chaisson, and Thomas R. Gingeras. 2013. "STAR: Ultrafast Universal RNA-Seq Aligner." *Bioinformatics* 29 (1): 15–21.
Duggan, Shane P., Catherine Garry, Fiona M. Behan, Sinead Phipps, Hiromi Kudo, Murat Kirca, Abdul Zaheer, et al. 2018. "siRNA Library Screening Identifies a Druggable Immune-Signature Driving Esophageal Adenocarcinoma Cell Growth." *Cellular and Molecular Gastroenterology and Hepatology* 5 (4): 569–90.
Ensan, Sherine, Angela Li, Rickvinder Besla, Norbert Degousee, Jake Cosme, Mark Roufaiel, Eric A. Shikatani, et al. 2016. "Self-Renewing Resident Arterial Macrophages Arise from Embryonic CX3CR1+ Precursors and Circulating Monocytes Immediately after Birth." *Nature Immunology* 17 (2): 159–68.
Febbraio, Maria, Ella Guy, and Roy L. Silverstein. 2004. "Stem Cell Transplantation Reveals That Absence of Macrophage CD36 Is Protective against Atherosclerosis." *Arteriosclerosis, Thrombosis, and Vascular Biology* 24 (12): 2333–38.
Febbraio, M., E. A. Podrez, J. D. Smith, D. P. Hajjar, S. L. Hazen, H. F. Hoff, K. Sharma, and R. L. Silverstein. 2000. "Targeted Disruption of the Class B Scavenger Receptor CD36 Protects against Atherosclerotic Lesion Development in Mice." *The Journal of Clinical Investigation* 105 (8): 1049–56.
Freigang, Stefan, Franziska Ampenberger, Adrienne Weiss, Thirumala-Devi Kanneganti, Yoichiro Iwakura, Martin Hersberger, and Manfred Kopf. 2013. "Fatty Acid–induced Mitochondrial Uncoupling Elicits Inflammasome-Independent IL-1α and Sterile Vascular Inflammation in Atherosclerosis." *Nature Immunology* 14 (10): 1045–53.
Galkina, Elena, Alexandra Kadl, John Sanders, Danielle Varughese, Ian J. Sarembock, and Klaus Ley. 2006. "Lymphocyte Recruitment into the Aortic Wall before and during Development of Atherosclerosis Is Partially L-Selectin Dependent." *The Journal of Experimental Medicine* 203 (5): 1273–82.
Gollan, Bridget, Grzegorz Grabe, Charlotte Michaux, and Sophie Helaine. 2019. "Bacterial Persisters and Infection: Past, Present, and Progressing." *Annual Review of Microbiology* 73 (September): 359–85.

Golumbeanu, Monica, Sara Cristinelli, Sylvie Rato, Miguel Munoz, Matthias Cavassini, Niko Beerenwinkel, and Angela Ciuffi. 2018. “Single-Cell RNA-Seq Reveals Transcriptional Heterogeneity in Latent and Reactivated HIV-Infected Cells.” *Cell Reports* 23 (4): 942–50.
Gomi, Kazunori, Vanessa Arbelaez, Ronald G. Crystal, and Matthew S. Walters. 2015. “Activation of NOTCH1 or NOTCH3 Signaling Skews Human Airway Basal Cell Differentiation toward a Secretory Pathway.” *PloS One* 10 (2): e0116507.
Goritzka, Michelle, Lydia R. Durant, Catherine Pereira, Samira Salek-Ardakani, Peter J. M. Openshaw, and Cecilia Johansson. 2014. “Alpha/beta Interferon Receptor Signaling Amplifies Early Proinflammatory Cytokine Production in the Lung during Respiratory Syncytial Virus Infection.” *Journal of Virology* 88 (11): 6128–36.
Goritzka, Michelle, Spyridon Makris, Fahima Kausar, Lydia R. Durant, Catherine Pereira, Yutaro Kumagai, Fiona J. Culley, Matthias Mack, Shizuo Akira, and Cecilia Johansson. 2015. “Alveolar Macrophage–derived Type I Interferons Orchestrate Innate Immunity to RSV through Recruitment of Antiviral Monocytes.” *The Journal of Experimental Medicine* 212 (5): 699–714.
Griffiths, Cameron, Steven J. Drews, and David J. Marchant. 2017. “Respiratory Syncytial Virus: Infection, Detection, and New Options for Prevention and Treatment.” *Clinical Microbiology Reviews* 30 (1): 277–319.
Guerreiro, Rita, Aleksandra Wojtas, Jose Bras, Minerva Carrasquillo, Ekaterina Rogaeva, Elisa Majounie, Carlos Cruchaga, et al. 2013. “TREM2 Variants in Alzheimer’s Disease.” *The New England Journal of Medicine* 368 (2): 117–27.
Hamada, Michito, Megumi Nakamura, Mai Thi Nhu Tran, Takashi Moriguchi, Cynthia Hong, Takayuki Ohsumi, Tra Thi Huong Dinh, et al. 2014. “MafB Promotes Atherosclerosis by Inhibiting Foam-Cell Apoptosis.” *Nature Communications* 5: 3147.
Han, Xiaoping, Renying Wang, Yincong Zhou, Lijiang Fei, Huiyu Sun, Shujing Lai, Assieh Saadatpour, et al. 2018. “Mapping the Mouse Cell Atlas by Microwell-Seq.” *Cell* 173 (5): 1307.
