UNIVERSITÀ DEGLI STUDI DI TORINO

PhD School of Health and Life Sciences

PhD Programme in Complex Systems for Life Sciences

XXXIV Cycle

Mining the complexity of host-pathogen interactions from dual RNA-Seq short time series on *Neisseria gonorrhoeae* infection models

PhD Candidate

Dr. Stefano Carello

Supervisors

Dr. Alessandro Brozzi

Prof. Michele Caselle

Coordinator

Prof. Michele De Bortoli

A.Y. 2021/2022

Index

ABSTRACT

With a worldwide incidence of around 100 million cases each year, gonorrhea still represents one of the major sexually transmitted infections. In the absence of an effective vaccine, an alarm is rising due to the insurgence of strains resistant to most available antibiotics, common asymptomatic infections and uncontrolled transmission, especially in low-income countries. Soon after its appearance, dual RNA-Seq technology began to be used to simultaneously analyze the transcriptome from a large number of pathogens and infected hosts, providing new insights into bacterial pathogenesis. Here, with such an approach, we characterized the global transcriptional changes leading the host-pathogen crosstalk during early stages of *Neisseria gonorrhoeae* WHO M infection. We elucidated the chemotactic and pro-inflammatory responsiveness following bacterial interaction of three described cell models representative of the main human anatomical sites infected by this pathogen, as well as *N. gonorrhoeae* factors regulated during adhesion, invasion and early intracellular persistence. Furthermore, we proposed a new analytical method to study host-pathogen interactions from short time series dual RNA-Seq data. Through a permutation test-based approach, this method aims at detecting non-casual transcriptional influences occurring during the infection process between the two organisms. Such dependencies are described with a weighted directed bipartite graph, whose complexity is mined through centrality measures and node removal tests in order to extract key regulations that support the interaction. The new strategy was first tested on a publicly available gene expression dataset concerning early infection of human lung cells by the *Streptococcus pneumoniae* well-studied strain D39, from which regulations of known factors needed by this pathogen to initiate the interaction and colonization of epithelial cells emerged. The application to the new *N. gonorrhoeae* WHO M infection data set confirmed the induction of cytoskeletal rearrangements in infected cells as a baseline host event during early interactions and identified bacterial genes belonging to known pathogenetic classes, as well as new uncharacterized ones, as putative factors required by gonococcus to successfully establish the infection in the three epithelia. In summary, we reported a first dual RNA-Seq comparative study on early *N. gonorrhoeae* WHO M infection of three described cell models to aid in characterizing gonococcal pathogenesis. This work also represented a first step towards developing new analytical methods that look at the integrated nature of the ever-growing dual RNA-Seq data.

Keywords: *N. gonorrhoeae* | Dual transcriptomics | Host-pathogen interactions.

CHAPTER 1: INTRODUCTION AND AIM OF THE WORK

1.1 SIMULTANEOUSLY PROFILING HOST AND PATHOGEN TRANSCRIPTOMES DURING INFECTIONS

1.1.1 -OMICS SCIENCES: THE NEW INSTRUMENTS OF THE POST-GENOMIC ERA

The completion of the first human genome sequence in 2004 not only posed a significant milestone for our knowledge, but also paved the way to a new perspective in biological studies (International Human Genome Sequencing Consortium, 2004). The amount of time and resources that was needed to reach such a goal pushed the efforts in developing cheaper and faster approaches. Since the appearance of the first next-generation sequencing (NGS) technology one year later by 454 Life Sciences (today owned by Roche) (Margulies *et al.*, 2005), many other companies joined the challenge, starting to release and upgrade their respective technologies. The field in few years reached impressive results, significantly reducing costs and time required by these studies as well as increasing the high-throughput of the sequencing machines. Such speed in the production of vast amounts of data opened the door to the -omics sciences, and to a new, system level way of thinking and asking ourselves biological questions.

With the revolution supported by NGS, gene expression studies started to move from hybridization-based microarray technologies to the new emerging RNA-Seq one, based on the conversion of total or fractionated (often, poly(A)+) RNA into a library of cDNA sequences characterized by adapters attached to the extremes that, upon amplification or not in case of single molecule sequencing, are deeply sequenced to obtain reads of ~30-400 bp (depending on the technology) from one (single-end sequencing) or both ends (paired-end sequencing). Particularly, the Illumina strategy, based on the so-called sequencing-by-synthesis, became popular in the field. Compared to microarrays, RNA-Seq demonstrated to be more accurate, to have higher dynamic range of expression levels detection and no interfering background, and importantly to require no previous knowledge on the existing cell transcripts (Mortazavi *et al.*, 2008; Nagalakshmi *et al.*, 2008). Previously, the introduction of tiling arrays, that contains probes that could in principle represent also the entire genome at high-resolution, tried to overcome this limitation, but with the naturally consequent caveat of excessive costs. The new technique, in addition to be particularly attractive for studies on non-model organisms, made also possible the discovery of new RNAs belonging to that portion of non-coding sequences that the Human Genome Project highlighted as significantly enriched in our genome (Cabili *et al.*, 2011; Djebali *et al.*, 2012). The advent of strand-specific RNA-Seq protocols, that retain the information about the strand of origin and allow to discriminate and quantify the common event concerning non-coding transcripts being antisense to the coding ones, particularly fostered such studies. Furthermore, this was even more relevant for studies on bacterial pathogens, characterized by smaller genomes but typically exploiting more their encoding potential, therefore presenting many overlapping genes on the two strands. However, it is important to notice that many human non-coding RNAs are characterized by both spatiotemporal specificity and low expression levels, so the choice of the appropriate sample preparation and sequencing method is critical in order to efficiently detect them. In particular, since classical RNA-Seq sample preparation didn’t work for small RNAs, dedicated library preparation and enrichment protocols, such as for miRNA-Seq, were introduced (Landgraf *et al.*, 2007).

1.1.2 TRANSCRIPTOMICS STARTS TO MEET INFECTION BIOLOGY

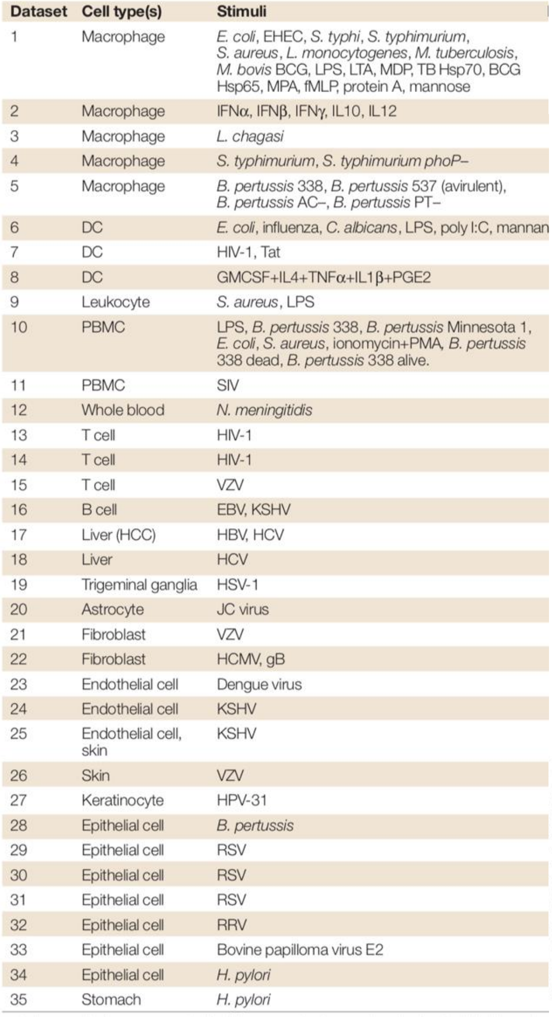
To successfully infect the host, bacterial pathogens typically pass through different steps, such as adhesion, colonization and usually invasion of host cells, experiencing different stresses due to the changing environment. Therefore, they finely tune their transcriptional programs in order to adapt, survive and replicate inside the host cells. At the same time, host cells respond to the exposure of pathogen-associated molecular patterns (PAMPs) activating specific signaling pathways, mainly driven by Toll-like receptors, that ultimately orchestrate gene expression in order to counteract the infection and lead to bacterial clearance by the immune system. The two start to share the same niche and nutrients and therefore begin, primarily from a metabolic point of view (Olive and Sassetti, 2016), a complex crosstalk that a transcriptional profiling along time in *in vitro* or *in vivo* models of infection can help in unraveling, with clear implications for the development of therapies and vaccines.

Microarray- and tiling array-based studies provided the first insights on transcriptome changes potentially relevant for infection in different pathogens, such as *Vibrio cholerae* (Merrell *et al.*, 2002), *Borrelia burgdorferi* (Revel, Talaat and Norgard, 2002), *Chlamydia trachomatis* (Belland *et al.*, 2003), *Chlamydia pneumoniae* (Mäurer *et al.*, 2007) and *Salmonella enterica* (Eriksson *et al.*, 2003; Hautefort *et al.*, 2008), and helped in the investigation of non-coding transcripts and new virulence genes in streptococci (Perez *et al.*, 2009; Kumar *et al.*, 2010; Zheng *et al.*, 2011). Similarly, these technologies were used to start the investigation on host transcriptional changes upon incubation with several pathogens. Human promyelocytic cell THP1 transcriptome was analyzed following infection with *Listeria monocytogenes* (Cohen *et al.*, 2000), while both macrophages and epithelial cells were investigated for their response to *Salmonella* infection (Rosenberger *et al.*, 2000; Eckmann *et al.*, 2000). Other studies reported transcriptional dynamics in A549 pneumocyte cell line and in the transformed human bronchial epithelial one BEAS-2B upon interaction with *Pseudomonas aeruginosa* and *Bordetella pertussis*, respectively (Ichikawa *et al.*, 2000; Belcher *et al.*, 2000). Interestingly, Jenner and Young collected in a review 32 of these published studies, representing 785 experiments and 77 different host-pathogen interactions, in order to define through cluster analysis a common host response to the infection (Jenner and Young, 2005). Despite the high variability in host cell type (macrophages, dendritic cells, T cells, B cells, Peripheral Blood Mononuclear Cells (PBMCs), endothelial cells, epithelial cells and others) and in pathogens considered (different species of bacteria, different viruses, yeasts and protozoa) (Table 1), the authors detected in total 511 commonly regulated genes, highlighting an existing shared transcriptional program that is activated in response to exposure to several pathogens. Figure 1 depicts the different functional groups in which these genes can be classified: as might be expected, genes that encodes cytokines, chemokines and in general pro-inflammatory mediators constituted one of these groups, others comprised IFN-stimulated genes, genes encoding transcriptional regulators and components of signaling pathways that might mediate the activation of the immune response, but also genes in some way limiting these events, probably contributing to negative feedback loops that allow the cells to go back to the inactivated state. A similar balancing effect was deduced also from the presence among the commonly regulated genes of factors both positively or negatively activating apoptosis, suggesting that it can be both initiated and prevented depending on how the infection process goes on. Finally, there were groups of genes concerning lymphocyte activation, antigen presentation, cell adhesion and tissue invasion.

In addition to the aforementioned excessive cost needed for tiling array experiments, dual transcriptomics for the elucidation of host-pathogen interactions was limited on these hybridization-based technologies due to probe cross-reactivity between host and pathogen cDNAs. To deal with such problem, either cross-hybridizing clones needed to be removed or the RNA from the pathogen and the host needed to be analyzed separately. However, in order to have the RNA from one of the two, usually that of the other was sacrificed (for instance, isolating RNA from bacteria meant to lose the eukaryotic one during the lysis of the host cells), so limiting our ability in precisely catching the concurrent response in the two interacting organisms. Despite this, different examples of the application of these methods to simultaneously assess mRNA changes in the host and in the pathogen are available in literature: separate arrays have been used to study from a transcriptional point of view *Aspergillus fumigatus* and human airway epithelial cells (Oosthuizen *et al.*, 2011), while arrays containing both host and pathogen probes were used to characterize *Escherichia coli* CP9 infection of mouse models (Motley *et al.*, 2004) and *Plasmodium berghei* infection of different mouse tissues (Lovegrove *et al.*, 2006). Some probe-independent, tag-based sequencing methods, such as serial or cap analysis of gene expression (SAGE or CAGE) and their variant SuperSAGE allowing longer tags, were applied to the field of infection biology, also for simultaneous characterizations, and partly solved the mentioned problems, but soon the numerous advantages provided by the new-emerging RNA-Seq imposed this technique as the elected one to routinely study transcriptomes.

RNA-Seq was initially used to characterize bacterial transcriptomes associated to particular stages and conditions experienced during host infection. In particular, a differential RNA-Seq approach, discriminating newly generating primary transcripts (most mRNAs and small non-coding RNAs) from processed RNA species (rRNAs and tRNAs) through exonuclease degradation of the latter (containing a 5’ monophosphate), was applied to investigate *Helicobater pylori* transcriptome during exponential phase and acid stress (Sharma *et al.*, 2010) and to different *Chlamydia* spp. to study the two differentiated states, elementary bodies and reticular bodies, observed during infections by this organism (Albrecht *et al.*, 2010, 2011). An RNA-Seq study of *Vibrio cholerae* isolated from the caecum of infected rabbit or the intestine of infected mice was also conducted in order to understand transcriptional changes characterizing the *in vivo* pathogenesis of this bacterium (Mandlik *et al.*, 2011). From the host point of view, examples of RNA-Seq applications to understand responses to infections are fewer, but in some cases also focused on non-coding transcriptional profiles, as for the attempt to elucidate miRNA response in macrophages or HeLa cells upon infection by *Salmonella enterica* (Schulte *et al.*, 2011).

Table 1 | List of datasets used to investigate the host common response to the infection. From Jenner and Young, 2005.



