Implementation and application of bioinformatical software for the analysis of dual RNA sequencing data of host and pathogen during infection

---

Implementierung und Anwendung bioinformatischer Software für die Analyse von dual RNA-Sequenzierdaten von Wirt und Erreger während Infektion

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

submitted by

Till Sauerwein
from
Aschaffenburg

Würzburg, 2022

Submitted on:

Members of the *Promotionskomitee*:

Chairperson: Prof. Dr. Manfred Gessler

Primary Supervisor: Prof. Dr. Konrad Förstner

Supervisor (Second): Prof. Dr. Thomas Dandekar

Supervisor (Third): Jun. Prof. Dr. Alexander Westermann

Date of Public Defense:

Date of Receipt of Certificates:

Abstract

Since the advent of high-throughput sequencing technologies in the mid-2010s, RNA sequencing (RNA-seq) has been established as the method of choice for studying gene expression. In comparison to microarray-based methods, which have mainly been used to study gene expression before the rise of RNA-seq, RNA-seq is able to profile the entire transcriptome of an organism without the need to predefine genes of interest. Today, a wide variety of RNA-seq methods and protocols exist, including dual RNA sequencing (dual RNA-seq) and multi RNA sequencing (multi RNA-seq). Dual RNA-seq and multi RNA-seq simultaneously investigate the transcriptomes of two or more species, respectively. Therefore, the total RNA of all interacting species is sequenced together and only separated *in silico*. Compared to conventional RNA-seq, which can only investigate one species at a time, dual RNA-seq and multi RNA-seq analyses can connect the transcriptome changes of the species being investigated and thus give a clearer picture of the interspecies interactions. Dual RNA-seq and multi RNA-seq have been applied to a variety of host-pathogen, mutualistic and commensal interaction systems.

We applied dual RNA-seq to a host-pathogen system of human mast cells and *Staphylococcus aureus* (*S. aureus*). *S. aureus*, a commensal gram-positive bacterium, can become an opportunistic pathogen and infect skin lesions of atopic dermatitis (AD) patients. Among the first immune cells *S. aureus* encounters are mast cells, which have previously been shown to be able to kill the bacteria by discharging antimicrobial products and releasing extracellular traps made of protein and deoxyribonucleic acid (DNA). However, *S. aureus* is known to evade the host's immune response by internalizing within mast cells. Our dual RNA-seq analysis of different infection settings revealed that mast cells and *S. aureus* need physical contact to influence each other's gene expression. We could show that *S. aureus* cells internalizing within mast cells undergo profound transcriptome changes to adjust their metabolism to survive in the intracellular niche. On the host side, we found out that infected mast cells elicit a type-I interferon (IFN-I) response in an autocrine manner and in a paracrine manner to non-infected bystander-cells. Our study provides the first evidence that mast cells are capable to produce IFN-I upon infection with a bacterial pathogen.

In order to facilitate the bioinformatical analysis of dual RNA-seq and multi RNA-seq we released a major update of the already existing RNA-seq analysis tool *READemption*. The new version *READemption 2* allows users to analyze dual RNA-seq and multi RNA-seq data of any number of species in a convenient way, while still being able to analyze conventional RNA-seq projects that investigate only one species. In the course of development, emphasis was placed on keeping the software quality high by following good practices for scientific software development.

Zusammenfassung

Seit dem Aufkommen von Hochdurchsatz-Sequenziertechnologien Mitte der 2010er Jahre hat sich RNA-Sequenzierung (RNA-seq) als Methode der Wahl für die Untersuchung von Genexpression etabliert. Im Vergleich zu Microarray-basierten Methoden, die vor dem Aufkommen von RNA-seq hauptsächlich zur Untersuchung der Genexpression verwendet wurden, kann mit RNA-seq das gesamte Transkriptom eines Organismus charakterisiert werden, ohne dass die Gene von Interesse vorab definiert werden müssen. Heute gibt es eine Vielzahl von RNA-seq-Methoden und Protokollen, darunter Dual RNA-seq und Multi RNA-seq. Dual RNA-seq und Multi RNA-seq untersuchen gleichzeitig die Transkriptome von zwei bzw. mehreren Arten. Dazu wird die gesamte RNA aller interagierenden Arten gemeinsam sequenziert und nur *in silico* aufgetrennt. Im Vergleich zur herkömmlichen RNA-seq, bei der jeweils nur eine Spezies untersucht wird, können Dual RNA-seq- und Multi RNA-seq-Analysen die Transkriptomveränderungen der untersuchten Spezies miteinander in Verbindung bringen und so ein klareres Bild der Wechselwirkungen zwischen den Spezies vermitteln. Dual RNA-seq und Multi RNA-seq wurden bereits auf eine Vielzahl von Wirt-Pathogen-, mutualistischen und kommensalen Interaktionssystemen angewendet.

Wir haben Dual RNA-seq auf ein Wirt-Pathogen-System aus menschlichen Mastzellen und *S. aureus* angewendet. *S. aureus*, ein kommensales grampositives Bakterium, kann zu einem opportunistischen Erreger werden und Hautläsionen von Patienten mit atopischer Dermatitis (AD) infizieren. Zu den ersten Immunzellen, auf die *S. aureus* trifft, gehören Mastzellen, die nachweislich in der Lage sind, das Bakterium abzutöten, indem sie antimikrobielle Produkte abgeben und extrazelluläre Fallen aus Proteinen und DNA freisetzen. Es ist jedoch bekannt, dass *S. aureus* die Immunantwort des Wirts umgehen kann, indem es in die Mastzellen internalisiert wird. Unsere Dual RNA-seq-Analyse verschiedener Infektionssituationen ergab, dass Mastzellen und *S. aureus* physischen Kontakt benötigen, um ihre Genexpression gegenseitig zu beeinflussen. Wir konnten zeigen, dass *S. aureus* Zellen, die von Mastzellen internalisiert werden, tiefgreifende Transkriptomveränderungen durchlaufen, um ihren Stoffwechsel für das Überleben in der intrazellulären Nische anzupassen. Auf Seite des Wirts fanden wir heraus, dass infizierte Mastzellen eine IFN-I (Interferon Typ I)-Antwort auf autokrine und auf parakrine Weise auf nicht-infizierte, in der Nähe befindliche Zellen auslösen. Unsere Studie liefert den ersten Beweis dafür, dass

Mastzellen bei einer Infektion mit einem bakteriellen Erreger in der Lage sind, IFN-I zu produzieren.

Um die bioinformatische Analyse von Dual RNA-seq und Multi RNA-seq zu erleichtern, haben wir ein umfangreiches Update des bereits existierenden RNA-seq-Analyseprogramms *READemption* veröffentlicht. Die neue Version *READemption 2* ermöglicht es den Nutzern, Dual RNA-seq- und Multi RNA-seq-Daten einer beliebigen Anzahl von Spezies auf bequeme Weise zu analysieren, während es weiterhin möglich ist, herkömmliche RNA-seq-Projekte zu analysieren, die nur eine Spezies untersuchen. Bei der Entwicklung wurde Wert darauf gelegt, die Qualität der Software durch die Einhaltung bewährter Verfahren für die Entwicklung wissenschaftlicher Software hoch zu halten.

Contents

1 Introduction

1.1 RNA - How our understanding has changed from a mere messenger molecule to a universal regulator

The central dogma of molecular biology describes the genetic information flow within a cell, which all life forms have in common. It states that information can be passed from deoxyribonucleic acid (DNA) to ribonucleic acid (RNA) and from RNA to protein and thus also from DNA to protein. While it is possible to translate the sequence information held by RNA back to DNA, meaning a DNA sequence can be derived by its transcribed RNA sequence, a protein sequence of amino acids can never be translated back to its originating RNA or DNA sequence (Crick, 1958). This can be explained by the fact that the genetic code is redundant and multiple three-base pair codons with different nucleic acid sequences result in the same amino acid. Therefore, a protein sequence consisting of amino acids can not unambigously be translated back to a nucleic acid sequence. Given the fact that an RNA sequence holds information about its originating DNA sequence and its resulting protein sequence and hence covers a large spectrum of the protein biosynthesis, makes RNA a molecule of high interest in biology and medicine. The essential RNA species taking part in protein biosynthesis are messenger RNA (mRNA), which holds the genetic information after transcription, and ribosomal RNA (rRNA) and transfer RNA (tRNA), which both play important roles in translation. These RNA species were first discovered in the 1950s and became the main focus in the research field of RNA in the following decades.

In 1965, the first complete nucleotide sequence of a tRNA, the alanine tRNA from yeast could be determined (Holley et al., 1965). In the late 1960s the discovery of precursors of mature mRNA and rRNA was the door opener for studies that later revealed the mechanisms of rRNA processing and splicing (Lewis et al., 1975; Berk, 2016). The first complete genome sequence of an organism, namely the Bacteriophage MS2 (*Emesvirus zinderi*) was published in 1976 (Fiers et al., 1976). Although the RNA genome is only 3,569 nucleotides long, it was considered a landmark in molecular biology. While essential

discoveries regarding the function, nucleotide sequence and structure of rRNA and tRNA were made in this period, RNA was mostly considered a mere messenger of the flow of genetic information from gene to protein (Jarroux et al., 2017). This view gradually changed in the late 1970s and 1980s with the discovery that RNA can function as a catalyst for chemical reactions (Kruger et al., 1982; Guerrier-Takada et al., 1983). In 1984, the first gene expression regulating non-coding RNA (ncRNA), micF in *E. coli*, was discovered. It was shown that micF repressed the translation of its target mRNA into a porin, an outer membrane protein, through base pairing with the mRNA. This new class of regulating, ncRNA in bacteria had been termed small RNA (sRNA) (Inouye and Delihas, 1988).

A similar concept of regulation by RNAs in eukaryotes was discovered in the early 1990s. It was found that the lin-4 gene of the nematode *Caenorhabditis elegans* produces two sRNAs of the size of 22 and 61 nucleotides. The shorter RNA that is cut from the longer RNA base pairs with the untranslated region (UTR) of the 3'-end of the lin-14 RNA and thus silences the gene expression post transcription. This was the first example of a eukaryotic micro RNA (miRNA) and RNA interference (Lee et al., 1993).

The technological progress of the past 15 years made in detecting and sequencing RNA and the knowledge gained from it reinforced the point of view that RNA is much more than a messenger of the flow of genetic information. RNA plays an important role in regulating gene expression at all levels - ranging from epigenetic chromatin modification to transcription and translation.

1.2 Eukaryotic and bacterial non-coding RNA

The following section gives a brief overview of the functions and mechanisms of the diverse classes of ncRNAs of eukaryotes and bacteria that have been discovered in the late 20th and early 21st century. Because tRNA and rRNA are commonly known, they will not be described, but it should be noted that they exist in eukaryotes and prokaryotes and also belong to the class of ncRNAs. The proportions of mass of all RNA classes in eukaryotic and bacterial cells (Westermann et al., 2012) are shown in Table 1.1.

1.2.1 Eukaryotic non-coding RNA

MiRNA

Micro RNAs (miRNA) are single-stranded RNA molecules of about 22 nucleotides length that form RNA-induced silencing complexes (RISC) with proteins of the Argonaute family and other proteins. The miRNA guides RISC to a target mRNA and binds it by base pairing. The translation of the target mRNA is then hindered by mRNA cleavage, inhibition of translation and initiation of mRNA decay (Bartel, 2009; Macfarlane and Murphy, 2010).

SiRNA

Small interfering RNAs (siRNA) have a similar length compared to miRNAs and also build RISC complexes to hinder translation of their target mRNA. In contrast to miRNAs, the complete siRNA sequence is fully complementary to its target mRNA, while miRNAs only bind with a seed region of up to seven nucleotides to their target mRNA, usually to its UTR at the 3'-end. Therefore, miRNAs have a general broader specificity compared to siRNAs (Lam et al., 2015).

PiRNA

P-element-induced wimpy testis (Piwi)-interacting RNA (piRNA) is the largest class of small ncRNA molecules with a length of 21 to 36 nucleotides. They form RNA-protein complexes with piwi-subfamily Argonaute proteins and are mainly involved in preserving genome integrity through silencing of transposable elements (Siomi et al., 2011; Diamantopoulos et al., 2018).

SnoRNA and scaRNA

Small nucleolar RNAs (snoRNA) are a class of regulatory small ncRNAs (60 to 250 nucleotides) that guide chemical modifications of rRNA and other RNA molecules. They can be further distinguished by their sequence motif and secondary structure. C/D box snoRNAs consist of a sequence motif called C box (RUGAUGA motif, where R is a purine) and one called D box (CUGA motif). H/ACA box snoRNAs consist of a two-hairpin structure that is connected by an H box region (ANANNA, N corresponds to nucleotide). The C/D box snoRNAs are associated with methylation of rRNA and the H/ACA box snoRNAs with pseudouridylation of rRNA. A third subclass are small Cajal body-specific RNAs (scaRNA), which possess both C/D and H/ACA boxes and an addi-

tional CAB box (UGAG motif). They are located in the Cajal body and are involved in the biogenesis and modification of small nuclear ribonucleoproteins through methylation and pseudouridylation (Henras et al., 2004; Reichow et al., 2007).

SnRNAs

Small nuclear RNAs (snRNA) exist in the nucleus and play an important role in intron splicing and RNA processing. They form ribonucleoproteins and, together with other proteins, build the spliceosome. A particular snRNA (U7) also plays a role in histone pre-mRNA processing (Valadkhan, 2005; Lesman et al., 2021).

LncRNA

The class of long non-coding RNAs (lncRNA) consists of a large and highly heterogeneous collection of transcripts, which differ in their biogenesis and genomic origin and carry out various functions in cells. Members of this class have a length of at least 200 nucleotides. The various functions of lncRNAs can be categorized into four broad groups: Mediation of chromatin modifications and methylation of DNA involved in epigenetic regulation; DNA and protein interactions involved in transcriptional level regulation; post-transcriptional mRNA processing and regulation of protein translation; and post-translation modification via interactions with proteins (Statello et al., 2021).

1.2.2 Bacterial non-coding RNA

SRNA

Bacterial sRNAs are 50 to 500 nucleotides long and regulate gene expression in various ways. Down-regulation of gene expression can happen via the following mechanisms: Through base pairing with target mRNAs, sRNAs prevent binding of ribosomes to the mRNAs and thus inhibit translation initiation. Base pairing with mRNAs can also lead to recruitment of ribonucleases that degrade both the sRNA and the mRNA and hence stops gene expression after transcription. SRNAs can also change the conformation of transcripts and thereby generate intrinsic terminators that prevent movement of the RNA polymerase. This attenuates gene expression by premature termination of transcription. SRNAs can also up-regulate gene expression with the following mechanisms: They can bind to mRNAs to protect them from degradation of ribonucleases and consequently increase protein output. Leader sequences of mRNAs can contain secondary structures that inhibit ribosomes to bind the ribosome binding sites. Some sRNAs have been shown to

bind to the mRNA, changing its secondary structure and unfolding the ribosome binding site to allow initiation of translation. Furthermore, sRNAs are able to up-regulate transcription by inhibiting Rho-dependent transcription termination. Another form of gene expression regulation by sRNAs is sequestration of proteins. The sRNAs form single or multiple protein binding folds that sequester proteins, so that the proteins are no longer available to exert their functions on their mRNA targets and thus down-regulating them (Dutta and Srivastava, 2018; Denham, 2020).

TmRNA

Transfer-messenger RNAs (tmRNA) are RNAs that have properties of both tRNA and mRNA. During translation, ribosomes stall when the mRNA is missing a stop codon. TmRNAs help releasing stalled ribosomes and making them available for translation again as well as causing degradation of the incomplete nascent polypeptide (Keiler and Ramadoss, 2011)

Table 1.1: Proportion of mass of RNA classes in eukaryotic and bacterial cells

RNA class	**Eukaryotic cell**	**Bacterial cell**
rRNA	~80%	~80%
tRNA	~15%	14-15%
mRNA		4-5%
snRNA		-
snoRNA		-
scaRNA	~5%	-
miRNA		-
siRNA		-
piRNA		-
lncRNA		-
tmRNA	-	<1%
sRNA	-	Varies

1.3 RNA sequencing

1.3.1 Historic development and state-of-the-art technologies

1.3.1.1 Historic development of RNA sequencing

The study of an organism's RNA is called transcriptomics and investigates the transcriptome, which is the entirety of RNA produced by specific cell types or under certain circumstances. To this day RNA-seq is the method of choice used for studying transcriptomes. Because of its ability to sequence full transcriptomes, RNA-seq has superseded microarray-based methods, which could only profile predefined genes and transcripts. The majority of RNA-seq technologies are closely linked to DNA sequencing (DNA-seq) methods, because usually the RNA of a sample is reverse transcribed to complementary DNA (cDNA) and then the cDNA is sequenced. The accomplishments in the field of RNA research made in the past 15 years are mainly based on the advances achieved in RNA-seq and DNA-seq at that time. Although the Sanger DNA-seq technology was available since 1977 and was predominant until the early 2000s, it was costly and time consuming compared to modern technologies. Sanger sequencing was also used for the human genome project that had the aim to sequence the first complete human genome in history. It ended in 2003 after 13 years, had costs of three billion U.S. dollars and yielded the first complete human genome, though the sequence was a patchwork of several people (Lander et al., 2001; *Human Genome Project Fact Sheet* 2022). Five years later, the first human genome of an individual was published. The sequencing took the researchers only two months and cost only one million U.S. dollars (Wheeler et al., 2008). This reduction of sequencing speed and costs was possible by the use of the first commercial high-throughput sequencing machine, the *454 System* by *454 Live Science Corp* launched in 2005 (Margulies et al., 2005). In the next 15 years other companies developed high-throughput sequencing machines (also called next-generation sequencing machines) that significantly reduced the costs and speed of sequencing. While the costs for sequencing one megabase of DNA in the year 2001 was more than 5,000 U.S. dollars, 20 years later it was only 0.6 cents (*National Human Genome Research Institute - Sequencing costs* 2022) (Figure 1.1). This decline in sequencing costs enabled researchers to perform en masse DNA and RNA sequencing. An indicator of the continuing interest in DNA and RNA sequencing is the number of bases uploaded in the past decade to the Sequence Read Archive (SRA), the largest publicly available repository of high-throughput sequencing data. The number of bases rose steadily since 2012 and amounted to 67 petabases by the end of 2021 (*Sequence*

Read Archive - Bases in database,
https://www.ncbi.nlm.nih.gov/sra/docs/sragrowth/ 2022) (Figure 1.2).

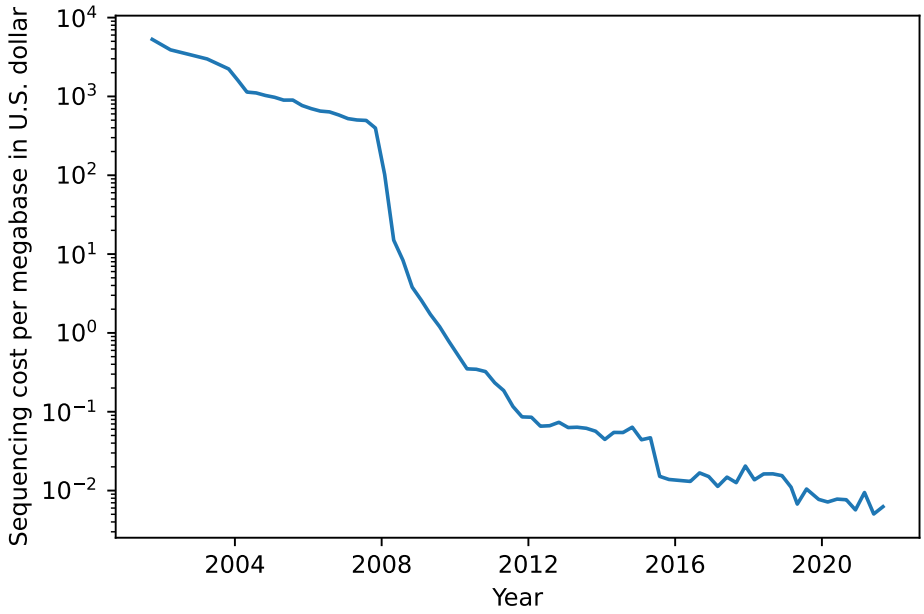

Figure 1.1: **Sequencing costs development over time.** Sequencing cost per megabase in U.S. dollar as estimated by the National Human Genome Research Institute (NHGRI) (*National Human Genome Research Institute - Sequencing costs* 2022)

.

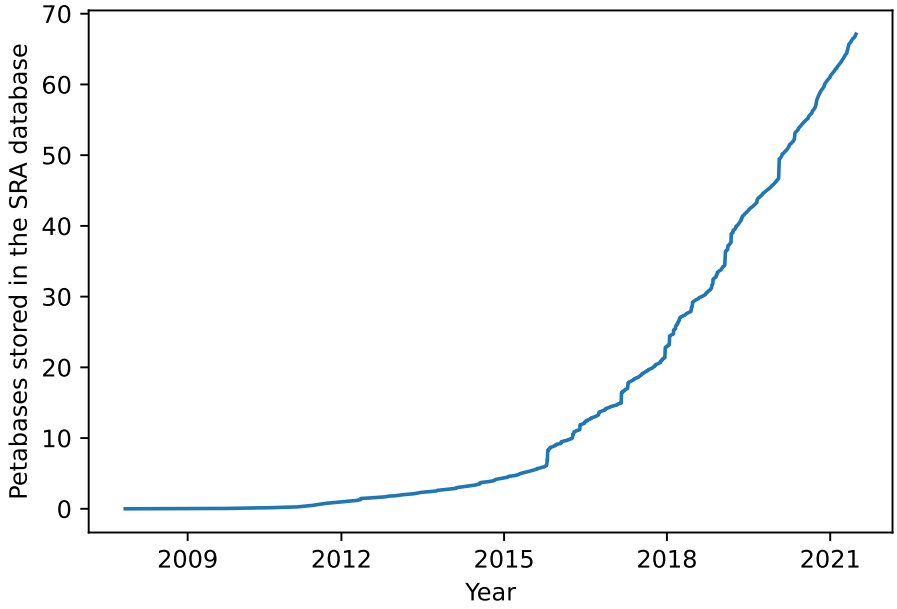

Figure 1.2: **Stored sequenced bases SRA over time.** Stored bases of DNA and RNA sequencing in petabases in the SRA database over time. (*Sequence Read Archive - Bases in database, https://www.ncbi.nlm.nih.gov/sra/docs/sragrowth/* 2022)

.

1.3.1.2 State-of-the-art RNA sequencing technologies

The state-of-the-art technology used nowadays for RNA-seq can be divided into three main categories: Short read cDNA sequencing, long read cDNA sequencing and long read direct RNA-seq. Except for the direct sequencing approach, RNA is reverse transcribed into cDNA and subsequently sequenced.

Short read cDNA sequencing

The Ion Torrent sequencing technology uses a sequencing-by-synthesis approach. A microwell containing a single-stranded DNA template to be sequenced is flooded successively with four different deoxyribonucleotide triphosphates (dNTP). When one of the dNTPs is incorporated by a DNA polymerase into the growing complementary strand, a hydrogen ion is released and the resulting pH change can be detected. Different Ion Torrent machines produce read lengths from 200 up to 600 nucleotides (Rothberg et al., 2011).

Illumina dye sequencing also relies on sequencing-by-synthesis. The process begins with fragmenting cDNA and adding adapters to the fragments. Then the fragments are loaded onto a flow cell where they bind with the added adapters to anchoring molecules. The fragments that were attached to the surface are amplified by bridge polymerase chain reaction (PCR) and clusters of about a thousand copies of each fragment are created. During the actual sequencing process that follows, fluorescently tagged dNTPs with a reversible blocking group are added to the flow cell. Once the matching dNTP is incorporated in the complementary strand, the specific fluorescence signal of this nucleotide is detected and the blocking group is removed. This cycle is repeated, detecting one nucleotide at a time. The various Illumina sequencing machines reach maximum read lengths from 150 up to 350 nucleotides (Bentley et al., 2008).

Long read cDNA sequencing

Pacific Biosciences' (PacBio) single-molecule real-time (SMRT) sequencing technology produces reads from 250 up to 50,000 nucleotides. The method also works by the sequencing-by-synthesis principal. Tiny holes called zero-mode waveguides with a diameter of 70 nanometer (nm) and a depth of 100 nm contain a DNA polymerase. Attached to the polymerase is a single-strand DNA template. Fluorescently tagged dNTPs in the solution travel into the zero-mode waveguide and leave it again, fast enough to not be excited by a light in the hole. However, when a matching dNTP is incorporated by the polymerase, the dNTP is excited and its color signal can be detected by a photodetector (Eid et al., 2009).

Nanopore sequencing by Oxford Nanopore Technologies does not rely on sequencing-by-synthesis, but on detecting the different sizes of nucleotides. A chamber is divided into two compartments by a lipid membrane that contains transmembrane proteins, called porins. A voltage is applied across the membrane, causing charged particles in the solution to travel through the porins and inducing a total charge flow that can be measured. After unwinding a double-stranded DNA molecule, one strand is pulled through the transmembrane protein and depending on various factors, like geometry, size and chemical composition of the nucleotides inside the pore, the base flow of particles changes. This change can be translated into the sequence of the DNA molecule. The read lengths that can be achieved usually range from 10,000 to 30,000 nucleotides (Jain et al., 2018).

Long read direct RNA sequencing

The nanopore technology can also be used to sequence RNA directly. This has the advantage that cDNA synthesis and PCR amplification can be omitted during library preparation and thus bias is reduced. Another advantage is that epigenetic information about nucleotide modifications can be retained by this approach (Schatz, 2017; Garalde et al., 2018).

Compared with short reads, long reads are better suited for analyzing long repetitive genomic regions, isoform detection and de-novo transcriptome analysis (Wright et al., 2022). Short-read sequencing technologies have a general higher throughput and lower error rates. While the Illumina platform can generate 10^9 to 10^{10} reads per run, the Pacific Bio and Oxford Nanopore machines only reach 10^6 to 10^7 (Stark et al., 2019). A higher throughput results in higher read depths, which is defined as the total number of reads obtained for a sample. High read depths allow detecting genes expressed at low levels and are indispensable when performing large-scale differential gene expression (DGE) analysis. Since DGE analysis is the most widely used application of RNA-seq and Illumina sequencing is well suited for this purpose due to its high throughput, more than 97% of all RNA-seq datasets in the SRA database have been constructed with the Illumina technology (Stark et al., 2019).

1.3.2 RNA sequencing analysis workflow

A typical RNA-seq analysis workflow begins with library preparation of extracted RNA, including enrichment or depletion of certain RNA classes and converting RNA into cDNA. After library preparation, the cDNA is sequenced by a sequencing machine, which creates

millions of raw sequencing reads. The raw reads serve as input for the following bioinformatical analysis. The following section describes the library preparation for Illumina sequencing and the bioinformatical analysis is described further below (1.4.2).

Prior to the actual RNA-seq, RNA has to be extracted from cells and except for direct RNA-seq, appropriate cDNA libraries have to be constructed. Library construction for Illumina machines consists of three steps: In the first step, specific RNA species can be enriched or depleted. In the second step, RNA is converted to cDNA because Illumina technologies can not directly sequence DNA. In the third step, sequencing adapters are added to the cDNA.

When performing DGE analysis, researchers are usually interested in mRNA and regulatory or functional ncRNAs, while the abundant rRNA and tRNA molecules are neglected. However, rRNA and tRNA account for about 80% and 15% of the total RNA mass in a cell, respectively. If this large difference is not considered, the majority of sequencing reads would originate from rRNA and other low expressed non-coding and coding RNAs would be at risk not being detected by the sequencer. To overcome this disparity, rRNA can be depleted from a sample with rRNA-specific probes. RNAs of interest can also be enriched. A common technique is to pull out polyadenlyated RNAs with oligo-dTs attached to magnetic beads. It is important to note that this technique yields mRNAs with different fates for eukaryotic and prokaryotic samples. In eukaryotes poly-A tails serve to stabilize transcripts, while in prokaryotes they serve as markers for degradation of transcripts.

In the second step, RNA transcripts are fragmented to obtain fragments that have a similar length as the sequencing reads. The majority of transcripts usually exceeds the maximum read length of short read sequencing technologies. If transcripts would not be fragmented, reads would only represent a short sequence (equal to the read length) from the beginning or the end of a transcript and the middle part of longer transcripts could not be sequenced. Thus, fragments stretching over different positions of a transcript lead to a clearer picture of the sequenced transcript. After the fragmentation, the RNA is reverse transcribed into single-stranded cDNA copies and then the cDNA is converted to double-stranded DNA.

In the third step, adapters are ligated to the double-stranded cDNA. The adapters are used to enable attachment of the DNA molecules to predefined positions of the flow cell inside the sequencing machine. In an optional step, the DNA can now be amplified before

the actual sequencing begins (Chao et al., 2019; *TruSeq DNA Sample Preparation Guide* 2022).

1.3.3 Paired-end reads and circular RNAs

In addition to the single-end sequencing protocols, where an RNA fragment is sequenced from its 5'-end, most sequencing technologies also offer paired-end sequencing protocols, where an RNA fragment is both sequenced from its 5'-end and its 3'-end, which results in a read pair. The first read represents the start of the fragment and the second one its end. Paired-end reads have the advantage over single-end reads that they improve mapping accuracy, because the reads of a pair are mapped together and thus represent the complete fragment instead of only one end of a fragment as in single-read protocols. Short paired-end reads have also been shown to outperform long single-end reads in regard to gene expression analysis (Freedman et al., 2020). After read alignment, mapped paired-end reads can be merged to represent the complete fragment, resulting in fragments that start from the beginning of the first read until the end of the second read. However, this approach is only suitable for linear RNA, but not for circular RNA (circRNA).