Han, Xiaoping, Ziming Zhou, Lijiang Fei, Huiyu Sun, Renying Wang, Yao Chen, Haide Chen, et al. 2020. “Construction of a Human Cell Landscape at Single-Cell Level.” *Nature* 581 (7808): 303–9.
Hasegawa, Hiromasa, Tetsu Watanabe, Shigehiko Kato, Taku Toshima, Miyuki Yokoyama, Yasuko Aida, Michiko Nishiwaki, et al. 2016. “The Role of Macrophage Transcription Factor MafB in Atherosclerotic Plaque Stability.” *Atherosclerosis* 250 (July): 133–43.
Heaton, Haynes, Arthur M. Talman, Andrew Knights, Maria Imaz, Daniel J. Gaffney, Richard Durbin, Martin Hemberg, and Mara K. N. Lawniczak. 2020. “Souporcell: Robust Clustering of Single-Cell RNA-Seq Data by Genotype without Reference Genotypes.” *Nature Methods* 17 (6): 615–20.
Helaine, Sophie, Angela M. Cheverton, Kathryn G. Watson, Laura M. Faure, Sophie A. Matthews, and David W. Holden. 2014. “Internalization of Salmonella by Macrophages Induces Formation of Nonreplicating Persisters.” *Science* 343 (6167): 204–8.
Hoffmann, Markus, Hannah Kleine-Weber, Simon Schroeder, Nadine Krüger, Tanja Herrler, Sandra Erichsen, Tobias S. Schiergens, et al. 2020. “SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor.” *Cell* 181 (2): 271–80.e8.
Howick, Virginia M., Andrew J. C. Russell, Tallulah Andrews, Haynes Heaton, Adam J. Reid, Kedar Natarajan, Hellen Butungi, et al. 2019. “The Malaria Cell Atlas: Single Parasite Transcriptomes across the Complete Plasmodium Life Cycle.” *Science* 365 (6455). https://doi.org/10.1126/science.aaw2619.

Hwang, Byungjin, Ji Hyun Lee, and Duhee Bang. 2018. “Single-Cell RNA Sequencing Technologies and Bioinformatics Pipelines.” *Experimental & Molecular Medicine* 50 (8): 1–14.
Imdahl, Fabian, and Antoine-Emmanuel Saliba. 2020. “Advances and Challenges in Single-Cell RNA-Seq of Microbial Communities.” *Current Opinion in Microbiology* 57 (November): 102–10.
Inchley, Christopher S., Tonje Sonerud, Hans O. Fjærli, and Britt Nakstad. 2015. “Nasal Mucosal microRNA Expression in Children with Respiratory Syncytial Virus Infection.” *BMC Infectious Diseases* 15 (March): 150.
Ioannidis, Ioannis, Beth McNally, Meredith Willette, Mark E. Peeples, Damien Chaussabel, Joan E. Durbin, Octavio Ramilo, Asuncion Mejias, and Emilio Flaño. 2012. “Plasticity and Virus Specificity of the Airway Epithelial Cell Immune Response during Respiratory Virus Infection.” *Journal of Virology* 86 (10): 5422–36.
Islam, Saiful, Amit Zeisel, Simon Joost, Gioele La Manno, Pawel Zajac, Maria Kasper, Peter Lönnerberg, and Sten Linnarsson. 2014. “Quantitative Single-Cell RNA-Seq with Unique Molecular Identifiers.” *Nature Methods* 11 (2): 163–66.
Jaitin, Diego Adhemar, Lorenz Adlung, Christoph A. Thaiss, Assaf Weiner, Baoguo Li, Hélène Descamps, Patrick Lundgren, et al. 2019. “Lipid-Associated Macrophages Control Metabolic Homeostasis in a Trem2-Dependent Manner.” *Cell* 178 (3): 686–98.e14.
Johnson, Sara M., Beth A. McNally, Ioannis Ioannidis, Emilio Flano, Michael N. Teng, Antonius G. Oomens, Edward E. Walsh, and Mark E. Peeples. 2015. “Respiratory Syncytial Virus Uses CX3CR1 as a Receptor on Primary Human Airway Epithelial Cultures.” *PLoS Pathogens* 11 (12): e1005318.
Jones, Jennifer E., Valerie Le Sage, and Seema S. Lakdawala. 2020. “Viral and Host Heterogeneity and Their Effects on the Viral Life Cycle.” *Nature Reviews. Microbiology*, October. https://doi.org/10.1038/s41579-020-00449-9.
Jonsson, Thorlakur, Hreinn Stefansson, Stacy Steinberg, Ingileif Jonsdottir, Palmi V. Jonsson, Jon Snaedal, Sigurbjorn Bjornsson, et al. 2013. “Variant of TREM2 Associated with the Risk of Alzheimer’s Disease.” *The New England Journal of Medicine* 368 (2): 107–16.
Kamari, Yehuda, Aviv Shaish, Shay Shemesh, Einav Vax, Itamar Grosskopf, Shahar Dotan, Malka White, et al. 2011. “Reduced Atherosclerosis and Inflammatory Cytokines in Apolipoprotein-E-Deficient Mice Lacking Bone Marrow-Derived Interleukin-1α.” *Biochemical and Biophysical Research Communications* 405 (2): 197–203.
Kang, Ju-Gyeong, Marcelo J. Amar, Alan T. Remaley, Jaeyul Kwon, Perry J. Blackshear, Ping-Yuan Wang, and Paul M. Hwang. 2011. “Zinc Finger Protein Tristetraprolin Interacts with CCL3 mRNA and Regulates Tissue Inflammation.” *Journal of Immunology* 187 (5): 2696–2701.