Dataset	Cell type(s)	Stimuli
1	Macrophage	*E. coli*, EHEC, *S. typhi*, *S. typhimurium*, *S. aureus*, *L. monocytogenes*, *M. tuberculosis*, *M. bovis* BCG, LPS, LTA, MDP, TB Hsp70, BCG Hsp65, MPA, fMLP, protein A, mannose
2	Macrophage	IFNα, IFNβ, IFNγ, IL10, IL12
3	Macrophage	*L. chagasi*
4	Macrophage	*S. typhimurium*, *S. typhimurium phoP–*
5	Macrophage	*B. pertussis* 338, *B. pertussis* 537 (avirulent), *B. pertussis* AC–, *B. pertussis* PT–
6	DC	*E. coli*, influenza, *C. albicans*, LPS, poly I:C, mannan
7	DC	HIV-1, Tat
8	DC	GMCSF+IL4+TNFα+IL1β+PGE2
9	Leukocyte	*S. aureus*, LPS
10	PBMC	LPS, *B. pertussis* 338, *B. pertussis* Minnesota 1, *E. coli*, *S. aureus*, ionomycin+PMA, *B. pertussis* 338 dead, *B. pertussis* 338 alive.
11	PBMC	SIV
12	Whole blood	*N. meningitidis*
13	T cell	HIV-1
14	T cell	HIV-1
15	T cell	VZV
16	B cell	EBV, KSHV
17	Liver (HCC)	HBV, HCV
18	Liver	HCV
19	Trigeminal ganglia	HSV-1
20	Astrocyte	JC virus
21	Fibroblast	VZV
22	Fibroblast	HCMV, gB
23	Endothelial cell	Dengue virus
24	Endothelial cell	KSHV
25	Endothelial cell, skin	KSHV
26	Skin	VZV
27	Keratinocyte	HPV-31
28	Epithelial cell	*B. pertussis*
29	Epithelial cell	RSV
30	Epithelial cell	RSV
31	Epithelial cell	RSV
32	Epithelial cell	RRV
33	Epithelial cell	Bovine papilloma virus E2
34	Epithelial cell	*H. pylori*
35	Stomach	*H. pylori*

+ indicates cells that were treated with all the agents simultaneously, rather than individually, as shown for other studies. BCG Hsp65, bacille Calmette–Guérin heat shock protein 65; *B. pertussis* AC–, *Bordetella pertussis* adenlylate cyclase mutant; *B. pertussis* PT–, *Bordetella pertussis* pertussis toxin mutant; *C. albicans, Candida albicans;* DC, dendritic cell; EBV, Epstein-Barr virus; *E. coli*, *Escherichia coli*; EHEC, enterohaemorrhagic *E. coli*; fMLP, formyl-methionine-leucine-phenylalanine; gB, glycoprotein B; GMCSF, granulocyte-macrophage colony-stimulating factor; HBV, hepatitis B virus; HCC, hepatocellular carcinoma; HCMV, human cytomegalovirus; HCV, hepatitis C virus; HPV, human papilloma virus; *H. pylori, Helicobacter pylori;* HSV, herpes simplex virus; IFN, interferon; IL, interleukin; KSHV, Kaposi's sarcoma-associated herpesvirus; *L. chagasi, Leishmania chagasi; L. monocytogenes; Listeria monocytogenes*, LPS, lipopolysaccharide; LTA, lipoteichoic acid; MDP, muramyl dipeptide; MPA, monophophoryl lipid A; *M. tuberculosis, Mycobacterium tuberculosis; M. bovis* BCG, *Mycobacterium bovis* bacille Calmette–Guérin; *N. meningitidis*, *Neisseria meningitidis;* PBMC, peripheral blood mononuclear cell; PGE2, prostaglandin E2; PMA, phorbol myristate acetate; poly I:C, polyinosinic-polycytidylic acid; RRV, rhesus rotavirus; RSV, respiratory syncytial virus; *S. aureus*, *Staphylococcus aureus;* SIV, simian immunodeficiency virus; *S. typhi*, *Salmonella typhi*; *S. typhimurium phoP–*, *Salmonella typhimurium phoP* mutant; TB Hsp70, *Mycobacterium tuberculosis* Hsp70; TNF, tumour-necrosis factor; VZV, varicella-zoster virus.

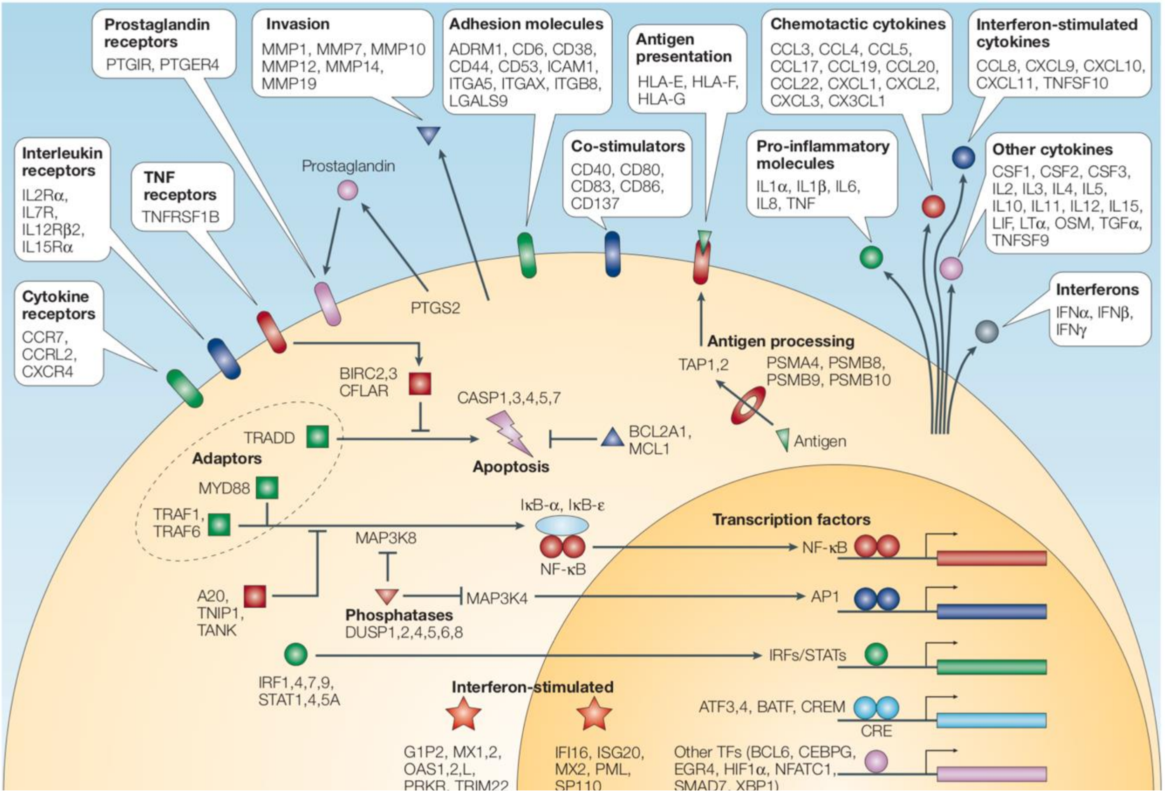

Prostaglandin receptors
PTGIR, PTGER4
Invasion
MMP1, MMP7, MMP10 MMP12, MMP14, MMP19
Adhesion molecules
ADRM1, CD6, CD38, CD44, CD53, ICAM1, ITGA5, ITGAX, ITGB8, LGALS9
Antigen presentation
HLA-E, HLA-F, HLA-G
Chemotactic cytokines
CCL3, CCL4, CCL5, CCL17, CCL19, CCL20, CCL22, CXCL1, CXCL2, CXCL3, CX3CL1
Interferon-stimulated cytokines
CCL8, CXCL9, CXCL10, CXCL11, TNFSF10
Interleukin receptors
IL2Rα, IL7R, IL12Rβ2, IL15Rα
TNF receptors
TNFRSF1B
Prostaglandin
Co-stimulators
CD40, CD80, CD83, CD86, CD137
Pro-inflammatory molecules
IL1α, IL1β, IL6, IL8, TNF
Other cytokines
CSF1, CSF2, CSF3, IL2, IL3, IL4, IL5, IL10, IL11, IL12, IL15, LIF, LTα, OSM, TGFα, TNFSF9
Cytokine receptors
CCR7, CCRL2, CXCR4
PTGS2
Interferons
IFNα, IFNβ, IFNγ
Antigen processing
TAP1,2
PSMA4, PSMB8, PSMB9, PSMB10
BIRC2,3 CFLAR
CASP1,3,4,5,7
TRADD
Adaptors
BCL2A1, MCL1
Antigen
Apoptosis
MYD88
TRAF1, TRAF6
IκB-α, IκB-ε
Transcription factors
MAP3K8
NF-κB
NF-κB
A20, TNIP1, TANK
MAP3K4
AP1
Phosphatases
DUSP1,2,4,5,6,8
IRFs/STATs
IRF1,4,7,9, STAT1,4,5A
Interferon-stimulated
ATF3,4, BATF, CREM
CRE
G1P2, MX1,2, OAS1,2,L, PRKR, TRIM22
IFI16, ISG20, MX2, PML, SP110
Other TFs (BCL6, CEBPG, EGR4, HIF1α, NFATC1, SMAD7, XBP1)

Figure 1 | Host common response to the infection. Genes belonging to the common host response, as individuated by cluster analysis on 785 hybridization-based experiments interrogating 77 different host-pathogen interactions, and their functions inside and outside a single cell (Jenner and Young, 2005).

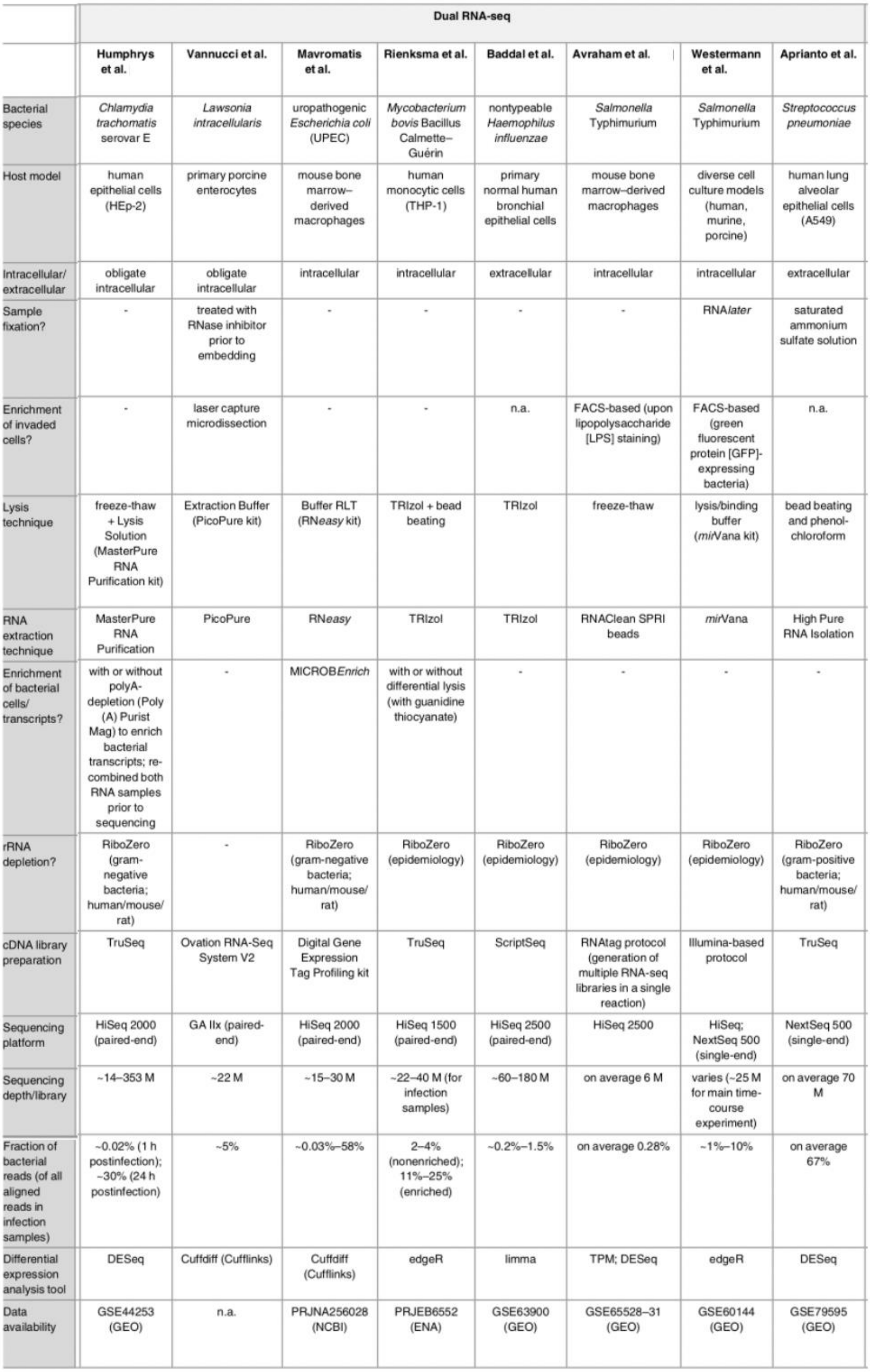
1.1.3 NEXT-GENERATION DUAL TRANSCRIPTOMICS

The first study successfully implying dual RNA-Seq investigated transcriptomes from a eukaryotic pathogen, the fungus *C. albicans*, and mouse dendritic cells (Tierney *et al.*, 2012). The scope of the work was primarily to build an interspecies regulatory network describing from a molecular point of view host-pathogen interactions, rather than an in-depth transcriptional characterization, therefore the sequencing depth was quite low (around 120 million reads of 36 bp in total for five time points) and previous knowledge from the literature was integrated in the method in order to select relevant candidates for interspecies regulatory interactions. Importantly, this demonstrates that since the first time this technique has been successfully applied, both the full potential behind these data and the need for new analysis strategies able to deeply interrogate them were already clear.

There are some potential limiting factors that need to be considered when discussing dual RNA-Seq experiments. The key technical issue surely concerns the difference in both nature and content between host and bacterial RNA. The human genome size is around 3.2 Gb, while bacterial genomes usually do not exceed 5 Mb. Such difference naturally translates into a different amount of RNA per cell. A mammalian cell typically contains 10-20 pg of total RNA, that means about two order of magnitude more than a single bacterial cell, typically presenting around 0.1 pg of RNA. However, in practice this difference decreases, since a single eukaryotic cell is associated with or invaded by multiple bacterial cells. Of course, this varies depending on the considered infection model, but on average it can be assumed that a single host cell is associated with at least ten bacterial ones, so that typically it is possible a decrease in the mentioned difference to one order of magnitude. However, considering also variable infection rates, all together this usually leaves a small fraction of informative bacterial reads (Table 2). Different solutions have been used to overcome this problem, for instance by sequencing libraries to high depth or enriching for invaded host cells by fluorescence-activated cell sorting (FACS) (Avraham *et al.*, 2015; Westermann *et al.*, 2016). It is advisable to estimate the relative concentration of host and bacterial RNA in samples before sequencing, for instance by quantitative real-time PCR, in order to guide decisions on required read depth or whether to increase the multiplicity of infection (MOI) of the infection protocol. Furthermore, there is also a high heterogeneity in RNA species characterizing the two organisms, with the highest diversity being in the non-coding repertoire of the two. Common species differ as well, such as mRNAs in their polyadenylation amount and function, so that enriching for the polyadenylated fraction will result in transcripts with different fates: stable transcripts for the host and transcripts undergoing degradation for the pathogen. As it is well-known, for both the organisms the RNA pool consists primarily of rRNA (>80%), with mRNA constituting only a minor fraction (<5%). Frequently, rRNA is depleted using kits specifically designed for the host and the pathogen, that usually exploits sequence-specific oligonucleotides bound to magnetic beads. However, commercial kits have different efficiencies (especially the kits for the pathogens usually are less efficient, leaving also more than half of the total bacterial reads mapping to rRNAs) and may add biases. Furthermore, in a dual RNA-Seq, the two different kits have to be used in succession and each depletion steps also decreases the final amount of non-rRNA transcripts. Other important considerations include the use of stranded protocols, critical for detection of non-coding and antisense transcriptions (particularly high in bacteria), and the sequencing depth, that at the beginning was theoretically assessed to be necessary quite high to capture bacterial gene expression in the host background. Today, because of improved strategies and sequencing of longer reads, many dual RNA-Seq protocols demonstrated to provide informative data with as few as ~25 millions of reads per sample (Table 2). Finally, dual RNA-Seq are complex experiments where samples are subjected to various treatments that may differ in their effect, for instance we need to consider variability in the infection, and may or may not be exposed to some protocols, such as cell sorting, therefore there are many possibilities of introducing unwanted variations and it could be advisable to assess them in order to introduce such factors into the design for differential expression analyses.

