Chlamydia trachomatis metabolism during infection and metatranscriptome analysis in *Neisseria gonorrhoeae* coinfected STD patients

Chlamydia trachomatis Metabolismus während der Infektion sowie die Analyse des Metatranskriptoms bei *Neisseria gonorrhoeae* koinfizierten STD-Patienten

submitted by

MANLI YANG

aus Ganzhou, China

Würzburg 2019

Submitted on:

Members of the Thesis Committee

Chairperson: Prof. Dr. Markus Sauer

Primary Supervisor: Prof. Dr. Thomas Dandekar

Second Supervisor: Prof. Dr. Thomas Rudel

Third Supervisor: Dr. Tobias Müller

Date of Public Defence:

Date of Receipt of Certificates:

Abstract

Chlamydia trachomatis (*Ct*) is an obligate intracellular human pathogen. It causes blinding trachoma and sexually transmitted disease such as chlamydia, pelvic inflammatory disease and lymphogranuloma venereum. *Ct* has a unique biphasic development cycle and replicates in an intracellular vacuole called inclusion. Normally it has two forms: the infectious form, elementary body (EB); and the non-infectious form, reticulate body (RB). *Ct* is not easily amenable to genetic manipulation. Hence, to understand the infection process, it is crucial to study how the metabolic activity of *Ct* exactly evolves in the host cell and what roles of EB and RB play differentially in *Ct* metabolism during infection. In addition, *Ct* was found regularly coinfected with other pathogens in patients who got sexually transmitted diseases (STDs). A lack of powerful methods to culture *Ct* outside of the host cell makes the detailed molecular mechanisms of coinfection difficult to study.

In this work, a genome-scale metabolic model with 321 metabolites and 277 reactions was first reconstructed by me to study *Ct* metabolic adaptation in the host cell during infection. This model was calculated to yield 84 extreme pathways, and metabolic flux strength was then modelled regarding 20hpi, 40hpi and later based on a published proteomics dataset. Activities of key enzymes involved in target pathways were further validated by RT-qPCR in both HeLa229 and HUVEC cell lines. This study suggests that *Ct*'s major active pathways involve glycolysis, gluconeogenesis, glycerolphospholipid biosynthesis and pentose phosphate pathway, while *Ct*'s incomplete tricarboxylic acid cycle and fatty acid biosynthesis are less active. EB is more activated in almost all these carbohydrate pathways than RB. Result suggests the survival of *Ct* generally requires a lot of acetyl-CoA from the host. Besides, both EB and RB can utilize folate biosynthesis to generate NAD(P)H but may use different pathways depending on the demands of ATP. When more ATP is available from both host cell and *Ct* itself, RB is more activated by utilizing energy providing chemicals generated by enzymes associated in the nucleic acid metabolism. The forming of folate also

suggests large glutamate consumption, which is supposed to be converted from glutamine by the glutamine-fructose-6-phosphate transaminase ($glmS$) and CTP synthase ($pyrG$).

Then, RNA sequencing (RNA-seq) data analysis was performed by me in a coinfection study. Metatranscriptome from patient RNA-seq data provides a realistic overview. Thirteen patient samples were collected and sequenced by our collaborators. Six male samples were obtained by urethral swab, and seven female samples were collected by cervicovaginal lavage. All the samples were *Neisseria gonorrhoeae* (GC) positive, and half of them had coinfection with Ct. HISAT2 and Stringtie were used for transcriptomic mapping and assembly respectively, and differential expression analysis by DESeq2, Ballgown and Cuffdiff2 are parallelly processed for comparison. Although the measured transcripts were not sufficient to assemble Ct's transcriptome, the differential expression of genes in both the host and GC were analyzed by comparing Ct positive group (Ct^+) against Ct-uninfected group. The results show that in the Ct^+ group, the host MHC class II immune response was highly induced. Ct infection is associated with the regulation of DNA methylation, DNA double-strand damage and ubiquitination. The analysis also shows Ct infection enhances host fatty acid β oxidation, thereby inducing mROS, and the host responds to reduce ceramide production and glycolysis. The coinfection upregulates GC's own ion transporters and amino acid uptake, while it downregulates GC's restriction and modification systems. Meanwhile, GC has the nitrosative and oxidative stress response and also increases the ability for ferric uptake especially in the Ct^+ group compared to Ct-uninfected group.

In conclusion, methods in bioinformatics were used here in analyzing the metabolism of Ct itself, and the responses of the host and GC respectively in a coinfection study with and without Ct. These methods provide metabolic and metatranscriptomic details to study Ct metabolism during infection and Ct associated coinfection in the human microbiota.

Zusammenfassung

Chlamydia trachomatis (*Ct*) ist ein obligater intrazellulärer Pathogen des Menschen. Er verursacht Trachoma und sexuell übertragbare Krankheiten, wie Chlamydiose, Unterleibsentzündung und Lymphogranuloma venereum. *Ct* besitzt einen biphasischen Entwicklungszyklus und vermehrt sich in intrazellulären Vakuolen, sogenannten Einschlusskörperchen. Normalerweise können zwei Formen beobachtete werden: Die infektiöse Form, Elementarkörperchen (EK); und die nicht-infektiöse Form, Retikularkörperchen (RK). *Ct* ist nicht einfach genetisch zu manipulieren. Um den Infektionsablauf besser zu verstehen, ist es wichtig, zu untersuchen, wie sich genau die metabolische Aktivität von *Ct* in der Wirtszelle entwickelt und welche Rolle EK und RK im Metabolismus von *Ct* während der Infektion spielen. Zusätzlich wurde *Ct* häufig bei Patienten mit sexuell übertragbaren Krankheiten (STD) in Co-Infektion mit anderen Erregern gefunden. Ein Mangel an leistungsfähigen Methoden zur Kultivierung von *Ct* außerhalb der Wirtszelle macht es schwierig die genauen molekularen Mechanismen von Co-Infektionen zu untersuchen.

In dieser Arbeit wurde erstmals ein genomweites metabolisches Model mit 321 Metaboliten und 277 Reaktionen aufgebaut, um die metabolische Adaption von *Ct* in der Wirtzelle während der Infektion zu untersuchen. Dieses Model wurde erstellt und umfasst 84 „extreme pathways" (Grenz-Stoffwechselwege). Darauf aufbauend wurde die metabolische Fluss-Stärke berechnet. Die Zeitpunkte 20 hpi (20 Stunden nach der Infektion), 40 hpi und die anschließende Infektionsphase wurden durch Nutzung von Proteom-Daten modelliert. Die Aktivitäten von Schlüsselenzymen, welche in wichtigen Stoffwechselwegen involviert sind, wurden zusätzlich durch RT-qPCR überprüft. Dabei wurden die Ergebnisse sowohl für HeLA229- als auch HUVEC-Zellen nachgemessen. Diese Untersuchungen zeigten, dass *Ct*'s wichtigste aktive Stoffwechselwege die Glykolyse, die Gluconeogenese und der Pentosephosphatweg sind, während der unvollständige Zitronensäurezyklus und die Fettsäuresynthese weniger aktiv sind. Gegenüber RK sind bei EK fast alle diese Kohlenhydratwege stärker aktiviert. Im Allgemeinen benötigt *Ct* eine größere Menge an Acetyl-CoA.

Außerdem können sowohl EK, als auch RK die Folsäurebiosynthese nutzen, um NAD(P)H zu generieren. Dabei werden möglicherweise unterschiedliche Pathways genutzt, abhängig vom Bedarf an ATP. Sobald mehr ATP sowohl durch die Wirtszellen als auch von der Ct-Zelle selbst zur Verfügung steht, wird die Nutzung von Energieträgern, produziert durch Enzyme des Nukleinsäurestoffwechsels, in RK stärker aktiviert. Die Bildung von Folsäure lässt den Schluss zu, dass große Mengen von Glutamat umgesetzt werden, welches vermutlich aus der Umwandlung von Glutamin durch die Glutamine-fructose-6-phosphate-transaminase ($glmS$) und CTP-Syntase ($pyrG$) stammt.

Anschließend wurde eine Analyse von RNA-Sequenzierungsdaten (RNA-seq) aus einer Co-Infektions-Studie (Chlamydien und andere Keime, insbesondere Gonokokken (GC)) durchgeführt. Dafür wurden Proben von dreizehn Patienten gesammelt und von Kollaborationspartnern sequenziert. Sechs Proben männlicher Patienten wurden durch Abstrich der Harnröhre und sieben Proben weiblicher Patientinnen durch cervicovaginale Lavage gewonnen. Alle Proben waren *Neisseria gonorrhoeae* (GC) positiv, wobei die Hälfte eine Co-Infektion mit Ct aufwies. Die Programme HISAT2 and Stringtie wurden zum Abbilden der transgenomischen Reads beziehungsweise zur Assemblierung des Genoms verwendet, und eine Analyse der differentiellen Expression wurde jeweils mit DESeq2, Ballgown und Cuffdiff2 durchgeführt und die Ergebnisse verglichen. Obwohl nicht ausreichend viele Transkripte von Ct gewonnen werden konnten, um das Transkriptom komplett assemblieren zu können, wurde die differentielle Expression der Gene sowohl von Wirt als auch von GC durch den Vergleich zwischen der Gruppe der Ct-positiven (Ct^+) der Gruppe der Ct-unifizierten Patienten analysiert. Die Ergebnisse zeigten, dass in der Ct^+-Gruppe die auf der MHC-Klasse-II basierte Immunantwort stark induziert war. Die Infektion von Ct ist mit der Regulation der DNA-Methylierung, DNA-Doppel-Strang-Schädigung und Ubiquitinierung verbunden. Die Analyse zeigte zusätzlich, dass die Infektion mit Ct die Fettsäure -Oxidation des Wirts steigert, dadurch mROS induziert, und sowohl die Ceramid-Produktion als auch die Glycolyse reduziert. Die Co-Infektion reguliert GC's eigene Eisentransporter und Aminosäureaufnahme hoch, während Restriktions- und Mod-

ifikationssysteme herunterreguliert werden. Gleichzeitig zeigt GC sowohl eine stickstoff-sensitve Stress Antwort als auch eine oxidative. Dies verstärkt zusätzlich die Fähigkeit für die Aufnahme von Eisen, insbesondere in der Ct^+-Gruppe.

Zusammenfassend wurden Methoden der Bioinformatik genutzt, um den Metabolismus von Ct selbst, und die Antwort des Wirtes respektive GC's in einer Co-Infektionsstudie mit und ohne Ct zu analysieren. Diese Methoden lieferten wichtige metabolische und metatranskriptomische Details, um den Metabolismus von Ct während der Infektion, aber auch das Mikrobiom während einer Ct assoziierter Co-Infektion zu untersuchen.

Acknowledgements

First of all, I would particularly like to express my appreciation and thanks to my supervisor, Prof. Dr. Thomas Dandekar, for offering me the opportunity to work in the Department of Bioinformatics. Sincerely thanks for his kind supervision, nice discussions and suggestions.

I would also thank Prof. Dr. Thomas Rudel to be my second supervisor. I am grateful for his support and discussions on cell biology and microbiology, and many help in the collaboration of the *Chlamydia* project in GRK2157 3D tissue program.

Thanks to my third supervisor Dr. Tobias Müller for the advice in statistical analysis.

The work presented would not be possible without others help. I would like to thank Dr. Karthika Rajeeve for her support with the experiments and nice talks about *Chlamydia*. Thanks to Dr. Elena Bencúrová and Dr. Mugdha Srivastava for the sweet discussions on proteins and pathogen infections. Thanks to Nadine Vollmuth for the help with both our project and program. And many thanks to Qian Yu and Tao Yang for the discussions of *Neisseria*. I would also thank Christian Luther.

Thanks to Prof. Dr. Caroline A. Genco and Dr. Paola Masssari for the collaboration, and also for their reception of my visiting in Tufts University. I would appreciate and never forget the precious time in Boston. Thanks to Dr. Albert Tai for the help of data analysis.

I would like to thank my colleagues in the Department of Bioinformatics. Also, I thank the people in the Department of Microbiology for unblocking the door and even sharing the coffee for me. Special thanks to the secretaries Eva and Simone, and the system administrator Stephan.

It was a great opportunity for me to join the program GRK2157 from DFG during my PhD study. I'm grateful to all the PhD students and other colleagues in the program. I should also thank the FOKUS Life Science program, especially thanks to the FOKUS 2013

Contents

1

Introduction

1.1 *Chlamydia trachomatis (Ct)* infection

Chlamydia trachomatis (*Ct*) is an obligate human pathogen. It infects particularly host mucosal epithelial cells and implicates several significant human diseases, such as trachoma and sexually transmitted diseases (STDs). It was first found by Halberstaedter and Prowazek in 1907 as the form of the inclusion body (Schlosser, 2004). At that time, scientists tried to figure out the cause of blindness and made efforts to eliminate trachoma. In 1954, Feifan Tang together with his colleagues successfully cultured and isolated *Ct* from the yolk sac of chicken eggs (Schlosser, 2004). *Ct* was regarded as a virus for its smallness and uncultivation without living cells. Until the 1970s, *Ct* was recognized as a bacterium because it could synthesize its own macromolecules, such as DNA and RNA (Fan and Zhong, 2015, Schlosser, 2004). Nowadays, eliminating blinding trachoma has made great achievements, however, the STDs caused by *Ct* is in a grim situation and researches about *Ct* as a human pathogen are still ongoing (Fan and Zhong, 2015, Schlosser, 2004).

1.1.1   Intracellular biology and metabolic genetics of *Ct*

The Gram-negative *Ct* is one of the medically significant species in the genus *Chlamydiae*. *Chlamydiae* grows in an inclusion body and has a unique biphasic developmental cycle: the elementary body (EB, ~0.3μm), a nonreplicating infectious form which is metabolically static, and the reticulate body (RB, ~1μm), which is a metabolic active and non-infectious form, differentiated after entry into the host cell (Abdelrahman and Belland, 2005). Figure 1.1.1 shows the life cycle of *Chlamydiae* and the general hours for each developmental stage after infection. EBs enter in the host cell and initially grow within inclusion, an intracellular vacuole, by host intracellular membrane in 6 hpi. They lose the infectivity and differentiate to RBs in 12 hpi. during 12-24 hpi, RBs start binary fusion and accumulate the maximum number of replicates. Then RBs asynchronously redifferentiate back to EBs and will be released by host cell lysis after 48 hpi. The released infectious EB will continue to subsequent rounds of infection. (Käding et al., 2014, Shaw et al., 2000, Stephens et al., 1998, Stephens, 1999)

Compared to other bacteria, *Ct* has a substantially reduced genome. The genome of the reference strain, *Chlamydia trachomatis D/UW-3/CX* (Genbank session: AE001273.1) (Stephens et al., 1998), is only 1.04Mb, with 935 genes, encoding 6 rRNAs, 37 tRNAs and 887 proteins. Comparative genomic analysis reveals about 67% of proteins in the genome of *Chlamydia* spp. are conserved (Tan and Bavoil, 2012). How *Chlamydia* becomes such successful intracellular pathogen with so small genome is a worthwhile discussion (Tan and Bavoil, 2012).

The lack of some key metabolic enzymes leads to the loss of some special synthetic abilities. However, it presented many unexpected metabolic features. For example, the absence of gene encoding ribose-phosphate diphosphokinase reveals no phosphoribosyl pyrophosphate (PRPP) converts from ribose 5-phosphate (Stephens et al., 1998), thereby it is theoretically impossible to produce further purine and pyrimidine metabolism which needs the import of PRPP from pentose phosphate pathway. However, it encodes lots of enzymes involves in nucleic acid metabolism. Also, it is assumed that no genes encoding citrate synthase, aconitate hydratase and isocitrate dehydrogenase were identified in the genome, which makes the tricarboxylic acid cycle (TCA) incomplete. However, *Ct* seems not only import ATP from their host cell, but also keep enzymes in limited ATP biosynthesis pathways (Stephens et al., 1998). Many genes in *Ct*'s genome exhibit the character of horizontal gene transfer from both bacterial and eukaryotic ancestors (Horn et al., 2004, Stephens et al., 1998). It will be helpful to understand the *Ct*'s atypical metabolic feature.
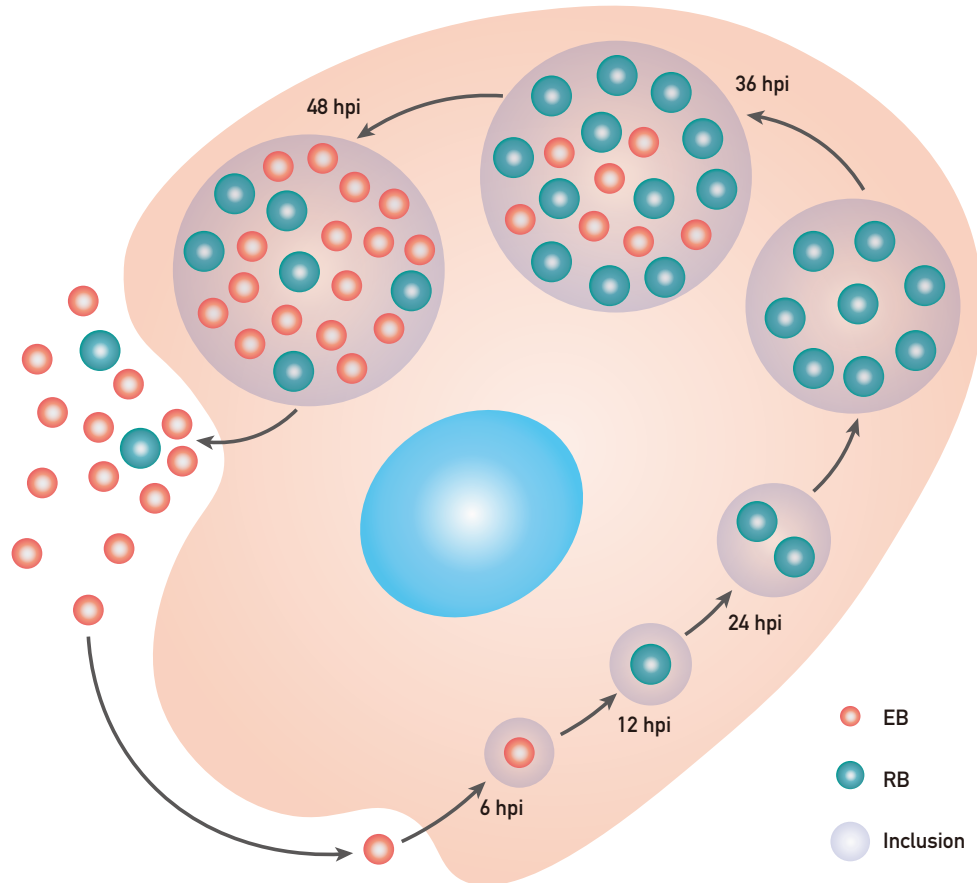
Figure 1.1.1: The developmental cycle of *Chlamydia trachomatis*. *Chlamydia trachomatis*(*Ct*) is an obligate intracellular human pathogen which has a biphasic developmental cycle. The infectious form EBs enter the host cell by an inclusion. It transforms into to RBs in the inclusion soon. After about 12 hpi, RBs start fission and growth. During the time, RBs will replicate several times and then differentiate back to EB. Roughly in 48 hpi (depends on cell types), the EBs come out from the host because of the cell lysis. EB: Elementary body; RB: Reticulate body; hpi:hour post infection. (source: own figure)

1.1.2   Pathogenesis and clinic significance of *Ct*

Based on the major outer membrane protein (MOMP) differentiation and whole genome sequence comparisons, *Ct* comprises three types of human serovars: 1) the trachoma biovar, serovar A, B, Ba and C, basically infects eyes and leads to non-congenital blindness; 2) the genital tract biovar, serovar D-K, is the most prevalent bacterium inducing genital infection; 3) the lymphogranuloma venereum (LGV), serovar L1, L2 and L3, causes invasive urogenital infections (Elwell et al., 2016, Seth-Smith and Thomson, 2013). It is regarded as a general rule for classification of strains of *Ct*, however, unusual features presented in some serovars improve difficulties for understanding and studying. The trachoma serovars B and Ba also lead to genital disease, and both genital and LGV serovars are capable infecting conjunctiva of the eye (Batteiger et al., 2010, Tan and Bavoil, 2012).

Although the World Health Organization (WHO) are trying to eliminate blinding trachoma by 2020 (Organization et al., 2012), STDs caused by *Ct* becomes more and more prevalent worldwide. The most common sexually transmitted infections (STIs) globally are chlamydia, gonorrhoea, syphilis and trichomoniasis. In 2016, it is estimated 357 million new infections were reported with one of these STIs, including 131 million chlamydia induced by *Ct* (WHO, 2016). Centres for Disease Control and Prevention (CDC) also claimed that chlamydia is the most prevalent disease in the United States and contains the largest proportion of STDs since 1994 (Braxton et al., 2018). *Ct* infection is usually asymptomatic and long-lasting but also cause symptomatic disease, which makes the disease difficult to be diagnosed and treated. *Ct* infection also induces pelvic inflammatory disease, which results in the ectopic pregnancy, tubal factor infertility and chronic pelvic pain (Braxton et al., 2018). Besides genitourinary cases, research also showed the association between *Ct* infection and reactive arthritis (Keat et al., 1987). In addition, coinfections are frequently found in *Ct* with many STI pathogens, such as *Neisseria gonorrhoeae (GC)*, *Trichomonas vaginalis*, *Human immunodeficiency virus* (HIV), *Herpes simplex virus* type 2 (HSV-2) and *Human papillomavirus* (HPV) (Alberts et al., 2013, Creighton et al., 2003, Kahn et al., 2005, Kilmarx et al., 2001, Tamim et al., 2002). Tamim et al. (2002) suggested *Ct* as a cofactor contributed to HPV induced cervical cancer together with HSV. Kimani et al. (1996) also reported that women got more serious chlamydia disease when infected with HIV. *Ct*'s coinfection is regarded as a risk factor for cervical intraepithelial neoplasia and even invasive cervical cancer (Anttila et al., 2001, Koskela et al., 2000). Antibiotic therapy is used to deal with chlamydia, however, it is risky to conclude the bacterium could be eliminated adequately and efficiently by drugs and vaccines for *Ct*'s persistence and relapse of infection (Elwell et al., 2016, Tan and Bavoil, 2012).

1.1.3 *Ct* coinfection with *Neisseria gonorrhoeae (GC)*

In the human urogenital tract, vaginal microbiota presents a diverse and dynamic population to keep the homeostatic environment with the host. However, coinfections among these vaginal bacterial communities had not been well understood yet (Ravel et al., 2011, Tan and Bavoil, 2012). *Ct* is found to coinfect with many STI pathogens as mentioned above. One of the most frequent coinfections *Ct* cooperated with is *Neisseria gonorrhoeae* (*GC*) infection.

GC, also known as gonococcus, is a Gram-negative bacteria which recognizes human as the unique natural hosts like *Ct*. *GC* is responsible for another widespread STD, gonorrhoea, one of the oldest human diseases. Strains of *GC* present multiple antibiotic resistance causing gonorrhoea widespread and untreatable (Quillin and Seifert, 2018). The infection

could be both symptomatic and asymptomatic (Quillin and Seifert, 2018, WHO, 2016). Generally, infections in males with urithritis are symptomatic, but female genital infections are usually asymptomatic (Detels et al., 2011, Quillin and Seifert, 2018). Sex specificity drives strikingly different molecular mechanisms of *GC* infection (Edwards and Apicella, 2004).

Together with *Ct*, *GC* reveals significant coinfection in both men and women. Kahn et al. (2005) reported 54% females and 51% males with gonorrhoea got infected by *Ct* in the USA according to the selected statistics of US Juvenile detention centres from 1997 to 2002. Detels et al. (2011) also described a high prevalence and incidence of infections by asymptomatic *Ct* and *GC* coinfection in selected populations among China, India, Peru, Russia and Zimbabwe. A large proportion of *GC-Ct* coinfected patients was also reported by Nudel et al. (2018) recently.

1.1.4   Challenges to study *Ct* induced host response during coinfection

Two main factors bring the obstacles in *Ct* study in the lab: the unculturable outside of the human host cell, and limitations in transformation for genetic manipulation (Tan and Bavoil, 2012). Because of that, comparisons of different *Ct* serovars in diverse host cell types become complicated. Besides, the EB-RB-EB differentiation during its life cycle is asynchronous because not all EBs or RBs convert to the other form at the same time. Thus, lacking efficient and precise methods to study the subpopulations makes the study of conversion more confusing (Tan and Bavoil, 2012).

Additionally, the existence of coinfection is not wise to be ignored but still tough to study. Like the interactions between *GC* and *Ct*, sexually transmitted coinfection occurs frequently among different pathogens. However, genetic patterns and molecular mechanisms of these concomitant infections are still less known due to the system complexity and difficulties for understanding (Cox, 2001). Individual *Chlamydia* is able to access the host cell by various mechanisms (Moulder, 1991, Stephens, 1999, Tan and Bavoil, 2012). Other STI pathogens may also utilize not only one mechanism for adhesion, invasion and evasion for the host cell, and it is widely believed that there are much more mechanisms for adaptation to keep homeostasis. Genes are likely to coregulate by pathogen-pathogen interaction and host-pathogen interaction for intracellular infection (Tan and Bavoil, 2012).

Nevertheless, the understanding of natural coinfections could not only rely on the researches applied in human cell lines. Generating 3D tissue utilized for coinfection study is a practical but challenging way for the consistency of diverse cell types in natural tissue. Cells within or without natural tissue may perform different immune response. It is important to offer

the natural host microenvironment to study pathogenic strategies and mechanisms for intracellular infection in order to promote affordable drug and vaccine design.

1.2 Sequencing and bioinformatics

Since James Watson and Francis Crick published the DNA double-helix structure in 1953 (Watson and Crick, 1953) based on Rosalind Franklin's crystallographic structure of DNA, the door of understanding molecular biology in the era of the genome was unblocked. Sequencing the primary structure for genetic molecules, such as nucleotide order of DNA and RNA, and amino acid order of protein, became the key to deeply understand genetic code. Later, the central dogma of molecular biology was stated by Crick, which well described the flow of genetic information: DNAs make RNAs, and RNAs then make proteins (Crick, 1958). More than half a century, scientists made great efforts in looking for methods and technologies in determining accurate genome information.

1.2.1 Sanger sequencing

Before introducing the next-generation sequencing, it is undisputable to talk about the "first-generation sequencing", which normally refer to Sanger sequencing and Maxam-Gilbert sequencing technologies. Frederick Sanger started his sequencing work with protein and initially performed amino acid order of insulin with his coworkers in 1955. After that, he explored techniques for sequencing RNA. In 1975, Sanger together with Alan Coulson published chain-termination sequencing which made a breakthrough of rapid determination of DNA sequence (Sanger et al., 1977). Sanger sequencing became the most common technology in DNA sequencing for almost 40 years (Heather and Chain, 2016). On the other side, Allan Maxam and Walter Gilbert developed the chemical sequencing method by partially breaking a terminally labelled DNA molecular at site adjacent of each repetition of a nucleobase in 1977 (Maxam and Gilbert, 1977). Although Maxam-Gilbert sequencing was very popular for a long time, this method was gradually replaced by the improved chain-termination method for reducing toxicity from chemicals and radioisotopes (Schuster, 2007).

Sanger sequencing requires a single-stranded DNA template, deoxynucleotide triphosphates (dNTPs, including dATP, dGTP, dCTP, dTTP), dideoxynucleotide triphosphates (ddNTPs, including ddATP, ddGTP, ddCTP, ddTTP), a specific DNA primer and DNA polymerase. The ddNTPs result in the termination of the extension of DNA for the lack of 3'-OH group and the incapacity of forming a phosphodiester bond between two nucleotides.

ddNTPs could be radionucleotide or labelled by fluorescence for automatic detection by sequencing machines. The process is similar to the DNA replication, which amplifying DNA by polymerase chain reaction (PCR). During PCR, DNAs are heat denatured from double-strand to single-strand. Primed DNA strands are added to four ddNTP vessels together with DNA polymerase and free nucleotides for replication. Then, DNA is separated by polyacrylamide gel electrophoresis. Each band shown in the gel represents a DNA fragment terminated by the ddNTP and is read from the gel bottom to the top together with the fluorescence signal (Figure 1.2.1).
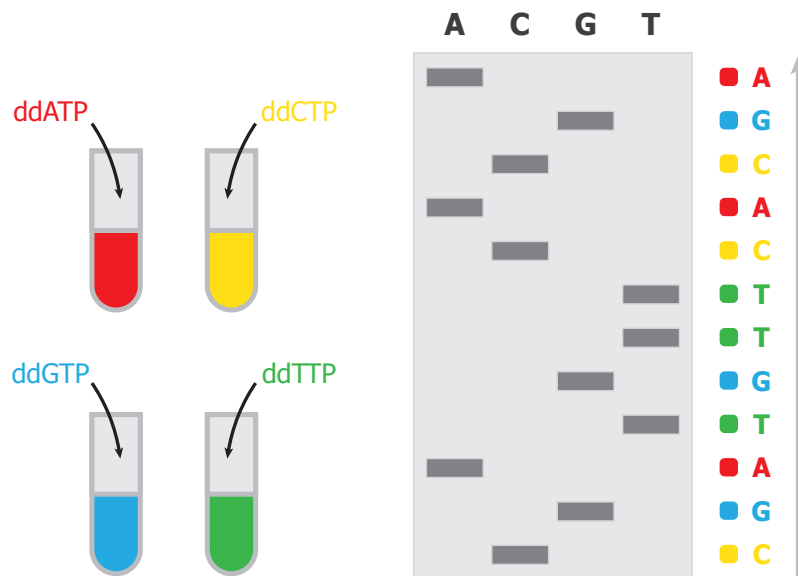
Figure 1.2.1: Method of Sanger sequencing. ddNTPs are used for chain termination by added in four prepared DNA reaction vessel. The new DNA fragments are labelled with fluorescence. Chain terminated reactions conduce DNA fragments in different length. The four vessels are then put on gel electrophoresis, results in different band patterns to show the DNA fragments in different length. The sequence should be read from the bottom to the top of the plate. (source: own figure)

Sanger sequencing was successfully used in the Human Genome Project from 1984 to 2003. With the improvement of automation and efficiency during the past four decades, Sanger sequencing achieved 1000bp read length and the lowest polymerase errors with the precision up to 99.99%. However, it is low throughput with the high cost. Thus, alternative approaches to rapid DNA sequencing sprout afterwards (Shendure and Ji, 2008).

1.2.2 Next-generation sequencing

Next-generation sequencing represents the high throughput, high sensitivity and massively parallel sequencing, green-lighting sequencing of the entire genome at one time. Typically, 454 pyrosequencing and Illumina (Solexa) sequencing are the most common sequencing

methods. There are also many other technologies such as Ion Torrent, SOLiD, PacBio sequencing.

The method of pyrosequencing was developed in 2005 by 454 Life Sciences (Margulies et al., 2005), a biotechnology company founded by Jonathan Rothberg, acquired by Roche Diagnostics in 2007 and shut down business in 2016. As the first commercial platform for NGS, Roche 454 system is capable of sequencing 350-600 Mb. In 454 sequencing (Figure 1.2.2a), DNA is fragmented, denatured, and then prepared to ligate with adaptors. Prepared fragments are assigned to DNA capture beads. Fragments are amplified by emulsion PCR within a water-in-oil drop. One bead contains one fragment with the replicates. Each DNA-capture bead is placed into a special well which only allows the single bead to fit in for sequencing by the luminescent signal.

Illumina sequencing was developed by Shankar Balasubramanian and David Klenerman who founded Solexa, a company formed in 1998 and acquired by Illumina in 2007 (Illumina). This technique was based on reversible terminated chemistry method created by Bruno Canard and Simon Sarfati in 1994 (Canard and Sarfati, 1994a,b). The process of Illumina sequencing could be divided into these steps: 1) NGS library preparation; 2) Cluster amplification; 3) Sequencing; 4) Alignment and data analysis (Illumina, 2015) (Figure 1.2.2b). In this method, purified DNA is prepared by random fragmentation with adapter ligation and PCR amplification. For cluster amplification, single-strand fragments are randomly loaded and captured onto a flow cell surface. Instead of emulsion PCR used in 454 sequencing, bridge amplification method serves for Illumina sequencing. Unlabelled nucleotide and DNA polymerase are added for amplification, which builds "bridges" for double-strand DNAs. These double-strand fragments are denatured by heat and leave single-strand fragments into identical DNA clusters. Fluorescent labelled terminators together with primers and DNA polymerase are added to the flow cell for the first sequencing cycle, and emitted fluorescence from each cluster is exported into the digital image. The sequencing cycle is repeated "n" times for "n" bases in reading length (Illumina, 2015). The sequencing could detect either short reads or long reads. For short reads, it could also be the single-end or paired-end according to technical choice.

Ion Torrent sequencing has a faster workflow and economical device with good portability. Fluorescence labelling and camera scanning are not necessary for this method. However its throughput is not as high as 454 or Illumina sequencing, and it has difficulties to identify repetitive sequences (Quail et al., 2012). SOLiD is the abbreviation of "Sequencing by Oligonucleotide Ligation and Detection", an NGS method developed by Life Technologies Corporation. Instead of using the method of DNA synthesis, it relies on DNA ligation by ligase. It has a lower cost but less read length and is much slower compared to other

methods. Also, SOLiD is reported to be difficult in sequencing palindromic region (Huang et al., 2012). PacBio, short for the company Pacific Biosciences, focuses on single molecule real-time sequencing (SMRT). This method is rapid and able to sequence long read, but is less accurate together with higher cost. PacBio had been acquired by Illumina on November 2018 (Illumina, 2018).

(a) 454   **(b)** Illumina
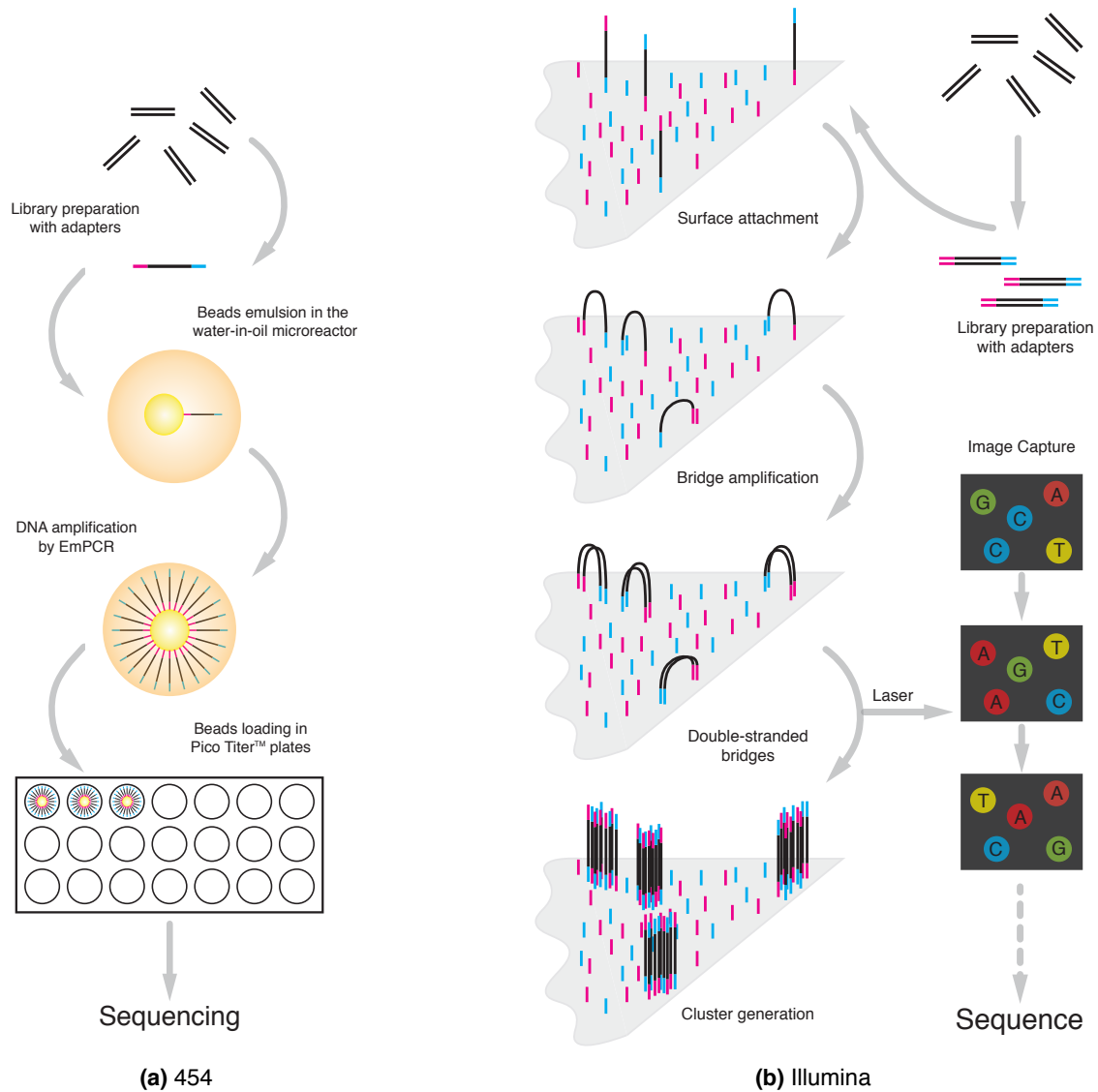
Figure 1.2.2: Schematic diagram of 454 and Illumina sequencing (source: own figure). (a) DNA fragments are amplified with adapters in the pool and captured by the beads. The beads are placed in micro-reactors for "water-in-oil" emulsion PCR (EmPCR). These beads are loaded onto Pico Titer™ plates and subsequently pyrosequenced. (b) DNA fragments are prepared with adapters and then attached to the flow cell surface. Fragments replicate into clusters via bridge amplification. During the sequencing cycle, the data is under laser excitation and output in the digital image, which will be extracted into the text file afterwards.

1.2.3   RNA sequencing

DNA sequencing for genome analysis is no longer to satisfy the researchers' curiosity with the development of scientific technologies. Scientists would like not only to know what and how many genes are encoded, but also to look for how the gene expressed in transcript level. RNA sequencing (RNA-Seq) represents an evolved approach for transcriptome profiling by using NGS methods to reveal the qualitative and quantitative features of the transcriptome (Ozsolak and Milos, 2011, Wang et al., 2009). It continues to use standard cDNA library preparation in genome sequencing but the sample is RNA before converted to cDNA (Illumina, 2015). RNA-Seq can be served for whole transcriptome, targeted RNAs, small RNA or noncoding RNAs.

High-throughput RNA-Seq becomes one of the most popular techniques for the whole transcriptome profiling. RNA-Seq has a broad range for novel transcripts detection based on known reference genomes and also possible for unknown genomes or unannotated species. It is highly sensitive for detecting genes expressed even with low abundance. Other traditional techniques are also designed for gene expression quantification. Quantitative real-time PCR (RT-qPCR) is a quantitative method of DNA amplification by PCR. The exact amounts of amplified DNA are monitored by fluorescent probes in real time. RT-qPCR is widely used in measuring the expression of small sets of sequence-known genes, with the features of extreme sensitivity, precision celerity and reproducibility. However, it is not suitable for whole transcriptome determination and sample contamination is an undeniable problem which highly influences the result. Microarray (or DNA chips) is another approach to examine gene expression based on nucleic acid hybridization. It simultaneously measures large amounts of genes and even for the genome designed on the probes. Because of less capable of detecting fusion genes and varied transcripts based on alternative splicing, the microarray was rapidly replaced by cost-reducing NGS method 20 years ago (Ledford, 2008). Today, places are still available for microarray, because microarray is affordable and could offer more stable and accurate results on differentially expressed genes in a reference genome when compared to RNA-Seq.

As mentioned above, RNA-Seq makes it possible for sequencing organisms with unknown genomes. However, the sequencing quality is diverse by different sequencing machines, library composition, preparation methods, and sequencing methods. Afterwards, transcriptional profiling and differential expression analysis rely on approaches by bioinformatics and statistics.

1.2.4 Bioinformatics and statistics in NGS analysis

Technical development drives many omics studies sprung out. Sequencing-based technologies accelerate the necessary of genomic, transcriptomic, epigenomic and proteomic analysis (Shapiro et al., 2013). Numerous data amounts with diverse data formats cast its longing to methods and tools in bioinformatics analysis. Varieties of applications are available or at lease under developing in bioinformatics analysis, especially in NGS analysis. However, no single pipeline is able to be used for all cases (Conesa et al., 2016). The downstream data analysis of NGS is facing opportunities and also challenges.

RNA-Seq data analysis, although supported by broad applications, is still complicated due to different workflows used for different purpose and dataset for the sequenced reads. Sahraeian et al. (2017) examined 39 analyzing tools combined to about 120 pipelines in RNA-Seq analysis, and compared the different workflows based on the sensitivity, precision, speed, and computing performance and cost. This study revealed the choice of the tools and pipelines used for analysis intensively affects the precision of the results. Features of tools should be balanced to take into consideration according to the purpose of analysis. No independent "golden standard" is achievable to support a comprehensive assessment (Su et al., 2014).

A basic but challenging question for RNA-Seq data analysis is the data normalization for different sequencing depth and coverage in each sample. Before the mapping step, a comparable dataset should be normalized by statistics. Dillies et al. (2013) summarized 7 normalization methods for reads counting, including: total count (TC), which using mapping reads divide gene counts; upper quartile (UQ), whose gene counts are divided by upper quartile of counts different from 0 in the normalization factor's computation (Bullard et al., 2010), instead of using total mapped reads; median (Med), which substitute upper quartile of counts in UQ method to median counts; quantile (Q), using distribution matching methods for gene counts and initially performed in analyzing microarray data (Bolstad et al., 2003, Yang and Thorne, 2003); reads per kilobase per million mapped reads (RPKM) (Mortazavi et al., 2008), trimmed mean of m-values (TMM) (Robinson and Oshlack, 2010), and relative log expression (RLE) (Anders and Huber, 2010, Anders et al., 2013, Love et al., 2014) which will be described in details as follows.

RPKM, as its name described, is a method of counting reads C_i per kilobase of a gene $L_i/10^3$ per million mapped reads $N/10^6$:

$$RPKM_i = \frac{C_i}{\left(\frac{L_i}{10^3}\right)\left(\frac{N}{10^6}\right)} = \frac{C_i * 10^9}{L_i * N} \tag{1.1}$$

RPKM is designed for single-end RNA-Seq, while FPKM is made for pair-end RNA-Seq,

which stands for fragments instead of counting reads. The nommalization method of RPKM and FPKM is the same. However, both of them introduce bias in differential expression analysis by correcting gene lengh differences (Dillies et al., 2013, Oshlack and Wakefield, 2009), which cause lower RPKM in longer genes among genes express the same number of reads. Later, transcripts per kilobase million (TPM) become an alternative method for transcripts abundance description to eliminate RPKM's inconsistency (Wagner et al., 2012):

$$TPM_i = \frac{C_i * 10^6}{L_i * \sum_j \frac{C_j}{L_k}} \tag{1.2}$$

RPKM, FPKM and TPM are all within-sample normalization, and TPM is comparatively more suggested.

TMM uses the weighted mean of ratios to estimate scale factors between samples (Robinson and Oshlack, 2010). It takes one lane as a reference sample, and consider other lanes as the test sample. A basic hypothesis of TMM is that most of the genes are not the differential expressed, thus theoretically the scaling factor of TMM should be adjusted around 1 (Dillies et al., 2013, Robinson and Oshlack, 2010). It normalizes the read counts in library size (Maza, 2016), and has been implemented in the R package, edgR (Robinson et al., 2010).

RLE was implemented in R package DESeq, which is also based on the hypothesis of no differentiation in most genes and take reference samples, similar to TMM (Anders and Huber, 2010). It takes the size factor to estimate the gene counts by the median count ratio among all samples. Unlike TMM, RLE does not directly work on raw reads in each sample, but performs a counting matrix based on a geometric mean of all sample values (Maza, 2016).

These normalization methods discussed above are primarily used for counting reads in order to describe the level of gene expression for a specific group with multiple samples. For differential expression(DE) analysis among groups under different condition, other sophisticated methods for normalization and comparison are obligatory. The statistical significance of DE is an important issue to introduce. The statistical test is used to determine the probability value (p-value) in order to describe the differential expressions. p-value, ranging from 0 to 1, is basically used to describe how possible the difference between two compared groups happens (Wasserstein, 2016). Besides, a value of false discovery rate (FDR) or adjusted p (*padj*) is also involved. FDR is used to test significant differences which do not exist in reality but also presented in statistical significance. For example, no difference should present in reality when the sample is in a self-self comparison, but the difference may display in the test, which regarded as a false positive (Su et al., 2014). FDR refers to an adjusted p-value to value the chance of p-value shown may be wrong. Most of the time, the threshold of a statistic significance is both p-value and *padj* should below

0.05. However, the threshold of p-value is arbitrary, and the statistical significance is not equal to biological significance.

1.3 Modelling in Systems biology

Systems biology is using interdisciplinary approaches to study complex biological systems. It holistically integrates methods of mathematics, computer science to comprehensively analyze the complex biological processes in different scales (organism, tissue, cell, subcell and molecule), and is widely used in various biological fields based on the ultrafast development of omic techniques and explosively increased biological data (Tavassoly et al., 2018). The main aspect of systems biology is computational modelling to understand how the cell works by the mechanisms of molecular interaction and cooperation. On the basis of information collecting, mathematical models are constructed to figure out the significant properties of the target system. By introducing experimental data sets, the model simulation could be curated and quantitatively calculated through mathematical algorithms to look for solution space in order to obtain accurate predictions for the biological questions. In the post-genomic era, network analysis of systems biology increasingly aims more at molecular networks to understand how genes, gene products and metabolites involved mandate cellular behaviour.

1.3.1 Introduction of graph theory

Using the method of the network to describe biological systems are basically derived from graph theory. Graph theory was firstly described by the mathematician Leonhard Euler in 1736, when he published his paper about the solution of the problem of *Seven Bridges of Königsberg* (Euler, 1736). The mathematical definition is: A **_graph_** is an ordered pair of disjoint sets (V, E) such that E is a subset of the set V of unordered pairs of V (Bollobás, 2013). The **_graph_** can be represented by a matrix G, $G = (V, E)$. V is a vertex, the sets of nodes in the graph, and E means the sets of edges between the nodes. The graph can be used to edges and nodes to describe relationships among different components, and features of these relationships are investigated through mathematical disciplines by graph theory. Graph theory is a favour for big data analysis especially to look for communities in networks. It is widely used in diverse fields, such as social networks, supply chain, telecommunications, programming, and also biological networks.

1.3.2 Types of biological networks

According to different components of nodes and the connectivities of edges in the network, the biological networks in the molecular level have different types. The most common types are gene/transcriptional regulatory network, genetic interaction network, protein-protein interaction network, metabolic network and signalling network.