Kenney, Adam D., James A. Dowdle, Leonia Bozzacco, Temet M. McMichael, Corine St Gelais, Amanda R. Panfil, Yan Sun, et al. 2017. “Human Genetic Determinants of Viral Diseases.” *Annual Review of Genetics* 51 (November): 241–63.
Keren-Shaul, Hadas, Amit Spinrad, Assaf Weiner, Orit Matcovitch-Natan, Raz Dvir-Szternfeld, Tyler K. Ulland, Eyal David, et al. 2017. “A Unique Microglia Type Associated with Restricting Development of Alzheimer’s Disease.” *Cell* 169 (7): 1276–90.e17.
Kleinberger, Gernot, Matthias Brendel, Eva Mracsko, Benedikt Wefers, Linda Groeneweg, Xianyuan Xiang, Carola Focke, et al. 2017. “The FTD-like Syndrome Causing TREM 2 T66M Mutation Impairs Microglia Function, Brain Perfusion, and Glucose Metabolism.” *The EMBO Journal* 36 (13): 1837–53.
Klünemann, H. H., B. H. Ridha, L. Magy, J. R. Wherrett, D. M. Hemelsoet, R. W. Keen, J. L. De

Bleecker, et al. 2005. “The Genetic Causes of Basal Ganglia Calcification, Dementia, and Bone Cysts: DAP12 and TREM2.” *Neurology* 64 (9): 1502–7.
Knudsen, Nelson H., and Chih-Hao Lee. 2016. “Identity Crisis: CD301b+ Mononuclear Phagocytes Blur the M1-M2 Macrophage Line.” *Immunity* 45 (3): 461–63.
Kober, Daniel L., and Tom J. Brett. 2017. “TREM2-Ligand Interactions in Health and Disease.” *Journal of Molecular Biology* 429 (11): 1607–29.
Kröger, Carsten, Aoife Colgan, Shabarinath Srikumar, Kristian Händler, Sathesh K. Sivasankaran, Disa L. Hammarlöf, Rocío Canals, et al. 2013. “An Infection-Relevant Transcriptomic Compendium for Salmonella Enterica Serovar Typhimurium.” *Cell Host & Microbe* 14 (6): 683–95.
Krusat, T., and H. J. Streckert. 1997. “Heparin-Dependent Attachment of Respiratory Syncytial Virus (RSV) to Host Cells.” *Archives of Virology* 142 (6): 1247–54.
Kuchina, Anna, Leandra M. Brettner, Luana Paleologu, Charles M. Roco, Alexander B. Rosenberg, Alberto Carignano, Ryan Kibler, Matthew Hirano, R. William DePaolo, and Georg Seelig. 2019. “Microbial Single-Cell RNA Sequencing by Split-Pool Barcoding.” *Cold Spring Harbor Laboratory*. https://doi.org/10.1101/869248.
Kwong, Jeffrey C., Kevin L. Schwartz, and Michael A. Campitelli. 2018. “Acute Myocardial Infarction after Laboratory-Confirmed Influenza Infection.” *The New England Journal of Medicine*. Mass Medical Soc.
Lathe, Richard, Alexandra Sapronova, and Yuri Kotelevtsev. 2014. “Atherosclerosis and Alzheimer--Diseases with a Common Cause? Inflammation, Oxysterols, Vasculature.” *BMC Geriatrics* 14 (March): 36.
Lavin, Yonit, Soma Kobayashi, Andrew Leader, El-Ad David Amir, Naama Elefant, Camille Bigenwald, Romain Remark, et al. 2017. “Innate Immune Landscape in Early Lung Adenocarcinoma by Paired Single-Cell Analyses.” *Cell* 169 (4): 750–65.e17.
Lawlor, N., J. George, M. Bolisetty, and R. Kursawe. 2017. “Single-Cell Transcriptomes Identify Human Islet Cell Signatures and Reveal Cell-Type–specific Expression Changes in Type 2 Diabetes.” *Genome / National Research Council Canada = Genome / Conseil National de Recherches Canada*. https://genome.cshlp.org/content/27/2/208.short.
Libby, Peter. 2013. “Mechanisms of Acute Coronary Syndromes and Their Implications for Therapy.” *The New England Journal of Medicine* 368 (21): 2004–13.
Lifland, Aaron W., Jeenah Jung, Eric Alonas, Chiara Zurla, James E. Crowe Jr, and Philip J. Santangelo. 2012. “Human Respiratory Syncytial Virus Nucleoprotein and Inclusion Bodies Antagonize the Innate Immune Response Mediated by MDA5 and MAVS.” *Journal of Virology* 86 (15): 8245–58.
Lin, Weikang Nicholas, Matthew Zirui Tay, Ri Lu, Yi Liu, Chia-Hung Chen, and Lih Feng Cheow. 2020. “The Role of Single-Cell Technology in the Study and Control of Infectious Diseases.” *Cells* 9 (6). https://doi.org/10.3390/cells9061440.
Liu, Shuang, Li Gao, Xia Wang, and Yan Xing. 2015. “Respiratory Syncytial Virus Infection Inhibits TLR4 Signaling via up-Regulation of miR-26b.” *Cell Biology International* 39 (12): 1376–83.
Love, Michael I., Wolfgang Huber, and Simon Anders. 2014. “Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2.” *Genome Biology* 15 (12): 550.