1.1.4 DUAL RNA-SEQ STUDIES AND PERSPECTIVES

Since the appearance of the first aforementioned study, dual RNA-Seq started to be applied to different infection models and the number of available works in literature implying this technique rapidly begun to grow. An early study of HEp-2 epithelial cells infected by *Chlamydia trachomatis* serovar E generated new hypotheses on the strategies adopted by this pathogen at the beginning of the infection, for instance on early iron acquisition (Humphrys *et al.*, 2013). Previously, these observations were not possible because of the small number of infecting organisms in the culture that the limitations of arrays made impracticable for accurate investigations. From the host point of view, this work helped in deeply assessing the transcriptional response to the infection, contrary to the previous array-based studies. Interestingly, the same year a publication appeared coupling laser capture microdissection with dual RNA-Seq to study host-pathogen interactions *in vivo* upon *Lawsonia intracellularis* infection (Vannucci, Foster and Gebhart, 2013). Soon, dual RNA-Seq started to be used to decipher also later stages of infections, as it is the case of a work on uropathogenic *Escherichia coli*-infected mouse macrophages followed over a 24h time course (Mavromatis *et al.*, 2015), and the one on nontypeable *Haemophilus influenzae* infecting for 72h a ciliated human bronchial epithelium reconstituted *in vitro* (Baddal *et al.*, 2015). A work by Rienksma et al. on *Mycobacterium tuberculosis* can be reported as one of the first examples of the need for the application of specific protocols to enrich bacterial RNA, since this intracellular pathogen is characterized by a particularly unfavorable host-to-pathogen ratio (Rienksma *et al.*, 2015). The new technique was also applied to study changes regarding the non-coding transcriptome of infecting bacteria, for instance for *Salmonella typhimurium* in a study by Westermann and colleagues (Westermann *et al.*, 2016). They focused on the highly induced, and previously uncharacterized, PinT small RNA, predicting its activation by the PhoP/Q two-component system and, through another dual RNA-Seq that used PinT deletion mutants, its activity as a post-transcriptional regulator for the expression of many important virulence genes. The field is rapidly moving on and the technique shows a considerable potential in unraveling the heterogeneity in responses and interactions regarding the infection processes. A recent study reported transcriptional signatures for the host and the pathogen varying with mouse lung macrophage ontology upon *Mycobacterium tuberculosis* infection (Pisu *et al.*, 2020). However, one of the most exciting advancement is surely represented by coupling dual RNA-Seq with the growing field of single cell RNA-Seq (scRNA-Seq), as already anticipated by a study published in 2015, where a functional heterogeneity in the response of individual host macrophages emerged because of the heterogeneous activity of bacterial factors in individual infecting *Salmonella* cells (Avraham *et al.*, 2015). Finally, the field is also expanding towards the simultaneous characterization of transcriptomes from the host and different co-infecting pathogens, a technique recently described as triple RNA-Seq (Seelbinder *et al.*, 2020), while examples of metatranscriptomic characterizations of the microbiota from a given anatomical site, together with the simultaneous profiling of the host transcriptional signature at that site, quite early started to appear in the literature (Pérez-Losada *et al.*, 2015).



	Dual RNA-seq							
	Humphrys et al.	**Vannucci et al.**	**Mavromatis et al.**	**Rienksma et al.**	**Baddal et al.**	**Avraham et al.**	**Westermann et al.**	**Aprianto et al.**
Bacterial species	*Chlamydia trachomatis* serovar E	*Lawsonia intracellularis*	uropathogenic *Escherichia coli* (UPEC)	*Mycobacterium bovis* Bacillus Calmette–Guérin	nontypeable *Haemophilus influenzae*	*Salmonella* Typhimurium	*Salmonella* Typhimurium	*Streptococcus pneumoniae*
Host model	human epithelial cells (HEp-2)	primary porcine enterocytes	mouse bone marrow–derived macrophages	human monocytic cells (THP-1)	primary normal human bronchial epithelial cells	mouse bone marrow–derived macrophages	diverse cell culture models (human, murine, porcine)	human lung alveolar epithelial cells (A549)
Intracellular/ extracellular	obligate intracellular	obligate intracellular	intracellular	intracellular	extracellular	intracellular	intracellular	extracellular
Sample fixation?	-	treated with RNase inhibitor prior to embedding	-	-	-	-	RNA*later*	saturated ammonium sulfate solution
Enrichment of invaded cells?	-	laser capture microdissection	-	-	n.a.	FACS-based (upon lipopolysaccharide [LPS] staining)	FACS-based (green fluorescent protein [GFP]-expressing bacteria)	n.a.
Lysis technique	freeze-thaw + Lysis Solution (MasterPure RNA Purification kit)	Extraction Buffer (PicoPure kit)	Buffer RLT (RN*easy* kit)	TRIzol + bead beating	TRIzol	freeze-thaw	lysis/binding buffer (*mir*Vana kit)	bead beating and phenol-chloroform
RNA extraction technique	MasterPure RNA Purification	PicoPure	RN*easy*	TRIzol	TRIzol	RNAClean SPRI beads	*mir*Vana	High Pure RNA Isolation
Enrichment of bacterial cells/ transcripts?	with or without polyA-depletion (Poly (A) Purist Mag) to enrich bacterial transcripts; re-combined both RNA samples prior to sequencing	-	MICROB*Enrich*	with or without differential lysis (with guanidine thiocyanate)	-	-	-	-
rRNA depletion?	RiboZero (gram-negative bacteria; human/mouse/ rat)	-	RiboZero (gram-negative bacteria; human/mouse/ rat)	RiboZero (epidemiology)	RiboZero (epidemiology)	RiboZero (epidemiology)	RiboZero (epidemiology)	RiboZero (gram-positive bacteria; human/mouse/ rat)
cDNA library preparation	TruSeq	Ovation RNA-Seq System V2	Digital Gene Expression Tag Profiling kit	TruSeq	ScriptSeq	RNAtag protocol (generation of multiple RNA-seq libraries in a single reaction)	Illumina-based protocol	TruSeq
Sequencing platform	HiSeq 2000 (paired-end)	GA IIx (paired-end)	HiSeq 2000 (paired-end)	HiSeq 1500 (paired-end)	HiSeq 2500 (paired-end)	HiSeq 2500	HiSeq; NextSeq 500 (single-end)	NextSeq 500 (single-end)
Sequencing depth/library	~14–353 M	~22 M	~15–30 M	~22–40 M (for infection samples)	~60–180 M	on average 6 M	varies (~25 M for main time-course experiment)	on average 70 M
Fraction of bacterial reads (of all aligned reads in infection samples)	~0.02% (1 h postinfection); ~30% (24 h postinfection)	~5%	~0.03%–58%	2–4% (nonenriched); 11%–25% (enriched)	~0.2%–1.5%	on average 0.28%	~1%–10%	on average 67%
Differential expression analysis tool	DESeq	Cuffdiff (Cufflinks)	Cuffdiff (Cufflinks)	edgeR	limma	TPM; DESeq	edgeR	DESeq
Data availability	GSE44253 (GEO)	n.a.	PRJNA256028 (NCBI)	PRJEB6552 (ENA)	GSE63900 (GEO)	GSE65528–31 (GEO)	GSE60144 (GEO)	GSE79595 (GEO)

Table 2 | List of selected dual RNA-Seq studies. "M", million; "TPM", transcripts per million; "NCBI", National Center for Biotechnology Information; "ENA", European Nucleotide Archive; "GEO", Gene Expression Omnibus. Modified from Westermann et al., 2017.

1.2 INTERROGATING THE HOST-PATHOGEN CROSSTALK FROM DUAL RNA-SEQ DATA

1.2.1 BIOINFORMATICS PIPELINE FOR DUAL RNA-SEQ EXPERIMENTS

The standard bioinformatics analysis pipeline for dual RNA-Seq experiments doesn't differ dramatically from the one typically applied to RNA-Seq studies, with some differences naturally due to the "duality" of the data (Figure 2). The basic idea behind these experiments is that the total RNA pool is collected from infected cells and the discrimination between host and pathogen reads is performed *in silico*. First, FastQ files originating from the sequencing machine are subjected to a deep quality check and eventually trimmed to remove low-quality bases and adapters. However, it should be beared in mind that over-trimming can introduce biases and reduce the statistical significance of differentially expressed (DE) genes (Williams *et al.*, 2016). Next, reads are mapped to the reference genome or transcriptome (if available from repositories such as NCBI, UCSC and Ensembl) for the two organisms. Different strategies can be used to minimize cross-mapping phenomena: for combined reads, the probability of belonging to the host are higher and eventually a mistake will influence less the results compared to erroneously assigning sequences to the usually small pool of bacterial reads, therefore reads can be first mapped to the host genome and next only the unmapped ones can be aligned to the bacterial genome (Baddal *et al.*, 2015); or, all reads can be mapped to both genomes in parallel, in order to precisely quantify cross-mapping sequences and remove them (Westermann *et al.*, 2016); finally, a chimeric sequence comprehensive of concatenated host and pathogen genomes can be assembled and used to directly map all reads against (Aprianto *et al.*, 2016). In practice, with Illumina reads of at least 75bp and considering infection models with mammalian host cells and bacterial pathogens, cross-mapping is expected to be usually negligible, with most of this originating from rRNA and tRNA loci (Westermann *et al.*, 2016). The non-splice-aware short-read aligner Bowtie2 (Langmead *et al.*, 2009) is usually sufficient for bacterial read mapping, but also other non-splice-aware alternatives can be used, such as SEAL (Pireddu, Leo and Zanetti, 2011) and SOAP2 (Li *et al.*, 2009). On the contrary, powerful splice-aware algorithms are typically used for read mapping on the host genome, examples are STAR (Dobin *et al.*, 2013), HISAT2 (Kim, Langmead and Salzberg, 2015) and TopHat2 (Trapnell, Pachter and Salzberg, 2009). After the mapping, alignment quality statistics are produced, particularly assessing relative mapping percentages for the two organisms, and it's also advisable to check a uniform read distribution along genomes, through genome viewers like Integrated Genome Viewer (IGV), as well as the percentages of read mapping to rRNA sequences, that is ribodepletion efficacy for both the host and the pathogen protocol. A supplementary, more than alternative, and widely used approach is the *in silico* removal of rRNA reads. However, many of the tools today used for DE analyses, such as DeSeq2, exploit normalization procedures that seem to be very robust to high and also variable levels of rRNAs in samples, and even applying the *in silico* removal, the critical issue always will be to understand if the remaining informative reads are enough to adequately cover the other genes. So, the wet lab depletion remains the most important step. From the alignment files, both host and pathogen count matrices are produced on the basis of a reference annotation, basically quantifying the number of reads aligned to biologically meaningful features, like genes.

Once that matrices of genes (rows) and conditions (columns) are obtained, differences between samples are investigated to detect potential outliers and to guide decisions for the design of the following DE analysis, typically using different diagnostic plots and Principal Component Analysis (PCA), a commonly used technique to reduce dimensionality of large data sets while preserving the highest variability that characterizes them. Next, statistically significant DE genes from the contrasts of interest can be detected using different tools, such as DESeq2 (Love, Huber and Anders, 2014), edgeR (Robinson, McCarthy and Smyth, 2010), Cufflinks (Trapnell *et al.*, 2013), BaySeq (Hardcastle and Kelly, 2010), Salmon (Patro *et al.*, 2017) and Kallisto (Bray *et al.*, no date), that usually model RNA-Seq counts via a negative binomial distribution and apply distinct statistical methods to calculate reliable dispersion estimates. For example, DESeq2 builds a generalized linear model (GLM) describing the dependency of the parameters that characterize the count distribution from the conditions under investigation. Alternatively, also limma can be used, a package originally developed for microarrays that attempts to correctly model the mean-variance relationship between samples to achieve a more probabilistic distribution of the counts (Smyth, 2005). Finally, functional enrichment analyses are usually conducted to search for Gene Ontology (GO) terms or Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways enriched for significantly regulated genes.