The regulatory network is made up by molecular regulators (usually proteins, sometimes also regulatory RNAs) and represents the regulatory interaction between regulators with their potential targets to show how gene expression is controlled (Emmert-Streib et al., 2014). It contains gene regulatory network (GRN) and transcriptional regulatory network (TRN) according to the target layer of DNA and RNA respectively. In the network, nodes are genes and transcriptional factors, and edges are usually directed and display the relationship between the nodes. Genetic interaction network (GIN) presents logical interactions between different genes and is used for comparing the significant difference among mutated genes according to effects shown in phenotype. Normally it describes the genes as nodes interact functionally rather than have physical interactions like GRN, and normally the edges are undirected. Protein-protein interaction (PPI) network is also linked by undirected edges but the nodes are all proteins. And the protein nodes connected have physical interaction. The metabolic network focuses on biochemical reactions of metabolites. The nodes are usually metabolites and enzymes involved, and the edges are not shown in the way of interaction but transformations in between. The transformations are displayed by directed edges and could be reversible depended on exact biochemical reactions. Signalling network studies the communication process of cellular activities. The signal nodes included could be nucleic acids, proteins and metabolites, and the signalling pathways are linked by directed edges.

1.3.3 Metabolic modelling and flux analysis *in silico*

To study metabolism in organisms, metabolic flux analysis (MFA) is a well-used approach. MFA initially uses experimental fluxomics technique to measure the production of metabolites in a biological system, in order to analyze the variation of metabolites quantitatively and demonstrate metabolism in the system. Typically, ^{13}C (or other isotopes) isotope labelling, mass spectrometry and NMR are common techniques in use.

Methods *in silico* provide efficient support for MFA to comprehensively understand molecular mechanisms (Francke et al., 2005). Metabolic network modelling is one of the most widely used approaches. It reconstructs the metabolic model (normally in genome-scale)

and breaks the whole metabolic systems into several metabolic pathways with enzymes and reactions. These metabolic pathways are compiled into a mathematical model shown into a stoichiometric matrix, where rows correspond to metabolites of the reactions, and columns show reactions. The simulation based on the stoichiometric matrix is called stoichiometric flux analysis. Stoichiometric flux analysis assumes the metabolic system in the stoichiometric steady state and quantitatively demonstrates the relationship between substrates in the reaction. This is a constraint-based modelling and primarily used for the single sample. Many approaches are used to achieve the flux analysis stoichiometrically, such as elementary mode analysis, extreme pathways and flux balance analysis.

As is defined by Orth et al. (2010), flux balance analysis (FBA) is a mathematical method aim on analyzing the flow of metabolites in genome-scale metabolic network simulation. Based on linear programming, FBA predicts phenotype responses by using metabolic models (Cuevas et al., 2016). FBA quantitatively calculates metabolic flows according to the network and predicts the production of the metabolites or activity of the enzymes. Based on constraints, FBA can optimize a phenotype in the solution space and predict all available metabolic pathways for a given set of enzymes. According to the optimized results of biomass production, or activities in each reaction, it can summarize the metabolic consumption or pathway activities in the system (Orth et al., 2010).

1.4 Aim of the work

The overall aim of the work is studying the metabolic adaptation of *Ct* especially the differentiated forms EBs and RBs during infection and also the coinfection analysis of the response from both the host and coinfected pathogen in *GC* infected patients by meta-transcriptome analysis.

As described in Chapter 1.1.4, the special characteristics of *Ct* cause many obstacles in using various microbiological methods and technologies in studying *Ct*. It is difficult to culture outside of the human cell and no easy for genetic modification. Although RNA-Seq is a mature technology, it is still very challenging in profiling the variation of *Ct*'s transcriptome during the biphasic developmental cycle. Microarray (Belland et al., 2003) and deep sequencing (Albrecht et al., 2009) techniques have been used for detect *Ct* transcriptome and the enriched genes in EB and RB. Also, *Ct* proteomics (Østergaard et al., 2016, Saka et al., 2011, Skipp et al., 2005) were for understanding the metabolic and pathogenic feature of *Ct* (even the difference between EB and RB) in protein level, however it can not clearly show details of the activities of pathways during EB-RB-EB transformation. Besides, lack of systematical methods to clearly show the host-pathogen response in *Ct* in-

fection. A technique called dual RNA-Seq (Westermann et al., 2012) is a well-established transcriptome quantification method for study both host and pathogen during infection. However, it is rigorous for *Ct*. Because fluorescence-activated cell sorter (FACS) is necessary for dual RNA-Seq to filtered out uninfected cells, but for *Ct* is not as easy as other microbes to be genetically added green fluorescent proteins (GFPs). In addition, current studies of *Ct* are mainly focused in the lab environment, but bacteria rarely live alone as isolation. Many reports (discussed in Chapter 1.1.2 and 1.1.3) announced *Ct*'s coinfection with other bacteria or viruses, but the comprehensive studies about the mechanisms of pathogens in natural infection and how they keep homeostasis with other organisms in the host niche are still ongoing.

This study is to use methods in bioinformatics and systems biology to study problems about *Ct* discussed above. The quantitative genome-scale metabolic modelling and metabolic flux analysis are tried to understand metabolic differences of pathways between EB and RB during infection according to proteomic data. Additionally, differential expression analysis directly based on *GC* infected STD patients is used to reveal the responses of the host and *GC* respectively in *Ct* positive group. This patient transcriptome project has collaborated with Genco lab, School of Medicine, Tufts University, USA.

2

Materials and Methods

2.1 Materials

2.1.1 Platforms, systems and programming languages

Table 2.1.1: Platforms, device and programming languages

High performance computing cluster (Julia), University of Würzburg, Germany

Tufts high performance computer cluster, Tufts University, MA, US

Linux (version 4.4.92-31, SUSE)

Macos (High Sierra and Mojave, x86_64-apple-darwin18.2.0)

Microsoft Windows 10

StepOnePlus device (Applied Biosystems)

Python (version 2.7.13 and 3.6.1)

R (version 3.5.2)

2.1.2 Databases and Software

Many database and software used for analysis or data arrangement, including: Ballgown (Frazee et al. (2015), https://www.bioconductor.org/packages/release/bioc/html/ballgown.html); Bedtools (Quinlan and Hall (2010), https://bedtools.readthedocs.io/en/latest); Bioconductor (Gentleman et al. (2004), https://www.bioconductor.org); Bowtie2 (Langmead and Salzberg (2012), http://bowtie-bio.sourceforge.net/bowtie2/index.shtml); BRENDA (Schomburg et al. (2002), https://www.brenda-enzymes.org); ClueGO (Bindea et al. (2009), http://apps.cytoscape.org/apps/cluego); Cufflinks (Trapnell et al. (2012), http://cole-trapnell-lab.github.io/cufflinks); Cytoscape (Shannon et al. (2003), https://cytoscape.org); DESeq2 (Love et al. (2014), https://bioconductor.org/packages/release/bioc/html/DESeq2.html); FastQC (Andrews et al. (2010), https://www.bioinformatics.babraham.ac.uk/projects/fastqc); GENCODE (Harrow et al. (2012), https://www.gencodegenes.org); gffread (Pertea and Rozenberg, https://github.com/gpertea/gffread); HISAT2 (Kim et al. (2015), https://ccb.jhu.edu/software/hisat2/index.shtml); KEGG (Kanehisa and Goto (2000), https://www.genome.jp/kegg); NCBI and Genbank (Benson et al. (2012), https://www.ncbi.nlm.nih.gov); One Codex (Minot et al. (2015), https://www.onecodex.com); Microsoft Office 365 (Excel, Word, (version 16.16.5), https://www.office.com); PANTHER (Mi et al. (2016), http://www.pantherdb.org); SAMtools (Cock et al. (2015), Li et al. (2009), http://www.htslib.org); StepOnePlus™software package (Applied Biosystems, https://www.thermofisher.com/de/en/home/brands/applied-biosystems.html); STRING (Szklarczyk et al. (2016), https://string-db.org); StringTie (Pertea et al. (2015, 2016), https://ccb.jhu.edu/software/stringtie); YANAsquare (Schwarz et al. (2007), https://www.biozentrum.uni-wuerzburg.de/bioinfo/computing/yanasquare); WikiPathways (Slenter et al. (2017), https://www.wikipathways.org/).

2.1.3 File types

FASTA format is a text-based format developed by Lipman and Pearson (1985) which used to describe the sequence of nucleotides or amino acids. It uses two lines per sequence. An identifier ">" is always at the beginning of the first line, and a summary description of the sequence follows afterwards. The subsequent line is the sequence. Here is an example:

> >chlamydia trachomatis unreal gene
> accgcagatcacaatattttcgctagatcacaata

FASTQ format is text-based and created by Wellcome Trust Sanger Institute (Cock et al., 2009). It is formed by four lines per sequence and added a quality value of sequencing in each nucleotide (or amino acid). The initial line starts with a "@" which is similar to

the ">" in Fasta format, with sequence description following. The second line records the sequence. The third line is "+". The fourth line is the quality value shown by letters in ASCII code, whose order presents sequencing quality from low to high. Here is an example:

> >chlamydia trachomatis unreal gene
> accgcagatcacaatattttcgctagatcacaata
> +
> _|OabcIJK-MN@PQRS!![00)))xxxII@II+I

SAM format is also text-based sequence alignment map format for storing the alignments between the reads and reference sequences (Li et al., 2009). The binary format is the **BAM** file, which stores the same data as in SAM. SAM format normally has 11 fields and separated by the tab. For details please check the description from SAMtools (Group, Li et al., 2009). Here is a short example of the SAM format:

> Queryname 4 ref 0 0 * *0 0 AAAGAGGTTGTAGT P-ccaacbabccb- YT:Z:UU

GFF file is a general feature format, and **GTF** file is general transfer format. Both of them are used for recording gene descriptions and other features of the gene structure. Here is an example of the GTF format:

> gene_id "gene111"; transcript_id "transcript222"; exon_number 1

SBML format is XML-based systems biology markup language format. It is used for representation and exchange of biochemical network (Hucka et al., 2003). It currently contains three levels.

2.1.4 Bacterial strains

Table 2.1.2: *Ct* strains

Species	Serovar	ATCC number
Chlamydia trachomatis	L2/434/Bu	ATCC®VR-902B™

2.1.5   Eukaryotic cell lines

Table 2.1.3: Human tissue culture cell lines

Name	Origin	ATCC number
HeLa229	Human cervix carcinoma	ATCC®CCL2.1™
HUVECs	Human umbilical Vein	ATCC®CRL-1730™

2.1.6   Primers

CT054, sucA, forward: 5'-tggagtttgaagacgctccc-3', reverse: tgacccgccaaggctgc;

CT055, sucB, forward: 5'-gcgatgcgcagattattgcc-3', reverse: ttcattaaatgtcgttaacattgctg;

CT121, araD, forward: 5'-gtccggagcagatcttattcat-3', reverse: tcaaaatgcacaatgatacgatcc;

CT183, pyrG, forward: 5'-aagtatacgtgaccgacgatg-3', reverse: ctgcgcacgattgaatgacat;

CT205, pfkA1, forward: 5'-ctctgttgcagaactatttccatg-3', reverse: ctgtgtagcaatcctaagatgac;

CT207, pfkA2, forward: 5'-cccagacctcactctttcaaaaa-3', reverse: ctcgaataagtccaagagggc;

CT238, fabD, forward: 5'-ctgggattagaaataaggtggaa-3', reverse: atacactaactcccctattacatta;

CT313, tal, forward: 5'-ctctgacttttgtttttagataagatc-3', reverse: ttttaactaataaacgtttcttatctcct;

CT332, pykF, forward: 5'-ggccggaccatcgctattc-3', reverse: gagagtaagacgatccccag;

CT376, mdhC, forward: 5'-atggtttctcaaacagtgagtgt-3', reverse: acctgcacacgctgtaataga;

CT378, pgi, forward: 5'-aggatgcttatctgaagagcgt-3', reverse: cttctatacgattcatgatggcg;

CT453, plsC, forward: 5'-gctaagtccggactgttttcta-3', reverse: aataacagggacattgcctttgata;

CT505, gapA, forward: 5'-ggagaaagaaagatccgtttctta-3', reverse: tcccataacaaacgtagggaca;

CT612, folA, forward: 5'-cattcgaaatcatcccatcattatg-3', reverse: tgtgacaaaacaagctttcagaag;

CT613, folP, forward: 5'-gggaaacgcagctatctcca-3', reverse: acgatacggacataaagaagcc;

CT614, folX, forward: 5'-gtctcagaacaagaacggcatt-3', reverse: gccaaagccttttctattttttcc;

CT714, gpdA, forward: 5'-agaatctttctttcacataccacat-3', reverse: cgatttcgcttaggagtaagc;

CT750, tktB, forward: 5'-ctaaagattctcgatgggtgata-3', reverse: tgtagtggcttctacaccatca;

CT775, tktB, forward: 5'-ggttaaagtttcacgtgcgac-3', reverse: aacgatctcttctttcccatttct;

CT807, plsB, forward: 5'-gatgaccttaagaatcctattattttc-3', reverse: tacatgagctgggggtctga;

CT811, plsX, forward: 5'-cctcgatggctctaggattg-3', reverse: tctcgtcaggattcacggac;

CT816, glmS, forward: 5'-ggcgtagtattttcatcagatac-3', reverse: acctcgtcttctcctaatccta;

CT821, sucC, forward: 5'-gagtattatcttgcgatcgtcat-3', reverse: atcggcaataggttgatccca;

CT822, sucD, forward: 5'-ccgtttgctgcagaagccat-3', reverse: atatacccaggcatgatgcca.

CT855, fumC, forward: 5'-atgcggcaagagaatgatagc-3', reverse: agcaacaatcatatccctgcg.

2.2 Methods

2.2.1 Metabolic modelling

2.2.1.1 Genome-scale metabolic model reconstruction

Genome-scale metabolic model (GEMM) is reconstructed for *Ct*'s genome. The procedure begins with the identification of enzyme-coding genes from the reference genome of strain *Chlamydia trachomatis D/UW-3/CX* (Stephens et al., 1998). Metabolites and reactions are collected from a variety of resources, including KEGG (Kanehisa and Goto, 2000), BRENDA (Placzek et al., 2016), NCBI. The model is established in YANAsquare (Schwarz et al., 2007) by input metabolites and the chemical reactions displayed as nodes and edges respectively in a graph. Metabolites are classified into internal nodes and external nodes. The internals is unique compounds involved in the system and the externals could be more feasible taken in (or out) of the system by calculation. Both KEGG reaction number (Begin with an "R" with five number followed, e.g. "R00001") and EC (Enzyme Commission) number (from EC 1 to EC 6 category, not included EC 7 which is just announced on August 2018 (Tipton, 2018)) if available. The model is saved in SBML format. In addition, gap filling for key pathways are based on physiological knowledge from previous study and published proteomics like the study by Saka et al. (2011) and Østergaard et al. (2016). If no evidence or data could support, the gap would not be filled concerning on *Ct*'s specificity.

2.2.1.2 Metabolic flux modelling and flux balance analysis

Extreme pathways (EPs) were firstly calculated in YANAsquare (Schwarz et al., 2007) by an internal convex basis algorithm (Kaleta et al., 2006). According to the methods described by Orth et al. (2010), the metabolic reactions are prepared in a stoichiometric matrix ($\boldsymbol{S}$) with the size of $\boldsymbol{m} \times \boldsymbol{n}$ ($\boldsymbol{S} = \boldsymbol{m} \times \boldsymbol{n}$). The rows and columns of the matrix respond to internal metabolites in the model ($\boldsymbol{m}$) and involved reactions or enzymes ($\boldsymbol{n}$), respectively. For each column, the entries are stoichiometrically correlated to involved metabolites. All possible phenotypes are involved in the non-negative linear combinations of the stoichiometric matrix. Fluxes calculated by the reactions are presented in vector $\boldsymbol{v}$. Solution of the model was based on the steady state of the network, which makes the equation:

$$\boldsymbol{S} \cdot \boldsymbol{v} = \boldsymbol{0} \tag{2.1}$$

In this case, it results in a solution space for the fluxes. After the steady state is calculated, enzyme activity collected from proteomics by Saka et al. (2011) and Østergaard et al. (2016) are calculated for the optimal pathway fluxes in a dynamic state according to different phenotype in different time points. These optimal pathway fluxes were calculated by using a square fitting. Least square fitting by gradient descent method was generated by YANAsquare to calculate the best-matched pathway fluxes including identification of enzymes not presented in proteomic data. The optimized solution can be found for balancing fluxes by calculating the actual flux intensities of the involved pathways. The simulation focuses on metabolic difference between EB and RB in 20 hpi, 40 hpi and 40 hpi post (cell lysis) respectively as described by Østergaard et al. (2016).

2.2.2   RNA-Seq and data analysis

2.2.2.1   Sample collection and RNA sequencing

Patient samples from were prepared with TruSeq RNA Preparation Kit from National Center for Sexually Transmitted Disease in Nanjing, China. RNA sequencing was produced by Illumina HiSeq 2500 instrument with High Output V3 chemistry in single read 100 formats by our collaborator, Genco lab, School of Medicine, Tufts University, Boston, USA. Details were described by Nudel et al. (2018)

2.2.2.2   Quality control of RNA-Seq raw data

The RNA sequencing raw data was single-end raw reads performed in FASTQ format. Reads quality was estimated and controlled by FastQC (version:0.11.7) (Andrews et al., 2010). FastQC summarized reads quality by the following parameters: basic statistics, per base sequence quality, per tile sequence quality, per sequence quality scores, per base sequence GC content, per base N content, sequence length distribution, sequence duplication levels, overrepresented sequences, adapter content and Kmer content.

2.2.2.3   Taxonomic identification

Online platform One Codex (Minot et al., 2015) was used to detect microbial abundance in each sample. Qualified reads were submitted to One Codex and got the taxonomic profiling in the default setting. Not only host and bacteria, but also fungi, virus and protists were shown.

2.2.2.4 Sequence alignment for qualified RNA-Seq reads

Sequence alignment was performed by HISAT2 (version:2.1.0) (Kim et al., 2015). Bowtie2 (version 2.3.3.1) and Bedtools (version:2.26.0) (Quinlan and Hall, 2010) were also used as a comparison. Reads were aligned to three reference genomes in FASTA format: *Homo sapiens* GRCh38.p11 (Consortium et al., 2001, McLaren et al., 2010, Zerbino et al., 2017), *Neisseria gonorrhoeae FA 1090* (Lewis et al.) and *Chlamydia trachomatis D/UW-3/CX* (Stephens et al., 1998). Sequence alignment map (SAM) files were output after genome alignment and then transferred and sorted to Binary format (BAM) files by SAMtools (version: 1.7) (Cock et al., 2015, Li et al., 2009).
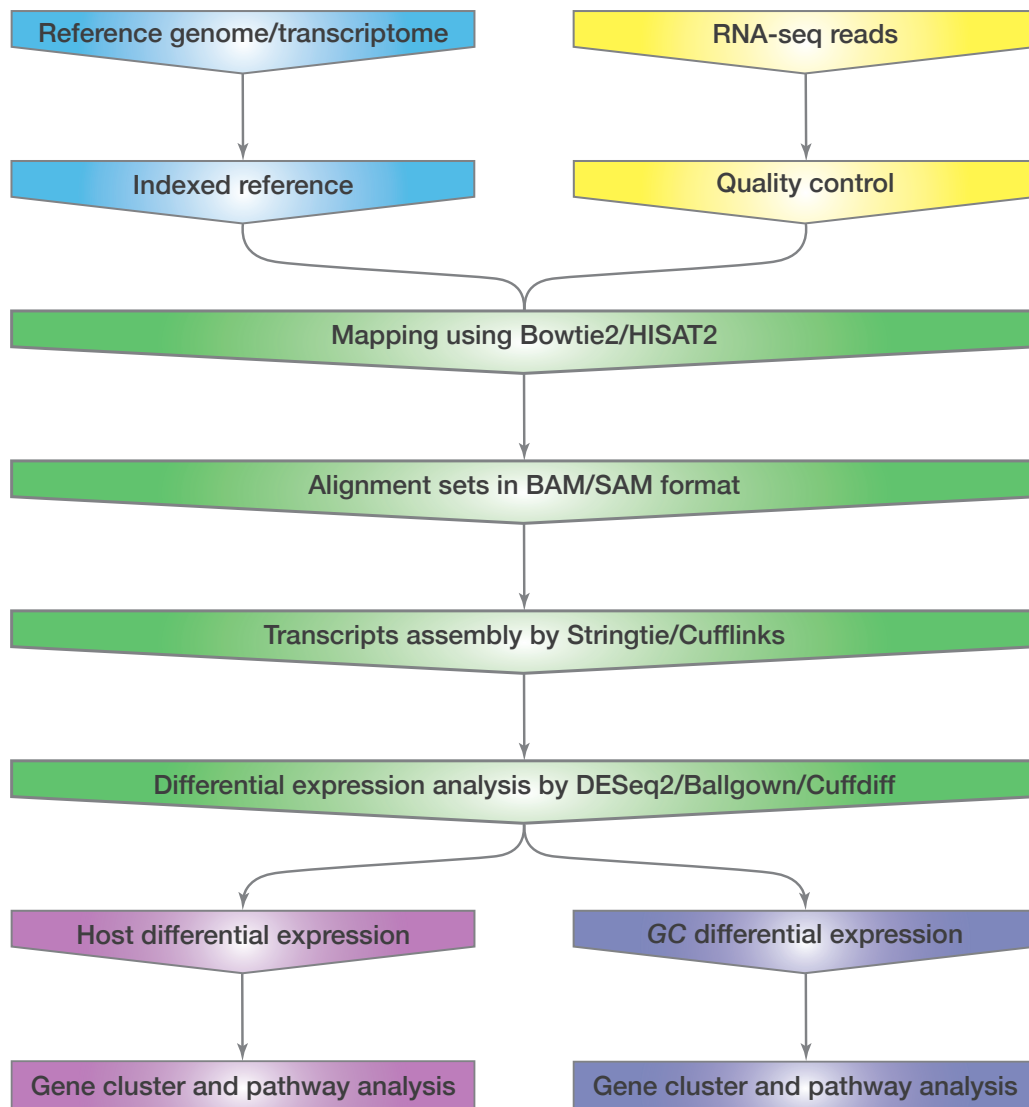
Figure 2.2.1: RNA-Seq data analysis pipeline for differential expression analysis (source: own figure).

2.2.2.5 Post-alignment analysis

Gene transfer format (GTF) file is necessary for annotation and expression extraction. GTF file of the human sample was prepared from GENCODE (Harrow et al., 2012) in a comprehensive gene annotation version of *Homo sapiens* GRCh38.p11 (*HS*). GTF files of *GC* and *Ct* were converted from general feature format (GFF) files from NCBI by gffread (Pertea and Rozenberg). Gene expression was extracted from sorted BAM files by using different pipelines respectively: Stringtie (version: 1.3.4) (Pertea et al., 2016)/Ballgown (version: 2.12.0) (Frazee et al., 2015), Stringtie (version: 1.3.4) (Pertea et al., 2016)/DESeq2 (version: 1.20.0) (Love et al., 2014) and Cufflinks/Cuffdiff (version: 2.2.1) (Trapnell et al., 2012).

2.2.2.6 Differential expression analysis

Three applications were mainly used for differential expression analysis: DESeq2 (Version: 1.20.0) (Love et al., 2014), Ballgown (version: 2.12.0) (Frazee et al., 2015) and Cuffdiff (version: 2.2.1) (Trapnell et al., 2012). DESeq2 and Ballgown are both R packages and meanwhile with the support of the following R packages: pheatmap (Kolde, 2012), genefilter (Gentleman et al., 2015), dplyr (Wickham et al., 2015), devtools (Wickham and Chang, 2016), ggplot2 (Wickham, 2011), pca3d (Weiner, 2013). In addition, the R package CummeRbund (Goff et al., 2013) is involved in analyzing the output from Cufflinks.

2.2.3 Quantitative real-time PCR (RT-qPCR) validation

2.2.3.1 Cell culture and *Ct* infection

Cell culture and *Ct* infection were performed by our collaborator Dr. Karthika Rajeeve. *Chlamydia trachomatis* (strain *L2/434/Bu*) was used in this study. *Chlamydia* EBs were verified to be free of *Mycoplasma* contamination via PCR. The *Homo sapiens* cervix adenocarcinoma cell line HeLa229 (ATCC®CCL2.1™) cell line was grown in RPMI medium (Gibco) supplemented with 10% FCS (Biochrome). Human Umbilical Vein Endothelial Cells (HUVEC) cell line was cultivated in Medium 200 supplemented with LSGS (Gibco). Transfected cells were grown in the presence of G418 (Invitrogen) at a final concentration of 1 mg/mL. Cell cultures were grown and maintained at 37°C in a humidified tissue culture incubator with 5% CO_2 using standard tissue culture procedures.

Chlamydia trachomatis was grown in HeLa229 cells and HUVECs. HeLa229 cells and HUVECs were lysed by using glass beads (2.85–3.45 mm) after 48 hpi and 36 hpi respectively.

The cell debris was removed by centrifuging at 2,000 g for 10 min at 4°C. The supernatant was collected and centrifuged at 24,000 g for 30 min at 4 °C. The pellet was resuspended in sucrose–phosphate–glutamic acid (SPG; 10 mM sodium phosphate (8 mM Na_2HPO_4, 2 mM NaH_2PO_4), 220 mM sucrose, 0.50 mM L-glutamic acid) buffer and passed through G20 and G18 syringes to dissociate clumps. The bacteria were aliquoted and frozen at -80°C until use.

2.2.3.2   RT-qPCR

RNA was isolated from uninfected and *Chlamydia*-infected HeLa229 cells and HUVECs using RNA easy kit (Qiagen, Germany). RNA was reverse transcribed using a Revert Aid First Strand Synthesis Kit (Fermentas) according to the manufacturer's instructions and diluted 1:10 with RNase free water. RT-qPCR was performed as previously described in Karunakaran et al. (2015a). Briefly, RT-qPCR reactions were prepared with Quanta SYBR (Quanta Bio) and PCR was performed on a Step One Plus device (Applied Biosystems). Data were analysed using ΔCt method, Step One Plus software package (Applied Biosystems) and Excel (Microsoft). Endogenous control was *Ct* 16S rRNA. Primers were designed by qPrimer Depot and ordered from SIGMA Life Science. Primers have been listed in Chapter 2.1.6

2.2.3.3   Statistical analysis

The mean values ($\pm$ SEM) was calculated for 3 independent experiments for each target gene. And Student *t* test was later used to determine the significance of the differences by *p* value ≤ 0.05 (*: $p \leq 0.05$; **: $p \leq 0.01$).

3

Results

3.1 Metabolic modelling and flux analysis of *Ct* during infection

3.1.1 Genome-scale metabolic model reconstruction

Table 3.1.1 shows characteristics of *Ct*'s reconstructed metabolic model. A genome-scale metabolic model of *Ct* is reconstructed with 321 unique metabolites (203 internal nodes and 118 external nodes) and 277 reactions based on gene-reaction association from the described database and literature in Chapter 2.2.1.1. The model involved 171 enzymes and distributed in six Enzyme Commission (EC) classifications (Figure 3.1.1). Transferases and hydrolyses together constitute 64% in the metabolic network, with the number of 72 and 37 respectively. Isomerases and ligases were not prevalent, with the number of 12 for each of them. Subsystems in the primary metabolism of *Ct* included: metabolism of carbohydrates, energy, nucleotides, amino acids, lipids, glycans, cofactors and vitamins. Missing gaps in the functional reconstruction were seriously treated according to literature. Gaps without evidence supported would not be filled for the specificity of *Ct*. It was settled that all amino acids nucleotides and biomass were external that could be taken from the

host, for the transporters on the inclusion membrane were still not known clearly.

Table 3.1.1: Characteristics of the reconstructed metabolic network of *Ct*

Genome size	1.04 Mb
Total ORFs	935
Total proteins	887
Number of enzymes used	171
Number of metabolic reactions	277
Number of metabolites	321
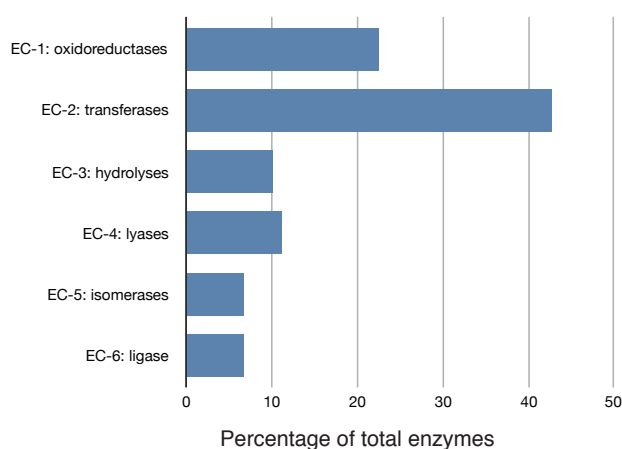
Figure 3.1.1: Enzymes involved in the reconstructed models grouped by Enzyme Commission (EC) classification.

3.1.2 Flux analysis based on extreme pathways

To predict the steady state initially, the reconstructed metabolic model was resulted in 84 enzyme-constrained pathways by internal convex basis algorithm (Kaleta et al., 2006) on the basis of the stoichiometric matrix. Average production loss was 1.38 out of 61, with a robustness score of 97.74%, and average modes lost of 2.06. Figure 3.1.2 shows the quantitative fluxes in steady state. The pathway with the strongest flux strength in the whole system is (the thickest blue pathway in Figure 3.1.2):

$$\text{pyruvate} \longrightarrow \text{acetyl}-\text{CoA} \longrightarrow \text{malonyl}-\text{CoA} \rightleftharpoons \text{malonyl}-\text{ACP} \tag{3.1}$$
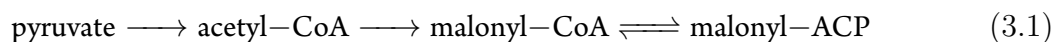
To optimize the solutions based on experimental data, genes encoded essential enzymes involved in the network were re-annotated and valued from the available proteomics dataset (Østergaard et al., 2016) with time tracing followed by 20 hpi, 40 hpi and later in cell lysis respectively. For each time point, enzyme activity for the EB group and RB group were
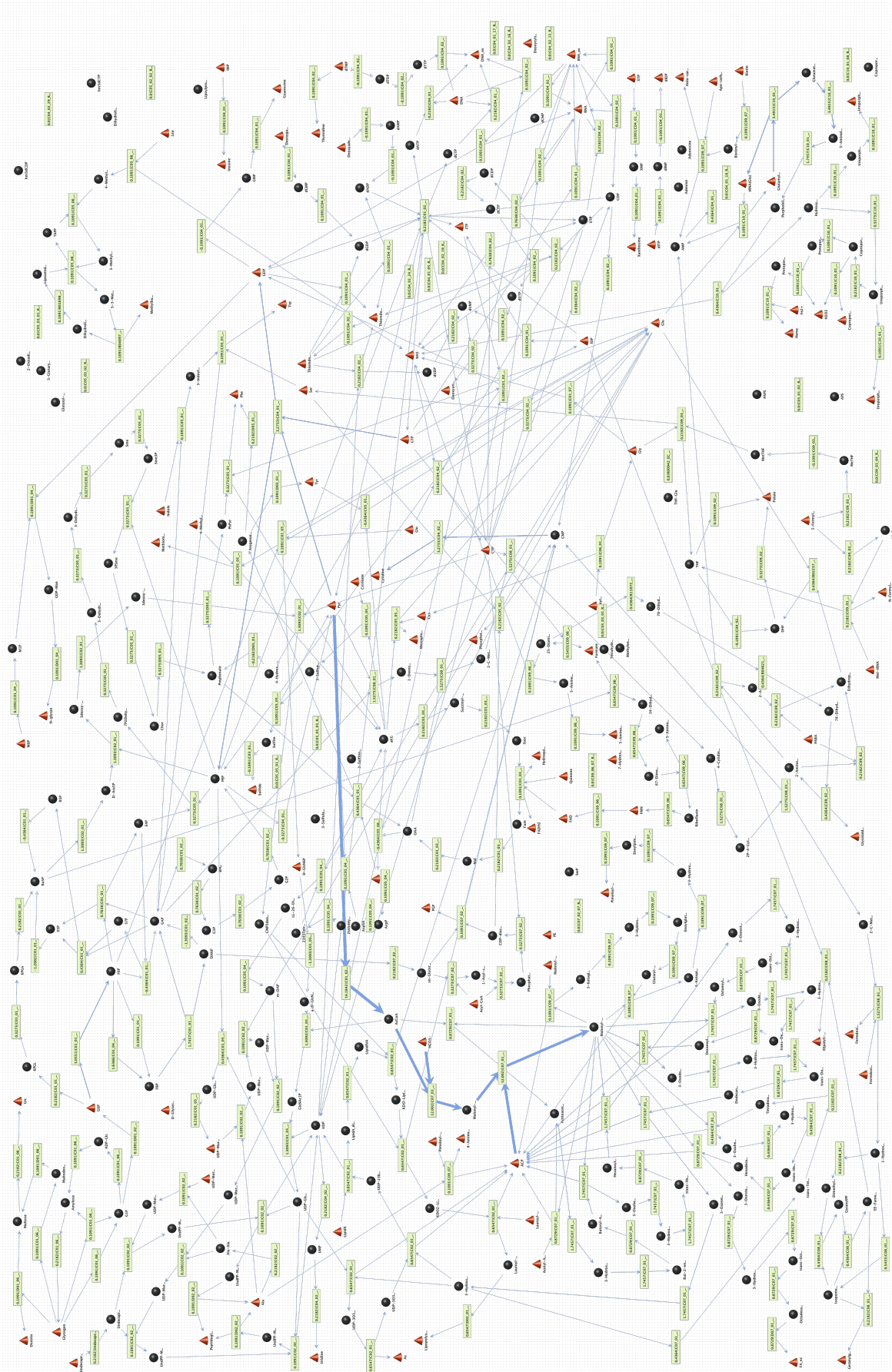
Figure 3.1.2: Steady state calculated from the reconstructed model. Internal and external metabolites are shown in black ball and red cone respectively. Valued paths refer to the flux intensity. The thickest blue flux is pyruvate $\longrightarrow$ acetyl$-$CoA $\longrightarrow$ malonyl$-$CoA $\rightleftharpoons$ malonyl$-$ACP.

input into the model for analyzing pathway activities quantitatively. Without the quantitative control of the ratio between EB and RB after infection in the original proteomic dataset, EB+RB group is also settled by the normalized combination of EB and RB to show a general feature of *Ct*. Here is the summary of the results with efficient fluxes in different metabolic session.
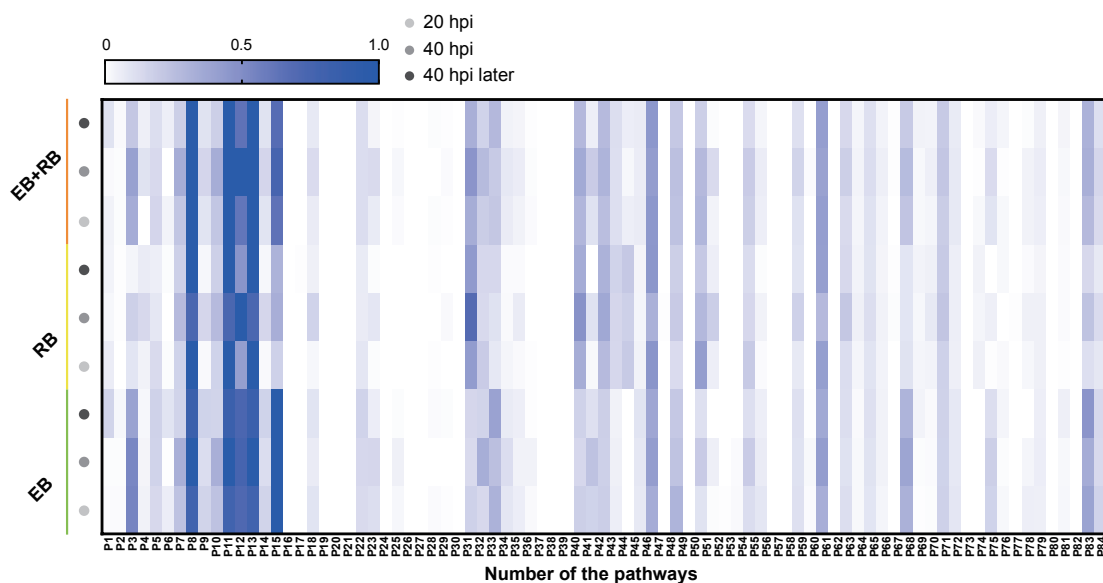
Figure 3.1.3: Heatmap overview of pathways differentially expressed in *Ct*'s different forms during different infection time. Shown is pathway intensities normalized to [0, 1] at 20 hpi, 40 hpi and 40 hpi later respectively by heatmap. hpi: hour post infection

3.1.2.1 Central pathways

Pentose phosphate pathway (PPP), glycolysis together with glycerophospholipids synthesis (EMP/GPL), two phosphofructokinases reversible in glycolysis (*pfkA_1* (CT_205) and *pfkA_2* (CT_207)), reduced tricarboxylic cycle (TCA) and fatty acid synthesis (FA) were activated and summarised in the central metabolism (Table 3.1.2). Figure 3.1.4 shows the pathway strengths. PPP, EMP/GPL and pfkA presented higher intensities in both EB and RB groups compared to that in TCA and FA. EB, the form of setting up carbohydrates, activated more in PPP, EMP/GPL and even TCA than that of RB. EMP/GPL pathway was much more activated in central metabolism with an average intensity of 0.40 in EB compared the 0.16 in RB. Phosphofructokinase, the pacemaker of glycolysis, revealed the average intensity of 0.32 in EB and 0.11 in RB. It suggests glycolysis and glycerophospholipids biosynthesis, which produced more by EB than RB, is of vital importance in *Ct*'s metabolism.

Table 3.1.2: Pathways in central metabolism[*]

Pathway	Reaction
PPP (P84)	$6\,PGn + 2\,G6P + 2\,GDP + 2\,H_2O + 5\,NAD^+ + 4\,NADP^+ + Orthophosphate \longrightarrow 6\,PGL + 4\,CO_2 + 2\,GTP + 9\,H^+ + 5\,NADH + 4\,NADPH + Pyr + Ru5P$
EMP/GPL (P83)	$2\,P + 2\,Acyl\text{–}CoA + CTP + 2\,G6P + 4\,GDP + NAD^+ + NADPH \longrightarrow 2\,CoA + Cytidine + 4\,GTP + H_2O + NADH + NADP^+ + Orthophosphate + PGP + 2\,Pyr$
pfkA (P34)	$ATP + Orthophosphate \longrightarrow 2\,P + ADP$
TCA (P46)	$2\,ADP + FAD + 2\,Glu + 2\,H_2O + 4\,NAD^+ + 2\,Orthophosphate + Quinone \longrightarrow 2\,ATP + 2\,Asp + FADH_2 + 4\,H^+ + Hydroquinone + 4\,NADH$
FA1** (P66)	$18\,ATP + 2\,Acetyl\text{–}ACP + 18\,H^+ + 18\,HCO_3^- + 6\,NAD^+ + 26\,NADPH + 18\,Pyr \longrightarrow 18\,ADP + 36\,CO_2 + 4\,FA\_ex + 18\,H_2O + 6\,NADH + 26\,NADP^+ + 18\,Orthophosphate$
FA2** (P67)	$2\,ACP + 18\,ATP + 16\,H^+ + 18\,HCO_3^- + 8\,NAD^+ + 26\,NADPH + 20\,Pyr \longrightarrow 18\,ADP + 38\,CO_2 + 4\,FA\_ex + 18\,H_2O + 8\,NADH + 26\,NADP^+ + 18\,Orthophosphate$
EMP/GNG (P5)	$2\,P + G6P + 2\,NAD^+ + Orthophosphate \rightleftharpoons 4\,GTP + 2\,H^+\,2\,H_2O + 2\,NADH + 2\,Pyr$

*:PPP: pentose phosphate pathway; EMP/GPL: glycolysis and glycerophospholipid biosynthesis; pfkA: 6-phosphofructokinase 1 (pfkA_1, EC 2.7.1.11) and diphosphate-fructose-6-phosphate 1-phosphotransferase (pfkA_2, EC 2.7.1.90) formed cyclic pathway; TCA: tricarboxylic acid cycle; FA1 and FA2: both are fatty acid biosynthesis. EMP/GNG: glycolysis and gluconeogenesis.

**: In fatty acid biosynthesis, FA1 utilizes 3-oxoacyl-ACP synthase II (FabF, CT_770, EC 2.3.1.179) on the step from malonyl-ACP to acetoacetyl-ACP, while FA2 use both FabF and 3-oxoacyl-ACP synthase III (FabH, CT_239, EC 2.3.1.180) for the same step. ACP: acyl carrier protein.

Besides the dominant glycolysis, *Ct* presented capability of synthesizing its own phospholipids and fatty acids (though in low intensity) in both EB and RB groups. *Ct* was reported to utilize host fatty acid for phospholipid synthesis (Yao et al., 2015) and accumulate lipid droplets (LDs) (Cocchiaro et al., 2008, Kumar et al., 2006). Also, the pathway 3.1 shows a strong intensity from the downstream of glycolysis and the upstream of fatty acid synthesizing (acetyl-CoA was set as external), while the whole pathway of biosynthesis of fatty acid is almost inert, whose average intensities are below 0.04 and 0.01 in FA1 and FA2 respectively in both EB and RB groups. It suggests *Ct* is supported by acyl-CoA from host LD to synthesize its own phospholipids and utilize host fatty acid directly maybe for forming its outer membrane.

The incomplete TCA cycle displays lower reaction activity in both EB and RB groups compared to the average, which was 0.08 and 0.04 respectively. The slight reducing of TCA in RB group was probably because It might be easier for *Ct* to take energy source directly from the host in the reticulate form.

In addition, gluconeogenesis was also observed in this study. Figure 3.1.4 (G) clearly shows during the infection from 20 hpi to 40 hpi and afterwards, EB went from gluconeogenesis to glycolysis steadily; RB, whose variation trend matched the general trend in EB+RB group, converted from glycolysis to gluconeogenesis, and transformed again to glycolysis afterwards. *Ct* was easier to take not only energy source but also carbon source (e.g. G6P)
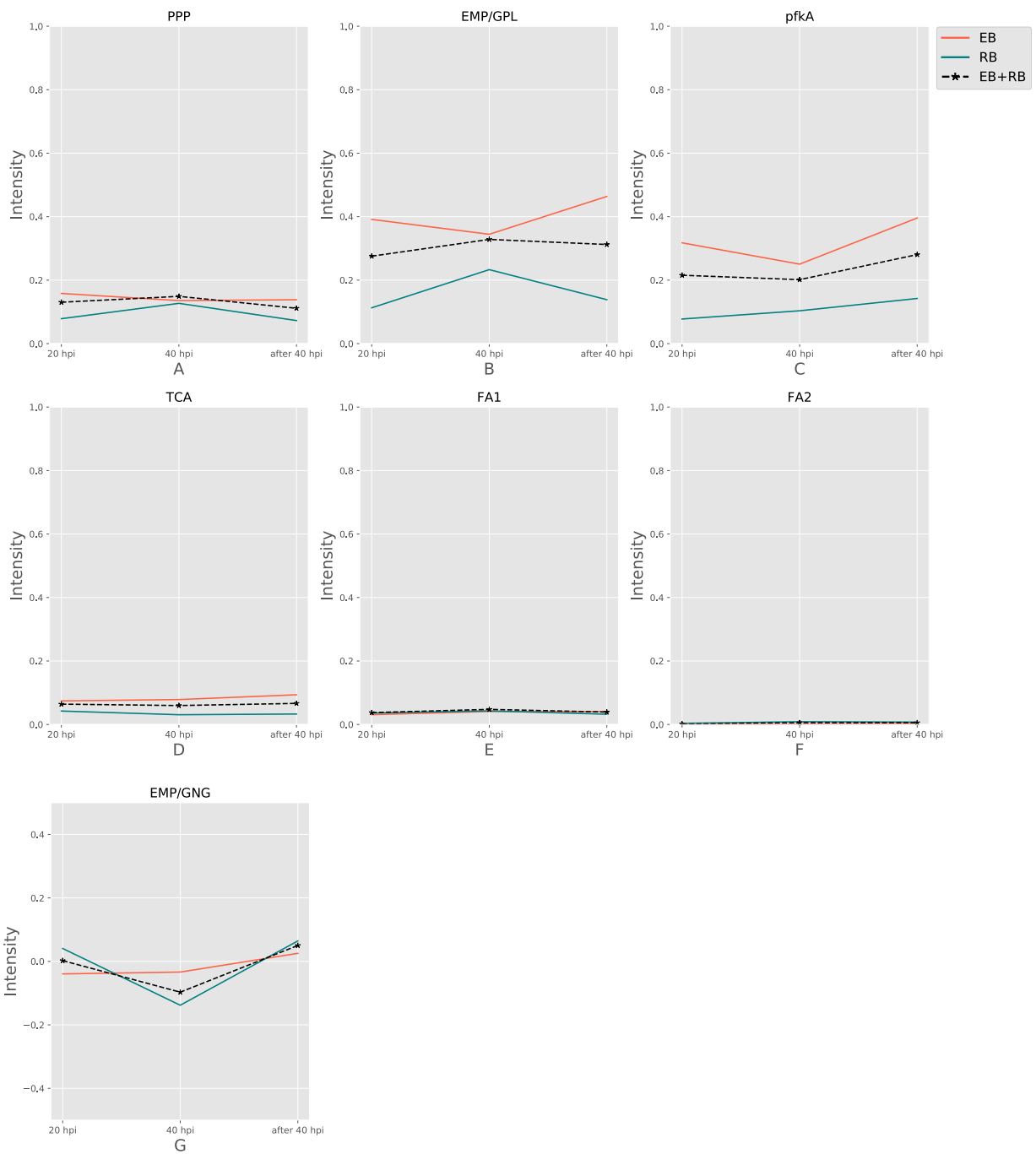
Figure 3.1.4: Pathway activities of central metabolism. EB is given in red solid line and RB is shown in green solid line; combined flux during infection group EB+RB is given in black dashed line. The time course indicates the variation over infection from 20 hpi, 40 hpi and later after 40hpi respectively. Flux intensity is normalized in [0, 1] EMP: glycolysis; GPL: glycerophospholipid metabolism; GNG: gluconeo- genesis; TCA: tricarboxylic acid cycle; FA1 and FA2: fatty acid biosynthesis.

from the host at 20 hpi when more RBs than EBs were available in the host. Once the transformation from RB to EB started, it stores carbon source for further usage. Though the intensity of RB (-0.14) is three times more than that of EB (-0.03) at 40 hpi, it is hard to conclude which stage of *Ct* prefers because of the unknown quantitative ratio between EB and RB. Nevertheless, the presence of gluconeogenesis could support carbon metabolism when *Ct* meets the necessary.