Luecken, M. D., M. Büttner, K. Chaichoompu, A. Danese, M. Interlandi, M. F. Mueller, D. C. Strobl, et al. 2020. “Benchmarking Atlas-Level Data Integration in Single-Cell Genomics.” *Cold Spring Harbor Laboratory*. https://doi.org/10.1101/2020.05.22.111161.
Luo, Geyang, Qian Gao, Shuye Zhang, and Bo Yan. 2020. “Probing Infectious Disease by Single-Cell RNA Sequencing: Progresses and Perspectives.” *Computational and Structural*

Biotechnology Journal 18 (October): 2962–71.
Maaten, Laurens van der, and Geoffrey Hinton. 2008. "Visualizing Data Using T-SNE." *Journal of Machine Learning Research: JMLR* 9 (86): 2579–2605.
Makowski, Liza, Jeffrey B. Boord, Kazuhisa Maeda, Vladimir R. Babaev, K. Teoman Uysal, Maureen A. Morgan, Rex A. Parker, et al. 2001. "Lack of Macrophage Fatty-Acid–binding Protein aP2 Protects Mice Deficient in Apolipoprotein E against Atherosclerosis." *Nature Medicine* 7 (6): 699–705.
Malhotra, Rajneesh, Malcolm Ward, Helen Bright, Richard Priest, Martyn R. Foster, Michael Hurle, Eddie Blair, and Michael Bird. 2003. "Isolation and Characterisation of Potential Respiratory Syncytial Virus Receptor(s) on Epithelial Cells." *Microbes and Infection / Institut Pasteur* 5 (2): 123–33.
Marchant, David, Gurpreet K. Singhera, Soraya Utokaparch, Tillie L. Hackett, John H. Boyd, Zongshu Luo, Xiaoning Si, Delbert R. Dorscheid, Bruce M. McManus, and Richard G. Hegele. 2010. "Toll-like Receptor 4-Mediated Activation of p38 Mitogen-Activated Protein Kinase Is a Determinant of Respiratory Virus Entry and Tropism." *Journal of Virology* 84 (21): 11359–73.
Martinez, Fernando O., and Siamon Gordon. 2014. "The M1 and M2 Paradigm of Macrophage Activation: Time for Reassessment." *F1000prime Reports* 6 (March): 13.
Martínez-Martínez, Sara, and Juan Miguel Redondo. 2004. "Inhibitors of the calcineurin/NFAT Pathway." *Current Medicinal Chemistry* 11 (8): 997–1007.
McNamara, P. S., A. M. Fonceca, D. Howarth, J. B. Correia, J. R. Slupsky, R. E. Trinick, W. Al Turaiki, R. L. Smyth, and B. F. Flanagan. 2013. "Respiratory Syncytial Virus Infection of Airway Epithelial Cells, in Vivo and in Vitro, Supports Pulmonary Antibody Responses by Inducing Expression of the B Cell Differentiation Factor BAFF." *Thorax* 68 (1): 76–81.
Menon, Rajasree, Edgar A. Otto, Austin Kokoruda, Jian Zhou, Zidong Zhang, Euisik Yoon, Yu-Chih Chen, et al. 2018. "Single-Cell Analysis of Progenitor Cell Dynamics and Lineage Specification in the Human Fetal Kidney." *Development* 145 (16). https://doi.org/10.1242/dev.164038.
Munir, Shirin, Cyril Le Nouen, Cindy Luongo, Ursula J. Buchholz, Peter L. Collins, and Alexander Bukreyev. 2008. "Nonstructural Proteins 1 and 2 of Respiratory Syncytial Virus Suppress Maturation of Human Dendritic Cells." *Journal of Virology* 82 (17): 8780–96.
Musher, Daniel M., Michael S. Abers, and Vicente F. Corrales-Medina. 2019. "Acute Infection and Myocardial Infarction." *The New England Journal of Medicine* 380 (2): 171–76.
Musher, Daniel M., Adriana M. Rueda, Anjum S. Kaka, and Sulaiman M. Mapara. 2007. "The Association between Pneumococcal Pneumonia and Acute Cardiac Events." *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* 45 (2): 158–65.
Muus, Christoph, Malte D. Luecken, Gökcen Eraslan, Lisa Sikkema, Avinash Waghray, Graham Heimberg, Yoshihiko Kobayashi, et al. 2021. "Single-Cell Meta-Analysis of SARS-CoV-2 Entry Genes across Tissues and Demographics." *Nature Medicine* 27 (3): 546–59.
Nahrendorf, Matthias, and Filip K. Swirski. 2016. "Abandoning M1/M2 for a Network Model of Macrophage Function." *Circulation Research* 119 (3): 414–17.
Nakahara, Takehiro, Marc R. Dweck, Navneet Narula, David Pisapia, Jagat Narula, and H. William Strauss. 2017. "Coronary Artery Calcification: From Mechanism to Molecular Imaging." *JACC. Cardiovascular Imaging* 10 (5): 582–93.
Nam, Hannah H., and Michael G. Ison. 2019. "Respiratory Syncytial Virus Infection in Adults." *BMJ* 366 (September): l5021.
N'Diaye, Elsa-Noah, Catherine S. Branda, Steven S. Branda, Lisette Nevarez, Marco Colonna,

Clifford Lowell, Jessica A. Hamerman, and William E. Seaman. 2009. “TREM-2 (triggering Receptor Expressed on Myeloid Cells 2) Is a Phagocytic Receptor for Bacteria.” *The Journal of Cell Biology* 184 (2): 215–23.