CircRNA is a type of single-stranded RNA found in all domains of life. Their common characteristic is that their 5'-end is covalently bound to their 3'-end, forming a closed circle of RNA. CircRNAs mainly modulate gene expression or translation of regulatory proteins but have also be found to be translated into protein (Yu and Kuo, 2019). When circRNAs are fragmented during library preparation, linear RNAs can emerge, where the ordering of their exons is reversed in comparison to the genome (Jeck and Sharpless, 2014). This reversed order will also result in a changed order of the first read and the second read, so that the second read aligns up-stream of the first read (Figure 1.3). This special case must be taken care of, when paired-end reads are merged to fragments.

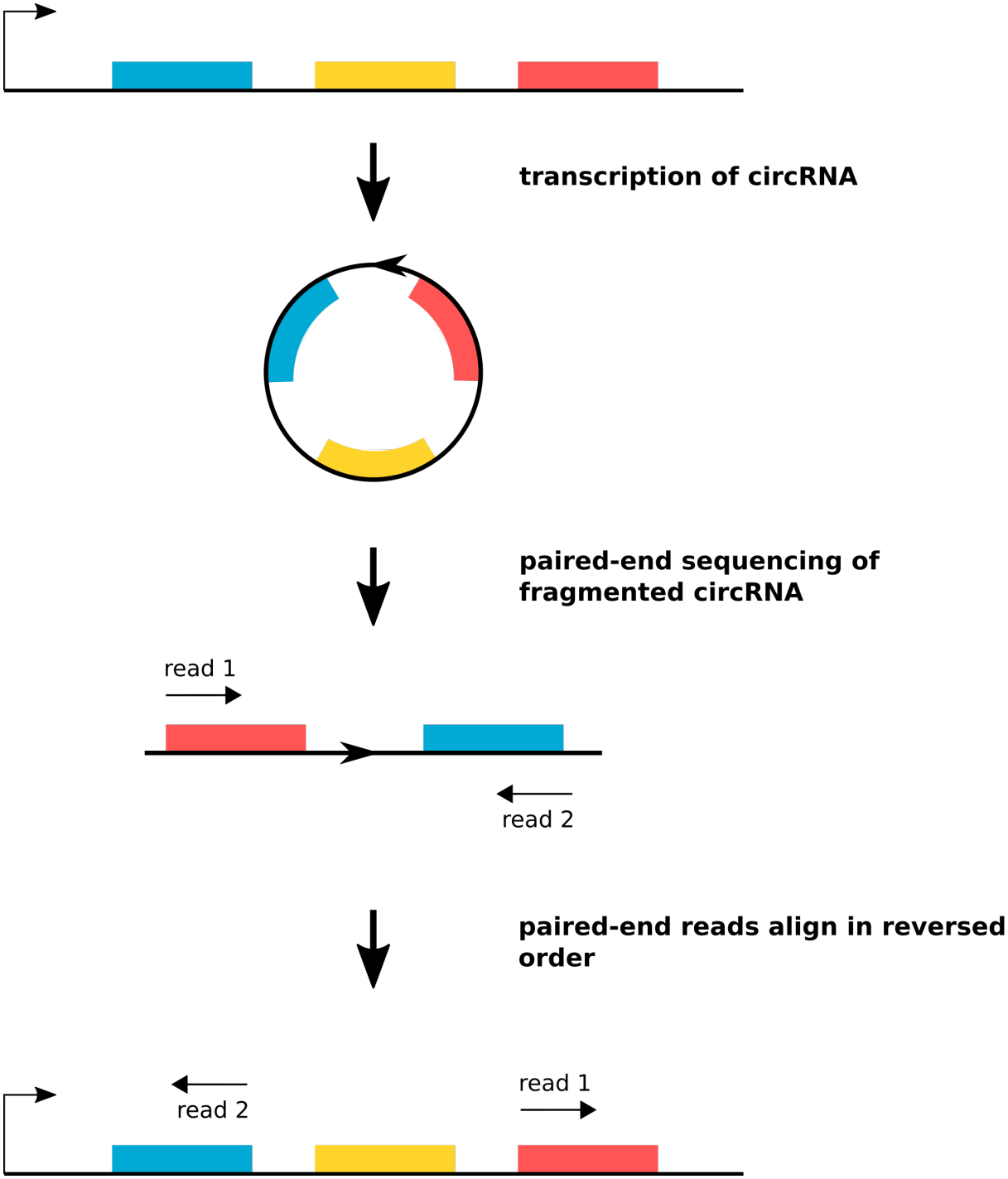

Figure 1.3: **Mechanism of paired-end sequencing of a circRNA resulting in a reversed order of aligned reads.** Three exons (blue, yellow and red) are transcribed into a circRNA (arrows indicate 5' to 3' direction of the genome and the transcript). The circRNA is fragmented during RNA-seq library preparation to a linear RNA with reversed exon order. After sequencing, read 1 and read 2 map in reverse order, where read 2 aligns upstream of read 1.

1.4 Dual and multi RNA sequencing

Dual RNA-seq is used to investigate the transcriptomes of two interacting species simultaneously. If more than two species are investigated, the term multi RNA-seq is used. Both methods work according to the same principal: The total RNA of samples that contain two (or more) interacting species is extracted. Then, the mixed RNA undergoes library preparation and RNA-sequencing, resulting in read files that contain reads from both (or more) species. Apart from their individual sequence, the reads do not contain any information about their origin species. The reads are only assigned *in silico* to their corresponding species and genomic position after sequencing. An overview of a dual RNA-seq workflow is depicted in Figure 1.4.

1.4.1 Applications and challenges of dual RNA sequencing

Since the first application of dual RNA-seq to a eukaryotic pathogen and host system in 2012 (Tierney et al., 2012), the number of publications performing dual RNA-seq steadily increased over the years (Figure 2 B in manuscript of chapter 3). The method has been applied to study a variety of host-pathogen, mutualistic and commensal interaction systems, however the majority of studies focused on host-pathogen systems involving eukaryotic hosts and prokaryotic pathogens (Westermann et al., 2017; Wolf et al., 2018).

Compared to conventional RNA-seq, which can only investigate one species at a time, dual RNA-seq gives a clearer picture of the interspecies interactions. For example, Westermann et al. (2016) performed dual RNA-seq of *Salmonella enterica* serovar Typhimurium and human host cells, using time course samples that were generated at different times after infection. An interspecies correlation analysis of the pathogen and host transcriptome changes over time identified bacterial and human genes that had similar expression kinetics across the time course of the infection. Thus, it was possible to link bacterial genes playing an important role during infection to the host's response. Another advantage of dual RNA-seq comes from the joint sequencing library preparation for both species. In contrast to conventional RNA-seq, where library preparation and sequencing has to be done once for every species being investigated, the joint approach of dual RNA-seq is cheaper, because costly library preparation and sequencing have to be done only once for each sample.

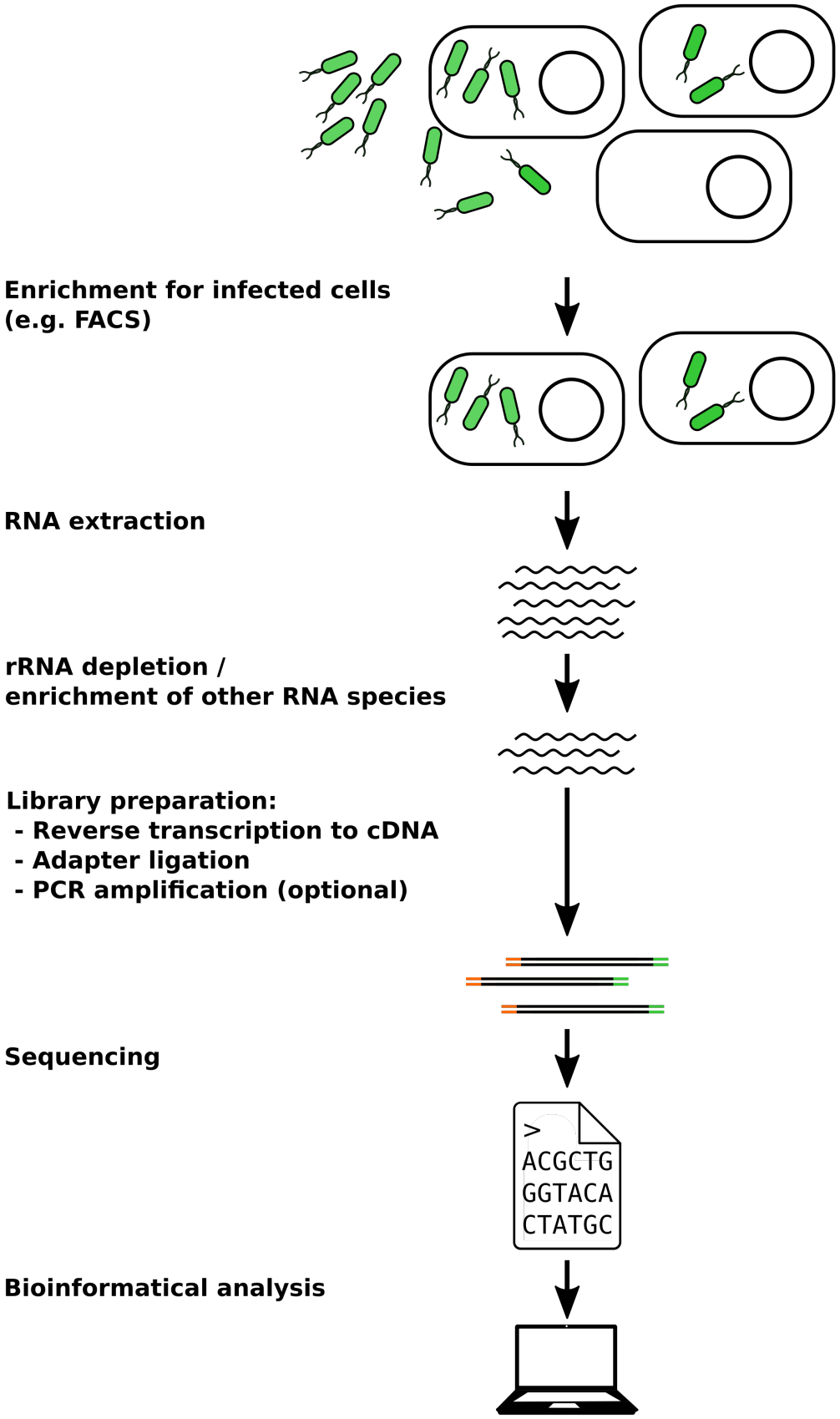

Figure 1.4: **Dual RNA-seq experiment workflow.** Workflow of a dual RNA-seq experiment with prokaryotic (green) and eukaryotic (white) cells: Samples are enriched for infected cells via fluorescence-activated cell sorting (FACS). Afterwards total RNA of all species is extracted, rRNA is depleted and other RNA-species are enriched. Then, library preparation takes place, followed by RNA-sequencing of cDNA. The resulting reads are subjected to bioinformatical analysis.

However, dual RNA-seq also entails some challenges caused by the different nature and mass of RNA of the interacting partners. Typical dual RNA-seq experiments investigate eukaryotic pathogens and mammalian host cells, which have a large difference in the total amount of RNA. While eukaryotic cells contain 10 to 20 picogram (pg) of RNA, bacterial cells contain around 0.1 pg. Taking into account that on average each infected eukaryotic cell is infected by ten bacterial cells, the mass of eukaryotic RNA in an infection experiment is 10 to 20 times higher than the bacterial one (Westermann et al., 2012). In order to guarantee that there is sufficient prokaryotic RNA to be sequenced, the following measures can be taken: Prior to RNA extraction from mixed species samples, infected cells can be separated from non-infected by-stander cells using laser microdissection (Vannucci et al., 2013) or fluorescence-activated cell sorting (FACS) (Avraham et al., 2015; Westermann et al., 2016). Prokaryotic RNA can be selectively enriched by depleting poly-A tailed mRNAs, due to the fact that prokaryotic RNAs, in contrast to eukaryotic RNAs, rarely possess poly-A tails (Humphrys et al., 2013). Depletion of rRNA of both species also increases the sequencing sensitivity of other RNA classes. Finally, increasing sequencing depth overall improves detection of low expressed genes (Westermann et al., 2017).

1.4.2 Bioinformatical analysis of dual and multi RNA sequencing

The bioinformatical analysis workflow of dual RNA-seq or multi RNA-seq is very similar to conventional RNA-seq. In the following, a brief overview is given before the steps are explained in detail. In the first step, raw reads are pre-processed by quality filtering and adapter trimming. Afterwards, the remaining, processed reads are aligned to reference sequences of all species that are part of the experiment. The alignment gives information about the species origin of each read and the genomic positions it aligns to. The following analysis steps are executed species-wise, meaning every step is done separately for each species. Standard analyses after alignment include gene quantification, DGE analysis and generation of nucleotide-wise coverage files. In order to perform DGE analysis, gene quantification has to be carried out before. Gene quantification sums up the amount of aligned reads for every genomic feature (e.g. genes, exons, coding DNA sequence (CDS), etc.) that they overlap with. The calculated counts per gene are the basis for DGE analysis. Nucleotides-wise coverage files are created by summing up the amount of reads that overlap with each genomic position. The above described analysis steps are standard in most experiments aiming to investigate gene expression. When many expressed genes

are found, gene set enrichment analysis (GSEA) is considered a common follow-up analysis step.

1.4.2.1 Pre-processing

The first step of raw read pre-processing before alignment is the removal of adapter sequences, called adapter trimming. Sequencing-by-synthesis RNA-seq methods add sequencing adapters to the ends of actual transcripts that need to be removed before alignment. Since the adapter sequences are synthetic and not part of any genome sequence, they complicate read alignment and can ultimately cause unaligned reads. Various tools exist to handle adapter trimming, e.g. *cutadapt* (Martin, 2011), *ngShoRT* (Chen et al., 2014), or *FastqPuri* (Pérez-Rubio et al., 2019). Next, reads can be trimmed by their sequencing quality. Sequencing machines are able to report base-calling error probabilities for each sequenced nucleotide. Nucleotides that fall beneath a certain threshold of base call accuracy are removed from the ends of the read. After adapter and quality trimming, reads can be filtered by their length, to avoid very short reads that due to their short sequence would be aligned to multiple genomic locations. A good minimum read length is around 20 nucleotides, since the majority of sRNAs exceeds this limit.

1.4.2.2 Read alignment

During read alignment, aligners find the best position for every read inside the genome reference sequences. One common format used to store every single alignment is the text-based format sequence alignment/map (SAM), which can be converted to its binary equivalent binary alignment/map (BAM) (Li et al., 2009). The majority of alignments is represented as a single line, containing information about each matching or mismatching base, insertions and deletions and other details of the alignment. Reads that can be aligned equally well to multiple locations produce an alignment entry for every location. Thus, reads can be classified as uniquely aligned or multiple aligned, depending on whether they align to a single location or to multiple locations. Regarding dual RNA-seq or multi RNA-seq it is important to additionally identify which of the multiple aligned reads align to different species, to be able to exclude them in the following analysis steps and to verify that species cross-aligned reads are low in numbers. Furthermore, care must be taken that splice-aware aligners, like *STAR* (Dobin et al., 2013), *HISAT2* (Kim et al., 2019) or *segemehl* (Hoffmann et al., 2014), which can align transcripts over splice junctions are used in experiments with eukaryotic species. For experiments, which only

investigate prokaryotic species aligners that are not splice-aware like *Bowtie* (Langmead et al., 2009) or *BWA* (Li and Durbin, 2009) are also suitable, because splicing rarely occurs in prokaryotes (Reinhold-Hurek and Shub, 1992).

1.4.2.3 Nucleotide-wise coverage

The genomic location of each read generated during read alignment can be presented as the number of reads that overlap with every single nucleotide of the genome reference sequences. The files that store this information are called coverage files and are usually in WIG (wiggle) file format (*ENSEMBL - WIG File Format - https://www.ensembl.org/info /website/upload/wig.html - 2022-10-07* 2022), which is a text-based format that consists of two columns, where the first column indicates the sorted genomic positions of the nucleotides of the reference sequence and the second the number of overlapping reads with the nucleotide of the first column. The read counts can be normalized e.g. by read depth, which is the total number of aligned reads for a given sequencing library and thus enable visual semi-quantitative comparison of transcripts of different libraries. Genome browsers like Integrated Genome Browser (IGB) (Freese et al., 2016) or Integrative Genomics Viewer (IGV) (Robinson et al., 2011) can be used to view coverage files. Coverage files are also useful for analyses that require exact transcript profiles, e.g. discovery of transcription start sites and processing start sites via differential RNA-seq (Sharma et al., 2010)

1.4.2.4 Gene quantification

Already known genomic locations and functions of annotated genes are stored in publicly accessible databases e.g. the *RefSeq* database (Pruitt et al., 2007). A common format to store annotation is the general feature format (GFF)3 format (*ENSEMBL - GFF3 File Format - https://www.ensembl.org/info/website/upload/gff3.html, 2022-10-07* 2022), which contains information about the strand specific location and type of each feature as well as a custom section that holds additional information of a feature e.g. unique identifiers for annotation databases. Comparing the genomic positions of every aligned read received from read alignment and the genomic positions of already annotated features, gene quantification sums up the number of reads that overlap with each feature. To account for differences in gene length and read depth, raw read counts are normalized by commonly used methods like transcripts per million (TPM) and reads per kilobase million (RPKM) or fragments per kilobase million (FPKM), when working with paired-end reads.

Although these methods have been used to draw conclusions about differentially expressed genes across different libraries, they are not reliable especially for lowly expressed genes and instead other methods (described in 1.4.2.5) should be used (Dillies et al., 2013; Zhao et al., 2021)

1.4.2.5 Differential gene expression analysis

DGE analysis aims to find differentially expressed genes between different biological conditions. Various methods and tools exist, which usually try to fit each expression value for a given gene into a particular distribution, like Poisson and negative binomial (*baySeq*, Hardcastle and Kelly, 2010; *DESeq2*, Love et al., 2014; *edgeR*, Robinson et al., 2010). In contrast to these parametric methods, tools that use non-parametric methods, like *SAM-seq* (Li and Tibshirani, 2013) and *NOIseq* (Tarazona et al., 2015) also exist. The different methods calculate fold changes, which indicate to which extend a gene is up- or down-regulated between two different conditions. The fold changes are usually accompanied by a value for each gene indicating its statistical significance (e.g. p-values). The p-values can be adjusted for multiple testing to control the false discovery rate of significantly regulated genes by applying correction methods such as Bonferroni or Benjamini-Hochberg, which can be further improved by introducing weights (Ignatiadis et al., 2016). To obtain statistically meaningful results, most tools require biological replicates of the individual conditions. When conducting RNA-seq experiments with the intention to perform DGE analysis, a decision has to be made whether the sequencing budget should be used for increasing sequencing depth or increasing the number of replicates. It was shown that increasing sequencing depth over certain thresholds gives diminishing returns for power of detecting differentially expressed genes, whereas increasing the number of replicates consistently increases detection power. Thus, increasing the number of replicates should be preferred over increasing sequencing depth when performing DGE analysis (Liu et al., 2014).

1.4.2.6 Gene set enrichment analysis

DGE analysis often yields long lists of differentially expressed genes that need to be interpreted. To avoid an impractically large amount of manual literature research, GSEA can help do identify differentially expressed sets of genes that share the same characteristics (Subramanian et al., 2005). These gene sets are defined *a priori* by known characteristics stored in databases. For example, Gene Ontology hosts annotations for genes regard-

ing their biological process, cellular component or molecular function (Ashburner et al., 2000), the Kyoto Encyclopedia of Genes and Genomes (KEGG) provides KEGG-terms that can be used to group genes by their biological pathway (Kanehisa et al., 2010) and Disease Ontology classifies genes by their association with human diseases (Osborne et al., 2009). In GSEA, genes are sorted by their level of expression changes, meaning the most up-regulated gene is at the top of the sorted list and the most down-regulated gene at the bottom of the list. Afterwards an enrichment score is calculated for the gene set that is investigated. The enrichment score represents the extent to which the genes of the set are overrepresented at the top or the bottom of the sorted list.

1.5 Aims of the study

The aims of this thesis were to carry out bioinformatical analysis of a dual RNA-seq data set generated from a host-pathogen system during infection and the development of scientific software to analyze dual RNA-seq and multi RNA-seq data. In chapter 1, dual RNA-seq data of three different infection settings of human mast cells and *S. aureus* were analyzed using bioinformatical methods like DGE and GSEA. The analysis was exploratory and aimed to gain new insides in the transcriptome changes on both the host's and the pathogen's side that take place during infection. In chapter 2, the aim was to further develop the RNA-seq analysis tool *READemption* to enable dual RNA-seq and multi RNA-seq in a convenient and user-friendly way. Additionally, further improvements, like fragment building for paired-end reads, adding TPM normalization to the gene quantification subcommand, increasing the system and unit test coverage and distributing the software as a Conda package have been carried out.

2 Chapter 1: Cytosolic sensing of intracellular Staphylococcus aureus by mast cells elicits a type I IFN response that enhances cell-autonomous immunity

2.1 Cover and author affiliations

Cytosolic Sensing of Intracellular Staphylococcus aureus by Mast Cells Elicits a Type I IFN Response That Enhances Cell-Autonomous Immunity

Oliver Goldmann[1,A] , Till Sauerwein[2,3,A], Gabriella Molinari[4], Manfred Rohde[4], Konrad U. Förstner[2,3,5] and Eva Medina[1]

[1]Infection Immunology Research Group, Helmholtz Centre for Infection Research, 38124 Braunschweig, Germany

[2]Institute for Molecular Infection Biology, University of Würzburg, 97080 Würzburg, Germany

[3]ZB MED-Information Centre for Life Science, 50931 Cologne, Germany

[4]Central Facility for Microscopy, Helmholtz Centre for Infection Research, 38124 Braunschweig, Germany

[5]TH Köln, University of Applied Sciences, Faculty of Information Science and Communication Studies, 50678 Cologne, Germany

[A]O.G. and T.S. contributed equally to this work

2.2 Publication

The following article has been published in *The Journal of Immunology*.

URL: `https://www.jimmunol.org/content/early/2022/03/23/jimmunol.2100622`

DOI: `https://doi.org/10.4049/jimmunol.2100622`

Accepted for publication on January 20, 2022.

Personal contribution: I conducted the bioinformatical analysis of the dual RNA-seq data set, including raw read pre-processing, alignment of processed reads and performing gene quantification, DGE, and GSEA for both species. I visualized the results of the mapping statistics and GSEA, and created the MA-plots. In order to make the bioinformatical analysis transparent and reproducible, I made the above described analysis publicly available at the *Repository for Life Sciences* (`https://repository.publisso.de/resource/frl:6427216`) (including all results, executable scripts and software containers) and uploaded the raw reads to the ENA (`https://www.ebi.ac.uk/ena/browser/view/PRJEB43874`).

The Journal of Immunology

Cytosolic Sensing of Intracellular *Staphylococcus aureus* by Mast Cells Elicits a Type I IFN Response That Enhances Cell-Autonomous Immunity

Oliver Goldmann,*[,1] Till Sauerwein,[†,‡,1] Gabriella Molinari,[§] Manfred Rohde,[§] Konrad U. Förstner,[†,‡,¶] and Eva Medina*

Strategically located at mucosal sites, mast cells are instrumental in sensing invading pathogens and modulating the quality of the ensuing immune responses depending on the nature of the infecting microbe. It is believed that mast cells produce type I IFN (IFN-I) in response to viruses, but not to bacterial infections, because of the incapacity of bacterial pathogens to internalize within mast cells, where signaling cascades leading to IFN-I production are generated. However, we have previously reported that, in contrast with other bacterial pathogens, *Staphylococcus aureus* can internalize into mast cells and therefore could trigger a unique response. In this study, we have investigated the molecular cross-talk between internalized *S. aureus* and the human mast cells HMC-1 using a dual RNA sequencing approach. We found that a proportion of internalized *S. aureus* underwent profound transcriptional reprogramming within HMC-1 cells to adapt to the nutrients and stress encountered in the intracellular environment and remained viable. HMC-1 cells, in turn, recognized intracellular *S. aureus* via cGMP–AMP synthase–STING–TANK-binding kinase 1 signaling pathway, leading to the production of IFN-I. Bacterial internalization and viability were crucial for IFN-I induction because inhibition of *S. aureus* internalization or infection with heat-killed bacteria completely prevented the production of IFN-I by HMC-1 cells. Feeding back in an autocrine manner in *S. aureus*–harboring HMC-1 cells and in a paracrine manner in noninfected neighboring HMC-1 cells, IFN-I promoted a cell-autonomous antimicrobial state by inducing the transcription of IFN-I–stimulated genes. This study provides unprecedented evidence of the capacity of mast cells to produce IFN-I in response to a bacterial pathogen. ***The Journal of Immunology*****, 2022, 208: 1675–1685.**

Mast cells are important effector cells of the innate immune system and contribute to the early host defense against pathogens (1–3). They are present in practically all tissues and are predominantly located at sites that interface with the external environment, such as mucosal surfaces, as well as in s.c. tissue in close proximity to blood vessels. Mast cells, therefore, may be among the first immune cells encountering invading pathogens and initiating the ensuing immune response. They are equipped with a variety of receptors, including TLRs, and several Fc and complement receptors that recognize specific bacterial components and enable them to tailor their response to the pathogen that they encounter (4, 5). Mast cells have been shown to be essential for containing pathogens at the sites of infection and prevent further dissemination (1, 2). They also play a major role in initiating both innate and adaptive immune responses to many bacterial pathogens (3).

A prominent feature of mast cells is the presence of abundant secretory granules in the cytoplasm, which contain large amounts of preformed mediators, including serotonin, histamine, heparin, TNF-α, and enzymes such as tryptase and chymase, and are rapidly released following activation (6). The release of preformed mediators initiates the recruitment and activation of effector immune cells to the sites of pathogen invasion (2). Mast cells can also release de novo synthesized mediators, such as proinflammatory leukotrienes, PGs, chemokines, and cytokines fitted to the specific pathogen (7). For example, although mast cells respond to dengue virus infection with the release of high amounts of CCL5 and low amounts of IL-1β and IL-6 (8), they produce large amounts of CCL20, IL-1α, IL-1β, CXCL8, and GM-CSF in response to *Pseudomonas aeruginosa* (9, 10). Furthermore, it has also been reported that, although mast cells can produce type I IFNs (IFN-Is) in response to viral infection, they elicit only proinflammatory cytokines, but not IFN-I responses, after infection

*Infection Immunology Research Group, Helmholtz Centre for Infection Research, 38124 Braunschweig, Germany; [†]Institute for Molecular Infection Biology, University of Würzburg, 97080 Würzburg, Germany; [‡]ZB MED-Information Centre for Life Science, 50931 Cologne, Germany; [§]Central Facility for Microscopy, Helmholtz Centre for Infection Research, 38124 Braunschweig, Germany; and [¶]TH Köln, University of Applied Sciences, Faculty of Information Science and Communication Studies, 50678 Cologne, Germany

[1]O.G. and T.S. contributed equally to this work.

ORCIDs: 0000-0003-4641-9782 (O.G.); 0000-0001-5830-4208 (T.S.); 0000-0002-6781-1292 (G.M.); 0000-0002-1481-2996 (K.U.F.); 0000-0002-5935-3260 (E.M.).

Received for publication June 24, 2021. Accepted for publication January 20, 2022.

This work was supported in part by the Helmholtz Center for Infection Research with a seed grant through funds from the Bavarian Ministry of Economic Affairs and Media, Energy and Technology (Grants 0703/68674/5/2017 and 0703/89374/3/2017) and in part by the Interdisciplinary Centre for Clinical Researchers Würzburg (Grant IZKF Z-6).

The raw read files presented in this article have been submitted to the European Nucleotide Archives (https://www.ebi.ac.uk/ena/browser/view/PRJEB43874) under accession number PRJEB43874. The bioinformatical workflow presented in this article has been submitted to the Repository for Life Sciences (https://repository.publisso.de/resource/frl:6427216).

Address correspondence and reprint requests to Prof. Eva Medina, Infection Immunology Research Group, Helmholtz Centre for Infection Research, Inhoffenstrasse 7, 38124 Braunschweig, Germany. E-mail address: eva.medina@helmholtz-hzi.de

The online version of this article contains supplemental material.