3.1.2.2 Folate Biosynthesis

Folate biosynthesis is a central pathway for *Ct*. This biosynthesis was divided into four subpathways (Figure 3.1.5, P33, P51, P55 and P56). Two subpathways, P33 and P51, converted glutamate and 4-aminobenzoic acid (PABA) to glycolaldehyde and folate (Table 3.1.3). Strengths of both pathways were enriched above average for all pathways, which indicated the importance of folate biosynthesis in *Ct* metabolism. The difference between P33 and P51 was the consumption of ATP and H2O. Less ATP was served in P33 with more activation in EB than RB. In contrast, P51 displayed higher activation in RB than EB when more ATP is available. The variable trend of EB+RB group also presented similar results, which increased first and then decreased from 20 hpi, 40 hpi and the lysis phase. It was similar to the quantity pattern of EB in P33, and steadily went down during the infection which fit for RB's quantity in P51. The ATP generated in P51 but not in P33 were relied on the ATP:AMP phosphotransferase (ATP + AMP $\rightleftharpoons$ 2 ADP). This suggests RB relies more on ATP consumption when it is easier to utilize from both the host and the self-production by activated nucleotides metabolism. There are other folate-related pathways with less ATP consuming in P55 and P56 though with lower intensity. P55 was the transformation among folate, tetrahydrofolate and dihydrofolate. And P56, although not as active as P33 and P51, could be an alternative pathway for folate biosynthesis, especially for EB.

Table 3.1.3: Extreme pathways in folate biosynthesis

Pathway	Reaction
P33[a]	$2\,\mathrm{ATP} + 2\,\mathrm{GTP} + 2\,\mathrm{Glu} + \mathrm{NAD}^+ + \mathrm{NADP}^+ + 2\,\mathrm{Orthophosphate} + 2\,\mathrm{PABA} \longrightarrow 4(2P) + 2\,\mathrm{ADP} + 2\,\mathrm{Folate} + 2\,\mathrm{Formate} + 2\,\mathrm{Glycolaldehyde} + 2\,\mathrm{H}^+ + \mathrm{NADH} + \mathrm{NADPH}$
P51[b]	$6\,\mathrm{ATP} + 2\,\mathrm{GTP} + 2\,\mathrm{Glu} + 2\,\mathrm{H_2O} + \mathrm{NAD}^+ + \mathrm{NADP}^+ + 2\,\mathrm{Orthophosphate} + 2\,\mathrm{PABA} \longrightarrow 62P + 6\,\mathrm{ADP} + 2\,\mathrm{Folate} + 2\,\mathrm{Formate} + 2\,\mathrm{Glycolaldehyde} + 2\,\mathrm{H}^+ + \mathrm{NADH} + \mathrm{NADPH}$
P55[c]	$\mathrm{Folate} \longrightarrow \mathrm{THF} \rightleftharpoons \mathrm{DHF} \rightleftharpoons \mathrm{Folate}$
P56[d]	$2(5-\mathrm{Formyl}-\mathrm{THF}) + 2\,\mathrm{ATP} + 2\,\mathrm{Gly} + 2\,\mathrm{H_2O} + \mathrm{NAD}^+ + \mathrm{NADP}^+ \longrightarrow 2\,\mathrm{ADP} + 2\,\mathrm{Folate} + 2\,\mathrm{H}^+ + \mathrm{NADH} + \mathrm{NADPH} + 2\,\mathrm{Orthophosphate} + 2\,\mathrm{Ser}$

Reaction involved based on KEGG reaction number:

a: R03504, R03066, R02235, R00425, R02237, R04621, R11072;

b: R00127, R03504, R03503, R03067, R02235, R00425, R02237, R04621, R011071;

c: R02235, R00937, R00936;

d: R02235, R00936, R00945, R01200, R02301.
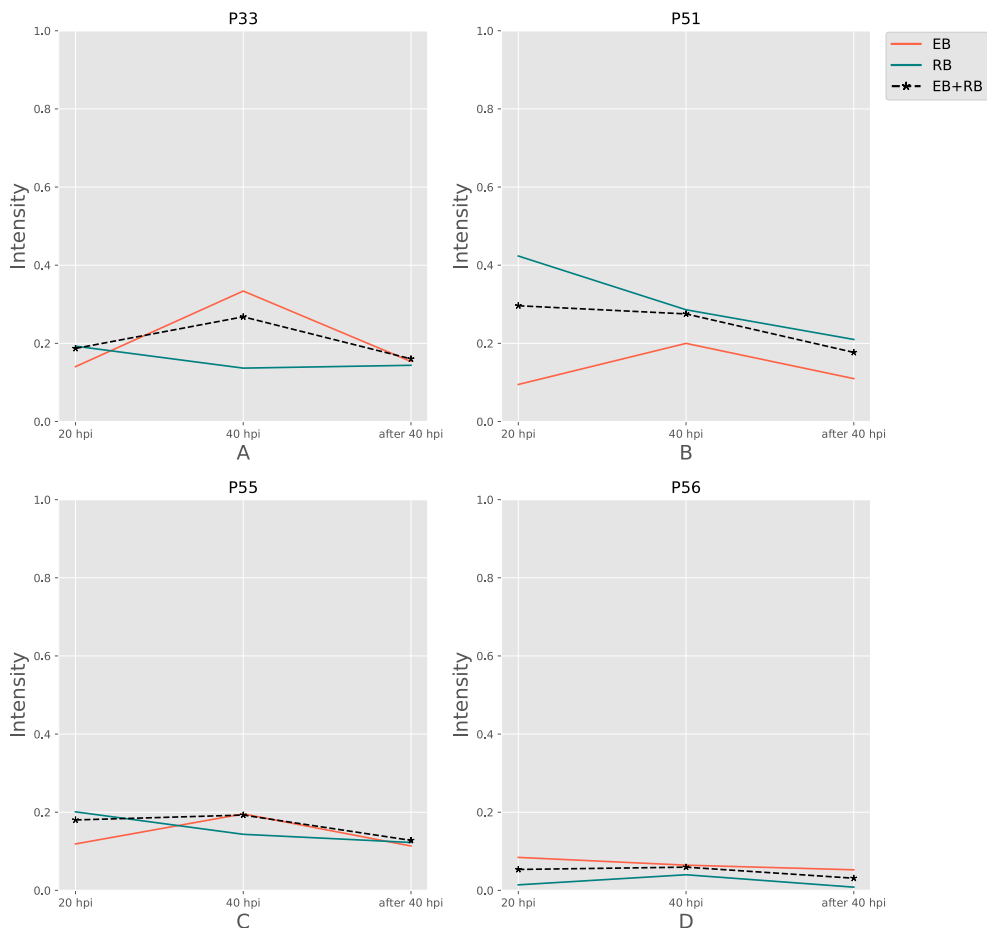
Figure 3.1.5: Pathway activities of Folate biosynthesis. P33, P51, P55 and P56 are all extreme pathways synthesizing folate. EB is given in red solid line and RB is shown in green solid line; combined flux during infection group EB+RB is given in black dashed line. The time course indicates the variation over infection from 20 hpi, 40 hpi and later after 40 hpi respectively. Flux intensity is normalized in [0, 1].

3.1.2.3 Amino acid metabolism

3.1.2.3.1 Cysteine metabolism

Biosynthesis of cysteine is important for the arrangement of the EB's outer membrane. Because *Ct* contains type three secretion system (T3SS) which composed by enriched disulfide bonds in infectious form. In our result, the intensities of cysteine synthesis pathways (Figure 3.1.6 A, B and C) were much higher than cysteine consumption pathways (Figure 3.1.6 D, E and F) on average. It presented 2-5 times more activities in EB than RB, which supported to the current consensus that cysteine should be enriched in EBs for the formation of the cysteine-rich outer membrane.
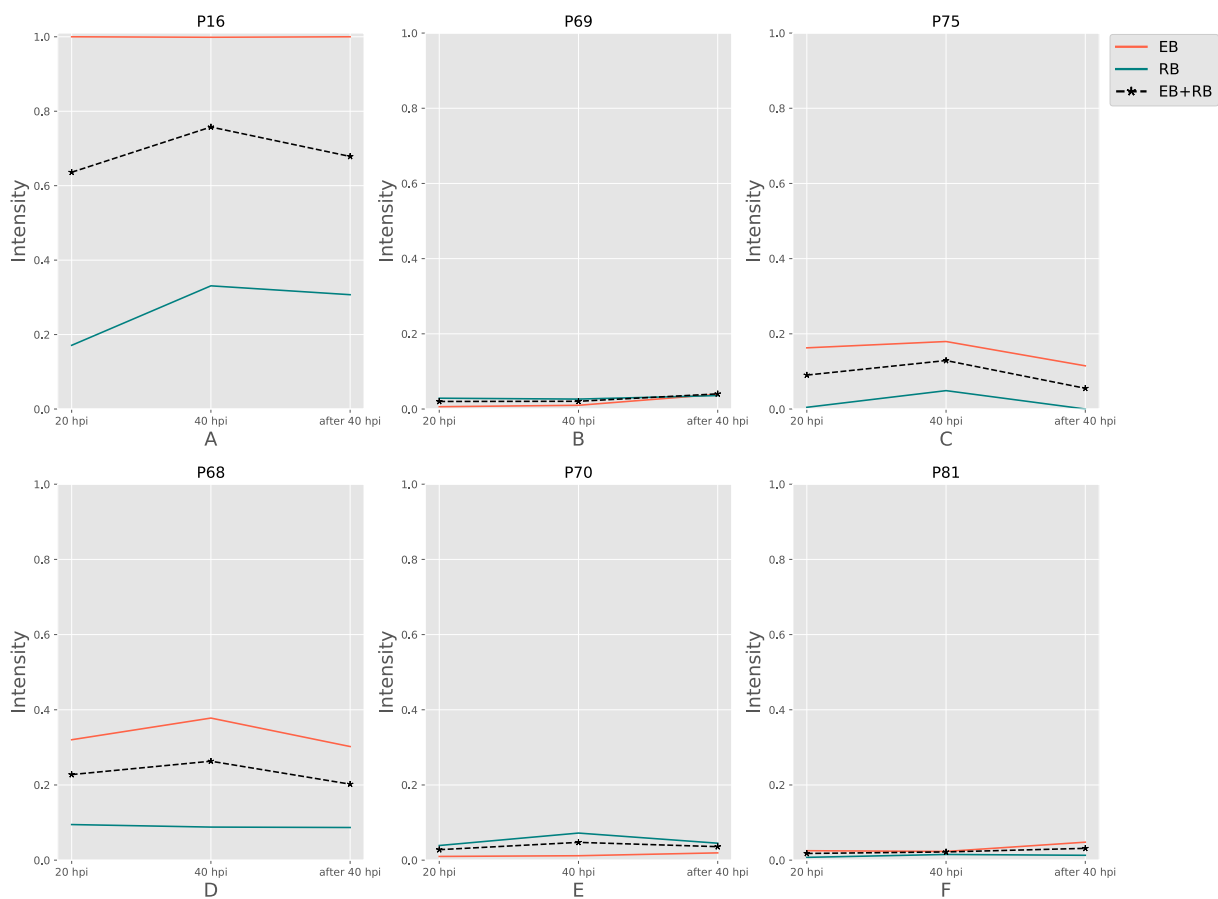
Figure 3.1.6: Cysteine and other amino acids associated pathways. A, B and C are cysteine biosynthesis pathways; D, E and F are cysteine consumptions. EB is given in red solid line and RB is shown in green solid line; combined flux during infection EB+RB is given in black dashed line. The time course indicates the variation over infection by from 20 hpi, 40 hpi and later after 40 hpi respectively. Flux intensity is normalized in [0, 1].

3.1.2.3.2 Peptidoglycan biosynthesis

Extreme pathway P70 (Table 3.1.4) was also combined with peptidoglycan (PG) biosynthesis. With the encoding *Mur* genes for PG synthesis and being sensitive to penicillin, the absence of PG in *Ct* is quite unusual until new labelling methods help to reveal PG (Liechti et al., 2014, Pilhofer et al., 2013). P70, though with weak intensity, was still active and slightly presented more in RB than EB (Figure 3.1.6 E).

Table 3.1.4: Cysteine and other amino acids associated metabolism

Pathway	Reaction
P16[a]	$\mathrm{Cysteinyl-glycine} + H_2O \longrightarrow \mathrm{Cys} + \mathrm{Gly}$
P68[b]	$\mathrm{4\text{-}Methylthio-2\text{-}oxobutanoate} + \mathrm{Cys} \longrightarrow \mathrm{Mercaptopyruvate} + \mathrm{Methionine-L}$
P69[c]	$\mathrm{Cysteate} + 3\,H^+ + \mathrm{Mercaptopyruvate} + 3\,\mathrm{NADPH} \longrightarrow \mathrm{Cys} + 2\,H_2O + 3\,\mathrm{NADP}^+ + \mathrm{Pyr} + \mathrm{Sulfide}$
P70[d]	$7\,\mathrm{ATP} + 2\,\mathrm{Ala} + \mathrm{Asp} + \mathrm{Cys} + \mathrm{GlcN6P} + \mathrm{Glu} + H^+ + \mathrm{NAD}^+ + 2\,\mathrm{NADPH} + 2\,\mathrm{Pyr} + \mathrm{UDP-MurNAc} + \mathrm{Undecaprenol} \longrightarrow 2\,\mathrm{P} + 7\,\mathrm{ADP} + CO_2 + H_2O + \mathrm{Mercaptopyruvate} + \mathrm{NADH} + 2\,\mathrm{NADP}^+ + 6\,\mathrm{Orthophosphate} + \mathrm{Peptidoglycan}$
P75[e]	$2\,\mathrm{ADP} + 2\,\mathrm{Asp} + \mathrm{GTP} + \mathrm{ITP} + 2\,\mathrm{Mercaptopyruvate} \longrightarrow 2\,\mathrm{ATP} + 2\,CO_2 + 2\,\mathrm{Cys} + \mathrm{GDP} + \mathrm{IDP} + 2\,\mathrm{Pyr}$
P81[f]	$2\,\mathrm{P} + 3\,\mathrm{ADP} + 3\,\mathrm{Cys} + 2\,\mathrm{G6P} + 12\,\mathrm{GTP} + 3\,H^+ + 3\,\mathrm{NADPH} + 6\,\mathrm{Pyr} \longrightarrow 3\,\mathrm{ATP} + 3\,CO_2 + 12\,\mathrm{GDP} + 3\,H_2O + 3\,\mathrm{Mercaptopyruvate} + 3\,\mathrm{NADP}^+ + 13\,\mathrm{Orthophosphate} + 3\,\mathrm{Phe}$

Reaction involved based on KEGG reaction number:
a: R00899;
b: R07396, R00895;
c: R00858, R02433, R00895, R04861;
d: R00014, R03270, R02569, R07618, R00416, R05332, R02060, R03193, R02783, R02788, R01150, R04617, R05630, R05032, R00480, R02291, R10147, R04199, R07613, R02735, R00895, R05626;
e: R00431, R00200, R00895, R00355;
f: R01056, R01529, R02740, R01070, R01830, R01827, R01641, R01015, R02073, R00330, R01826, R03083, R03084, R02413, R02412, R03460, R01714, R00694, R00895, R01715, R01373.

3.1.2.3.3 Uptake of glutamate and glutamine

It is assumed that the TCA cycle in *Ct* is abbreviate, for the lack of genes encoding citrate synthase, aconitate hydratase and isocitrate dehydrogenase. It enters into incomplete TCA cycle by 2-oxoglutarate and turns out to oxaloacetate. With the appearance of aspartate transaminase (EC 2.6.1.1), the oxaloacetate could exchange with aspartate accompanied by barter between glutamate and 2-oxoglutarate. However, the flux prefers to drive from glutamate to aspartate (Figure 3.1.4, D; Table 3.1.2, TCA (P46)). It is possible that an anaplerosis shown in Figure 3.1.4 D is commonly completed in almost all aerobic organisms. Nevertheless, because of the low flux strengths, this unfeasible cycle is not significantly responsible for gaining chemical energy in the form of ATP.

In most microorganisms, glutamine synthetase (EC 6.3.1.2), glutaminase (EC 3.5.2.1) and glutamate synthase (EC 1.4.1.13/1.4.1.14) are important for switching between glutamate and glutamine. Normally, glutamine could be converted to glutamate after taken from the environment or host cell. Without encoding genes for these enzymes, *Ct* presents glutamine-fructose-6-phosphate transaminase (*glmS*, EC 2.6.1.16), whose protein product converts glutamine and fructose 6-phosphate to glutamate and glucosamine 6-phosphate. Besides, for the existence of CTP synthase (*pyrG*, EC 6.3.4.2), glutamate could be transformed from glutamine. *pyrG* is not involved in long essential pathways like *glmS*, because no real pyrimidine metabolism was formed in *Ct*. Both *glmS* and *pyrG* were detected with expressions in the proteomics data used (Østergaard et al., 2016).

Although glutamate and glutamine are both necessary in the metabolism of *Ct*, it was still unknown that *Ct* imports only glutamine or both glutamine and glutamate from the host cell. Interestingly, three glutamate-related transporters, CT_216, CT_230 and CT_401 are encoded in the genome, and both CT_216 and CT_401 are expressed in the protein level (Østergaard et al., 2016). However, glutamine transporters are not confirmed yet.

3.1.2.4 Fragmentation pathways of purine and pyrimidine

It is known that no PRPP is detected in *Ct*, neither through biosynthesis nor from host cell via membrane transporter (Stephens et al., 1998). The whole purine and pyrimidine metabolism does not strictly exist in *Ct*. However, many genes encoding enzymes involved in nucleic acid metabolism are detected in *Ct*'s genome (Stephens et al., 1998) and could form theoretical metabolic pathways (Figure 3.1.8). Besides, some of these enzymes were represented in the proteomics data (Østergaard et al., 2016) with high expressions when two or three reactions build small fragments of pathways. Most of these enzymes were quite active in RB. It is still not clear why these enzymes are encoded in *Ct*'s degenerate genome. However, it potentially offers small-scale energy providing and also helps with the amino acid transformation (e.g. *pyrG* as described) when it is required. In short, the nucleotide acid metabolism may play the role of assistant for pathway alternation during infection.

3.1.3 RT-qPCR validation

To validate the results of the metabolic model, gene expression of candidate enzymes in *Ct* were examined by RT-qPCR as shown in Figure 3.1.7. RT-qPCR is a sensitive, efficient and precise approach for validation based on *Ct*'s specificity. HeLa229 cell line was used as the host and infected in different duration: 12 hpi (as the control of the fold change), 24 hpi

and 48 hpi. Because RT-qPCR is very sensitive and hard to control, here we only presented results from our experienced collaborator Dr. Karthika Rajeeve for data accuracy (though I did some of the validation).

Table 3.1.5: Candidate genes for RT-qPCR validation

CT no.	Gene	Reaction**	EC no.	annotation	pathway***
CT121*	*araD*	R01529	5.1.3.1	ribulose-phosphate 3-epimerase	
CT750*	*tktB*	R01830	2.2.1.1	transketolase	PPP
CT313*	*tkl*	R01827	2.2.1.2	transaldolase	
CT378*	*pgi*	R02740	5.3.1.9	glucose-6-phosphate isomerase	
CT205	*pfkA_1*	R04779	2.7.1.11	6-phosphofructokinase	
CT505*	*gapA*	R01061	1.2.1.12	glyceraldehyde 3-phosphate dehydrogenase	EMP
CT332	*pykF*	R00200	2.7.1.40	pyruvate kinase	
CT207*	*pfkA_2*	R02073	2.7.1.90	pyrophosphate–fructose-6-phosphate 1-phosphotransferase	
CT714*	*gpdA*	R00842	1.1.1.94	glycerol-3-phosphate dehydrogenase (NAD(P)+)	
CT807	*plsB*	R00851	2.3.1.15	glycerol-3-phosphate O-acyltransferase	GPL
CT811	*plsX*				
CT775	*plsC*	R00241	2.3.1.51	1-acyl-sn-glycerol-3-phosphate acyltransferase	
CT376*	*mdhC*	R00342	1.1.1.37	malate dehydrogenase	
CT855*	*fumC*	R01082	4.2.1.2	fumarate hydratase class II	
CT821*	*sucC*	R00405	6.2.1.5	succinyl-CoA synthetase beta subunit	TCA
CT822*	*sucD*			succinyl-CoA synthetase alpha subunit	
CT054	*sucA*	-	1.2.4.2	2-oxoglutarate dehydrogenase	
CT055	*sucB*		2.3.1.61	2-oxoglutarate dehydrogenase	
CT238*	*fabD*	R01626	2.3.1.39	acyl-carrier-protein] S-malonyltransferase	FA
CT816	*glmS*	R00768	2.6.1.16	glucosamine-fructose-6-phosphate aminotransferase	Gln/Glu
CT183	*pyrG*	R00571	6.3.4.2	CTP synthase	
CT614	*folX*	R03504	4.1.2.25	dihydroneopterin aldolase	
CT613	*folP*	R03067	2.5.1.15	dihydropteroate synthase	Folate
CT612*	*folA*	R02235	1.5.1.3	dihydrofolate reductase	

*: this enzyme triggers the reversible reaction. **: Reaction is performed in KEGG Reaction number. ***:PPP: pentose phosphate pathway; EMP: glycolysis; GPL: glycerophospholipid biosynthesis; TCA: tricarboxylic acid cycle; FA: fatty acid biosynthesis; Gln/Glu: converting glutamine to glutamate; Folate: Folate biosynthesis.

In HeLa229 cell lines (Figure 3.1.7), most of the enzymes in glycolysis were clearly upregulated at 24 hpi, and downregulated at 48 hpi. The two phosphofructokinases were very active, and *pfkA_2*, functional on reversible reactions, expressed nearly two folds more at 24 hpi compared to *pfkA_1*, which triggers the one-way step in glycolysis. Expressions of glucose-6-phosphate isomerase (*pgi*) and pyruvate kinase (*pykF*) were slightly increased with the folds of 1.25 and 1.2 respectively. It indicates that glycolysis is still strongly

(a) *Ct* gene expression in HeLa229 cell lines

(b) *Ct* gene expression in HUVECs
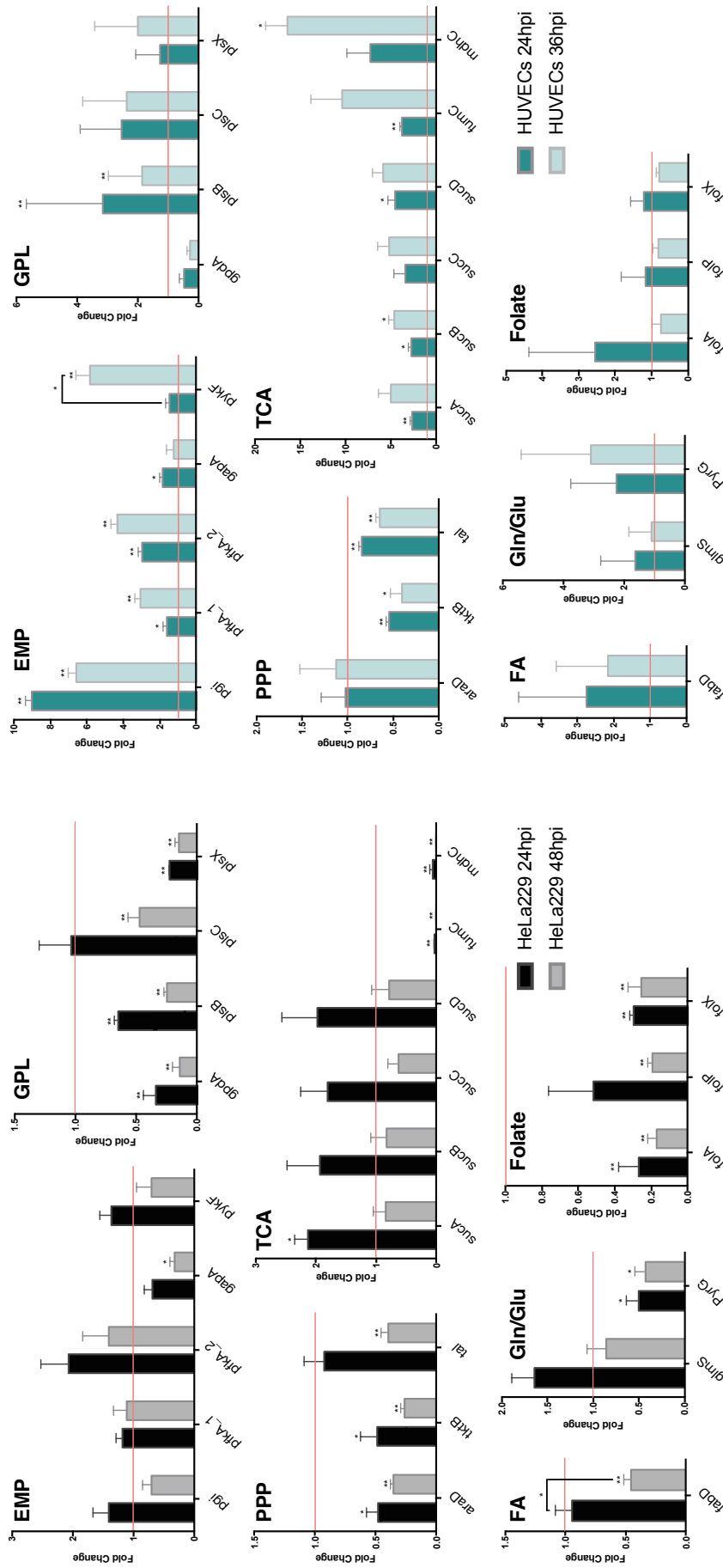
Figure 3.1.7: Quantitative RT-PCR analysis of gene expression in major metabolic pathways of *Ct* during infection in HeLa229 cell lines and HUVECs cell lines. The gene expressions were analyzed at 24 hpi and 48 hpi for HeLa229, and 24 hpi and 36 hpi for HUVECs. Each gene was quantified in 3 independent experiments with the mean values ($\pm$ SEM) compared to the expression at 12 hpi shown as a control in red line. Statistical analysis was performed by SEM and Student *t* test (*: $p \leq 0.05$; **: $p \leq 0.01$). PPP: pentose phosphate pathway; EMP: glycolysis; GPL: glycerophospholipid biosynthesis; TCA: tricarboxylic acid cycle; FA: fatty acid biosynthesis; Folate: Folate biosynthesis. Figure was prepared based on the raw data from Dr. Karthika Rajeeve.

available during RB-to-EB stage and increases activities as the simulation shown in the (EB+RB) group in Figure 3.1.4B. *pfkA_2* was supposed to be more efficient in the glycolysis of *Ct* for its utilization of diphosphate instead of ATP. In glycerolphospholipids pathway, most of the enzymes for phospholipid biosynthesis reduced expressions during infection except *plsC*. *plsC* increased expression at 24 hpi. It suggests the biosynthesis of glycerolphospholipids is probably more relied on the downstream of glycolysis and the Acyl-CoA, rather than directly went from GAPDH in upstream of glycolysis (Figure 3.1.8), which fit for the calculation as described in pathway 3.1. Expressions of PPP-related genes were not as intense as that in glycolysis. The expression of L-ribulose-5-phosphate 4-epimerase (*araD*) and transketolase (*tktB*) decreased by nearly 50% from 12 hpi to 24 hpi, while transaldolase (*tal*) slightly reduced the expression. Expressions of the genes in TCA presented various results. The *sucABCD* exhibited about 2 folds upregulation at 24 hpi, while fumarate hydratase (*fumC*) and malate dehydrogenase (*mdhC*) decreased at 24 hpi and almost invisibly expressed at 48 hpi. Moreover, the malonyl-CoA-ACP transacylase (*fabD*) was involved in the upstream of fatty acid biosynthesis (Table 3.1.2, FA1 (P67) and FA2 (P68); Figure 3.1.8) and associated with the formation of glycerolipids (Table 3.1.2, EMP/GPL (P84); Figure 3.1.8). It slightly downregulated at 24 hpi (Figure 3.1.7). In addition, glutamine-fructose-6-phosphate transaminase (*glmS*) and CTP synthase (*pyrG*) are able to irreversibly convert glutamine to glutamate. *glmS* increased its expression by over 1.5 fold change at 24 hpi compared to that at 12 hpi, and it was still very active even at 48 hpi, the lysis phase. *pyrG* presented around 50% activity in both 24 hpi and 48 hpi compared to the 12 hpi. It could assume that both of the enzymes may be actively functional on transfer glutamine to glutamate. *glmS* could serve for both EB and RB, while *pyrG* mainly activated in RBs rather than EBs. Three genes in folate biosynthesis were validated. They were significantly decreased by even more than 50% in both 24 hpi and 48 hpi.

The human umbilical vein endothelial cells (HUVECs) were also used as the host for *Ct* infection to make a comparison (Figure 3.1.7). Target genes generally exhibited more expressions (and also more deviations) in HUVECs than in HeLa229. Glycolysis, pentose phosphate pathway and TCA in HUVECs presented similar trends of expression compared to these pathways in HeLa229. In glycolysis, *pfkA_2* expressed twice more than *pfkA_1* at 24 hpi, and presented almost triple folds at 24 hpi than in 36 hpi. Unlike in HeLa229, *gapA* was upregulated in HUVECs and expressed two times more at 24 hpi than at 12 hpi. In the pentose phosphate pathway, *tktB* and *tal* showed almost the same expression but *araD* was slightly upregulated in HUVECs. In TCA, the *suc* genes were overexpressed in both 24 hpi and 36 hpi compared to that in HeLa229. The other genes involved in glycerophospholipid biosynthesis, fatty acid biosynthesis, glutamine/glutamate

transformation and folate biosynthesis, showed different regulation in HUVECs versus to in HeLa229. *plsB*, *plsC*, *plsX* together with *fabD* highly increased the regulation at 24hpi and 36hpi, which suggested the intensive activity of synthesizing glycerophospholipid in HUVECs. The active biosynthesis was also presented by all increased regulations of *folA*, *folP* and *folX* for folate, which were differential expressions compared to that in HeLa229. In addition, the expression of *glmS* was similar in both cell lines. However, *pyrG*, which was downregulated in HeLa229, displayed two folds more upregulations at both 24 hpi and 36 hpi in HUVECs. It suggested the *Ct*'s *pyrG* was potentially able to convert glutamine to glutamate as well as *glmS*.

3.1.4   Summary

A genome-scale metabolic model of *Ct* with 321 metabolites and 277 reactions was first reconstructed to study *Ct* metabolism in the host cell during infection. The model was calculated and then performed to 84 extreme pathways to fulfil a steady state condition. These pathways were modelled at 20 hpi, 40 hpi and later in the lysis according to their metabolic flux strengths. The RT-qPCR experiments validated the modelled pathways with strength changes for key enzymes during *Ct* infection in HeLa229 cell lines at 12 hpi, 24 hpi and 48 hpi. Major active pathways in the carbohydrate metabolism of *Ct* are glycolysis, gluconeogenesis, pentose phosphate pathway and glycerolphospholipid biosynthesis pathway, whereas the incomplete TCA and fatty acid biosynthesis are less active. EB is more activated in almost all carbohydrate pathways than RB. *Ct* generally requires a lot of acetyl-CoA from the host for synthesizing phospholipids instead of fatty acid. Besides, both EB and RB are quite active in folate biosynthesis. When more ATP is available from both host cell and *Ct* itself, RB is more activated by utilizing energy providing chemicals generated by enzymes associated in the nucleic acid metabolism. The forming of folate suggests large glutamate consumption in *Ct*, which is supposed to be converted from glutamine by the glutamine-fructose-6-phosphate transaminase (*glmS*) and maybe also by CTP synthase (*pyrG*). Nucleic acid metabolism breaks into fragile pathways by two or three reactions based on calculation, whereas the enzymes involved were more active in RB than EB.
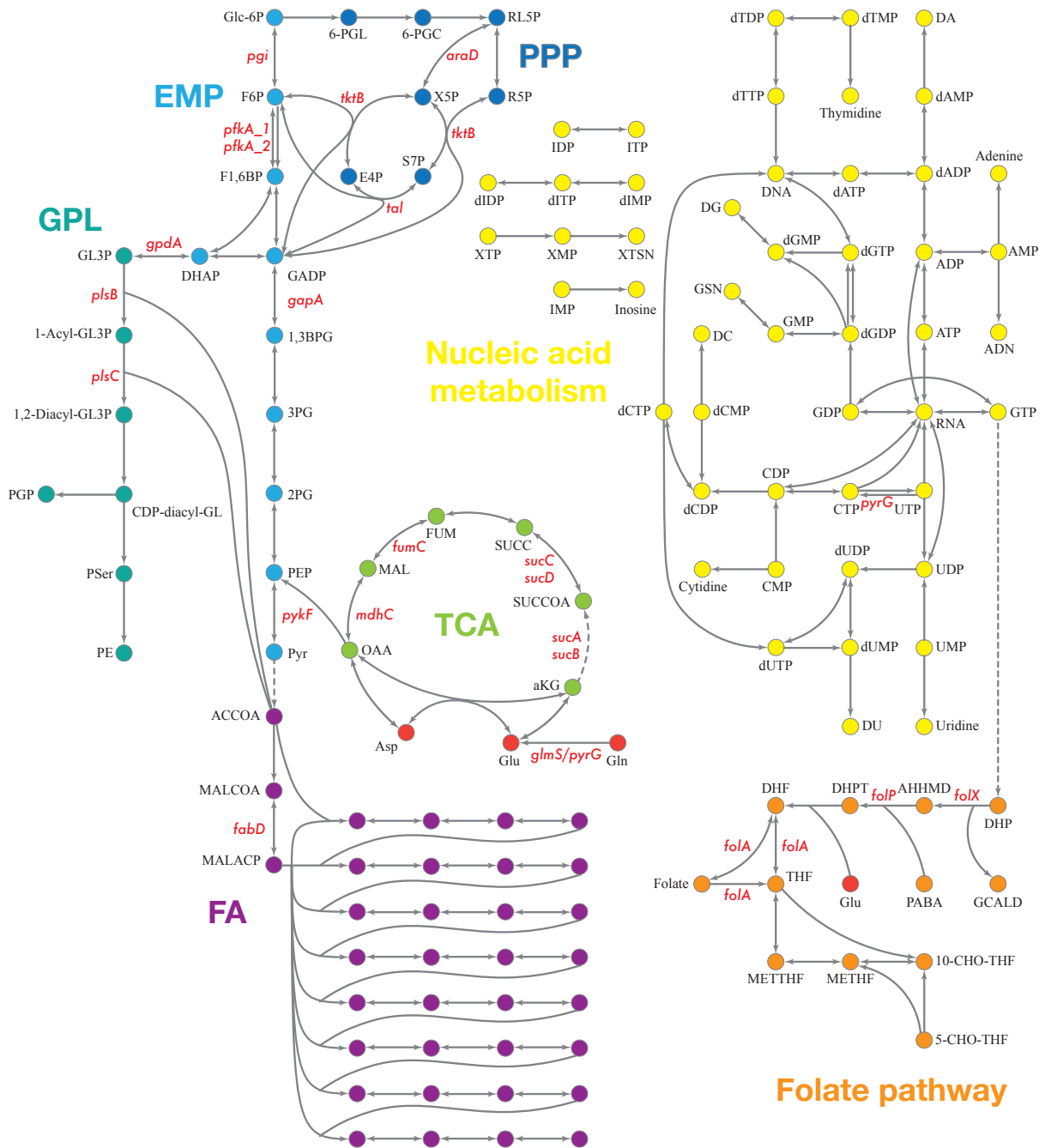
Figure 3.1.8: An overview of *Ct* central metabolism. Red labelled enzymes are validated by RT-qPCR. EMP: glycolysis; PPP: pentose phosphate pathway; GPL: glycerophospholipid metabolism; GNG: gluconeo- genesis; TCA: tricarboxylic acid cycle; FA: fatty acid biosynthesis. *araD*: ribulose-phosphate 3-epimerase; *tktB*: trans- ketolase; *tkl*: transketolase; *pgi*: glucose-6-phosphate isomerase; *pfkA_1*: 6-phosphofructokinase; *gapA*: glyc- eraldehyde 3-phosphate dehydrogenase; *pykF*: pyruvate kinase; *pfkA_2*: pyrophosphate–fructose-6-phosphate 1-phosphotransferase; *gpdA*: glycerol-3-phosphate dehydrogenase (NAD(P)+); *plsB*: glycerol-3-phosphate O- acyltransferase; *plsX*: glycerol-3-phosphate O-acyltransferase; *plsC*: 1-acyl-sn-glycerol-3-phosphate acyltrans- ferase; *mdhC*: malate dehydrogenase; *fumC*: fumarate hydratase class II; *sucC*: succinyl-CoA synthetase beta subunit; *sucD*: succinyl-CoA synthetase alpha subunit; *sucA*: 2-oxoglutarate dehydrogenase; *sucB*: 2-oxoglutarate dehydrogenase; *fabD*: acyl-carrier-protein S-malonyltransferase; *glmS*: glutamine-fructose-6-phosphate aminotrans- ferase; *pyrG*: CTP synthase; *folX*: dihydroneopterin aldolase; *folP*: dihydropteroate synthase; *folA*: dihydrofolate reductase.

3.2 Metatranscriptome and differential expression analysis of the host and *Neisseria gonorrhoeae* with *Chlamydia trachomatis* infection

This project is a collaboration with Prof. Dr. Caroline Genco (School of Medicine, Tufts University, Boston, USA). Samples for RNA-Seq was originally for studying *GC* infection. The samples were collected and anonymized from the National Center for Sexually Transmitted Disease in Nanjing, China. Thirteen of them were sequenced, with 6 male samples obtained by urethral swab and 7 cervicovaginal lavage female samples. 3 out of 7 women were reported they were monogamous and sexual contacts with gonococcal urithritis (Nudel et al., 2018). However, whether the patients had taken antibiotics was not reported. Physiological conditions (e.g. days since menstrual cycle in female samples) were not recorded for the analysis. PCR identified that all samples are infected by *GC*, within 7 of them coinfected with *Ct*. All samples were shifted and sequenced by the Genco lab (Nudel et al., 2018). Here, the purpose was to analyze the complex patient metatranscriptome and to find out what role did *Ct* play in the host natural environment of microbial coinfection. In addition, the responses of human host and *GC* with the presence of *Ct* were also interesting to analyze. The analysis started from the sequenced sample reads in FASTQ format. The naming convention for each sample includes the sample number (e.g. "S01"), barcode (e.g."03A2"), *Ct* positive ("cton") or not ("ctoff") and the sex (male: "m", female: "f").

3.2.1 Quality control

The metatranscriptome could not fit for all assessments designed for single transcriptome. For example, "GC content" should not assess reads from multiple organisms. Here, two indexes were used to remove unqualified samples: "per tile sequencing quality", used for filtering out contaminated samples; and the "sequence length distribution", aimed at eliminating samples with over short reads. Figure 3.2.1 shows "S11_06HT_cton_f" should be excluded for the low precision of the reads. Sample "S13_05NK_ctoff_f" did not present bad quality in Figure 3.2.2 but was still omitted for the length distribution as its short reads with around only 20bp would cause bias compared to the 100bp of other samples. Sample "S12_8369_ctoff_f" is not qualified enough (Figure 3.2.1) and with a 72bp leak in the length distribution. In short, 10 sample dataset, including 6 male samples (from S01 to S06) and 4 female samples (from S07 to S10) were involved for further analysis.
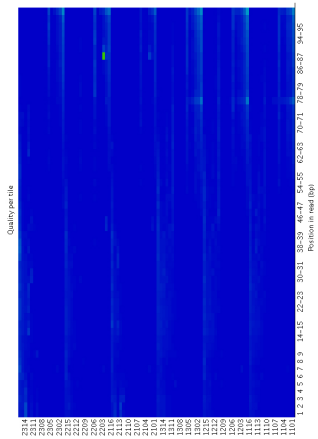
(a) S01_03A2_cton_m
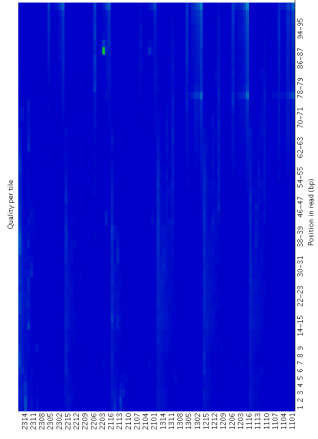
(b) S02_0745_cton_m
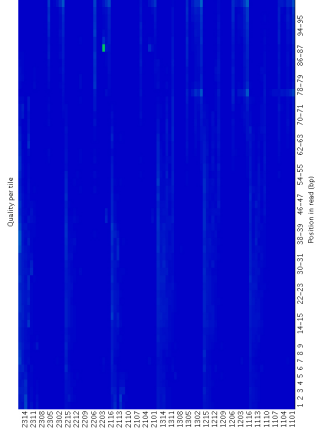
(c) S03_07KJ_ctoff_m
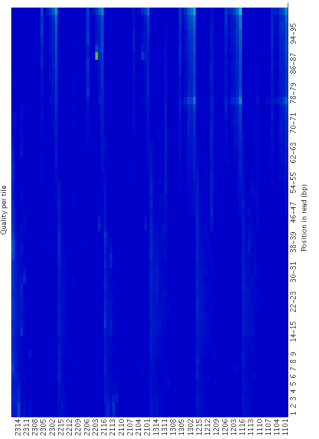
(d) S04_O365_cton_m
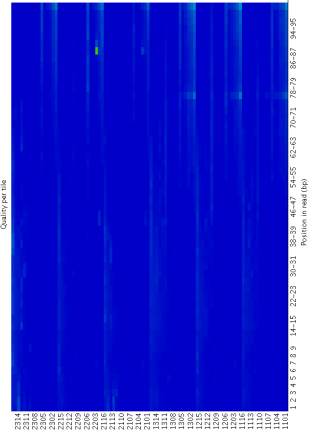
(e) S05_07D4_ctoff_m
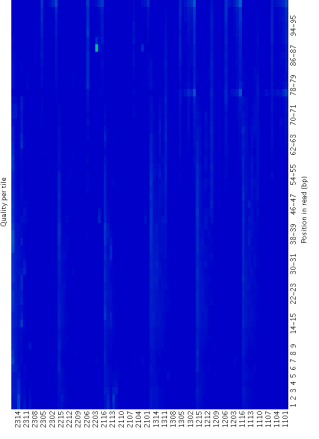
(f) S06_078H_ctoff_m
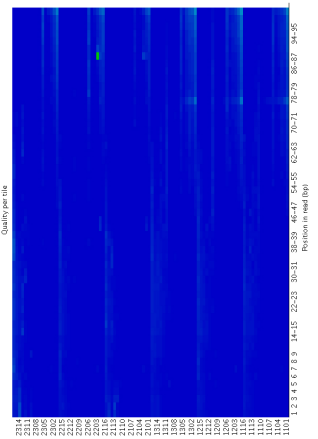
(g) S07_073F_cton_f
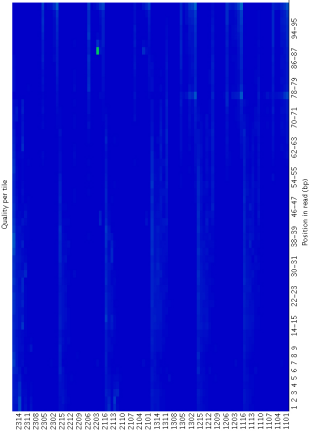
(h) S08_039K_cton_f
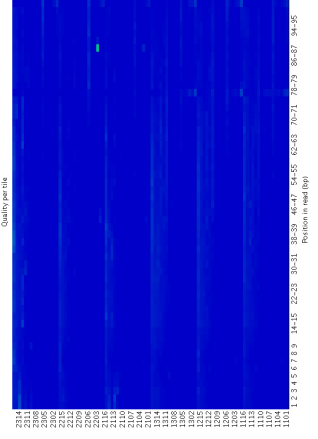
(i) S09_078N_ctoff_f

Figure 3.2.1: Average quality for each tile. The sequencing quality per tile is shown in the heatmap. The X-axis is the read position, and Y-axis is the tile. Quality is shown in the range of blue to red. Blue colour indicates the read in this tile has a good quality, and read colour presents the opposite.

(j) S10_07JW_cton_f

(k) S11_06HT_cton_f

(l) S12_8369_ctoff_f

(m) S13_05NK_ctoff_f
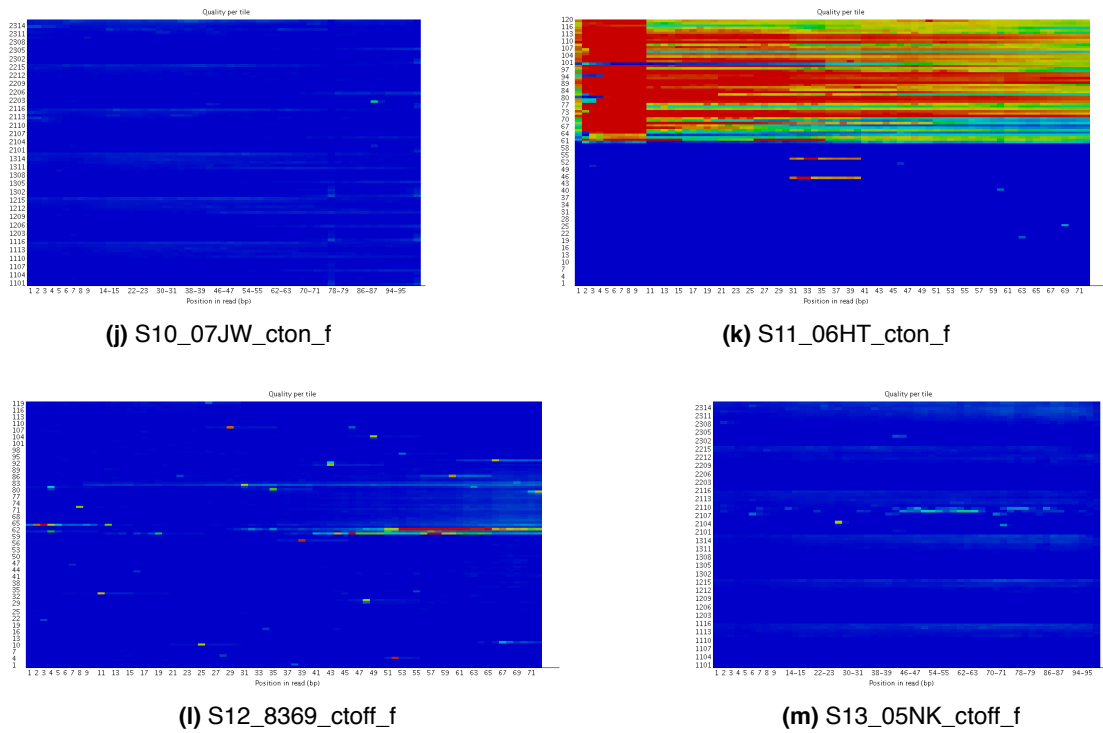
Figure 3.2.1: Average quality for each tile (continue)

(a) S01_03A2_cton_m

(b) S02_0745_cton_m

(c) S03_07KJ_ctoff_m

(d) S04_O365_cton_m
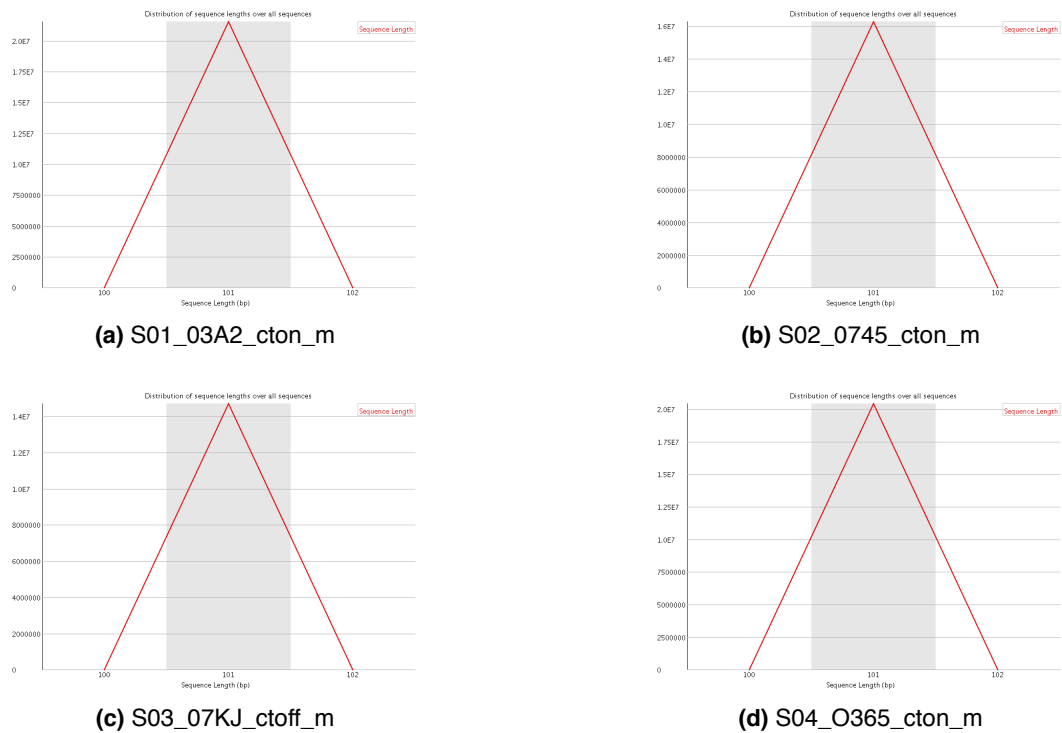
Figure 3.2.2: Distribution of sequence length over all sequences. The graph presents the distribution of fragment length in each sample. In many cases, only the peak of the distribution will be shown by FastQC (Babraham_Bioinformatics).