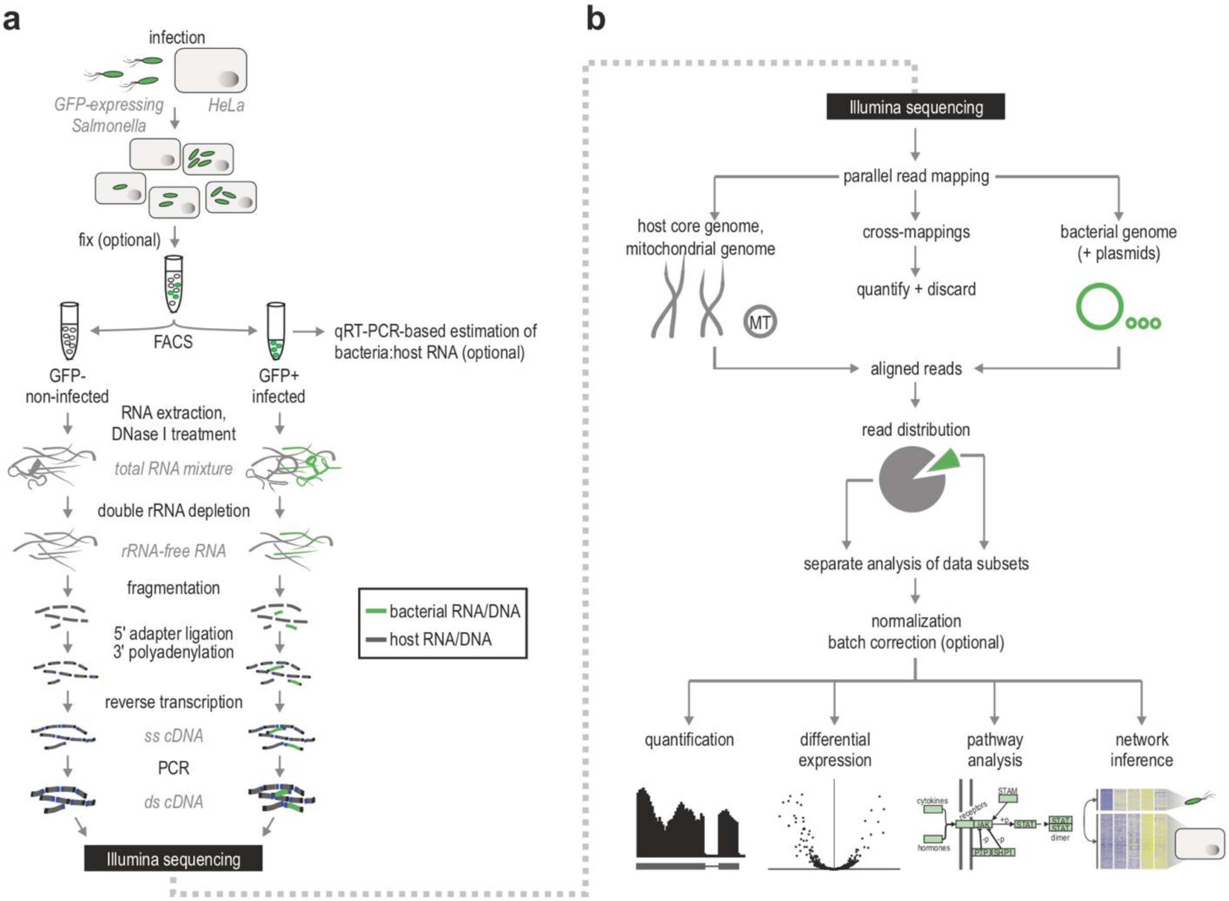


Figure 2 | General protocol and pipeline for dual RNA-Seq data preparation and analysis. A) Example of a wet lab workflow for dual RNA-Seq data preparation for the sequencing. **B)** General pipeline for dual RNA-Seq data analysis. (Westermann and Vogel, 2018).

1.2.2 INTEGRATING HOST AND PATHOGEN TRANSCRIPTOMICS INFORMATION

As described, the classical bioinformatics pipeline, after discriminating between host and pathogen reads, proceeds in parallel, always keeping separate the two transcriptomes in the analysis. The transcriptomic picture for both the organisms derived by DE analyses clearly represent the standard way to start to decipher how the two are responding to the different stresses and events that concern the infection process. However, it's clear that strongly relevant messages could emerge from the integration of the two information, and several studies in literature already report some attempts in this direction.

The first aforementioned dual RNA-Seq study was also the first one performing an integrated analysis of host and pathogen transcriptomic data (Tierney *et al.*, 2012). The authors inferred an interspecies regulatory network characterizing C. albicans infection of Mus musculus dendritic cells by NetGenerator tool (Guthke *et al.*, 2005; Linde *et al.*, 2010). On the basis of a set of linear differential equations, this tool models the temporal change of the expression intensity of a gene at a certain time point as the weighted sum of the expression intensities of all the other genes and an external stimulus, modeled as a stepwise constant function representing the change from no host-pathogen interaction to the onset of it, at the same time point. The aim is to identify the best network structure that fits the measured expression data, while minimizing the predicted interactions. Prior knowledge on putative regulatory interactions can be integrated by the tool, that scores the proposed interaction based on the confidence of that knowledge. However, such an approach based on differential equations is inappropriate for large-scale modeling, because incorporating a large number of genes would result in a big number of parameters to be identified and could lead to over-fitting. Therefore, the number of genes in this study was restricted based on available biological knowledge. However, one of the potentialities of dual RNA-Seq data is the ability to capture relevant signal also from the (numerous, in the case for instance of non-reference bacterial strains) uncharacterized genes.

Interestingly, Zhu and colleagues developed a mixture model-based framework to cluster host and pathogen genes on the basis of their co-expression pattern (Zhu *et al.*, 2013). The idea was to capture the reciprocal plasticity in host and pathogen gene expression to unravel the crosstalk occurring between the two during the course of the infection. However, this approach doesn't take in consideration the fact that host and pathogen temporal expression patterns can be shifted one compared to the other, because of a delay in the response by one of the two, and the possibility also of opposite regulations that can be in some way linked to each other. Furthermore, the developed quantitative model was based on Poisson distribution, while today we know that RNA-Seq data are more correctly modeled using negative binomial distributions. However, determining such parametric model for gene expression data is complicated by the many noise factors that affect this kind of measurements. On the contrary, Westermann and colleagues clustered host and pathogen genes with similar expression kinetics across the time course of HeLa-S3 cells infected by *Salmonella enterica* by Pearson correlation (Westermann *et al.*, 2016).

Graph theory already revealed its potential in many fields of biological studies, and different works in literature tried to apply graph theory to address also this problem. Jae Lee and colleagues, in the attempt of studying malaria through dual transcriptomics on blood of infected patients, built a co-expression network comprehensive of both host and pathogen genes (Lee *et al.*, 2018). Co-expression networks are widely used to capture correlations in the measured gene expression levels, that usually indicate functional correlation and involvement in the same process. However, co-expression networks are typically quite noisy, without directly conferring information about causality, especially when considering together two different systems like the host cell and the pathogen one. Using $\log_2$ ratios of *Haemophilus ducreyi* and host differentially expressed genes, Griesenauer et al. incorporated the duality of the dataset in the analysis by the reconstruction of bipartite graphs connecting genes from the two organisms showing positive or negative correlation along the observed time points (Griesenauer *et al.*, 2019). Notably, a general issue related to all these correlation-based strategies is that the sample size in these studies is usually very small, and the authors of this work, for example, were compelled to use stringent p-values and r cutoff values.

1.2.3 MAIN CHALLENGES IN DEVELOPING INTEGRATIVE METHODS FOR THE ANALYSIS OF DUAL RNA-SEQ DATA SETS

One of the main challenges in developing new methods that look at this kind of data in a more integrative way is linked to the nature of these experiments. Usually, the aim of these studies is to assess host and pathogen transcriptional dynamics as the infection proceeds, therefore they are designed in a time course fashion, but the number of time points, as is the case also of many standard RNA-Seq studies, usually is very small, mainly due to the higher cost that bigger experiments would require or to difficulties in obtaining large amounts of biological materials. This results in a strongly unbalanced matrix, where thousands of variables (expression of a large number of genes) are profiled in very few samples, adding an extra layer of complexity to any exploratory approach. Interestingly, Ernst and his group developed a strategy to cluster short time series gene expression data, that makes use of permutation tests to avoid the risk that some patterns of regulation arise at random (Ernst, Nau and Bar-Joseph, 2005a; Ernst and Bar-Joseph, 2006). Initially, data are properly normalized and transformed to represent them as time point fold changes respect to the time 0 experiment (that will always be 0), and a set of representative temporal profiles is defined from all the possible ones through a maximum-minimum approach, that basically implies the iteration of a pre-defined number of selections, each time choosing the temporal pattern that is farthest, on the basis of a certain distance measure, from all those selected so far. Next, the genes, filtered for those that show a modulation along the time course and consistency among replicates, are assigned to the most similar profile by correlation. Finally, a permutation test is used to remove dependency between time points and quantify the number of expected genes that would have been assigned to each of the model profiles if data were generated at random, a value that is basically derived by the mean of the number of genes assigned to a profile in all the possible permutations (since the number of time points is low, all the permutations can be computed). At the end, since each gene has been assigned to a pattern, it is assumed that the number of genes in each profile is distributed as a binomial random variable, with the parameters being represented by the total number of genes and, as probability of success, the ratio of expected genes assigned to that profile compared to the total number of genes. In this way, on the basis of a pre-defined significance threshold and correcting for multiple testing, p-values can be computed to detect non-casual temporal patterns of gene expression that are enriched in the data set of interest.

Another problem in developing these methods is more a biological/theoretical one. With the advent of high-throughput technologies, that made rapidly available a large amount of data, complexity, together with strategies and attempts to tackle it, suddenly pervaded biological fields. Properly defining a complex system is quite a difficult task. Briefly, we can describe it as an entity made up of many smaller sub-entities that shows properties or behaviors that cannot be attributed to, or studied from, the single components, but they are emerging from their interaction, and so uniquely affordable by looking at the complete picture. Infection is a composite event initially triggered by the encounter between two complex systems, the host cell and the pathogen one, that once they begin to interact, they cannot be anymore considered as separate entities, since in some way during the process they start to co-evolve and influence each other, and this is even more evident when considering pathogens that are also characterized by an intracellular lifestyle. Therefore, such a complex picture is not so easy to be extrapolated and interpreted from data. A powerful way to represent and investigate complex systems is through the branch of mathematics called graph theory. Properties that can describe graphs, or networks, and investigations that can be conducted on them are many, here just the basis of graph theory will be disclosed, particularly referring to "Network Science" by A. L. Barabàsi. A graph, or network, is a representation of system's components, called nodes or vertices, and the interactions occurring between them, called links or edges. The total number of nodes is usually indicated with N, while the total number of links among them with L. Both nodes and edges present properties, that can vary depending on the network. Edges, for instance, can be weighted on the basis of a defined measure, so that a link between node i and j carries a weight w_{ij}, and can be directed or not. A key and commonly used property of a node is its degree, usually indicated by k_i for node i, describing how much it is connected with other nodes in the graph. If we move at the network level, we can define an average degree, that for undirected graphs is given by

$$< k >= \frac{1}{N}\sum_{i=1}^{N} k_i = \frac{2L}{N} \tag{1}$$

In directed networks, it is possible to distinguish between incoming degree, k_i^{in}, so the number of links pointing to node i, and outgoing degree, k_i^{out}, so the number of links that exit from node i. Therefore, the node's total degree is given by

$$k_i = k_i^{in} + k_i^{out} \tag{2}$$

while the average degree of the directed graph can be described with

$$< k^{in} >= \frac{1}{N}\sum_{i=1}^{N} k_i^{in} =< k^{out} >= \frac{1}{N}\sum_{i=1}^{N} k_i^{out} = \frac{L}{N} \tag{3}$$

If the graph is weighted, this can be taken into account and we can refer to the strength of a node, given by the sum of weights on its edges. Many networks properties and insights are derived through calculation of the degree distribution p_k, that is given by

$$p_k = \frac{N_k}{N} \tag{4}$$

where N_k is the number of degree-k nodes.

Furthermore, we refer to a complete graph if each node is connected to every other node, but in real networks L is usually much smaller than L_{max}, defined by

$$L_{max} = \frac{N(N-1)}{2} \tag{5}$$

and we call them sparse networks.

Another widely used concept related to graph theory is the one of paths, that are routes that run along the links of the network. A path's length is the number of links inside the path, and we called shortest path between node i and j the one with the fewest number of links, or it can also be called distance between i and j, indicated by d_{ij}. Of course, in directed graphs, the paths follow the direction of links. The network diameter is the maximum shortest path in the network, and similarly to what described before for node degree, we can also define an average path length, that is the average distance between all pairs of nodes in the network, described by

$$< d > = \frac{1}{N(N-1)} \sum_{i,j=1,N;i\neq j} d_{i,j} \qquad (6)$$

Different solutions and centrality measures can be used to prioritize nodes and links in networks. Hubs, for instance, indicating nodes with high degree, have clear implications for the topology of a network. Another frequently used centrality measure is named betweenness centrality, to refer to the extent to which a certain node lies on the shortest path between other nodes. The betweenness centrality for node k is defined by

$$b_k = \frac{1}{2} \sum_{i\neq k}^{n} \sum_{j\neq k,i}^{n} \frac{g_{ikj}}{g_{ij}} \qquad (7)$$

where g_{ij} is the number of shortest paths from node i to node j and g_{ikj} is the number of shortest paths from i to j that pass through k. Often, betweenness centrality is normalized by dividing by

$$(N-1)(N-2) = N^2 - 3N + 2 \qquad (8)$$

that is the maximum betweenness that any node can achieve in a network with N nodes.

In some particular cases, nodes can be divided in two disjoint sets and a link can be allowed only between nodes belonging to the two different sets, in these cases we refer to a bipartite graph. Many real-world networks can be described by this special class of graphs, one example being represented by the human disease networks (Goh *et al.*, 2007), and their applications are varied. Maybe the most common method to deal with this kind of networks is to generate two projections, one of them connecting nodes of the first set if they are connected to at least one shared node in the second set, usually weighting their link on the basis of the number of nodes in the other set that they shared, and viceversa for the other projection. Next, the usual analysis techniques can be used on the resulting networks. However, even the double projection approach will necessarily lose the information regarding the interplay between the two sets of nodes. Therefore, an alternative strategy is to directly investigate the bipartite graph, redefining measures and approaches originally developed for classical, unipartite networks. In this direction, the normalization in calculating betweenness centrality needs to be modified to take in account the size of each node set. Therefore, we can calculate the two maxima to be used for normalization of the betweenness centrality value for the nodes in the two disjoint sets V_1 and V_2 by

$$\begin{aligned} bV_1 \max &= \frac{1}{2}\,[m^2(s+1)^2 + m(s+1)(2t-s-1) - t(2s-t+3)] \\ bV_2 \max &= \frac{1}{2}[n^2(p+1)^2 + n(p+1)(2r-p-1) - r(2p-r+3)] \end{aligned} \tag{9}$$

with n indicating the number of nodes in set V_1, m the number of nodes in set V_2, and

$$s = (n-1)\; div\; m, \qquad t = (n-1)\; mod\; m$$
$$p = (m-1)\; div\; n, \qquad r = (m-1)\; mod\; n$$

where *x div y* refers to the integer division of *x* by *y* and *x mod y* to the remainder of an integer division of *x* by *y* (Borgatti, no date).