Abbreviations used in this article: CAT#, catalog number; cGAS, cGMP–AMP synthase; EM, electron microscopy; for., forward; HZI, Helmholtz Centre for Infection Research; IF, immunofluorescence; IFNAR, IFN-α/β receptor; IFN-I, type I IFN; IRF, IFN regulatory factor; KEGG, Kyoto Encyclopedia of Genes and Genomes; MOI, multiplicity of infection; PCA, principal-component analysis; RIG-I, retinoic acid–inducible gene I; RNA-seq, RNA sequencing; rev., reverse; sRNA, small RNA; TBK1, TANK-binding kinase 1.

www.jimmunol.org/cgi/doi/10.4049/jimmunol.2100622

with Gram-positive or Gram-negative bacteria (11). The authors argued that the lack of IFN-I responses was owed to the incapacity of bacterial pathogens to internalize within mast cells because signaling cascades leading to IFN-I production are triggered by receptors located in intracellular compartments (11). However, the incapacity to internalize into mast cells seems not to be a general phenomenon for all bacteria because we (12, 13) and others (14) have shown that the Gram-positive bacterium *Staphylococcus aureus* is capable of internalizing and surviving within mast cells. Mast cells are commonly found at sites of the body used as portals of entry by *S. aureus*, including the skin and the respiratory tract, and they will probably be one of the first cells of the innate immune system that sense and respond to this pathogen after invasion of the host. In previous studies, we reported that mast cells respond to *S. aureus* by releasing antimicrobial granule compounds, as well as extracellular trap in an attempt to kill the pathogen in an extracellular manner (12). However, *S. aureus* is able to subvert the extracellular antimicrobial mechanisms of the mast cells by promoting its internalization within these cells using α5β1 integrins expressed on the mast cell surface (12). In humans, mast cells harboring internalized *S. aureus* have been observed in nasal polyps isolated from patients with chronic rhinosinusitis (15). Because *S. aureus* internalization and intracellular survival could affect the ensuing response of the infected mast cells, the objective of this study was to investigate how the human mast cells HMC-1 and *S. aureus* respond to each other by assessing simultaneously gene expression changes taking place in the infected host cell and in the intracellular bacteria using a dual RNA sequencing (RNA-seq) approach (16). The results of this study highlight the plasticity of *S. aureus* to reprogram its transcriptional response to adapt to the intracellular environment and survive within HMC-1 cells. More importantly, we also show that intracellular viable *S. aureus* triggers the cytosolic DNA-sensing cGMP–AMP synthase (cGAS)–STING pathway within HMC-1 cells and leads to the production and release of IFN-I. Released IFN-Is act via the surface receptor, IFN-α/β receptor (IFNAR), in an autocrine fashion on the infected HMC-1 cells to enhance cell-autonomous host defenses and in a paracrine fashion to sensitize noninfected neighboring cells and thereby amplifying the immune response.

Materials and Methods

Cell lines

The human mast cell line HMC-1 was provided by J.H. Butterfield (Mayo Foundation for Medical Education and Research, Rochester, MN) (17).

Bacterial strains

The following *S. aureus* bacterial strains were used in this study: *S. aureus* strain SH1000 (18), *S. aureus* strain Newman (NCTC 8178), *S. aureus* strain 6850 (19), GFP-expressing *S. aureus* SH1000 (20), and *S. aureus* hla-deficient mutant strain (Δhla) (21). *Salmonella enterica* subsp. enterica serotype Typhimurium (NTCC 12023) was also used in this study. *S. aureus* strains were grown to midlog phase in brain-heart infusion medium (Roth) at 37°C with shaking (120 rpm), and *Salmonella typhimurium* was grown in lysogeny broth (Roth) also at 37°C with shaking. Bacteria were collected by centrifugation, washed with sterile PBS, and diluted to the required concentration.

For heat inactivation, bacteria were heated to 95°C for 2 h using an Eppendorf thermomixer.

Lysostaphin/gentamicin protection assay to assess intracellular viable bacteria

HMC-1 cells were adjusted to 2×10^6 cells/ml in IMDM (Life Technologies) supplemented with 5% FCS and infected with *S. aureus* at a multiplicity of infection (MOI) of five bacteria per one HMC-1 cell. After 2 h of infection, lysostaphin (2.5 μg/ml) (Sigma-Aldrich) was added and HMC-1 cells were incubated for 10 min to remove noninternalized extracellular bacteria. HMC-1 cells were then washed twice with sterile PBS and further incubated in medium containing 100 μg/ml gentamicin. At the indicated times, infected HMC-1 cells were centrifuged at 1500 × *g* for 5 min, and cells in the pellet were lysed by incubating them with 0.1% Triton X-100 in double-distilled H_2O for 5 min. The numbers of viable bacteria were determined by plating serial dilutions on blood agar plates. The cell culture supernatants were used for determination of IFN-α by ELISA.

In some experiments, HMC-1 cells were incubated 1 h before infection with 1 μg/ml of the irreversible STING inhibitor H-151 (InvivoGen) or with 100 nM for the TANK-binding kinase 1 (TBK1)/IKKε inhibitor BX-795 (Cayman Chemicals). Control HMC-1 cells were incubated with a similar concentration of vehicle DMSO. HMC-1 cells transfected with the retinoic acid–inducible gene I (RIG-I) ligand 5′ppp dsRNA using the transfection reagent LyoVec according to the manufacturer's instructions (Invivogen) were used to confirm that H-151 (1 μg/ml) is specific for STING and does not affect RIG-I signaling.

For blocking the IFNAR, HMC-1 cells were incubated in the presence of 500 ng/ml anti-IFNAR Ab or isotype-matching IgG Abs as control (Sigma-Aldrich).

In stimulation experiments, 5×10^3 IU/ml rIFN-α (Abcam) was added to HMC-1 cells 1 h before infection.

S. typhimurium *infection assay*

HMC-1 cells were infected with *S. typhimurium* at an MOI of 5:1 for 2 h. Gentamicin was then added at a concentration of 100 μg/ml to kill extracellular bacteria, and HMC-1 cells were further incubated at 37°C and 5% CO_2. After 24 h, HMC-1 cells were harvested, and supernatants were collected for determination of IFN-α.

Infection assay for RNA-seq analysis

HMC-1 cells were adjusted to 2×10^6 cells/ml in IMDM supplemented with 5% FCS and infected with *S. aureus* strain SH1000-GFP for 2 h at an MOI of 5:1. After 2 h of infection, 2.5 μg/ml lysostaphin was added, and HMC-1 cells were incubated for 10 min to remove noninternalized extracellular bacteria. HMC-1 cells were then washed twice with sterile PBS and further incubated for 24 h in medium containing 100 μg/ml gentamicin. HMC-1 cells harboring intracellular *S. aureus* (GFP^+) were separated from noninfected bystander HMC-1 cells (GFP^-) by FACS using a BD FACSAria III (Becton Dickinson) and resuspended in RNA*later* (Ambion). Sorted HMC-1 cells were centrifuged for 10 min at 1000 × *g*, washed twice with sterile prewarmed PBS, and carefully resuspended in 600 μl per 5×10^6 cells of cell lysis buffer included in mirVANA miRNA Isolation Kit (Ambion). Cell lysates were then transferred to FastPrep 24 lysing matrix tubes (mechanical lysis with FastPrep at 5′, 1000 × *g*), and RNA was isolated following the recommendations provided in the mirVANA miRNA Isolation Kit (Ambion).

rRNA depletion

RNA integrity was determined using a 2100 Bioanalyzer and the RNA 6000 Nano kit (Agilent Technologies, Santa Clara, CA). RNA integrity values for all samples ranged from 8.5 to 10.0. In accordance with the manufacturer's instructions, rRNA was depleted using Illumina's RiboZero Epidemiology Kit (Illumina). In brief, rRNA-specific biotinylated DNA probes were added to the total RNA. After hybridization of the probes and the rRNA, magnetic beads were added that bind to the rRNA–DNA hybrids. By placing the samples on a magnetic stand, the rRNA–DNA hybrids that are bound to magnetic beads were pulled down. The rRNA-depleted RNA was then purified using RNA Clean & Concentrator 5 kit (Zymo Research) following the manufacturer's protocol (manual version 2.2.1).

RNA fragmentation and cDNA library preparation

RNA was fragmented using NEB Next Magnesium RNA fragmentation module (New England Biolabs) following the manufacturer's protocol. The following modifications were introduced in the protocol: Mg^{2+} was used to fragment RNA for 3 min at 94°C using ABI 9700 PCR System. The fragmented RNA was purified with the RNA Clean & Concentrator kit 5 (Zymo Research), and RNA quality was determined using a 2100 Bioanalyzer and the RNA 6000 Pico kit (Agilent Technologies). Prior to adapter ligation, RNA was dephosphorylated at the 3′ end and phosphorylated at the 5′ end using 10 U T4-PNK ± 10 mM ATP (New England Biolabs). RNA was then decapped twice using 5 U RppH (New England Biolabs) following the manufacturer's protocol for eukaryotic cells and prokaryotic cells, respectively. RNA was purified with RNA Clean & Concentrator kit 5 (Zymo Research) after each enzymatic treatment as described earlier. cDNA synthesis was performed using NEBNext Small RNA Library Prep Set for Illumina (Illumina). In brief, RNA fragments were ligated to the 3′ SR and 5′ SR adapters prediluted 1:4 with nuclease-free water. PCR amplification to add Illumina adaptors and indices was performed for 15 cycles with 1:4 prediluted primers. Prior to sequencing, cDNA libraries were purified using the magnetic

MagSi-NGS^{PREP} Plus beads (magtivio) at a 1.8:1 ratio of beads to sample volume and afterward quantified with the Qubit 2.0 Fluorometer using Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific). The libraries' quality and size distribution were checked with a 2100 Bioanalyzer using HS DNA 7500 kit.

IFN-α ELISA

The amount of IFN-α was quantified in the culture supernatants using a human IFN-α Instant ELISA System according to the manufacturer's instructions (Invitrogen).

Quantitative RT-PCR

Total RNA was isolated from HMC-1 cells at the indicated time points using the GeneJET RNA purification kit (Fisher Scientific). RNA samples were reverse transcribed and amplified using a SensiFAST SYBR No-ROX Kit (Bioline) following the manufacturer's recommendations. The primers used for quantitative RT-PCR were for IFNA1 (IFN-α), forward [for.]: 5′-GTG AGG AAA TAC TTC CAA AGA ATC AC-3′, reverse [rev.]: 5′-TCT CAT GAT TTC TGC TCT GAC AA-3′; IFNB1 (IFN-β), for.: 5′-CGC CGC ATT GAC CAT CTA-3′, rev.: 5′-TTA GCC AGG AGG TTC TCA ACA ATA GTC TCA CTA-3′; and for the gene encoding β-actin (ACTB), for.: 5′-AAC TCC ATC ATG AAG TGT GAC G-3′; rev.: 5′-GAT CCA CAT CTG CTG GAA GG-3′. Thermal cycling conditions for IFNA1 and ACTB mRNA quantification consisted of reverse transcription for 20 min at 45°C, initial denaturation for 5 min at 95°C, followed by 40 cycles of 20 s at 95°C (denaturation), 20 s at 58°C (annealing), and 20 s at 72°C (elongation). Primers for RT-PCR quantification of selected IFN-I–induced genes mRNA and for RT-PCR quantification of TNFA (TNF-α) mRNA were purchased from OriGene and used following the conditions recommended by the manufacturer (OriGene). The following qPCR Primer Pairs were used: RSAD2 (Viperin) (NM_080657; catalog number [CAT#]: HP216708), IFN regulatory factor (IRF) 7 (NM_004031; CAT#: HP231979, IFI6 (NM_022873; CAT#: HP225644), IFI27 (NM_005532; CAT#: HP208651), MX2 (NM_002463; CAT#: HP206143), and TNF-α (TNF) (NM_000594; CAT#: HP200561). Data were normalized against the housekeeping gene β-actin. Fold change values were calculated by the Pfaffl equation, in which the expression ratio is estimated by (Etarget)ΔCt, target (control − experimental)/(Eref)ΔCt ref (control − experimental).

Inhibition of S. aureus *internalization within HMC-1 cells*

HMC-1 cells (2×10^6 cells/ml) were preincubated for 1 h with 1 μg/ml anti–β1-integrin blocking Abs (Santa Cruz biotechnology) or for 30 min with 5 μg/ml cytochalasin D (Sigma-Aldrich). Control cells received medium alone. HMC-1 cells were washed to remove unbound Abs or cytochalasin D and infected for 2 h with *S. aureus* at an MOI of 5:1. Lysostaphin was added at a concentration of 2.5 μg/ml for 10 min to eliminate noninternalized extracellular bacteria, and HMC-1 cells were washed and used either to determine the amount of intracellular viable bacteria as described earlier or further incubated for 24 h in medium containing 100 μg/ml gentamicin to determine the concentration of IFN-α in the culture supernatant.

Infection assay for microscopy

HMC-1 cells in suspension at a density of 1×10^6 cells/ml in IMDM supplemented with 5% FCS were infected with *S. aureus* at an MOI of 10:1 for the immunofluorescence (IF) staining and MOI 20:1 for electron microscopy (EM). At different infection times, parallel samples of infected and uninfected HMC-1 cells were fixed for IF or EM. Fixation was performed in the IF samples by adding the same volume of a 6% paraformaldehyde solution in PBS and incubating during 20 min at room temperature. Cells were centrifuged at $1000 \times g$ for 10 min, and the pellet was used for the IF labeling. Cells processed for EM were first centrifuged at $1000 \times g$ for 10 min, washed with PBS and resuspended in PBS, and immediately fixed for field emission scanning EM or transmission EM.

Confocal microscopy examination of a total of 45 HMC-1 cells was used to calculate the percentage of HMC-1 cells harboring internalized *S. aureus* and the mean number of bacteria per cell.

IF microscopy of HMC-1 cells in suspension

The staining of HMC-1 cells in suspension was performed following a modified protocol (22). HMC-1 cells were fixed as mentioned earlier and transferred to microcentrifuge tubes where the labeling was performed. Generally, after each step, cells were washed with 1200 μl of PBS, centrifuged at $1000 \times g$, and the supernatant was discarded by aspiration. The different labeling solutions were added to the pellet and after mixing, each labeling was performed on an Eppendorf thermomixer set at $700 \times g$ with the temperature control off. Cells were first washed with 900 μl of 10 mM glycine in PBS and after centrifugation were permeabilized with 0.1% Triton X-100 in PBS during 5 min and then washed twice with PBS. The pellet was resuspended in 100 ml of PBS, transferred to a fresh microcentrifuge tube, and 800 μl of 10% FBS-PBS was added for blocking during 45 min. After centrifugation, 120 μl of custom-produced anti–*S. aureus* rabbit serum diluted 1:100 was added to the cells and incubated during 1 h. After washing twice, HMC-1 cells were incubated with 1:500 secondary Ab Alexa Fluor 488–conjugated goat anti-rabbit (Thermo Fisher Scientific) for 45 min at room temperature. After washing twice, cells were stained with 10 μl of Alexa Fluor 633 phalloidin (Thermo Fisher Scientific) in 500 μl of PBS for 45 min and washed three times. ProLong Gold Antifade Mountant with DAPI (Thermo Fisher Scientific) was added to the pellet and carefully mixed. A total of 7 μl of sample was applied to the center of a 22 × 22-mm coverslip, and a microscope slice was placed on top. Mounted cells were allowed to dry overnight, and the edges of the coverslips were sealed before microscopic observation. Imaging was performed with a confocal laser-scanning upright microscope Leica SP5 equipped with an HC PL APO 63×/1,40 oil-immersion objective using three lasers, diode (405), argon (488 nm), and He-Ne (633 nm), and the LAS AF software. After the confocal laser-scanning upright microscope measurement, the image stacks were processed with Fiji-ImageJ.

Field emission scanning EM

HMC-1 cells were fixed with 4% paraformaldehyde, washed with TE buffer (20 mM Tris, 1 mM EDTA [pH 6.9]) and dehydrated after incubation with a graded series of ethanol (10, 30, 50, 70, 90, 100%) on ice for 15 min. HMC-1 cells were then critical-point dried with liquid CO_2 and covered with a gold film by sputter coating (SCD 40; Balzers Union). HMC-1 cells were examined in a field emission scanning electron microscope (Zeiss DSM 982 Gemini) using the Everhart Thornley SE detector and the inlens detector in a 50:50 ratio at an acceleration voltage of 5 kV.

Transmission EM

HMC-1 cells were fixed with 2% glutaraldehyde and 3% formaldehyde in cacodylate buffer for 1 h on ice, washed with cacodylate buffer, and osmificated with 1% aqueous osmium for 1 h at room temperature. HMC-1 cells were then dehydrated with a graded series of acetone (10, 30, 50, 70, 90, and 100%) for 30 min at each step. The 70% acetone dehydration step was performed in 2% uranyl acetate overnight. HMC-1 cells were then infiltrated with an epoxy resin, and ultrathin 70-nm sections were cut with a diamond knife. Sections were counterstained with uranyl acetate and lead citrate and examined in a TEM910 transmission electron microscope (Carl Zeiss) at an acceleration voltage of 80 kV. Images were taken at calibrated magnifications using a line replica and recorded digitally with a Slow-Scan CCD-Camera (ProScan) with ITEM Software (Olympus Soft Imaging Solutions). Brightness and contrast were adjusted with Adobe Photoshop CS3.

Bioinformatical procedure

Illumina reads were trimmed using cutadapt (version: 1.16) (23). Illumina's TruSeq "Read 1" adapter sequence was removed from the 3′ end. Nucleotides with a Phred quality score <20 and their following downstream (5′–3′) bases were also cut off. Further filtering steps including read mapping and downstream analysis, such as gene quantification, generation of coverage files, and differential gene expression analysis, were made by the RNA-seq tool READemption (version: 0.4.3, doi: 10.5281/zenodo.250598) (24). Additional reads filtering included clipping of poly(A) sequences and discarding of reads that had a read length <20 nucleotides after performing the trimming steps. The read mapping was performed using the short read mapper segemehl (version: 0.2.0) (25), which is integrated into READemption. The mapping was performed with an accuracy of 95% and segemehl's aligner lack (26). The human genome and annotation were obtained from GENCODE (version: 27, NCBI assembly name: GRCh38.p10) and the bacterial ones from NCBI's RefSeq database (accession number: NC_007795.1, https://www.ncbi.nlm.nih.gov/nuccore/88193823; RefSeq assembly accession number: GCF_000013425.1, https://www.ncbi.nlm.nih.gov/assembly/GCF_000013425.1). The *S. aureus* annotation was extended with small RNAs (sRNAs) predicted by ANNOgesic (27). Transcripts that were not associated with any of RefSeq's annotated features were determined as sRNA candidates based on their predicted folding energy. Candidates that had homologs in NCBI's nonredundant protein database (https://www.ncbi.nlm.nih.gov/refseq/about/nonredundantproteins) were discarded, while candidates with homologs in sRNA database BSRD (28) were accepted. The gene quantification files (i.e., the number of reads overlapping with an annotated feature) and the coverage files in wiggle format (i.e., the number of reads overlapping with each base of the genome) were created using READemption. Afterward both file types were split up by species. The coverage was normalized by the total number of aligned reads of a given

replicate and multiplied by 1,000,000. Differential gene expression analysis was performed with the R package DESeq2 (version: 1.20.0) based on raw read counting (29). Genes with an adjusted (Benjamini–Hochberg corrected) $p < 0.05$ were defined as differentially expressed.

Raw read files can be found at the European Nucleotide Archives under the project ID PRJEB43874 (https://www.ebi.ac.uk/ena/browser/view/PRJEB43874). The complete bioinformatical workflow is available at the Repository for Life Sciences (https://repository.publisso.de/resource/frl:6427216, doi: 10.4126/FRL01-006427216). A shell script can be executed step by step or in one go to reproduce the analysis. Singularity images are provided that contain all required programs.

Other data analysis

Heatmaps, hierarchical clustering dendrograms, and principal-component analysis (PCA) plots were generated using the corresponding function of the platform MetaboAnalyst v.3.0 (30). Gene lists of all significantly expressed genes between the different conditions were used as input for the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway (*S. aureus*) or Reactome (HMC-1) analysis using DAVID (31). Comparisons between groups were made using a parametric ANOVA test with Tukey's posttest or a *t* test. The *p* values <0.05 were considered significant.

Results

Transcriptional response of S. aureus *and HMC-1 cells under different infection settings*

In contrast with what has been previously reported for other Gram-positive bacteria (11), *S. aureus* is capable of internalizing within HMC-1 cells. The IF and EM photographs depicted in Fig. 1A show the capacity of *S. aureus* to adhere to the surface of HMC-1 cells (Fig. 1A, upper panel) and internalize within the HMC-1 cells (Fig. 1A, lower panel). In our experimental setting, ~66% of HMC-1 cells cells were found to harbor internalized *S. aureus* with a mean of 5.6 ± 6.5 bacteria per cell. Within HMC-1 cells, *S. aureus* could be found within membrane-bound vacuoles (Fig. 1B, lower panel, red arrows) or free in the cell cytosol (Fig. 1B, lower panel, insert). Evaluation of intracellular bacterial viability indicated that, although HMC-1 cells has the capacity to kill a proportion of the internalized *S. aureus*, a subpopulation of bacteria was capable of escaping the intracellular antimicrobial mechanisms of HMC-1 cells and remained viable after 24 h of infection (Fig. 1C).

A dual RNA-seq approach was then used to investigate the strategies used by *S. aureus* to survive within HMC-1 cells, as well as the functional consequences of harboring intracellular *S. aureus* for the HMC-1 cells responses. For this purpose, HMC-1 cells cells were infected with GFP-expressing *S. aureus* for 2 h, noninternalized bacteria were removed by lysostaphin treatment, and HMC-1 cells were further incubated for 24 h in the presence of antibiotics. HMC-1 cells harboring intracellular *S. aureus* (GFP^+) were then separated from noninfected bystander HMC-1 cells (GFP^-) cells by FACS and subjected to dual RNA-seq for parallel gene expression analysis of HMC-1 cells and intracellular *S. aureus*. The transcriptional response of uninfected HMC-1 cells and of *S. aureus* in the input infection inoculum were used as control for host and intracellular pathogen, respectively. We also determined the transcriptional response of noninfected bystander HMC-1 cells (GFP^-) cells, as well as of HMC-1 cells cocultured with *S. aureus* in separated chambers using a permeable transwell system. The different infection settings are summarized in the scheme depicted in Fig. 1D. Total RNA was isolated from the different samples and subjected to RNA-seq analysis. The distribution of RNA classes from HMC-1 cells indicated that between 40 and 50% of the HMC-1 cell reads mapped to coding sequences in the different samples (Fig. 1E). Regarding the RNA classes distribution from *S. aureus,* tRNAs were more represented in intracellular *S. aureus* than in *S. aureus* in the infection inoculum (Fig. 1F), probably suggesting a more active protein synthesis in the intracellular bacteria. The dual RNA-seq

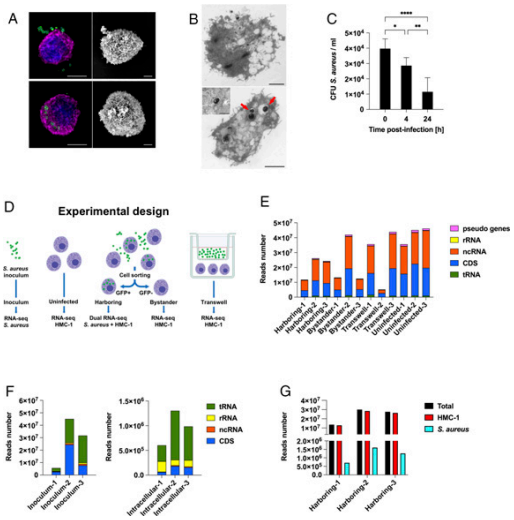

FIGURE 1. Experimental design and mapping of RNA-seq reads. (**A**) Confocal (left panels) and scanning electron microscope (right panels) photographs showing *S. aureus* attached to the surface of an HMC-1 cell at 1 h of infection (upper panels) and internalizing within HMC-1 cells at 2 h of infection (lower panels). HMC-1 cells were stained with Alexa Fluor 633 phalloidin for actin (magenta) and DAPI for DNA (blue) and *S. aureus* labeled with primary rabbit anti–*S. aureus* Ab followed by secondary Alexa Fluor 488–conjugated goat anti-rabbit Ab (green). Scale bars: 10 µm. (**B**) Transmission EM photographs showing *S. aureus* located within an HMC-1 cell at 2 h (lower panel) and 4 h (lower panel, insert) of infection. Bacteria can be found either in membrane-bound vacuoles (lower panel, red arrows) or free in the HMC-1 cells cytoplasm (lower panel, insert). An uninfected HMC-1 cell is shown in the upper panel for comparison. (**C**) Numbers of viable bacteria within HMC-1 cells at progressing times postinfection with *S. aureus* (MOI = 5). The data are presented as mean ± SD of three replicates from three independent experiments. $*p < 0.05$, $**p < 0.005$, $****p < 0.0001$. (**D**) Experimental design scheme. HMC-1 cells were infected with GFP-expressing *S. aureus* for 2 h, the remaining noninternalized bacteria were removed, and HMC-1 cells were further incubated for 24 h. HMC-1 cells "harboring" intracellular bacteria (GFP^+) were separated from noninfected "bystander" HMC-1 (GFP^-) cells by FACS sorter. HMC-1 cells were also cocultured with *S. aureus* in separated chambers in a "transwell" system. "Uninfected" HMC-1 cells and *S. aureus* in the infection "inoculum" were used as control for host cells and pathogen, respectively. Total RNA was isolated from the different samples and subjected to RNA-seq analysis. (**E**) Distribution of RNA-seq reads mapped to the human reference genome in each sample over the main RNA classes. (**F**) Distribution of RNA-seq reads mapped to the *S. aureus* reference genome over the main RNA classes in intracellular (right) and inoculum (left) *S. aureus*. (**G**) Number of reads mapped either to the human or to the *S. aureus* reference genome in RNA-seq libraries generated from HMC-1 cells harboring intracellular *S. aureus*. CDS, coding sequences; ncRNA, noncoding RNA.

analysis of HMC-1 cells harboring *S. aureus* showed that ~95% of the reads could be mapped to the human genome and 5% to the bacteria genome in each of the three replicates (Fig. 1G).

Intracellular survival of S. aureus *within HMC-1 cells is associated with metabolic reprogramming and upregulation of stress responses*

To gain a better understanding of the strategies used by *S. aureus* to survive and persist within HMC-1 cells, we compared the expression profile of protein coding genes from intracellular *S. aureus* with that of *S. aureus* in the infection inoculum. Hierarchical clustering (Fig. 2A), PCA (Fig. 2B), and heatmap of gene expression levels (Fig. 2C) showed a clear separation between the

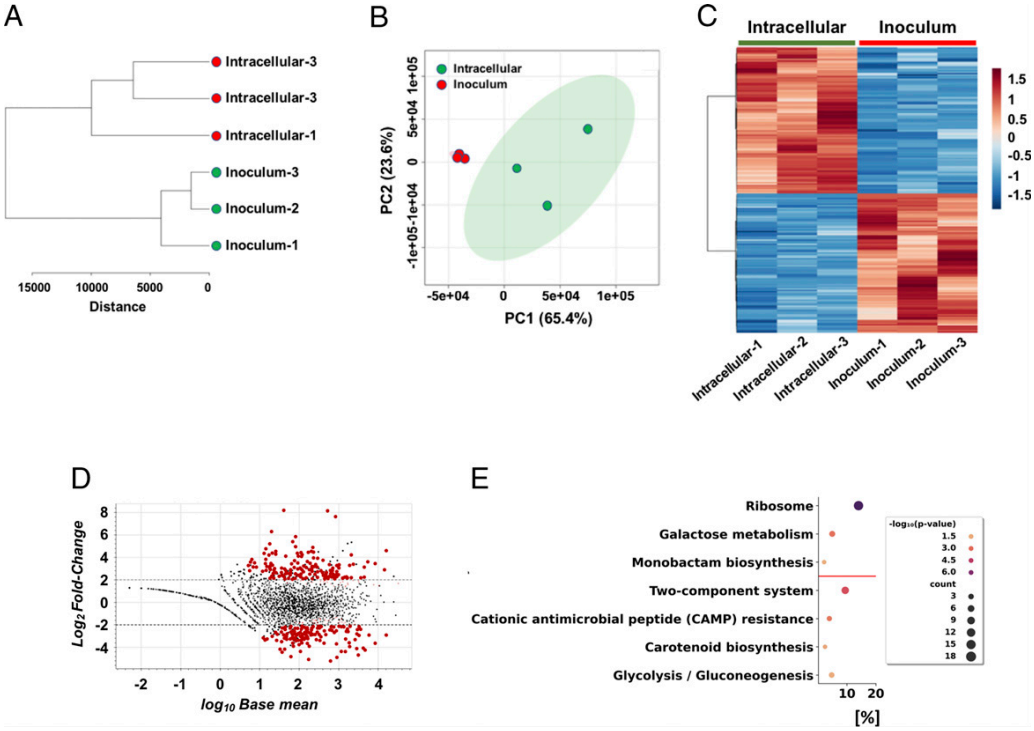

FIGURE 2. Analysis of gene expression in intracellular *S. aureus* versus *S. aureus* in the infection inoculum. (**A**) Hierarchical clustering dendrogram of intracellular *S. aureus* and *S. aureus* in the infection inoculum RNA-seq datasets based on Euclidean distance metric. (**B**) PCA of the RNA-seq datasets of intracellular *S. aureus* and *S. aureus* in the infection inoculum. Ellipse surrounds the 95% confidence limit of the centroid of the group. Replicates of the same samples group are indicated by the same color as shown in the legend. (**C**) Heatmap showing gene expression levels (top 200) in intracellular *S. aureus* and *S. aureus* in the infection inoculum. Color coding shows the *z* score normalized transcripts per million of each sample. (**D**) MA plots showing the transcripts abundance (log10 base mean) versus log2 fold change in gene expression between intracellular *S. aureus* and *S. aureus* in the infection inoculum. Genes with adjusted $p < 0.05$ and log2 fold change > 2 or log2 fold change < -2 are labeled in dark red. (**E**) KEGG pathways enriched in genes with significantly greater expression (over the red line) or with significantly lower expression (under the red line) in intracellular *S. aureus* versus *S. aureus* in the infection inoculum. The color of the dots reflects the *p* values calculated by DAVID software program using a modification of the Fisher's exact test, and the size of the dots reflects the number of genes in the pathway (count).

transcriptome datasets of *S. aureus* located within HMC-1 cells (intracellular) and *S. aureus* in the infection inoculum (inoculum). This indicated that *S. aureus* underwent profound remodeling of the transcriptional response on internalization within HMC-1 cells. We performed differential gene expression analysis on the RNA-seq datasets and focused on transcripts with differential expression of $\log_2$ fold change > 2 (upregulated) or $\log_2$ fold change < -2 (downregulated) (Benjamini-Hochberg adjusted $p < 0.05$) for further analysis. We found 143 genes upregulated and 126 downregulated by intracellular *S. aureus* in comparison with the bacteria in the inoculum, with many of them encoding hypothetical proteins (Fig. 2D, Supplemental Table I). KEGG pathway enrichment analysis of differentially expressed genes showed "ribosome," followed by "galactose metabolism" and "monobactam biosynthesis" as the most predominant enriched pathway in genes upregulated by intracellular *S. aureus* (Fig. 2E). Indeed, many genes encoding components of the protein translation machine, such as ribosomal proteins and ribonucleoproteins, were expressed to a significantly higher extent by intracellular *S. aureus* than by *S. aureus* in the input inoculum (Supplemental Table I). Interestingly, the genes of the lactose operon *lacABCD* operon, which are implicated in the catabolism of lactose and D-galactose, as well as the cotranscribed genes *lacFEG*, which encode the proteins for transport, phosphorylation, and cleavage of these carbon sources (32), were expressed by intracellular *S. aureus*, but not by the bacteria in the infection inoculum (Supplemental Table I). These genes are inducible by lactose or galactose (33) and repressed in the presence of glucose (34), indicating that lactose or galactose, but not glucose, are the carbon sources available to the bacterium in the intracellular compartment. Furthermore, the gene encoding the ROK family protein, which is involved in the metabolism of the amino sugar *N*-acetylglucosamine and the sialic acid *N*-acetylneuraminate (35), was also induced by *S. aureus* in the intracellular environment (Supplemental Table I). In addition to these pathways, genes involved in the stress response, such as the genes encoding components of the classical chaperones DnaK/DnaJ and GroES/GroEL (*dnaK*, *groES*, *groEL*) (36), as well as those coding for Clp chaperones (*clpB* and *clpC)* (37) and genes encoding virulence factors, such as superantigen-like protein SSL6 (*ssl6*), coagulase (*coa*), fibronectin-binding proteins (*fnbA*, *fnbB*), the extracellular matrix, and plasma binding protein Emp (*emp*), were also upregulated by intracellular *S. aureus* (Supplemental Table I).