47

(e) S05_07D4_ctoff_m

(f) S06_078H_ctoff_m

(g) S07_073F_cton_f

(h) S08_039K_cton_f

(i) S09_078N_ctoff_f

(j) S10_07JW_cton_f

(k) S11_06HT_cton_f

(l) S12_8369_ctoff_f

(m) S13_05NK_ctoff_f
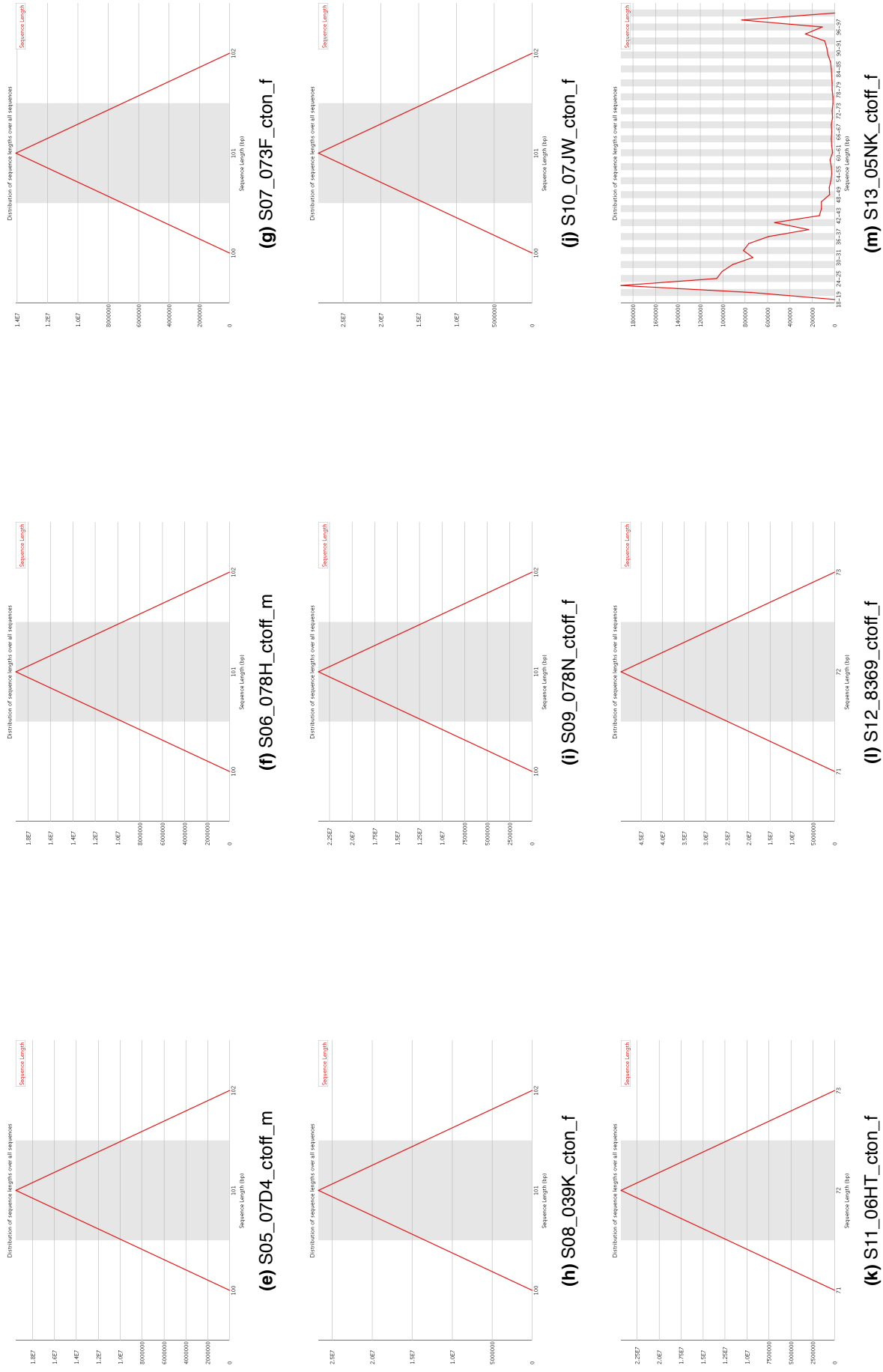
Figure 3.2.2: Distribution of sequence length over all sequences (continue)

3.2.2 Parallel Read Mapping

To quantify gene expression, all reads in each patient sample were aligned to target reference genomes: *Homo sapiens* GRCh38.p11 (Consortium et al., 2001, McLaren et al., 2010, Zerbino et al., 2017), *Neisseria gonorrhoeae FA 1090* (Lewis et al.) and *Chlamydia trachomatis D/UW-3/CX* (Stephens et al., 1998) (Table 3.2.1 and Figure 3.2.3). The amount of reads in each sample shows sex-specific differences. Generally, female samples (n=4) contain more reads sequenced than that of male samples (n=6). Female samples also present more fluctuations than male samples. "S12_8369_ctoff_f" holds the most reads, with a number of 49,619,138 which is 3 times more than that in "S07_073F_cton_f", with the number of 14,054,498, and is almost 5 times of the "S13_05NK_ctoff_f". The target reference stains allowed the mapping of about 90% of the reads in the male sample. However, there are more mixed infections in female samples, hence it only takes 11.09% in "S09_078N_ctoff_f", and even less in "S10_07JW_cton_f" (4.05%). Sample "S13_05NK_ctoff_f" has 93.01% reads mapped for the unqualified shorter reads. It also suggests the importance of quality control of the sequenced reads, for reducing deviation by removing unqualified dataset (from S10 to S13). Other organisms are more dominant in female samples. Reads from human are the most dominant in each sample among the three target organisms, and aligned reads in GC present sex-specific difference. In each female sample, the mapped GC reads occupy less than 1% of the total reads. This indicates a more complex species composition in female samples than in male samples. It may due to the different complexity between cervicovaginal niche and urethral environment. The sex-specific feature should be taken into consideration in further statistical analysis. Due to the lower microbial variation and better quality, male samples were specifically chosen for statistical analysis. For female samples, only 4 datasets were available after quality control. These were not enough for standard statistical analysis with the necessity of at least 3 replicates in each group.

Only a few reads from Ct were able to be aligned to Ct's reference genome. After quality control, samples with the most mapped reads from Ct were "S04_O365_cton_m" and "S10_07JW_cton_f", with the number of 5540 and 5598 reads respectively. Ct's reads in each sample were all less than 0.1%, which was impossible to get its whole transcriptome and most of the coding gene expressions. It is probably caused by sample complexity, methods of sequence preparation and sequencing depth.

Table 3.2.1: Overview of reads mapping statistics

Sample no.	Reads	Alignment*							
		HS		*GC*		*CT*		*sum*	
		reads	rate(%)	reads	rate(%)	reads	rate(%)	reads	rate(%)
S01_03A2_cton_m	21,533,437	17,410,269	80.85	1,742,960	8.09	176	0	19,153,405	88.95
S02_0745_cton_m	16,269,358	11,388,122	70	3,273,431	20.12	13	0	14,661,566	90.12
S03_07KJ_ctoff_m	14,694,597	12,877,494	87.63	243,932	1.66	2	0	13,121,428	89.29
S04_O365_cton_m	20,409,820	17,296,161	84.74	1,250,955	6.13	5540	0.03	18,552,656	90.9
S05_07D4_ctoff_m	19,503,923	16,096,437	82.53	1,368,462	7.02	2	0	17,464,901	89.55
S06_078H_ctoff_m	19,079,707	16,447,703	86.21	624,311	3.27	0	0	17,072,014	89.48
S07_073F_cton_f	14,054,498	12,600,514	89.65	26,976	0.19	893	0.01	12,628,383	89.85
S08_039K_cton_f	26,673,653	20,000,156	74.98	165,251	0.62	565	0	20,165,972	75.6
S09_078N_ctoff_f	23,719,512	2,615,941	11.03	13,778	0.06	13	0	2,629,732	11.09
S10_07JW_cton_f	28,248,413	1,123,299	3.98	14,494	0.05	5598	0.02	1,143,391	4.05
S11_06HT_cton_f	24,323,815	2,469,016	10.15	172,300	0.71	616	0	2,641,932	10.86
S12_8369_ctoff_f	49,619,138	4,254,165	8.57	91,438	0.19	414	0	4,346,017	8.76
S13_05NK_ctoff_f	11,691,682	10,298,046	88.08	573,813	4.91	2792	0.02	10,874,651	93.01

*: *HS*: *Homo sapiens*; *GC*: *Neisseria gonorrhoeae*; *CT*: *Chlamydia trachomatis*.
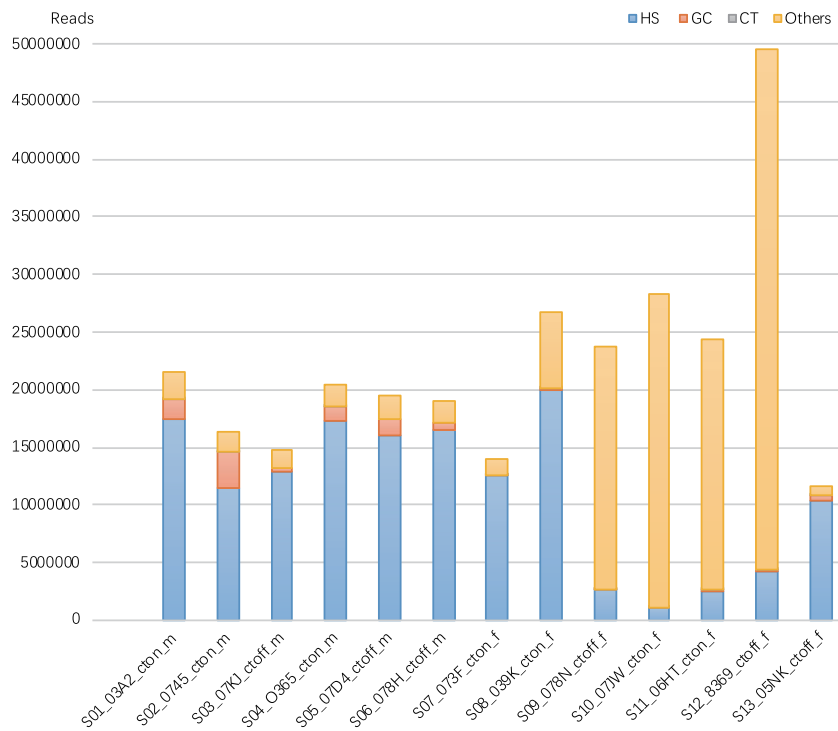
Figure 3.2.3: Comparison of mapping statistics among 13 samples. *HS*: *Homo sapiens*; *GC*: *Neisseria gonorrhoeae*; *CT*: *Chlamydia trachomatis*

3.2.3   Microbial Census

The presence of mixed infection in both male and female samples brings the requirement of the microbial census. Figure 3.2.4 shows the constitution of the human host cell, bacteria, fungi, virus, and protists, which could be detected by the database in One Codex platform for each sample. Male samples (from S01 to S06) were primarily made up by human host and *GC*, with about 90% of total reads together. Female samples displayed the complexity of the abundance. In sample "S07_073F_cton_f", reads from human host occupied nearly 90% of the total reads in the mapping (Table 3.2.1); Other female samples presented at least 10 more microbial species. Compared to the well-known vaginal microbiota (van de Wijgert et al., 2014) *Gardnerella vaginalis* and *Prevotella* in female samples, *Sneathia* was predominant in "S08_039K_cton_f" and even shared more than 50% of the total aligned reads in "S09_078N_ctoff_f" and "S10_07JW_cton_f". The species of *Sneathia* formerly belong to *Leptotrichis* (Collins et al., 2001), and was regarded as an opportunistic pathogen in the genitourinary tract, however very little is known about its biological potential (Harwich et al., 2012) currently. *Sneathia* is reported to be able to ferment glycogen, maltose and glucose but not mannose, mucin and starch according to the genome sequence (Collins et al., 2001). Many potential STI pathogens like *Sneathia* are expected more studies for the pathogenesis and infectious mechanisms.

Reads of fungi, virus and protists were also detected and shown in Figure 3.2.4.c. *Herpes alphaherpesvirus 2* (HSV-2) infected "S08_039K_cton_f" with around 3000 mapped reads. "S02_0745_cton_m" also exhibited HSV-2 infection but not dominant. "S03_07KJ_ctoff_m" and "S07_073F_cton_f" shared similar patterns of infections by *Malassezia globosa* and *Malassezia restricta*. *Puccinia striiformis*, originally a plant pathogen, was identified in "S07_073F_cton_f". Other microbes, detected less than 100 reads, were not presented dominantly or concluded in "Other" catalogue. These included: *Valsa mali*, *Ramularia collo-cygni*, *Periglandula ipomoeae*, *Coccidioides immitis*, *Alternaria alternata*, *Cyberlindnera jadinii*, *Cladonia metacorallifera*, and *Pseudogymnoascus destructans* among Fungi; *Dyolambdapapillomavirus 1*, *Alphapapillomavirus 9*, *Choristoneura occidentalis granulovirus*, *Chimpanzee alpha-1 herpesvirus*, and *Staphylococcus phage 92* in viruses; and *Toxoplasma gondii* and *Paramecium tetraurelia* in protists. It is hard to summarize whether these microbes were really presented from the sample or not when only rely on the sequence mapping with their low reads appearance. However, it provides the possibility of horizontal gene transfer even among different domains in STIs. The complexity of microbial abundance presented in human samples suggests the increased risk of potential coinfections among diverse microorganisms in a real microbiota.

What should not be ignored is the undetected reads in each sample. It could because of no

reads mapping for the limited collections of the reference database, or mapping reads with the lower abundance which is not enough for precise species identification. Other unknown microorganisms especially extracellular bacteria were also not involved in the study due to the library preparation method (e.g. lavage).

(a) Host and microbes in each sample

(b) Presence of bacteria in each sample (filtered by reads > 1%)

(c) Presence of fungi, virus, and protists in each sample (filtered by reads >5%)
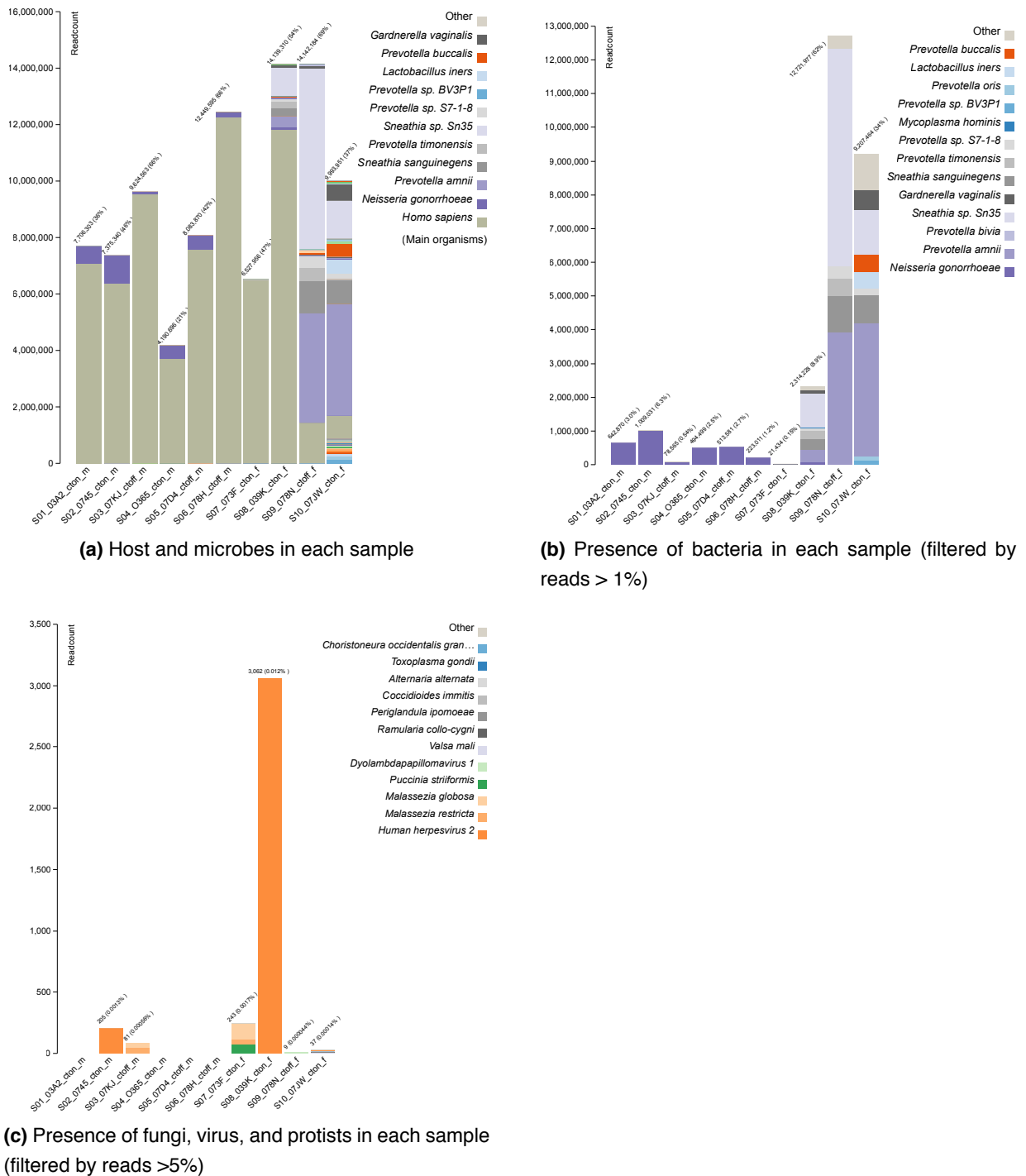
Figure 3.2.4: Microbial census in the STD patient samples

3.2.4 Human transcriptome profiling and differential expression analysis

The transcriptome of *HS* and *GC* were successfully processed, however, *Ct*'s transcriptome was impossible to obtain based on inadequate reads of *Ct* in current samples, sequenced from the *Neisseria* infected patients. Two conditions are taken into consideration to extract differentially expressed genes (DEGs): metatranscriptomic sample groups of male (n=6) vs. female (n=4), and *Ct* positive (Ct^+, n=6) vs. *Ct*-negative (control, n=4). To avoid sample bias caused by a sex-specific population and to make comparisons in the different normalized dataset, here the analysis in Ct^+ group versus control group was designed by using all the sample populations and using only male subjects. Female datasets served three samples in Ct^* group, but only one in the control group. So the female group would not be discussed individually here. Three different pipelines for transcriptome analysis (as mentioned in Chapter 2.2.2) were applied for comparison. The main purpose of this study is to look for the host and *GC*'s responses in Ct^+ group based on differential expression analysis, in order to reveal the influences caused by *Ct* infection.
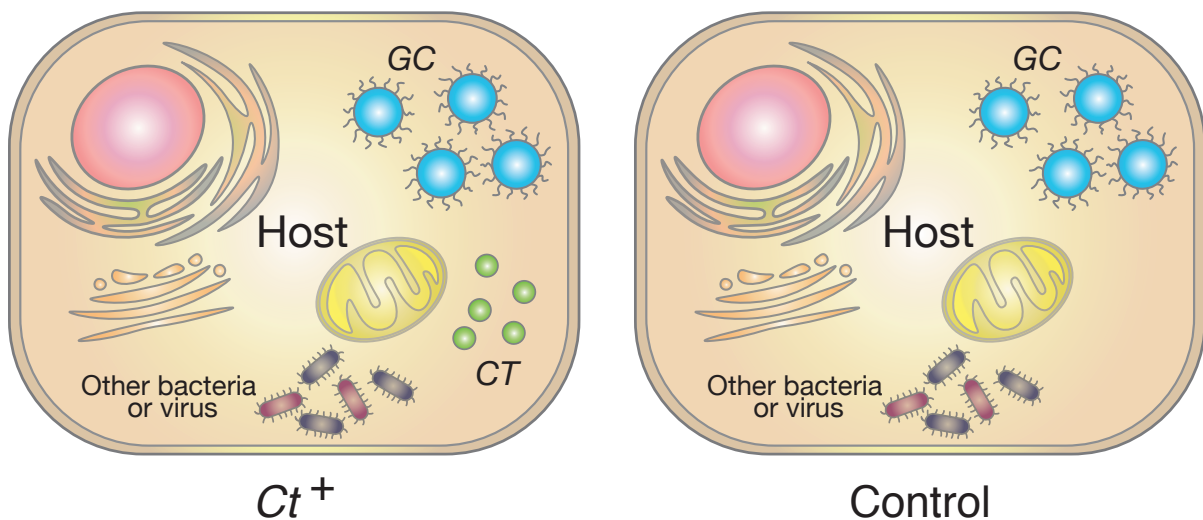
Figure 3.2.5: Ct^+ group vs. control group. *CT*: *Chlamydia trachomatis*, *GC*: *Neisseria gonorrhoeae* (source: own figure).

3.2.4.1 Overview of the human transcriptome

The eukaryotic transcriptome is more complicated than prokaryotic transcriptome, because the eukaryotic genome has a comparatively large size and contains both coding sequences and large amounts of noncoding sequences. In the human genome, only 2% is encoding sequences for proteins, the other 98% are noncoding sequences, including introns, regions for noncoding or regulatory RNAs, and DNA repeats (Consortium et al., 2001, McLaren

et al., 2010, Zerbino et al., 2017). The noncoding region drives variation of the transcripts in RNA splicing. Thus, the expressions of genes and transcripts are different in eukaryotes. It is disproportionately and statistically difficult to analyze human transcriptome.

The human transcriptome in all samples are shown in Figure 3.2.6(a), which suggests the gene expressions in different samples are uncomparable without normalization. Generally, principal component analysis (PCA) would be achieved before the differential expression analysis for the clear demonstration of whether the variables were possibly correlated in the condition of *Ct* infection. Figure 3.2.7 shows the PCA plot of gene and transcript expression in different sample groups by two dimensions. A read count matrix in DESeq2 was generated for PCA and presented in Figure 3.2.7.A-D. Another counting matrix based on FPKM in Ballgown package were used for Figure 3.2.7 (e and f). The community of "S03_07KJ_ctoff_m" and "S06_078H_ctoff_m" are highly similar. No significant scatter clusters of both gene and transcript expressions when considering both male and female samples. However, based on the condition of *Ct* infection presented in Figure 3.2.7 (b, d and e), a three dimensional PCA was applied (Figure 3.2.6b) and clearly showed samples in Ct^+ group were clustered interior.

(a) human gene expression in all samples.

(b) 3D PCA plot of transcript expression in male samples
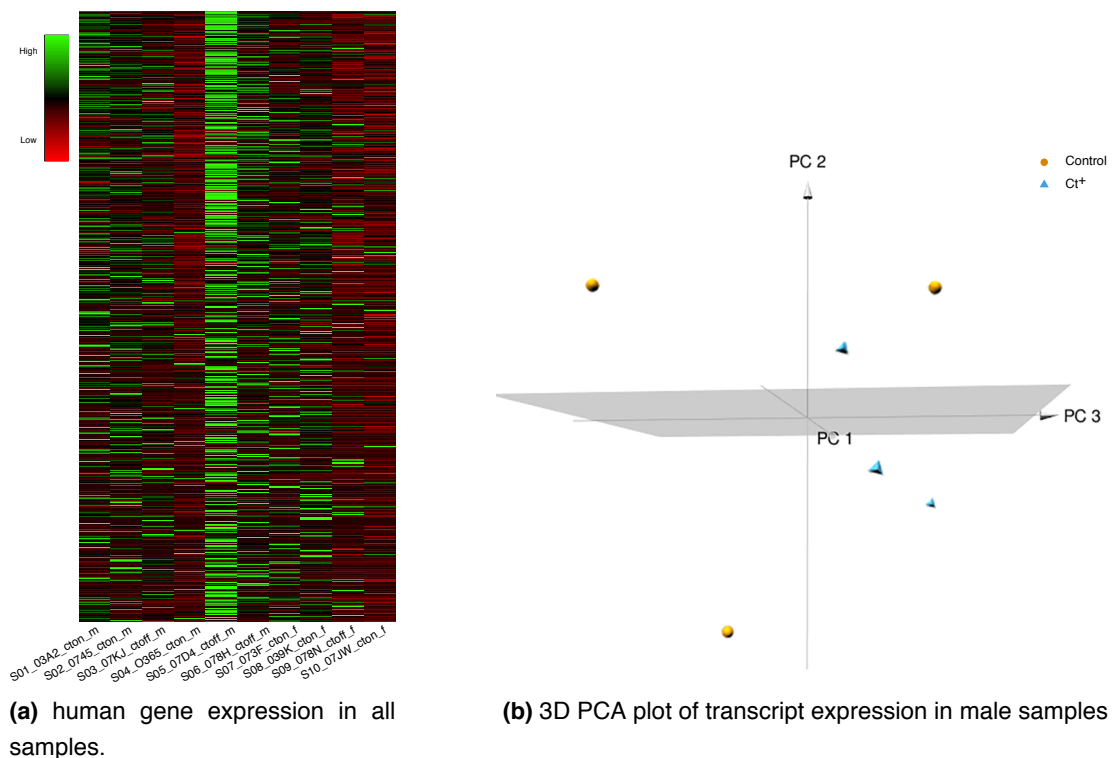
Figure 3.2.6: Human gene expression and 3D PCA analysis in male samples.
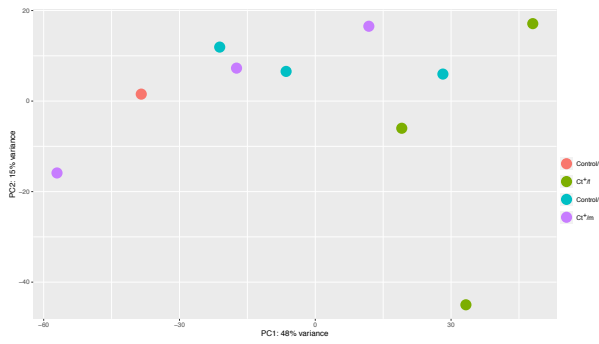
(a) Gene expression in all samples by DESeq2
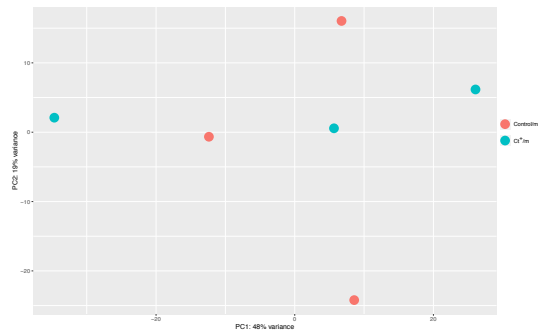
(b) Gene expression in male samples by DESeq2
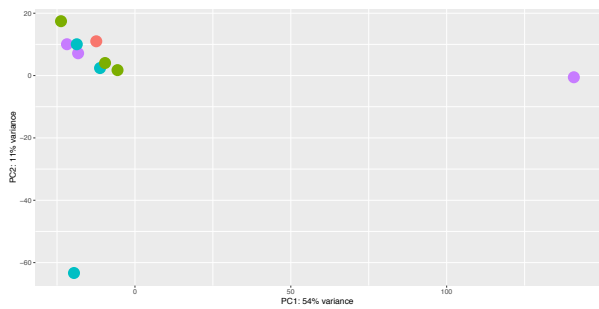
(c) Transcript expression in all samples by DESeq2
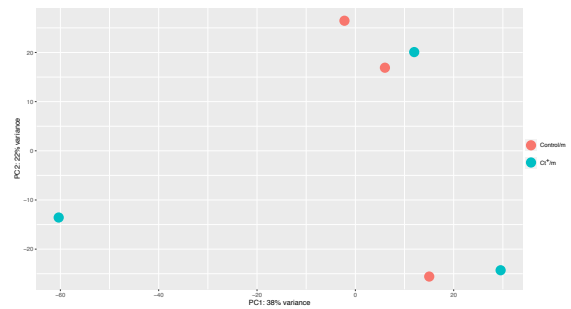
(d) Transcript expression in male samples by DE-Seq2
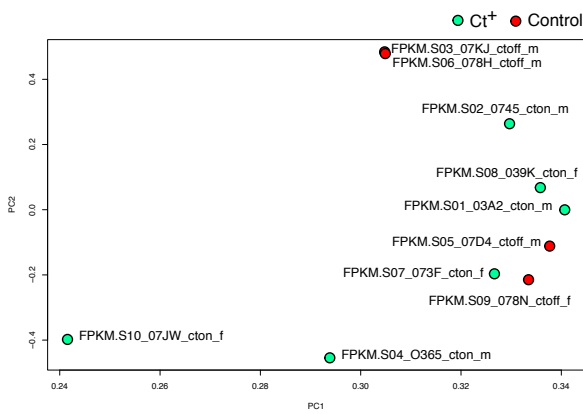
(e) Gene expression in all samples by Ballgown
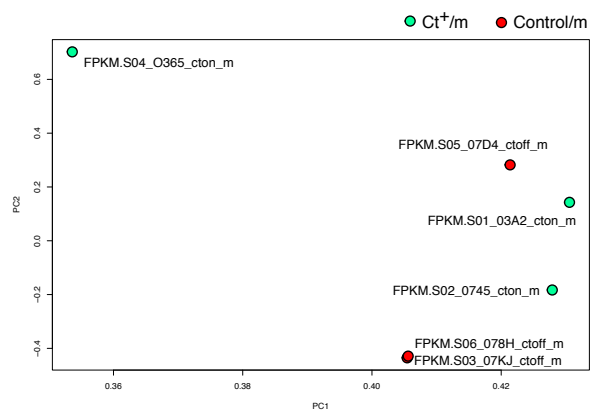
(f) Gene expression in male samples by Ballgown

Figure 3.2.7: 2D PCA plot of human gene and transcript expression

3.2.4.2 Host differential expressions in all samples

In order to study the influence caused by *Ct* infection in the host cell, differential expression analysis was applied to present the differences in both gene and transcript expressions between Ct^+ group and control group. Although no significant PCA clusters were separated by using both male and female samples (Figure 3.2.7), differentially expressed genes (DEGs) and differentially expressed transcripts (DETs) were extracted as comparisons. There were in total 10 samples, with 6 samples in Ct^+ group (including 3 males and 3 females), and 4 samples in the control group (including 3 males and only 1 female). The MA plot (Figure 3.2.8) shows the variation of genes and transcripts in using both male and female samples by DESeq2. The DEGs and DETs were marked in red under the threshold of both *p*-value and adjusted *p*-value (*padj*) below 0.05 for statistical significance.

(a) Gene expression in all samples by DESeq2   **(b)** Transcript expression in male samples by DESeq2
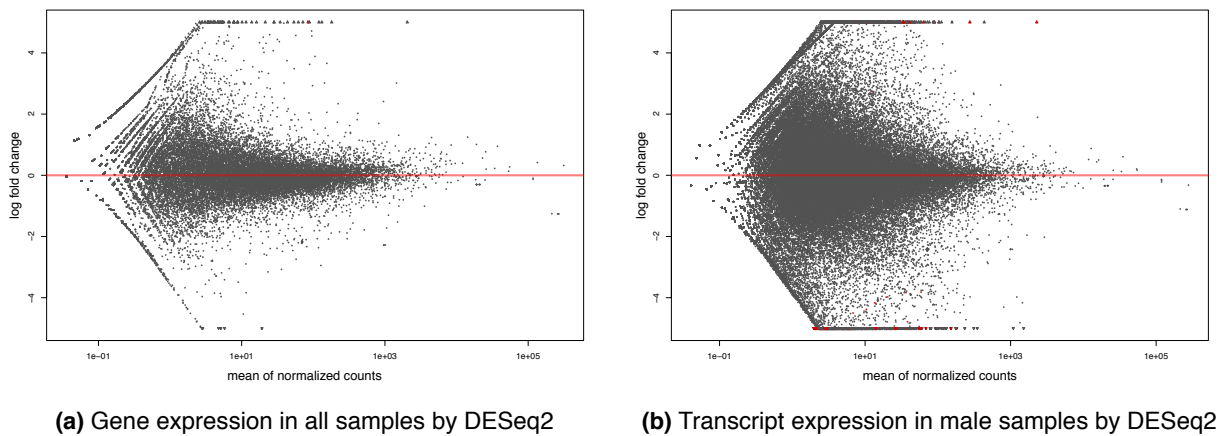
Figure 3.2.8: MA plot of gene and transcript expression in all samples. The statistically significant expressions (both *p* and *padj* $\leq$ 0.05) are marked in red.

The DEGs and DETs in details are shown in Table 3.2.2. Under the threshold of both *p*-value < 0.05 and *padj* < 0.05, 1 DEG and 39 DETs (with 31 up-regulation and 8 down-regulation) were extracted. The only DEG presented was a noncoding RNA gene, *AC019349.1*, described as a novel transcript located on *HS* chromosome 17. Gene *AC019349.1* highly downregulated by almost 22 times log2 fold change in gene expression level with the presence of *Ct*. However, no further information could support the understanding of *AC019349.1*'s function. The 39 DETs were associated with the molecular and biological function based on the gene ontology analysis. In the host cells, the presence of *Ct* influenced more on molecular activities of binding (GO: 0005488, 43%) and catalysis (GO: 0003824, 35%) (Figure 3.2.9 a). The host response was more focused on the metabolic process (25%) and cellular process (18%) than the general immune system process (5%) (Figure 3.2.9 b). The cellular component organization or biogenesis (GO: 00071840) and

response to the stimulus (GO: 0050896) presented the second dominance with 6 DETs (15%) in each of them. Most of the DETs were transcriptional factors (PC00218, 18%) and enzyme modulators (PC00095, 18%), some of them were hydrolase (PC00121, 12%) and associated with the nucleic acid binding (PC00171, 15%). In short, *Ct* infection influenced more adhesion and binding molecules, transcriptional factors, and metabolic-related proteins in the host cell.

Table 3.2.2: Differential expressions with statistical significance in all samples by DESeq2

Type	Name	Ensemble ID	Annotation	log2FC*	pvalue	padj
Gene	AC019349.1	ENSG00000229732.1	novel transcript	-21.98369702	8.18E-17	2.40E-12
	ADGRE2-201	ENST00000315576.7	adhesion G protein-coupled receptor E2	25.06218308	4.12E-16	1.15E-11
	MYADM-203	ENST00000391769.2	myeloid associated differentiation marker	22.93680624	1.31E-14	1.22E-10
	AHR-202	ENST00000463496.1	aryl hydrocarbon receptor	22.78661858	2.70E-13	1.83E-09
	SCARB2-215	ENST00000639145.1	scavenger receptor class B member 2	22.69911033	3.33E-13	1.83E-09
	PKM-205	ENST00000561609.5	pyruvate kinase M1/2	22.63064764	3.93E-13	1.83E-09
	GZF1-202	ENST00000377051.2	GDNF inducible zinc finger protein 1	22.40224251	6.76E-13	2.71E-09
	FAM91A1-205	ENST00000519721.5	family with sequence similarity 91 member A1	22.28575025	1.21E-14	1.22E-10
	USP32-211	ENST00000590133.5	ubiquitin specific peptidase 32	22.1205508	1.31E-12	4.60E-09
	CALR-202	ENST00000586760.1	calreticulin	21.95505752	1.93E-12	6.02E-09
	LGALS8-203	ENST00000341872.10	galectin 8	21.82650619	2.61E-12	6.31E-09
	PRPF4B-201	ENST00000337659.10	pre-mRNA processing factor 4B	21.82028444	2.64E-12	6.31E-09
	SEPT9-203	ENST00000427177.5	septin 9	21.81085513	2.70E-12	6.31E-09
	GRINA-204	ENST00000527194.5	glutamate ionotropic receptor NMDA type subunit associated protein 1	21.77143823	2.96E-12	6.38E-09
	ADD1-207	ENST00000446856.5	adducin 1	21.6562891	3.86E-12	7.73E-09
	WDR74-204	ENST00000525239.5	WD repeat domain 74	11.37781309	1.97E-06	0.003073402
	ENPP2-201	ENST00000075322.10	ectonucleotide pyrophosphatase/phosphodiesterase 2	8.883085729	1.64E-05	0.019098115
	PRPF8-201	ENST00000304992.10	pre-mRNA processing factor 8	8.743363721	3.97E-05	0.037114096
	SHOC2-201	ENST00000265277.9	SHOC2, leucine rich repeat scaffold protein	8.710474728	6.89E-05	0.049480784
	CDK12-202	ENST00000447079.4	cyclin dependent kinase 12	8.613727618	4.26E-06	0.006189968
Transcript	HLA-DQA1-201	ENST00000343139.9	major histocompatibility complex, class II, DQ alpha 1	8.027606229	4.79E-05	0.041903556
	ARPP19-208	ENST00000566423.5	cAMP regulated phosphoprotein 19	7.81226898	6.87E-05	0.049480784
	EML4-202	ENST00000401738.3	echinoderm microtubule associated protein like 4	7.761851933	6.81E-05	0.049480784
	ZNF493-203	ENST00000392288.6	zinc finger protein 493	7.311045236	5.61E-07	0.000981611
	XRRA1-203	ENST00000431210.2	X-ray radiation resistance associated 1	7.266042518	1.16E-06	0.001913678
	GBP3-202	ENST00000370481.8	guanylate binding protein 3	7.078537699	2.77E-05	0.028727396
	USP47-202	ENST00000339865.9	ubiquitin specific peptidase 47	7.047231677	1.60E-05	0.019098115
	ASAH1-253	ENST00000637609.1	N-acylsphingosine amidohydrolase 1	6.587620852	3.66E-05	0.035557997
	ZEB2-236	ENST00000636471.1	zinc finger E-box binding homeobox 2	6.492282061	9.40E-06	0.012547301
	RASA1-202	ENST00000456692.6	RAS p21 protein activator 1	5.741760225	1.09E-05	0.013939435
	PHF3-210	ENST00000509876.5	PHD finger protein 3	5.689306606	4.42E-06	0.006189968
	NCKAP5L-201	ENST00000335999.6	NCK associated protein 5 like	2.599292502	5.00E-05	0.041905186
	SMAP2-206	ENST00000614549.4	small ArfGAP2	-3.791536715	2.24E-05	0.025124077
	RUNX1-207	ENST00000437180.5	runt related transcription factor 1	-4.252718261	6.46E-05	0.049480784
	BAG6-249	ENST00000441793.5	BCL2 associated athanogene 6	-4.331085604	4.71E-05	0.041903556
	RBPJ-225	ENST00000514380.5	recombination signal binding protein for immunoglobulin kappa J region	-4.628466732	5.08E-05	0.041905186
	TRAPPC3-203	ENST00000373163.5	trafficking protein particle complex 3	-5.86916212	5.24E-05	0.041969968
	UBE4A-202	ENST00000431736.6	ubiquitination factor E4A	-7.894866632	3.68E-05	0.035557997
	EIF4G1-216	ENST00000427845.5	eukaryotic translation initiation factor 4 gamma 1	-8.048987767	2.86E-07	0.000534669
	ALDH1A3-202	ENST00000346623.6	aldehyde dehydrogenase 1	-9.457467945	2.46E-05	0.026549094

Under the threshold of both *p* and *padj* $\leq$ 0.05. *: log2 fold change.

(a) Molecular function **(b)** Biological process **(c)** Protein class
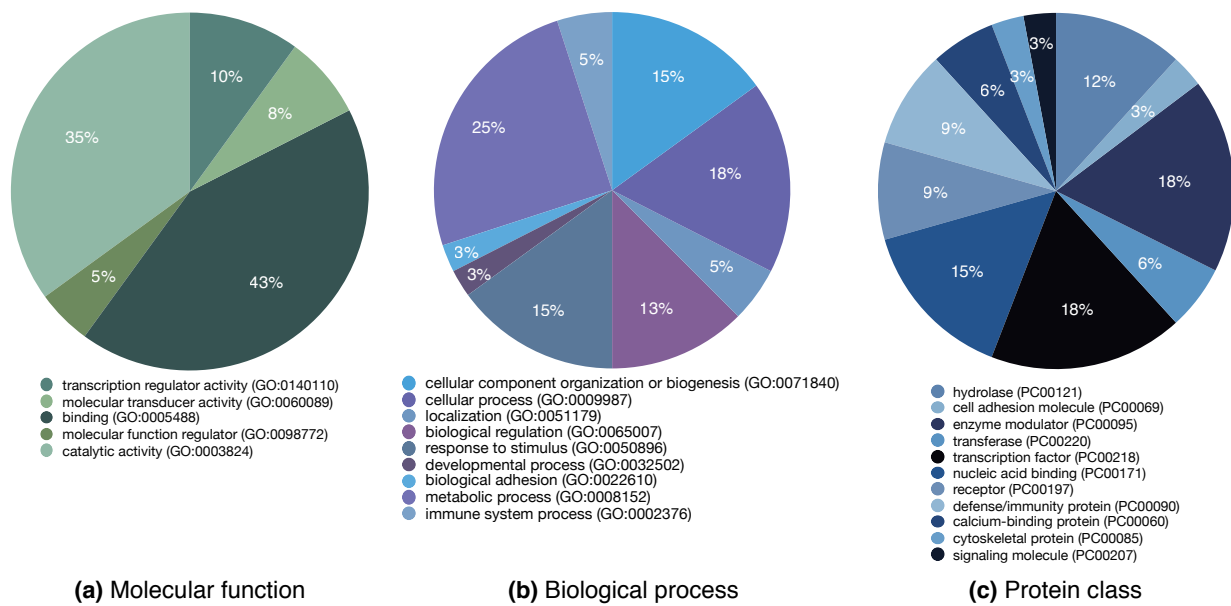
Figure 3.2.9: Gene ontology (GO) of differentially expressed genes in all samples by DESeq2

Potential interactions among these DETs in Table 3.2.2 were analyzed by protein-protein interaction network (PPI) from the String database (Figure 3.2.10). The PPI analysis showed *Ct* infection affected on host DNA repair, ubiquitin system, and MHC class II immune response.

3.2.4.2.1 DNA repair and ubiquitin system

The prior interactions were clustered in protein ubiquitin and deubiquitin system. USP47-202 and USP32-211 are protein-coding transcripts in the ubiquitin-specific proteases USP47 and USP32 respectively. Both of them were highly upregulated in Ct^+ group. USP47 is functionally involved in DNA base excision repair (BER) process and deubiquitinate the key enzyme DNA polymerase II (Polβ) (Parsons et al., 2011), which induced only by widespread and persistent DNA damage (Žgur-Bertok, 2013). Parsons et al. (2011) reported the knockdown of USP47 resulted in an upregulation of ubiquitylated Polβ, downregulation of Polβ, and the deficiency in BER afterwards, which causing DNA strand break. Interestingly, Gulve et al. (2018) reported *Ct* induced the downregulation of Polβ and the reducing of BER. Besides, silencing USP47 was reported to suppress cell growth and survival by promoting the accumulation of Cdc25A (Peschiaroli et al., 2010). USP32, a ubiquitin protease whose silencing decreased cell proliferation, was observed overexpression in breast cancers (Akhavantabasi et al., 2010). The E3/E4 ubiquitin ligase UBE4A, participating in the optimal repair of DNA double-strand break by acclimating ubiquitylation (Baranes-Bachar et al., 2018), decreased the expression of transcript UBE4A-202. It helps to protect the
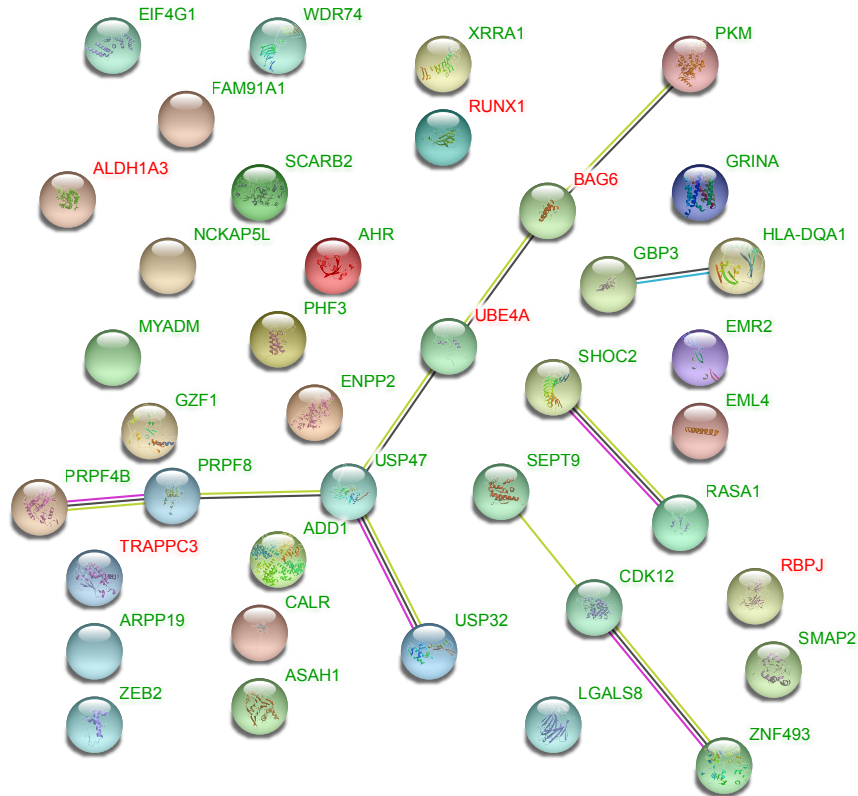
Figure 3.2.10: Potential protein interactions of differential expressions in all samples. Upregulated genes are marked in green, and the downregulated are shown in red.

cell in stress stimulus and to suppress cancer expansion (Sun et al., 2017). Transcripts from two pre-mRNA processing factor, PRPF8 and PRPF4B, augmented the expression with *Ct*'s company. PRPF8 is highly conserved and plays as a core role of the spliceosome assembly and recycling (Grainger and Beggs, 2005). Knockdown of PRPF8 together with another ubiquitin-like protein 5 was reported to activate p53 (Allende-Vega et al., 2013). PRPF4B is associated with cell differentiation and mitosis, and it even drives the metastasis in breast cancer cell migration (Liu et al., 2013, van de Water et al., 2018). In addition, a protein-coding transcript BAG6-219 was downregulated. However, BAG6-219 contains only 244 amino acid, much smaller than another protein-coding transcript BAG6-201 (with 1143 amino acid). PKM-205, a transcript from pyruvate kinase isoform M1/2 presented significant upregulation. In short, these interactions suggest the infection of *Ct* is associated to host DNA repair by ubiquitin system.