1.3 THE HOST-ADAPTED HUMAN PATHOGEN *Neisseria gonorrhoeae*

1.3.1 SITES AND CHARACTERISTICS OF *N. gonorrhoeae* INFECTIONS

Neisseria gonorrhoeae, or gonococcus, is a Gram-negative diplococcus bacterium known to be the cause of gonorrhea, one of the major sexually transmitted infections (STIs), commonly spread via sexual contact. The population at higher risk includes individuals with multiple sexual partners having unprotected sex, with the highest incidence being associated to less developed countries. A first report of this infection disease can be found in a passage on sexual practices in the Book of Leviticus (chapter 15, verses 1 to 3) in the Old Testament, where women are warned to avoid men with discharges, probably referring to the purulent exudate from the penis representing the typical sign of the inflammatory response in men with gonorrhea (Shafer and Ohneck, 2011). Upon thousands of years of relationship, today this bacterium is a strict, host-adapted human pathogen, able to resist to many natural host defenses. Because of the estimate by World Health Organization (WHO) of a worldwide incidence of 86.9 million cases each year (Unemo *et al.*, 2019), a loud alarm related to the insurgence of strains resistant to most available antibiotics (including sulfonamides, penicillins, tetracyclins, macrolides and fluoroquinolones) (Unemo and Shafer, 2014) (Figure 3) and uncontrolled transmission in low-income countries or poorer communities, it is arising the idea about a risk of an imminent epidemic scenario concerning widespread untreatable gonorrhea, together with increased complications from the infection. Such complications in women can include pelvic inflammatory disease, infertility and ectopic pregnancy, and lead to neonatal blindness because of transmission to children (Little, 2006; Sandstrom 1987). Furthermore, untreated *N. gonorrhoeae* infection can also culminate in disseminated gonococcal infection, followed by other complications like infectious arthritis and endocarditis (Birrell *et al.,* 2019).

The main site that is colonized by *Neisseria gonorrhoeae* is the genital mucosa of both men and women. However, this bacterium can also colonize the anal mucosa and the oropharyngeal one, like the other well-known pathogenic species from the same genus, *Neisseria meningitidis*. Male and female urogenital tracts undergo different embryological developments that provide the cells lining the epithelial surface of this anatomical site in the two genders with a different arsenal of membrane molecules that can be used as receptors and coreceptors for invasive microorganisms like the gonococcus. Since the microenvironment differs, *N. gonorrhoeae* evolved different strategies to survive in the male or female urogenital niche (Edwards and Apicella, 2004). Furthermore, it is traditionally considered that female genital infections are mostly asymptomatic, on the contrary of male ones. However, this is probably biased by the fact that the result of immune cell influx and inflammation in men is more evident and easier to be noticed, because of the purulent exudate from the penis and painful urination. Such clinical manifestations lack in women, where the response does not occur in the same site of urination, therefore usually being not painful and more invisible (Quillin and Seifert, 2018). Actually, it seems that asymptomatic infections are common in both sexes (Xiong *et al.*, 2016). If this asymptomatic condition results from neutrophils recruitment in a number that is insufficient to generate observable symptoms or because in these situations neutrophils are not recruited at all at the site of infection is still not clear.

1.3.2 INFECTION MODELS AND MOLECULAR INSIGHTS ON THE ESTABLISHMENT OF *N. gonorrhoeae* INFECTION

Since *N. gonorrhoeae* is a human-restricted pathogen, it’s quite difficult to set up appropriate animal and tissue culture models to study its pathogenesis. This is further complicated if we think to the multifaceted nature of the tissues composing male and female genital tract, that make them difficult to be properly translated into a model. Such a model to be complete should in some way includes a variety of factors and conditions that the pathogen encounter inside the host, particularly referring to different types and local concentrations of nutrients, such as iron and zinc, to different oxygen concentrations and to the heterogeneity in immune and microbiota profiles that exist between different individuals. Human male volunteers have been challenged with gonococcus infection to understand different aspects of the *in vivo* pathogenesis of this bacterium (Cohen and Cannon, 1999), however conclusions from these studies cannot be translated to women, for which human experimentation is ethically limited by the severe complications that they can undergo. Primary cell cultures and organ systems have been developed to recapitulate men urethral epithelium (Harvey *et al.*, 1997) and also women upper and lower genital sites, however the already mentioned complexity of these tracts, particularly for females, where changes in relation to the menses cycle represent an extra layer of variability, cannot be appropriately reproduced in such models. So, probably they provided us with only partial insights on the gonococcal infection. Importantly, a female mouse model of genital infection continues to represent a reference model to study *N. gonorrhoeae* pathogenesis in an immunologically defined environment (Jerse, 1999). However, several limitations have to be taken in consideration also in this case, the main one being the lack in such model of human-specific receptors that have been highlighted as important ones for the gonococcal disease, such as CR3, CD46 and CEACAM. Finally, immortal and malignant cell lines are typically used. The clear advantages in their availability and manipulation often make them the model of choice, even if their protein expression is altered to some extent and, of course, they cannot be considered as completely representative of the tissues encountered *in vivo* by the pathogen. Three immortalized epithelial cell lines from vagina (VK2/E6E7), ectocervix (Ect1/E6E7) and endocervix (End1/E6E7), morphologically and immunocytochemically similar to their respective tissue of origin and primary culture, were developed by the expression of E6 and E7 genes from human papillomavirus type 16 (Rheinwald and Anderson, no date). End1 cell line demonstrated to be more active in cytokine expression compared to the other two (Fichorova and Anderson, 1999), and all the three cell lines upon 4h of infection by piliated or non-piliated gonococci up-regulated chemokines (IL-8), cytokines (IL-6) and intercellular adhesion molecule 1 (ICAM-1) independently from bacteria uptake and the secretion of the amplifier of inflammation IL-1, that was later increased only by End1 cells infected by the more invasive piliated gonococci (Fichorova *et al.*, 2001). Primary urethral epithelial cells were also immortalized using a retroviral vector expressing human papillomavirus E6 and E7 genes (Harvey, Post and Apicella, 2002). Phenotypically, these cells showed similarity with the primary culture and were described as keratinocytes. Furthermore, they expressed higher levels of IL-6 and IL-8 cytokines upon challenge with gonococci, like the primary cells. Finally, the pharyngeal Detroit 562 carcinoma cell monolayer model is used to study both *N. gonorrhoeae* and *N. meningitidis* adhesion and invasion of epithelia, particularly referring to the respiratory tract (Kibble *et al.*, 2019). The pharyngeal infection is the least considered one among the gonococcal infections in the literature, since it is mostly asymptomatic, however more efforts should be directed in studying it, because of its potential role as reservoir of infection in the population, particularly in men that have sex with men (Bernstein *et al.*, 2009), and as reservoir of antimicrobial-resistant gonococcal infection, as well (Weinstock and Workowski, 2009).

All these models increased our knowledge on the establishment of the infection by this pathogen (Figure 4). *N. gonorrhoeae* twitching motility, first attachment to and subsequent colonization of the mucosal surface are largely mediated by type IV pili (Higashi *et al.*, 2007), next retracted through the activity of the ATPase PilT (Aukema *et al.*, 2005). Microcolonies are formed by the bacteria on the epithelial cell surface, inducing a remodeling of the microvilli and modulating host cell signaling to aid the bacterial invasion. Indeed, it has also been demonstrated that pili activity and adherence play a role in host cell cytoskeletal rearrangements (Griffiss *et al.*, 1999; Grassmé, Ireland and van Putten, 1996a; Merz and So, 1997; Merz, Enns and So, 1999a). Notably, in some cells the human-specific membrane cofactor receptor CD46 has been individuated as a receptor for the gonococcal pilus (Kallstrom *et al.*, 2001). Other essential factors for the gonococcal colonization, strongly helping the bacterial adherence after the initial contact by pili, are represented by Opa proteins. They are mainly divided in two classes, Opa$_{50}$ that recognize host cell heparin sulfate proteoglycans (HSPG), and Opa$_{52}$ that recognize members of the carcinoembryonic antigen-related family of cell adhesion molecules (CEACAM or CD66) ('Redefined Nomenclature for Members of the Carcinoembryonic Antigen Family', 1999). Vitronectin (Gómez-Duarte *et al.*, 1997) and fibronectin (Van Putten, Duensing and Cole, 1998) act as a bridge between the bacteria and the target HSPG, and the association with an integrin coreceptor triggers a signaling cascade in the host cell dependent upon protein kinase C (PKC) activation (Dehio *et al.*, 1998), probably leading to modulations in the cytoskeleton that favor bacterial internalization. Also, the Opa-interacting lipooligosaccharide (LOS) contributes to adherence and invasion by the gonococcus. On the contrary of the majority of the other Gram-negative bacteria that present a lipopolysaccharide (LPS), the LOS of pathogenic Neisseriae lacks the repeating O-antigen sugar characterizing the polysaccharide side chain of LPS. The oligosaccharide side chains of the LOS present at their terminus epitopes that mimic sugar moieties of host glycosphingolipids, hence providing the bacterium with a strategy of immune avoidance and with the possibility to exploit host molecules normally associating with the mimicked structure (Mandrell, 1992; Yamasaki *et al.*, 1999). Furthermore, LOS sialylation can occur by gonococcus sialyltransferases that are present in the outer membrane (Shell *et al.*, 2002), conferring serum resistance but impairing Opa-mediated entry into non-ciliated cervical and urethral epithelial cells (van Putten, 1993). Therefore, LOS antigenic phase variation has been suggested as a mechanism that allows the bacterium to fluctuate between invasive and serum-resistant phenotypes (van Putten, 1993). It seems that LOS-mediated interaction with asialoglycoproteins receptors promotes the urethral invasion in men. In women, complement receptor 3 (CR3) is the receptor that mediates the invasion in the lower cervical tract, while lutropin-choriogonadotropic hormone receptor is the one involved in the endometrium and in the fallopian sites. Finally, also porins, water-filled channels that allow the passage of small molecules through the gonococcal outer membrane and that are among the most abundant gonococcal outer membrane proteins, have been implicated in potentiating multiple aspects of the disease, comprising the early phases of adhesion and invasion. It has been demonstrated that they can help the actin-mediated entry of the bacterium into the epithelial cells by acting as actin-nucleating proteins. Once that bacteria adhere and are internalized by the host cells, in order to efficiently replicate and maintain the infection, they need to meet their nutritional requirements competing with both the resident microbiota and the host for the available nutrients. Particularly, *N. gonorrhoeae* must acquire nutrients like iron, zinc and manganese, all factors that the host as a defense can sequester and limit in the so-called nutritional immunity process (Cassat and Skaar, 2013). Importantly, *Neisseria* spp. do not secrete siderophores, therefore they obtain iron directly from host-bound complexes through a series of membrane transport machineries that transport it into the bacterial cell (Hagen and Cornelissen, 2006).

1.3.3 TRANSCRIPTIONAL STUDIES ON *N. gonorrhoeae*

Due to problems and limitations described above concerning *N. gonorrhoeae* infection models, different gene expression studies focused on mimicking environmental conditions that the pathogen experiences in the host and in characterizing the transcriptional profiles in such conditions. Analyses on the genes encoding the molecular chaperones DnaK, DnaJ and GrpE following the bacterial exposure to heat stress suggested their transcription by σ^{32}

(RpoH)-dependent promoters in stress conditions (Laskos *et al.*, 2004). *rpoH*, together with other genes of the RpoH regulon, *groEL* and *groES*, was also found induced upon adherence to human epithelial cells in culture, and it was demonstrated to be important for the invasion of the host cells (Du, Lenz and Arvidson, 2005). The gonococcal transcriptomics response to anaerobiosis was also investigated, and results suggested that the anaerobic stimulon in gonococci is larger than previously considered and it is intertwined also to other aspects of the pathogen response to the various stresses encountered in the host environment (Isabella and Clark, 2011). Interestingly, Mercante et al. using microarray studies investigated the interconnected regulatory system that in gonococcus modulates the expression of the *mtrCDE*-encoded efflux pump, that confers this pathogen with resistance to hydrophobic antimicrobials, in response to iron availability. Indeed, an iron-dependent mechanism involving Fur is responsible for modulating MpeR transcriptional regulator that represses *mtrF*, an accessory protein of the aforementioned pump. In their work, they showed that in condition of iron limitation, the MpeR-mediated repression of MtrR, a direct repressor of *mtrCDE*, is enhanced, resulting in the increased expression of the efflux pump operon (Mercante *et al.*, 2012). Few regulatory small non-coding RNAs (sRNAs) have been characterized also for gonococcus, one of them, NrrF, being under the control of iron availability. Jackson et al. used a global approach to study NrrF effect on gonococcal transcription in response to iron and the emerging 12 genes were linked to energy metabolism, oxidative stress, antibiotic resistance, amino acid synthesis or uncharacterized functions (Jackson *et al.*, 2013). One year later, a transcriptomics study investigating specifically the sRNAome of *N. gonorrhoeae* cultured under different *in vitro* growth conditions appeared (McClure, Tjaden and Genco, 2014), and later the gonococcal Fur-controlled transcriptional program was deciphered as well (Yu *et al.*, 2016). Transcriptomics studies also reported gonococcus transcriptional changes linked to DNA methyltransferases (mod genes) phase variation (Srikhanta *et al.*, 2009). The whole transcriptome of *N. gonorrhoeae* strain MS11 was investigated by Remmele et al. (Remmele *et al.*, 2014). This work identified numerous new transcripts and suggested a considerable antisense transcription, that may have regulatory functions, occurring along the gonococcus genome, particularly for the phase variable *opa* genes. Furthermore, the authors performed a transposon insertion site sequencing (Tn-Seq) experiment to identify a first set of 827 essential genes for gonococcal survival. The majority of these genes was involved in fundamental biological processes, such as amino acid transport, DNA replication, translation and cell wall biosynthesis, however a good portion (135 genes) included also hypothetical proteins. Like for other bacteria, iron acquisition is essential for the ability of *N. gonorrhoeae* to successfully establish and continue the host infection. TonB, in complex with ExbB and ExbD, transduces energy generated at the cytoplasmic membrane to specific TonB-dependent transporter (TdTs) in the outer membrane, comprehensive of TbpA, LbpA and HpuB and four additional putative transporters encoded by *tdfF*, *tdfG*, *tdfH* and *tdfJ*, that directly interact with host iron-binding proteins (transferrin, lactoferrin and haemoglobin) to internalize iron. Hagen and Cornelissen demonstrated that TonB expression was necessary for gonococcal strain FA1090 survival within cervical epithelial cells, together with the expression of the single putative transporter *tdfF* (Hagen and Cornelissen, 2006). Interestingly, the latter was not expressed in gonococci grown in bacterial growth media, but it was detected when they were grown in cell culture media supplemented with fetal bovine serum or co-incubated with cervical epithelial cells. Importantly, also the gonococcus transcriptome from *in vivo* infections was characterized, both regarding female lower genital tract (McClure *et al.*, 2015) and urethra specimens from men (Nudel *et al.*, 2018). Large differences were found comparing gene expression during *in vivo* female infection with that one from bacteria grown *in vitro* in a chemically defined media and comparing infected men and women. Genes exclusively expressed in men were involved in host immune cell interactions, while in women they included phage-associated genes. Interestingly, a 4-fold higher expression of the Mtr efflux pump genes was observed in men (Nudel *et al.*, 2018). Finally, gonococcus transcriptomic data sets, including the aforementioned data about the natural infection of the human genital tract, were used to build the first global gene co-expression network for this pathogen, in order to infer central genes critical for gonococcal growth and virulence and to assign additional putative categories to different proteins (McClure *et al.*, 2020).