Nuss, Aaron Mischa, Franziska Schuster, Louisa Roselius, Johannes Klein, René Bücker, Katharina Herbst, Ann Kathrin Heroven, et al. 2016. “A Precise Temperature-Responsive Bistable Switch Controlling Yersinia Virulence.” *PLoS Pathogens* 12 (12): e1006091.
Openshaw, Peter J. M., Chris Chiu, Fiona J. Culley, and Cecilia Johansson. 2017. “Protective and Harmful Immunity to RSV Infection.” *Annual Review of Immunology* 35 (April): 501–32.
Otero, Karel, Masahiro Shinohara, Haibo Zhao, Marina Cella, Susan Gilfillan, Angela Colucci, Roberta Faccio, et al. 2012. “TREM2 and β-Catenin Regulate Bone Homeostasis by Controlling the Rate of Osteoclastogenesis.” *Journal of Immunology* 188 (6): 2612–21.
Otero, Karel, Isaiah R. Turnbull, Pietro Luigi Poliani, William Vermi, Elisa Cerutti, Taiki Aoshi, Ilaria Tassi, et al. 2009. “Macrophage Colony-Stimulating Factor Induces the Proliferation and Survival of Macrophages via a Pathway Involving DAP12 and β-Catenin.” *Nature Immunology* 10 (7): 734–43.
Ozbudak, Ertugrul M., Mukund Thattai, Iren Kurtser, Alan D. Grossman, and Alexander van Oudenaarden. 2002. “Regulation of Noise in the Expression of a Single Gene.” *Nature Genetics* 31 (1): 69–73.
Packer, Jonathan S., Qin Zhu, Chau Huynh, Priya Sivaramakrishnan, Elicia Preston, Hannah Dueck, Derek Stefanik, et al. 2019. “A Lineage-Resolved Molecular Atlas of C. Elegans Embryogenesis at Single-Cell Resolution.” *Science* 365 (6459). https://doi.org/10.1126/science.aax1971.
Park, Yune-Jung, Seung-Ah Yoo, Mingyo Kim, and Wan-Uk Kim. 2020. “The Role of Calcium–Calcineurin–NFAT Signaling Pathway in Health and Autoimmune Diseases.” *Frontiers in Immunology* 11: 195.
Pastey, M. K., J. E. Crowe Jr, and B. S. Graham. 1999. “RhoA Interacts with the Fusion Glycoprotein of Respiratory Syncytial Virus and Facilitates Virus-Induced Syncytium Formation.” *Journal of Virology* 73 (9): 7262–70.
Picelli, Simone, Åsa K. Björklund, Omid R. Faridani, Sven Sagasser, Gösta Winberg, and Rickard Sandberg. 2013. “Smart-seq2 for Sensitive Full-Length Transcriptome Profiling in Single Cells.” *Nature Methods* 10 (11): 1096–98.
Polioudakis, Damon, Luis de la Torre-Ubieta, Justin Langerman, Andrew G. Elkins, Xu Shi, Jason L. Stein, Celine K. Vuong, et al. 2019. “A Single-Cell Transcriptomic Atlas of Human Neocortical Development during Mid-Gestation.” *Neuron* 103 (5): 785–801.e8.
Prinz, Marco, Daniel Erny, and Nora Hagemeyer. 2017. “Erratum: Ontogeny and Homeostasis of CNS Myeloid Cells.” *Nature Immunology* 18 (8): 951.
Qi, Furong, Shen Qian, Shuye Zhang, and Zheng Zhang. 2020. “Single Cell RNA Sequencing of 13 Human Tissues Identify Cell Types and Receptors of Human Coronaviruses.” *Biochemical and Biophysical Research Communications* 526 (1): 135–40.
Raj, Arjun, and Alexander van Oudenaarden. 2008. “Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences.” *Cell* 135 (2): 216–26.
Rajewsky, Nikolaus, Geneviève Almouzni, Stanislaw A. Gorski, Stein Aerts, Ido Amit, Michela G. Bertero, Christoph Bock, et al. 2020. “LifeTime and Improving European Healthcare through Cell-Based Interceptive Medicine.” *Nature* 587 (7834): 377–86.
Ramos, Irene, Gregory Smith, Frederique Ruf-Zamojski, Carles Martínez-Romero, Miguel Fribourg, Edwin A. Carbajal, Boris M. Hartmann, et al. 2019. “Innate Immune Response to Influenza Virus at Single-Cell Resolution in Human Epithelial Cells Revealed Paracrine Induction of Interferon Lambda 1.” *Journal of Virology* 93 (20).

https://doi.org/10.1128/JVI.00559-19.
Ramsköld, Daniel, Shujun Luo, Yu-Chieh Wang, Robin Li, Qiaolin Deng, Omid R. Faridani, Gregory A. Daniels, et al. 2012. “Full-Length mRNA-Seq from Single-Cell Levels of RNA and Individual Circulating Tumor Cells.” *Nature Biotechnology* 30 (8): 777–82.
Regev, Aviv, Sarah Teichmann, Orit Rozenblatt-Rosen, Michael Stubbington, Kristin Ardlie, Ido Amit, Paola Arlotta, et al. 2018. “The Human Cell Atlas White Paper.” *arXiv [q-bio.TO]*. arXiv. http://arxiv.org/abs/1810.05192.