3.2.4.2.2 DNA damage and MHC class II immune response

Another significant response caused by *Ct* infection was the major histocompatibility complex (MHC) class II, exhibiting increased regulation of the surface receptor HLA-DQA1 and guanylate binding protein 3 (GBP3). The upregulations were also in a leucine-rich

repeat protein SHOC2 and Ras p21 protein activator RASA1, as well as the network linked with SEPT9, CDK12 and ZNF493. Generally, fewer DETs are downregulated in the experiments, including transcripts of EIF4G1 and ALDHA3 which displayed dominance besides UBE4A described above. EIF4G1, the eukaryotic initiation factor $4\gamma1$, was reported mediating and augmenting the translation of mRNAs involved in cell survival and DNA damage response in the breast cancer cell (Badura et al., 2012). Many factors exhibited host response of DNA damage, cell survival and apoptosis with the presence of *Ct*. However, the conclusion was still not clear for the statistical limitations based on the sample complexity as described before. Nevertheless, the analysis in this subsection could be taken as a comparison.

3.2.4.2.3 Differential expressions in other pipelines

The same dataset was analyzed by another software Ballgown. Though many DEGs and DETs were carried out by p-value ≤ 0.05, none of them was significantly different by the threshold q-value ≤ 0.05 (Table 3.2.3). In addition, though better performing in prokaryotic transcriptome analysis than that in eukaryotes, Cuffdiff was also tested and got no statistically significant DEGs and DETs (data not shown here).

Table 3.2.3: Differential expressions in all samples by Ballgown

Type	Gene name	Annotation	log2FC*	pval	qval
Gene	NUPL2	Nucleoporin like 2	-1.690789433	8.04E-05	0.644290496
	NBR2	Neighbor of BRCA1 lncRNA 2	-1.696289264	0.000166878	0.644290496
	PILRB	Paired immunoglobin-like type 2 receptor beta	-2.617291031	0.00017849	0.644290496
Transcript	EEF1D	Eukaryotic translation elongation factor 1 delta	2.177734656	4.49E-05	0.973620582
	RALGAPB	Ral GTPase activating protein non-catalytic beta subunit	-1.268323915	6.72E-05	0.973620582
	NUPL2	Nucleoporin like 2	-1.735027279	0.000117301	0.973620582
	NBR2	Neighbor of BRCA1 lncRNA 2	-1.669441039	0.000191667	0.973620582
	PILRB	Paired immunoglobin-like type 2 receptor beta	-2.543890431	0.000255409	0.973620582
	PPIG	Peptidylprolyl isomerase G	-1.164661369	0.000290951	0.973620582
	UBE2G2	Ubiquitin conjugating enzyme E2 G2	-1.704203113	0.000296502	0.973620582

DEG: differentially expressed genes; DET: differentially expressed transcripts. Top 10 expressions are listed under the threshold of $p \leq 0.05$. *: log2 fold change.

3.2.4.3 Differential expression in male samples

Male samples statistically presented better than by using all samples with bias caused by varied female samples (Figure 3.2.7). Here the same differential expression was applied but only using male samples, with the same condition: Ct^+ group ("S01_03A2_cton_m", "S02_0745_cton_m", and "S04_O365_cton_m") versus the control group ("S03_07KJ _ctoff_m", "S05_07D4_ctoff_m", and "S06_078H_ctoff_m"). Figure 3.2.11 shows the expression in male samples by MA plot, and the significantly differential expressions are marked in red.

(a) Gene expression in male samples by DESeq2

(b) Transcript expression in male samples by DESeq2
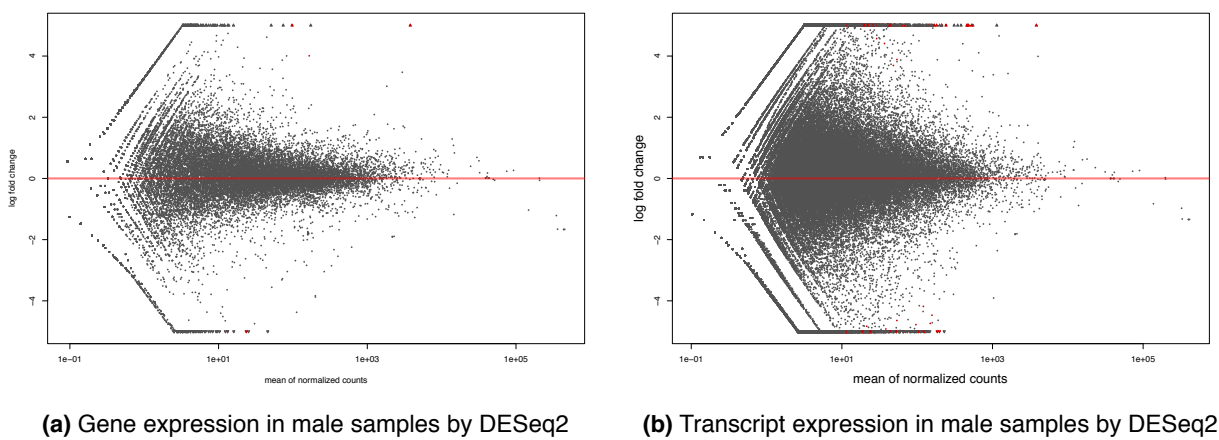
Figure 3.2.11: MA plot of gene and transcript expression in male samples. The statistically significant expressions (both *p* and *padj* $\leq$ 0.05) are marked in red.

(a) Molecular function

(b) Biological process

(c) Protein class
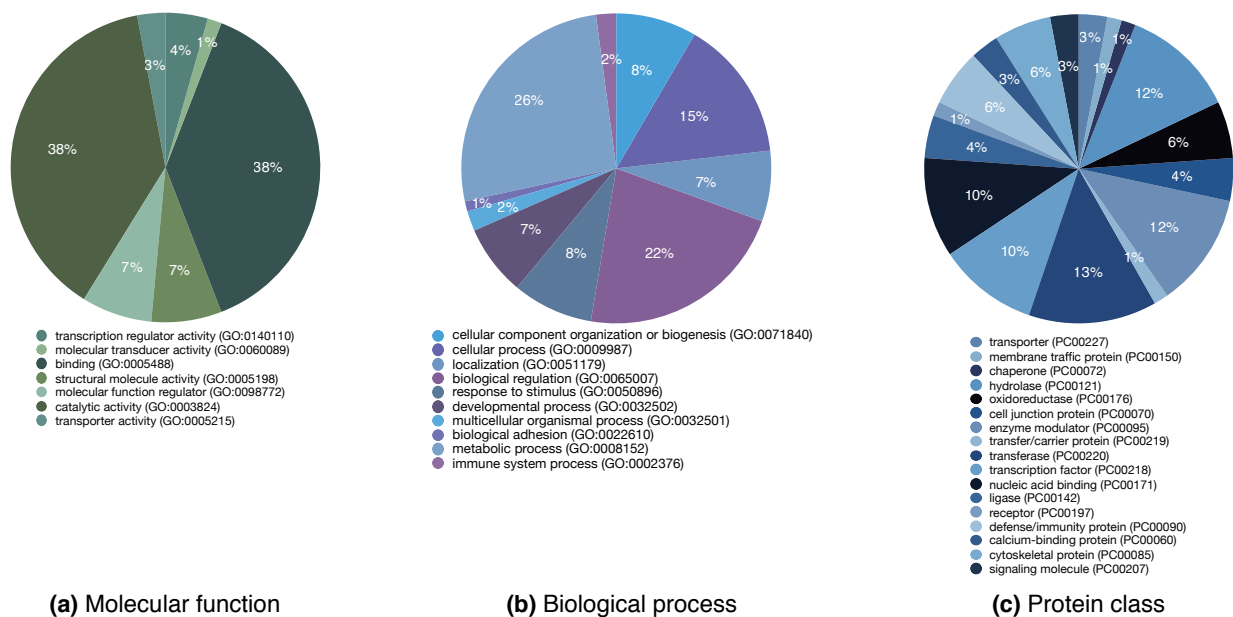
Figure 3.2.12: Gene ontology (GO) of differentially expressed human genes in male samples by DESeq2

The analysis focused on male samples extracted 4 DEGs and 77 DETs in total (Table 3.2.4). This number was twice of previous results by using all samples. Gene ontology of these differential expressions was summarized in Figure 3.2.12. Binding (GO:0005488) and catalytic activity (GO:0003824) were remained the most dominant clusters in Ct positive group, occupied 38% in each and was similar as the results of all-sample analysis. Structural molecule activity (GO:0005198) was not shown in all-sample analysis but among 7% of analyzed genes. The biological process aimed mainly on the metabolic process (GO:0022610), biological regulation (GO:0065007) and cellular process (GO:0009987) were 26%, 22% and 15% respectively. Similar to the appearance in all-sample analysis, protein class of these DETs were focused on hydrolase (PC00121, 12%), enzyme modulator (PC00095, 12%), transferase (PC00220, 13%), transcription factor (PC00218, 10%) and nucleic acid binding (PC00171, 10%).

Four DEGs come out in the male-sample analysis (Table 3.2.4). Three of them were highly expressed in both gene and transcript level (HLA-DQA1, HLA-DQB1 and WDR74) in Ct^+ group, and associated with DETs shown in the all-sample analysis. Additionally, GBP3, ZNF493, ENPP2 and EIF4G1 shared similar regulation and stood out in both all-sample and male-sample analyses (Figure 3.2.13).

Table 3.2.4: Differential expressions with statistical significance in male samples by DESeq2

Type	Name	Ensembl ID	Annotation	log2FC	pvalue	padj
Gene	HLA-DQB1	ENSG00000179344.16	major histocompatibility complex, class II, DQ beta 1	9.887177399	2.49E-13	7.11E-09
	WDR74	ENSG00000133316.15	WD repeat domain 74	8.668361839	6.01E-13	8.58E-09
	HLA-DQA1	ENSG00000196735.11	major histocompatibility complex, class II, DQ alpha 1	4.008661229	1.45E-09	1.38E-05
	TRIM6-TRIM34	ENSG00000258588.3	TRIM6-TRIM34 readthrough	-7.340112625	7.39E-06	0.052746454
Transcript	DDX3X-203	ENST00000457138.6	DEAD-box helicase 3 X-linked	12.38612505	9.69E-05	0.039641135
	WDR74-204	ENST00000525239.5	WD repeat domain 74	11.89116998	1.06E-06	0.001555192
	ADAR-201	ENST00000368471.7	adenosine deaminase, RNA specific	11.19594983	6.01E-05	0.029533483
	HUWE1-214	ENST00000612484.4	HECT, UBA and WWE domain containing 1, E3 ubiquitin protein ligase	11.07387352	3.15E-06	0.003716203
	PLEKHB2-201	ENST00000234115.10	pleckstrin homology domain containing B2	10.84571354	2.08E-10	6.12E-06
	SLC8A1-206	ENST00000406391.2	solute carrier family 8 member A1	10.46670872	2.03E-09	3.00E-05
	NBN-203	ENST00000409330.5	nibrin	10.36445223	9.75E-06	0.007761598
	PLAUR-210	ENST00000597107.1	plasminogen activator, urokinase receptor	10.08306163	9.18E-08	0.000225405
	ENPP2-201	ENST00000075322.10	ectonucleotide pyrophosphatase/phosphodiesterase 2	9.476972265	2.54E-08	8.31E-05
	DAPK1-209	ENST00000472284.5	death associated protein kinase 1	9.293829437	3.08E-08	9.08E-05
	PKN1-202	ENST00000342216.8	protein kinase N1	9.035007927	1.01E-07	0.00022944
	PPP6R3-205	ENST00000524845.5	protein phosphatase 6 regulatory subunit 3	8.797526217	6.87E-07	0.00112492
	AL049629.2-201	ENST00000534312.5	CD59 glycoprotein	8.779560873	1.85E-07	0.000363651
	HLA-DQA1-201	ENST00000343139.9	major histocompatibility complex, class II, DQ alpha 1	8.723030695	2.49E-07	0.000457956
	DCUN1D1-208	ENST00000632685.1	defective in cullin neddylation 1 domain containing 1	8.388342007	1.10E-05	0.007887513
	SLC4A2-202	ENST00000413384.6	solute carrier family 4 member 2	8.313706962	5.23E-06	0.005708148
	GBP3-202	ENST00000370481.8	guanylate binding protein 3	8.277627216	1.72E-07	0.000361881
	SH2D3A-201	ENST00000245908.10	SH2 domain containing 3A	8.150021634	1.77E-06	0.002487394
	EPG5-208	ENST00000590854.5	ectopic P-granules autophagy protein 5 homolog	8.125021811	3.08E-05	0.01711235
	HLA-DQB1-205	ENST00000434651.6	major histocompatibility complex, class II, DQ beta 1	8.099237239	5.43E-05	0.028085277
	ACSL5-205	ENST00000433418.5	acyl-CoA synthetase long chain family member 5	8.018422056	9.09E-06	0.007441678
	ACADVL-225	ENST00000581378.5	acyl-CoA dehydrogenase very long chain	7.948984287	3.89E-05	0.021208666
	MYO1G-207	ENST00000495831.5	myosin IG	7.9064581	5.70E-06	0.005727942
	SRSF2-205	ENST00000508921.7	serine and arginine rich splicing factor 2	7.812491331	6.98E-06	0.006431312
	MED15-204	ENST00000406969.5	mediator complex subunit 15	7.797260718	6.16E-05	0.029761719

Table 3.2.4 continued from previous page

Type	Name	Ensembl ID	Annotation	log2FC*	pvalue	padj
	TNRC6C-202	ENST00000335749.4	trinucleotide repeat containing 6C	7.759968179	6.74E-06	0.006407526
	SHTN1-202	ENST00000355371.8	shootin 1	7.647845488	9.91E-05	0.039991456
	HLA-DQB1-209	ENST00000487676.1	major histocompatibility complex, class II, DQ beta 1	7.616168059	6.49E-05	0.02990342
	USF2-201	ENST00000222305.7	upstream transcription factor 2, c-fos interacting	7.582111101	2.25E-05	0.014194351
	SPECC1L-204	ENST00000437398.5	sperm antigen with calponin homology and coiled-coil domains 1 like	7.55664426	2.60E-05	0.015613094
	EMB-204	ENST00000508934.5	embigin	7.53571038	2.26E-05	0.014194351
	ARSA-205	ENST00000453344.6	arylsulfatase A	7.456199175	6.30E-05	0.02990342
	ANKRD13D-216	ENST00000514166.5	ankyrin repeat domain 13D	7.306467555	3.06E-05	0.01711235
	ZNF493-203	ENST00000392288.6	zinc finger protein 493	7.144942005	7.24E-05	0.031459953
	ZRANB2-201	ENST00000254821.10	zinc finger RANBP2-type containing 2	7.127713873	7.26E-05	0.031459953
	PDXK-210	ENST00000468090.5	pyridoxal kinase	7.035039886	6.92E-05	0.031388209
	TFRC-202	ENST00000392396.7	transferrin receptor	7.027963835	1.66E-05	0.011643121
	MBOAT7-206	ENST00000437868.5	membrane bound O-acyltransferase domain containing 7	6.995644058	5.14E-05	0.027041812
	CPT1B-204	ENST00000405237.7	carnitine palmitoyltransferase 1B	6.829013243	0.000115863	0.044925839
Transcript	PHF20L1-220	ENST00000622263.4	PHD finger protein 20 like 1	6.776777093	5.83E-06	0.005727942
	MAP4-207	ENST00000426837.6	microtubule associated protein 4	6.577114591	1.08E-05	0.007887513
	ZNF407-207	ENST00000582337.5	zinc finger protein 407	5.151382025	5.66E-06	0.005727942
	RIF1-206	ENST00000453091.6	replication timing regulatory factor 1	5.015489578	0.000119314	0.045663337
	MCM3AP-207	ENST00000486937.5	minichromosome maintenance complex component 3 associated protein	4.571133619	8.43E-05	0.035484296
	ZNF638-219	ENST00000487638.5	zinc finger protein 638	3.691777639	0.000109416	0.042991845
	PLAUR-204	ENST00000593396.1	plasminogen activator, urokinase receptor	-4.180950635	2.23E-05	0.014194351
	TYMP-208	ENST00000487577.5	thymidine phosphorylase	-4.469264016	2.49E-06	0.003053623
	SMCHD1-211	ENST00000584897.5	structural maintenance of chromosomes flexible hinge domain containing 1	-4.642774533	5.91E-05	0.029524526
	RAP1GAP2-201	ENST00000254695.12	RAP1 GTPase activating protein 2	-4.777782784	3.72E-07	0.000645137
	PGS1-215	ENST00000592043.5	phosphatidylglycerophosphate synthase 1	-4.831080241	0.000109358	0.042991845
	ZYX-207	ENST00000457235.5	zyxin	-4.919750914	7.90E-05	0.033726357
	PPP1R21-203	ENST00000416913.5	protein phosphatase 1 regulatory subunit 21	-5.39169818	1.02E-05	0.007887513
	ZNF33A-202	ENST00000374618.7	zinc finger protein 33A	-6.018538423	6.46E-05	0.02990342
	CAP1-218	ENST00000479759.5	cyclase associated actin cytoskeleton regulatory protein 1	-6.178017189	2.07E-05	0.013866761
	ZNF721-201	ENST00000338977.5	zinc finger protein 721	-6.773011749	2.82E-05	0.016648468

Table 3.2.4 continued from previous page

Type	Name	Ensembl ID	Annotation	log2FC*	pvalue	padj
	SLC4A2-214	ENST00000485713.5	solute carrier family 4 member 2	-6.966848556	2.38E-05	0.014612744
	PTGES3-203	ENST00000436399.6	prostaglandin E synthase 3	-7.058628033	8.66E-05	0.035937101
	LRRC37A4P-201	ENST00000579913.5	leucine rich repeat containing 37 member A4, pseudogene	-7.061224877	7.24E-05	0.031459953
	ZNF24-201	ENST00000261332.10	zinc finger protein 24	-7.21572215	4.79E-08	0.000128309
	NCLN-205	ENST00000590671.5	nicalin	-7.399755895	3.00E-05	0.01711235
	SPP1-209	ENST00000509659.5	secreted phosphoprotein 1	-7.435168587	5.87E-05	0.029524526
	SFPQ-205	ENST00000468598.5	splicing factor proline and glutamine rich	-7.564491161	2.23E-06	0.002859069
	SCO2-206	ENST00000638670.1	SCO2, cytochrome c oxidase assembly protein	-7.662843253	1.91E-05	0.013097744
	PIK3CG-201	ENST00000359195.3	phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit gamma	-7.665284933	3.62E-09	3.55E-05
	CBLL1-205	ENST00000440859.7	Cbl proto-oncogene like 1	-7.843465771	7.62E-06	0.006801784
Transcript	CERS2-201	ENST00000271688.10	ceramide synthase 2	-7.882251745	9.02E-06	0.007441678
	MAL-203	ENST00000353004.7	mal, T cell differentiation protein	-7.91729657	8.87E-06	0.007441678
	GLYR1-210	ENST00000589389.5	glyoxylate reductase 1 homolog	-8.101934631	3.59E-06	0.004072641
	NHS-204	ENST00000615422.1	NHS actin remodeling regulator	-8.136278335	2.13E-06	0.002857325
	TPR-205	ENST00000474852.1	translocated promoter region, nuclear basket protein	-8.220346974	5.03E-05	0.026932277
	EIF4G1-216	ENST00000427845.5	eukaryotic translation initiation factor 4 gamma 1	-9.01725429	1.07E-05	0.007887513
	CREBBP-202	ENST00000382070.7	CREB binding protein	-10.55253344	9.81E-07	0.001520882
	ALDH1A3-202	ENST00000346623.6	aldehyde dehydrogenase 1 family member A3	-10.58984523	2.03E-08	7.48E-05
	REST-208	ENST00000619101.4	RE1 silencing transcription factor	-22.26722887	1.22E-08	5.16E-05
	LPP-222	ENST00000640853.1	LIM domain containing preferred translocation partner in lipoma	-22.32908136	1.11E-08	5.16E-05
	RARA-201	ENST00000254066.9	retinoic acid receptor alpha	-22.42825259	9.59E-09	5.16E-05
	HK1-203	ENST00000360289.6	hexokinase 1	-22.44691451	9.32E-09	5.16E-05

Under the threshold of both *p* and *padj* $\leq$ 0.05. *: log2 fold change.

ZNF493
GBP3
ENPP2
EIF4G1
HLA-DQA1
WDR74
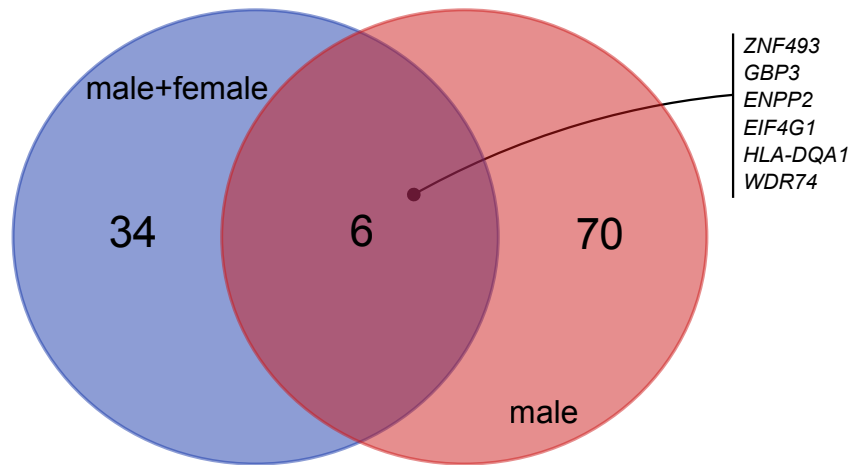
Figure 3.2.13: Venn diagram exibits the comparison of differential expression genes in all samples (including male and female samples) and male samples by DESeq2.
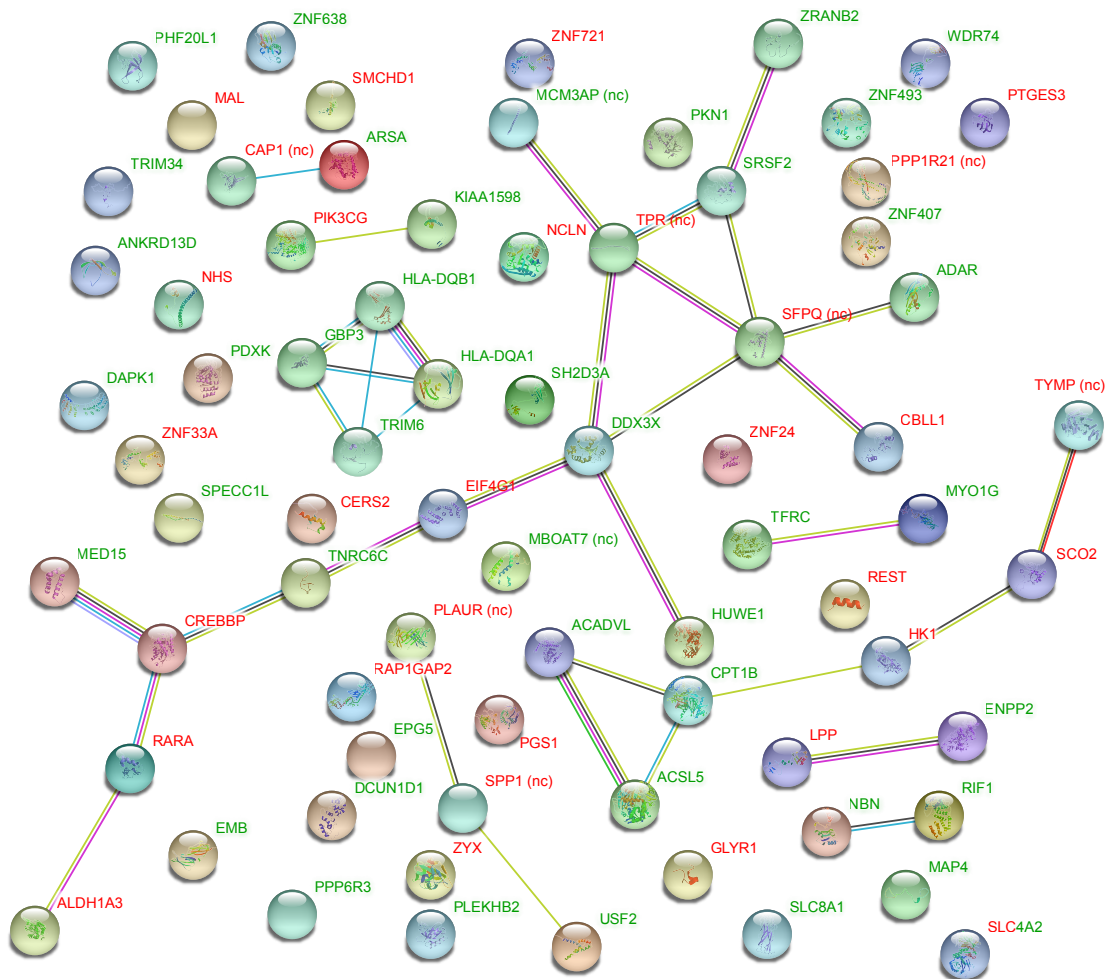
Figure 3.2.14: Potential protein interations of differential expressions in male samples. Upregulated genes marked in green, and the downregulated is shown in red. (nc): noncoding.

3.2.4.3.1 MHC class II immune response

The significant upregulation of the HLA-DQ cluster (HLA-DQA1, HLA-DQB1, GBP3 and TRIM6) in Ct^+ group suggests the acquired immune response mediated by CD4$^+$ T cells with the recognition of the MHC class II molecules, as described in Chapter 3.2.4.2.2 and based on the reports of (Loomis and Starnbach, 2002) and (Karunakaran et al., 2015b). Meanwhile, the transcript of T cell differentiation protein (MAL) highly reduced the regulation after Ct infection.

3.2.4.3.2 RNA binding and transcriptomic regulation

Genes associated with RNA binding, regulation and splicing were differentiated with the presence of Ct. Significant upregulation of DDX3X was presented together in Ct^+ group with the downregulation of EIF4G1. DDX3X is X-linked DEAD (Asp-Glu-Ala-Asp)-box RNA helicase 3, which reported blocking eIF4E-eIF4G complex formation at the 5'UTR cap site of RNA element by inhibiting eIF4E (Shih et al., 2008). DDX3X displays both functions as an oncogene and a tumour suppressor candidate (He et al., 2018). On the 3'UTR site of the RNA element, the TNRC6 family (trinucleotide repeat containing 6) normally interacts with the miRNA (microRNA) associated argonaute (Ago) and together with functional on targeting mRNA degradation (Trabucchi, 2018). The miRNA-target mRNA degradation regulator TNRC6C was downregulated in Ct^+ group. Containing several components (A, B and C), TNRC6 is involved in controlling miRNA-dependent gene expression by miRNA-target sequestration and inducing deadenylation on target RNAs (Trabucchi, 2018). Besides, the RNA-binding promoter SFPQ (splicing factor proline and glutamine-rich) regulates miRNA binding and controls miRNA-induced mRNA silencing (Bottini et al., 2017, Trabucchi, 2018). In Ct positive group, the non-coding transcript of SFPQ (205) highly decreased the regulation, as well as the TPR-205, a non-coding transcript of nuclear basket protein translocated promoter region. In addition, protein-coding transcripts of alternative splicing factors SRSF2 and ZRANB2 presented promoted regulation in Ct^+ group. Both of SRSF2 and ZRANB2 are involved in mRNA metabolism and functionally increased exon skipping and exon inclusion respectively (Giono et al., 2016, Ip et al., 2011, Solier et al., 2010).

3.2.4.3.3 DNA methylation and DNA damage

Deficiency of DNA methylation was revealed by the significantly decreased regulation of the protein-coding transcripts of the histone acetyltransferases CREB (cAMP-response element-binding protein) binding protein (CREBBP, also called CBP), the retinoic acid

receptor a (RARa), and the aldehyde dehydrogenase 1 family member A3 (ALDH1A3). These genes are also coregulated with some transcription of cell proliferation and antiapoptosis related genes (e.g. c-Myc, CyclinD1) (Tomita et al., 2017, Uribesalgo et al., 2012). For example, RARa directly interacts with c-Myc (Hörlein et al., 1995) and forms the c-Myc/RARa complex to regulate RARa-dependent targets via switching phosphorylation in leukaemia cells (Uribesalgo et al., 2012). Additionally, the E3 ubiquitin protein ligase HUWE1 (HECT, UBA and WWE domain containing 1, also known as MULE, ARF-BP1 or HectH9) was significantly upregulated in Ct^+ group. HUWE1 is a Myc ubiquitin E3 enhances the transcriptional activity of Myc instead of inducing degradation with the recruitment of the cofactor p300 (Adhikary et al., 2005, Sun et al., 2015). The HUWE1 is involved in mediating ubiquitination of many other substrates, including Miz-1, Mcl-1, p53, and H1.3, which are functional on cell proliferation, DNA-damage induced apoptosis and transformation of the ovarian epithelial cell (Kao et al., 2018, Yang et al., 2017). Moreover, the DNA-damage repair protein nibrin (NBN) was also highly upregulated, which enhances the association between Ct infection and DNA damage.

3.2.4.3.4 Fatty acid β oxidation induced mROS and the response

In mitochondria, enhancement of fatty acid β oxidative degradation was highlighted according to the upregulated genes of very long chain acyl-CoA dehydrogenase (ACADVL), acyl-CoA synthetase long chain family member 5 (ACSL5) and carnitine palmitoyltransferase 1B (CBT1B). Besides, decreased regulations were found in protein-coding transcripts of hexokinase 1 (HK1) and cytochrome c oxidase assembly protein (SCO2), which reveals depression of glycolysis and oxidative phosphorylation (OXPHOS). HK1 was downregulated the most in all the DEGs and DETs with more than 22 log2 fold change. The accumulation of fatty acids β oxidation induces the production of mitochondrial reactive oxygen species (mROS). HK1 and SCO2 are not only participated in the metabolic activity but also regulated by oncogene Myc and p53 respectively (Kroemer and Pouyssegur, 2008). In addition, downregulation of ceramide synthase 2 (CERS2) and upregulation of arylsulfatase A (ARSA), probably suggests the shrinkage of ceramide, which wields on the oxidative stress in mitochondria (Kogot-Levin and Saada, 2014). The reducing of ROS through regulating glycolysis, OXPHOS and ceramide synthesis are supposed to be the host response to inhibit mROS induced by Ct.

3.2.4.3.5 Solute carrier family transporters

Host responses were also shown in ion transporters. AE2 (SLC4A2, solute carrier family 4 member 2) presented contrary regulation between its two splice variants: the upregulated

SLC4A2-202 and the downregulated SLC4A2-214. AE2 is regarded as the "house-keeping transporter and sensitively regulated by cytosolic pH (Bonar and Casey, 2008). It converts intracellular HCO_3^- to the outside of the plasma membrane and takes extracellular Cl^-. In addition, another solute carrier transporter like SLC8A1 was highly upregulated.

3.2.4.3.6 Zinc fingers

Several zinc fingers were detected with significant variations in expression. Belonging to the protein family of KRAB box transcription factors, ZNF493 and ZNF407 exhibited increasing regulation, while ZNF721 and ZNF33A were downregulated. Besides, ZNF638 and ZNF24 displayed differential expression, with up and down regulation respectively. Less is known about the mechanisms of how these zinc fingers functionally regulated transcriptions. Nevertheless, ZNF407 was demonstrated to cause intellectual disability and autism when mutated to loss of function (Ren et al., 2013), and was reported to associate with tumour progression and essentially bound by WDR5 to affect the metastasis of colorectal cancer (Tan et al., 2017). Additionally, the regulation of REST (RE1 silencing transcription factor) was highly decreased.

3.2.4.3.7 Differential expressions in other pipelines

Cuffdiff and Ballgown were also applied in the analysis for comparison and the results were shown in Table 3.2.5 and Table 3.2.6 respectively. No statistically significant DEGs and DETs were presented in Ballgown's results. However, Cuffdiff extracted 38 DEGs, with 28 upregulation and 10 downregulation. Small proline-rich protein 2A, 2D, 2E and 2F (SPRR2A, SPRR2D, SPRR2E and SPRR2F) were presented increased regulation in Ct^+ group, together with keratin 4 (KRT4) and cornifelin (CNF). It suggests that Ct infection may be associated with host cell cornification. Besides, genes functional on DNA damage and cell apoptosis were also shown, such as TP53. However, regulated differentiation is questionable. For example, the gene *mal* results in upregulation in Cuffdiff but displays downregulation in DESeq2 as a transcript. eclark28 (2017) discussed the bias caused in Cuffdiff for the performance of positive fold change more than negative fold change when also analyzing human transcriptome with the reference of GRCh38. Also, more bias driven by Cuffdiff compared to Ballgown were reported (Frazee et al., 2015).

Table 3.2.5: Differentially expressed genes in male samples by Cuffdiff

Gene name	Annotation	Gene_id	log2FC*	p_value	q_value
OPHN1	oligophrenin 1	ENSG00000079482.12	8.57411	5.00E-05	0.0145286
CR2	complement C3d receptor 2	ENSG00000117322.17	6.51682	5.00E-05	0.0145286
KIAA1524	cell proliferation regulating inhibitor of protein phosphatase 2A	ENSG00000163507.13	5.80052	5.00E-05	0.0145286
SOX5	SRY-box 5	ENSG00000134532.15	5.17098	5.00E-05	0.0145286
LINC00598	long intergenic non-protein coding RNA 598	ENSG00000215483.10	4.74225	5.00E-05	0.0145286
IDH3A	isocitrate dehydrogenase 3 (NAD(+)) alpha	ENSG00000166411.13	4.50067	5.00E-05	0.0145286
PAX5	paired box 5	ENSG00000196092.12	3.83298	5.00E-05	0.0145286
FCRL5	Fc receptor like 5	ENSG00000143297.18	3.67423	5.00E-05	0.0145286
TP53	tumor protein p53	ENSG00000141510.16	3.51496	5.00E-05	0.0145286
LRMDA	leucine rich melanocyte differentiation associated	ENSG00000148655.14	3.48335	5.00E-05	0.0145286
IGKC	immunoglobulin kappa constant	ENSG00000211592.8	3.39935	5.00E-05	0.0145286
RDH13	retinol dehydrogenase 13	ENSG00000160439.15	3.28959	5.00E-05	0.0145286
SPRR2E	small proline rich protein 2E	ENSG00000203785.8	3.19034	5.00E-05	0.0145286
FASN	fatty acid synthase	ENSG00000169710.8	3.11648	5.00E-05	0.0145286
CRCT1	cysteine rich C-terminal 1	ENSG00000169509.5	3.0114	2.00E-04	0.042375
BANK1	B cell scaffold protein with ankyrin repeats 1	ENSG00000153064.11	2.92464	1.00E-04	0.0260769
RAD51D	RAD51 paralog D	ENSG00000185379.20	2.88922	5.00E-05	0.0145286
KRT4	keratin 4	ENSG00000170477.12	2.81372	5.00E-05	0.0145286
MS4A1	membrane spanning 4-domains A1	ENSG00000156738.17	2.74006	2.00E-04	0.042375
SPRR2F	small proline rich protein 2F	ENSG00000244094.1	2.65952	5.00E-05	0.0145286
CNFN	cornifelin	ENSG00000105427.9	2.65381	5.00E-05	0.0145286
DMKN	dermokine	ENSG00000161249.20	2.56377	5.00E-05	0.0145286
ECM1	extracellular matrix protein 1	ENSG00000143369.14	2.4274	0.00015	0.0339
SPRR2D	small proline rich protein 2D	ENSG00000163216.6	2.36241	5.00E-05	0.0145286
PVT1	Pvt1 oncogene	ENSG00000249859.9	2.27973	5.00E-05	0.0145286
MAL	mal, T cell differentiation protein	ENSG00000172005.10	2.09795	5.00E-05	0.0145286
SPRR2A	small proline rich protein 2A	ENSG00000241794.1	2.09468	0.00015	0.0339
IKZF3	IKAROS family zinc finger 3	ENSG00000161405.16	1.86705	1.00E-04	0.0260769
HNRNPH1	heterogeneous nuclear ribonucleoprotein H1	ENSG00000169045.17	-1.74486	5.00E-05	0.0145286
NPIPB4	nuclear pore complex interacting protein family member B4	ENSG00000185864.16	-1.8724	0.00015	0.0339
TCF7L2	transcription factor 7 like 2	ENSG00000148737.16	-1.90931	5.00E-05	0.0145286
BSCL2	BSCL2, seipin lipid droplet biogenesis associated	ENSG00000168000.14	-2.63481	5.00E-05	0.0145286
MXRA5	matrix remodeling associated 5	ENSG00000101825.7	-2.66488	5.00E-05	0.0145286
C3AR1	complement C3a receptor 1	ENSG00000171860.4	-3.17092	5.00E-05	0.0145286
AC110079.1	uncharacterized LOC729218	ENSG00000260404.3	-3.71756	2.00E-04	0.042375
NAV2	NAV2 antisense RNA 3	ENSG00000166833.19	-4.15425	5.00E-05	0.0145286
SLC2A14	solute carrier family 2 member 14	ENSG00000173262.11	-4.24983	5.00E-05	0.0145286
DDIAS	DNA damage induced apoptosis suppressor	ENSG00000165490.12	-5.43433	5.00E-05	0.0145286

Under the threshold of both *p* and *padj* $\leq$ 0.05. *: log2 fold change.

Table 3.2.6: Differential expressions in male samples by Ballgown

Type	Gene name	Annotation	log2FC*	pval	qval
Gene	RNU6-1316P	RNA, U6 small nuclear 1316, pseudogene	-2.151923587	9.21E-05	0.51427453
Transcript	HBP1	HMG-box transcription factor 1	-2.401999815	6.49E-07	0.018501949
	PPP1R10	protein phosphatase 1 regulatory subunit 10	1.700641923	2.30E-05	0.218079323
	CRNKL1	crooked neck pre-mRNA splicing factor 1	-2.334258813	3.03E-05	0.218079323
	SERPING1	Serpin family G member 1	-3.419376111	3.45E-05	0.218079323
	KPNA4	karyopherin subunit alpha 4	-2.139232248	5.71E-05	0.218079323
	KDM6B	Lysine demethylase 6B	3.402868404	5.82E-05	0.218079323
	ANP32A	Acidic nuclear phosphoprotein 32 family member A	-3.079927823	6.70E-05	0.218079323
	IFI30	IFI30, lysosomal thiol reductase	-1.050845047	7.16E-05	0.218079323
	PLEKHB2	Pleckstrin homology domain containing B2	-1.861493669	7.33E-05	0.218079323
	YTHDC1	YTH domain containing 1	-1.685911404	7.65E-05	0.218079323

Top 10 expressions are listed under the threshold of $p \leq 0.05$. *: log2 fold change.

3.2.4.4 Summary: host response in Ct^+ group

In summary (Figure 3.2.15 a), the Ct^+ group significantly shows immunoglobulin-mediated host immune response, including immunoglobulin production, diversification and recombination of related genes and segments. The immune response is associated with DNA methylation and the site of DNA damage especially double-strand break supposed to be induced by Ct infection. Besides, Ct influences host regulation of interferon-β production and transcriptional inhibition exhibited by protein-coding DETs of DDX3X, EIF4G1, RARA and MAL. In addition, Ct infection accelerates fatty acid β oxidation in mitochondria and thereby inducing mROS, and the host responds to the increased mROS via decreasing ceramide production and impairing activity of glycolysis. Ct infection is also associated with viral myocarditis pathways by the IFNγ-stimulated genes GBP3, EIF4G1 and both α and β subunits of HLA-DQ (Bachmaier et al., 1999, Grayston et al., 1981). Ct may also induce host response in spliceosomal alternative mRNA splicing (REST, RARA), endoplasmic reticulum organization (ARSR, BGP3, SHTN1), retinoic acid signalling (ALDH1A3, RARA, CPT1B) and protein kinase C binding (MAL, ACSL5, CPT1B). Although not shown in the signalling cluster, ubiquitination is highly affected by Ct infection according to the significant upregulation of HUWE1 ($p \leq 0.05$, $q \leq 0.05$.) and the MDM2 ($p \leq 0.05$, $q \leq 0.1$).

Additionally, results from Cuffdiff is likely to cause some bias although, it could still regarded as a secondary reference (Figure 3.2.15 b). The reinforcement of corneodesmosin bound cornified envelope is specially presented by Cuffdiff but not DESeq2. Many factors do not form large pathways themselves but promote the functional network generated from results by DESeq2. Besides, transcriptional regulation is contributed by the appearance of BANK1, TCF7L2, PVT1. preference of CR2 and MS4A1 together with TFRC and HLA-DQ may indicate the association with hematopoietic cell linkage. This is also related to the discussed protein C kinase binding in DESeq2 via genes associated with B cell proliferation and regulation. Many genes are found in cysteine-type endopeptidase activity involved in the apoptotic process, including DDX3X, DAPK1, FASN, NBN, PVT1, TFRC and REST. Nuclear receptors in signalling pathway are displayed by ALDH1A3, CPT1B, CREBBP, PTGES3, RARA, RDH13, TNRC6C and USF2. Most of the other pathways are related to the tumour repressor protein 53 (TP53). TP53 is shown an upregulation in transcript level resulted in Cuffdiff but not in DESeq2, and it has been reported degradation of protein P53 was induced by Ct infection for the ubiquitin (Siegl et al., 2014). It is still questionable whether TP53 is exactly increased in transcript level during Ct infection or just caused by bias in Cuffdiff.

(a) Clusters from DESeq2

(b) Clusters summarized from DESeq2 and Cuffdiff
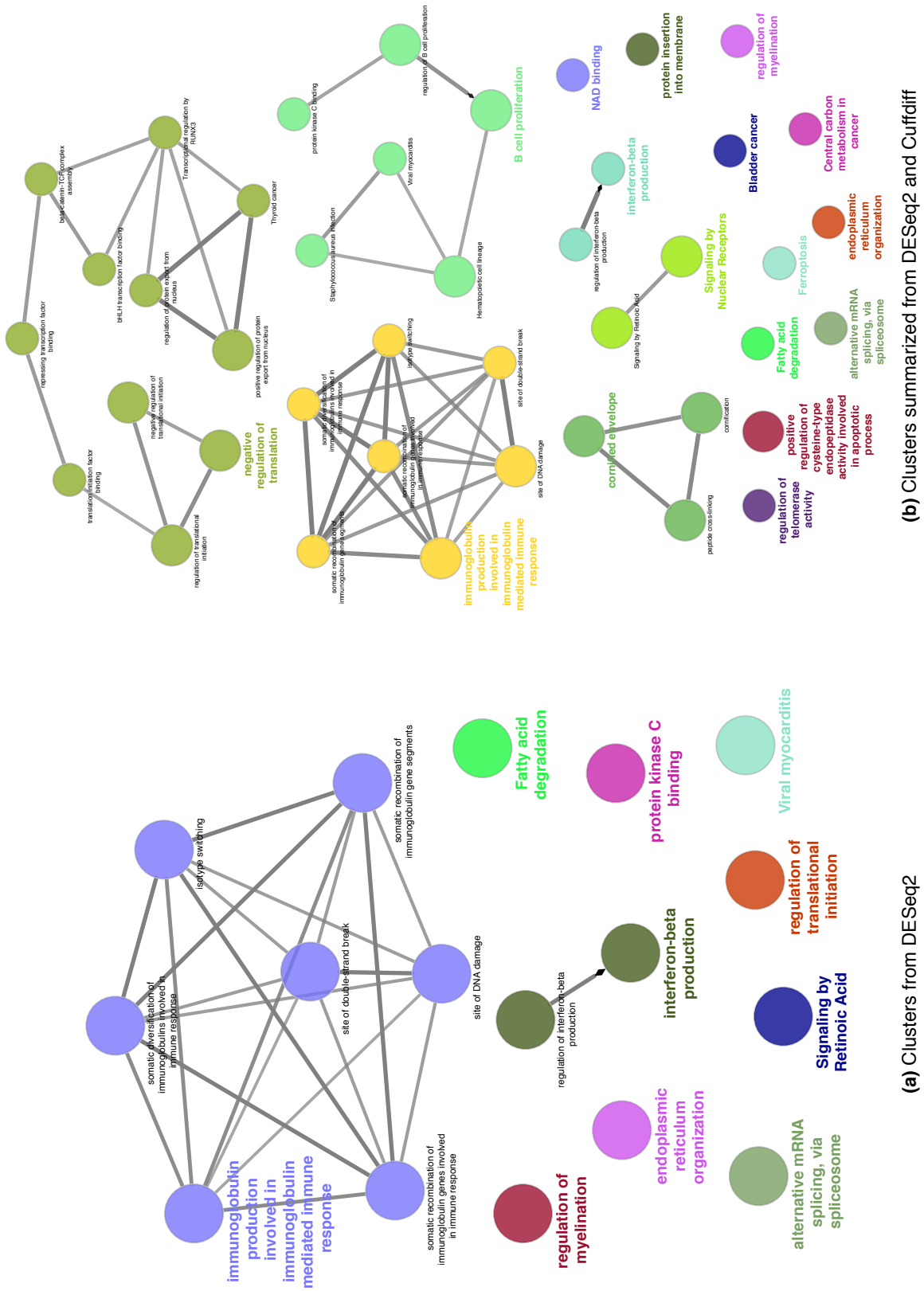
Figure 3.2.15: functional cluster analysis of differentially expressed genes

3.2.5 *GC* transcriptome profiling and differential analysis

Compared to the eukaryotic organisms, *GC*, the bacterial pathogen has no introns in the genome, thereby not processing sophisticated splicing as in the human host. Thus, only gene expression but not transcript expression would be discussed here. Figure 3.2.16 shows *GC*'s gene expression variance among different samples. The PCA plot showed *GC*'s gene expression clusters based on the conditions. The factor of host gender caused big variance (Figure 3.2.16 a) and the female communities were not well clustered (Figure 3.2.16 b). In male samples, the communities were separated into *Ct* positive and negative groups (Figure 3.2.16 c). Thus, the results would mainly focus on the influences *Ct* caused in *GC*'s gene expression only according to male samples.