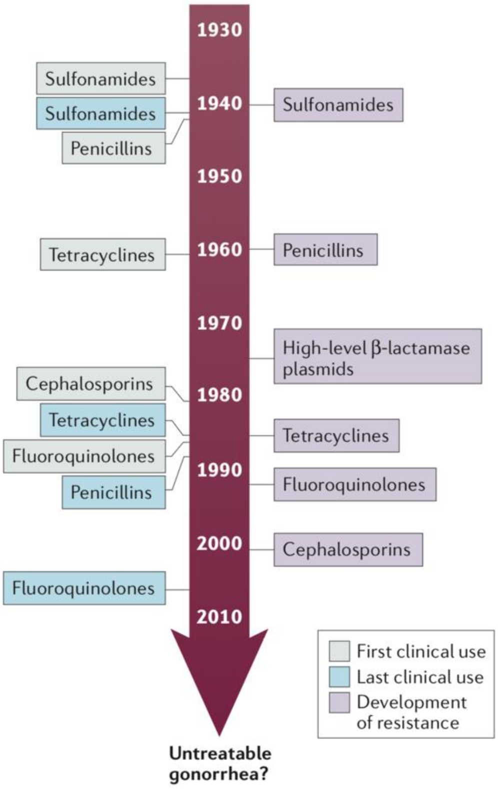
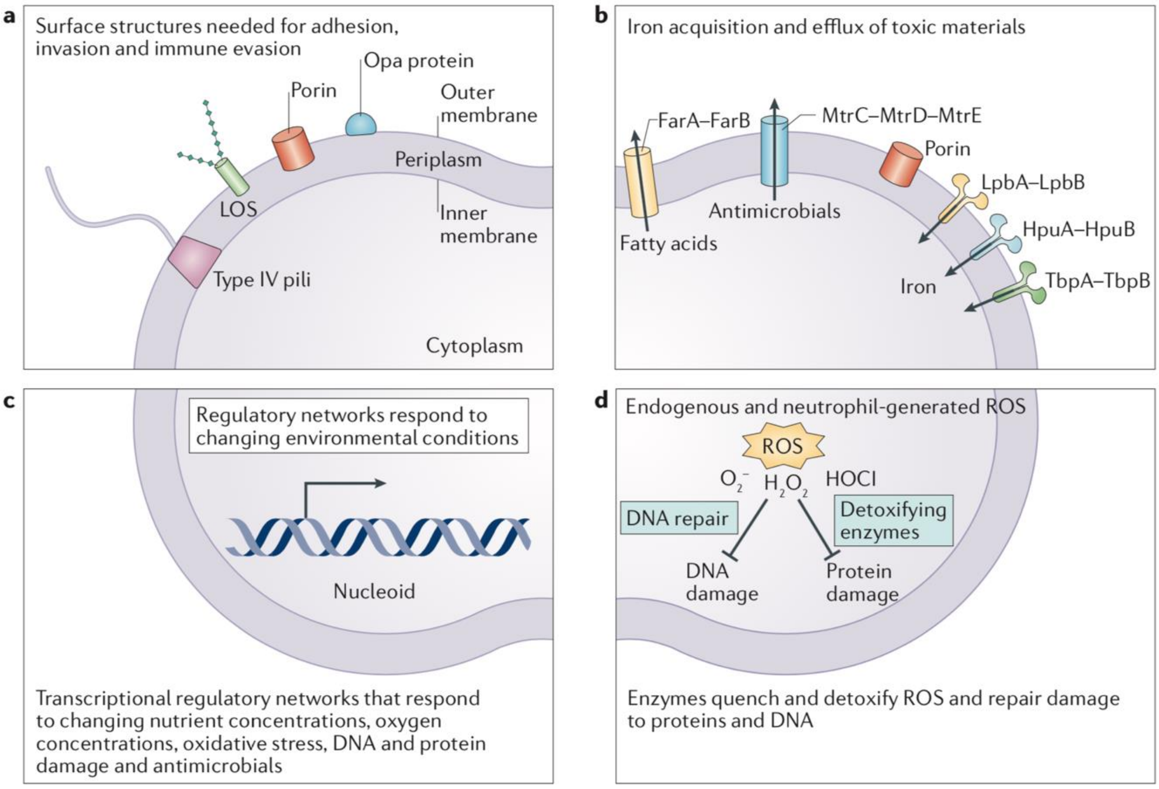


Figure 3 | Timeline of antibiotics introduction for gonorrhea treatment and the parallel resistance development by *N. gonorrhoeae*. The figure shows as each new class of antibiotics that was used as first-line strategy to treat gonorrhea soon was stopped because of the insurgence of resistant bacterial strains, until the last available treatment, the extended-spectrum cephalosporins, for which the bacterium recently gained resistance. As a consequence, WHO termed this pathogen "superbug", with the fear that, if new therapies or a vaccine will not be developed soon, we will face the rising of an untreatable antibiotics-resistant gonorrhea epidemics (Quillin and Seifert, 2018).



a
Surface structures needed for adhesion, invasion and immune evasion
Opa protein
Porin
Outer membrane
Periplasm
LOS
Inner membrane
Type IV pili
Cytoplasm
b
Iron acquisition and efflux of toxic materials
FarA–FarB
MtrC–MtrD–MtrE
Porin
LpbA–LpbB
Antimicrobials
Fatty acids
HpuA–HpuB
Iron
TbpA–TbpB
c
Regulatory networks respond to changing environmental conditions
Nucleoid
Transcriptional regulatory networks that respond to changing nutrient concentrations, oxygen concentrations, oxidative stress, DNA and protein damage and antimicrobials
d
Endogenous and neutrophil-generated ROS
ROS
O_2^- H_2O_2 HOCl
DNA repair
Detoxifying enzymes
DNA damage
Protein damage
Enzymes quench and detoxify ROS and repair damage to proteins and DNA

Figure 4 | Main *N. gonorrhoeae* pathogenesis factors. A) Surface factors to adhere and invade host cells and to evade the immune system. **B)** Efflux pumps that protect the pathogen from antimicrobials and fatty acids stress and membrane transporters for the nutrient uptake from the environment. **C)** Quite few transcriptional regulators, if compared to the *E. coli* arsenal, drive transcriptional responses to environmental stresses experienced during the infection. **D)** Protective enzymes are used by the pathogen to deal with bactericidal reactive oxygen species (ROS) generated by the host immune cells (Quillin and Seifert, 2018).

1.4 AIM OF THE WORK

The recent outbreak of Covid-19 pandemics reminded us how much microbial epidemics can impact our lives.

A major public health concern is globally rising about *Neisseria gonorrhoeae*, the bacterium causing the major sexually transmitted infection known as gonorrhea, since this pathogen is evolving high levels of resistance to antibiotics, the only available medical instruments that we have today to counteract the infection in absence of an effective vaccine. In the past nine years since its first appearance due to the increase in sensitivity of RNA sequencing, dual RNA-Seq showed a fast-growing trend in infection biology studies having the purpose to better characterize the molecular dynamics that accompany bacterial infections of eukaryotic cells. Furthermore, few attempts in literature exist to develop exploratory approaches able to fully exploit the potential that is behind this kind of data and their ability to simultaneously capture the rewiring in both host and pathogen transcriptome during the infection process. Therefore, the main aims of our work were:

i) Characterize the pro-inflammatory responsiveness of three described host cell models to gonococcal infection;

ii) Simultaneously profile *N. gonorrhoeae* WHO M transcriptional program during adhesion, invasion and early persistence in the host cells;

iii) Develop a new strategy to analyze short time series dual RNA-Seq data in order to extract host and pathogen key regulations sustaining the interaction observed between the two organisms.

In order to answer these questions, we conducted a first dual RNA-Seq comparative study of early *N. gonorrhoeae* infection of three described *in vitro* models, with the final goal to progress our current understanding of the host-gonococcus interaction.

CHAPTER 2: MATERIALS AND METHODS

2.1 BACTERIAL STRAIN AND GROWTH CONDITIONS

The bacterial strain used for the infection models was Neisseria gonorrhoeae WHO M (PorB 1 b serotype), belonging to both 2008 and 2016 published WHO panel of *Neisseria gonorrhoeae* reference strains, internationally validated as representative for *Neisseria* *gonorrhoeae* species (Unemo *et al.*, 2016). Bacteria were grown on gonococcus medium (GC) agar plates (Difco) or in liquid GC broth supplemented with 1% isovitalex (BBL) at 37°C and 5% CO_2.

2.2 CELL CULTURES

Detroit 562 (oropharyngeal) cell line (ATCC CCL-138) was maintained in Eagle's Minimum Essential Medium (EMEM) supplemented with fetal bovine serum to a final concentration of 10% and antibiotics at 37°C and 5% CO_2.

End1/E6E7 (endocervical) cell line (ATCC CRL-2615) was maintained in keratinocyte serum-free medium (KSFM, Gibco) supplemented with 50 μg/ml bovine pituitary extract (BPE), 0.1 ng/ml epidermal growth factor (EGF), 0.4 mM $CaCl_2$ and antibiotics at 37°C and 5% CO_2.

t-UEC/E6E7 (urethral) cell line (kindly provided by M. Apicella, Department of Microbiology, University of Iowa) was maintained in Prostate Epithelial Cell Growth Medium (PrEGM, Clonetics) supplemented with growth factors included in PrEGM BulletKit and antibiotics at 37°C and 5% CO_2. See also Supplementary Table 5.

2.3 CONFOCAL MICROSCOPY

Detroit562, End1 and t-UEC cells were cultured on HTS Transwell ® 24-well pearmeable supports with 0,4 μm pore polyester membrane (Corning, CLS3397-12EA) for 7 days. In order to reach confluence at the same time, cells were seeded on collagen-coated transwell insert (Collagen I, Rat Tail, Coring 354236) at different densities: 1.2*10^5cells/cm^2 for Detroit562, 1*10^5 cells/cm^2 for End1, 9*10^5 cells/cm^2 for t-UEC. The medium was replaced every 2 days. The day before the experiment, cells were incubated in antibiotic-free medium. After polarization, cells were infected by fluorescent bacteria (Oregon Green, 1:200 in PBS, 15 minutes at 37°C). Samples were fixed at two different time points: 15 minutes post-infection (mpi) and 120 mpi. In order to monitor persistence of bacteria during time, after 15 minutes of infection all the samples were washed to remove unbound bacteria. Cells were fixed with 4% (v/v) formaldehyde for 10 minutes, then blocked with D-PBS containing 1% (w/v) BSA for 30 minutes and incubated with anti-*N. gonorrhoeae* antibodies (1:100) for 60 minutes at room temperature. Next, wells were washed with PBS and incubated with goat anti-mouse Alexa Fluor 568 conjugated antibodies (ThermoFisher, A-11004). After washing in PBS to remove the excess of dye, samples were permeabilize using Triton 100 0,25% in PBS-BSA 1% for 30 minutes and stained with CellMask Green Actin Tracking Stain (TermoFisher, A57243) and DAPI (Invitrogen, D1306) for 15 minutes at room temperature. Samples were analyzed with Zeiss LSM710 confocal microscope at magnification 100X (Oil Immersion Objective).

2.4 ADHESION/INVASION ASSAY

The assay was carried out on glass-bottomed 96 well plates (Eppendorf cell imaging plate) to allow the analysis by confocal microscope and Opera Phenix instruments. A specific number of cells for cell line (Detroit562: $1*10^6$ cells/well; End1: $9*10^5$ cells/well; t-UEC: $8*10^5$ cells/well) was seed the day before to have a complete monolayer of cells the following day. Starting from agar plate, bacteria were grown in GC isovitalex 1% medium for 60 and 45 minutes until they reach the exponential phase, OD_{600} 0.5. Next, bacteria were fluorophored with Oregon Green 488 (1:200 in PBS) for 15 minutes at 37°C. In order to remove the excess of dye, samples were washed with D-PBS and then bacteria were diluted in cell medium w/o P/S to reach the OD of interest (final OD_{600} 0.1). Cells were infected for 15 minutes and the unbound bacteria were removed with a D-PBS wash. For the first time point, some wells were fixed using PFA, while the others were filled with fresh medium to monitor the infection over time, fixing at 30 minutes, 60 minutes and 120 minutes post-infection (mpi). The use of a secondary antibody, interacting only with adherent bacteria on the cell surface, allowed us to assess the percentage of adherent and intracellular bacteria. Infected samples were analyzed with a high content screening (HCS) fluorescence microscope platform, Opera Phenix. Images were acquired with a 40X water immersion objective in confocal mode performing a z-stack image acquisition. Total number of bacteria was defined through the find image region building block (channel 488). Total bacterial fluorescence area was calculated as a value of bacterial adhesion (a). Another building block based on channel 568 was created on the output to define the percentage of extracellular bacteria (b). Finally, the percentage of intracellular bacteria were obtained by (a-b)*100/a.

2.5 FLOW CYTOMETRY

Detroit562 ($5*10^5$ cells/well), End1 ($4*10^5$ cells/well) and t-UEC ($2.5*10^5$ cells/well) cells were cultured in a 24 well-plate. Cells reached confluence in 3-4 days. For each time point

(15 mpi and 120 mpi), cells were detached from the bottom using Trypsin-EDTA. Upon pelleting (7 min at 1800 g, RT), the vitality of cells upon infection was defined using Live/Dead Fixable Aqua Dead cells stain kit (L34957) (1:500 in PBS, 20 min in the dark at RT). After fixing with PFA 4%, samples were resuspended in FACS buffer (1% BSA in PBS) and analyzed by flow cytometry using the Becton Dickinson FACSCantoII and the software FACSDiva 8.0.1. Upon selecting for intact and viable cells by gating based on cell diameter (forward-scatter), granularity (side-scatter), and vitality (linear scale), infected (GFP-positive) and non-infected (GFP-negative) sub-fractions were further discriminated based on GFP signal intensity (FITC channel) vs granularity (side scatter) (logarithmic scale). To discriminate between intracellular and extracellular bacteria, quenching properties of Trypan Blue 0.4% (Gibco, 15250061) were exploited to reduce the fluorescence of the extracellular ones.