Ren, J., T. Liu, L. Pang, K. Li, and R. P. Garofalo. 2011. “A Novel Mechanism for the Inhibition of Interferon Regulatory Factor-3-Dependent Gene Expression by Human Respiratory Syncytial Virus NS1 Protein.” *The Journal of*. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3353388/.
Reyfman, Paul A., James M. Walter, Nikita Joshi, Kishore R. Anekalla, Alexandra C. McQuattie-Pimentel, Stephen Chiu, Ramiro Fernandez, et al. 2019. “Single-Cell Transcriptomic Analysis of Human Lung Provides Insights into the Pathobiology of Pulmonary Fibrosis.” *American Journal of Respiratory and Critical Care Medicine* 199 (12): 1517–36.
Roche, Béatrice, and Dirk Bumann. 2021. “Single-Cell Reporters for Pathogen Responses to Antimicrobial Host Attacks.” *Current Opinion in Microbiology* 59 (February): 16–23.
Rosenberg, Alexander B., Charles M. Roco, Richard A. Muscat, Anna Kuchina, Paul Sample, Zizhen Yao, Lucas T. Graybuck, et al. 2018. “Single-Cell Profiling of the Developing Mouse Brain and Spinal Cord with Split-Pool Barcoding.” *Science* 360 (6385): 176–82.
Rossi, Giovanni A., and Andrew A. Colin. 2015. “Infantile Respiratory Syncytial Virus and Human Rhinovirus Infections: Respective Role in Inception and Persistence of Wheezing.” *The European Respiratory Journal: Official Journal of the European Society for Clinical Respiratory Physiology* 45 (3): 774–89.
———. 2017. “Respiratory Syncytial Virus--Host Interaction in the Pathogenesis of Bronchiolitis and Its Impact on Respiratory Morbidity in Later Life.” *Pediatric Allergy and Immunology: Official Publication of the European Society of Pediatric Allergy and Immunology* 28 (4): 320–31.
Rossi, Giovanni A., Michela Silvestri, and Andrew A. Colin. 2015. “Respiratory Syncytial Virus Infection of Airway Cells: Role of microRNAs.” *Pediatric Pulmonology* 50 (7): 727–32.
Saliba, Antoine-Emmanuel, Sara C Santos, and Jörg Vogel. 2017. “New RNA-Seq Approaches for the Study of Bacterial Pathogens.” *Current Opinion in Microbiology* 35 (February): 78–87.
Saliba, Antoine-Emmanuel, Alexander J. Westermann, Stanislaw A. Gorski, and Jörg Vogel. 2014. “Single-Cell RNA-Seq: Advances and Future Challenges.” *Nucleic Acids Research* 42 (14): 8845–60.
Sasagawa, Yohei, Itoshi Nikaido, Tetsutaro Hayashi, Hiroki Danno, Kenichiro D. Uno, Takeshi Imai, and Hiroki R. Ueda. 2013. “Quartz-Seq: A Highly Reproducible and Sensitive Single-Cell RNA Sequencing Method, Reveals Non-Genetic Gene-Expression Heterogeneity.” *Genome Biology* 14 (4): R31.
Schiller, Herbert B., Daniel T. Montoro, Lukas M. Simon, Emma L. Rawlins, Kerstin B. Meyer, Maximilian Strunz, Felipe A. Vieira Braga, et al. 2019. “The Human Lung Cell Atlas: A High-Resolution Reference Map of the Human Lung in Health and Disease.” *American Journal of Respiratory Cell and Molecular Biology* 61 (1): 31–41.
Schlepckow, Kai, Gernot Kleinberger, Akio Fukumori, Regina Feederle, Stefan F. Lichtenthaler, Harald Steiner, and Christian Haass. 2017. “An Alzheimer-Associated TREM2 Variant Occurs at the ADAM Cleavage Site and Affects Shedding and Phagocytic Function.” *EMBO*

Molecular Medicine 9 (10): 1356–65.
Shang, Jian, Gang Ye, Ke Shi, Yushun Wan, Chuming Luo, Hideki Aihara, Qibin Geng, Ashley Auerbach, and Fang Li. 2020. "Structural Basis of Receptor Recognition by SARS-CoV-2." *Nature* 581 (7807): 221–24.
Sheng, Kuanwei, Wenjian Cao, Yichi Niu, Qing Deng, and Chenghang Zong. 2017. "Effective Detection of Variation in Single-Cell Transcriptomes Using MATQ-Seq." *Nature Methods* 14 (3): 267–70.
Shingai, Masashi, Masahiro Azuma, Takashi Ebihara, Miwa Sasai, Kenji Funami, Minoru Ayata, Hisashi Ogura, Hiroyuki Tsutsumi, Misako Matsumoto, and Tsukasa Seya. 2008. "Soluble G Protein of Respiratory Syncytial Virus Inhibits Toll-like Receptor 3/4-Mediated IFN-Beta Induction." *International Immunology* 20 (9): 1169–80.
Shi, Weibin, Xuping Wang, Jack Wong, Catherine C. Hedrick, Howard Wong, Lawrence W. Castellani, and Aldons J. Lusis. 2004. "Effect of Macrophage-Derived Apolipoprotein E on Hyperlipidemia and Atherosclerosis of LDLR-Deficient Mice." *Biochemical and Biophysical Research Communications* 317 (1): 223–29.