(a) *GC* gene expression in all samples

(b) PCA in all samples

(c) PCA in male samples
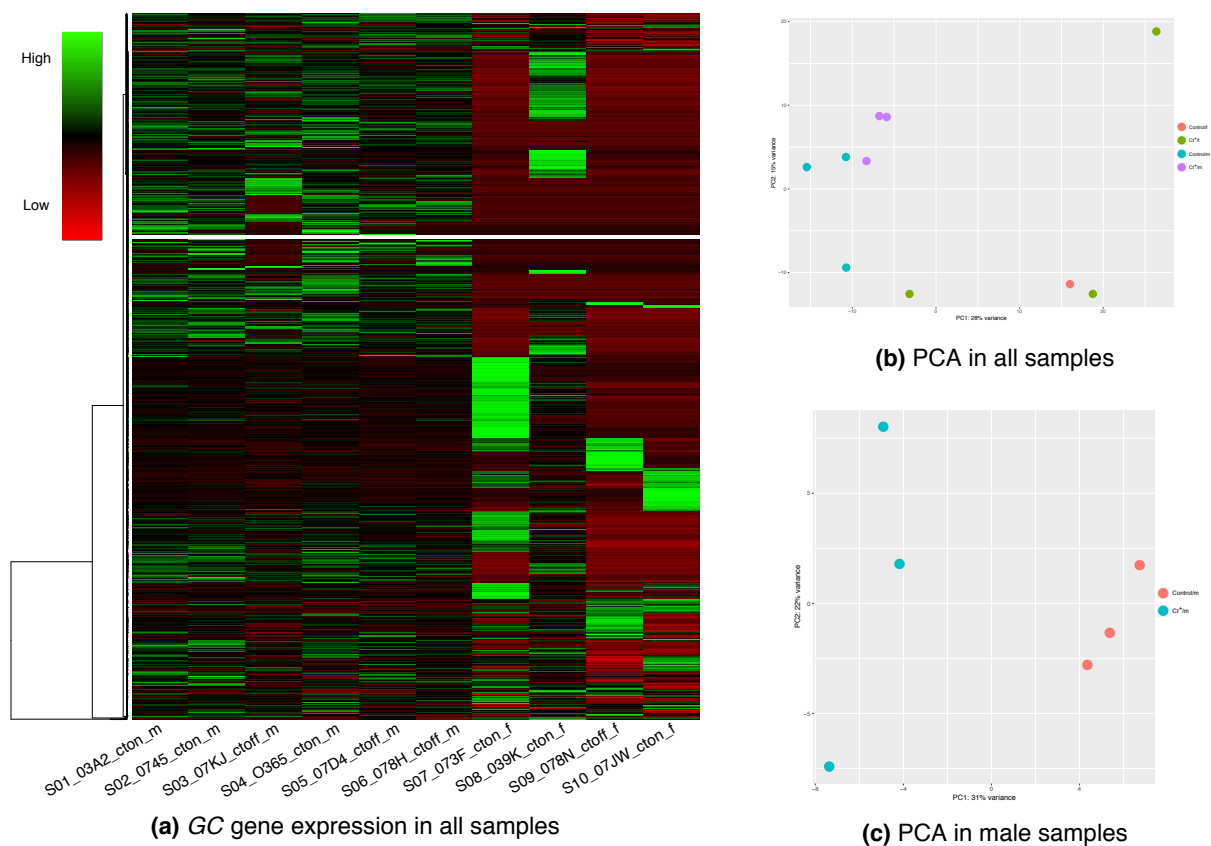
Figure 3.2.16: *Neisseria gonorrhoeae* (*GC*) gene expression and PCA plot in all samples and male samples.

3.2.5.1 Differential expressions of *GC* in Ct^+ male samples

The *GC*'s differentially expressed genes (DEGs) were shown in Table 3.2.7 and Table 3.2.8 analyzed by DESeq2 and Cuffdiff respectively. DESeq2 processed 17 DEGs by both *p* and *padj* $\leq$ 0.05, and 10 DEGs more when *padj* between 0.05 and 0.1. The overall 27 DEGs

had upregulations and downregulations equally. On the other side, 47 DEGs were filtered out by Cuffdiff under the threshold of both p and $padj \leq 0.05$, with 30 of them were upregulated and the rest 17 were downregulated. The two different methods shared a high similarity, with 20 DEGs in total presented in both lists (Figure 3.2.17). Unlike the host transcriptome, 74% GC's DEGs extracted from DESeq2 were found in the results of Cuffdiff as well. Though Cuffdiff had unsatisfying performance in host transcriptome analysis, it is experientially applied in microbial transcriptomics analysis. Ballgown was designed for eukaryotic organisms and caused a data-loading error in analyzing GC's transcriptome. The Ballgown's failure in analyzing GC transcriptome was caused by the necessary intron data import for downstream processing but prokaryote has no introns. Thus, here we performed both results from DESeq2 and Cuffdiff workflows but not Ballgown for GC, and summarized results together for further analysis. These summarized DEGs suggest Ct's infection affected on GC's expression of ion transporters, amino acid and ferric absorption, restriction-modification system and responses to nitrosative and oxidative stress.
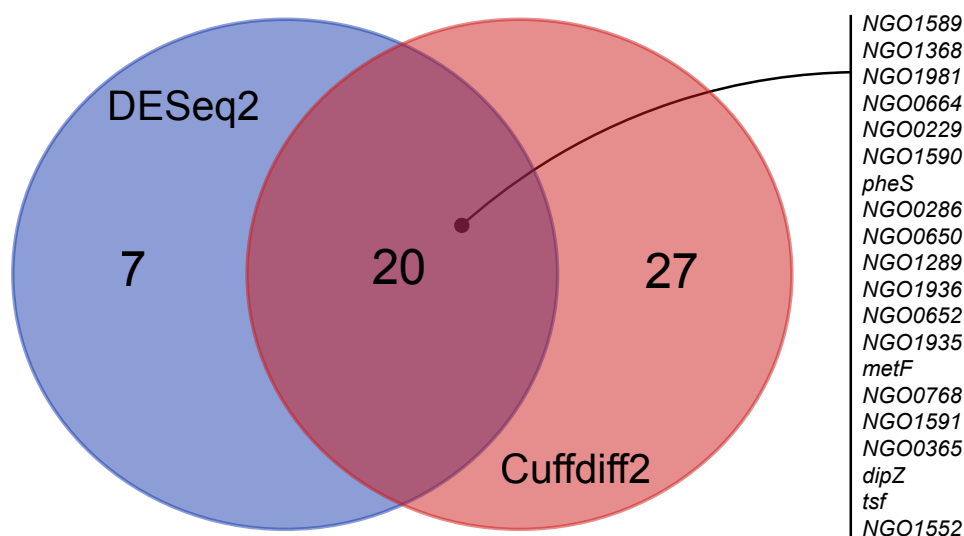
Figure 3.2.17: Venn diagram exhibits the comparison of differential expression genes in male samples analyzed by DESeq2 and Cuffdiff.

3.2.5.1.1 Ion transporters and amino acid absorption

Ion transporters are the sensitive targets for the acquisition of corresponding carbon or nitrogen source directly from the host niche. NGO1368 is annotated as the efflux pump component MtrF in the strain FA1090, which belongs to p-Aminobenzoyl-glutamate transporter family. MtrF was significantly upregulated with 5times fold change in Ct^+ group. MtrF is associated with resistance of antimicrobial hydrophobic agents (Folster and Shafer,

Table 3.2.7: Differential expressed genes of *Neisseria gonorrhoeae* with statistical significance in male samples by DESeq2

Gene_name	Annotation	log2FC*	pvalue	padj
dnaG	DNA primase	7.548708371	1.18E-08	9.88E-06
NGO1967	hypothetical protein	6.016102892	1.15E-05	0.003196226
lpxK	tetraacyldisaccharide 4'-kinase	5.686508071	4.91E-05	0.009121497
NGO1384	hypothetical protein	3.797396978	0.001595457	0.098859226
NGO0768	hypothetical protein	2.796343446	1.67E-05	0.00399196
NGO0229	hypothetical protein	2.374455637	6.95E-06	0.002325753
NGO1368	efflux pump component MtrF	2.28142944	2.67E-05	0.005593272
NGO1289	hemolysin III family channel protein	2.082552054	0.000936508	0.063695852
NGO1552	sodium/proline symporter	1.859861232	0.000862773	0.063695852
NGO1981	hypothetical protein	1.832949171	0.000755883	0.063695852
NGO0664	membrane protein	1.726731561	0.000371295	0.041411824
NGO0286	hypothetical protein	1.483749269	0.000440808	0.04338073
NGO0650	ATP-dependent RNA helicase	1.398146652	0.001566384	0.098859226
dipZ	thiol:disulfide interchange protein	1.248554095	0.000843023	0.063695852
tsf	elongation factor Ts	-1.096541908	0.000318848	0.041411824
NGO1935	electron transfer flavoprotein subunit beta	-1.308690168	0.000552712	0.051371516
NGO0652	thioredoxin I	-1.327465594	0.000923631	0.063695852
NGO1936	electron transfer flavoprotein subunit alpha	-1.356225126	0.000873456	0.063695852
metF	5,10-methylenetetrahydrofolate reductase	-1.415944144	0.000363311	0.041411824
pheS	phenylalanine–tRNA ligase subunit alpha	-1.466223002	9.19E-05	0.013970565
NGO1565	nicotinate-nucleotide pyrophosphorylase	-1.785270034	0.000951821	0.063695852
NGO1589	pretoxin HINT domain-containing protein	-1.917299947	7.91E-05	0.013239528
NGO1590	hypothetical protein	-1.960032831	1.94E-06	0.000810901
NGO0365	site-specific DNA-methyltransferase	-1.983549212	0.000326465	0.041411824
NGO1591	hypothetical protein	-2.48791197	6.84E-07	0.000381637
NGO0087	hypothetical protein	-5.355544737	0.000427626	0.04338073
NGO0228	hypothetical protein	-9.237828451	8.46E-13	1.42E-09

Under the threshold of both *p* and *padj* $\leq$ 0.05. *: log2 fold change

Table 3.2.8: Differential expressed genes of *Neisseria gonorrhoeae* with statistical significance in male samples by Cuffdiff

Gene name	Annotation	log2FC*	p_value	q_value
NGO2109	hemoglobin-haptoglobin utilization protein B	3.13907	0.00145	0.0434396
NGO0768	hypothetical protein	2.84593	1.00E-04	0.00553077
NGO1562	ArsR family transcriptional regulator	2.58171	5.00E-05	0.00378421
NGO0229	hypothetical protein	2.3972	5.00E-05	0.00378421
NGO1628	phage associated protein	2.38895	5.00E-05	0.00378421
NGO0757	hypothetical protein	2.35543	2.00E-04	0.00845882
NGO1368	efflux pump component MtrF	2.26523	5.00E-05	0.00378421
NGO1981	hypothetical protein	2.25948	5.00E-05	0.00378421
NGO0664	membrane protein	2.14507	1.00E-04	0.00553077
NGO1289	hemolysin III family channel protein	2.1232	1.00E-04	0.00553077
NGO1776	glyceraldehyde-3-phosphate dehydrogenase	2.00418	5.00E-05	0.00378421
NGO1552	sodium/proline symporter	1.99663	5.00E-05	0.00378421
NGO1386	hypothetical protein	1.98648	0.00105	0.0351511
NGO0372	amino acid ABC transporter substrate-binding	1.95379	5.00E-05	0.00378421
NGO1059	membrane protein	1.95093	5.00E-05	0.00378421
NGO1040a	opacity protein	1.9192	5.00E-05	0.00378421
NGO0374	amino acid ABC transporter ATP-binding protein	1.76608	1.00E-04	0.00553077
NGO1347	hypothetical protein	1.7364	0.00145	0.0434396
NGO1439	macrolide ABC transporter ATP-binding	1.64969	0.00015	0.00719
NGO0286	hypothetical protein	1.58922	0.00035	0.0132447
NGO0340	cysteine synthase	1.52193	5.00E-05	0.00378421
NGO0650	ATP-dependent RNA helicase	1.46194	5.00E-05	0.00378421
dnaA	chromosomal replication initiation protein DnaA	1.44653	0.00025	0.00998611
NGO2127	cadmium resistance protein	1.38806	0.0011	0.0351511
NGO0399	protease HtpX	1.30438	4.00E-04	0.0147487
dipZ	thiol:disulfide interchange protein	1.29577	0.00025	0.00998611
NGO0795	bacterioferritin B	1.23819	0.0011	0.0351511
NGO1404	glycine cleavage system protein H	1.14035	0.00115	0.03595
NGO1659	intracellular septation protein A	1.11763	0.0017	0.048892
NGO1947	hypothetical protein	1.03447	0.00175	0.0493431
tsf	elongation factor Ts	-1.10195	9.00E-04	0.0315659
NGO1767	catalase	-1.10217	0.0017	0.048892
NGO1935	electron transfer flavoprotein subunit beta	-1.26029	1.00E-04	0.00553077
NGO1936	electron transfer flavoprotein subunit alpha	-1.30626	0.00015	0.00719
NGO0652	thioredoxin I	-1.30859	2.00E-04	0.00845882
metF	5,10-methylenetetrahydrofolate reductase	-1.40192	0.00015	0.00719
pheS	phenylalanine–tRNA ligase subunit alpha	-1.48695	5.00E-05	0.00378421
NGO1592	pretoxin HINT domain-containing protein	-1.52282	5.00E-04	0.017975
NGO0055	pilus assembly protein	-1.67525	0.0011	0.0351511
NGO1769	cytochrome-c peroxidase	-1.707	1.00E-04	0.00553077
NGO0545	type III restriction-modification system	-1.71207	2.00E-04	0.00845882
NGO1589	pretoxin HINT domain-containing protein	-2.09034	3.00E-04	0.0116595
NGO1590	hypothetical protein	-2.2343	5.00E-05	0.00378421
NGO1276	nitrite reductase	-2.29543	1.00E-04	0.00553077
NGO0365	site-specific DNA-methyltransferase	-2.38617	0.00015	0.00719
NGO0363	DEAD/DEAH box helicase	-2.58688	5.00E-05	0.00378421
NGO1591	hypothetical protein	-2.66464	5.00E-05	0.00378421

Under the threshold of both *p* and *padj* $\leq$ 0.05. *: log2 fold change

2005, Veal and Shafer, 2003). It suggests Ct infection may induce the antimicrobial response from the host cell, thereby modifying the surface hydrophobicity which revises bacterial pathogenic adhesion. Interestingly, NGO0372 and NGO0374 (amino acid ABC transporter) were observed significant upregulation with almost 4 times fold change in Ct^+ group. Together with MtrF, NGO0372 and NGO0374 were reported the upregulation caused by glutamine (Friedrich et al., 2007). It reflects the increased level of glutamine in the host niche after Ct infection, as glutamine uptake is beneficial or even necessary for the metabolism of Ct in host adaptation. NGO1552, regarded as sodium/proline symporter PutP, was also significantly upregulated in Ct^+ group. A similar transporter OpuE in *Bacillus subtilis* was reported to be associated with osmotic stress response (Spiegelhalter and Bremer, 1998), and PupT in *Staphylococcus aureus* supplied proline scavenging by the pathogen from the host cell for *in vivo* survival (Jung, 2002, Schwan et al., 1998). The mechanism of the sodium/proline transporter in GC was unclear, however, it showed the presence of Ct may promote the proline level in the host. The Ct-caused increasing of glutamine and proline in the niche revealed potential competition of metabolic resource among different pathogens under the osmotic stress.

The influence of the amino acid was not only for glutamine and proline, but also significantly shown in cysteine. NGO340 (cysteine synthase), utilizing acetyl-serine and hydrogen sulfide to form cysteine and acetate, was highly upregulated. Besides, NGO0372 and NGO0374 were also reported functional as the transporters for absorbing cystine and cysteine (Bulut et al., 2012), and both of them were upregulated as described above. Because Ct needs cysteine to form its cysteine-rich type III secretion system (T3SS) during infection, the level of cysteine in the host niche was in shortage status for other pathogens. GC displayed different competitive strategies by upgrading its own synthesis of cysteine and increasing cysteine imported from the host.

3.2.5.1.2 DNA replication and restriction-modification system

DNA primase ($dnaG$) was the most significantly upregulated gene in results from DESeq2. The upregulation of $dnaG$ and chromosomal replication initiation protein ($dnaA$) indicated the accelerated DNA replication in GC when Ct is available. Besides, genes related to restriction-modification (RM) system were dominantly downregulated in Ct^+ group, including NGO0545 (type III restriction-modification system methyltransferase), NGO0363 (DEAD/DEAH box helicase), and NGO0365 (site-specific DNA methyltransferase). The detailed molecular mechanisms could not be clarified only according to these DEGs. However, it implied that GC as a natural competent, would prefer to adapt by replication when the environment changed by the involvement of Ct, rather than systematically defenced.

3.2.5.1.3 Nitrosative and oxidative stress response

Transcriptional regulations related to the well-known fumarate and nitrate regulator (FNR) and stress response were differentially expressed in the Ct positive group. Three FNR activated regulons had reduced regulations, including nitrate reductase (NGO1276, $aniA$), cytochrome-c peroxidase (NGO1769, ccp), and NGO0087 (hypothetical protein, functionally predicted as glycosyl transferase). The gene $aniA$ encodes an outer membrane lipoprotein which catalyzes anaerobic respiration in GC (Mellies et al., 1997). It converts NO_2^- to $NO\cdot$ and belongs to both reactive oxygen species (ROS) and reactive nitrogen species (RNS) regulator. The ccp is also involved in ROS by converting H_2O_2 to H_2O. Besides, electron transfer flavoproteins (NGO1935 and NGO1936, ETFs) became inferior together with ccp. It suggests the electron transfer was deficient in GC's respiratory chain. In addition, NGO1767, a catalase catalyzes either H_2O_2 to O_2 (KEGG reaction: R0009) or O_2 to $O_2\cdot^-$ (KEGG reaction: R02670), was about 2 folds downregulated from analysis of Cuffdiff. Similar to the host response, the glyceraldehyde-3-phosphate dehydrogenase (GAPDH) was significantly upregulated, which would be the GC's response to the increased nitrosative and oxidative stress. In addition, thiol:disulfide interchange protein ($dipZ$) was shown upregulated. $dipZ$ contributes to the reduction of $dsbC$, a gene encoding prokaryotic disulfide bond isomerase which helps to correct the disulfide bond in folded protein (Rietsch et al., 1996). The reducing system protects bacteria from the oxidative damage (Arts et al., 2015), implied Ct may induce the ROS and increase the oxidative stress to GC. In summary, the nitrosative stress of GC in Ct^+ group was increased by defective regulation of related reductase, and the oxidative stress was rising synergistically which regulated by RNS/ROS system. GC utilized the increased regulation of GAPDH and augmented the reducing system to respond to RNS/ROS in order to maintain cellular redox homeostasis.

3.2.5.1.4 Ferric iron acquisition

Ferric uptake regulator (Fur) related regulations reflected the iron uptake was affected in Ct infection. The increased expression of $arsR$ (NGO1562) presented the Fur activated regulation. Fur repressed regulation was revealed by the downregulation of thioredoxin I (NGO0652). Hemoglobin-haptoglobin utilization protein B (NGO2109, $hpuB$), an iron-acquisition Fur-repressed regulon, presented the most significant upregulation with more than 8 times fold change according to the analysis of Cuffdiff. Besides, the bacterioferritin B (NGO0795, $bfrB$) and hemolysin III family channel protein (NGO1289) were upregulated. $bfrB$ utilizes ferritin-like molecules for iron storage (Rivera, 2017), while NGO1289 is helpful in transporting hemolysin and functional on lysing red blood cells which releasing

haemoglobin to the iron-containing heme (Runyen-Janecky, 2013). The participation of *Ct* drove the competition in *GC*'s iron uptake from the host environment.

3.2.5.1.5 Opacity protein

One opacity protein (*opa*) was highly upregulated with the appearance of *Ct*. This protein was regarded homologous with NGO1040a in the reference strain of *Neisseria gonorrhoeae FA 1090*. According to the "CTTCT" repeats in the nucleic acid sequence, this protein supposed to be *opa54* (Stern et al., 1986). While, in the classification system of Bhat et al. (1991), this protein is also likely to be *opaE*(*opa55*). Which *opa* protein exactly it is and how it paticipates in the coinfection between *GC* and *Ct* are still questionable.

4

Discussion

4.1 Constraint-based metabolic modelling enables the study of *Ct* metabolism in quantitative pathways

In this work, *Ct*'s genome-scale metabolic network was first reconstructed based on the constraints, with 321 unique metabolites, 171 enzymes and 277 reactions. Many efforts have been made on studying the metabolic properties of *Chlamydia trachomatis* developmental forms during its biphasic life cycle (Käding et al., 2014, Omsland et al., 2012, Østergaard et al., 2016, Saka et al., 2011). König et al. (2017) reported the biphasic metabolism of a chlamydial symbiont, the *Protochlamydia amoebophila* UWE25 whose host is *Acanthamoeba*, by using RNA-Seq method. For the first time, our work compared the differences between EB and RB by quantitative pathways and not only based on the view of expressed genes or gene clusters. According to our results, both EBs and RBs are metabolically dynamic. Their metabolic differentiations are highly variable based on adaptation to the human host environment. With the support of the omics data, the network analysis helps to give an overview of *Ct*'s metabolism according to quantitative pathways. The network was calculated to have 84 pathways and modelled for both EB and RB in the time points of 20 hpi, 40 hpi and the lysis phase respectively. According to our analysis, EB is more

active than RB in most of the central metabolism, including pentose phosphate pathway, glycolysis and glycerophospholipid biosynthesis. Tricarboxylic acid cycle and fatty acid biosynthesis are relatively ineffective. Downstream of glycolysis and upstream of fatty acid biosynthesis effectively flux to phospholipid metabolism. Enzymes involved in nucleic acid metabolism are strongly expressed in RB, which may offer necessary GTP and ATP for folate biosynthesis and amino acid transformation. EB and RB are both active in folate biosynthesis, which utilizes glutamate and PABA to form folate and glycolaldehyde. The flux of this pathway is stronger in EB than RB, however, the more intensive flux will be presented in RB than EB when more ATP and GTP are available imported from the host or generated by purine and pyrimidine metabolism.

The genome-scale metabolic modelling is a useful method for microbial engineering and studying pathogenic metabolism. From single organism (e.g. *Mycoplasma genitalium* (Suthers et al., 2009) and Methicillin-resistant *Staphylococcus aureus* (Choe et al., 2018)) to microbial communities (Thiele et al., 2013), metabolic modelling offers a new option to reductionistically analyze complex biological system via the chemical and mathematical calculation of stoichiometric matrix.

4.1.1   Gene expression in different cell lines

The famous HeLa229 cell line is frequently used as the host of *Ct* infection. It has a high frequency in cell division which is good for infection study. However, there are many limitations in gene expression study with HeLa cell lines. It has been noticed that some metabolic activities are influenced by the host cell (Stephens, 1999, Tan and Bavoil, 2012). When in HeLa cell lines, pathogen's tricarboxylic acid cycle is supposed to be highly downregulated and nucleic acid metabolism is particularly enriched. The results observed in Chapter 3.1.2 about TCA cycle, purine and pyrimidine metabolism is also probably due to the influence of the host HeLa cells. Because of that, the human umbilical vein endothelial cells (HUVECs) were used as the host for *Ct* infection. *Ct* generally presented more upregulated gene expressions in HUVECs than in HeLa229, especially in the biosynthesis of glycerophospholipid and folate (*pls* and *fol*).

The different gene expressions in different cell lines drive unstable and unpredictable deviations especially in studying pathogenic metabolism. Whether the previous metabolic observations (Käding et al., 2014) in HeLa229 cell lines revealed the real metabolism of *Ct* in a natural infection or not is still questionable. Nevertheless, one approach to improve the metabolic study is to use the 3D human tissue model instead of cell lines. 3D tissue model restores more appropriate environments for studying infection caused by human obligate

pathogens infection, and offers safer and more efficient support for therapeutic strategies and drug design in the future compared to animal experiments or human cell lines.

4.1.2 Carbon and nitrogen source uptake and energy production

Ct lacks an adaptive response in the carbon source changing from glucose to glutamate or *α*-ketoglutarate in transcription level (Nicholson et al., 2004). Thereby the appearance of gluconeogenesis would be very important for storage of carbon source in *Ct* when it encounters different niches during evolution. However, gluconeogenesis is a process with high energy cost, which is also regarded as an abandon choice. *Ct* could take ATP from the host cell by ATP/ADP transporter and it is also able to generate ATP itself according to the genome and the proteome. This indicates *Ct* is supposed to balance the homeostasis between the uptake of carbon source and energy. It suggests *Ct* needs a lot of glutamates if host cell could support enough glutamate or glutamine. The glutamate even could play as a role of energy source for triggering folate biosynthesis and the incomplete TCA cycle.

One question is where the glutamate comes from? Does *Ct* directly take glutamate from the host by glutamate transporter, or it takes glutamine by glutamine transporter, and converts glutamine to glutamate by enzymes? On one side, three glutamate transporters (CT_216, CT_230 and CT_401) are encoded in the genome and could express in protein level Østergaard et al. (2016), and Iliffe-Lee and McClarty (2000) also mentioned the glutamate transporter *gltT* (CT_401) in its carbon metabolism; however, transport of glutamine is still unclear. On the other side, two enzymes encoded by gene *glmS* and *pyrG* present capabilities of utilizing glutamine to transfer to glutamate in both cell lines of HeLa229 and HUVECs (Figure 3.1.7). *glmS* upregulated nearly two-fold change in both cell lines from 12 hpi to 24 hpi, and it is still active in the late lysis phases. While *pyrG*, though downregulated in HeLa229 cell lines during infection, comparatively upregulated in HUVECs. Also, the CTP synthase, product of *pyrG*, functionally transcribed and translated during the mid and late stage of the *Ct*'s life cycle (Wylie et al., 1996). According to both the genome information (Stephens et al., 1998) and many old-day observations (Weiss, 1967), *Ct* takes glutamate with the known transporters in preference to glutamine. Not only in *Ct*, but glutamate may also be a necessary metabolite for many obligate intracellular parasites. However, *Ct* lacks an adaptive carbon catabolite repression when carbon source changes from glucose to glutamate, which means *Ct* may not directly take glutamate from the host (Nicholson et al., 2004). Taken into together, *Ct*'s utilization of glutamate is supposed to be converted from glutamine by *glmS* and *pyrG*.

Pathways generated many energetic products (e.g. NADH and NADPH) were: glycoly-

sis/gluconeogenesis (P5), folate biosynthesis (P33 and P51), glycerophospholipid biosynthesis (P83) and PPP (P84). It is highly hypothetical that *Ct* has quite flexible metabolic capabilities, not only between different forms but also in different carbon source adaptations. When carbon and energy condition changes, it could adjust its metabolism into an adaptive status, which is not as similar as general metabolism. How exact of the adaptation still need further study. What we were sure is that both EB and RB should be metabolically active, for playing different characters respectively.

4.1.3 Kinase and intermediates in central metabolism

Two phosphofructokinases are involved in glycolysis, which are 6-phosphofructokinase 1 (EC 2.7.1.11) and diphosphate-fructose-6-phosphate 1-phosphotransferase (EC 2.7.1.90) encoded by *pfkA_1* (CT_205) and *pfkA_2* (CT_207) respectively. The PfkA_2 is similar to PfkA_1 but utilizing diphosphate instead of ATP to catalyze a reversible reaction. Pfk_2 has been found in higher plants, eukaryotes bacterial and archaea, however, not in *Homo sapiens* and most of the bacteria. In human cells, this reaction is functional by fructose-1,6-bisphosphatase (EC 3.1.3.11). It is potential of PfkA_2 to be a drug target for treating with *Ct* because of its specificity.

Because of the expression of 1,6-fructose biphosphate aldolase (EC 4.1.2.13), fructose 1-phosphate (F1P) is supposed to reversibly exchange with glycerone-P and glyceraldehyde. However, for the lack of fructose kinase, F1P is not able to be transformed from fructose. Thereby theoretically F1P is possible to be generated and might be an intermediate for glycolysis, though it is not sure that whether the 1,6-fructose biphosphate aldolase will works in this reaction or not. The similar situation happened in ribose 1-phosphate (R1P) for the appearance of ribose 5-phosphate isomerase A (EC 5.3.1.6) which encoded by *rpiA* (CT_213). Whether F1P and R1P exist in *Ct* and what're the functions of them are still questionable.

4.2 Metatranscriptomics in host-pathogen interaction and pathogenic coinfection

Isolation and cultivation are the basic methods in standard microbiology. Culture isolated bacteria *in vitro* essentially contributes to the golden age of microbiology in the 19th century (Wade, 2002). Once study bacteria, the initial way for the experiment is prepared an isolated bacteria in order to reduce unnecessary errors and unclear results influenced

by other microbes. Most of the known bacteria are also identified based on isolation. However, what we should not ignore is that bacteria do not grow in isolation in nature. Unculturable bacterial community occupied a huge fraction of the total diversity. It is estimated only 2% bacteria in the environment is able to be cultured, and about half of the human oral microbes are unculturable (Cox et al., 2013, Wade, 2002, Wilson et al., 1997). On the other side, it is effective to study bacterial biofilm and other interactions in microbial communities. Metagenomics and metatranscriptomes offer new aspects to study bacterial communities. New techniques such as dual RNA-Seq promote to study host and pathogen simultaneously. Patient transcriptome profiling and analysis would help to study both host-pathogen and pathogen-pathogen interactions, and pave the way for new strategies for therapy. There was not yet any publication on the clinical Ct-coinfected transcriptomics.

4.2.1 Carrot and stick: Ct-host interaction

According to the differential expression analysis, the host has significantly differential responses in Ct^+ group. MHC class II immune response is highly upregulated whenever within or without female samples. The appearance of Ct increases host fatty acid β oxidation, thereby inducing mROS, which drives the host responses by downregulating glycolysis and ceramide synthesis. Flux analysis shows that Ct's metabolism needs a large amount of acyl-CoA, which could be converted from host acetyl-CoA or from lipid drops. Besides, genes related to DNA methylation and DNA double-strand damage are differentiated in Ct^+ group. Differentiations are also found in factors associated with RNA binding and transcriptional regulation. Moreover, ubiquitin and deubiquitin systems are involved in the host response, however current data cannot support the clearance of intricate mechanisms.

Chlamydia and host apoptosis are worthy to be discussed for its various strategies for surviving and affecting on host apoptotic signalling pathways (Byrne and Ojcius, 2004). The induction of apoptosis caused by Ct on some circumstances (Gibellini et al., 1998, Schöier et al., 2001). However, posterior studies broadly reveal that Ct infection causes inhibition of host apoptosis by regulating host anti-apoptotic proteins (Rudel et al., 2010, Sharma and Rudel, 2009). Many strategies of Ct are observed for apoptosis inhibition, such as stabilization of Mcl-1 (Fischer et al., 2017, Rajalingam et al., 2008). The Mcl-1 ubiquitin ligase E3 (MULE), which called HUWE1 also and normally directly interacted with Mcl-1, has no alterations between the uninfected cell and Ct infected cells 24 hpi (Fischer et al., 2017). However, HUWE1 presents statistically significant upregulation in the patient transcriptome with Ct infection. The infected host cells could be classified as the status

of neutral, anti-apoptosis and cell death according to *Chlamydia* developmental status, and they also pointed out the difficulty to understand the correlation of pro-apoptotic and anti-apoptotic effects with clinical trials (Byrne and Ojcius, 2004). The observation of HUWE1's upregulation in the patient transcriptome may help to understand *Ct*'s capability of interactions with the host cells.

Interestingly, HUWE1 not only binds to Mcl-1 but also regulates other factors, in order to mediate cell proliferation, differentiation, apoptosis and DNA repair (Kao et al., 2018). In HUWE1-mediated apoptosis, HUWE1 mediates ubiquitination of Miz-1 and induce apoptosis by the degradation of Miz-1 when in TNFα stimulation; while under the stress caused DNA damage, HUWE1 could both inducing apoptosis by interacting with HDAC2, Mcl-1 and Cdc6, and inhibiting apoptosis by the degradation of p53 via ubiquitination (Kao et al., 2018). Besides, HUWE1 activates c-Myc and the binding promotes cell growth (Adhikary et al., 2005). Siegl et al. (2014) reported the degradation of p53 induced by *Chlamydia* by activation of HDM2 in protein level. In our result, HDM2 shows upregulation with nearly 7 times log2 fold change in $p = 0.0004$ and *padj* $=0.0982$ in the STI patient DETs, which reveals both of the HUWE1 and HDM2 are highly upregulated. Canfield et al. (2016) observed that more patients with high expressions of both HUWE1 and HDM2 than that with high expression of HUWE1 but low expression with HDM2, which indicate a positive correlation of HUWE1 and HDM2 in breast cancer. Kurokawa et al. (2013) also found HDM2 in some circumstance will promote the degradation of HUWE1, thereby inducing the accumulation of Mcl-1 to antiapoptosis. It is very difficult to clarify the insight of the dynamic and variable regulation with molecular basis during *Ct* infection. However, current results point out that the infection affects the host (anti-)apoptosis via mediating the regulation balanced by ubiquitylation of c-myc-HUWE1 and p53-hdm2, the stasis regulation between oncoprotein and tumour suppressor. Sharma and Rudel (2009) used "be murdered" to describe the infected host cell, which though not directly brings to "suicide". Above all, *Ct* infection for host seems to follow the carrot-and-stick principle, and the most amazing part of the game is the mechanism that when the carrot is decided and when the stick is chosen.

Shown in Chapter 3.2.4.3, the results of DEGs and DETs are statistically significant setting by a threshold of $p \leq 0.05$ and *padj* ≤ 0.05. However, statistical significance does not totally equal to biological significance. When under the threshold of $p \leq 0.05$ and *padj* ≤ 0.1, more differences with less significance come out (Figure 4.2.1). These augment the networks of DNA damage and recombination analyzed by DESeq2 in Figure 3.2.15(a). Interestingly, this shares some signals with the result of Cuffdiff2 in Figure 3.2.15(b), such as NAD binding and factors related to bladder cancer. This indicates different statistical algorithms from various software cause different weighted result in differential expression analysis,

and it should be carefully taken into consideration before the conclusion. Definitely, good experimental design for qualified sample preparation before RNA-Seq fundamentally determines the quality of results. Nevertheless, improvement of statistical and computational analyses according to sample diversity and quality is necessary for data analysis under challenging biological conditions such as the patient data.
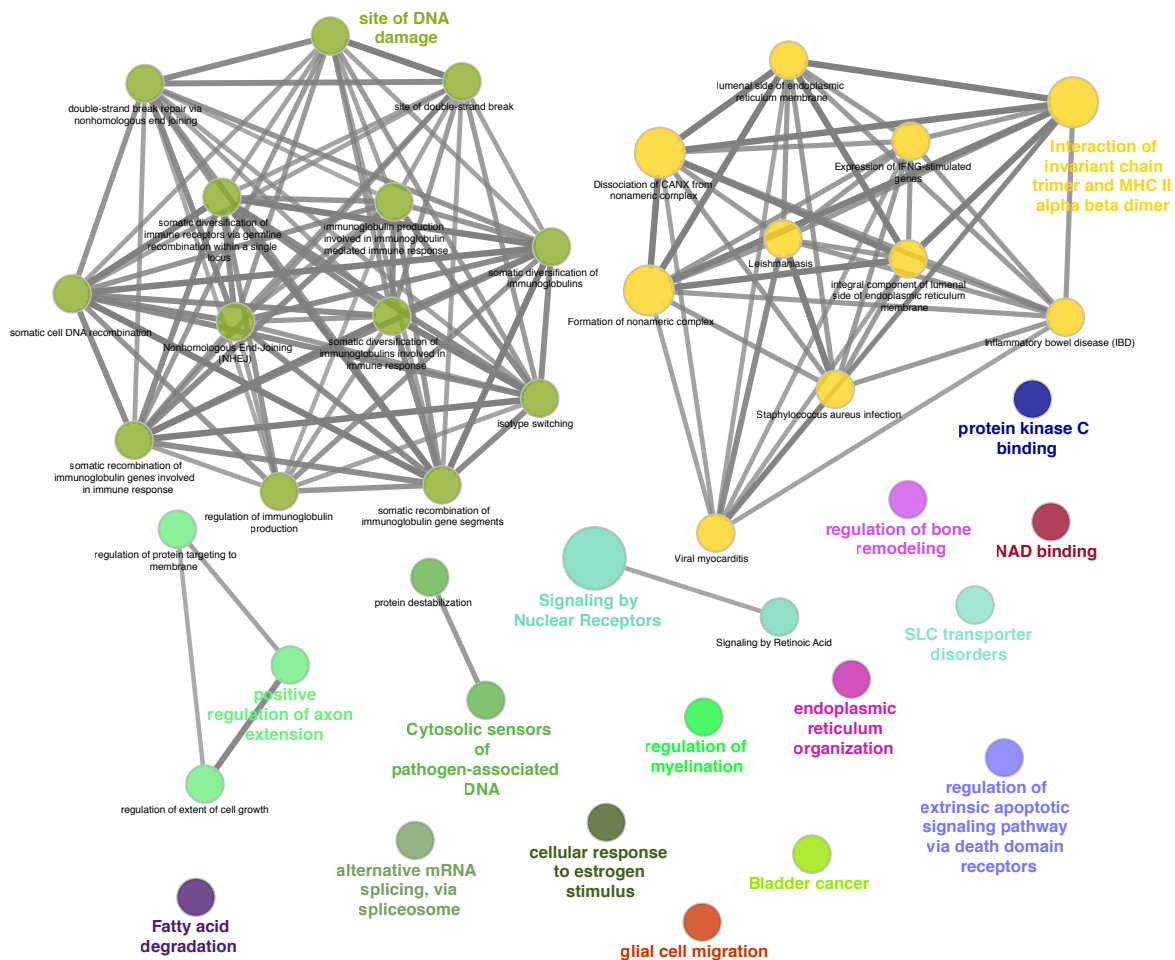
Figure 4.2.1: Results from DESeq2 in $p \leq 0.05$ and *padj* ≤ 0.1

4.2.2 Competitors or cooperators? *GC-Ct* interaction

Compared to the analysis of the host results based on the sophisticated dataset, results of *GC* present a more clear response in ion and amino acid transportation, DNA replication and modification, RNS/ROS system, and ferric uptake. Although how are the influences exactly involved by other organisms, *GC* trends to be adapted in the host niche when *Ct* is available. Are these pathogens involved competitors or cooperators? Theoretically, almost all the STI pathogens especially intracellular parasites like *Chlamydia* are competitors to

each other, for the limited resources offered by the host cell. Also, the pathogen communities are supposed to reach a status of homeostasis with the host so that they could persistently infect and survive in the host under the protection of the immune system. For host-pathogen interaction, pathogens would not just gain resources from the host, but also try to regulate host cell growth and apoptosis as described. Therefore, there should be some strategies for pathogens treat with the other pathogens in the niche to avoid to break the counterbalance among different pathogens and the host. In this case, the appearance of *Ct* prefers to cooperate with *GC*, rather than to be defended by *GC*. *GC* increases the regulation of many transporters in Ct^+ group for the competition of the ion and amino acid imported from the host. As a natural competent, *GC* alters adaptive regulation for DNA modification and stress response. Similar like the host, nitrosative and oxidative stresses are also rising, and the strategy of *GC* to maintain cellular redox homeostasis is the enhancement of reducing systems to respond to RNS/ROS, which is beneficial for both the host and *GC* itself.

4.2.3 Transferrin uptake in host-*GC*-*Ct* interaction

Another featured differentiation of *GC* in *Ct*-coinfected group is the regulation of ferric iron uptake. It is very clear that *GC* could take host iron compounds, such as transferrin, lactoferrin, and haemoglobin (Wooldridge and Williams, 1993). As an obligate but not strictly intracellular human pathogen, *GC* also gains exogenous siderophore-mediated iron acquisition (Schryvers and Stojiljkovic, 1999, West and Sparling, 1985). The iron uptake mechanisms of *GC* is well studied and used for vaccine designs (Cornelissen, 2008). Not only *GC* but also other pathogens must acquire iron for surviving and replicating in the host cell. What revealed in Ct^+ group is that, *GC* additionally increase the Fur activator *arsR*, the bacterioferritin B (*bfrB*), and the hemolysin related channel protein. These responses facilitate the acquisition of ferrin. It could either be the competition of the limited ferric in the host cell with other pathogens or promotion of host iron imported outside of the cell, which is beneficial for other pathogens.

The iron-associated response also exhibits in the host site. The basic strategies for host defence to the pathogen infection is to suppress nutrients uptake in order to prevent the outgrowth of the pathogen (Skaar, 2010), and one approach is targeting TfR-mediated iron acquisition. Transferrin receptor (TfR) is the janitor controlling the iron uptake in most of the cells and regulating cellular iron homeostasis (Gammella et al., 2017). Also, transferrin is reported required for *Ct*'s growth (Ouellette and Carabeo, 2010). Normally when under infection, host cells reduce the TfR's expression in order to offer less iron for pathogens surviving (Wooldridge and Williams, 1993). However, the TfR-encoding tran-

script of TFRC is strongly upregulated (Table 3.2.4) in the patient results, which seems to enrich the iron in the intracellular environment. The accumulation of iron actually induces the oxidative stress in the cell via the Haber-Weiss reaction, increasing $\cdot OH$ redoxed from H_2O_2 and $\cdot O_2^-$ (Haber and Weiss, 1932, Koppenol, 2001).

Unlike *GC*, *Ct* lacks the Fur regulator for iron metabolism, whereas it encodes a series of genes for heme biosynthesis pathways in its degenerate genome. Heme could be used to maintain iron storage. Heme biosynthesis starts with the utilization of glutamate to form uroporphyrin and Heme (Figure 4.2.2). However, this pathway in both the host and *GC* recruits glycine instead of glutamate. Proteomics data (Østergaard et al., 2016) shows this pathway is slightly active.
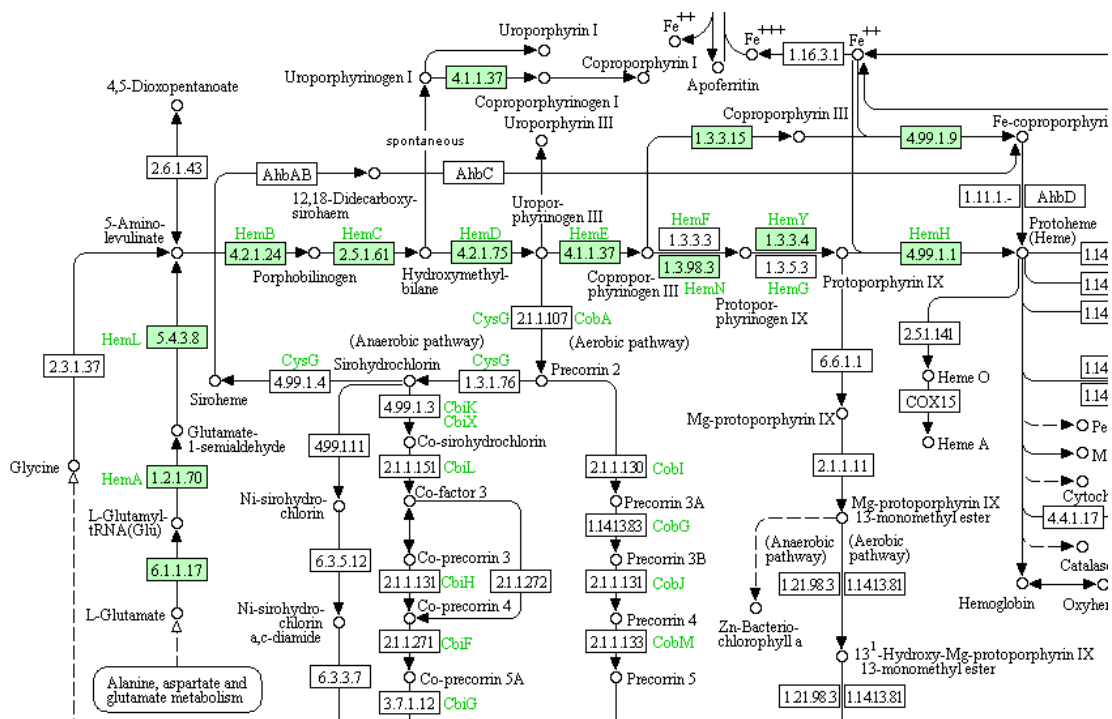
Figure 4.2.2: *Ct*'s heme biosynthesis, screenshot from KEGG database (KEGG).

4.2.3.1 Potential ferroptosis

Iron homeostasis is studied based on different pathogens, for the performances of various iron acquisition and maintenance systems (Wooldridge and Williams, 1993). The mechanisms for both Gram-positive and Gram-negative bacteria had been well described (Brown and Holden, 2002, Cornelissen and Sparling, 1994, Schryvers and Stojiljkovic, 1999). However, how multiple pathogens coinfected together mediate iron homeostasis is complicated and still unclear.