KEGG pathway enrichment analysis of genes exhibiting lower expression in intracellular *S. aureus* than in *S. aureus* in the infection inoculum identified high enrichment of pathways involved in "two-components system" followed by "cationic antimicrobial peptide (CAMP) resistance," "carotenoid biosynthesis," and "glucolysis/gluconeogenesis" (Fig. 2E). The genes *dltB*, *dltD*, and *dltABCD*, which encode factors involved in cationic antimicrobial peptide resistance

(38) as well as *vraF*, which encodes part of the VraFG ABC transporter that potentially enhances export of cell wall/teichoic acid precursors (39) were also downregulated by intracellular *S. aureus*.

Transcriptional analysis reveals expression of IFN-I–induced genes in HMC-1 cells cells harboring intracellular S. aureus, *as well as in noninfected bystander HMC-1 cells*

To investigate the consequences of harboring intracellular *S. aureus* for the HMC-1 cells responses, we compared the gene expression profile of HMC-1 cells harboring intracellular bacteria with the gene expression profile of either noninfected bystander HMC-1 cells, uninfected HMC-1 cells, or HMC-1 cells cocultured with *S. aureus* but separated by a transwell system. Hierarchical clustering of the transcriptome of HMC-1 cells in the different infection settings showed that *S. aureus*–harboring HMC-1 and noninfected bystander HMC-1 cell samples clustered together but away from uninfected HMC-1 and transwell HMC-1 cell samples (Fig. 3A). This clustering was also reflected by the heatmap depicted in Fig. 3B showing the pattern of gene expression across the samples. The results of these analyses indicated that the transcriptional response of *S. aureus*–harboring HMC-1 cells was highly similar to that of noninfected bystander HMC-1 cells but significantly different from the

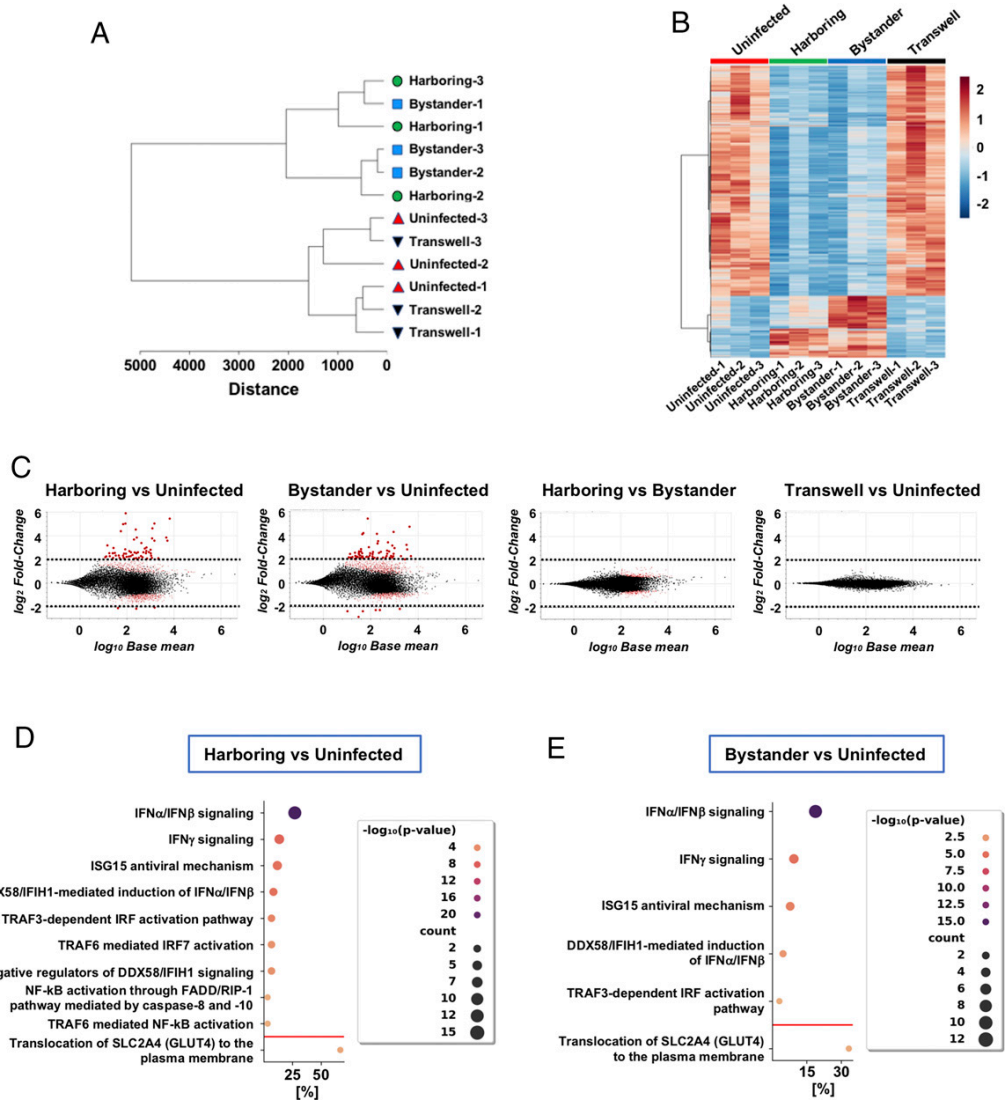

FIGURE 3. Gene expression analysis of HMC-1 cells under different infection conditions. (**A**) Hierarchical clustering dendrogram of RNA-seq datasets from HMC-1 under different infection conditions based on Euclidean distance metric. (**B**) Heatmap showing gene expression levels (top 500) in HMC-1 cells under different infection conditions. Color coding shows the *z* score normalized transcripts per million of each sample. (**C**) MA plots showing the transcripts abundance ($\log_{10}$ base mean) versus $\log_2$ fold change in gene expression for the indicated transcriptomes comparisons. Genes with adjusted $p < 0.05$ and $\log_2$ fold change > 2 or $\log_2$ fold change < -2 are labeled in dark red. (**D**) Enriched Reactome pathways in genes with significantly greater expression and $\log_2$ fold change > 2 (over the red line) or with significantly lower expression and $\log_2$ fold change < -2 (under the red line) in HMC-1 cells harboring intracellular *S. aureus* versus uninfected HMC-1 cells. (**E**) Enriched Reactome pathways in genes with significantly greater expression and $\log_2$ fold change > 2 (over the red line) or with significantly lower expression and $\log_2$ fold change < -2 (under the red line) in HMC-1 bystander versus uninfected HMC-1 cells. The color of the dots in (D) and (E) reflects the *p* values calculated by DAVID software program using a modification of the Fisher's exact test, and the size of the dots reflects the number of genes in the pathway (count).

transcriptional response of uninfected or transwell samples. The fact that the gene expression profile of HMC-1 cells in a transwell system, where they are separated from *S. aureus* by a permeable membrane, did not differ from that of uninfected HMC-1 cells excluded a potential effect of soluble factors released by *S. aureus* on the transcriptional response of HMC-1 cells.

To get further insights into the transcriptional changes taking place in HMC-1 cells harboring intracellular *S. aureus* and in noninfected bystander HMC-1 cells, we performed differential gene expression analysis of these samples in comparison with uninfected HMC-1 cells. The results of this analysis showed 59 genes with significantly higher expression ($\log_2$ fold change > 2, adjusted $p < 0.05$) and 3 genes with significantly lower expression ($\log_2$ fold change < -2, adjusted $p < 0.05$) in *S. aureus*–harboring HMC-1 cells with respect to uninfected HMC-1 cells (Fig. 3C, Supplemental Table II) and 66 genes with significantly greater expression and 6 genes with significantly lower expression in noninfected bystander HMC-1 cells in comparison with uninfected HMC-1 cells (Fig. 3C, Supplemental Table III). No differentially expressed genes with $\log_2$ fold change > 2, adjusted $p < 0.05$ or $\log_2$ fold change < -2, adjusted $p < 0.05$ were identified between *S. aureus*–harboring HMC-1 cells and noninfected bystander HMC-1 cells (Fig. 3C). Likewise, no differentially expressed genes with adjusted $p < 0.05$ were found between transwell HMC-1 cells and uninfected HMC-1 cells (Fig. 3C). Reactome pathway enrichment analysis performed in differentially expressed genes with greater expression in *S. aureus*–harboring HMC-1 cells than in uninfected HMC-1 cells indicated a robust transcriptional signature related to genes induced by IFN-I (Figs. 3D, 4A). A similar overlapping IFN-I–induced transcriptional response was observed in noninfected bystander HMC-1 cells in comparison with uninfected HMC-1 cells (Figs. 3E, 4A). The induction of IFN-I target genes in *S. aureus*–infected HMC-1 cells was confirmed by RT-PCR (Fig. 4B).

IFN-Is comprise a family of highly pleiotropic cytokines that includes IFN-α and IFN-β (40). Because the induction of IFN-I target genes in *S. aureus*–harboring HMC-1 and noninfected bystander HMC-1 cells was observed in the transcriptional analysis performed after 24 h of infection, we speculated that IFN-I proteins may already be present in the culture supernatant at this time of infection; consequently, the induction of the genes encoding IFN-I may take place at earlier times of infection. To investigate whether this is the case, we first determined the expression levels of the genes encoding IFN-α and IFN-β in HMC-1 cells at 2 and 4 h postinfection by RT-PCR. The results show that both genes were induced at 2 h postinfection and their level of expression substantially increased at 4 h postinfection, although the gene encoding IFN-α was expressed to a significantly greater extent than the gene encoding IFN-β (Fig. 4C). In addition to IFN-I, NF-κB target genes, such as TNF-α, were also upregulated by HMC-1 cells in response to *S. aureus* infection (Fig. 4D). At the protein level, significant amounts of IFN-α were detectable in the supernatant of *S. aureus*–infected HMC-1 cells at 24 h of infection, but not in the supernatant from uninfected HMC-1 cells, HMC-1 cells cocultured with *S. aureus* in a transwell system, or HMC-1 cells infected with *Salmonella typhimurium*, which have been previously reported to be incapable of eliciting IFN-I in human mast cells (11) (Fig. 4E). Bacterial viability was required for the production of IFN-I by HMC-1 cells because IFN-α was under detection levels in the supernatant of HMC-1 cells incubated with heat-inactivated *S. aureus* (Fig. 4E). We also demonstrated that IFN-I production by HMC-1 cells was not bacterial strain dependent because they produced significant amounts of IFN-α not only postinfection with *S. aureus* strain SH1000, which is the strain that has been used in all the earlier-described experiments, but also postinfection with *S. aureus* strain Newman and strain 6850 (Fig. 4F).

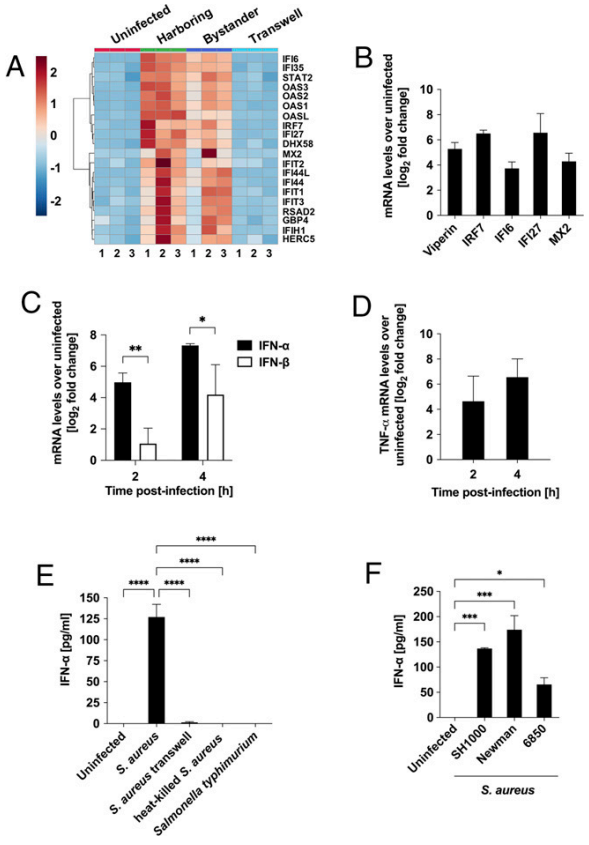

FIGURE 4. Production of IFN-I by HMC-1 cells in response to *S. aureus*. (**A**) Heatmap showing expression levels of IFN-I target genes in HMC-1 cells under different infection conditions. Color coding shows the *z* score normalized transcripts per million of each sample. (**B**) mRNA levels of selected IFN-I target genes in *S. aureus*–infected HMC-1 cells at 24 h postinfection determined by RT-PCR. Values are expressed as $\log_2$ fold change between the mRNA levels in infected versus uninfected HMC-1 cells. (**C**) Expression levels of the gene encoding IFN-α and of the gene encoding IFN-β in HMC-1 cells at 2 and 4 h postinfection with *S. aureus* determined by RT-PCR. Values are expressed as $\log_2$ fold change of gene expression between infected and uninfected HMC-1 cells. (**D**) Levels of TNF-α gene expression in *S. aureus*–infected HMC-1 cells at 2 and 4 h postinfection determined by RT-PCR. Values are expressed as $\log_2$ fold change between the mRNA levels in infected versus uninfected HMC-1 cells. (**E**) Levels of IFN-α in the supernatant of HMC-1 cells either uninfected or after 24 h of infection with either viable or heat-killed *S. aureus*, cocultured with *S. aureus* in separated chambers in a transwell system or infected with *S. typhimurium*. (**F**) Levels of IFN-α in the supernatant of HMC-1 cells either uninfected or after 24 h of infection with *S. aureus* strain SH1000, *S. aureus* strain Newman, or *S. aureus* strain 6850. The data are presented as mean ± SD of three replicates from three independent experiments. $^*p < 0.05$, $^{**}p < 0.005$, $^{***}p < 0.001$, $^{****}p < 0.0001$.

Production of IFN-α by HMC-1 cells requires S. aureus *internalization and involves the cGAS–STING signaling pathway*

We next explored the signaling pathway leading to IFN-I induction in *S. aureus*–infected HMC-1 cells. Cytosolic signaling pathways, such as the cGAS–STING pathway that recognizes DNA (41) and RIG-I that recognizes RNA (42), have emerged as the major sensing systems driving IFN-I responses. Because these pathways are largely

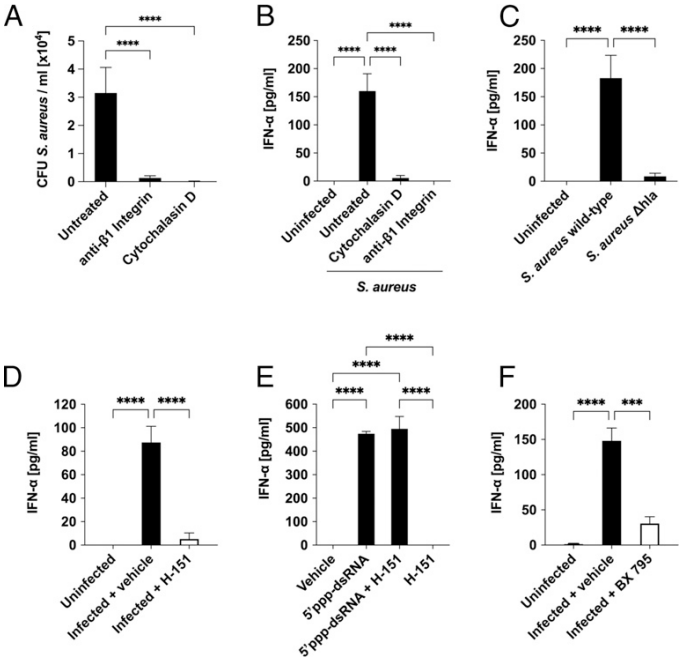

FIGURE 5. Production of IFN-α by HMC-1 cells requires *S. aureus* internalization and involves the cytosolic cGAS–STING signaling pathway. (**A**) Quantification of *S. aureus* bacteria internalized within untreated HMC-1 cells or treated with anti–β1-integrin Abs or with cytochalasin D. HMC-1 cells were infected with *S. aureus* for 2 h, treated with lysostaphin/gentamicin to kill extracellular bacteria, washed, and the amount of internalized viable bacteria was determined 2 h thereafter after lysis of HMC-1 cells. (**B**) Levels of IFN-α in the supernatant of *S. aureus*–infected HMC-1 cells (24 h postinfection) either untreated or treated with anti–β1-integrin Abs or with cytochalasin D. (**C**) Levels of IFN-α in the supernatant of HMC-1 cells after 24 h of infection with *S. aureus* wild-type or *S. aureus* Δhla mutant strain. (**D**) Levels of IFN-α in the supernatant of *S. aureus*–infected HMC-1 cells (24 h postinfection) treated with the STING inhibitor H-151 (1 μg/ml) or with vehicle alone. (**E**) Levels of IFN-α in the supernatant of HMC-1 cells at 24 h after transfection with the RIG-1 agonist 5′ppp-dsRNA and incubated in the presence or absence of H-151 (1 μg/ml). (**F**) Levels of IFN-α in the supernatant of *S. aureus*–infected HMC-1 cells cells (24 h postinfection) treated with the TBK1 inhibitor BX 795 (100 nM) or with vehicle alone. The data are presented as mean ± SD of three replicates from three independent experiments. ***$p < 0.001$, ****$p < 0.0001$.

triggered by recognition of pathogen-derived nucleic acids in the cell cytosol, we first determined the requirement of bacterial internalization for IFN-I production by HMC-1 cells. Inhibition of *S. aureus* internalization using either the actin polymerization inhibitor cytochalasin D or β1-integrin blocking Abs prevented *S. aureus* internalization within HMC-1 cells (Fig. 5A) as previously reported (12, 13) and resulted in complete abrogation of IFN-α production (Fig. 5B). Furthermore, HMC-1 cells failed to produce IFN-α postinfection with a *S. aureus* mutant strain deficient in the production of α-hemolysin (Δhla), which has been reported to be impaired in its capacity to internalize and survive within mast cells (13) (Fig. 5C). Because *S. aureus* has been reported to activate IFN-I responses in macrophages via the cGAS–STING pathway (43), we next explored the relevance of this pathway in the production of IFN-I by infected HMC-1 cells. In the cGAS–STING pathway, pathogen-derived DNA present in the cell cytosol binds to the cGAS, resulting in conformational changes that induce enzymatic activity (44). Activation of cGAS leads to the production of the second messenger cGMP–AMP, which binds to the endoplasmic reticulum–localized adaptor protein STING. After activation, STING translocates from the endoplasmic reticulum to the Golgi, where it recruits kinases such as TANK-binding kinase 1 (TBK1), which phosphorylates IRF3 and triggers the expression of IFN-I (41). STING can also directly bind bacterial c-di-AMP in the host cytosol and induce an IFN-I response (45, 46). To determine the potential involvement of the cGAS–STING pathway in the production of IFN-I by *S. aureus*–infected HMC-1 cells, we blocked this pathway using the STING-specific inhibitor H-151 (47). As shown in Fig. 5D, treatment with H-151 almost completely abrogated the production of IFN-α by *S. aureus*–infected HMC-1 cells. Treatment with H-151 did not affect the capacity of HMC-1 cells to produce IFN-α after stimulation of the alternative signaling pathway RIG-I with the agonist 5′ppp-dsRNA (Fig. 5E). These results corroborated the specificity of H-151 for STING inhibition.

Furthermore, we also demonstrated that treatment with BX795, a potent inhibitor of TBK1 (48), resulted in profound reduction of IFN-α production by *S. aureus*–infected HMC-1 cells (Fig. 5F). These results indicated that the cGAS–STING–TBK1 axis was involved in the induction of IFN-I in HMC-1 cells by *S. aureus*.

IFN-Is enhance HMC-1 cell-autonomous immunity

Although the concerted activation of IFN-I–stimulated genes is a key component of the innate immune response against viruses (49), it has become increasingly evident that they also play an important role in the control of intracellular bacterial pathogens (50). IFN-I molecules bind to a common surface receptor named IFNAR, which comprises two subunits, IFNAR1 and IFNAR2, forming a ternary complex that leads to the activation of the Jak tyrosine kinase 2 and Jak1 (51). After activation, these kinases propagate downstream signaling leading to the activation of transcription factors such as STAT1 and STAT2 that after dimerization translocate to the nucleus, where they assemble with IRFs and mediate the transcription of a large number of IFN-I–stimulated genes involved in cell-autonomous immunity (51).

To determine the relevance of IFN-I–induced response on the capacity of HMC-1 cells to control intracellular *S. aureus*, we disrupted IFN-I signaling by blocking IFNAR1 with specific Abs. Disruption of IFN-I/IFNAR1 signaling did not influence the amount of *S. aureus* internalizing within HMC-1 cells cells (Fig. 6A, 0 h), but reduced considerably the capacity of HMC-1 cells to control intracellular *S. aureus* because significantly higher numbers of intracellular *S. aureus* were detected in HMC-1 after inhibition of IFN-I/IFNAR signaling in comparison with untreated HMC-1 cells at 2, 4, and 24 h postinfection (Fig. 6A). To discard that the effect of blocking IFNAR1 on the capacity of HMC-1 cells to reduce intracellular *S. aureus* was due to an unspecific effect of the Ab, we determined the level of expression of a set of IFN-I–induced genes in *S. aureus*–infected HMC-1 cells treated with either anti-IFNAR1 Abs or with an isotype-matched (IgG) control Ab. As shown in Fig. 6B, whereas the expression levels of the genes encoding IFI27, IFR7, and MX2 in *S. aureus*–infected HMC-1 cells treated with isotype control Abs were comparable to those observed in untreated

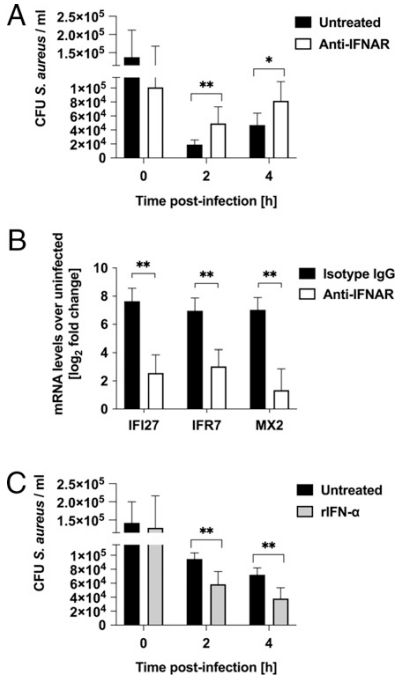

FIGURE 6. IFN-Is enhance HMC-1 cell-autonomous immunity. (**A**) Quantification of viable *S. aureus* within untreated HMC-1 cells (black bars) or treated with anti-IFNAR blocking antibodies (white bars). (**B**) Levels of IFI27, IFR7, and MX2 mRNA in *S. aureus*–infected HMC-1 cells at 24 h of infection either pretreated with anti-IFNAR blocking antibodies or with isotype-matching IgG1 control determined by RT-PCR. Values are expressed as $\log_2$ fold change between the mRNA levels in infected versus uninfected HMC-1 cells. (**C**) Quantification of viable *S. aureus* within HMC-1 cells either untreated (black bars) or treated with rIFN-α (5×10^3 IU/ml) (gray bars). The data are presented as mean ± SD of three replicates from three independent experiments. $^*p < 0.05$, $^{**}p < 0.01$.

S. aureus–infected HMC-1 cells (Fig. 4B), the expression levels of these genes were significantly lower in *S. aureus*–infected HMC-1 cells treated with anti-IFNAR1 Abs. These results corroborate the specific effect of anti-IFNAR1 blocking Abs.

We also investigated the effect of stimulating HMC-1 cells with rIFN-α prior to infection on their capacity to control intracellular *S. aureus*. As shown in Fig. 6C, treatment with rIFN-α enhanced the capacity of HMC-1 cells to control intracellular *S. aureus* because they exhibited significantly lower numbers of intracellular viable bacteria than untreated HMC-1 cells.

Altogether, these results indicated that IFN-α released by HMC-1 cells harboring intracellular *S. aureus* signaled back in an autocrine manner resulting in the induction of IFN-I target genes and improved cell-autonomous host defenses. The observation that IFN-I target genes were also upregulated in bystander HMC-1 that did not harbor intracellular *S. aureus* indicated that IFN-I released by *S. aureus*–harboring HMC-1 cells signaled also in a paracrine manner to induce an IFN-I signature in these cells.

Discussion

Mast cells are generally located at host sites used by *S. aureus* for invasion of the host; therefore, they may be among the first innate immune cells recognizing and fighting this pathogen. We have previously reported the capacity of murine and human HMC-1 cells mast cells to recognize extracellular *S. aureus* and respond by releasing extracellular traps and antimicrobial compounds in an attempt to immobilize and kill the pathogen (52). We have also reported that *S. aureus* was able to induce its own internalization within mast cells to escape the extracellular antimicrobial mechanisms of these cells (12, 13). In this study, we show that HMC-1 cells responded to *S. aureus* internalization by activating intracellular antimicrobial defense mechanisms that resulted in a significant reduction of intracellular bacteria within a few hours after bacterial internalization. However, a subpopulation of internalized *S. aureus* was capable of circumventing these antimicrobial mechanisms and survived within HMC-1 cells for long periods. Therefore, the interactions between *S. aureus* and HMC-1 cells during infection involve a series of events as each part deploys mechanisms of defense and survival. Because the outcome of these interactions can influence the ensuing immune response, we investigated in this study how the human mast cells HMC-1 cells and *S. aureus* respond to each other by assessing simultaneously gene expression changes taking place in the infected HMC-1 cells and in the harbored bacteria using dual RNA-seq analysis.

The results of the bacterial gene expression analysis indicated that, to survive within HMC-1 cells, *S. aureus* undergoes profound transcriptional reprogramming to readjust its metabolism to the nutritional changes and to counteract the stress conditions encountered in the intracellular niche. Thus, the genes encoding enzymes and transport systems involved in the galactose/lactose and D-tagatose-6-phosphate metabolic pathways were upregulated in intracellular *S. aureus* in comparison with the bacteria in the infection inoculum, whereas the genes associated with glycolysis were downregulated. This is interesting because *S. aureus* is one of the few microorganisms known to exclusively use enzymes of the D-tagatose-6-phosphate pathway to metabolize D-galactose, which is imported into the bacterial cell by a transport system encoded by genes *lacFEG* and metabolized by proteins encoded by the lactose operon, *lacABCD* (53). The lac operon has been shown to be inducible by the presence of D-galactose or lactose (33), suggesting that these sugars may be the carbon source available to the bacteria within the HMC-1 cells. Furthermore, the gene encoding the ROK family protein, which is involved in the metabolism of the amino sugar *N*-acetylglucosamine and the sialic acid *N*-acetylneuraminate by *S. aureus* (35), was also induced in the intracellular environment. We speculated that peptidoglycans, which are complex macromolecules comprising disaccharides, such as *N*-acetyl-glucosamine and galactose, and are abundant within mast cells because they play an important role in the tight packaging of compounds within secretory granules (54), could provide a source of galactose and amino sugars for intracellular *S. aureus*. Increased expression of the genes belonging to the heat shock stimulon, including the DnaK and GroESL chaperones, was also observed in intracellular *S. aureus*, most probably required for the bacteria to deal with the highly stressful conditions encountered within HMC-1 cells.