Sica, Antonio, and Alberto Mantovani. 2012. "Macrophage Plasticity and Polarization: In Vivo Veritas." *The Journal of Clinical Investigation* 122 (3): 787–95.
Smeeth, Liam, Sara L. Thomas, Andrew J. Hall, Richard Hubbard, Paddy Farrington, and Patrick Vallance. 2004. "Risk of Myocardial Infarction and Stroke after Acute Infection or Vaccination." *The New England Journal of Medicine* 351 (25): 2611–18.
Spann, Kirsten M., Kim-C Tran, Bo Chi, Ronald L. Rabin, and Peter L. Collins. 2004. "Suppression of the Induction of Alpha, Beta, and Lambda Interferons by the NS1 and NS2 Proteins of Human Respiratory Syncytial Virus in Human Epithelial Cells and Macrophages." *Journal of Virology* 78 (12): 6705.
Spann, Kirsten M., Kim C. Tran, and Peter L. Collins. 2005. "Effects of Nonstructural Proteins NS1 and NS2 of Human Respiratory Syncytial Virus on Interferon Regulatory Factor 3, NF-kappaB, and Proinflammatory Cytokines." *Journal of Virology* 79 (9): 5353–62.
Stark, Rory, Marta Grzelak, and James Hadfield. 2019. "RNA Sequencing: The Teenage Years." *Nature Reviews. Genetics* 20 (11): 631–56.
Steuerman, Yael, Merav Cohen, Naama Peshes-Yaloz, Liran Valadarsky, Ofir Cohn, Eyal David, Amit Frishberg, et al. 2018. "Dissection of Influenza Infection In Vivo by Single-Cell RNA Sequencing." *Cell Systems* 6 (6): 679–91.e4.
Stewart, Mary K., and Brad T. Cookson. 2012. "Non-Genetic Diversity Shapes Infectious Capacity and Host Resistance." *Trends in Microbiology* 20 (10): 461–66.
Svensson, Valentine, Roser Vento-Tormo, and Sarah A. Teichmann. 2018. "Exponential Scaling of Single-Cell RNA-Seq in the Past Decade." *Nature Protocols* 13 (4): 599–604.
Swedan, Samer, Alla Musiyenko, and Sailen Barik. 2009. "Respiratory Syncytial Virus Nonstructural Proteins Decrease Levels of Multiple Members of the Cellular Interferon Pathways." *Journal of Virology* 83 (19): 9682–93.
Tabas, Ira, and Karin E. Bornfeldt. 2016. "Macrophage Phenotype and Function in Different Stages of Atherosclerosis." *Circulation Research* 118 (4): 653–67.
Tabula Muris Consortium, Overall coordination, Logistical coordination, Organ collection and processing, Library preparation and sequencing, Computational data analysis, Cell type annotation, Writing group, Supplemental text writing group, and Principal investigators. 2018. "Single-Cell Transcriptomics of 20 Mouse Organs Creates a Tabula Muris." *Nature* 562 (7727): 367–72.
Takahashi, Kazuya, Christian D. P. Rochford, and Harald Neumann. 2005. "Clearance of Apoptotic Neurons without Inflammation by Microglial Triggering Receptor Expressed on

Myeloid Cells-2." *The Journal of Experimental Medicine* 201 (4): 647–57.
Tal, Guy, Avigdor Mandelberg, Ilan Dalal, Karine Cesar, Eli Somekh, Asher Tal, Anat Oron, et al. 2004. "Association between Common Toll-like Receptor 4 Mutations and Severe Respiratory Syncytial Virus Disease." *The Journal of Infectious Diseases* 189 (11): 2057–63.
Tayyari, Farnoosh, David Marchant, Theo J. Moraes, Wenming Duan, Peter Mastrangelo, and Richard G. Hegele. 2011. "Identification of Nucleolin as a Cellular Receptor for Human Respiratory Syncytial Virus." *Nature Medicine* 17 (9): 1132–35.
Troy, Niamh M., and Anthony Bosco. 2016. "Respiratory Viral Infections and Host Responses; Insights from Genomics." *Respiratory Research* 17 (1): 156.
Turnbull, Isaiah R., Susan Gilfillan, Marina Cella, Taiki Aoshi, Mark Miller, Laura Piccio, Maristela Hernandez, and Marco Colonna. 2006. "Cutting Edge: TREM-2 Attenuates Macrophage Activation." *Journal of Immunology* 177 (6): 3520–24.
Ulland, Tyler K., and Marco Colonna. 2018. "TREM2—a Key Player in Microglial Biology and Alzheimer Disease." *Nature Reviews. Neurology* 14 (11): 667–75.
Ulland, Tyler K., Wilbur M. Song, Stanley Ching-Cheng Huang, Jason D. Ulrich, Alexey Sergushichev, Wandy L. Beatty, Alexander A. Loboda, et al. 2017. "TREM2 Maintains Microglial Metabolic Fitness in Alzheimer's Disease." *Cell* 170 (4): 649–63.e13.
Veening, Jan-Willem, Wiep Klaas Smits, and Oscar P. Kuipers. 2008. "Bistability, Epigenetics, and Bet-Hedging in Bacteria." *Annual Review of Microbiology* 62: 193–210.
Vieira Braga, Felipe A., Gozde Kar, Marijn Berg, Orestes A. Carpaij, Krzysztof Polanski, Lukas M. Simon, Sharon Brouwer, et al. 2019. "A Cellular Census of Human Lungs Identifies Novel Cell States in Health and in Asthma." *Nature Medicine* 25 (7): 1153–63.