As is discussed, GC expressed more hemolysins in Ct positive group for the lysis red blood cells which are rich for iron-containing heme. GC promotes host iron uptake via transferrin, but the purpose is unsure for iron competition or assistance of iron acquisition by Ct. Generally, the existence of other pathogens should not be forgotten in the natural infection. Whenever Ct is coinfected or not, GC definitely under a status of iron competition with other pathogens. Thus, the iron-uptake promotion by GC specially presented in Ct^+ is quite reasonable for cooperation rather than competition. Accordingly, a programmed cell death called ferroptosis may be a potential strategy for pathogen cooperatively mediate host cell death (Figure 4.2.3).
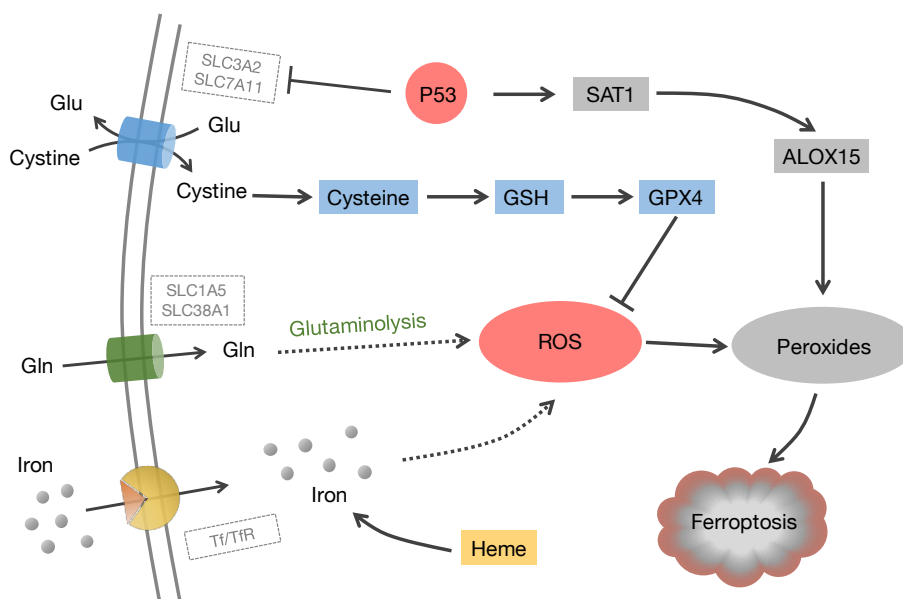
Figure 4.2.3: Schematic diagram of ferroptosis (source: own figure)

Ferroptosis is a caspase-independent but iron-dependent cell death resulting from the accumulation of oxidized lipids (Gnanapradeepan et al., 2018). The sign of ferroptosis is upregulation of cellular tumour antigen p53 (P53) and down-regulation of glutathione (GSH) accompanied by glutaminolysis and upregulation of TfR (Gao et al., 2015). In this work, the host TfR shows significant upregulation when Ct is available, and Ct participated in P53 regulation and potentially would utilize myc-HUWE1 and p53-hdm2 systems to mediate cell apoptosis. Moreover, although no evidence reported Ct induced host glutaminolysis, the necessity of glutamate in Ct' metabolism should be converted from glutamine imported from the host cell, thereby potentially resulting in glutaminolysis. It is challenging to prove the existence of ferroptosis for it may only happen in the coinfected environment when pathogens have cooperations in iron acquisition. Also, pathogens may be able to induce ferroptosis, however, the host may finally win the game, wf6phich makes the detection of ferroptosis more difficult.

4.3 Prospects

In this work, metabolic modelling and flux analysis are initially used to study the biphasic variation of *Ct* metabolism in the scope of the quantitative pathway during different hours post infection. And the differential expression analysis from the patient metatranscriptome provides potential targets interacted with *Ct* infection from both the host and *GC*. *Ct* metabolism is reliant on the host condition. Coinfection study brings a more natural infected environment and response compared to the typical study focused on only one pathogen. Patient metatranscriptome and differential expression analysis persuade coinfection study to be possible. Improvement of data precision in the metatranscriptomics is a challenge in the long-term future. Due to the specificity of the patient data, large amounts of datasets are necessary for statistical analysis. As discussed, pathogenic gene expression varies in different host cell lines. Thus, comparison between the *Ct*'s gene expression from the patient metatranscriptome and from the *Ct*-infected cell line would clearly show the limitations relying on cell lines. For host and pathogen interaction, more appropriate material instead of cell lines required if the patient data is not available. Using 3D human tissue models is an applicable approach and allows us to control compared to patient data. This is also challenging but will develop new strategies in clinical therapy for infection treatment. The upcoming study of coinfection could be initiated with two pathogens (e.g. *Ct* and *GC*) by infecting 3D human tissue model and accompany with RNA sequencing or single-cell sequencing.

Coinfection study offers a more comprehensive and reliable way to understand both the pathogenic metabolism and the host apoptosis. For example, Siegl et al. (2014) reported the degradation of p53 via the PI3K-Akt-signalling pathway activated by *Chlamydia*. However, the regulation of myc-HUWE1 to host apoptosis affected by *Ct* infection and its relationship with p53-hdm2 have not studied yet. Also, metabolic modelling pointed out the demand for host acyl-CoA for *Ct* metabolic surviving. Thus, it explains the host fatty acid β oxidation is the reasonable cause but not the result of the host glycolysis deficiency. Not only host-pathogen interaction but also pathogen-pathogen interaction should be noticed. Although the *GC*'s response to *Ct* infection has been discussed in this work, the influence of other microorganisms involved needs further discussion. For example, the Gram-negative pathogen *Sneathia* is less characterized but dominant in the female reproductive tract. Manhart et al. (2013) reported the *Sneathia* may also be a urethral pathogen in male urethritis. Similar to the *Ct-GC* interaction discussed in this work, interactions among more pathogens are worth further research.

References

Yasser M Abdelrahman and Robert J Belland. The chlamydial developmental cycle. *FEMS microbiology reviews*, 29(5):949–959, 2005.

Sovana Adhikary, Federica Marinoni, Andreas Hock, Esther Hulleman, Nikita Popov, Rudi Beier, Sandra Bernard, Micaela Quarto, Maria Capra, Stephan Goettig, et al. The ubiquitin ligase hecth9 regulates transcriptional activation by myc and is essential for tumor cell proliferation. *Cell*, 123(3):409–421, 2005.

Shiva Akhavantabasi, Hesna B Akman, Aysegul Sapmaz, Jennifer Keller, Elizabeth M Petty, and Ayse E Erson. Usp32 is an active, membrane-bound ubiquitin protease over-expressed in breast cancers. *Mammalian genome*, 21(7-8):388–397, 2010.

Catharina Johanna Alberts, Maarten F Schim van der Loeff, Mary R Papenfuss, Roberto José Carvalho da Silva, Luisa Lina Villa, Eduardo Lazcano-Ponce, Alan G Nyitray, and Anna R Giuliano. Association of chlamydia trachomatis infection and herpes simplex virus type 2 serostatus with genital human papillomavirus infection in men: the him study. *Sexually transmitted diseases*, 40(6):508, 2013.

Marco Albrecht, Cynthia M Sharma, Richard Reinhardt, Joerg Vogel, and Thomas Rudel. Deep sequencing-based discovery of the chlamydia trachomatis transcriptome. *Nucleic acids research*, 38(3):868–877, 2009.

N Allende-Vega, S Dayal, U Agarwala, A Sparks, JC Bourdon, and MK Saville. p53 is activated in response to disruption of the pre-mrna splicing machinery. *Oncogene*, 32(1): 1, 2013.

Simon Anders and Wolfgang Huber. Differential expression analysis for sequence count data. *Genome biology*, 11(10):R106, 2010.

Simon Anders, Davis J McCarthy, Yunshun Chen, Michal Okoniewski, Gordon K Smyth, Wolfgang Huber, and Mark D Robinson. Count-based differential expression analysis of rna sequencing data using r and bioconductor. *Nature protocols*, 8(9):1765, 2013.

Simon Andrews et al. Fastqc: a quality control tool for high throughput sequence data. 2010.

Tarja Anttila, Pekka Saikku, Pentti Koskela, Aini Bloigu, Joakim Dillner, Irma Ikäheimo, Egil Jellum, Matti Lehtinen, Per Lenner, Timo Hakulinen, et al. Serotypes of chlamydia trachomatis and risk for development of cervical squamous cell carcinoma. *Jama*, 285 (1):47–51, 2001.

References

Isabelle S Arts, Alexandra Gennaris, and Jean-François Collet. Reducing systems protecting the bacterial cell envelope from oxidative damage. *FEBS letters*, 589(14):1559–1568, 2015.

Babraham_Bioinformatics.

Kurt Bachmaier, Nikolaus Neu, M Luis, Sukumar Pal, Andrew Hessel, and Josef M Penninger. Chlamydia infections and heart disease linked through antigenic mimicry. *Science*, 283(5406):1335–1339, 1999.

Michelle Badura, Steve Braunstein, Jiri Zavadil, and Robert J Schneider. Dna damage and eif4g1 in breast cancer cells reprogram translation for survival and dna repair mrnas. *Proceedings of the National Academy of Sciences*, page 201203853, 2012.

Keren Baranes-Bachar, Adva Levy-Barda, Judith Oehler, Dylan A Reid, Isabel Soria-Bretones, Ty C Voss, Dudley Chung, Yoon Park, Chao Liu, Jong-Bok Yoon, et al. The ubiquitin e3/e4 ligase ube4a adjusts protein ubiquitylation and accumulation at sites of dna damage, facilitating double-strand break repair. *Molecular cell*, 69(5):866–878, 2018.

Byron E Batteiger, Wanzhu Tu, Susan Ofner, Barbara Van Der Pol, Diane R Stothard, Donald P Orr, Barry P Katz, and J Dennis Fortenberry. Repeated chlamydia trachomatis genital infections in adolescent women. *The Journal of infectious diseases*, 201(1):42–51, 2010.

Sam Behjati and Patrick S Tarpey. What is next generation sequencing? *Archives of Disease in Childhood-Education and Practice*, 98(6):236–238, 2013.

Robert J Belland, Guangming Zhong, Deborah D Crane, Daniel Hogan, Daniel Sturdevant, Jyotika Sharma, Wandy L Beatty, and Harlan D Caldwell. Genomic transcriptional profiling of the developmental cycle of chlamydia trachomatis. *Proceedings of the National Academy of Sciences*, 100(14):8478–8483, 2003.

Dennis A Benson, Mark Cavanaugh, Karen Clark, Ilene Karsch-Mizrachi, David J Lipman, James Ostell, and Eric W Sayers. Genbank. *Nucleic acids research*, 41(D1):D36–D42, 2012.

Sahely Bhadra and Juho Rousu. Analysis of fluxomic experiments with principal metabolic flux mode analysis. In *Data Mining for Systems Biology*, pages 141–161. Springer, 2018.

KS Bhat, CP Gibbs, O Barrera, SG Morrison, F Jähnig, A Stern, E-M Kupsch, TF Meyer, and J Swanson. The opacity proteins of neisseria gonorrhoeae strain ms11 are encoded by a family of 11 complete genes. *Molecular microbiology*, 5(8):1889–1901, 1991.

Gabriela Bindea, Bernhard Mlecnik, Hubert Hackl, Pornpimol Charoentong, Marie Tosolini, Amos Kirilovsky, Wolf-Herman Fridman, Franck Pagès, Zlatko Trajanoski, and Jérôme Galon. Cluego: a cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics*, 25(8):1091–1093, 2009.

Béla Bollobás. *Modern graph theory*, volume 184. Springer Science & Business Media, 2013.

Benjamin M Bolstad, Rafael A Irizarry, Magnus Åstrand, and Terence P. Speed. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*, 19(2):185–193, 2003.

Pamela T Bonar and Joseph R Casey. Plasma membrane cl-/hco3-exchangers: structure, mechanism and physiology. *Channels*, 2(5):337–345, 2008.

Silvia Bottini, Nedra Hamouda-Tekaya, Raphael Mategot, Laure-Emmanuelle Zaragosi, Stephane Audebert, Sabrina Pisano, Valerie Grandjean, Claire Mauduit, Mohamed Benahmed, Pascal Barbry, et al. Post-transcriptional gene silencing mediated by micrornas is controlled by nucleoplasmic sfpq. *Nature communications*, 8(1):1189, 2017.

Jim Braxton, Darlene W Davis, Brian Emerson, Elaine W Flagg, Jeremy Grey, LaZetta Grier, Alesia Harvey, Sarah Kidd, Jennifer Kim, Kristen Kreisel, et al. Sexually transmitted disease surveillance 2017. 2018.

Jeremy S Brown and David W Holden. Iron acquisition by gram-positive bacterial pathogens. *Microbes and infection*, 4(11):1149–1156, 2002.

James H Bullard, Elizabeth Purdom, Kasper D Hansen, and Sandrine Dudoit. Evaluation of statistical methods for normalization and differential expression in mrna-seq experiments. *BMC bioinformatics*, 11(1):94, 2010.

Haydar Bulut, Sebastien Moniot, Anke Licht, Frank Scheffel, Stephanie Gathmann, Wolfram Saenger, and Erwin Schneider. Crystal structures of two solute receptors for l-cystine and l-cysteine, respectively, of the human pathogen neisseria gonorrhoeae. *Journal of molecular biology*, 415(3):560–572, 2012.

Gerald I Byrne and David M Ojcius. Chlamydia and apoptosis: life and death decisions of an intracellular pathogen. *Nature Reviews Microbiology*, 2(10):802, 2004.

Bruno Canard and Robert S Sarfati. Dna polymerase fluorescent substrates with reversible 3 -tags. *Gene*, 148(1):1–6, 1994a.

Bruno Canard and Simon Sarfati. Novel derivatives usable for the sequencing of nucleic acids. *CA Patent*, 2:13, 1994b.

Kaleigh Canfield, Wendy Wells, Joseph Geradts, William B Kinlaw, Chao Cheng, and Manabu Kurokawa. Inverse association between mdm2 and huwe1 protein expression levels in human breast cancer and liposarcoma. *International journal of clinical and experimental pathology*, 9(6):6342, 2016.

Jose A Chavez and Scott A Summers. A ceramide-centric view of insulin resistance. *Cell metabolism*, 15(5):585–594, 2012.

Harrell W Chesson and Steven D Pinkerton. Sexually transmitted diseases and the increased risk for hiv transmission: implications for cost-effectiveness analyses of sexually transmitted disease prevention interventions. *Journal of acquired immune deficiency syndromes (1999)*, 24(1):48–56, 2000.

Donghui Choe, Richard Szubin, Samira Dahesh, Suhyung Cho, Victor Nizet, Bernhard Palsson, and Byung-Kwan Cho. Genome-scale analysis of methicillin-resistant staphylococcus aureus usa300 reveals a tradeoff between pathogenesis and drug resistance. *Scientific reports*, 8(1):2215, 2018.

Jordan L Cocchiaro, Yadunanda Kumar, Elizabeth R Fischer, Ted Hackstadt, and Raphael H Valdivia. Cytoplasmic lipid droplets are translocated into the lumen of the chlamydia trachomatis parasitophorous vacuole. *Proceedings of the National Academy of Sciences*, 105(27):9379–9384, 2008.

Peter JA Cock, Christopher J Fields, Naohisa Goto, Michael L Heuer, and Peter M Rice. The sanger fastq file format for sequences with quality scores, and the solexa/illumina fastq variants. *Nucleic acids research*, 38(6):1767–1771, 2009.

Peter JA Cock, James K Bonfield, Bastien Chevreux, and Heng Li. Sam/bam format v1. 5 extensions for de novo assemblies. *bioRxiv*, page 020024, 2015.

Matthew D Collins, Lesley Hoyles, Eva Törnqvist, Robert von Essen, and Enevold Falsen. Characterization of some strains from human clinical sources which resemble "leptotrichia sanguinegens": Description of sneathia sanguinegens sp. nov., gen. nov. *Systematic and applied microbiology*, 24(3):358–361, 2001.

Daniel F Comiskey Jr, Aishwarya G Jacob, Ravi K Singh, Aixa S Tapia-Santos, and Dawn S Chandler. Splicing factor srsf1 negatively regulates alternative splicing of mdm2 under damage. *Nucleic acids research*, 43(8):4202–4218, 2015.

Ana Conesa, Pedro Madrigal, Sonia Tarazona, David Gomez-Cabrero, Alejandra Cervera, Andrew McPherson, Michał Wojciech Szcześniak, Daniel J Gaffney, Laura L Elo, Xuegong Zhang, et al. A survey of best practices for rna-seq data analysis. *Genome biology*, 17(1):13, 2016.

International Human Genome Sequencing Consortium et al. Initial sequencing and analysis of the human genome. *Nature*, 409(6822):860, 2001.

CN Cornelissen. Identification and characterization of gonococcal iron transport systems as potential vaccine antigens. 2008.

Cynthia Nau Cornelissen and P Frederick Sparling. Iron piracy: acquisition of transferrin-bound iron by bacterial pathogens. *Molecular microbiology*, 14(5):843–850, 1994.

FEG Cox. Concomitant infections, parasites and immune responses. *Parasitology*, 122(S1): S23–S38, 2001.

Michael J Cox, William OCM Cookson, and Miriam F Moffatt. Sequencing the human microbiome in health and disease. *Human molecular genetics*, 22(R1):R88–R94, 2013.

Sarah Creighton, Melinda Tenant-Flowers, Christopher B Taylor, Rob Miller, and Nicola Low. Co-infection with gonorrhoea and chlamydia: how much is there and what does it mean? *International journal of STD & AIDS*, 14(2):109–113, 2003.

Francis HC Crick. On protein synthesis. In *Symp Soc Exp Biol*, volume 12, page 8, 1958.

Daniel A Cuevas, Janaka Edirisinghe, Chris S Henry, Ross Overbeek, Taylor G O'Connell, and Robert A Edwards. From dna to fba: how to build your own genome-scale metabolic model. *Frontiers in microbiology*, 7:907, 2016.

Mu-Shui Dai, Yetao Jin, Jayme R Gallegos, and Hua Lu. Balance of yin and yang: ubiquitylation-mediated regulation of p53 and c-myc. *Neoplasia*, 8(8):630–644, 2006.

Roger Detels, Annette M Green, Jeffrey D Klausner, David Katzenstein, Charlotte Gaydos, H Hunter Handsfield, Willo Pequegnat, Kenneth Mayer, Tyler D Hartwell, and Thomas C Quinn. The incidence and correlates of symptomatic and asymptomatic chlamydia trachomatis and neisseria gonorrhoeae infections in selected populations in five countries. *Sexually transmitted diseases*, 38(6):503, 2011.

dgelsin and alyssafrazee.

Veena Dhawan. Reactive oxygen and nitrogen species: general considerations. In *Studies on respiratory disorders*, pages 27–47. Springer, 2014.

Marie-Agnès Dillies, Andrea Rau, Julie Aubert, Christelle Hennequet-Antier, Marine Jeanmougin, Nicolas Servant, Céline Keime, Guillemette Marot, David Castel, Jordi Estelle, et al. A comprehensive evaluation of normalization methods for illumina high-throughput rna sequencing data analysis. *Briefings in bioinformatics*, 14(6):671–683, 2013.

eclark28. Question: Cuffdiff bias towards positive fold changes, 2017.

Jennifer L Edwards and Michael A Apicella. The molecular mechanisms used by neisseria gonorrhoeae to initiate infection differ between men and women. *Clinical microbiology reviews*, 17(4):965–981, 2004.

Cherilyn Elwell, Kathleen Mirrashidi, and Joanne Engel. Chlamydia cell biology and pathogenesis. *Nature Reviews Microbiology*, 14(6):385, 2016.

Frank Emmert-Streib, Matthias Dehmer, and Benjamin Haibe-Kains. Gene regulatory networks and their applications: understanding biological and medical problems in terms of networks. *Frontiers in cell and developmental biology*, 2:38, 2014.

Leonhard Euler. Solutio problematis ad geometriam situs pertinentis. *Commentarii Academiae Scientiarum Imperialis Petropolitanae*, 8:128–140, 1736.

Benjamin Ezraty, Alexandra Gennaris, Frédéric Barras, and Jean-François Collet. Oxidative stress, protein damage and repair in bacteria. *Nature Reviews Microbiology*, 15(7): 385, 2017.

Huizhou Fan and Guangming Zhong. Chlamydia trachomatis. In *Molecular Medical Microbiology (Second Edition)*, pages 1449–1MDM469. Elsevier, 2015.

Jerad Fields, Jesse J Hanisch, Jeong W Choi, and Paul M Hwang. How does p53 regulate mitochondrial respiration? *IUBMB life*, 59(10):682–684, 2007.

Annette Fischer, Kelly S Harrison, Yesid Ramirez, Daniela Auer, Suvagata Roy Chowdhury, Bhupesh K Prusty, Florian Sauer, Zoe Dimond, Caroline Kisker, P Scott Hefty, et al. Chlamydia trachomatis-containing vacuole serves as deubiquitination platform to stabilize mcl-1 and to interfere with host defense. *Elife*, 6:e21465, 2017.

Jason P Folster and William M Shafer. Regulation of mtrf expression in neisseria gonorrhoeae and its role in high-level antimicrobial resistance. *Journal of bacteriology*, 187 (11):3713–3720, 2005.

JM Fraile, V Quesada, David Rodríguez, JMP Freije, and Carlos López-Otín. Deubiquitinases in cancer: new functions and therapeutic options. *Oncogene*, 31(19):2373, 2012.

Christof Francke, Roland J Siezen, and Bas Teusink. Reconstructing the metabolic network of a bacterium from its genome. *Trends in microbiology*, 13(11):550–558, 2005.

Alyssa C Frazee, Geo Pertea, Andrew E Jaffe, Ben Langmead, Steven L Salzberg, and Jeffrey T Leek. Ballgown bridges the gap between transcriptome assembly and expression analysis. *Nature biotechnology*, 33(3):243, 2015.

Alexandra Friedrich, Cindy G Arvidson, William M Shafer, Eun-Hee Lee, and Magdalene So. Two abc transporter operons and the antimicrobial resistance gene mtrf are pilt responsive in neisseria gonorrhoeae. *Journal of bacteriology*, 189(14):5399–5402, 2007.

Elena Gammella, Paolo Buratti, Gaetano Cairo, and Stefania Recalcati. The transferrin receptor: the cellular iron gate. *Metallomics*, 9(10):1367–1375, 2017.

Minghui Gao, Prashant Monian, Nosirudeen Quadri, Ravichandran Ramasamy, and Xuejun Jiang. Glutaminolysis and transferrin regulate ferroptosis. *Molecular cell*, 59(2): 298–308, 2015.

R Gentleman, V Carey, W Huber, and F Hahne. Genefilter: methods for filtering genes from high-throughput experiments. *R package version*, 1(1), 2015.

Robert C Gentleman, Vincent J Carey, Douglas M Bates, Ben Bolstad, Marcel Dettling, Sandrine Dudoit, Byron Ellis, Laurent Gautier, Yongchao Ge, Jeff Gentry, et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome biology*, 5(10):R80, 2004.

Davide Gibellini, Roberto Panaya, and Fabio Rumpianesi. Induction of apoptosis by chlamydia psittaci and chlamydia trachomatis infection in tissue culture cells. *Zentralblatt für Bakteriologie*, 288(1):35–43, 1998.

CC Ginocchio, K Chapin, JS Smith, J Aslanzadeh, J Snook, CS Hill, and CA Gaydos. Prevalence of trichomonas vaginalis and coinfection with chlamydia trachomatis and neisseria gonorrhoeae in the united states as determined by the aptima® trichomonas vaginalis nucleic acid amplification assay. *Journal of clinical microbiology*, pages JCM–00748, 2012.

Luciana E Giono, Nicolás Nieto Moreno, Adrián E Cambindo Botto, Gwendal Dujardin, Manuel J Muñoz, and Alberto R Kornblihtt. The rna response to dna damage. *Journal of molecular biology*, 428(12):2636–2651, 2016.

Keerthana Gnanapradeepan, Subhasree Basu, Thibaut Barnoud, Anna Budina-Kolomets, Che-Pei Kung, and Maureen E Murphy. The p53 tumor suppressor in the control of metabolism and ferroptosis. *Frontiers in endocrinology*, 9:124, 2018.

L Goff, C Trapnell, and D Kelley. cummerbund: Analysis, exploration, manipulation, and visualization of cufflinks high-throughput sequencing data. *R package version*, 2(0), 2013.

Richard J Grainger and Jean D Beggs. Prp8 protein: at the heart of the spliceosome. *Rna*, 11(5):533–557, 2005.

J Thomas Grayston, Carl H Mordhorst, and San-pin Wang. Childhood myocarditis associated with chlamydia trachomatis infection. *JAMA*, 246(24):2823–2827, 1981.

The SAM/BAM Format Specification Working Group. Sequence alignment/map format specification.

Nitish Gulve, Bhupesh K Prusty, and Thomas Rudel. Chlamydia trachomatis impairs host base excision repair by downregulating polymerase beta. *Cellular microbiology*, page e12986, 2018.

F Haber and J Weiss. Über die katalyse des hydroperoxydes. *Naturwissenschaften*, 20(51): 948–950, 1932.

Catherine L Haggerty, Sami L Gottlieb, Brandie D Taylor, Nicola Low, Fujie Xu, and Roberta B Ness. Risk of sequelae after chlamydia trachomatis genital infection in women. *The Journal of infectious diseases*, 201(Supplement_2):S134–S155, 2010.

Simon R Harris, Ian N Clarke, Helena MB Seth-Smith, Anthony W Solomon, Lesley T Cutcliffe, Peter Marsh, Rachel J Skilton, Martin J Holland, David Mabey, Rosanna W Peeling, et al. Whole-genome analysis of diverse chlamydia trachomatis strains identifies phylogenetic relationships masked by current clinical typing. *Nature genetics*, 44(4):413, 2012.

Jennifer Harrow, Adam Frankish, Jose M Gonzalez, Electra Tapanari, Mark Diekhans, Felix Kokocinski, Bronwen L Aken, Daniel Barrell, Amonida Zadissa, Stephen Searle, et al. Gencode: the reference human genome annotation for the encode project. *Genome research*, 22(9):1760–1774, 2012.

Michael D Harwich, Myrna G Serrano, Jennifer M Fettweis, João MP Alves, Mark A Reimers, Gregory A Buck, and Kimberly K Jefferson. Genomic sequence analysis and characterization of sneathia amnii sp. nov. *BMC genomics*, 13(8):S4, 2012.

Yu He, Dan Zhang, Yanfang Yang, Xixi Wang, Xinyu Zhao, Peng Zhang, Hongxia Zhu, Ningzhi Xu, and Shufang Liang. A double-edged function of ddx3, as an oncogene or tumor suppressor, in cancer progression. *Oncology reports*, 39(3):883–892, 2018.

James M Heather and Benjamin Chain. The sequence of sequencers: the history of sequencing dna. *Genomics*, 107(1):1–8, 2016.

Brian Henderson, Michael Wilson, Rod McNab, and Alistair J Lax. *Cellular Microbiology: bacteria-host interactions in health and disease*, volume 1. John Wiley & Sons Ltd., 1999.

Andreas J Hörlein, Anders M Näär, Thorsten Heinzel, Joseph Torchia, Bernd Gloss, Riki Kurokawa, Aimee Ryan, Yasutomi Kamei, Mats Söderström, Christopher K Glass, et al. Ligand-independent repression by the thyroid hormone receptor mediated by a nuclear receptor co-repressor. *Nature*, 377(6548):397, 1995.

Matthias Horn, Astrid Collingro, Stephan Schmitz-Esser, Cora L Beier, Ulrike Purkhold, Berthold Fartmann, Petra Brandt, Gerald J Nyakatura, Marcus Droege, Dmitrij Frishman, et al. Illuminating the evolutionary history of chlamydiae. *Science*, 304(5671): 728–730, 2004.

Yu-Feng Huang, Sheng-Chung Chen, Yih-Shien Chiang, Tzu-Han Chen, and Kuo-Ping Chiu. Palindromic sequence impedes sequencing-by-ligation mechanism. In *BMC systems biology*, volume 6, page S10. BioMed Central, 2012.

Michael Hucka, Andrew Finney, Herbert M Sauro, Hamid Bolouri, John C Doyle, Hiroaki Kitano, Adam P Arkin, Benjamin J Bornstein, Dennis Bray, Athel Cornish-Bowden, et al. The systems biology markup language (sbml): a medium for representation and exchange of biochemical network models. *Bioinformatics*, 19(4):524–531, 2003.

Emma R Iliffe-Lee and Grant McClarty. Regulation of carbon metabolism in chlamydia trachomatis. *Molecular microbiology*, 38(1):20–30, 2000.

I Illumina. An introduction to next-generation sequencing technology. 2015.

Inc Illumina. History of illumina sequencing.

Inc Illumina. Illumina to acquire pacific biosciences, 2018.

Isabelle Iost, Thierry Bizebard, and Marc Dreyfus. Functions of dead-box proteins in bacteria: current knowledge and pending questions. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, 1829(8):866–877, 2013.

Joanna Y Ip, Dominic Schmidt, Qun Pan, Arun K Ramani, Andrew G Fraser, Duncan T Odom, and Benjamin J Blencowe. Global impact of rna polymerase ii elongation inhibition on alternative splicing regulation. *Genome research*, 21(3):390–401, 2011.

Lydgia A Jackson, Thomas F Ducey, Michael W Day, Jeremy B Zaitshik, Joshua Orvis, and David W Dyer. Transcriptional and functional analysis of the neisseria gonorrhoeae fur regulon. *Journal of bacteriology*, 192(1):77–85, 2010.

Stephen P Jackson and Jiri Bartek. The dna-damage response in human biology and disease. *Nature*, 461(7267):1071, 2009.

Xavier Jacq, Mark Kemp, Niall MB Martin, and Stephen P Jackson. Deubiquitylating enzymes and dna damage response pathways. *Cell biochemistry and biophysics*, 67(1): 25–43, 2013.

Heinrich Jung. The sodium/substrate symporter family: structural and functional features. *FEBS letters*, 529(1):73–77, 2002.

Nadja Käding, Márta Szaszák, and Jan Rupp. Imaging of chlamydia and host cell metabolism. *Future microbiology*, 9(4):509–521, 2014.

Richard H Kahn, Debra J Mosure, Susan Blank, Charlotte K Kent, Joan M Chow, Melina R Boudov, Jeffrey Brock, Scott Tulloch, et al. Chlamydia trachomatis and neisseria gonorrhoeae prevalence and coinfection in adolescents entering selected us juvenile detention centers, 1997–2002. *Sexually transmitted diseases*, 32(4):255–259, 2005.

Christoph Kaleta, Florian Centler, and Peter Dittrich. Analyzing molecular reaction networks. *Molecular biotechnology*, 34(2):117–123, 2006.

Axel von Kamp and Stefan Schuster. Metatool 5.0: fast and flexible elementary modes analysis. *Bioinformatics*, 22(15):1930–1931, 2006.

Minoru Kanehisa and Susumu Goto. Kegg: kyoto encyclopedia of genes and genomes. *Nucleic acids research*, 28(1):27–30, 2000.

Shih-Han Kao, Han-Tsang Wu, and Kou-Juey Wu. Ubiquitination by huwe1 in tumorigenesis and beyond. *Journal of biomedical science*, 25(1):67, 2018.

Karthika Karunakaran, Prema Subbarayal, Nadine Vollmuth, and Thomas Rudel. Chlamydia-infected cells shed gp 96 to prevent chlamydial re-infection. *Molecular microbiology*, 98(4):694–711, 2015a.

Karuna P Karunakaran, Hong Yu, Xiaozhou Jiang, Queenie Chan, Kyung-Mee Moon, Leonard J Foster, and Robert C Brunham. Outer membrane proteins preferentially load mhc class ii peptides: implications for a chlamydia trachomatis t cell vaccine. *Vaccine*, 33(18):2159–2166, 2015b.

Andrew Keat, Josh Dixey, Chris Sonnex, Brenda Thomas, Mary Osborn, and David Taylor-Robinson. Chlamydia trachomatis and reactive arthritis: the missing link. *The Lancet*, 329(8524):72–74, 1987.

KEGG.

Peter H Kilmarx, Philip A Mock, and William C Levine. Effect of chlamydia trachomatis coinfection on hiv shedding in genital tract secretions. *Sexually transmitted diseases*, 28 (6):347–348, 2001.

Daehwan Kim, Ben Langmead, and Steven L Salzberg. Hisat: a fast spliced aligner with low memory requirements. *Nature methods*, 12(4):357, 2015.

Man S Kim, Huan Zhang, and Won Bo Shim. Application of game theory to explore the dynamics of host- pathogen association in phytobiomes. *Phytobiomes*, 2(3):111–116, 2018.

Joshua Kimani, Ian W Maclean, Job J Bwayo, Kelly MacDonald, Julius Oyugi, Gregory M Maitha, Rosanna W Peeling, Mary Cheang, Nicolaas JD Nagelkerke, Francis A Plummer, et al. Risk factors for chlamydia trachomatis pelvic inflammatory disease among sex workers in nairobi, kenya. *Journal of Infectious Diseases*, 173(6):1437–1444, 1996.

Aviram Kogot-Levin and Ann Saada. Ceramide and the mitochondrial respiratory chain. *Biochimie*, 100:88–94, 2014.

Raivo Kolde. Pheatmap: pretty heatmaps. *R package version*, 61, 2012.

Lena König, Alexander Siegl, Thomas Penz, Susanne Haider, Cecilia Wentrup, Julia Polzin, Evelyne Mann, Stephan Schmitz-Esser, Daryl Domman, and Matthias Horn. Biphasic metabolism and host interaction of a chlamydial symbiont. *MSystems*, 2(3):e00202–16, 2017.

Willem H Koppenol. The haber-weiss cycle–70 years later. *Redox Report*, 6(4):229–234, 2001.

Pentti Koskela, Tarja Anttila, Tone Bjørge, Anne Brunsvig, Joakim Dillner, Matti Hakama, Timo Hakulinen, Egil Jellum, Matti Lehtinen, Per Lenner, et al. Chlamydia trachomatis infection as a risk factor for invasive cervical cancer. *International journal of cancer*, 85 (1):35–39, 2000.

Guido Kroemer and Jacques Pouyssegur. Tumor cell metabolism: cancer's achilles' heel. *Cancer cell*, 13(6):472–482, 2008.

Yadunanda Kumar, Jordan Cocchiaro, and Raphael H Valdivia. The obligate intracellular pathogen chlamydia trachomatis targets host lipid droplets. *Current biology*, 16(16): 1646–1651, 2006.

Cho-chou Kuo, Noriko Takahashi, Albertina F Swanson, Yasuhiro Ozeki, and Sen-itiroh Hakomori. An n-linked high-mannose type oligosaccharide, expressed at the major outer membrane protein of chlamydia trachomatis, mediates attachment and infectivity of the microorganism to hela cells. *The Journal of clinical investigation*, 98(12):2813–2818, 1996.

Manabu Kurokawa, Jiyeon Kim, Joseph Geradts, Kenkyo Matsuura, Liu Liu, Xu Ran, Wenle Xia, Thomas J Ribar, Ricardo Henao, Mark W Dewhirst, et al. A network of substrates of the e3 ubiquitin ligases mdm2 and huwe1 control apoptosis independently of p53. *Sci. Signal.*, 6(274):ra32–ra32, 2013.

Ben Langmead and Steven L Salzberg. Fast gapped-read alignment with bowtie 2. *Nature methods*, 9(4):357, 2012.

Heidi Ledford. The death of microarrays?, 2008.

LA Lewis, AF Gillaspy, RE McLaughlin, M Gipson, T Ducey, T Ownbey, K Hartman, C Nydick, M Carson, J Vaughn, et al. The gonococcal genome sequencing project. *unpublished*.

Heng Li, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, and Richard Durbin. The sequence alignment/map format and samtools. *Bioinformatics*, 25(16):2078–2079, 2009.

GW Liechti, E Kuru, E Hall, A Kalinda, YV Brun, M VanNieuwenhze, and AT Maurelli. A new metabolic cell-wall labelling method reveals peptidoglycan in chlamydia trachomatis. *Nature*, 506(7489):507, 2014.

David J Lipman and William R Pearson. Rapid and sensitive protein similarity searches. *Science*, 227(4693):1435–1441, 1985.

Lin Liu, Yinhu Li, Siliang Li, Ni Hu, Yimin He, Ray Pong, Danni Lin, Lihua Lu, and Maggie Law. Comparison of next-generation sequencing systems. *BioMed Research International*, 2012, 2012.

Rui-Yan Liu, Cai-Feng Diao, Yi Zhang, Nan Wu, Hai-Ying Wan, Xiang-Yang Nong, Min Liu, and Hua Tang. mir-371-5p down-regulates pre mrna processing factor 4 homolog b (prpf4b) and facilitates the g1/s transition in human hepatocellular carcinoma cells. *Cancer letters*, 335(2):351–360, 2013.

Wendy P Loomis and Michael N Starnbach. T cell responses to chlamydia trachomatis. *Current opinion in microbiology*, 5(1):87–91, 2002.

Michael I Love, Wolfgang Huber, and Simon Anders. Moderated estimation of fold change and dispersion for rna-seq data with deseq2. *Genome biology*, 15(12):550, 2014.

Lisa E Manhart, Christine M Khosropour, Congzhu Liu, Catherine W Gillespie, Kevin Depner, Tina Fiedler, Jeanne M Marrazzo, and David N Fredricks. Bacterial vaginosis–associated bacteria in men: Association of leptotrichia/sneathia spp. with nongonococcal urethritis. *Sexually transmitted diseases*, 40(12):944, 2013.

Marcel Margulies, Michael Egholm, William E Altman, Said Attiya, Joel S Bader, Lisa A Bemben, Jan Berka, Michael S Braverman, Yi-Ju Chen, Zhoutao Chen, et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, 437(7057):376, 2005.

Allan M Maxam and Walter Gilbert. A new method for sequencing dna. *Proceedings of the National Academy of Sciences*, 74(2):560–564, 1977.

Elie Maza. In papyro comparison of tmm (edger), rle (deseq2), and mrn normalization methods for a simple two-conditions-without-replicates rna-seq experimental design. *Frontiers in genetics*, 7:164, 2016.

William McLaren, Bethan Pritchard, Daniel Rios, Yuan Chen, Paul Flicek, and Fiona Cunningham. Deriving the consequences of genomic variants with the ensembl api and snp effect predictor. *Bioinformatics*, 26(16):2069–2070, 2010.

J Mellies, J Jose, and TF Meyer. The neisseria gonorrhoeae gene ania encodes an inducible nitrite reductase. *Molecular and General Genetics MGG*, 256(5):525–532, 1997.

Huaiyu Mi, Xiaosong Huang, Anushya Muruganujan, Haiming Tang, Caitlin Mills, Diane Kang, and Paul D Thomas. Panther version 11: expanded annotation data from gene ontology and reactome pathways, and data analysis tool enhancements. *Nucleic acids research*, 45(D1):D183–D189, 2016.

Samuel S Minot, Niklas Krumm, and Nicholas B Greenfield. One codex: a sensitive and accurate data platform for genomic microbial identification. *BioRxiv*, page 027607, 2015.

Yoshihiko Miyata and Eisuke Nishida. Distantly related cousins of map kinase: biochemical properties and possible physiological functions. *Biochemical and biophysical research communications*, 266(2):291–295, 1999.

References

Ali Mortazavi, Brian A Williams, Kenneth McCue, Lorian Schaeffer, and Barbara Wold. Mapping and quantifying mammalian transcriptomes by rna-seq. *Nature methods*, 5(7): 621, 2008.

JAMES W Moulder. Interaction of chlamydiae and host cells in vitro. *Microbiological reviews*, 55(1):143–190, 1991.

Tracy L Nicholson, Karen Chiu, and Richard S Stephens. Chlamydia trachomatis lacks an adaptive response to changes in carbon source availability. *Infection and immunity*, 72 (7):4286–4289, 2004.

Kathleen Nudel, Ryan McClure, Matthew Moreau, Emma Briars, A. Jeanine Abrams, Brian Tjaden, Xiao-Hong Su, David Trees, Peter A. Rice, Paola Massari, and Caroline A. Genco. Transcriptome analysis of neisseria gonorrhoeae during natural infection reveals differential expression of antibiotic resistance determinants between men and women. *mSphere*, 3(3), 2018.

Matthew A Oberhardt, Bernhard Ø Palsson, and Jason A Papin. Applications of genome-scale metabolic reconstructions. *Molecular systems biology*, 5(1):320, 2009.

Anders Omsland, Janet Sager, Vinod Nair, Daniel E Sturdevant, and Ted Hackstadt. Developmental stage-specific metabolic and transcriptional activity of chlamydia trachomatis in an axenic medium. *Proceedings of the National Academy of Sciences*, 109 (48):19781–19785, 2012.

World Health Organization et al. Global who alliance for the elimination of blinding trachoma by 2020. *Weekly Epidemiological Record= Relevé épidémiologique hebdomadaire*, 87(17):161–168, 2012.

Jeffrey D Orth, Ines Thiele, and Bernhard Ø Palsson. What is flux balance analysis? *Nature biotechnology*, 28(3):245, 2010.

Alicia Oshlack and Matthew J Wakefield. Transcript length bias in rna-seq data confounds systems biology. *Biology direct*, 4(1):14, 2009.

Ole Østergaard, Frank Follmann, Anja W Olsen, Niels H Heegaard, Peter Andersen, and Ida Rosenkrands. Quantitative protein profiling of chlamydia trachomatis growth forms reveals defense strategies against tryptophan starvation. *Molecular & Cellular Proteomics*, pages mcp–M116, 2016.

Scot Ouellette and Rey Carabeo. A functional slow recycling pathway of transferrin is required for growth of chlamydia. *Frontiers in microbiology*, 1:112, 2010.

Fatih Ozsolak and Patrice M Milos. Rna sequencing: advances, challenges and opportunities. *Nature reviews genetics*, 12(2):87, 2011.

Jason L Parsons, Irina I Dianova, Svetlana V Khoronenkova, Mariola J Edelmann, Benedikt M Kessler, and Grigory L Dianov. Usp47 is a deubiquitylating enzyme that regulates base excision repair by controlling steady-state levels of dna polymerase β. *Molecular cell*, 41(5):609–615, 2011.

Geo Pertea and Andrey Rozenberg.

Mihaela Pertea, Geo M Pertea, Corina M Antonescu, Tsung-Cheng Chang, Joshua T Mendell, and Steven L Salzberg. Stringtie enables improved reconstruction of a transcriptome from rna-seq reads. *Nature biotechnology*, 33(3):290, 2015.

Mihaela Pertea, Daehwan Kim, Geo M Pertea, Jeffrey T Leek, and Steven L Salzberg. Transcript-level expression analysis of rna-seq experiments with hisat, stringtie and ballgown. *Nature protocols*, 11(9):1650, 2016.

A Peschiaroli, JR Skaar, M Pagano, and G Melino. The ubiquitin-specific protease usp47 is a novel *β*-trcp interactor regulating cell survival. *Oncogene*, 29(9):1384, 2010.

Thomas Pfeiffer, I Sánchez-Valdenebro, JC Nuño, Francisco Montero, and Stefan Schuster. Metatool: for studying metabolic networks. *Bioinformatics (Oxford, England)*, 15(3): 251–257, 1999.

Martin Pilhofer, Karin Aistleitner, Jacob Biboy, Joe Gray, Erkin Kuru, Edward Hall, Yves V Brun, Michael S VanNieuwenhze, Waldemar Vollmer, Matthias Horn, et al. Discovery of chlamydial peptidoglycan reveals bacteria with murein sacculi but without ftsz. *Nature communications*, 4:2856, 2013.

Sandra Placzek, Ida Schomburg, Antje Chang, Lisa Jeske, Marcus Ulbrich, Jana Tillack, and Dietmar Schomburg. Brenda in 2017: new perspectives and new tools in brenda. *Nucleic acids research*, page gkw952, 2016.

Michael A Quail, Miriam Smith, Paul Coupland, Thomas D Otto, Simon R Harris, Thomas R Connor, Anna Bertoni, Harold P Swerdlow, and Yong Gu. A tale of three next generation sequencing platforms: comparison of ion torrent, pacific biosciences and illumina miseq sequencers. *BMC genomics*, 13(1):341, 2012.

Sarah Jane Quillin and H Steven Seifert. Neisseria gonorrhoeae host adaptation and pathogenesis. *Nature Reviews Microbiology*, 16(4):226, 2018.

Aaron R Quinlan and Ira M Hall. Bedtools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6):841–842, 2010.

Krishnaraj Rajalingam, Manu Sharma, Christine Lohmann, Monique Oswald, Oliver Thieck, Christopher J Froelich, and Thomas Rudel. Mcl-1 is a key regulator of apoptosis resistance in chlamydia trachomatis-infected cells. *PloS one*, 3(9):e3102, 2008.

Karthik Raman and Nagasuma Chandra. Flux balance analysis of biological systems: applications and challenges. *Briefings in bioinformatics*, 10(4):435–449, 2009.

Jacques Ravel, Pawel Gajer, Zaid Abdo, G Maria Schneider, Sara SK Koenig, Stacey L McCulle, Shara Karlebach, Reshma Gorle, Jennifer Russell, Carol O Tacket, et al. Vaginal microbiome of reproductive-age women. *Proceedings of the National Academy of Sciences*, 108(Supplement 1):4680–4687, 2011.

Cong-mian Ren, Yan Liang, Fengxiang Wei, Ya-nan Zhang, Shou-qiang Zhong, Heng Gu, Xing-Sheng Dong, Yang-Yu Huang, Hua Ke, Xin-ming Son, et al. Balanced translocation t (3; 18)(p13; q22. 3) and points mutation in the znf407 gene detected in patients with both moderate non-syndromic intellectual disability and autism. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, 1832(3):431–438, 2013.

Arne Rietsch, Dominique Belin, Nancy Martin, and Jonathan Beckwith. An in vivo pathway for disulfide bond isomerization in escherichia coli. *Proceedings of the National Academy of Sciences*, 93(23):13048–13053, 1996.

Mario Rivera. Bacterioferritin: structure, dynamics, and protein–protein interactions at play in iron storage and mobilization. *Accounts of chemical research*, 50(2):331–340, 2017.

Mark D Robinson and Alicia Oshlack. A scaling normalization method for differential expression analysis of rna-seq data. *Genome biology*, 11(3):R25, 2010.

Mark D Robinson, Davis J McCarthy, and Gordon K Smyth. edger: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1): 139–140, 2010.

Thomas Rudel, Oliver Kepp, and Vera Kozjak-Pavlovic. Interactions between bacterial pathogens and mitochondrial cell death pathways. *Nature Reviews Microbiology*, 8(10): 693, 2010.

Laura Jane Runyen-Janecky. Role and regulation of heme iron acquisition in gram-negative pathogens. *Frontiers in cellular and infection microbiology*, 3:55, 2013.

Sayed Mohammad Ebrahim Sahraeian, Marghoob Mohiyuddin, Robert Sebra, Hagen Tilgner, Pegah T Afshar, Kin Fai Au, Narges Bani Asadi, Mark B Gerstein, Wing Hung Wong, Michael P Snyder, et al. Gaining comprehensive biological insight into the transcriptome by performing a broad-spectrum rna-seq analysis. *Nature communications*, 8 (1):59, 2017.

Hector A Saka, J Will Thompson, Yi-Shan Chen, Yadunanda Kumar, Laura G Dubois, M Arthur Moseley, and Raphael H Valdivia. Quantitative proteomics reveals metabolic and pathogenic properties of chlamydia trachomatis developmental forms. *Molecular microbiology*, 82(5):1185–1203, 2011.

Fred Sanger and Alan R Coulson. A rapid method for determining sequences in dna by primed synthesis with dna polymerase. *Journal of molecular biology*, 94(3):441–448, 1975.

Frederick Sanger, Steven Nicklen, and Alan R Coulson. Dna sequencing with chain-terminating inhibitors. *Proceedings of the national academy of sciences*, 74(12):5463–5467, 1977.

Katherine Schlosser. Trachoma through history. *Trachoma Matters, International Trachoma Initiative*, 2004.