On the host cell side, we observed that HMC-1 cells produced IFN-I in response to internalized *S. aureus*. The production of IFN-I by HMC-1 in response to *S. aureus* contrasts with another study where the authors claimed that only viruses and not bacterial pathogens can induce an IFN-I response in mast cells because of the incapacity of bacteria to internalize into these cells (11). In that study, the authors used the Gram-positive *Listeria monocytogenes* and *Streptococcus pyogenes* and the Gram-negative *Salmonella typhimurium* in their mast cells infection assays (11). The results of our study indicate that this is not the case for all bacterial pathogens but probably only for those that fail to internalize within mast cells. Indeed, inhibition of *S. aureus* internalization within HMC-1 cells after treatment with cytochalasin or β1-integrin blocking Abs or infection of HMC-1 cells using a mutant *S. aureus* strain unable to

internalize within HMC-1 cells (13) completely prevented the production of IFN-I. Our study, therefore, provides compelling evidence that mast cells can indeed produce IFN-I in response to bacterial infection and argues against the concept that mast cells can elicit IFN-I responses only to viral infections as previously reported (11). IFN-I produced and released by HMC-1 cells harboring intracellular *S. aureus* can then bind to the IFNAR either on the same infected HMC-1 cells cells in an autocrine fashion or on noninfected bystander neighboring MHC-1 cells in a paracrine way, resulting in the induction of a large number of IFN-I–stimulated genes. It has been reported that the product of these IFN-I–stimulated genes contributes to enhanced cell-autonomous host defense against intracellular pathogens in infected cells (55). In our study, the autocrine stimulation of IFN-I–stimulated response in infected HMC-1 cells seems to contribute, at least to some extent, to the proper control of internalized *S. aureus* because interfering with IFN-α/IFNAR signaling using blocking Abs significantly reduced the capacity of HMC-1 cells to kill intracellular *S. aureus* and resulted in much lower expression of IFN-I–induced genes. The transcriptional analysis also indicated that IFN-I signaled in a paracrine manner in noninfected bystander HMC-1 cells and induced IFN-I–stimulated genes, probably instructing them to enter a state of enhanced resistance toward *S. aureus*. Indeed, pretreatment of HMC-1 cells with rIFN-α increased the capacity of these cells to control intracellular *S. aureus*.

We also found that the cGAS–STING–TBK1 signaling pathway was involved in the recognition of intracellular *S. aureus* by HMC-1 cells and in the induction of IFN-I. The role of STING in detection of cytosolic DNA, such as those from viral or bacterial infections, is well known (56–58). In bacterial infections, STING-dependent induction of IFN-I has been reported for both intracellular and extracellular pathogens (58). In the particular case of *S. aureus*, cGAS–STING signaling activated an IFN-I response in macrophages after infection with live but not killed bacteria (43). This is in line with our data showing that HMC-1 cells incubated with heat-killed *S. aureus* failed to produce IFN-α. Activation of STING in infected HMC-1 cells can ensue either after recognition of bacterial DNA or most probably through its direct activation by c-di-AMP produced by *S. aureus*. In this regard, it has been reported that c-di-AMP released from *S. aureus* biofilms can activate STING and induce an IFN-I response in macrophages (59).

In summary, the results of this study provide a scenario where, after invasion of the host, mast cells recognize extracellular *S. aureus*, most probably via pattern recognition receptors on the cell surface or by sensing bacterial toxins such as δ toxin as previously reported (60), and respond by undergoing degranulation with the concomitant discharge of prepackaged antimicrobial compounds or by releasing extracellular traps to kill the extracellular bacteria (52). *S. aureus*, in turn, induces its own internalization within mast cells, most probably to escape their extracellular killing mechanisms, and establishes a survival niche within these cells. *S. aureus*–infected mast cells sense the intracellular bacteria by cytosolic receptors and produce IFN-Is that act in an autocrine manner to enhance cell-autonomous host defense in the infected mast cells and in a paracrine way to sensitize neighboring cells and amplify the immune response. Our study thus has provided important information about the strategy used by mast cells to recognize *S. aureus* and how they contribute to the induction and propagation of an antimicrobial immune response.

Acknowledgments

We thank Sabine Beyer (Infection Immunology Research Group/Helmholtz Centre for Infection Research), Ina Schleicher and Melanie Tillig (Central Facility for Microscopy/Helmholtz Centre for Infection Research), and Elena Katzowitsch (Core Unit Systems Medicine Würzburg) for excellent technical assistance.

Disclosures

The authors have no financial conflicts of interest.

References

1. Dudeck, A., M. Köberle, O. Goldmann, N. Meyer, J. Dudeck, S. Lemmens, M. Rohde, N. G. Roldán, K. Dietze-Schwonberg, Z. Orinska, et al. 2019. Mast cells as protectors of health. *J. Allergy Clin. Immunol.* 144(4S): S4–S18.
2. Urb, M., and D. C. Sheppard. 2012. The role of mast cells in the defence against pathogens. *PLoS Pathog.* 8: e1002619.
3. Abraham, S. N., and A. L. St John. 2010. Mast cell-orchestrated immunity to pathogens. *Nat. Rev. Immunol.* 10: 440–452.
4. Marshall, J. S. 2004. Mast-cell responses to pathogens. *Nat. Rev. Immunol.* 4: 787–799.
5. Sandig, H., and S. Bulfone-Paus. 2012. TLR signaling in mast cells: common and unique features. *Front. Immunol.* 3: 185.
6. Wernersson, S., and G. Pejler. 2014. Mast cell secretory granules: armed for battle. *Nat. Rev. Immunol.* 14: 478–494.
7. Moon, T. C., A. D. Befus, and M. Kulka. 2014. Mast cell mediators: their differential release and the secretory pathways involved. *Front. Immunol.* 5: 569.
8. King, C. A., R. Anderson, and J. S. Marshall. 2002. Dengue virus selectively induces human mast cell chemokine production. *J. Virol.* 76: 8408–8419.
9. Lin, T. J., R. Garduno, R. T. Boudreau, and A. C. Issekutz. 2002. *Pseudomonas aeruginosa* activates human mast cells to induce neutrophil transendothelial migration via mast cell-derived IL-1 alpha and beta. *J. Immunol.* 169: 4522–4530.
10. Lin, T. J., L. H. Maher, K. Gomi, J. D. McCurdy, R. Garduno, and J. S. Marshall. 2003. Selective early production of CCL20, or macrophage inflammatory protein 3alpha, by human mast cells in response to *Pseudomonas aeruginosa*. *Infect. Immun.* 71: 365–373.
11. Dietrich, N., M. Rohde, R. Geffers, A. Kröger, H. Hauser, S. Weiss, and N. O. Gekara. 2010. Mast cells elicit proinflammatory but not type I interferon responses upon activation of TLRs by bacteria. *Proc. Natl. Acad. Sci. USA* 107: 8748–8753.
12. Abel, J., O. Goldmann, C. Ziegler, C. Höltje, M. S. Smeltzer, A. L. Cheung, D. Bruhn, M. Rohde, and E. Medina. 2011. *Staphylococcus aureus* evades the extracellular antimicrobial activity of mast cells by promoting its own uptake. *J. Innate Immun.* 3: 495–507.
13. Goldmann, O., L. Tuchscherr, M. Rohde, and E. Medina. 2016. α-Hemolysin enhances Staphylococcus aureus internalization and survival within mast cells by modulating the expression of β1 integrin. *Cell. Microbiol.* 18: 807–819.
14. Rocha-de-Souza, C. M., B. Berent-Maoz, D. Mankuta, A. E. Moses, and F. Levi-Schaffer. 2008. Human mast cell activation by *Staphylococcus aureus*: interleukin-8 and tumor necrosis factor alpha release and the role of Toll-like receptor 2 and CD48 molecules. *Infect. Immun.* 76: 4489–4497.
15. Hayes, S. M., R. Howlin, D. A. Johnston, J. S. Webb, S. C. Clarke, P. Stoodley, P. G. Harries, S. J. Wilson, S. L. Pender, S. N. Faust, et al. 2015. Intracellular residency of *Staphylococcus aureus* within mast cells in nasal polyps: A novel observation. *J. Allergy Clin. Immunol.* 135: 1648–1651.
16. Westermann, A. J., K. U. Förstner, F. Amman, L. Barquist, Y. Chao, L. N. Schulte, L. Müller, R. Reinhardt, P. F. Stadler, and J. Vogel. 2016. Dual RNA-seq unveils noncoding RNA functions in host-pathogen interactions. *Nature* 529: 496–501.
17. Butterfield, J. H., D. Weiler, G. Dewald, and G. J. Gleich. 1988. Establishment of an immature mast cell line from a patient with mast cell leukemia. *Leuk. Res.* 12: 345–355.
18. Horsburgh, M. J., J. L. Aish, I. J. White, L. Shaw, J. K. Lithgow, and S. J. Foster. 2002. sigmaB modulates virulence determinant expression and stress resistance: characterization of a functional rsbU strain derived from *Staphylococcus aureus* 8325-4. *J. Bacteriol.* 184: 5457–5467.
19. Fraunholz, M., J. Bernhardt, J. Schuldes, R. Daniel, M. Hecker, and B. Sinha. 2013. Complete genome sequence of *Staphylococcus aureus* 6850, a highly cytotoxic and clinically virulent methicillin-sensitive strain with distant relatedness to prototype strains. *Genome Announc.* 1: e00775-13.
20. Pollitt, E. J. G., P. T. Szkuta, N. Burns, and S. J. Foster. 2018. *Staphylococcus aureus* infection dynamics. *PLoS Pathog.* 14: e1007112.
21. O'Reilly, M., J. C. de Azavedo, S. Kennedy, and T. J. Foster. 1986. Inactivation of the alpha-haemolysin gene of *Staphylococcus aureus* 8325-4 by site-directed mutagenesis and studies on the expression of its haemolysins. *Microb. Pathog.* 1: 125–138.
22. Wang, C. C., S. S. Bajikar, L. Jamal, K. A. Atkins, and K. A. Janes. 2014. A time- and matrix-dependent TGFBR3-JUND-KRT5 regulatory circuit in single breast epithelial cells and basal-like premalignancies. *Nat. Cell Biol.* 16: 345–356.
23. Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal* 17: 10–12 .
24. Förstner, K. U., J. Vogel, and C. M. Sharma. 2014. READemption-a tool for the computational analysis of deep-sequencing-based transcriptome data. *Bioinformatics* 30: 3421–3423.
25. Hoffmann, S., C. Otto, S. Kurtz, C. M. Sharma, P. Khaitovich, J. Vogel, P. F. Stadler, and J. Hackermüller. 2009. Fast mapping of short sequences with mismatches, insertions and deletions using index structures. *PLOS Comput. Biol.* 5: e1000502.

26. Otto, C., P. F. Stadler, and S. Hoffmann. 2014. Lacking alignments? The next-generation sequencing mapper segemehl revisited. *Bioinformatics* 30: 1837–1843.
27. Yu, S. H., J. Vogel, and K. U. Förstner. 2018. ANNOgesic: a Swiss army knife for the RNA-seq based annotation of bacterial/archaeal genomes. *Gigascience* 7: giy096.
28. Li, L., D. Huang, M. K. Cheung, W. Nong, Q. Huang, and H. S. Kwan. 2013. BSRD: a repository for bacterial small regulatory RNA. *Nucleic Acids Res.* 41(D1): D233–D238.
29. Love, M. I., W. Huber, and S. Anders. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15: 550.
30. Xia, J., I. V. Sinelnikov, B. Han, and D. S. Wishart. 2015. MetaboAnalyst 3.0–making metabolomics more meaningful. *Nucleic Acids Res.* 43(W1): W251–W257 .
31. Huang, D.W., B. T. Sherman, and R. A. Lempicki. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4: 44–57.
32. Rosey, E. L., B. Oskouian, and G. C. Stewart. 1991. Lactose metabolism by *Staphylococcus aureus*: characterization of lacABCD, the structural genes of the tagatose 6-phosphate pathway. *J. Bacteriol.* 173: 5992–5998.
33. Morse, M. L., K. L. Hill, J. B. Egan, and W. Hengstenberg. 1968. Metabolism of lactose by *Staphylococcus aureus* and its genetic basis. *J. Bacteriol.* 95: 2270–2274.
34. Oskouian, B., and G. C. Stewart. 1990. Repression and catabolite repression of the lactose operon of *Staphylococcus aureus*. *J. Bacteriol.* 172: 3804–3812.
35. Olson, M. E., J. M. King, T. L. Yahr, and A. R. Horswill. 2013. Sialic acid catabolism in *Staphylococcus aureus*. *J. Bacteriol.* 195: 1779–1788.
36. Gottesman, S., S. Wickner, and M. R. Maurizi. 1997. Protein quality control: triage by chaperones and proteases. *Genes Dev.* 11: 815–823.
37. Frees, D., U. Gerth, and H. Ingmer. 2014. Clp chaperones and proteases are central in stress survival, virulence and antibiotic resistance of *Staphylococcus aureus*. *Int. J. Med. Microbiol.* 304: 142–149.
38. Peschel, A., M. Otto, R. W. Jack, H. Kalbacher, G. Jung, and F. Götz. 1999. Inactivation of the dlt operon in *Staphylococcus aureus* confers sensitivity to defensins, protegrins, and other antimicrobial peptides. *J. Biol. Chem.* 274: 8405–8410.
39. Meehl, M., S. Herbert, F. Götz, and A. Cheung. 2007. Interaction of the GraRS two-component system with the VraFG ABC transporter to support vancomycin-intermediate resistance in *Staphylococcus aureus*. *Antimicrob. Agents Chemother.* 51: 2679–2689.
40. Pestka, S., C. D. Krause, and M. R. Walter. 2004. Interferons, interferon-like cytokines, and their receptors. *Immunol. Rev.* 202: 8–32.
41. Ishikawa, H., Z. Ma, and G. N. Barber. 2009. STING regulates intracellular DNA-mediated, type I interferon-dependent innate immunity. *Nature* 461: 788–792.
42. Rehwinkel, J., and M. U. Gack. 2020. RIG-I-like receptors: their regulation and roles in RNA sensing. *Nat. Rev. Immunol.* 20: 537–551.
43. Scumpia, P. O., G. A. Botten, J. S. Norman, K. M. Kelly-Scumpia, R. Spreafico, A. R. Ruccia, P. K. Purbey, B. J. Thomas, R. L. Modlin, and S. T. Smale. 2017. Opposing roles of Toll-like receptor and cytosolic DNA-STING signaling pathways for *Staphylococcus aureus* cutaneous host defense. *PLoS Pathog.* 13: e1006496.
44. Civril, F., T. Deimling, C. C. de Oliveira Mann, A. Ablasser, M. Moldt, G. Witte, V. Hornung, and K. P. Hopfner. 2013. Structural mechanism of cytosolic DNA sensing by cGAS. *Nature* 498: 332–337.
45. Barker, J. R., B. J. Koestler, V. K. Carpenter, D. L. Burdette, C. M. Waters, R. E. Vance, and R. H. Valdivia. 2013. STING-dependent recognition of cyclic di-AMP mediates type I interferon responses during *Chlamydia trachomatis* infection. *MBio* 4: e00018-13.
46. Moretti, J., S. Roy, D. Bozec, J. Martinez, J. R. Chapman, B. Ueberheide, D. W. Lamming, Z. J. Chen, T. Horng, G. Yeretssian, et al. 2017. STING senses microbial viability to orchestrate stress-mediated autophagy of the endoplasmic reticulum. *Cell* 171: 809–823.e13.
47. Haag, S. M., M. F. Gulen, L. Reymond, A. Gibelin, L. Abrami, A. Decout, M. Heymann, F. G. van der Goot, G. Turcatti, R. Behrendt, and A. Ablasser. 2018. Targeting STING with covalent small-molecule inhibitors. *Nature* 559: 269–273.
48. Feldman, R. I., J. M. Wu, M. A. Polokoff, M. J. Kochanny, H. Dinter, D. Zhu, S. L. Biroc, B. Alicke, J. Bryant, S. Yuan, et al. 2005. Novel small molecule inhibitors of 3-phosphoinositide-dependent kinase-1. *J. Biol. Chem.* 280: 19867–19874.
49. Schneider, W. M., M. D. Chevillotte, and C. M. Rice. 2014. Interferon-stimulated genes: a complex web of host defenses. *Annu. Rev. Immunol.* 32: 513–545.
50. Snyder, D. T., J. F. Hedges, and M. A. Jutila. 2017. Getting "inside" type I IFNs: type I IFNs in intracellular bacterial infections. *J. Immunol. Res.* 2017: 9361802.
51. Piehler, J., C. Thomas, K. C. Garcia, and G. Schreiber. 2012. Structural and dynamic determinants of type I interferon receptor assembly and their functional interpretation. *Immunol. Rev.* 250: 317–334.
52. von Köckritz-Blickwede, M., O. Goldmann, P. Thulin, K. Heinemann, A. Norrby-Teglund, M. Rohde, and E. Medina. 2008. Phagocytosis-independent antimicrobial activity of mast cells by means of extracellular trap formation. *Blood* 111: 3070–3080.
53. Miallau, L., W. N. Hunter, S. M. McSweeney, and G. A. Leonard. 2007. Structures of *Staphylococcus aureus* D-tagatose-6-phosphate kinase implicate domain motions in specificity and mechanism. *J. Biol. Chem.* 282: 19948–19957.
54. Humphries, D. E., G. W. Wong, D. S. Friend, M. F. Gurish, W. T. Qiu, C. Huang, A. H. Sharpe, and R. L. Stevens. 1999. Heparin is essential for the storage of specific granule proteases in mast cells. *Nature* 400: 769–772.
55. MacMicking, J. D. 2012. Interferon-inducible effector mechanisms in cell-autonomous immunity. *Nat. Rev. Immunol.* 12: 367–382.
56. Li, X. D., J. Wu, D. Gao, H. Wang, L. Sun, and Z. J. Chen. 2013. Pivotal roles of cGAS-cGAMP signaling in antiviral defense and immune adjuvant effects. *Science* 341: 1390–1394.
57. Ahn, J., and G. N. Barber. 2019. STING signaling and host defense against microbial infection. *Exp. Mol. Med.* 51: 1–10.
58. Marinho, F. V., S. Benmerzoug, S. C. Oliveira, B. Ryffel, and V. F. J. Quesniaux. 2017. The emerging roles of STING in bacterial infections. *Trends Microbiol.* 25: 906–918.
59. Gries, C. M., E. L. Bruger, D. E. Moormeier, T. D. Scherr, C. M. Waters, and T. Kielian. 2016. Cyclic di-AMP released from *Staphylococcus aureus* biofilm induces a macrophage type I interferon response. *Infect. Immun.* 84: 3564–3574.
60. Nakamura, Y., J. Oscherwitz, K. B. Cease, S. M. Chan, R. Muñoz-Planillo, M. Hasegawa, A. E. Villaruz, G. Y. Cheung, M. J. McGavin, J. B. Travers, et al. 2013. *Staphylococcus* δ-toxin induces allergic skin disease by activating mast cells. *Nature* 503: 397–401.

3 Chapter 2: READemption 2: Multi-species RNA-Seq made easy

3.1 Cover and author affiliations

READemption 2: Multi-species RNA-Seq made easy

Till Sauerwein[1], Thorsten Bischler[2], Konrad U. Förstner[1,3]

[1]ZB MED - Information Centre for Life Science, 50931 Cologne, Germany
[2]Core Unit Systems Medicine, University of Würzburg, 97080 Würzburg, Germany
[3]TH Köln - University of Applied Sciences, 50578, Cologne, Germany

3.2 Manuscript

The following manuscript has been published as a pre-print in *bioRxiv*.

URL: `https://www.biorxiv.org/content/10.1101/2022.09.30.510338v1`

DOI: `https://doi.org/10.1101/2022.09.30.510338`

Publication of pre-print on October 03, 2022.

Personal contribution: I designed and developed the changes that were necessary for *READemption 2* to be able to carry out dual RNA-seq and multi RNA sequencing (multi RNA-seq). The development included creating software system and unit tests and adding test data sets, as well as updating the documentation and creating software packages of *READemption 2*. I wrote the manuscript for *READemption 2* and created all figures presented in the manuscript.

READemption 2: Multi-species RNA-Seq made easy

Till Sauerwein[1], **Thorsten Bischler**[2], **Konrad U. Förstner**[1,3]

***For correspondence:**
foerstner@zbmed.de (KUF)

[1]ZB MED-Information Centre for Life Science, 50931 Cologne, Germany; [2]Core Unit Systems Medicine, University of Würzburg, 97080 Würzburg; [3]TH Köln – University of Applied Sciences, 50578, Cologne, Germany

+

Abstract Dual or Multi RNA-seq simultaneously analyze the transcriptomes of two or more interacting species to gain insights about their interplay. The RNA of the interacting species is collected and sequenced together and only separated *in silico* by mapping the reads to the corresponding genomes. We developed READemption 2.0, to our knowledge the first tool that performs all necessary steps to handle RNA-seq data from any number of species. These steps comprise basic quality filtering and adapter trimming of raw reads, aligning the reads to reference genomes, generating nucleotide-wise coverage files, creating gene-wise read counts and performing differential gene expression analysis. These results can be visualized by additional subcommands of the software. READemption 2.0 allows users to produce meaningful results with default settings that follow conventional standards. Furthermore, many parameters can be adjusted to meet the users' specific needs, e.g. keeping or discarding species cross-mapped reads or normalizing the data.

Introduction

Dual RNA-sequencing (Dual RNA-seq) is the simultaneous transcriptome profiling of two interacting species (***Westermann et al., 2012***). If more than two species are investigated the term Multi RNA-sequencing (Multi RNA-seq) is used. The distinctive feature of these methods, compared to conventional RNA-seq, is that the RNA of all interacting partners like a pathogen and its host is extracted and sequenced without physical separation. Since the RNA of different species is sequenced together, assigning each read to its originating species only happens *in silico* (Figure 1). The simultaneous investigation of two (or more) species allows researchers to correlate the transcriptome profiles and thus gain new insights of the molecular interplay of the interacting species.

Since the first application of Dual RNA-seq to an eukaryotic pathogen and host system (***Tierney et al., 2012***), and its theoretically assessment of the general feasibility in pathogen host systems in the early 2010s (***Westermann et al., 2012***), the method has been applied to a variety of host-pathogen, mutualistic and commensal interaction systems (***Wolf et al., 2018***).

Several recent studies investigated host-pathogen interactions: For example, ***Aulicino et al. (2022)*** revealed a dynamic adaption of iron metabolism during *Salmonella* infection of dendritic cells for both the human host and the bacterial pathogen. Different Salmonella strains used different evasive strategies to counteract the iron-driven antimicrobial defense of the host, which in turn showed unique responses depending on the infecting strain. *Staphylococcus aureus* showed differential expression of virulence factors during infection of two mice strains. The virulence was influenced by the host's different level of resistance to the bacteria (***Thänert et al., 2017***). A recent study applied Dual RNA-seq to different SARS-CoV-2-infected patient samples and cell lines that

revealed co-expressed viral and human genes. A consensus network derived from co-expression highlighted a host response characterized by increased chemokine and cytokine activity (***Maulding et al., 2022***).

Here we present READemption 2.0, an open source command line tool that allows users to analyze Dual or Multi RNA-seq data. To our knowledge, READemption 2.0 is the first tool that allows researchers to perform multi-species RNA-seq analysis with any number of species.

Results

Application and usage of READemption and the need for a Dual/Multi RNA-seq analysis tool

Since READemption's initial release in 2014 (***Förstner et al., 2014***) it has been used by numerous publications for analyzing data from different RNA-seq applications. Among these applications are conventional RNA-seq (***Aguilar et al., 2020***; ***Lee et al., 2021***), differential RNA-seq (***Ponath et al., 2021***; ***Ryan et al., 2020***), Grad-seq (***Hör et al., 2020***; ***Smirnov et al., 2016***), RIP-seq (***Kavita et al., 2022***; ***Liao et al., 2022***), CLIP-seq (***Bauriedl et al., 2020***; ***Holmqvist et al., 2016***), TIER-seq (***Hoyos et al., 2020***; ***Chao et al., 2017***) and metatranscriptomics (***Krohn-Molt et al., 2017***). The essential RNA-seq results that can be generated with READemption, like alignment files (BAM file format https://samtools.github.io/hts-specs/SAMv1.pdf), mapping statistics, nucleotide-wise coverage files, gene-wise quantification counts and differential gene expression analysis also serve as input for follow-up analysis tools: ANNOgesic (***Yu et al., 2018***), a tool for annotating bacterial and archaeal genomes uses coverage files as input e.g. for transcript start site and processing site detection, sRNA (small RNA) detection and sRNA target detection. GRADitude (https://github.com/foerstner-lab/GRADitude) uses gene-wise quantification counts and mapping statistics for RNA-RNA and RNA-protein prediction of GRAD-seq experiments. ClusterProfiler (***Wu et al., 2021***) performs gene set enrichment analysis (GSEA), which requires tables containing genes and their corresponding differential gene expression fold changes calculated by e.g. DESeq2 (***Love et al., 2014***), which is integrated into READemption's subcommand 'deseq'. PEAKachu (https://github.com/tbischler/PEAKachu), a peak-calling tool for CLIP-seq data, needs BAM files as input that can be generated with READemption's 'align' subcommand.

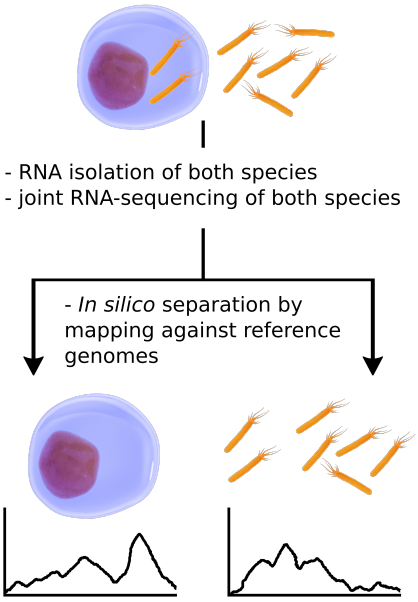

Figure 1. General Dual RNA-sequencing workflow

The number of publications using READemption for RNA-seq analysis has increased over the years (Figure 2A, ***PubMed*** (***2022a***)). This increase and the different RNA-seq protocols READemption has been applied to, show the need for an RNA-seq analysis tool that covers a broad spectrum of RNA-seq applications. As the number of publications applying Dual RNA-seq also increased over the years (Figure 2B, ***PubMed*** (***2022b***)) and READemption could not handle Dual or Multi RNA-seq data without additional manual manipulation of input and output files, we developed READemption 2.0.

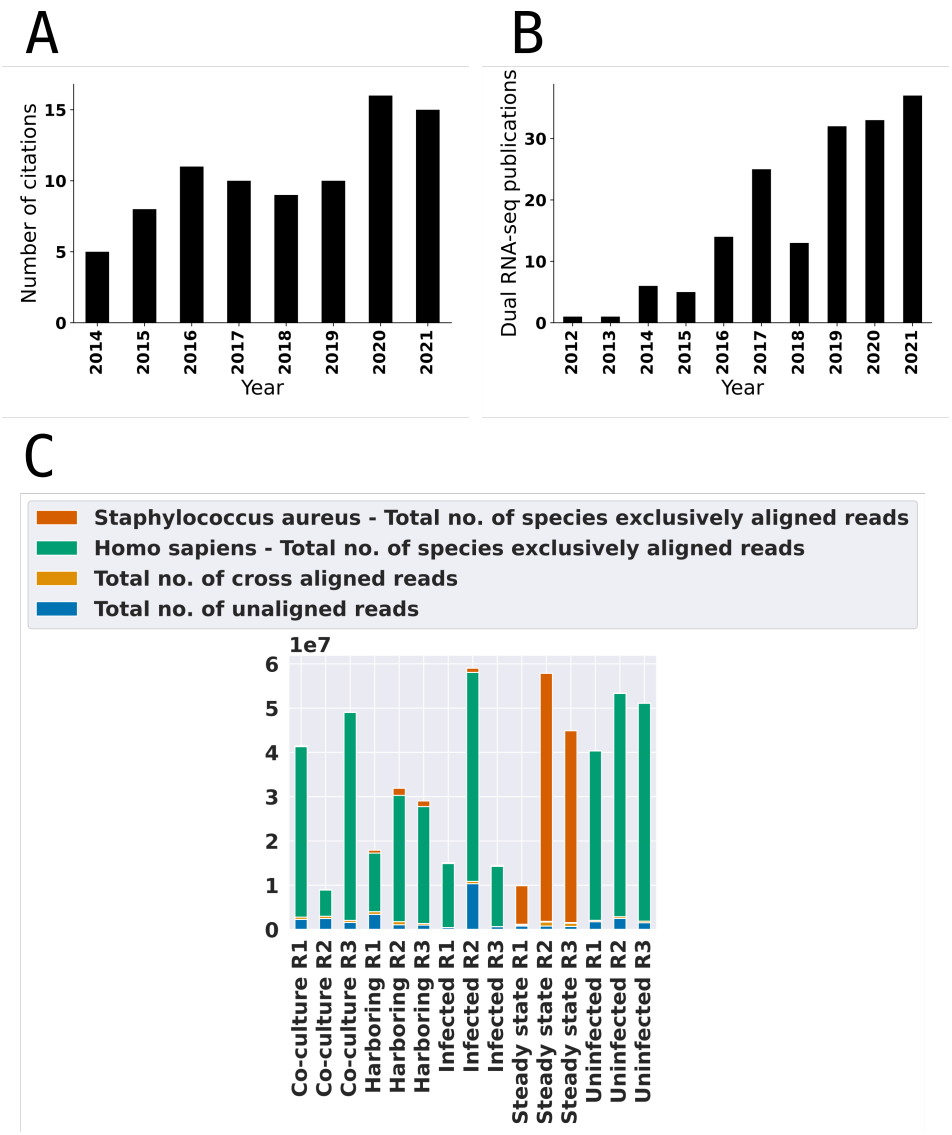

Figure 2. (A) Number of publications citing READemption 1.0 or earlier versions per year. **(B)** Number of publications having "Dual RNA-seq" in their title or abstract per year. **(C)** Alignment statistics plot of a Dual RNA-seq experiment with 15 libraries generated with READemption 2.0's 'viz_align' subcommand. The plot shows the number of species exclusive aligned reads for each species, the species cross-mapped reads and the unaligned reads.