Villenave, Rémi, Lindsay Broadbent, Isobel Douglas, Jeremy D. Lyons, Peter V. Coyle, Michael N. Teng, Ralph A. Tripp, Liam G. Heaney, Michael D. Shields, and Ultan F. Power. 2015. "Induction and Antagonism of Antiviral Responses in Respiratory Syncytial Virus-Infected Pediatric Airway Epithelium." *Journal of Virology* 89 (24): 12309–18.
Wang, Qihui, Yanfang Zhang, Lili Wu, Sheng Niu, Chunli Song, Zengyuan Zhang, Guangwen Lu, et al. 2020. "Structural and Functional Basis of SARS-CoV-2 Entry by Using Human ACE2." *Cell* 181 (4): 894–904.e9.
Weigel, W. A., and P. Dersch. 2018. "Phenotypic Heterogeneity: A Bacterial Virulence Strategy." *Microbes and Infection / Institut Pasteur* 20 (9-10): 570–77.
Wen, Wen, Wenru Su, Hao Tang, Wenqing Le, Xiaopeng Zhang, Yingfeng Zheng, Xiuxing Liu, et al. 2020. "Immune Cell Profiling of COVID-19 Patients in the Recovery Stage by Single-Cell Sequencing." *Cell Discovery* 6 (May): 31.
Westermann, Alexander J., Konrad U. Förstner, Fabian Amman, Lars Barquist, Yanjie Chao, Leon N. Schulte, Lydia Müller, Richard Reinhardt, Peter F. Stadler, and Jörg Vogel. 2016. "Dual RNA-Seq Unveils Noncoding RNA Functions in Host–pathogen Interactions." *Nature* 529 (7587): 496–501.
Whelan, Jillian N., Kim C. Tran, Damian B. van Rossum, and Michael N. Teng. 2016. "Identification of Respiratory Syncytial Virus Nonstructural Protein 2 Residues Essential for Exploitation of the Host Ubiquitin System and Inhibition of Innate Immune Responses." *Journal of Virology* 90 (14): 6453–63.
Yao, Yinan, Hequan Li, Junjun Chen, Weiyi Xu, Guangdie Yang, Zhang Bao, Dajing Xia, Guohua Lu, Shuwen Hu, and Jianying Zhou. 2016. "TREM-2 Serves as a Negative Immune Regulator through Syk Pathway in an IL-10 Dependent Manner in Lung Cancer." *Oncotarget* 7 (20): 29620–34.
Yeh, Felix L., Yuanyuan Wang, Irene Tom, Lino C. Gonzalez, and Morgan Sheng. 2016. "TREM2 Binds to Apolipoproteins, Including APOE and CLU/APOJ, and Thereby Facilitates Uptake

of Amyloid-Beta by Microglia." *Neuron* 91 (2): 328–40.
Yu, Guangchuang, Li-Gen Wang, Yanyan Han, and Qing-Yu He. 2012. "clusterProfiler: An R Package for Comparing Biological Themes Among Gene Clusters." *Omics: A Journal of Integrative Biology* 16 (5): 284–87.
Zanini, Fabio, Szu-Yuan Pu, Elena Bekerman, Shirit Einav, and Stephen R. Quake. 2018. "Single-Cell Transcriptional Dynamics of Flavivirus Infection." *eLife* 7 (February). https://doi.org/10.7554/eLife.32942.
Zanotti, Ilaria, Matteo Pedrelli, Francesco Potì, Grazia Stomeo, Monica Gomaraschi, Laura Calabresi, and Franco Bernini. 2011. "Macrophage, but Not Systemic, Apolipoprotein E Is Necessary for Macrophage Reverse Cholesterol Transport in Vivo." *Arteriosclerosis, Thrombosis, and Vascular Biology* 31 (1): 74–80.
Zernecke, Alma. 2015. "Dendritic Cells in Atherosclerosis: Evidence in Mice and Humans." *Arteriosclerosis, Thrombosis, and Vascular Biology* 35 (4): 763–70.
Zhang, Ji-Yuan, Xiang-Ming Wang, Xudong Xing, Zhe Xu, Chao Zhang, Jin-Wen Song, Xing Fan, et al. 2020. "Single-Cell Landscape of Immunological Responses in Patients with COVID-19." *Nature Immunology* 21 (9): 1107–18.
Zhang, Liqun, Mark E. Peeples, Richard C. Boucher, Peter L. Collins, and Raymond J. Pickles. 2002. "Respiratory Syncytial Virus Infection of Human Airway Epithelial Cells Is Polarized, Specific to Ciliated Cells, and without Obvious Cytopathology." *Journal of Virology* 76 (11): 5654–66.
Zhang, Xiaojing, Wei Wang, Peng Li, Xudong Wang, and Kan Ni. 2018. "High TREM2 Expression Correlates with Poor Prognosis in Gastric Cancer." *Human Pathology* 72 (February): 91–99.
Zhang, Zheren, Dennis Claessen, and Daniel E. Rozen. 2016. "Understanding Microbial Divisions of Labor." *Frontiers in Microbiology* 7 (December): 2070.
Zheng, Grace X. Y., Jessica M. Terry, Phillip Belgrader, Paul Ryvkin, Zachary W. Bent, Ryan Wilson, Solongo B. Ziraldo, et al. 2017. "Massively Parallel Digital Transcriptional Profiling of Single Cells." *Nature Communications* 8 (January): 14049.