Johan Schöier, Karin Öllinger, Maria Kvarnström, Gustaf Söderlund, and Erik Kihlström. Chlamydia trachomatis-induced apoptosis occurs in uninfected mccoy cells late in the developmental cycle and is regulated by the intracellular redox state. *Microbial pathogenesis*, 31(4):173–184, 2001.

Ida Schomburg, Antje Chang, and Dietmar Schomburg. Brenda, enzyme data and metabolic information. *Nucleic acids research*, 30(1):47–49, 2002.

Anthony B Schryvers and Igor Stojiljkovic. Iron acquisition systems in the pathogenic neisseria. *Molecular microbiology*, 32(6):1117–1123, 1999.

Stephan C Schuster. Next-generation sequencing transforms today's biology. *Nature methods*, 5(1):16, 2007.

William R Schwan, Silvija N Coulter, Eva YW Ng, Michael H Langhorne, Heather D Ritchie, Linnea L Brody, Shannon Westbrock-Wadman, Arnold S Bayer, Kim R Folger, and C Kendall Stover. Identification and characterization of the putp proline permease that contributes to in vivo survival ofstaphylococcus aureus in animal models. *Infection and immunity*, 66(2):567–572, 1998.

Roland Schwarz, Chunguang Liang, Christoph Kaleta, Mark Kühnel, Eik Hoffmann, Sergei Kuznetsov, Michael Hecker, Gareth Griffiths, Stefan Schuster, and Thomas Dandekar. Integrated network reconstruction, visualization and analysis using yanasquare. *BMC bioinformatics*, 8(1):313, 2007.

Helena MB Seth-Smith and Nicholas R Thomson. Whole-genome sequencing of bacterial sexually transmitted infections: implications for clinicians. *Current opinion in infectious diseases*, 26(1):90–98, 2013.

Paul Shannon, Andrew Markiel, Owen Ozier, Nitin S Baliga, Jonathan T Wang, Daniel Ramage, Nada Amin, Benno Schwikowski, and Trey Ideker. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research*, 13(11):2498–2504, 2003.

Ehud Shapiro, Tamir Biezuner, and Sten Linnarsson. Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nature Reviews Genetics*, 14(9):618, 2013.

Manu Sharma and Thomas Rudel. Apoptosis resistance in chlamydia-infected cells: a fate worse than death? *FEMS Immunology & Medical Microbiology*, 55(2):154–161, 2009.

EI Shaw, CA Dooley, ER Fischer, MA Scidmore, KA Fields, and T Hackstadt. Three temporal classes of gene expression during the chlamydia trachomatis developmental cycle. *Molecular microbiology*, 37(4):913–925, 2000.

Jay Shendure and Hanlee Ji. Next-generation dna sequencing. *Nature biotechnology*, 26 (10):1135, 2008.

JW Shih, TY Tsai, CH Chao, and YH Wu Lee. Candidate tumor suppressor ddx3 rna helicase specifically represses cap-dependent translation by acting as an eif4e inhibitory protein. *Oncogene*, 27(5):700, 2008.

Christine Siegl, Bhupesh K Prusty, Karthika Karunakaran, Jörg Wischhusen, and Thomas Rudel. Tumor suppressor p53 alters host cell metabolism to limit chlamydia trachomatis infection. *Cell reports*, 9(3):918–929, 2014.

Randhir Singh, Sushma Devi, and Rakesh Gollen. Role of free radical in atherosclerosis, diabetes and dyslipidaemia: larger-than-life. *Diabetes/metabolism research and reviews*, 31(2):113–126, 2015.

Eric P Skaar. The battle for iron between bacterial pathogens and their vertebrate hosts. *PLoS pathogens*, 6(8):e1000949, 2010.

Paul Skipp, Jo Robinson, C David O'Connor, and Ian N Clarke. Shotgun proteomic analysis of chlamydia trachomatis. *Proteomics*, 5(6):1558–1573, 2005.

Paul JS Skipp, Chris Hughes, Thérèse McKenna, Richard Edwards, James Langridge, Nicholas R Thomson, and Ian N Clarke. Quantitative proteomics of the infectious and replicative forms of chlamydia trachomatis. *PLoS One*, 11(2):e0149011, 2016.

Denise N Slenter, Martina Kutmon, Kristina Hanspers, Anders Riutta, Jacob Windsor, Nuno Nunes, Jonathan Mélius, Elisa Cirillo, Susan L Coort, Daniela Digles, et al. Wikipathways: a multifaceted pathway database bridging metabolomics to other omics research. *Nucleic acids research*, 46(D1):D661–D667, 2017.

Stéphanie Solier, Jennifer Barb, Barry R Zeeberg, Sudhir Varma, Michael Ryan, Kurt W Kohn, John N Weinstein, Peter J Munson, and Yves Pommier. Genome-wide analysis of novel splice variants induced by topoisomerase i poisoning shows preferential occurrence in genes encoding splicing factors. *Cancer research*, pages canres–2491, 2010.

Rotem Sorek and Pascale Cossart. Prokaryotic transcriptomics: a new view on regulation, physiology and pathogenicity. *Nature Reviews Genetics*, 11(1):9, 2010.

Frank Spiegelhalter and Erhard Bremer. Osmoregulation of the opue proline transport gene from bacillus subtilis: contributions of the sigma a-and sigma b-dependent stress-responsive promoters. *Molecular microbiology*, 29(1):285–296, 1998.

Richard S Stephens, Sue Kalman, Claudia Lammel, Jun Fan, Rekha Marathe, L Aravind, Wayne Mitchell, Lynn Olinger, Roman L Tatusov, Qixun Zhao, et al. Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis. *Science*, 282 (5389):754–759, 1998.

Richard Scott Stephens. *Chlamydia: intracellular biology, pathogenesis, and immunity.* Zondervan, 1999.

Anne Stern, Melissa Brown, Peter Nickel, and Thomas F Meyer. Opacity genes in neisseria gonorrhoeae: control of phase and antigenic variation. *Cell*, 47(1):61–71, 1986.

Zhenqiang Su, Paweł P Łabaj, Sheng Li, Jean Thierry-Mieg, Danielle Thierry-Mieg, Wei Shi, Charles Wang, Gary P Schroth, Robert A Setterquist, John F Thompson, et al. A comprehensive assessment of rna-seq accuracy, reproducibility and information content by the sequencing quality control consortium. *Nature biotechnology*, 32(9):903, 2014.

Agathe Subtil and Alice Dautry-Varsat. Chlamydia: five years ag (after genome). *Current opinion in microbiology*, 7(1):85–92, 2004.

Xiao-Xin Sun, Rosalie C Sears, and Mu-Shui Dai. Deubiquitinating c-myc: Usp36 steps up in the nucleolus. *Cell cycle*, 14(24):3786–3793, 2015.

Yanan Sun, Xiaopeng Jia, Qiang Gao, Xing Liu, and Lianguo Hou. The ubiquitin ligase ube4a inhibits prostate cancer progression by targeting interleukin-like emt inducer (ilei). *IUBMB life*, 69(1):16–21, 2017.

Patrick F Suthers, Madhukar S Dasika, Vinay Satish Kumar, Gennady Denisov, John I Glass, and Costas D Maranas. A genome-scale metabolic reconstruction of mycoplasma genitalium, ips189. *PLoS computational biology*, 5(2):e1000285, 2009.

Damian Szklarczyk, John H Morris, Helen Cook, Michael Kuhn, Stefan Wyder, Milan Simonovic, Alberto Santos, Nadezhda T Doncheva, Alexander Roth, Peer Bork, et al. The string database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic acids research*, page gkw937, 2016.

Hala Tamim, Ramzi R Finan, Huda E Sharida, Mooza Rashid, and Wassim Y Almawi. Cervicovaginal coinfections with human papillomavirus and chlamydia trachomatis. *Diagnostic microbiology and infectious disease*, 43(4):277–281, 2002.

Ming Tan and Patrik Bavoil. *Intracellular Pathogens I: Chlamydiales*, volume 1. American Society for Microbiology Press, 2012.

Xin Tan, Shuai Chen, Jiangxue Wu, Jiaxin Lin, Changchuan Pan, Xiaofang Ying, Zhizhong Pan, Lin Qiu, Ranyi Liu, Rong Geng, et al. Pi3k/akt-mediated upregulation of wdr5 promotes colorectal cancer metastasis by directly targeting znf407. *Cell death & disease*, 8(3):e2686, 2017.

Iman Tavassoly, Joseph Goldfarb, and Ravi Iyengar. Systems biology primer: the basic methods and approaches. *Essays in biochemistry*, 62(4):487–500, 2018.

Ines Thiele, Neil Swainston, Ronan MT Fleming, Andreas Hoppe, Swagatika Sahoo, Maike K Aurich, Hulda Haraldsdottir, Monica L Mo, Ottar Rolfsson, Miranda D Stobbe, et al. A community-driven global reconstruction of human metabolism. *Nature biotechnology*, 31(5):419, 2013.

Keith Tipton. Translocases (ec 7): A new ec class, 2018.

Hiroyuki Tomita, Tomohiro Kanayama, Ayumi Niwa, Kei Noguchi, Kazuhisa Ishida, Masayuki Niwa, and Akira Hara. Cancer stem cells and aldehyde dehydrogenase 1 in liver cancers. In *Updates in Liver Cancer*. InTech, 2017.

Michele Trabucchi. Subcellular heterogeneity of the microrna machinery. *Trends in Genetics*, 2018.

Cole Trapnell, Adam Roberts, Loyal Goff, Geo Pertea, Daehwan Kim, David R Kelley, Harold Pimentel, Steven L Salzberg, John L Rinn, and Lior Pachter. Differential gene and transcript expression analysis of rna-seq experiments with tophat and cufflinks. *Nature protocols*, 7(3):562, 2012.

Iris Uribesalgo, Salvador Aznar Benitah, and Luciano Di Croce. From oncogene to tumor suppressor: the dual role of myc in leukemia. *Cell cycle*, 11(9):1757–1764, 2012.

Bob van de Water, Erik Danen, Esmee Koedoot, Maria Rogkoti, Michiel Fokkelman, Sylvia Le Devedec, John Martens, Peter Stoilov, and John Foekens. Uncovering the signaling landscape controlling breast cancer cell migration identifies splicing prpf4b as a metastasis driver, 2018.

Janneke HHM van de Wijgert, Hanneke Borgdorff, Rita Verhelst, Tania Crucitti, Suzanna Francis, Hans Verstraelen, and Vicky Jespers. The vaginal microbiota: what have we learned after a decade of molecular characterization? *PloS one*, 9(8):e105998, 2014.

Wendy L Veal and William M Shafer. Identification of a cell envelope protein (mtrf) involved in hydrophobic antimicrobial resistance in neisseria gonorrhoeae. *Journal of Antimicrobial Chemotherapy*, 51(1):27–37, 2003.

William Wade. Unculturable bacteria—the uncharacterized organisms that cause oral infections. *Journal of the Royal Society of Medicine*, 95(2):81–83, 2002.

Günter P Wagner, Koryu Kin, and Vincent J Lynch. Measurement of mrna abundance using rna-seq data: Rpkm measure is inconsistent among samples. *Theory in biosciences*, 131(4):281–285, 2012.

Zhong Wang, Mark Gerstein, and Michael Snyder. Rna-seq: a revolutionary tool for transcriptomics. *Nature reviews genetics*, 10(1):57, 2009.

Ronald L Wasserstein. Asa statement on statistical significance and p-values, 2016.

James D Watson and Francis HC Crick. The structure of dna. In *Cold Spring Harbor symposia on quantitative biology*, volume 18, pages 123–131. Cold Spring Harbor Laboratory Press, 1953.

January Weiner. pca3d: Three dimensional pca plots. *R. Available online: https://CRAN. R-project. org/package= pca3d (accessed on 25 March 2018)*, 2013.

Emilio Weiss. Transaminase activity and other enzymatic reactions involving pyruvate and glutamate in chlamydia (psittacosis-trachoma group). *Journal of bacteriology*, 93 (1):177–184, 1967.

SE West and P FREDERICK Sparling. Response of neisseria gonorrhoeae to iron limitation: alterations in expression of membrane proteins without apparent siderophore production. *Infection and Immunity*, 47(2):388–394, 1985.

Alexander J Westermann, Stanislaw A Gorski, and Jörg Vogel. Dual rna-seq of pathogen and host. *Nature Reviews Microbiology*, 10(9):618, 2012.

Rebekah N Whitehead, Tim W Overton, Lori AS Snyder, Simon J McGowan, Harry Smith, Jeff A Cole, and Nigel J Saunders. The small fnr regulon of neisseria gonorrhoeae: comparison with the larger escherichia coli fnr regulon and interaction with the narqnarp regulon. *BMC genomics*, 8(1):35, 2007.

WHO. Sexually transmitted infections (stis), 2016.

Hadley Wickham. ggplot2. *Wiley Interdisciplinary Reviews: Computational Statistics*, 3 (2):180–185, 2011.

Hadley Wickham and Winston Chang. Devtools: Tools to make developing r packages easier. *R package version*, 1(0), 2016.

Hadley Wickham, Romain Francois, Lionel Henry, and K Müller. dplyr: A grammar of data manipulation. *R package version 0.4*, 3, 2015.

Melanie J Wilson, AJ Weightman, and WG Wade. Applications of molecular ecology in the characterization of uncultured microorganisms associated with human disease. *Reviews in Medical Microbiology*, 8(2):91–102, 1997.

Karl G Wooldridge and Peter H Williams. Iron uptake mechanisms of pathogenic bacteria. *FEMS microbiology reviews*, 12(4):325–348, 1993.

John L Wylie, Jody D Berry, and Grant McClarty. Chlamydia trachomatis ctp synthetase: molecular characterization and developmental regulation of expression. *Molecular microbiology*, 22(4):631–642, 1996.

Dong Yang, Bin Sun, Xiaohong Zhang, Daomei Cheng, Xiaoping Yu, Lanzhen Yan, Lei Li, Sanqi An, Hua Jiang, Anna Lasorella, et al. Huwe1 sustains normal ovarian epithelial cell transformation and tumor growth through the histone h1. 3-h19 cascade. *Cancer research*, 2017.

Yee Hwa Yang and Natalie P Thorne. Normalization for two-color cdna microarray data. *Lecture Notes-Monograph Series*, pages 403–418, 2003.

Jiangwei Yao, V Joshua Dodson, Matthew W Frank, and Charles O Rock. Chlamydia trachomatis scavenges host fatty acids for phospholipid synthesis via an acyl-acyl carrier protein synthetase. *Journal of Biological Chemistry*, pages jbc–M115, 2015.

Chunxiao Yu, Ryan McClure, Kathleen Nudel, Nadine Daou, and Caroline Attardo Genco. Characterization of the neisseria gonorrhoeae iron and fur regulatory network. *Journal of bacteriology*, pages JB–00166, 2016.

Mindaugas Zaremba, Paulius Toliusis, Rokas Grigaitis, Elena Manakova, Arunas Silanskas, Giedre Tamulaitiene, Mark D Szczelkun, and Virginijus Siksnys. Dna cleavage by cgii and ngoavii requires interaction between n-and r-proteins and extensive nucleotide hydrolysis. *Nucleic acids research*, 42(22):13887–13896, 2014.

Daniel R Zerbino, Premanand Achuthan, Wasiu Akanni, M Ridwan Amode, Daniel Barrell, Jyothish Bhai, Konstantinos Billis, Carla Cummins, Astrid Gall, Carlos García Girón, et al. Ensembl 2018. *Nucleic acids research*, 46(D1):D754–D761, 2017.

Darja Žgur-Bertok. Dna damage repair and bacterial pathogens. *PLoS pathogens*, 9(11): e1003711, 2013.

Hong Zhang and Eric H Baehrecke. Eaten alive: novel insights into autophagy from multicellular model systems. *Trends in cell biology*, 25(7):376–387, 2015.

Appendix

Table A: Extreme pathways calculated from *Ct* metabolic model

No.	Flux	Length	Reaction No.	Net reaction
P1	1	1	R00722	ATP + IDP = ADP + ITP
P2	2	2	R02340;R02722	Indole + Ser = H2O + Trp
P3	7	4	R00694;R00734; R01373; add5-1-3	2 H+ + NADH + NADPH + 2 Tyr = 2 H2O + NAD+ + NADP+ + 2 Phe
P4	2	2	R00330;R00200	2 ADP + 2 GTP = 2 ATP + 2 GDP
P5	14	9	R02740;R01070;R01015; R01061;R01512;R01518; R00658;R02073;R00330	2P + G6P + 4 GDP + 2 NAD+ + Orthophosphate = 4 GTP + 2 H+ + 2 H2O + 2 NADH + 2 Pyr
P6	1	1	R_GDC	Gly = NH3
P7	1	1	R00086	ATP + H2O = ADP + Orthophosphate
P8	1	1	R00439	Orthophosphate + RNA_ex = GDP + RNA
P9	1	1	R00441	GTP + RNA = 2P + RNA_ex
P10	1	1	R01126	H2O + IMP = Inosine + Orthophosphate
P11	1	1	R00437	Orthophosphate + RNA_ex = ADP + RNA
P12	1	1	R00435	ATP + RNA = 2P + RNA_ex
P13	1	1	R02016	H+ + NADPH + Thioredoxin_disulfide = NADP+ + Thioredoxin
P14	1	1	R00442	CTP + RNA = 2P + RNA_ex
P15	1	1	R01569	H2O + dTMP = Orthophosphate + Thymidine
P16	1	1	R00899	Cysteinyl-glycine + H2O = Cys + Gly
P17	1	1	R00161	ATP + FMN = 2P + FAD
P18	1	1	R10124	Ala + Pimeloyl-ACP = 8-Amino-7-oxononanoate + ACP + CO2
P19	2	2	R02720;R02719	2 H2O + XTP = 2P + Orthophosphate + Xanthosine
P20	3	3	R00425;R03459;R03458	GTP + 4 H2O + NADP+ = 2P + e5-Amino-6-5-P-D-ribitylamino-uracil + Formate + H+ + NADPH + NH3
P21	2	2	R03530;R03531	ADP + H2O + dITP = 2P + ATP + dIDP
P22	2	2	R00332;R01227	ADP + GDP + H2O = ATP + Guanosine + Orthophosphate
P23	5	5	R01088;R07601;R07602; R01698;R04097	CoA + H2O + Leu + 2 NAD+ = CO2 + 2 H+ + Methylbutanoyl-CoA + 2 NADH + NH3
P24	3	3	R01818;R00883;add1-4-2	GDP + M6P = N-glycan + Orthophosphate
P25	2	2	R05196;R02112	Glycogen = Dextrin + Glc

No.	Flux	Length	Reaction No.	Net reaction
P26	3	3	R11261;R05196;R11262	Glycogen + 2 H2O = Glc
P27	23	6	R04109;R02272;R00036; R00084;R03166;R04971	8 Glutamyl-tRNA + 2 H+ + 8 NADPH = 8 H2O + 8 NADP+ + 4 NH3 + Uroporphyrin_I + 8 tRNA(Glu)
P28	23	6	R04109;R02272;R00036; R00084;R03165;R03195	8 Glutamyl-tRNA + 2 H+ + 8 NADPH = 8 H2O + 8 NADP+ + 4 NH3 + Uroporphyrin_III + 8 tRNA(Glu)
P29	24	7	R04109;R02272;R00036; R00084;R03165;R03197; R04178	8 Glutamyl-tRNA + 8 H+ + 8 NADPH + 3 O2 = 4 CO2 + Coproporphyrin_III + 8 H2O + 3 H2O2 + 8 NADP+ + 4 NH3 + 8 tRNA(Glu)
P30	26	9	R04109;R02272;R00036; R00084;R03165;R03197; R03220;R03222;R00310	Fe2+ + 8 Glutamyl-tRNA + 6 H+ + 8 NADPH + 4 O2 = 6 CO2 + 10 H2O + 3 H2O2 + Heme + 8 NADP+ + 4 NH3 + 8 tRNA(Glu)
P31	3	3	R00378;R02093;R02094	2 ATP + DNA + dTMP = 2P + 2 ADP + DNA_ex
P32	3	3	R02331;R02100;R02098	2 ATP + H2O = 2P + 2 ADP
P33	13	7	R03504;R03066;R02235; R00425;R02237;R04621; R11072	2 ATP + 2 GTP + 2 Glu + NAD+ + NADP+ + 2 Orthophosphate + 2 PABA = 4 2P + 2 ADP + 2 Folate + 2 Formate + 2 Glycolaldehyde + 2 H+ + NADH + NADPH
P34	2	2	R04779;R02073	ATP + Orthophosphate = 2P + ADP
P35	2	2	R01821;R00959	Glycogen + Orthophosphate = G6P
P36	4	4	R00948;R02421;R02110; R00959	ATP + G6P = 2P + 2 ADP + Glycogen
P37	3	3	R02110;R02111;R00959	Orthophosphate = G6P
P38	7	4	R03940;R00945;R01200; R01655	2 Met-tRNA + NAD+ + NADP+ + 2 Ser = 2 Gly + 2 H+ + 2 N-Formylmethionyl-tRNA + NADH + NADPH
P39	10	10	combi1-2-8__11;R00416; R05332;R02060;R01150; R04573;R05629;R05662; add2-2-1;R05626	3 ATP + Ala + D-GlcN6P + NAD+ + Pyr + UDP-MurNAc-L-Ala-D-Glu-L-Lys + Undecaprenol = 2P + 3 ADP + CO2 + H+ + NADH + 2 Orthophosphate+ Peptidoglycan
P40	4	4	R00570;R02024;R02326; R00377	CTP + DNA + Thioredoxin = 2P + DNA_ex + Thioredoxin_disulfide
P41	7	7	R00570;R02024;R02326; R02325;R02331;R02098; R02102	2 ADP + CTP + 2 H2O + Thioredoxin = 2 ATP + Deoxyuridine + NH3 + Orthophosphate + Thioredoxin_disulfide
P42	2	2	R00570;R00440	ATP + Orthophosphate + RNA_ex = ADP + CTP + RNA
P43	3	3	R02017;R01547;R02088	2 ADP + Thioredoxin = ATP + Deoxyadenosine + Orthophosphate + Thioredoxin_disulfide
P44	7	4	R01137;R00375;R01547; R02088	3 ATP + 2 DNA + 2 Deoxyadenosine + 2 Orthophosphate = 2 2P + 3 ADP + 2 DNA_ex + 2 H2O

No.	Flux	Length	Reaction No.	Net reaction
P45	3	3	R02090;R01968;R02019	ADP + GDP + Thioredoxin = ATP + Deoxyguanosine + Orthophosphate + Thioredoxin_disulfide
P46	11	6	R00342;R01082;R00408; R00405;combi1-3-5_8; R00355	2 ADP + FAD + 2 Glu + 2 H2O + 4 NAD+ + 2 Orthophosphate + Quinone = 2 ATP + 2 Asp + FADH2 + 4 H+ + Hydroquinone + 4 NADH
P47	2	2	R00443;R00568	CTP + H2O + RNA = 2P + NH3 + RNA_ex
P48	3	2	R00571;R00568	2 ATP + Gln + 3 H2O = 2 ADP + Glu + NH3 + 2 Orthophosphate
P49	2	2	R00127;R05578	2 ATP + Glu + tRNA(Glu) = 2P + 2 ADP + Glutamyl-tRNA
P50	3	3	R00127;R01047;R05145	2 ATP + Apo-carboxylase + Biotin = 2P + 2 ADP + Holo-carboxylase
P51	17	9	R00127;R03504;R03503; R03067;R02235;R00425; R02237;R04621;R11072	6 ATP + 2 GTP + 2 Glu + 2 H2O + NAD+ + NADP+ + 2 Orthophosphate + 2 PABA = 6 2P + 6 ADP + 2 Folate + 2 Formate + 2 Glycolaldehyde + 2 H+ + NADH + NADPH
P52	3	3	R00570;R00511	2 ADP + CTP + H2O = 2 ATP + Cytidine + Orthophosphate
P53	9	5	R01731;R00734;R00355; add5-1-3;R00691	2 H+ + NADH + NADPH + 2 Tyr = 2 H2O + NAD+ + NADP+ + 2 Phe
P54	9	5	R01731;R00694;R00355; R01373;add5-1-5	2 H2O + NAD+ + NADP+ + 2 Phe = 2 H+ + NADH + NADPH + 2 Tyr
P55	3	3	R02235;R00937;R00936	no net reaction
P56	7	5	R02235;R00936;R00945; R01200;R02301	2 5-Formyl-THF + 2 ATP + 2 Gly + 2 H2O + NAD+ + NADP+ = 2 ADP + 2 Folate + 2 H+ + NADH + NADPH + 2 Orthophosphate + 2 Ser
P57	9	6	R02235;R00936;R00945; R01200;R01655;add9-3-6	2 5-Formyl-THF + 2 Gly + NAD+ + NADP+ = 2 Folate + 2 H+ + NADH + NADPH + 2 Ser
P58	2	2	R02740;R00768	G6P + Gln = D-GlcN6P + Glu
P59	3	3	R00438;R00158;R00963	ADP + H2O + RNA_ex = ATP + RNA + Uridine
P60	4	4	R00438;R02098;R02102; R02018	ADP + RNA_ex + Thioredoxin = ATP + Deoxyuridine + RNA + Thioredoxin_disulfide
P61	3	3	R00438;R00443;R00156	ATP + Orthophosphate = 2P + ADP
P62	5	3	R00438;R00156;R00571	4 ATP + Gln + H2O + NH3 + 2 RNA_ex = 4 ADP + 2 CTP + Glu + 2 RNA
P63	4	4	R00156;R00158;R00963; R00568	2 ADP + CTP + 2 H2O = 2 ATP + NH3 + Orthophosphate + Uridine
P64	5	5	R00156;R02098;R02102; R02018;R00568	2 ADP + CTP + H2O + Thioredoxin = 2 ATP + Deoxyuridine + NH3 + Orthophosphate + Thioredoxin_disulfide

No.	Flux	Length	Reaction No.	Net reaction
P65	14	11	combi1-2-8__11;R00742; R01626;R10115;R10116; R10117;R10118;R10119; R10120;R10121;R10122	2 ATP + 2 H+ + 2 HCO3- + Malonyl-ACP-methyl-ester + 2 NAD+ + 4 NADPH+ 2 Pyr = 2 ADP + 4 CO2 + 2 H2O + 2 NADH + 4 NADP+ + 2 Orthophosphate + Pimeloyl-ACP-methyl-ester
P66	124	36	combi1-2-8__11;R00742; R01626;R04355;R04533; R04428;R04429;R04952; R04953;R04954;R04955; R04957;R04536;R04537; R04958;R04960;R04534; R04535;R04961;R04963; R04964;R04965;R04724; R04726;R04566;R04568; R04966;R04968;R04543; R04544;R04969;R07762; R07763;R07764;R07765; add7-1-1	18 ATP + 2 Acetyl-ACP + 18 H+ + 18 HCO3- + 6 NAD+ + 26 NADPH + 18 Pyr = 18 ADP + 36 CO2 + 4 FA__ex + 18 H2O + 6 NADH + 26 NADP+ + 18 Orthophosphate
P67	126	36	combi1-2-8__11;R00742; R01626;R01070;R04533; R04428;R04429;R04952; R04953;R04954;R04955; R04957;R04536;R04537; R04958;R04960;R04534; R04535;R04961;R04963; R04964;R04965;R04724; R04726;R04566;R04568; R04966;R04968;R04543; R04544;R04969;R07762; R07763;R07764;R07765; add7-1-1	2 ACP + 18 ATP + 16 H+ + 18 HCO3- + 8 NAD+ + 26 NADPH + 20 Pyr = 18 ADP + 38 CO2 + 4 FA__ex + 18 H2O + 8 NADH + 26 NADP+ + 18 Orthophosphate
P68	2	2	R07396;R00895	4-Methylthio-2-oxobutanoate + Cys = Mercaptopyruvate + Methionine-L
P69	4	4	R00858;R02433;R00895; R04861	Cysteate + 3 H+ + Mercaptopyruvate + 3 NADPH = Cys + 2 H2O + 3 NADP+ + Pyr + Sulfide
P70	20	20	combi1-2-8__11;R00416; R05332;R02060;R03193; R02783;R02788;R01150; R04617;R05630;R05032; R00480;R02291;R10147; R04199;R07613;R02735; R00895;add2-2-2;R05626	7 ATP + 2 Ala + Asp + Cys + D-GlcN6P + Glu + H+ + NAD+ + 2 NADPH + 2 Pyr + UDP-MurNAc + Undecaprenol = 2P + 7 ADP + CO2 + H2O + Mercaptopyruvate + NADH + 2 NADP+ + 6 Orthophosphate + Peptidoglycan

114

No.	Flux	Length	Reaction No.	Net reaction
P71	8	5	R00376;R01857;R02090; R01968;R00330	2 ATP + 2 DNA + 2 Deoxyguanosine + 2 GTP + 2 Orthophosphate = 2 2P + 2 ADP + 2 DNA_ex + 2 GDP + 2 H2O
P72	21	6	R00200;R01857;R00330; R01137;R00156R00570	21 ADP + 24 GTP + 2 IDP = 21 ATP + 24 GDP + 2 ITP
P73	15	8	combi1-2-8_11;R03191; R00660;R00416;R05332; R02060;R00330;R00438	2 ADP + 2 D-GlcN6P + 4 GTP + NAD+ + NADPH + 4 Pyr + 2 RNA_ex = 2 2P + 2 ATP + 2 CO2 + 4 GDP + NADH + NADP+ + 2 RNA + 2 UDP-MurNAc
P74	17	9	combi1-2-8_11;R03191; R00660;R00416;R05332; R02060;R00330;R00156; R00568	4 ADP + 2 CTP + 2 D-GlcN6P + 4 GTP + 2 H2O + NAD+ + NADPH + 4 Pyr = 2 2P + 4 ATP + 2 CO2 + 4 GDP + NADH + NADP+ + 2 NH3 + 2 Orthophosphate + 2 UDP-MurNAc
P75	7	4	R00431;R00200;R00895; R00355	2 ADP + 2 Asp + GTP + ITP + 2 Mercaptopyruvate = 2 ATP + 2 CO2 + 2 Cys + GDP + IDP + 2 Pyr
P76	82	16	R02740;R01070;R01015; R02073;R00511;R05636; R05688;R05633;R05634; R05637;R08689;R05884; R07219;R01658;R02033; R02061	8 ATP + 8 CTP + 16 Ferredoxin-reduced + 4 G6P + 16 H+ + 4 NADH + 12 NADPH + 8 Pyr = 10 2P + 8 ADP + 8 CO2 + 8 Cytidine + 16 Ferredoxin-oxidized + 2 GeranylgeranylPP + 8 H2O + 4 NAD+ + 12 NADP+ + 12 Orthophosphate
P77	63	16	R02740;R01070;R01015; R02073;R00511;R05636; R05688;R05633;R05634; R05637;R08689;R05884; R07219;R01658;R02033; R06447	6 ATP + 6 CTP + 12 Ferredoxin-reduced + 3 G6P + 12 H+ + 16 IDP + 3 NADH + 9 NADPH + 6 Pyr = 23 2P + 6 ADP + 6 CO2 + 6 Cytidine + 2 EEpolyS-UndecaprenylPP + 12 Ferredoxin-oxidized + 6 H2O + 3 NAD+ + 9 NADP+ + 9 Orthophosphate
P78	44	13	R01056;R01529;R02740; R01070;R01830;R01827; R01641;R01015;R02073; R07281;R04457;R00066; R00549	2P + 6 ATP + 5 G6P = 6 ADP + 6 FMN + 6 Formate + 12 H2O + 7 Orthophosphate

No.	Flux	Length	Reaction No.	Net reaction
P79	320	52	R01056;R01529;R02740; R01070;R01830;R01827 ;R01641;R01015; combi1-2-8_11;R02073; R00416;R05332;R02060; R04567;R04587;R04550; R04606;R04657;R04658; R05074;R05146;R01530; R03254;R03350;R03351; R00330;R00511;R00742; R01626;R04355;R04533; R04428;R04429;R04952; R04953;R04954;R04955; R04957;R04536;R04537; R04958;R04960;R04534; R04535;R04961;R04963; R04964;R04965;R04724; R04726;R04566;add2-1-1	33 ATP + 6 Acetyl-ACP + 6 CTP + 3 D-GlcN6P + 5 G6P + 12 GTP + 21 H+ + 36 HCO3- + 3 Lauroyl-ACP + 3 LipidX + 24 NAD+ + 51 NADPH + 45 Pyr = 8 2P + 9 ACP + 33 ADP + 3 Ac + 75 CO2 + 6 Cytidine + 12 GDP + 15 H2O + 3 Lipopolysacchride + 24 NADH + 51 NADP+ + 55 Orthophosphate
P80	326	52	R01056;R01529;R02740; R01070;R01830;R01827; R01641;R01015; combi1-2-8_11;R02073; R00416;R05332;R02060; R04567;R04587;R04550; R04606;R04657;R04658; R05074;R05146;R01530; R03254;R03350;R03351; R00330;R00511;R00742; R01626;R01070;R04533; R04428;R04429;R04952; R04953;R04954;R04955; R04957;R04536;R04537; R04958;R04960;R04534; R04535;R04961;R04963; R04964;R04965;R04724; R04726;R04566;add2-1-1	33 ATP + 6 CTP + 3 D-GlcN6P + 5 G6P + 12 GTP + 15 H+ + 36 HCO3- + 3 Lauroyl-ACP + 3 LipidX + 30 NAD+ + 51 NADPH + 51 Pyr = 8 2P + 3 ACP + 33 ADP + 3 Ac + 81 CO2 + 6 Cytidine + 12 GDP + 15 H2O + 3 Lipopolysacchride + 30 NADH + 51 NADP+ + 55 Orthophosphate
P81	50	21	R01056;R01529;R02740; R01070;R01830;R01827; R01641;R01015;R02073; R00330;R01826;R03083; R03084;R02413;R02412; R03460;R01714;R00694; R00895;R01715;R01373	2P + 3 ADP + 3 Cys + 2 G6P + 12 GTP + 3 H+ + 3 NADPH + 6 Pyr = 3 ATP + 3 CO2 + 12 GDP + 3 H2O + 3 Mercaptopyruvate + 3 NADP+ + 13 Orthophosphate + 3 Phe

116

No.	Flux	Length	Reaction No.	Net reaction
P82	27	14	R02740;R01070;R01061; R01512;R01518;R00658; R02073;R00330;R00511; R01800;R00842;R00851; R00241;R01799	4 Acyl-CoA + 2 CTP + 2 G6P + 4 GDP + NAD+ + NADPH + 2 Ser = 4 CoA + 2 Cytidine + 4 GTP + NADH + NADP+ + 2 Orthophosphate + 2 Phosphatidylserine + 2 Pyr
P83	22	14	R02740;R01070;R01061; R01512;R01518;R00658; R02073;R00330;R00511; R00842;R00851;R00241; R01799;R01801	2P + 2 Acyl-CoA + CTP + 2 G6P + 4 GDP + NAD+ + NADPH = 2 CoA + Cytidine + 4 GTP + H2O + NADH + NADP+ + Orthophosphate + PGP + 2 Pyr
P84	20	14	R02736;R02035;R01528; R01056;R01529;R02740; R01830;R01827;R01641; R01061;R01512;R01518; R00658;R00330	6PGn + 2 G6P + 2 GDP + 2 H2O + 5 NAD+ + 4 NADP+ + Orthophosphate = 6PGL + 4 CO2 + 2 GTP + 9 H+ + 5 NADH + 4 NADPH + Pyr + Ru5P

List of Abbreviations

1,2-Diacyl-GL3P	1,2-diacyl-sn-glycerol 3-phosphate
1-Acyl-GL3P	Acyl-sn-glycerol 3-phosphate
1,3BPG	1,3-bisphosphoglycerate
2PG	2-phosphoglycerate
3PG	3-phosphoglycerate
6-PGC	Phospho-D-gluconate
6-PGL	Glucono-1,5-lactone 6-phosphate
5-CHO-THF	5-Formyltetrahydrofolate
10-CHO-THF	10-Formyltetrahydrofolate
ACADVL	Very long chain acyl-CoA dehydrogenase
ACCOA	Acetyl-CoA
ACSL5	Acyl-CoA synthetase long chain family member 5
ACP	Acyl carrier protein
Ago	Argonaute
AHHMD	2-Amino-4-hydroxy-6-hydroxymethyl-7,8-dihydropteridine diphosphate
aKG	2-oxoglutarate
ALDH1A3	Aldehyde dehydrogenase 1 family member A3
aniA	Nitrate reductase
araD	L-ribulose-5-phosphate 4-epimerase
ARSA	Arylsulfatase A
BAM	Sequence alignment map in binary format
BER	DNA base excision repair
bfrB	Bacterioferritin B
CBT1B	Carnitine palmitoyltransferase 1B
ccp	Cytochrome-c peroxidase
CDP-diacyl-GL	CDP-diacylglycerol
CDC	Disease Control and Prevention
CERS2	Ceramide synthase 2
CNF	Cornifelin

CREB	cAMP-response element-binding protein) binding protein	
CREBBP	CREB binding protein	
Ct	*Chlamydia trachomatis*	
dNTP	Deoxynucleotide triphosphates	
ddNTP	Dideoxynucleotide triphosphates	
DDX3X	X-linked DEAD (Asp-Glu-Ala-Asp)-box RNA helicase 3	
DEG	Differentially expressed gene	
DET	Differentially expressed transcript	
DHAP	Dihydroxyacetone phosphate	
DHF	Dihydrofolate	
DHP	Dehydropantoate	
DHPT	Dihydropteroate	
dipZ	Thiol:disulfide interchange protein	
dnaA	Chromosomal replication initiation protein	
dnaG	DNA primase	
E4P	Erythrose 4-phosphate	
EB	Elementary body	
EC	Enzyme Commission	
EMP	Glycolysis	
EP	Extreme pathway	
EmPCR	Emulsion PCR	
ETF	Electron transfer flavoprotein	
F1P	Fructose 1-Phosphate	
F1,6BP	Fructose 1,6-bisphosphate	
F6P	Fructose 6-phosphate	
FA	Fatty acid	
fabD	acyl-carrier-protein S-malonyltransferase	
fabF	3-oxoacyl-ACP synthase II	
fabG	3-oxoacyl-ACP synthase III	
FACS	Fluorescence-activated cell sorter	
FDR	False discovery rate	
FNR	Fumarate and nitrate regulator	
folA	Dihydrofolate reductase	
folP	Dihydropteroate synthase	
folX	Dihydroneopterin aldolase	
FPKM	Fragments per kilobase per million mapped reads	
FUM	Fumarate	

fumC	Fumarate hydratase
Fur	Ferric uptake regulator
GADP	Glyceraldehyde 3-phosphate
gapA/**GAPDH**	Glyceraldehyde 3-phosphate dehydrogenase
GBP3	Guanylate binding protein 3
GC	*Neisseria gonorrhoeae*
GCALD	Glycolaldehyde
GEMM	Genome-scale metabolic model
GFF	General feature format
GFP	Green fluorescent protein
GIN	Genetic interaction network
GL3P	sn-Glycerol 3-phosphate
Glc-6P	Glucose 6-phosphate
glmS	Glutamine-fructose-6-phosphate transaminase
gltT	Glutamate trasporter
GNG	Gluconeogenesis
GO	Gene ontology
gpdA	Glycerol-3-phosphate dehydrogenase (NAD(P)+)
GPL	Glycerophospholipids synthesis
GRN	Gene regulatory network
GSH	Glutathione
GTF	Gene transfer format
HIV	*Human immunodeficiency virus*
HK1	Hexokinase 1
HPV	*Human papillomavirus*
hpuB	Hemoglobin-haptoglobin utilization protein B
HS	*Homo sapiens*
HSV-2	*Herpes simplex virus* type 2
HUVECs	Human umbilical vein endothelial cells
HUWE1	HECT, UBA and WWE domain containing 1
KRT4	Keratin 4
LD	Lipid droplet
LGV	Lymphogranuloma venereum
mal	T cell differentiation protein
MAL	Malate
MALACP	Malonyl-[acyl-carrier protein]
MALCOA	Malonyl-CoA

mdhC	Malate dehydrogenase
METHF	5,10-Methenyltetrahydrofolate
METTHF	5,10-Methylenetetrahydrofolate
MFA	Metabolic flux analysis
MHC	Major histocompatibility complex
miRNA	microRNA
MOMP	Major outer membrane protein
mROS	Mitochondrial reactive oxygen species
MULE	Mcl-1 ubiquitin ligase E3
NBN	Nibrin
OAA	Oxaloacetate
opa	Opacity protein
OXPHOS	Oxidative phosphorylation
PABA	4-aminobenzoic acid
PCA	Principal component analysis
PCR	Polymerase chain reaction
PE	Phosphatidylethanolamine
PEP	2-phosphoenolpyruvate
pfkA_1	6-phosphofructokinase 1
pfkA_2	diphosphate-fructose-6-phosphate 1-phosphotransferase
PG	Peptidoglycan
pgi	Glucose-6-phosphate isomerase
PGP	Phosphatidylglycerophosphate
plsB	Glycerol-3-phosphate O-acyltransferase
plsC	1-acyl-sn-glycerol-3-phosphate acyltransferase
Polβ	DNA polymerase II
PPI	Protein-protein interaction
PPP	Pentose phosphate pathway
PRPP	Phosphoribosyl pyrophosphate
PSer	Phosphatidylserine
***p*-value**	Probability value
pykF	Pyruvate kinase
Pyr	Pyruvate
pyrG	CTP synthase
R1P	Ribose 1-phosphate
R5P	Ribose 5-phosphate
RAR*a*	Retinoic acid receptor *a*

RB	Reticulate body
REST	RE1 silencing transcription factor
RL5P	Ribulose 5-phosphate
RLE	Relative log expression
RM	Restriction-modification
RNS	Reactive nitrogen species
ROS	Reactive oxygen species
RNA-seq	RNA sequencing
RPKM	Reads per kilobase per million mapped reads
RT-qPCR	Quantitative real-time PCR
S7P	Sedoheptulose 7-phosphate
SAM	Sequence alignment map
SCO2	Cytochrome c oxidase assembly protein
SEM	Standard error of the mean
SFPQ	Splicing factor proline and glutamine-rich
SMRT	Single molecule real-time sequencing
SPRR	Small proline-rich protein
STD	Sexually transmitted disease
STI	Sexually transmitted infection
sucA/B	2-oxoglutarate dehydrogenase
sucC	succinyl-CoA synthetase beta subunit
SUCC	Succinate
SUCCOA	Succinyl-CoA
sucD	succinyl-CoA synthetase alpha subunit
T3SS	type three secretion system
tal	Transaldolase
TCA	Tricarboxylic acid cycle
TfR	Transferrin receptor
THF	Tetrahydrofolate
tktB	Transketolase
TMM	Trimmed mean of m-values
TNRC6	Trinucleotide repeat containing 6
TP53	Tumour repressor protein 53
TPM	Transcripts per kilobase million
TRN	transcriptional regulatory network
WHO	World Health Organization
X5P	Xylose 5-phosphate

Curriculum Vitae | Manli Yang

Education

University of Würzburg, Graduate School of Life Science   *Würzburg, Germany*

PH.D. IN DEPARTMENT OF BIOINFORMATICS, SECTION: INFECTION & IMMUNITY   *May. 2016 - PRESENT*

- GRK 2157 3D tissue models for studying microbial infections by human pathogens
  Chlamydia trachomatis metabolism during infection & metatranscriptome analysis in *Neisseria gonorrhoeae* coinfected STD patients

University of Würzburg, FOKUS Life Science   *Würzburg, Germany*

M.SC. IN DEPARTMENT OF BIOINFORMATICS   *Oct. 2013 - Jun. 2015*

- Genomic identification of essential genes involved in the thermal-tolerant LAB colony isolated from Algerian dromedary's raw milk

Shandong University, School of Life Science   *Jinan, China*

B.SC. IN STRUCTURAL BIOINFORMATICS AND BIOCHEMISTRY   *Sept. 2009 - Jun. 2013*

- Active site architecture and evolution of catalytic mechanism in glycoside hydrolase family 5

Research Experience

Department of Bioinformatics, University of Würzburg   *Würzburg, Germany*

3D TISSUE MODELS FOR STUDYING MICROBIAL INFECTIONS BY HUMAN PATHOGENS   *Oct. 2015 - PRESENT*

- *Chlamydia trachomatis* genome-scale metabolic modelling and flux analysis of biphasix differentiation during infection
- Metatranscriptome analysis of host and *Neisseria gonorrhoeae* responses by *Chlamydia trachomatis* in STD patients
- Functional assay and structural analysis of 6-phosphofructokinase in *Chlamydia trachomatis*

GENOMIC IDENTIFICATION OF THERMAL-TOLERANT LACTIC ACID BACTERIA STRAINS   *Oct. 2014 - Jun. 2015*

- Metagenomic assebmly and scaffolding of low-coverage mixed reads among multiple bacterial strains
- Functional analysis on heat shock gene and extended clusters
- Comparatie analysis on polysaccharide, glycosyltransferase and ABC transporters clusters
- SEM investigation

PROTEIN STRUCTURE HOMOLOGY MODELLING   *May. 2014 - Aug. 2014*

- Observing special folds: Src Kinase, TIM Barrel, Klenow fragment, Hexapeptide repeat
- Collecting candidates with low sequence identity but high structure identity
- Comparing entropy and sequence variation among candidates in same fold

Department of Integrative Physiology and Pathobiology, Tufts University   *Boston, MA, USA*

RNA SEQUENCING ANALYSIS OF THE HOST AND STI PATHOGENS FROM PATIENT TRANSCRIPTOME   *Oct. 2017 - Dec. 2017*

- Microbial census detection
- Quality control, reads mapping and assembly
- Differential expression analysis

State Key laboratory of microbial technology, Shandong University   *Jinan, China*

FUNCTIONAL EVOLUTION ANALYSIS OF CELLULASE IN GLYCOSIDE HYDROLASE FAMILIES   *Sept. 2012 - Jun. 2013*

- Homology modelling based on sequence and structural alignment
- Phylogenetic analysis
- Analysis of semi-conserved residues in active site
- Analysis of CH-Pi interaction between Trp and saccharides with QM/MM

ACTIVE SITE ARCHITECHTURE ANALYSIS OF CELLULASE IN GLYCOSIDE HYDROLASE FAMILIES   *Jun. 2012 - Aug. 2012*

- Classification of conserved, semi-conserved and unconserved residues in active site structure
- Structural profile in sequence logo based on structure alignment

NATIONAL UNDERGRADUATE SCIENTIFIC AND TECHNOLOGICAL INNOVATION PROJECT   *Sept. 2011 - Jun. 2012*

- Evolution of disulfide bond correspond with thermal stability of cellulase

Affidavit

I hereby confirm that my thesis entitled "*Chlamydia trachomatis* metabolism during infection and metatranscriptome analysis in *Neisseria gonorrhoeae* coinfected STD patients" 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 "*Chlamydia trachomatis* Metabolismus während der Infektion sowie die Analyse des Metatranskriptoms bei *Neisseria gonorrhoeae* koinfizierten STD-Patienten" 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