READemption 2.0's general workflow

READemption 2.0 is a major upgrade of the previous READemption RNA-seq analysis tool. The new version enables users to analyze projects with more than one interacting species, while keeping all of READemption 1.0's core functionalities, including the ability to analyze RNA-seq data of a single species. To allow users an easy transition to the new version, the main workflow has not been changed and is as follows (see Figure 3): The first step of every new project is creating the input folder structure for reads, reference sequences and annotations with the subcommand 'create'. READemption 2.0 adds the possibility to create annotation and reference sequence input folders for each species by providing individual names for each species being part of the current RNA-seq analysis project (the two species of the example workflow in Figure 3 are "Human" and "Staphylococcus"). Then, the input files can be copied to their corresponding input folders. After the input files have been provided, READemption 2.0 automatically manages the input and output of all following subcommands. The next step is aligning the reads to the combined reference sequences of all species, using the subcommand 'align'. It has been shown that aligning read pools, containing sequences from multiple species, to combined reference sequences instead of aligning the reads subsequently to each species reference genome avoids introducing mapping bias (***Espindula et al., 2020***), which makes this combined approach READemption's method of choice when analyzing Dual RNA-seq data. The mapping statistics generated by the 'align' subcommand were updated to include mapping statistics by species. These include counts for reads that align to a single species and reads that cross-align to multiple species (Figure 2C). Another new feature of the 'align' subcommand is the possibility to merge the two aligned reads of a read pair and build template fragments when analyzing paired-end data. The derived fragments are stored in a BAM file as single-end alignments and can be used for further analysis instead of the BAM files that include the un-merged paired-end reads. After running the 'align' subcommand the user can perform the subcommands 'coverage' or 'gene quanti' followed by 'deseq'. The subcommand 'coverage' creates strand specific coverage files in wiggle format, containing nucleotide-wise read counts for the genomic positions of the reference sequences. The counts are provided with and without normalization and can be viewed in a genome browser for further inspection. The subcommand 'gene quanti' calculates the number of reads overlapping with each feature listed in the annotation files. The feature types to be used for the calculation can be specified by the user. The results are presented as raw counts and normalized counts, including transcripts per million (TPM, ***Wagner et al. (2012)***), reads per kilobase million (RPKM, ***Mortazavi et al. (2008)***) and normalized by the total number of aligned reads of the given library (TNOAR). After the gene quantification has been completed the subcommand 'deseq' can be used to perform differential gene expression analysis using the R package DESeq2 (***Love et al., 2014***), which is integrated into READemption. The subcommand also produces PCA (principal component analysis) plots and heatmaps of the library compositions. Finally, READemption offers subcommands for further visualization. 'Viz_align' generates histograms of the read length distributions, 'viz_gene_quanti' bar plots of the feature distribution and scatter plots comparing raw gene-wise quantification values for each library pair and 'viz_deseq' MA and Volcano plots.

Species cross-mapped reads and normalization

During the alignment the majority of reads can be unambigously assigned to their species. These species exclusive reads are then used for down-stream analysis of the corresponding species. Each subcommand that is being called after the initial alignment produces independent results for the different species. E.g. in a Dual RNA-seq experiment containing human cells and *Staphylococcus aureus* cells, READemption 2.0 creates coverage files once for the human genome and once for the bacterial genome, while only taking reads into account that map to the respective species (Figure 3: 'Coverage'-box). However, a typical Dual or Multi RNA-seq experiment contains a small fraction of reads that map equally well to two or more species. These species cross-mapped reads pose a problem, since discarding them causes information loss while keeping them results in potential

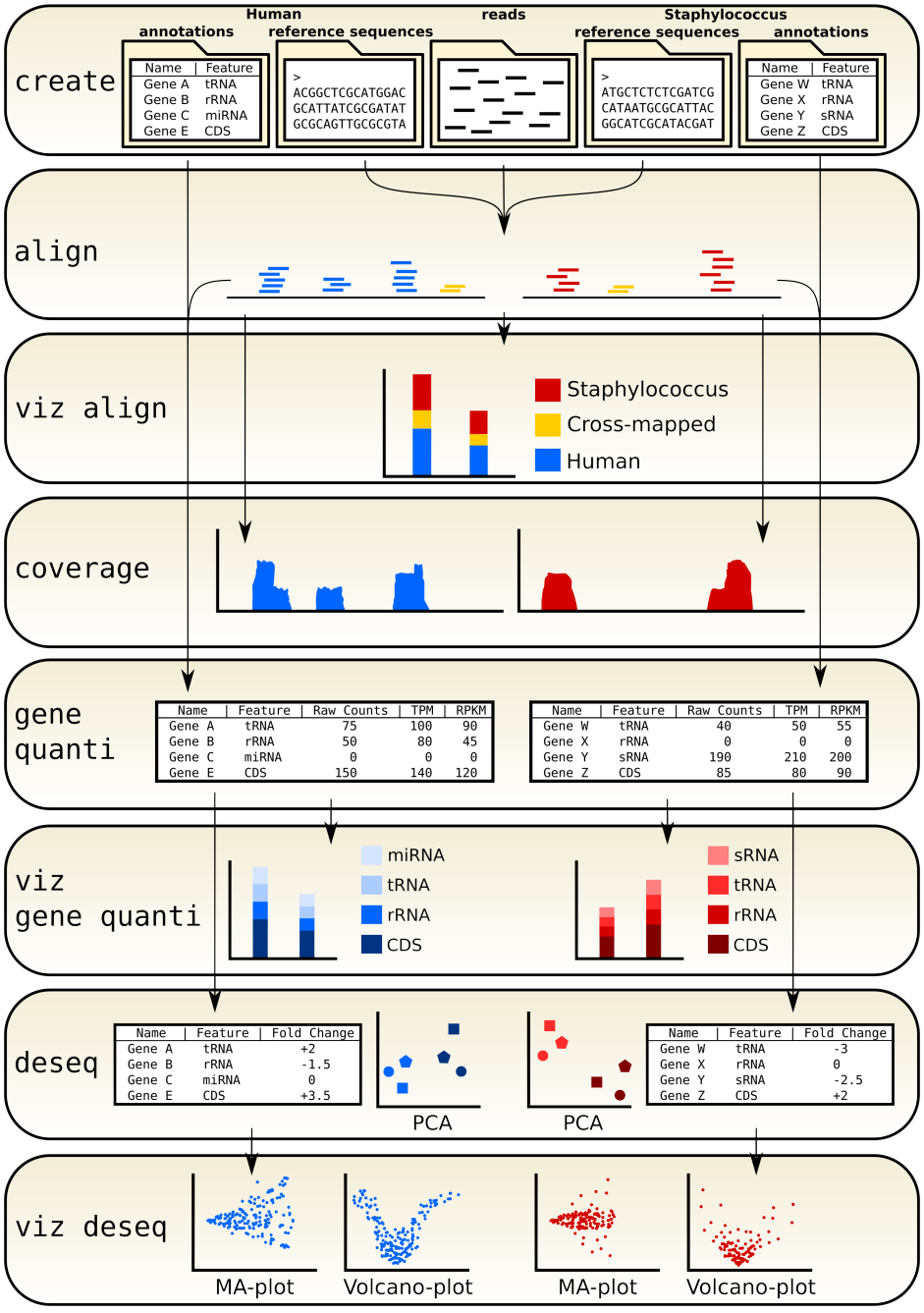

Figure 3. READemption 2.0 data and workflow overview of a Dual RNA-seq example analysis. Each subcommand is depicted as one box. The arrows indicate the data flow of the input and output

false positives. Although species cross-mapped reads are usually discarded, there is no gold standard of how to handle them (***Espindula et al., 2020***). To give users full control over cross-mapped reads and normalization, we added options to include or exclude cross-mapped reads for the nucleotide-wise counts of 'coverage' and the gene-wise counts of 'gene quanti', as well as including or excluding them in the values used for normalization for these subcommands. As cross-mapped reads are usually discarded, the default setting of READemption 2.0 is to exclude cross-mapped reads for both individual counts and normalization. READemption 2.0 provides three different ways for DESeq2's size factor calculation that is used for normalizing read counts over different libraries. The project-wise approach takes all feature counts of a species of all libraries into account when comparing conditions with each other, the species-wise approach uses only the libraries of the given species, and the comparison-wise approach only the libraries of the two conditions that are currently compared. We chose the species-wise approach as default setting, since the 'deseq' subcommand also generates PCA plots based on the libraries used for size factor estimation and usually the first quality control step of differential gene expression analysis is confirming via PCA, whether the libraries of the same condition cluster together.

Fragment building

Some manufacturers, e.g. Illumina or Applied Biosystems offer RNA-seq protocols that generate paired-end reads, where each cDNA template fragment is sequenced from both ends, resulting in a read pair. After the alignment the mapped pairs can be used to derive the genomic start and end position of the template they originate from. READemption 2.0 uses the alignment files (BAM files) of the initial alignment to generate template fragments from paired-end reads and writes them to a new BAM file containing the template fragments represented as single-end reads. Building these fragments is the default option, but can also be turned off to use the individual reads of a pair as input for the down-stream analysis.

Discussion

The growing number of research articles using Dual RNA-seq (***PubMed, 2022b***) shows the need in the scientific community for a tool that can conveniently analyze the data of such experiments. We present READemption 2.0, the first tool that can handle RNA-seq data of any number of species and any domain of life. Other tools that already exist and are suitable for the analysis of Multi-species RNA-seq either provide only basic alignment functionalities or can only handle a maximum of two different species. FastQ-screen (https://stevenwingett.github.io/FastQ-Screen/) generates read files that contain information about to which species a read could be aligned. These reads can be filtered and used as input for other third-party tools. However, FastQ-screen does not provide coverage-file creation, gene-wise quantification or differential gene expression analysis. The nf-core/dualrnaseq pipeline (https://nf-co.re/dualrnaseq) is able to perform read alignment and gene-wise quantification, but lacks the ability to analyze more than two species. Although READemption 2 has been developed with the intent to analyze Dual or Multi RNA-seq data of interacting species, its application in other areas of RNA-seq is conceivable. E.g. metatranscriptome analysis similar to ***Krohn-Molt et al.*** (***2017***) could be conveniently analyzed with READemption 2.0. During the development of READemption 2.0 we focused on easy accessibility, to ensure that researchers can run analyses with little prior knowledge of bioinformatics. We did this by choosing default parameters that are most common for the analysis of either Dual and Multi RNA-seq or conventional RNA-seq and by providing comprehensive tutorials, explanations and solutions for convenient installation of the tool and all its dependencies. However, parameters can be changed in different ways (e.g. different normalization approaches, use of single reads or fragment building for paired-end reads etc.) to meet the users' specific needs. This principal is called "convention of configuration" and has been applied to the default settings of all subcommands.

Methods and Materials

READemption 2.0 is written in Python and the source code is freely available under the ISC license on GitHub (https://github.com/foerstner-lab/READemption). The short read mapper segemehl (***Hoffmann et al., 2009***) and the R package DESeq2 (***Love et al., 2014***) are integrated into READemption 2.0. Software unit and system tests were created to guarantee READemption 2.0 runs as intended. The tests cover 85 % of the code, including all core functions. READemption can be installed via Conda (https://anaconda.org/Till_Sauerwein/reademption), PyPi (https://pypi.org/project/READemption/) or using a pre-installed Docker image (https://hub.docker.com/r/tillsauerwein/reademption). READemption 2.0's Documentation website (https://reademption.readthedocs.io/en/latest/index.html) hosts detailed descriptions of the subcommands, information about fragment building for paired-end reads, installation instructions and tutorials for beginners. The tutorials offer step-by-step instructions, input data and executable code to perform single or dual RNA-seq example analyses.

Acknowledgments

We thank Silvia Di Giorgio and Muhammad Elhossary for testing READemption 2.0 and giving us feedback. This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project number 460129525 (NFDI4Microbiota – National Research Data Infrastructure for Microbiota Research) and was supported by the IZKF at the University of Würzburg (project Z-6).

References

Aguilar C, Cruz AR, Rodrigues Lopes I, Maudet C, Sunkavalli U, Silva RJ, Sharan M, Lisowski C, Zaldívar-López S, Garrido JJ, Giacca M, Mano M, Eulalio A. Functional screenings reveal different requirements for host microRNAs in Salmonella and Shigella infection. Nature Microbiology. 2020 Jan; 5(1):192–205. doi: 10.1038/s41564-019-0614-3.

Aulicino A, Antanaviciute A, Frost J, Sousa Geros A, Mellado E, Attar M, Jagielowicz M, Hublitz P, Sinz J, Preciado-Llanes L, Napolitani G, Bowden R, Koohy H, Drakesmith H, Simmons A. Dual RNA sequencing reveals dendritic cell reprogramming in response to typhoidal Salmonella invasion. Communications Biology. 2022 Feb; 5(1):1–17. https://www.nature.com/articles/s42003-022-03038-z, doi: 10.1038/s42003-022-03038-z, number: 1 Publisher: Nature Publishing Group.

Bauriedl S, Gerovac M, Heidrich N, Bischler T, Barquist L, Vogel J, Schoen C. The minimal meningococcal ProQ protein has an intrinsic capacity for structure-based global RNA recognition. Nature Communications. 2020 Jun; 11(1):2823. doi: 10.1038/s41467-020-16650-6.

Chao Y, Li L, Girodat D, Förstner KU, Said N, Corcoran C, Śmiga M, Papenfort K, Reinhardt R, Wieden HJ, Luisi BF, Vogel J. In Vivo Cleavage Map Illuminates the Central Role of RNase E in Coding and Non-coding RNA Pathways. Molecular Cell. 2017 Jan; 65(1):39–51. doi: 10.1016/j.molcel.2016.11.002.

Espindula E, Sperb ER, Bach E, Passaglia LMP. The combined analysis as the best strategy for Dual RNA-Seq mapping. Genetics and Molecular Biology. 2020 Feb; 42(4):e20190215. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7249662/, doi: 10.1590/1678-4685-GMB-2019-0215.

Förstner KU, Vogel J, Sharma CM. READemption-a tool for the computational analysis of deep-sequencing-based transcriptome data. Bioinformatics (Oxford, England). 2014 Dec; 30(23):3421–3423. doi: 10.1093/bioinformatics/btu533.

Hoffmann S, Otto C, Kurtz S, Sharma CM, Khaitovich P, Vogel J, Stadler PF, Hackermüller J. Fast mapping of short sequences with mismatches, insertions and deletions using index structures. PLoS computational biology. 2009 Sep; 5(9):e1000502. doi: 10.1371/journal.pcbi.1000502.

Holmqvist E, Wright PR, Li L, Bischler T, Barquist L, Reinhardt R, Backofen R, Vogel J. Global RNA recognition patterns of post-transcriptional regulators Hfq and CsrA revealed by UV crosslinking in vivo. The EMBO journal. 2016 May; 35(9):991–1011. doi: 10.15252/embj.201593360.

Hoyos M, Huber M, Förstner KU, Papenfort K. Gene autoregulation by 3' UTR-derived bacterial small RNAs. eLife. 2020 Aug; 9:e58836. doi: 10.7554/eLife.58836.

Hör J, Di Giorgio S, Gerovac M, Venturini E, Förstner KU, Vogel J. Grad-seq shines light on unrecognized RNA and protein complexes in the model bacterium Escherichia coli. Nucleic Acids Research. 2020 Sep; 48(16):9301–9319. doi: 10.1093/nar/gkaa676.

Kavita K, Zhang A, Tai CH, Majdalani N, Storz G, Gottesman S. Multiple in vivo roles for the C-terminal domain of the RNA chaperone Hfq. Nucleic Acids Research. 2022 Feb; 50(3):1718–1733. doi: 10.1093/nar/gkac017.

Krohn-Molt I, Alawi M, Förstner KU, Wiegandt A, Burkhardt L, Indenbirken D, Thieß M, Grundhoff A, Kehr J, Tholey A, Streit WR. Insights into Microalga and Bacteria Interactions of Selected Phycosphere Biofilms Using Metagenomic, Transcriptomic, and Proteomic Approaches. Frontiers in Microbiology. 2017 Oct; 8:1941. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5641341/, doi: 10.3389/fmicb.2017.01941.

Lee SH, Cho SY, Yoon Y, Park C, Sohn J, Jeong JJ, Jeon BN, Jang M, An C, Lee S, Kim YY, Kim G, Kim S, Kim Y, Lee GB, Lee EJ, Kim SG, Kim HS, Kim Y, Kim H, et al. Bifidobacterium bifidum strains synergize with immune checkpoint inhibitors to reduce tumour burden in mice. Nature Microbiology. 2021 Mar; 6(3):277–288. https://www.nature.com/articles/s41564-020-00831-6, doi: 10.1038/s41564-020-00831-6, number: 3 Publisher: Nature Publishing Group.

Liao C, Sharma S, Svensson SL, Kibe A, Weinberg Z, Alkhnbashi OS, Bischler T, Backofen R, Caliskan N, Sharma CM, Beisel CL. Spacer prioritization in CRISPR-Cas9 immunity is enabled by the leader RNA. Nature Microbiology. 2022 Apr; 7(4):530–541. doi: 10.1038/s41564-022-01074-3.

Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology. 2014; 15(12):550. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4302049/, doi: 10.1186/s13059-014-0550-8.

Maulding ND, Seiler S, Pearson A, Kreusser N, Stuart JM. Dual RNA-Seq analysis of SARS-CoV-2 correlates specific human transcriptional response pathways directly to viral expression. Scientific Reports. 2022 Jan; 12(1):1329. https://www.nature.com/articles/s41598-022-05342-4, doi: 10.1038/s41598-022-05342-4, number: 1 Publisher: Nature Publishing Group.

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nature Methods. 2008 Jul; 5(7):621–628. https://www.nature.com/articles/nmeth.1226, doi: 10.1038/nmeth.1226, number: 7 Publisher: Nature Publishing Group.

Ponath F, Tawk C, Zhu Y, Barquist L, Faber F, Vogel J. RNA landscape of the emerging cancer-associated microbe Fusobacterium nucleatum. Nature Microbiology. 2021 Aug; 6(8):1007–1020. doi: 10.1038/s41564-021-00927-7.

PubMed, PubMed search for articles citing READemption; 2022. [Online; accessed 02-September-2022]. https://pubmed.ncbi.nlm.nih.gov/?linkname=pubmed_pubmed_citedin&from_uid=25123900.

PubMed, PubMed search for articles having 'Dual RNA-seq' in title or abstract; 2022. [Online; accessed 02-September-2022]. https://pubmed.ncbi.nlm.nih.gov/?term=dual+RNA-seq%5BTitle%2FAbstract%5D&sort=pubdate.

Ryan D, Jenniches L, Reichardt S, Barquist L, Westermann AJ. A high-resolution transcriptome map identifies small RNA regulation of metabolism in the gut microbe Bacteroides thetaiotaomicron. Nature Communications. 2020 Jul; 11(1):3557. doi: 10.1038/s41467-020-17348-5.

Smirnov A, Förstner KU, Holmqvist E, Otto A, Günster R, Becher D, Reinhardt R, Vogel J. Grad-seq guides the discovery of ProQ as a major small RNA-binding protein. Proceedings of the National Academy of Sciences of the United States of America. 2016 Oct; 113(41):11591–11596. doi: 10.1073/pnas.1609981113.

Thänert R, Goldmann O, Beineke A, Medina E. Host-inherent variability influences the transcriptional response of Staphylococcus aureus during in vivo infection. Nature Communications. 2017 Feb; 8(1):14268. https://www.nature.com/articles/ncomms14268, doi: 10.1038/ncomms14268, number: 1 Publisher: Nature Publishing Group.

Tierney L, Linde J, Müller S, Brunke S, Molina JC, Hube B, Schöck U, Guthke R, Kuchler K. An Interspecies Regulatory Network Inferred from Simultaneous RNA-seq of Candida albicans Invading Innate Immune Cells. Frontiers in Microbiology. 2012; 3:85. doi: 10.3389/fmicb.2012.00085.

Wagner GP, Kin K, Lynch VJ. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. Theory in Biosciences. 2012 Dec; 131(4):281–285. https://doi.org/10.1007/s12064-012-0162-3, doi: 10.1007/s12064-012-0162-3.

288 **Westermann AJ**, Gorski SA, Vogel J. Dual RNA-seq of pathogen and host. Nature Reviews Microbiology. 2012
289 Sep; 10(9):618–630. https://www.nature.com/articles/nrmicro2852, doi: 10.1038/nrmicro2852, number: 9
290 Publisher: Nature Publishing Group.

291 **Wolf T**, Kämmer P, Brunke S, Linde J. Two's company: studying interspecies relationships with dual RNA-
292 seq. Current Opinion in Microbiology. 2018 Apr; 42:7–12. https://www.sciencedirect.com/science/article/
293 pii/S1369527417301327, doi: 10.1016/j.mib.2017.09.001.

294 **Wu T**, Hu E, Xu S, Chen M, Guo P, Dai Z, Feng T, Zhou L, Tang W, Zhan L, Fu X, Liu S, Bo X, Yu G. clusterProfiler 4.0:
295 A universal enrichment tool for interpreting omics data. The Innovation. 2021 Aug; 2(3). https://www.cell.
296 com/the-innovation/abstract/S2666-6758(21)00066-7, doi: 10.1016/j.xinn.2021.100141, publisher: Elsevier.

297 **Yu SH**, Vogel J, Förstner KU. ANNOgesic: a Swiss army knife for the RNA-seq based annotation of bacte-
298 rial/archaeal genomes. GigaScience. 2018 Sep; 7(9). doi: 10.1093/gigascience/giy096.

3.3 Additional results: Fragment building of paired-end reads

Due to the different fragment lengths of the complementary DNA (cDNA) library, the insert size between two reads of a pair is not known before the alignment. Therefore, both reads are mapped independently to the reference sequence. After both reads have been mapped, the aligner can derive a template length for the sequenced fragment. The calculated template length is the distance from the leftmost mapping position until the rightmost mapping position of the alignments of a pair.

Two main layouts can occur when both reads of a pair have been aligned. The reads can either be *in order* or in *reverse order*. Two reads are *in order* if the position of read 2 is downstream or equal to the position of read 1. To determine the order of a read pair, it is important to consider the orientation of the mapped reads. The orientation of a read is defined as the strand it maps to. One strand of the two strands of a DNA molecule is named template-, forward-, plus- or sense-strand, while the other strand is named reverse-, minus- or anti-sense-strand. Because the fragment that a single-end read originates from is always sequenced in 5' to 3' direction, the orientation of a read is the same as the strand it maps to. For paired-end reads the orientation of read 1 is also the same orientation as the strand it maps to, but read 2 has the opposite orientation of the strand it maps to. This can be explained by the fact that every fragment is first sequenced from its 5'-end, which results in read 1 and than from its 3'-end, which results in read 2. Since mapping positions are always indicated in relation to the forward strand of the genome, a read pair is *in order* if read 1 maps to the forward strand and the first aligned base of read 2 is equal or greater than the first aligned base of read 1 (Figure 3.1 A, B and C). Whereas if instead read 1 maps to the reverse strand, the pair is *in order* if the first aligned base of read 1 is equal or greater than the first aligned base of read 2 (Figure 3.2 A, B and C).

If the alignment positions do not apply to the rules described above, the reads are in *reverse order*. Three different layout categories can occur for reads *in order* and two different ones for reads in *reverse order*. For reads *in order*, the reads of a pair can either overlap (Figure 3.1 B, Figure 3.2 B), are identical (Figure 3.1 C, Figure 3.2 C) or don't overlap at all (Figure 3.1 A, Figure 3.2 A). For reads in *reverse order*, the reads can overlap (Figure 3.1 D, Figure 3.2 D) or don't overlap (Figure 3.1 E, Figure 3.2 E), which can also be described as the reads exceeding each other. The two layouts of pairs in *reverse order* represent special cases. The layout where the reads overlap represents an

alignment, where the read length is greater than the actual fragment length (Figure 3.1 D, Figure 3.2 D). To obtain the boundaries of the fragment, only the part where both reads overlap is being kept and the exceeding ends are cut off. The layout in *reverse order* where two reads overlap and exceed each other (Figure 3.1 E, Figure 3.2 E) can indicate a circRNA as explained in the introduction 1.3.3.

For all layouts except the one where the fragment length is smaller than the read length, the fragment length is equal to the template length calculated by the aligner. Furthermore, the start position of the fragment is the leftmost position of the two start positions of read 1 and read 2. And the end position of the fragment is the start position of the fragment plus the template length. Because of this and the fact that each alignment has the information of the start position of read 1 and read 2 as well as the template length, the alignment information of only one read of the pair is sufficient to build the template. The only exception is the layout where the fragment length is smaller than the read length. To calculate the fragment length, either the start of a read and the end of its mate or the end of a read and the start of its mate are needed. This information can only be obtained when the information of both reads is present.

Retrieving the mate of one read of a pair slows down the process of parsing a SAM file, because first the position of the mate inside the file has to be searched and then the position of the file has to be changed to the position of the mate. A much faster alternative, which has been applied to *READemption's* fragment building algorithm, is sorting the SAM file in a way that read 2 is always presented one line after read 1 of the same pair before parsing the file. This ensures that the SAM file can be parsed from top to bottom. Instead of looking up the mate of a read somewhere in the file, the two consecutive lines containing both reads of a pair can be cached and afterwards processed together.