2.6 FLUORESCENCE-ACTIVATED CELL SORTING (FACS)

RNA Protect Bacteria Reagent (Qiagen, 76506) was used as fixative for FACS sorting. The protocol described by Westermann A. has been optimized for our work (Westermann, Barquist and Vogel, 2017). Starting from a 6 wells/plate, cells were infected as described in the section about the adhesion/invasion assay, detached from the bottom of the well using Trypsin/EDTA (Gibco, 25200056) and stained with Live/Dead Fixable Aqua Dead cell stain kit (Invitrogen, L34957). Next, samples were pelleted (7 minutes at 1800 g, RT) and resuspended in PBS-BSA 2% + 10% RNAlater (1 ml for $6*10^6$ cells). To avoid the formation of aggregates, samples were fixed under shaking 400 RPM at 4°C for at least 60 minutes. The cell suspension was first passed through MACS Pre-Separation Filters (30 μm exclusion size; Miltenyi Biotec) and then sorted using FACSAria II device (Becton Dickinson) at 4°C (cooling both the input tube holder and the collection tube rack) and at low flow rate (~2), using the following gating strategy. Upon selecting for intact and viable cells by gating based on cell diameter (forward-scatter), granularity (side-scatter), and vitality (linear scale), infected (GFP-positive) and non-infected (GFP-negative) sub-fractions were further discriminated based on GFP signal intensity (FITC channel) vs granularity (side scatter) (logarithmic scale) (Supplementary figure 1). To avoid RNA degradation, the instrument was decontaminated through two washes with HCL 10% and H_2O DNAse/RNAse-free (10 minutes each). Samples were isolated directly in lysis buffer. Typically, ~$2.5*10^5$ cells were collected in each tube to maintain constant the ratio between sample and Tri-reagent (Zymo).

The RNA was isolated starting from at least $8*10^5$ positive events using the Direct-zol RNA miniprep kit (Zymo, R2073) and following protocol instructions.

2.7 ISOLATION OF TOTAL RNA

Total RNA was isolated lysing the cells with an appropriate volume of TRI-Reagent and purifying samples by Direct-zol RNA miniprep kit (Zymo, R2073). RNA was treated with the Turbo DNA-free DNAse kit (Invitrogen, AM1907) according to the manufacturer's protocol to complete DNA removal. Purity of extracted RNA was evaluated by nanodrop considering the ratio of absorbance reading at 260 and 280 nm (A260:A280 > 1.8). Integrity of RNA was determined by Agilent 2100 Bioanalyzer and RNA 6000 Nano LabChip (RIN > 7).

2.8 qRT-PCR

To estimate bacterial RNA concentrations in mixed samples a Real Time Spike-in was performed. Bacterial RNA concentration was defined amplifying the 16S gene (F: GTCATTAGTTGCCATCATTCGG; R: GGAAGGTTCAGGTTGTTTTCTG) starting from 100 ng of DNase I-treated total RNA and comparing the obtained CT values to a bacterial RNA standard curve composed by fixed concentrations of eukaryotic RNA and defined dilution of bacterial RNA. The analysis was performed using the SuperScrip II Platinum SYBR Green One-Step qRT-PCR Kit with ROX (Invitrogen, 11746-100) and the Applied Biosystems 7500 Real-Time PCR System.

2.9 cDNA LIBRARY PREPARATION AND ILLUMINA SEQUENCING

cDNA libraries for Illumina sequencing were generated starting from at least 100 ng of total RNA sample using the Ovation Universal RNA-Seq System Kit by NuGEN. The double-stranded cDNA was produced using a mixture of random and poly(T) priming and then fragmented by Covaris Sonication System S2 (10dc, 5i, 200cpb, 90'') to generate on average ~200-400 bp fragmentation products. Fragmented cDNA was concentrated using Agencourt RNAClean XP Beads (Beckman Coulter, Cat. #A63987). Adaptors used were unique single indexes enabling multiplexing. Ribo-depletion was used instead of polyA selection to have better coverage of whole transcriptomic RNA and because of the differences in both function and amount existing between the two polyA fractions from the two organisms. Samples were ribo-depleted from eukaryotic and bacterial rRNA by the AnyDeplete technology. Resulting libraries were PCR-amplified, purified using the Agencourt RNAClean XP Bead and analyzed by capillary electrophoresis Agilent High Sensitivity DNA Kit (Agilent, Cat. #5067-4626). Finally, they were run on Illumina NovaSeq 6000 sequencing platform for 101 cycles in paired-end mode.

2.10 READ MAPPING AND DIFFERENTIAL EXPRESSION ANALYSES

Raw reads were first quality-checked using FastQC (version 0.11.3), with all samples showing a peak in the distribution of the mean sequence quality (Phred score) over all sequences around 36. In order to discriminate between host and bacterial reads, and to limit cross-mapping phenomena, before any mapping procedure to the bacterial genome all samples were aligned to the human genome (Baddal *et al.*, 2015). Reads were mapped in a paired-end mode to the human genome (GRCh38) with the splice-aware aligner STAR (version 2.7.0f). Next, previously unmapped reads were aligned to *Neisseria gonorroheae* WHO M assembled genome using Bowtie2 (version 2.2.6) (Marsh *et al.*, 2017). Reads were also mapped to Bowtie2 indexes created using host and bacterial rRNA sequences in order to evaluate the extent of ribodepletion efficacy. Finally, mapped reads were also visualized with IGV (version 2.6.0). Between 0.3% and 2.2% of the total reads from infected samples mapped to the bacterial genome. Unassigned reads, i.e. reads that didn’t map to either genome, were discarded (around 5% of total reads from each sample). Next, gene counts were obtained using featureCounts function by Rsubread Bioconductor package (R version 3.4.4, Rsubread version 1.28.1) and the Ensembl annotation for the host or the NCBI one for the pathogen, excluding multi-mapping reads. Differential expression analysis was performed with DESeq2 Bioconductor package (R version 3.6.0, DESeq2 version 1.24.0) (Love, Huber and Anders, 2014). Before to build host and pathogen generalized linear models, the remaining counts associated to human or bacterial rRNA genes were removed from the count matrices. Both an examination of gene dispersion values in relation to the mean of normalized counts and a principal component analysis (PCA) on transformed and normalized read counts were conducted to better guide decisions on the model design. The transformation before PCA was applied to make counts data approximately homoskedastic, so to have constant variance along the range of mean values. For the host, to control for the introduction of unwanted technical variation (as evidenced by PCA not separating efficiently many samples based on the different treatments inside the three cell groups and histograms of p-values upon model building in many cases being not as expected, with an unusual trend) and because of the huge diversity in the within-group variability between the different cell lines, differentially expressed genes were called upon building separate generalized linear models, comparing the data for the different infected samples (considered as independent conditions) to the data for the uninfected controls with a design of the type ~ W_1 + W_2…+ W_n + condition (where W_1 + W_2…+ W_n represents factors of unwanted variations) and considering the following threshold: Benjamini-Hochberg (BH)-adjusted P value ≤ 0.01. After filtering genes that didn't have more than 5 reads in at least 3 samples and after checking for the need of between-sample normalization inspecting boxplots of relative log expression (log-ratio of read count to median read count across samples) and PCA plot, an adequate number of factors of unwanted variation for each cell line was estimated using all genes from control samples (for which the covariates of interest are constant) through RUVs function by RUVSeq Bioconductor package (R version 3.6.0, RUVSeq version 1.18.0) (Supplementary Figure 2) (Risso *et al.*, 2014). Three factors of unwanted variations were introduced in the design for Detroit562 and End1 cell lines, four in the case of tUEC one. For the pathogen, in order to compare the transcriptome data for the different samples of infecting bacteria (considered as in independent conditions) to the transcriptome of bacteria adapted to the different host cell media (30 min of adaptation), differentially expressed genes were called upon building a generalized linear model with a design of the type ~ cell + condition + cell:condition and considering the following threshold: BH-adjusted P value ≤ 0.05. Remarkably, this kind of design also allowed us to look for genes showing a different condition effect upon infection of the different cell lines, even if the majority of these genes probably derived from the diversity in the initial, pre-infection adaptation to the three different cell media. Supplementary Figure 3 reports the histograms of p-values obtained from DE tests on host and pathogen models. Results were visualized in different ways. Heatmaps depicted mean-centered variance-stabilized (see above) read counts (normalized with respect to library size), with colours from yellow to red representing up-regulated genes and colours from light blue to blue representing down-regulated ones. Before any other visualization, like for volcano plots, the log2 fold changes were shrunken in order to correct for poorly reliable large estimates related to genes having low or highly variable read counts, using apeglm method (Zhu, Ibrahim and Love, 2019) as far as it was possible, or the ashr one (Stephens, 2016).

2.11 ENRICHMENT ANALYSES

Enrichment analyses on the host genes were conducted performing both a hypergeometric test on gene ontology (GO)-terms via goana function or KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways via kegga function by limma Bioconductor package (R version 3.6.0, limma version 3.40.6) and a Gene Set Enrichment Analysis (GSEA) for the gene sets of GO biological process by fgsea Bioconductor package (R version 3.6.0, fgsea version 1.10.1) (Subramanian *et al.*, 2005) (Sergushichev, 2016). In both cases, a significant enrichment was called with a BH-adjusted P value ≤ 0.05. Furthermore, a possible enrichment in TF regulations was also evaluated by GSEA for TF target prediction gene sets from the GTRD database (version 19.10) or for a combination of those and older gene sets (BH-adjusted P value ≤ 0.05) (Yevshin *et al.*, 2019). Transcription factor binding sites (TFBSs) significantly enriched inside the promoters of all DE genes emerging at 15 mpi compared to uninfected controls for Detroit562 cells were found by Simple Enrichment Analysis of motif (SEA) included in the MEME Suite of Sequence Analysis tools, considering "Vertebrates (*in vivo* and *in silico*)" database of motifs and control sequences obtained by shuffling the primary ones conserving 3-mer frequencies (Bailey and Grant, 2021). The promoters were defined considering the 1000 bp 5'-flanking sequences respect to the first exon of the considered genes retrieved from the Ensembl website. For the pathogen, WHO M genes were first annotated by homology to KEGG database through blastKOALA search and mapped into KEGG metabolic pathways. Next, a hypergeometric test to evaluate enrichment in KEGG pathways was performed, considering as significant the terms associated to a BH-adjusted P value ≤ 0.05.

2.12 HUMAN TF-TARGET GENES REGULATORY NETWORK RECONSTRUCTION

A whole human TF-target genes regulatory network was built using TRRUST version 2 database (Han *et al.*, 2018) by igraph package (R version 3.6.0, igraph version 1.2.6). From the previous GSEA analysis, we defined the leading-edge subsets to be those genes in the gene set that appeared in the ranked list at, or before, the point where the running sum reaches its maximum deviation from zero, so gene set cores accounting for the observed enrichment signal (Subramanian *et al.*, 2005). The genes inside the leading edge for the top 20 significantly enriched GO terms with positive normalized enrichment score (NES) by GSEA analysis shared by all the three cell lines upon 120 minutes of infection without sorting (inflammatory response, response to bacterium, positive regulation of MAPK cascade, cellular response to lipid, taxis and response to toxic substances) were mapped into the resulting network. Only genes showing regulations with each other were maintained (so, to have only genes that were inside the leading edges) and the three regulatory subnetworks were visualized through Cytoscape (version 3.7.2).

2.13 *Neisseria gonorrhoeae* WHO M CO-EXPRESSION NETWORK ANALYSIS

A *Neisseria gonorrhoeae* WHO M co-expression network was built using all samples from our data set (in total, 36). First, rRNA genes and genes whose counts were consistently low (≤ 5 in at least 25 samples, representing ~70% of all samples) were filtered out. Next, gene counts were normalized for library size and variance-stabilized using vst function by DESeq2 Bioconductor package (R version 3.6.0, DESeq2 version 1.24.0) (Love, Huber and Anders, 2014). Finally, after checking for correlation in sample quantile scatterplots, quantile normalization was performed to remove systematic shifts between samples using normalize.quantile function by preprocessCore package (R version 3.6.0, preprocessCore version 1.46.0). The input matrix was used to infer nine co-expression networks through context likelihood of relatedness varying the Z-score of mutual information used to define an edge by minet package (R version 3.6.0, minet version 3.42.0), following the strategy adopted by McClure et al. (McClure *et al.*, 2020). The nine networks were visualized through Cytoscape (version 3.7.2) and a network with 1315 nodes and 1500 edges (Z-score cutoff 8.418426) was selected because of its structure for the following analyses. The degree distribution was calculated and hub genes detected imposing node degree ≥ 6, using respectively degree_distribution and degree functions by igraph package (R version 3.6.0, igraph version 1.2.6). Communities, or modules, of co-expressed genes in the network were found imposing as minimum number of genes to define them 12, through the fast greedy approach using fastgreedy.community function by igraph package (R version 3.6.0, igraph version 1.2.6). Modules of co-expressed genes significantly enriched in DE genes from previous comparisons were tested with an hypergeometric test using phyper function by stats package, then the Benjamini-Hochberg-adjusted p-values were calculated through p.adjust function by stats package (R version 3.6.0, stats version 3.6.0). Hypothetical proteins being among the top 30 statistically significant DE genes in any of the analyzed contrasts were inspected for their belonging to the detected co-expressed modules.

2.14 GENE CLUSTERING AND HOST-PATHOGEN INTERACTIONS TESTING

Host and pathogen genes were grouped on the basis of their temporal expression pattern following the strategy published by Ernst et al. to deal with short time series gene expression data (Ernst, Nau and Bar-Joseph, 2005b). Next, as original contribution of this thesis, permutation tests were also used to test for influences occurring between host and pathogen significant profiles, then described and studied through weighted directed bipartite graphs. Such strategy was first tested on a publicly available dual RNA-Seq data set concerning, similarly to our study, an early infection of human type II lung epithelial cell line A549 by both *Streptococcus pneumoniae* D39 wild-type bacteria and mutant, unencapsulated ones (Aprianto *et al.*, 2016). The table with raw counts was downloaded from Gene Expression Omnibus (GEO) (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE79595). First, a set of representative temporal patterns or model profiles was defined through a pre-defined number of selections that from the total $(2c + 1)^{n-1}$ possible patterns, where *c* represents the amount of change a gene can exhibit between successive time points and *n* the number of time points, each time chooses the one that is farthest from all those selected so far (maximum-minimum approach). The total number of investigated time points in the aforementioned study was 5, so *n* was equal to 5, while *c* was set to 1. A set of 44 representative model profiles was selected from the 81 possible ones. Reads assigned to rRNAs were removed from host and pathogen raw counts. Next, counts were normalized (median of ratio normalization) and variance-stabilized using rlog function by DESeq2 Bioconductor package (R version 3.6.0, DESeq2 version 1.24.0) (Love, Huber and Anders, 2014). With data already in log space, the difference between each time point and the control (time 0), that was set as 0, was calculated, in order to represent data as time point fold changes compared to the pre-infection experiment. Genes that didn’t show consistency among the two replicates, i.e. being characterized by a value of Pearson correlation between the two time course lower than 0, were filtered out. Since just two replicates were analyzed in the study, gene time series were described by the mean of the values along the two. In order to select genes showing a modulation along the infection, genes showing as largest expression changes between two time points (not necessarily consecutive) a value lower than 0.8, were filtered out. The remaining genes were assigned to the most similar profile among the representative ones by Pearson correlation (cor function, R version 3.6.0, stats version 3.6.0). At the end, genes assigned to each profile were counted. If a gene was assigned to more than one profile, it was counted in each of these profiles as 1/*n*, where *n* is the number of profiles the gene was assigned to. Next, a permutation test, computing all the possible permutations, was performed to remove dependency between the time points and quantify the number of genes that would have been assigned to each of the model profiles if data were generated at random. At the end, assuming that *X*, the number of genes assigned to profile *i*, is distributed like a binomial random variable

$$X \sim Bin\left(|T|, \frac{E_i}{|T|}\right) \tag{10}$$

with $S_i = \sum_j s_i^j$ and $E_i = S_i/(n!)$,

where s_i^j indicates the number of genes assigned to *i* in permutation *j*, one of the *n!* possible permutations, and T the total number of genes, for *t(m$_i$)* genes assigned to model profile *i* the significance is called if

$$P\big(X \geq t(mi)\big) < a/m \tag{11}$$

where α = 0.05 and *m* = number of representative patterns or model profiles (Bonferroni adjustement).