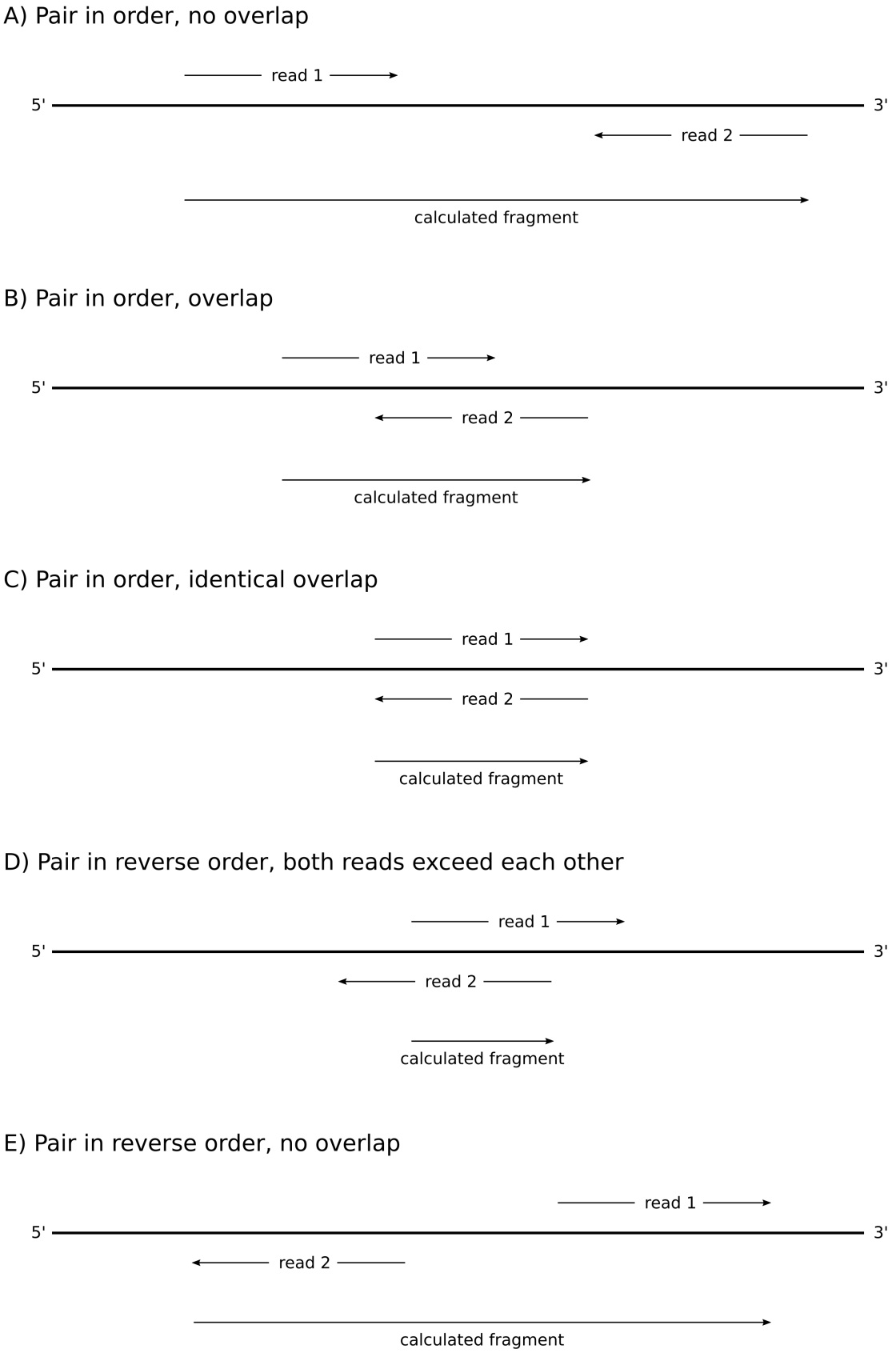

Figure 3.1: **Layouts of paired-end reads, where the originating fragment maps to the forward strand.** Each layout shows read 1 and read 2, the reference sequence in the middle of both and the calculated fragment at the bottom. The arrows indicate the direction from 5'-end to 3'-end

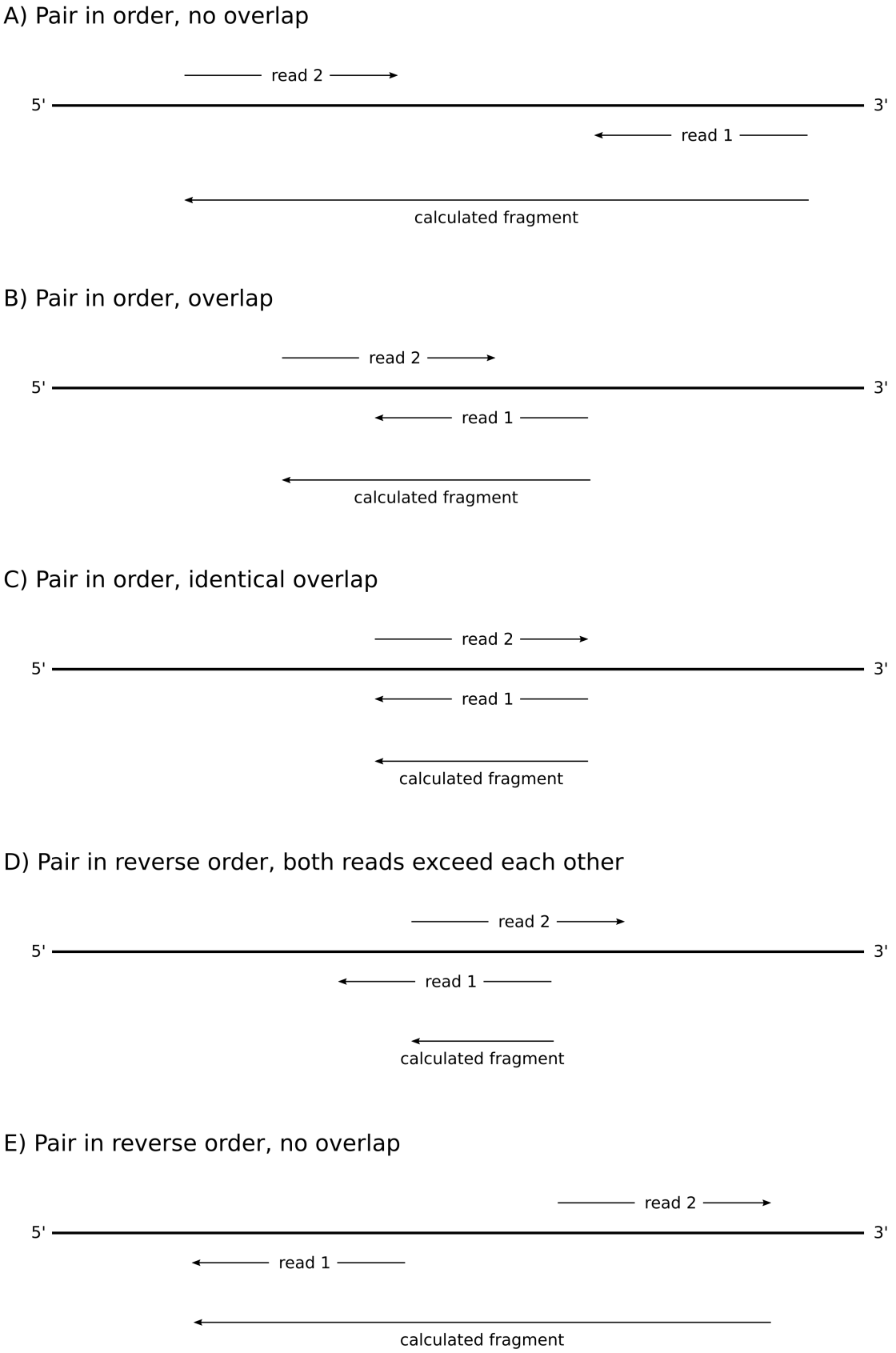

Figure 3.2: **Layouts of paired-end reads, where the originating fragment maps to the reverse strand.** Each layout shows read 1 and read 2, the reference sequence in the middle of both and the calculated fragment at the bottom. The arrows indicate the direction from 5'-end to 3'-end

4 Discussion and Outlook

4.1 Bioinformatical analysis of dual RNA sequencing of human mast cells and *S. aureus*

In our study from chapter 1 a dual RNA-seq approach has been conducted to investigate the cross-talk of human mast cells and *Staphylococcus aureus* (*S. aureus*) in an infection model. *S. aureus* is a commensal Gram-positive bacteria that can become an opportunistic pathogen, causing community and hospital-associated pathologies, like bacteremia-sepsis, endocarditis, pneumonia, osteomyelitis, arthritis and skin diseases like atopic dermatitis (AD) (Dayan et al., 2016). AD is a chronic inflammatory skin disease that causes acute and chronic skin lesions. In 75% up to 100% of the cases the skin lesions are colonized by *S. aureus* (Higaki et al., 1999; Breuer et al., 2002; Gong et al., 2006; Lin et al., 2007). Among the first immune host cells, *S. aureus* encounters during infection, may be mast cells. Mast cells are tissue-sentinel cells that are dispersed throughout most tissues and can be found at interfaces with the host's environment, like mucosae and skin. (Abraham and St John, 2010). In fact, *S. aureus* has been found to be internalized by mast cells in nasal polyps isolated from patients with chronic rhinosinusitis (Hayes et al., 2015). The defense mechanisms of mast cells against *S. aureus* involve release of extracellular traps composed of granule proteins and DNA that immobilize and kill the bacteria, as well as discharge of antimicrobial products that have a toxic effect on the bacteria (Abel et al., 2011). However, *S. aureus* can evade the host's defense mechanisms by directing its own uptake into mast cells, where they persist (Abel et al., 2011).

To investigate possible transcriptome changes of *S. aureus* and the host's response, different infection settings were subjected to dual RNA-seq. When planning RNA-seq experiments that aim to discover differentially expressed genes by DGE, determining the correct sequencing depth for libraries is of great importance and subject of an ongoing debate. A saturation analysis carried out by generating 214 million paired-end reads from H1 human embryonic stem cells came to the conclusion that 36 million reads are sufficient to quantify 80% of transcripts that are expressed at an expression level of fragments per kilobase million (FPKM) greater than 10. However, to quantify low expressed genes with FPKM values below 10, around 80 million mapped reads were needed (ENCODE Project

Consortium, 2011). Others have estimated that more than 200 million paired-end reads are required to detect all transcripts and possible isoforms of the human transcriptome (Tarazona et al., 2011). On the other hand, a study investigating the effects of sequencing depth and replicate number comes to the conclusion that exceeding a sequencing depth of 10 million reads generates diminishing returns for power of detecting differentially expressed genes in samples from MCF-7 breast cancer cells (Liu et al., 2014). Summing up, 10 to 40 million reads can be enough to generate meaningful DGE results for human cells, especially with increased numbers of replicates (Liu et al., 2014). For bacteria, findings suggest that 5 to 10 million rRNA depleted fragments are sufficient to detect the majority of transcripts. Even when the read number was lowered to 2 million fragments, 96% of open reading frames (ORFs) were covered by at least 1 fragment and 85% by at least 5 fragments (Haas et al., 2012).

In our study the three replicates of the condition containing human mast cells and intracellular *S. aureus* had 13 million, 28 million and 26 million mapped human reads and 700,000, 1.6 million and 1.2 million mapped bacterial reads, respectively. Hence, the total numbers of human mapped reads per library are in the above mentioned range of 10 to 40 million mapped reads, which have proven to be sufficient for DGE of human cells. Indeed, infected human cells harboring *S. aureus* and also bystander-cells that were infected with bacteria, but did not harbor them yielded differentially expressed genes compared to the uninfected control samples. The Reactome pathway (Gillespie et al., 2022) enrichment analysis revealed a transcriptional signature related to genes induced by type-I interferon (IFN-I) for both infection settings. Previous findings suggested that, though mast cells can elicit an IFN-I response upon viral infection, an infection with Gram-positive or Gram-negative bacteria does not trigger an IFN-I response, because of the bacteria's inability to internalize within mast cells (Dietrich et al., 2010). Our results show that human mast cells are in fact capable of governing an IFN-I immune response in an autocrine manner after uptake of *S. aureus* cells and that they signal non-infected bystander-cells in a paracrine manner to also trigger an IFN-I response.

The total number of the mapped reads per library of the intracellular bacteria were lower than the above described minimum of 2 million reads, but nevertheless resulted in robust transcriptome changes compared to the control condition. This can be explained by the relatively small genome size of *S. aureus* of 2.8 megabases compared with the genomes of *Escherichia coli* (4.6 megabases), *Mycobacterium tuberculosis* (4.4 megabases) and *Vibrio cholerae* (4.0 megabases), which were used to access sufficient sequencing depth for DGE as described above, because in general smaller genomes need fewer reads for a sufficient coverage (Haas et al., 2012). The KEGG pathway enrichment analysis revealed that genes

associated with enzymes and transport systems of galactose/lactose and D-tagatose-6-phosphate metabolic pathways were enriched in the up-regulated genes and genes involved in glycolysis were enriched in the down-regulated set of genes, indicating that D-galactose or lactose are potential carbon sources for the internalized bacteria. These findings suggest that *S. aureus* needs to readjusts its metabolism to the intracellular niche in order to survive. The DGE analysis of the transwell approach, where mast cells and *S. aureus* were separated by a permeable transwell did not yield any differentially expressed genes, which leads to the conclusion that the two species need physical contact to influence their gene expression. Taken together, our study has provided new insights about how human mast cells recognize intracellular *S. aureus* and how in turn *S. aureus* adapts its metabolism to persist inside mast cells to evade the immune response.

Since we did not include ncRNAs in our DGE analysis, follow-up research might focus on ncRNA to reveal possible regulatory ncRNAs that govern the gene expression changes observed during infection and internalization. We uploaded the raw reads to the ENA (Cummins et al., 2021) and made the bioinformatical analysis consisting of executable scripts, a singularity image containing every used software, and results, including mapped reads, gene-quantification, coverage-files and sRNA predictions generated with *ANNOgesic* (Yu et al., 2018), publicly available at the Repository for Life Sciences (`https://www.publisso.de/en/publishing/repositories/repository-for-life-sciences/`). Hence, our bioinformatical analysis is completely reproducible and can serve as a starting point for follow-up analysis of the dual RNA-seq data.

To gain further insights into the molecular interplay of mast cells and *S. aureus*, a dual RNA-seq approach based on single-cell RNA sequencing (scRNA-seq) might reveal possible subpopulations among internalized *S. aureus*, similar to Avital et al. (2017) where different subpopulations for infection stages were identified for intracellular *Salmonella typhimurium* and its host, namely mouse macrophages.

4.2 Software development for dual and multi RNA sequencing analysis

To carry out the bioinformatical analysis for the research article of chapter 1, various existing bioinformatical software tools were combined with Shell-, Python- and R-scripts. In principal, the workflow that was created can be adapted to other dual RNA-seq experiments with different species. However, adapting the workflow for other species requires

bioinformatical knowledge and manually changing the different scripts at specific locations, mainly to provide the genome sequence reference IDs of the respective species. In order to facilitate dual RNA-seq and multi RNA-seq analysis, we released a major update (manuscript in chapter 2) of the existing RNA-seq analysis tool *READemption* (Förstner et al., 2014).

The new version *READemption 2* retains all features of the previous *READemption* versions, while adding the option to analyze multi-species projects of any number of species chosen by the user. The basic workflow and subcommands have not been changed to allow users that have already used *READemption* an easy transition to *READemption 2*. However, some minor syntax and behavior changes were necessary. For instance, the first subcommand *create*, which creates the input folder structure for reference genomes, annotation files and reads, now requires the names of the species that are part of the project, in order to create reference genome and annotation input folders for each species. Furthermore, the *deseq* subcommand, in addition to the information about the condition and replicate number for every library, now requires information about which species are expected in each library.

One goal for *READemption 2* was to uphold the principal of 'convention over configuration' for running analyses, which has already been applied to the earlier versions of *READemption*. The aim of the principal is to reduce the number of decisions that have to be made by a user when executing a software. This aim can be achieved by setting the default behavior of the software to the most used conventional standards, without the need for the user to explicitly configure these settings. For example, species cross-aligned reads are usually discarded when analyzing dual RNA-seq data (Espindula et al., 2020). therefore, *READemption 2* also discards these reads by default when the subcommands for gene quantificaton or coverage file creation are called. However, to cover specific user needs, it is possible to include species cross-aligned reads when calling these subcommands by adding a predefined parameter ('count_cross_aligned_reads'). By applying the principal of 'convention over configuration', *READemption 2* lowers the hurdles for new users with little bioinformatics knowledge to perform data analysis of dual RNA-seq and multi RNA-seq data.

Further measurements that have been taken to help new users to get started with *READemption 2* are an updated documentation and the provision of software packages. Besides detailed descriptions about each subcommand and their adjustable parameters, the documentation provides tutorials for executing example analyses. The two tutorials cover a conventional RNA-seq analysis with one species and a dual RNA-seq analysis with two

species. Both tutorials provide step-by-step executable commands and explanations, as well as the required input data consisting of reads, annotations and reference sequences. *READemption 2* comes with a Conda package, which includes all necessary dependencies like Python packages, R with *DESeq2* and the aligner *segemehl*. The installation via Conda is explained in detail at *READemption 2*'s documentation website and is much more convenient and stable than installing *READemption 2* and its dependencies individually. Because even the Conda installation can result in conflicts, we also provide a Docker image with *READemption 2* that is in general more reliable than a Conda installation. The Docker image is also accompanied by a step-by-step tutorial for running an RNA-seq analysis with *READemption 2*.

Scientists spend more than 30% of their time to develop scientific software, which usually can not be outsourced, due to the domain-specific knowledge required and 90% of the scientists developing software are primarily self-taught (Wilson et al., 2014). To ensure appropriate software quality and to guide scientists during software development, a number of good practices have been worked out (Leprevost et al., 2014). During the development of *READemption 2* and its earlier versions, the following good practices for scientific software development have been applied. Each released version of *READemption* follows semantic versioning, has been assigned a digital object identifier (DOI) and its source code is stored publicly available on the open research database Zenodo (European Organization For Nuclear Research and OpenAIRE, 2013). A changelog file records all notable changes that have been made for each version. Storing different versions of a software and tracking the changes helps users that have older versions embedded in their bioinformatics workflow to maintain their workflow with a given older version and to decide when they should update their workflow to a newer version of the software. As described above, *READemption 2* comes with a comprehensive documentation to explain its features and workflow to new users. For developers it is also important that the source code itself is documented to get a deep understanding of the internal processes of the software. For this purpose, plain text explanations, which explain the intended function of code as well as Python type hints that annotate the arguments and the return value of functions have been introduced during development. Another recommendation for scientific software development is testing the software to make sure it runs as intended. Therefore, software unit tests and system test have been created for *READemption 2*'s core functionalities. The tests run parts of the software with a pre-defined input and output and test whether the actual output produced during the test is the same as the expected pre-defined output. The tests are included in *READemption 2* and can also be run by users to verify that the software's installation process was successful. Additionally, software tests are another way to explain to developers how the software is expected to

behave depending on the input data of tests. In regard to continuous integration (CI), whenever changes made to the software are published on *GitHub*, tests run automatically and indicate whether they failed or succeeded at *READemption 2*'s *GitHub* repository. *READemption 2* and its earlier versions are open-source to ensure full transparency and to provide clarity about how the results of its RNA-seq analysis are generated. Open-source projects also allow others to spot software bugs and to participate in development.

READemption 2 is the first tool that can comprehensively handle RNA-seq data of any amount of species and of any domain of life. Other existing tools are also suited to analyze dual RNA-seq or multi RNA-seq data but lack certain functions of *READemption 2*. The nf-core (Ewels et al., 2020) *dualrnaseq* pipeline (`https://nf-co.re/dualrnaseq/1.0.0`) is able to generate gene expression values of two interacting species, but lacks the ability to analyze more than two species and does not perform DGE or nucleotide-wise coverage calculation. *FastQ-Screen* (`https://stevenwingett.github.io/FastQ-Screen/`) can also be used as an entry point for dual RNA-seq or multi RNA-seq as it aligns reads to reference genomes of any amount of species and separates the initial files into species-specific files that only contain reads of one species. The species-specific read files can then be used to perform RNA-seq analysis for each species. However, *FastQ-Screen* does only provide the basic alignment and generation of species-wise alignment statistics but doesn't implement gene quantification, nucleotide-wise coverage calculation or DGE.

READemption 2's ability to analyze any number of species could also be used to analyze RNA-seq data of metatrancscriptomics projects that investigate a large number of different species (Shakya et al., 2019). More closely related species increase the risk of cross-species mapped reads and require a detailed inspection of cross-mapping between the species of a project. To find out which pairs of species share large amounts of cross-mapped reads, the cross-mapped reads for each possible pair of a project must be calculated. This feature is currently not implemented in *READemption 2*, instead only the total species cross-mapped reads for a library are calculated. In principal, the feature can be implemented in the future to make *READemption 2* more suitable for metatranscriptomics.

Further improvements for *READemption 2* that might be implemented in the future are described in the following. The aligner used by *READemption 2*, *segemehl* (Hoffmann et al., 2014), though having high accuracy in terms of correctly mapped reads compared to other aligners, has one of the longest run times (Otto et al., 2014; Donato et al., 2021). An alternative could be *STAR* (Dobin et al., 2013), a widely used mapper, which needs less than half of *segemehl's* memory consumption and tremendously outperforms it in terms

of mapping speed (Otto et al., 2014; Donato et al., 2021). *STAR* might either replace *segemehl* or be added as an alternative. Adding *STAR* as an alternative has the advantage that users can choose the aligner that best suits their needs, but also comes with the disadvantage for developers of *READemption 2* to maintain both aligners. Although *READemption 2* already covers a wide spectrum of the standard workflow of conventional RNA-seq, dual RNA-seq and multi RNA-seq that ranges from aligning to DGE, GSEA is not included in *READemption 2*. As GSEA is a common method to study gene expression and needs the results of DGE, which is already implemented in *READemption 2*, it would be a useful addition to *READemption 2*'s features. GSEA might be added as an additional subcommand that wraps already existing tools that implement GSEA, e.g. the Python package *GSEApy* (`https://github.com/zqfang/GSEApy`), or the R package *ClusterProfiler* (Wu et al., 2021). As described above, *READemption 2* features convenient installation via Conda or Docker and the usage is described in detail by tutorials. However, since *READemption 2* has a command-line interface that requires users to have a basic understanding of working with the command line, some potential users may not be able to use the tool. A solution might be to implement a graphical user interface (GUI), e.g. with *Gooey* (`https://github.com/chriskiehl/Gooey`), which makes use of the argument parser *argparse* (`https://docs.python.org/3/library/argparse.html`) that is already used by *READemption 2*. Another possible solution to make *READemption 2* available to users that are not familiar with working in a command line is the *Galaxy* platform (Afgan et al., 2018), which enables users to use a large set of bioinformatics tools via a browser-based GUI. Though, implementing *READemption 2* as a *Galaxy* workflow would be laborious compared to creating a GUI, since the complete input and output flow of each subcommand has to be implemented in the *Galaxy* environment.

5 Bibliography

Abel, J., O. Goldmann, C. Ziegler, C. Höltje, M. S. Smeltzer, A. L. Cheung, D. Bruhn, M. Rohde, and E. Medina (2011). "Staphylococcus aureus Evades the Extracellular Antimicrobial Activity of Mast Cells by Promoting Its Own Uptake". In: *Journal of Innate Immunity* 3.5. Publisher: Karger Publishers, pp. 495–507.

Abraham, S. N. and A. L. St John (2010). "Mast cell-orchestrated immunity to pathogens". In: *Nature reviews. Immunology* 10.6, pp. 440–452.

Afgan, E., D. Baker, B. Batut, M. van den Beek, D. Bouvier, M. Čech, J. Chilton, D. Clements, N. Coraor, B. A. Grüning, A. Guerler, J. Hillman-Jackson, S. Hiltemann, V. Jalili, H. Rasche, N. Soranzo, J. Goecks, J. Taylor, A. Nekrutenko, and D. Blankenberg (2018). "The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update". In: *Nucleic Acids Research* 46.W1, W537–W544.

Ashburner, M., C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, and G. Sherlock (2000). "Gene ontology: tool for the unification of biology. The Gene Ontology Consortium". eng. In: *Nature Genetics* 25.1, pp. 25–29.

Avital, G., R. Avraham, A. Fan, T. Hashimshony, D. T. Hung, and I. Yanai (2017). "scDual-Seq: mapping the gene regulatory program of Salmonella infection by host and pathogen single-cell RNA-sequencing". In: *Genome Biology* 18.1, p. 200.

Avraham, R., N. Haseley, D. Brown, C. Penaranda, H. B. Jijon, J. J. Trombetta, R. Satija, A. K. Shalek, R. J. Xavier, A. Regev, and D. T. Hung (2015). "Pathogen Cell-to-Cell Variability Drives Heterogeneity in Host Immune Responses". eng. In: *Cell* 162.6, pp. 1309–1321.

Bartel, D. P. (2009). "MicroRNAs: target recognition and regulatory functions". eng. In: *Cell* 136.2, pp. 215–233.

Bentley, D. R. et al. (2008). "Accurate whole human genome sequencing using reversible terminator chemistry". eng. In: *Nature* 456.7218, pp. 53–59.

Berk, A. J. (2016). "Discovery of RNA splicing and genes in pieces". In: *Proceedings of the National Academy of Sciences* 113.4. Publisher: Proceedings of the National Academy of Sciences, pp. 801–805.

Breuer, K., S. HAussler, A. Kapp, and T. Werfel (2002). "Staphylococcus aureus: colonizing features and influence of an antibacterial treatment in adults with atopic dermatitis". eng. In: *The British Journal of Dermatology* 147.1, pp. 55–61.

Chao, H.-P., Y. Chen, Y. Takata, M. W. Tomida, K. Lin, J. S. Kirk, M. S. Simper, C. D. Mikulec, J. E. Rundhaug, S. M. Fischer, T. Chen, D. G. Tang, Y. Lu, and J. Shen (2019). "Systematic evaluation of RNA-Seq preparation protocol performance". In: *BMC Genomics* 20.1, p. 571.

Chen, C., S. S. Khaleel, H. Huang, and C. H. Wu (2014). "Software for pre-processing Illumina next-generation sequencing short read sequences". In: *Source Code for Biology and Medicine* 9.1, p. 8.

Crick, F. H. (1958). "On protein synthesis". eng. In: *Symposia of the Society for Experimental Biology* 12, pp. 138–163.

Cummins, C., A. Ahamed, R. Aslam, J. Burgin, R. Devraj, O. Edbali, D. Gupta, P. W. Harrison, M. Haseeb, S. Holt, T. Ibrahim, E. Ivanov, S. Jayathilaka, V. Kadhirvelu, S. Kay, M. Kumar, A. Lathi, R. Leinonen, F. Madeira, N. Madhusoodanan, M. Mansurova, C. O'Cathail, M. Pearce, S. Pesant, N. Rahman, J. Rajan, G. Rinck, S. Selvakumar, A. Sokolov, S. Suman, R. Thorne, P. Totoo, S. Vijayaraja, Z. Waheed, A. Zyoud, R. Lopez, T. Burdett, and G. Cochrane (2021). "The European Nucleotide Archive in 2021". In: *Nucleic Acids Research* 50.D1, pp. D106–D110.

Dayan, G. H., N. Mohamed, I. L. Scully, D. Cooper, E. Begier, J. Eiden, K. U. Jansen, A. Gurtman, and A. S. Anderson (2016). "Staphylococcus aureus: the current state of disease, pathophysiology and strategies for prevention". In: *Expert Review of Vaccines* 15.11. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/14760584.2016.1179583, pp. 1373–1392.

Denham, E. L. (2020). "The Sponge RNAs of bacteria – How to find them and their role in regulating the post-transcriptional network". en. In: *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* 1863.8, p. 194565.

Diamantopoulos, M. A., P. Tsiakanikas, and A. Scorilas (2018). "Non-coding RNAs: the riddle of the transcriptome and their perspectives in cancer". en. In: *Annals of Translational Medicine* 6.12. Number: 12 Publisher: AME Publishing Company, pp. 241–241.

Dietrich, N., M. Rohde, R. Geffers, A. Kröger, H. Hauser, S. Weiss, and N. O. Gekara (2010). "Mast cells elicit proinflammatory but not type I interferon responses upon activation of TLRs by bacteria". eng. In: *Proceedings of the National Academy of Sciences of the United States of America* 107.19, pp. 8748–8753.

Dillies, M.-A., A. Rau, J. Aubert, C. Hennequet-Antier, M. Jeanmougin, N. Servant, C. Keime, G. Marot, D. Castel, J. Estelle, G. Guernec, B. Jagla, L. Jouneau, D. Laloë,

C. Le Gall, B. Schaëffer, S. Le Crom, M. Guedj, F. Jaffrézic, and on behalf of The French StatOmique Consortium (2013). "A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis". In: *Briefings in Bioinformatics* 14.6, pp. 671–683.

National Human Genome Research Institute - Sequencing costs (2022). *DNA Sequencing Costs: https://www.genome.gov/about-genomics/fact-sheets/ DNA-Sequencing-Costs-Data*. en.

Dobin, A., C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, and T. R. Gingeras (2013). "STAR: ultrafast universal RNA-seq aligner". eng. In: *Bioinformatics (Oxford, England)* 29.1, pp. 15–21.

Donato, L., C. Scimone, C. Rinaldi, R. D'Angelo, and A. Sidoti (2021). "New evaluation methods of read mapping by 17 aligners on simulated and empirical NGS data: an updated comparison of DNA- and RNA-Seq data from Illumina and Ion Torrent technologies". en. In: *Neural Computing and Applications* 33.22, pp. 15669–15692.

Dutta, T. and S. Srivastava (2018). "Small RNA-mediated regulation in bacteria: A growing palette of diverse mechanisms". eng. In: *Gene* 656, pp. 60–72.

Eid, J., A. Fehr, J. Gray, K. Luong, J. Lyle, G. Otto, P. Peluso, D. Rank, P. Baybayan, B. Bettman, A. Bibillo, K. Bjornson, B. Chaudhuri, F. Christians, R. Cicero, S. Clark, R. Dalal, A. Dewinter, J. Dixon, M. Foquet, A. Gaertner, P. Hardenbol, C. Heiner, K. Hester, D. Holden, G. Kearns, X. Kong, R. Kuse, Y. Lacroix, S. Lin, P. Lundquist, C. Ma, P. Marks, M. Maxham, D. Murphy, I. Park, T. Pham, M. Phillips, J. Roy, R. Sebra, G. Shen, J. Sorenson, A. Tomaney, K. Travers, M. Trulson, J. Vieceli, J. Wegener, D. Wu, A. Yang, D. Zaccarin, P. Zhao, F. Zhong, J. Korlach, and S. Turner (2009). "Real-time DNA sequencing from single polymerase molecules". eng. In: *Science (New York, N.Y.)* 323.5910, pp. 133–138.

ENCODE Project Consortium (2011). "A user's guide to the encyclopedia of DNA elements (ENCODE)". eng. In: *PLoS biology* 9.4, e1001046.

ENSEMBL - GFF3 File Format - https://www.ensembl.org/info/website/upload/gff3.html, 2022-10-07 (2022).

ENSEMBL - WIG File Format - https://www.ensembl.org/info /website/upload/wig.html - 2022-10-07 (2022).

Espindula, E., E. R. Sperb, E. Bach, and L. M. P. Passaglia (2020). "The combined analysis as the best strategy for Dual RNA-Seq mapping". In: *Genetics and Molecular Biology* 42.4, e20190215.

European Organization For Nuclear Research and OpenAIRE (2013). *Zenodo*. en.

Ewels, P. A., A. Peltzer, S. Fillinger, H. Patel, J. Alneberg, A. Wilm, M. U. Garcia, P. Di Tommaso, and S. Nahnsen (2020). "The nf-core framework for community-curated bioinformatics pipelines". en. In: *Nature Biotechnology* 38.3, pp. 276–278.

Fiers, W., R. Contreras, F. Duerinck, G. Haegeman, D. Iserentant, J. Merregaert, W. Min Jou, F. Molemans, A. Raeymaekers, A. Van den Berghe, G. Volckaert, and M. Ysebaert (1976). "Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene". en. In: *Nature* 260.5551. Number: 5551 Publisher: Nature Publishing Group, pp. 500–507.

Förstner, K. U., J. Vogel, and C. M. Sharma (2014). "READemption-a tool for the computational analysis of deep-sequencing-based transcriptome data". eng. In: *Bioinformatics (Oxford, England)* 30.23, pp. 3421–3423.

Freedman, A. H., J. M. Gaspar, and T. B. Sackton (2020). "Short paired-end reads trump long single-end reads for expression analysis". In: *BMC Bioinformatics* 21.1, p. 149.

Freese, N. H., D. C. Norris, and A. E. Loraine (2016). "Integrated genome browser: visual analytics platform for genomics". eng. In: *Bioinformatics (Oxford, England)* 32.14, pp. 2089–2095.

Garalde, D. R., E. A. Snell, D. Jachimowicz, B. Sipos, J. H. Lloyd, M. Bruce, N. Pantic, T. Admassu, P. James, A. Warland, M. Jordan, J. Ciccone, S. Serra, J. Keenan, S. Martin, L. McNeill, E. J. Wallace, L. Jayasinghe, C. Wright, J. Blasco, S. Young, D. Brocklebank, S. Juul, J. Clarke, A. J. Heron, and D. J. Turner (2018). "Highly parallel direct RNA sequencing on an array of nanopores". en. In: *Nature Methods* 15.3. Number: 3 Publisher: Nature Publishing Group, pp. 201–206.

Gillespie, M., B. Jassal, R. Stephan, M. Milacic, K. Rothfels, A. Senff-Ribeiro, J. Griss, C. Sevilla, L. Matthews, C. Gong, C. Deng, T. Varusai, E. Ragueneau, Y. Haider, B. May, V. Shamovsky, J. Weiser, T. Brunson, N. Sanati, L. Beckman, X. Shao, A. Fabregat, K. Sidiropoulos, J. Murillo, G. Viteri, J. Cook, S. Shorser, G. Bader, E. Demir, C. Sander, R. Haw, G. Wu, L. Stein, H. Hermjakob, and P. D'Eustachio (2022). "The reactome pathway knowledgebase 2022". In: *Nucleic Acids Research* 50.D1, pp. D687–D692.

Gong, J. Q., L. Lin, T. Lin, F. Hao, F. Q. Zeng, Z. G. Bi, D. Yi, and B. Zhao (2006). "Skin colonization by Staphylococcus aureus in patients with eczema and atopic dermatitis and relevant combined topical therapy: a double-blind multicentre randomized controlled trial". eng. In: *The British Journal of Dermatology* 155.4, pp. 680–687.

Guerrier-Takada, C., K. Gardiner, T. Marsh, N. Pace, and S. Altman (1983). "The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme". English. In: *Cell* 35.3. Publisher: Elsevier, pp. 849–857.

Haas, B. J., M. Chin, C. Nusbaum, B. W. Birren, and J. Livny (2012). "How deep is deep enough for RNA-Seq profiling of bacterial transcriptomes?" In: *BMC Genomics* 13.1, p. 734.

Hardcastle, T. J. and K. A. Kelly (2010). "baySeq: Empirical Bayesian methods for identifying differential expression in sequence count data". In: *BMC Bioinformatics* 11.1, p. 422.

Hayes, S. M., R. Howlin, D. A. Johnston, J. S. Webb, S. C. Clarke, P. Stoodley, P. G. Harries, S. J. Wilson, S. L. F. Pender, S. N. Faust, L. Hall-Stoodley, and R. J. Salib (2015). "Intracellular residency of Staphylococcus aureus within mast cells in nasal polyps: A novel observation". English. In: *Journal of Allergy and Clinical Immunology* 135.6. Publisher: Elsevier, 1648–1651.e5.

Henras, A. K., C. Dez, and Y. Henry (2004). "RNA structure and function in C/D and H/ACA s(no)RNPs". eng. In: *Current Opinion in Structural Biology* 14.3, pp. 335–343.

Higaki, S., M. Morohashi, T. Yamagishi, and Y. Hasegawa (1999). "Comparative study of staphylococci from the skin of atopic dermatitis patients and from healthy subjects". eng. In: *International Journal of Dermatology* 38.4, pp. 265–269.

Hoffmann, S., C. Otto, G. Doose, A. Tanzer, D. Langenberger, S. Christ, M. Kunz, L. M. Holdt, D. Teupser, J. Hackermüller, and P. F. Stadler (2014). "A multi-split mapping algorithm for circular RNA, splicing, trans-splicing and fusion detection". eng. In: *Genome Biology* 15.2, R34.

Holley, R. W., J. Apgar, G. A. Everett, J. T. Madison, M. Marquisee, S. H. Merrill, J. R. Penswick, and A. Zamir (1965). "Structure of a Ribonucleic Acid". In: *Science* 147.3664. Publisher: American Association for the Advancement of Science, pp. 1462–1465.

Human Genome Project Fact Sheet (2022). en.

Humphrys, M. S., T. Creasy, Y. Sun, A. C. Shetty, M. C. Chibucos, E. F. Drabek, C. M. Fraser, U. Farooq, N. Sengamalay, S. Ott, H. Shou, P. M. Bavoil, A. Mahurkar, and G. S. A. Myers (2013). "Simultaneous Transcriptional Profiling of Bacteria and Their Host Cells". en. In: *PLOS ONE* 8.12. Publisher: Public Library of Science, e80597.

Ignatiadis, N., B. Klaus, J. B. Zaugg, and W. Huber (2016). "Data-driven hypothesis weighting increases detection power in genome-scale multiple testing". en. In: *Nature Methods* 13.7. Number: 7 Publisher: Nature Publishing Group, pp. 577–580.

Inouye, M. and N. Delihas (1988). "Small RNAs in the prokaryotes: a growing list of diverse roles". eng. In: *Cell* 53.1, pp. 5–7.

Jain, M., S. Koren, K. H. Miga, J. Quick, A. C. Rand, T. A. Sasani, J. R. Tyson, A. D. Beggs, A. T. Dilthey, I. T. Fiddes, S. Malla, H. Marriott, T. Nieto, J. O'Grady, H. E. Olsen, B. S. Pedersen, A. Rhie, H. Richardson, A. R. Quinlan, T. P. Snutch, L. Tee, B. Paten, A. M. Phillippy, J. T. Simpson, N. J. Loman, and M. Loose (2018). "Nanopore

sequencing and assembly of a human genome with ultra-long reads". en. In: *Nature Biotechnology* 36.4. Number: 4 Publisher: Nature Publishing Group, pp. 338–345.

Jarroux, J., A. Morillon, and M. Pinskaya (2017). "History, Discovery, and Classification of lncRNAs". en. In: 1008. Ed. by M. Rao. Series Title: Advances in Experimental Medicine and Biology, pp. 1–46.

Jeck, W. R. and N. E. Sharpless (2014). "Detecting and characterizing circular RNAs". In: *Nature biotechnology* 32.5, pp. 453–461.

Kanehisa, M., S. Goto, M. Furumichi, M. Tanabe, and M. Hirakawa (2010). "KEGG for representation and analysis of molecular networks involving diseases and drugs". eng. In: *Nucleic Acids Research* 38.Database issue, pp. D355–360.

Keiler, K. C. and N. S. Ramadoss (2011). "Bifunctional transfer-messenger RNA". In: *Biochimie* 93.11, pp. 1993–1997.

Kim, D., J. M. Paggi, C. Park, C. Bennett, and S. L. Salzberg (2019). "Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype". en. In: *Nature Biotechnology* 37.8. Number: 8 Publisher: Nature Publishing Group, pp. 907–915.

Kruger, K., P. J. Grabowski, A. J. Zaug, J. Sands, D. E. Gottschling, and T. R. Cech (1982). "Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of tetrahymena". English. In: *Cell* 31.1. Publisher: Elsevier, pp. 147–157.

Lam, J. K. W., M. Y. T. Chow, Y. Zhang, and S. W. S. Leung (2015). "siRNA Versus miRNA as Therapeutics for Gene Silencing". English. In: *Molecular Therapy - Nucleic Acids* 4. Publisher: Elsevier.

Lander, E. S. et al. (2001). "Initial sequencing and analysis of the human genome". en. In: *Nature* 409.6822. Number: 6822 Publisher: Nature Publishing Group, pp. 860–921.

Langmead, B., C. Trapnell, M. Pop, and S. L. Salzberg (2009). "Ultrafast and memory-efficient alignment of short DNA sequences to the human genome". In: *Genome Biology* 10.3, R25.

Lee, R. C., R. L. Feinbaum, and V. Ambros (1993). "The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14". eng. In: *Cell* 75.5, pp. 843–854.

Leprevost, F. d. V., V. C. Barbosa, E. L. Francisco, Y. Perez-Riverol, and P. C. Carvalho (2014). "On best practices in the development of bioinformatics software". In: *Frontiers in Genetics* 5.

Lesman, D., Y. Rodriguez, D. Rajakumar, and N. Wein (2021). "U7 snRNA, a Small RNA with a Big Impact in Gene Therapy". eng. In: *Human Gene Therapy* 32.21-22, pp. 1317–1329.

Lewis, J. B., J. F. Atkins, C. W. Anderson, P. R. Baum, and R. F. Gesteland (1975). "Mapping of late adenovirus genes by cell-free translation of RNA selected by hybridization to specific DNA fragments". eng. In: *Proceedings of the National Academy of Sciences of the United States of America* 72.4, pp. 1344–1348.

Li, H. and R. Durbin (2009). "Fast and accurate short read alignment with Burrows-Wheeler transform". eng. In: *Bioinformatics (Oxford, England)* 25.14, pp. 1754–1760.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin (2009). "The Sequence Alignment/Map format and SAMtools". In: *Bioinformatics* 25.16, pp. 2078–2079.

Li, J. and R. Tibshirani (2013). "Finding consistent patterns: a nonparametric approach for identifying differential expression in RNA-Seq data". eng. In: *Statistical Methods in Medical Research* 22.5, pp. 519–536.

Lin, Y.-T., C.-T. Wang, and B.-L. Chiang (2007). "Role of bacterial pathogens in atopic dermatitis". eng. In: *Clinical Reviews in Allergy & Immunology* 33.3, pp. 167–177.

Liu, Y., J. Zhou, and K. P. White (2014). "RNA-seq differential expression studies: more sequence or more replication?" In: *Bioinformatics* 30.3, pp. 301–304.

Love, M. I., W. Huber, and S. Anders (2014). "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2". In: *Genome Biology* 15.12, p. 550.

Macfarlane, L.-A. and P. R. Murphy (2010). "MicroRNA: Biogenesis, Function and Role in Cancer". eng. In: *Current Genomics* 11.7, pp. 537–561.

Margulies, M., M. Egholm, W. E. Altman, S. Attiya, J. S. Bader, L. A. Bemben, J. Berka, M. S. Braverman, Y.-J. Chen, Z. Chen, S. B. Dewell, L. Du, J. M. Fierro, X. V. Gomes, B. C. Godwin, W. He, S. Helgesen, C. H. Ho, G. P. Irzyk, S. C. Jando, M. L. I. Alenquer, T. P. Jarvie, K. B. Jirage, J.-B. Kim, J. R. Knight, J. R. Lanza, J. H. Leamon, S. M. Lefkowitz, M. Lei, J. Li, K. L. Lohman, H. Lu, V. B. Makhijani, K. E. McDade, M. P. McKenna, E. W. Myers, E. Nickerson, J. R. Nobile, R. Plant, B. P. Puc, M. T. Ronan, G. T. Roth, G. J. Sarkis, J. F. Simons, J. W. Simpson, M. Srinivasan, K. R. Tartaro, A. Tomasz, K. A. Vogt, G. A. Volkmer, S. H. Wang, Y. Wang, M. P. Weiner, P. Yu, R. F. Begley, and J. M. Rothberg (2005). "Genome sequencing in microfabricated high-density picolitre reactors". en. In: *Nature* 437.7057. Number: 7057 Publisher: Nature Publishing Group, pp. 376–380.

Martin, M. (2011). "Cutadapt removes adapter sequences from high-throughput sequencing reads". en. In: *EMBnet.journal* 17.1. Number: 1, pp. 10–12.

Osborne, J. D., J. Flatow, M. Holko, S. M. Lin, W. A. Kibbe, L. J. Zhu, M. I. Danila, G. Feng, and R. L. Chisholm (2009). "Annotating the human genome with Disease Ontology". eng. In: *BMC genomics* 10 Suppl 1, S6.

Otto, C., P. F. Stadler, and S. Hoffmann (2014). "Lacking alignments? The next-generation sequencing mapper segemehl revisited". eng. In: *Bioinformatics (Oxford, England)* 30.13, pp. 1837–1843.

Pérez-Rubio, P., C. Lottaz, and J. C. Engelmann (2019). "FastqPuri: high-performance preprocessing of RNA-seq data". In: *BMC Bioinformatics* 20.1, p. 226.

Pruitt, K. D., T. Tatusova, and D. R. Maglott (2007). "NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins". In: *Nucleic Acids Research* 35.suppl_1, pp. D61–D65.

Reichow, S. L., T. Hamma, A. R. Ferré-D'Amaré, and G. Varani (2007). "The structure and function of small nucleolar ribonucleoproteins". eng. In: *Nucleic Acids Research* 35.5, pp. 1452–1464.

Reinhold-Hurek, B. and D. A. Shub (1992). "Self-splicing introns in tRNA genes of widely divergent bacteria". eng. In: *Nature* 357.6374, pp. 173–176.

Robinson, J. T., H. Thorvaldsdóttir, W. Winckler, M. Guttman, E. S. Lander, G. Getz, and J. P. Mesirov (2011). "Integrative genomics viewer". en. In: *Nature Biotechnology* 29.1. Number: 1 Publisher: Nature Publishing Group, pp. 24–26.

Robinson, M. D., D. J. McCarthy, and G. K. Smyth (2010). "edgeR: a Bioconductor package for differential expression analysis of digital gene expression data". In: *Bioinformatics* 26.1, pp. 139–140.

Rothberg, J. M., W. Hinz, T. M. Rearick, J. Schultz, W. Mileski, M. Davey, J. H. Leamon, K. Johnson, M. J. Milgrew, M. Edwards, J. Hoon, J. F. Simons, D. Marran, J. W. Myers, J. F. Davidson, A. Branting, J. R. Nobile, B. P. Puc, D. Light, T. A. Clark, M. Huber, J. T. Branciforte, I. B. Stoner, S. E. Cawley, M. Lyons, Y. Fu, N. Homer, M. Sedova, X. Miao, B. Reed, J. Sabina, E. Feierstein, M. Schorn, M. Alanjary, E. Dimalanta, D. Dressman, R. Kasinskas, T. Sokolsky, J. A. Fidanza, E. Namsaraev, K. J. McKernan, A. Williams, G. T. Roth, and J. Bustillo (2011). "An integrated semiconductor device enabling non-optical genome sequencing". eng. In: *Nature* 475.7356, pp. 348–352.

Schatz, M. C. (2017). "Nanopore sequencing meets epigenetics". en. In: *Nature Methods* 14.4. Number: 4 Publisher: Nature Publishing Group, pp. 347–348.

Sequence Read Archive - Bases in database, https://www.ncbi.nlm.nih.gov/sra/docs/sragrowth/ (2022).

Shakya, M., C.-C. Lo, and P. S. G. Chain (2019). "Advances and Challenges in Metatranscriptomic Analysis". In: *Frontiers in Genetics* 10.

Sharma, C. M., S. Hoffmann, F. Darfeuille, J. Reignier, S. Findeiss, A. Sittka, S. Chabas, K. Reiche, J. Hackermüller, R. Reinhardt, P. F. Stadler, and J. Vogel (2010). "The primary transcriptome of the major human pathogen Helicobacter pylori". eng. In: *Nature* 464.7286, pp. 250–255.

Siomi, M. C., K. Sato, D. Pezic, and A. A. Aravin (2011). "PIWI-interacting small RNAs: the vanguard of genome defence". eng. In: *Nature Reviews. Molecular Cell Biology* 12.4, pp. 246–258.

Stark, R., M. Grzelak, and J. Hadfield (2019). "RNA sequencing: the teenage years". en. In: *Nature Reviews Genetics* 20.11. Number: 11 Publisher: Nature Publishing Group, pp. 631–656.

Statello, L., C.-J. Guo, L.-L. Chen, and M. Huarte (2021). "Gene regulation by long non-coding RNAs and its biological functions". en. In: *Nature Reviews Molecular Cell Biology* 22.2. Number: 2 Publisher: Nature Publishing Group, pp. 96–118.

Subramanian, A., P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, and J. P. Mesirov (2005). "Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles". In: *Proceedings of the National Academy of Sciences* 102.43. Publisher: Proceedings of the National Academy of Sciences, pp. 15545–15550.

Tarazona, S., P. Furió-Tarí, D. Turrà, A. D. Pietro, M. J. Nueda, A. Ferrer, and A. Conesa (2015). "Data quality aware analysis of differential expression in RNA-seq with NOISeq R/Bioc package". In: *Nucleic Acids Research* 43.21, e140.

Tarazona, S., F. García-Alcalde, J. Dopazo, A. Ferrer, and A. Conesa (2011). "Differential expression in RNA-seq: a matter of depth". eng. In: *Genome Research* 21.12, pp. 2213–2223.

Tierney, L., J. Linde, S. Müller, S. Brunke, J. C. Molina, B. Hube, U. Schöck, R. Guthke, and K. Kuchler (2012). "An Interspecies Regulatory Network Inferred from Simultaneous RNA-seq of Candida albicans Invading Innate Immune Cells". In: *Frontiers in Microbiology* 3, p. 85.

TruSeq DNA Sample Preparation Guide (2022).

Valadkhan, S. (2005). "snRNAs as the catalysts of pre-mRNA splicing". eng. In: *Current Opinion in Chemical Biology* 9.6, pp. 603–608.

Vannucci, F. A., D. N. Foster, and C. J. Gebhart (2013). "Laser microdissection coupled with RNA-seq analysis of porcine enterocytes infected with an obligate intracellular pathogen (Lawsonia intracellularis)". In: *BMC Genomics* 14, p. 421.

Westermann, A. J., L. Barquist, and J. Vogel (2017). "Resolving host–pathogen interactions by dual RNA-seq". en. In: *PLOS Pathogens* 13.2. Publisher: Public Library of Science, e1006033.

Westermann, A. J., K. U. Förstner, F. Amman, L. Barquist, Y. Chao, L. N. Schulte, L. Müller, R. Reinhardt, P. F. Stadler, and J. Vogel (2016). "Dual RNA-seq unveils noncoding RNA functions in host–pathogen interactions". en. In: *Nature* 529.7587. Number: 7587 Publisher: Nature Publishing Group, pp. 496–501.

Westermann, A. J., S. A. Gorski, and J. Vogel (2012). "Dual RNA-seq of pathogen and host". en. In: *Nature Reviews Microbiology* 10.9. Number: 9 Publisher: Nature Publishing Group, pp. 618–630.

Wheeler, D. A., M. Srinivasan, M. Egholm, Y. Shen, L. Chen, A. McGuire, W. He, Y.-J. Chen, V. Makhijani, G. T. Roth, X. Gomes, K. Tartaro, F. Niazi, C. L. Turcotte, G. P. Irzyk, J. R. Lupski, C. Chinault, X.-z. Song, Y. Liu, Y. Yuan, L. Nazareth, X. Qin, D. M. Muzny, M. Margulies, G. M. Weinstock, R. A. Gibbs, and J. M. Rothberg (2008). "The complete genome of an individual by massively parallel DNA sequencing". en. In: *Nature* 452.7189. Number: 7189 Publisher: Nature Publishing Group, pp. 872–876.

Wilson, G., D. A. Aruliah, C. T. Brown, N. P. C. Hong, M. Davis, R. T. Guy, S. H. D. Haddock, K. D. Huff, I. M. Mitchell, M. D. Plumbley, B. Waugh, E. P. White, and P. Wilson (2014). "Best Practices for Scientific Computing". en. In: *PLOS Biology* 12.1. Publisher: Public Library of Science, e1001745.

Wolf, T., P. Kämmer, S. Brunke, and J. Linde (2018). "Two's company: studying interspecies relationships with dual RNA-seq". en. In: *Current Opinion in Microbiology*. Cell Regulation 42, pp. 7–12.

Wright, D. J., N. A. L. Hall, N. Irish, A. L. Man, W. Glynn, A. Mould, A. D. L. Angeles, E. Angiolini, D. Swarbreck, K. Gharbi, E. M. Tunbridge, and W. Haerty (2022). "Long read sequencing reveals novel isoforms and insights into splicing regulation during cell state changes". In: *BMC Genomics* 23.1, p. 42.

Wu, T., E. Hu, S. Xu, M. Chen, P. Guo, Z. Dai, T. Feng, L. Zhou, W. Tang, L. Zhan, X. Fu, S. Liu, X. Bo, and G. Yu (2021). "clusterProfiler 4.0: A universal enrichment tool for interpreting omics data". English. In: *The Innovation* 2.3. Publisher: Elsevier.

Yu, C.-Y. and H.-C. Kuo (2019). "The emerging roles and functions of circular RNAs and their generation". In: *Journal of Biomedical Science* 26.1, p. 29.

Yu, S.-H., J. Vogel, and K. U. Förstner (2018). "ANNOgesic: a Swiss army knife for the RNA-seq based annotation of bacterial/archaeal genomes". eng. In: *GigaScience* 7.9.

Zhao, Y., M.-C. Li, M. M. Konaté, L. Chen, B. Das, C. Karlovich, P. M. Williams, Y. A. Evrard, J. H. Doroshow, and L. M. McShane (2021). "TPM, FPKM, or Normalized Counts? A Comparative Study of Quantification Measures for the Analysis of RNA-seq Data from the NCI Patient-Derived Models Repository". In: *Journal of Translational Medicine* 19.1, p. 269.

A Abbreviations

S. aureus	*Staphylococcus aureus*
AD	atopic dermatitis
BAM	binary alignment/map
cDNA	complementary DNA
CDS	coding DNA sequence
CI	continuous integration
circRNA	circular RNA
DGE	differential gene expression
DNA	deoxyribonucleic acid
DNA-seq	DNA sequencing
dNTP	deoxyribonucleotide triphosphates
DOI	digital object identifier
dual RNA-seq	dual RNA sequencing
ENA	European Nucleotide Archive
FACS	fluorescence-activated cell sorting
FPKM	fragments per kilobase million
GFF	general feature format
GSEA	gene set enrichment analysis
GUI	graphical user interface
IFN-I	type-I interferon
IGB	Integrated Genome Browser
IGV	Integrative Genomics Viewer
KEGG	Kyoto Encyclopedia of Genes and Genomes
lncRNA	long non-coding RNA
miRNA	micro RNA
mRNA	messenger RNA
multi RNA-seq	multi RNA sequencing
ncRNA	non-coding RNA
nm	nanometer
ORFs	open reading frames
PCR	polymerase chain reaction

pg	picogram
piRNA	P-element-induced wimpy testis (Piwi)-interacting RNA
RISC	RNA-induced silencing complexes
RNA	ribonucleic acid
RNA-seq	RNA sequencing
RPKM	reads per kilobase million
rRNA	ribosomal RNA
SAM	sequence alignment/map
scaRNA	small Cajal body-specific RNA
scRNA-seq	single-cell RNA sequencing
siRNA	small interfering RNA
snoRNA	small nucleolar RNA
snRNA	small nuclear RNA
SRA	Sequence Read Archive
sRNA	small RNA
tmRNA	transfer-messenger RNA
TPM	transcripts per million
tRNA	transfer RNA
UTR	untranslated region

B List of Figures

C List of Tables

D Statement of individual author contributions and of legal second publication rights to manuscripts included in the dissertation

Manuscript 1 (complete reference): Oliver Goldmann , Till Sauerwein, Gabriella Molinari, Manfred Rohde, Konrad U. Förstner and Eva Medina, " Cytosolic Sensing of Intracellular Staphylococcus aureus by Mast Cells Elicits a Type I IFN Response That Enhances Cell-Autonomous Immunity" J Immunol March 23, 2022, ji2100622;

DOI: https://doi.org/10.4049/jimmunol.2100622

Participated in	**Author Initials,** Responsibility decreasing from left to right				
Study Design Methods Development	O.G.	K.F.	E.M.		
Data Collection	O.G.	M.R.	E.M.	K.F.	
Data Analysis and Interpretation	O.G.	T.S.	E.M.	K.F.	M.R.&G.M.
Manuscript Writing	E.M.	O.G.	T.S.	K.F.	M.R.&G.M.
Writing of Introduction	E.M.	O.G.			
Writing of Materials & Methods	O.G.	T.S.	E.M.	K.F.	M.R.&G.M.
Writing of Discussion	E.M.	O.G.			
Writing of First Draft	E.M.	O.G.			

Explanations (if applicable):

Manuscript 2 (complete reference): Till Sauerwein, Konrad U. Förstner, Thorsten Bischler, "READemption 2: Multi-species RNA-Seq made easy" bioRxiv, October 03, 2022,

DOI:https://doi.org/10.1101/2022.09.30.510338

Participated in	**Author Initials,** Responsibility decreasing from left to right				
Study Design Methods Development	T.S.	K.F.			
Software Development	T.S.	K.F.	T.B.		
Manuscript Writing	T.S.	K.F.			
Writing of Introduction	T.S.	K.F.			
Writing of Materials & Methods	T.S.	K.F.			
Writing of Discussion	T.S.	K.F.			
Writing of First Draft	T.S.	K.F.			

Explanations (if applicable):

If applicable, the doctoral researcher confirms that she/he has obtained permission from both the publishers (copyright) and the co-authors for legal second publication.

The doctoral researcher and the primary supervisor confirm the correctness of the above mentioned assessment.

Doctoral Researcher's Name	Date	Place	Signature
Konrad Förstner	24.10.2022	Köln	

Primary Supervisor's Name	Date	Place	Signature

Manuscript 1 (complete reference): Oliver Goldmann , Till Sauerwein, Gabriella Molinari, Manfred Rohde, Konrad U. Förstner and Eva Medina, " Cytosolic Sensing of Intracellular Staphylococcus aureus by Mast Cells Elicits a Type I IFN Response That Enhances Cell-Autonomous Immunity" J Immunol March 23, 2022, ji2100622;

DOI: https://doi.org/10.4049/jimmunol.2100622

Figure	**Author Initials,** Responsibility decreasing from left to right				
1	T.S.	E.M.	M.R.	G.M.	O.G.
2	E.M.	T.S.			
3	T.S.	E.M.			
4	O.G.	E.M.			
5	O.G.	E.M.			

Explanations (if applicable):

Manuscript 2 (complete reference): Till Sauerwein, Konrad U. Förstner, Thorsten Bischler, "READemption 2: Multi-species RNA-Seq made easy" bioRxiv, October 03, 2022, DOI:https://doi.org/10.1101/2022.09.30.510338

Figure	**Author Initials,** Responsibility decreasing from left to right				
1	T.S.	K.F.			
2	T.S.	K.F.			
3	T.S.	K.F.			

Explanations (if applicable):

I also confirm my primary supervisor's acceptance.

Doctoral Researcher's Name	Date	Place	Signature

E Publications

Lang JC, Seiß EA, Moldovan A, Müsken M, **Sauerwein T**, Fraunholz M, Müller AJ, Goldmann O, Medina E.
A Photoconvertible Reporter System for Bacterial Metabolic Activity Reveals That Staphylococcus aureus Enters a Dormant-Like State to Persist within Macrophages.
mBio. 2022 Sep 14:e0231622. DOI: 10.1128/mbio.02316-22. Online ahead of print.

Goldmann O, **Sauerwein T**, Molinari G, Rohde M, Förstner KU, Medina E.
Cytosolic Sensing of Intracellular Staphylococcus aureus by Mast Cells Elicits a Type I IFN Response That Enhances Cell-Autonomous Immunity.
J Immunol. 2022 Apr 1;208(7):1675-1685. DOI: 10.4049/jimmunol.2100622. Epub 2022 Mar 23.

Martins Gomes SF, Westermann AJ, **Sauerwein T**, Hertlein T, Förstner KU, Ohlsen K, Metzger M, Shusta EV, Kim BJ, Appelt-Menzel A, Schubert-Unkmeir A.
Induced Pluripotent Stem Cell-Derived Brain Endothelial Cells as a Cellular Model to Study Neisseria meningitidis Infection.
Front Microbiol. 2019 May 29;10:1181. DOI: 10.3389/fmicb.2019.01181. eCollection 2019.

F Curriculum Vitae

G Danksagung

Ich bedanke mich bei Konrad Förstner dafür, dass er mir ermöglicht hat diese Arbeit zu schreiben und mich auf diesem Weg begleitet und angeleitet hat. Er stand mir immer mit Rat und Tat zur Seite, gab mir das Vertrauen den Anforderungen gewachsen zu sein und wertschätzte stets meine erbrachte Arbeit. Außerdem möchte ich mich dafür bedanken, dass er mir vor ein paar Jahren die wunderbare Welt der Bioinformatik gezeigt hat und so den Weg zur Erstellung dieser Arbeit geebnet hat.

Außerdem möchte ich mich bei meinen weiteren Betreuern Thomas Dandekar und Alexander Westermann für die konstruktiven und ermutigenden Besprechungen meiner Ergebnisse bedanken.

Weiterhin gilt mein Dank allen Kooperationspartnern, insbesondere Eva Medina und Oliver Goldmann für die gute Zusammenarbeit in Chapter 1.

Ich möchte mich auch bei allen aktuellen und ehemaligen Kollegen bedanken. Sung-Huan, Malvika und Thorsten dafür, dass sie mir meinen Einstieg in die Bioinformatik und das Programmieren erleichtert haben, Silvia dafür, dass sie mir bei allen administrativen Aufgaben zu Beginn der Promotion weiter half und Rabea, Vanessa, Klaus, Mandela, Muhammad, Arindam, Eva, Richa und Panagiota für die schöne Zeit.

Julia möchte ich dafür danken, dass sie mich vor allem in der letzten Phase der Promotion in allen Belangen unterstützt hat und die richtigen Worte fand, wenn ich sie benötigte.

Großer Dank geht auch an meine Eltern Ingrid und Gerd und meinem Bruder Kai, die mich bei allen wichtigen Entscheidungen in meinem Leben unterstützt haben und immer für mich da sind, wenn ich sie brauche.

Affidavit

I hereby confirm that my thesis entitled "Implementation and application of bioinformatical software for the analysis of dual RNA sequencing data of host and pathogen during infection" 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 bioinformatischer Software für die Analyse von dual RNA-Sequenzierdaten von Wirt und Erreger während Infektion" 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