Once statistically significant, non-casual temporal expression patterns were defined for both the host and the pathogen, a second permutation test similar to the one described above was used to test for transcriptional influences occurring between them during the infection process. One by one, each of the pathogen significant profiles was fixed, host significant temporal profiles and assigned genes were described by pointwise differences between them and the fixed bacterial one, genes were assigned to the most similar profile of difference by Pearson correlation, and then again all the permutations were computed as before and the same procedure just described was repeated to test for significance as reported for the previous permutation test, and viceversa fixing each of the host significant profiles. A pseudocount of 0.001 was added at each time point excluding the first, pre-infection value to deal with situations in which the compared host and pathogen profiles were exactly the same. At the end, significant interactions between a fixed pathogen (host) profile and host (pathogen) ones were represented with an edge going from the first to the seconds in a weighted directed bipartite graph, where the two disjoint sets of nodes represented host and pathogen significant profiles and the links the influences occurring between them, weighted on the basis of the Bonferroni-adjusted p-value from the permutation test. To retrieve relevant information from the network, centrality measures and node removal tests were used. In particular, betweenness centrality, calculated through betweenness function by igraph package (R version 3.6.0, igraph version 1.2.6) and normalized by the formula in (9) (Section 1.2.3), and derivatives were used. Betweenness-related measures used in the analysis were retrieved from the work of Scardoni et al. (Scardoni *et al.*, 2014). In particular, the betweenness interference of a node *i* with respect to another node *n* in the network *G* is described by

$$Int_{Btw}(i,n,G) = relBtw(G,n) - relBtw(G_{|i},n) \qquad (12)$$

where $G_{|i}$ is the network obtained from *G* removing *i*. The global interference of a node is given by the sum of all the absolute interference values of that node with respect to all the other nodes, and it is a measure describing how much the removal of a node from a network affects globally the structure of that network. The betweenness robustness of node n is obtained from all the interference values from all the other nodes respect to node n and it is described by

$$Rob_{Btw}(n,G) = 1/max_{i \in N_{|n}}\{|Int_{Btw}(i,n,G)|\} \qquad (13)$$

indicating how much a node is resistant to network modifications. From this definition, the dependence value of a node n is described as

$$Dep_{Btw}(n,G) = max_{i \in N_{|n}}\{Int_{Btw}(i,n,G)\} \qquad (14)$$

where $\mathrm{Int}_{\mathrm{Btw}}(i,n,G) \geq 0$, and the competition value of a node n as

$$Comp_{Btw}(n,G) = max_{i \in N_{|n}}\{|Int_{Btw}(i,n,G)|\} \qquad (15)$$

where $\mathrm{Int}_{\mathrm{Btw}}(i,n,G) \leq 0$. If dependence is high, this means that the node is central because of the presence of at least another node and once this node is removed from the network the betwenness centrality of the considered node strongly decreases. On the contrary, a high competition value means that the betweenness centrality of the considered node increases consistently upon the removal of a particular node from the network.

Enrichment analyses on the host significant profiles were performed as described in 2.11, while the ones on pathogen significant profiles were performed through an hypergeometric test on KEGG metabolic pathways using kegga function by limma Bioconductor package (R version 3.6.0, limma version 3.40.6). The described analysis was performed both on the wild-type infection data set and the mutant one. The intersections between genes in pathogen significant profiles emerging from the analysis on the wild-type data and those from the analysis on the mutant one were tested with an hypergeometric test using phyper function by stats package, then the Bonferroni-adjusted p-values were calculated through p.adjust function by stats package (R version 3.6.0, stats version 3.6.0). Pathogen projections in the two conditions were computed through bipartite_projection function by igraph package (R version 3.6.0, igraph version 1.2.6).

The analysis in case of the gonococcus data set, for all the three infections, was conducted in the same way, considering as 120 mpi samples the ones that were subjected to sorting. *n*, number of time points, was equal to 3 and *c*, amount of change a gene can exhibit between successive time points, was set to 3. This led to a total of 48 possible temporal profiles, excluding the one being characterized by all 0, therefore a selection procedure was not applied. Since the number of replicates was 3 in this case, genes showing a Pearson mean value, obtained from the Pearson correlation coefficients calculated between all the possible couples of time series replicates, lower than 0 were filtered out, and gene time series were described by the median of the values along the replicates.

CHAPTER 3: RESULTS

3.1 PHENOTYPICAL CHARACTERIZATION OF INFECTION MODELS

3.1.1 *N. gonorrhoeae* WHO M SHOWS PHENOTYPICALLY DISTINCT TEMPORAL DYNAMICS OF ADHESION AND INVASION OF THE THREE CELL LINES

We first aimed to characterize phenotypically the dynamics of infection that *N. gonorrhoeae* WHO M shows in relation to Detroit 562, End1 and t-UEC cell lines. Confocal microscopy reported considerable adhesion to urethral and endocervical cells upon 15 minutes of infection, while a lower presence of bacterial cells was observed for oropharyngeal ones (Figure 5A). Consistently with reports of gonococcus intracellular phases of infection, at the later stage of 120 minutes post infection (mpi), bacteria were detected in large amounts clustered inside Detroit562 cells, in a quite detectable amount they were observed also inside End1 cell line, even if in this case the infection appeared to be more widespread with many extracellular bacterial cells, while only few of them seemed to be internalized by urethral cells (Figure 5A). These results were confirmed by Opera Phenix analysis on six time points of infection (15, 30, 60, 120, 240, 360 mpi), showing less bacterial area (μm^2) at 15 mpi considering the oropharyngeal infection compared to the cervical and especially the urethral one (Figure 5B). As expected, the bacterial area slightly increased over the time course for all the three cell models, with the proportion of internalized bacteria at 120 mpi being higher for oropharyngeal cells compared to End1 and t-UEC cell lines (Figure 5B). Finally, FACS analysis confirmed a significantly higher number of infected cells at 120 mpi compared to

15 mpi for all the three cell models and a more spread infection over the entire epithelium in case of End1 cells compared to the other two cell lines (Figure 5C).

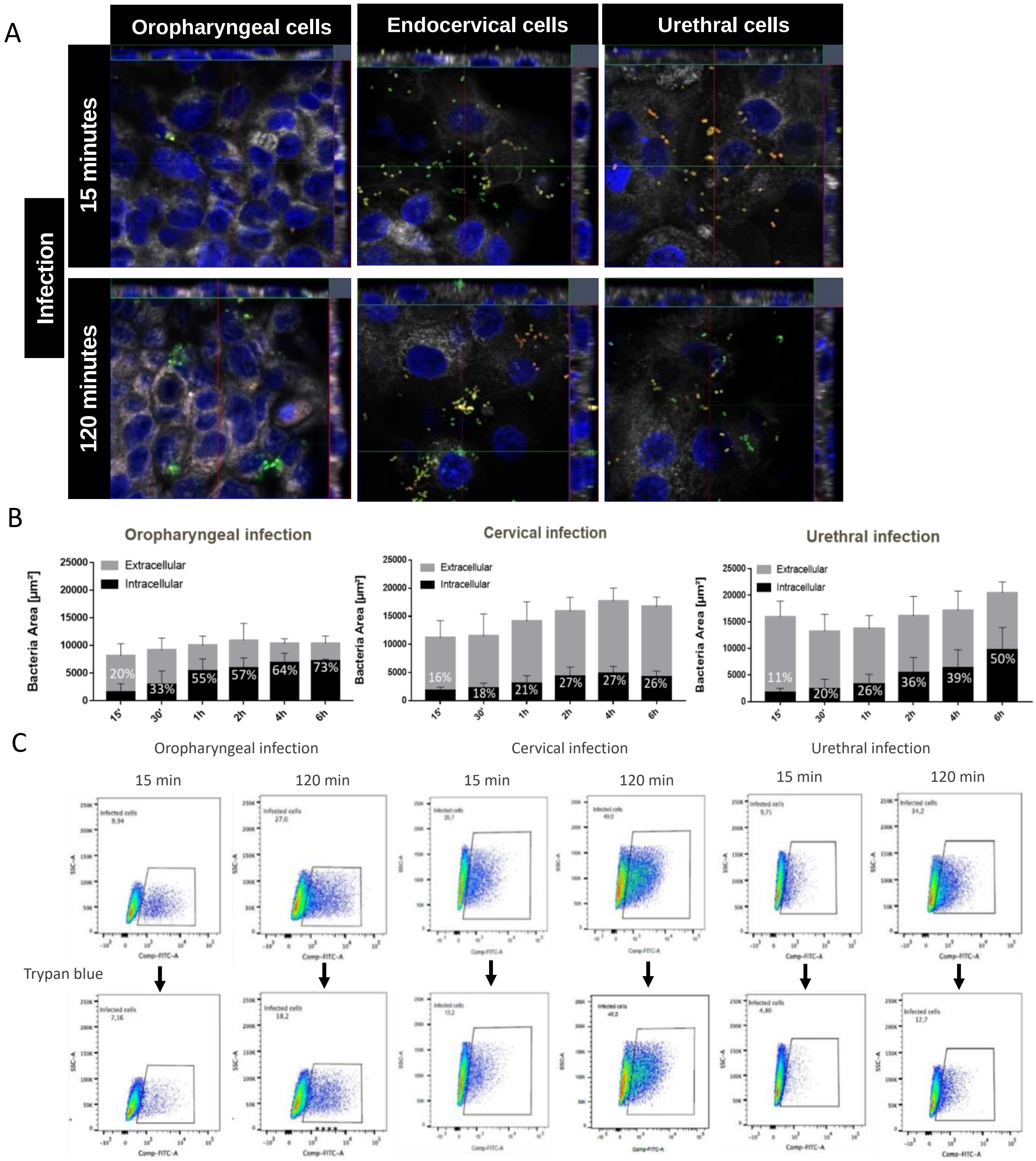


Figure 5 | Different temporal dynamics of *N. gonorrhoeae* WHO M infection of the three cell models. A) Confocal microscopy of Detroit562 (oropharyngeal), End1 (endocervical) and t-UEC (urethral) cells infected by Oregon green-conjugated *N. gonorrhoeae* WHO M bacteria for 15 minutes and 120 minutes. Double staining is used to discriminate between intracellular (green) and extracellular (green + red) bacterial cells. Host nuclei (Dapi, blue) and cytoskeleton (CellMask, gray) are visible. Images are 3D view of Z-stack acquisition. **B)** Total bacterial fluorescence area calculated through Opera Phenix analysis (Harmony software) to assess the number of infecting (extracellular and intracellular) pathogens from 15min up to 6h in the three cell models. Values inside the bars represent the percentage of intracellular bacteria compared to the total number of bacteria. Data are represented as mean and standard error of the mean of at least four biological replicates. **C)** Dot plot graphs showing the percentages of cells infected by gonococcus (as GFP positive events) at 15mpi and 120mpi in the three cell models by flow cytometry. Treatment with Trypan blue is used to define percentages of cells infected by intracellular bacteria (lower panel).

3.2 PRE-PROCESSING OF DUAL RNA-SEQ DATA

3.2.1 THE DUAL RNA-SEQ EXPERIMENT GENERATED A HIGH-QUALITY DATA SET AND ACCEPTABLE RELATIVE MAPPING PERCENTAGES

Starting from the aforementioned results, we wanted to characterize the global transcriptional changes driving the host-pathogen crosstalk during *N. gonorrhoeae* WHO M early interaction with the three cell models. Therefore, we set up the first dual RNA-Seq experiment concerning gonococcal infection. Consistently with the previous observations, we choose to investigate a really early time point, 15 mpi, to detect those changes guiding the first encounter between the host and the pathogen, and a later one, 120 mpi, characterized by a significantly higher bacterial intracellular lifestyle, to study pathogen adaptation to the host environment and, on the other side, the host response to counteract the invasion and the infection progression. Illumina NovaSeq 6000 generated 101 nucleotides-long paired-end reads. Excluding bacterial control samples (bacteria in the host cell media), we obtained on average ~50 millions of reads per library (~11-150 millions of reads) (Supplementary table 1). Libraries for the bacterial control samples were composed on average by ~18 millions of reads (~8-40 millions of reads) (Supplementary table 1). The overall quality of the data set was really high, with all samples showing a narrow peak in the distribution of the mean sequence quality over all sequences around 36 (Phred score). Importantly, the percentage of bacterial reads in the infected samples was reported to be between 0.3 and 2.2% (on average ~600 000 reads), so in line with the previous dual RNA-Seq experiments in the literature (Figure 6A, Table 2). Discarded reads, i.e. reads mapping neither to the host genome or to the bacterial one, were less than 5% for the majority of samples, usually mapping to specific human haplotypes not included in the generation of the human genome index (Figure 6A). PCA analysis for the host clearly separated the tumoral cell line representing the oropharyngeal tract from the immortalized ones representing the urogenital anatomical site (PC1), and the distinct cell lines in general (PC1 and PC2), so highlighting the diversity between cell lines as the driving source of variance (Figure 6B). Furthermore, we also observed that the within-group variability of the host response was strongly cell line-dependent, with Detroit562 tumoral cells being more coherent in their transcriptional behavior compared to End1 and t-UEC cells (Figure 6B). This can be due in part to the fact that we are studying an extreme treatment, like an infection, that in some cases can also culminate in cell damage or cell death. However, the difficulty to discriminate between the different conditions by PCA plots for the three cell lines suggested us the introduction of technical unwanted variations that we checked for and corrected with the recently developed RUV method. On the contrary, the transcriptional response in the pathogen from the PCA picture more clearly clustered in a condition-dependent way (Figure 6C). Interestingly, the bacterial response upon 120 minutes of infection (no sorting) appeared to be more similar to the controls compared to the response at 15 mpi, suggesting a huge transcriptome rewiring as soon as the bacteria start to interact with the host, followed by a later situation of higher adaptation that resembles more the pre-infection one to the medium, if we consider the bulk population where indeed the majority of cells was still extracellular (Figure 6C).

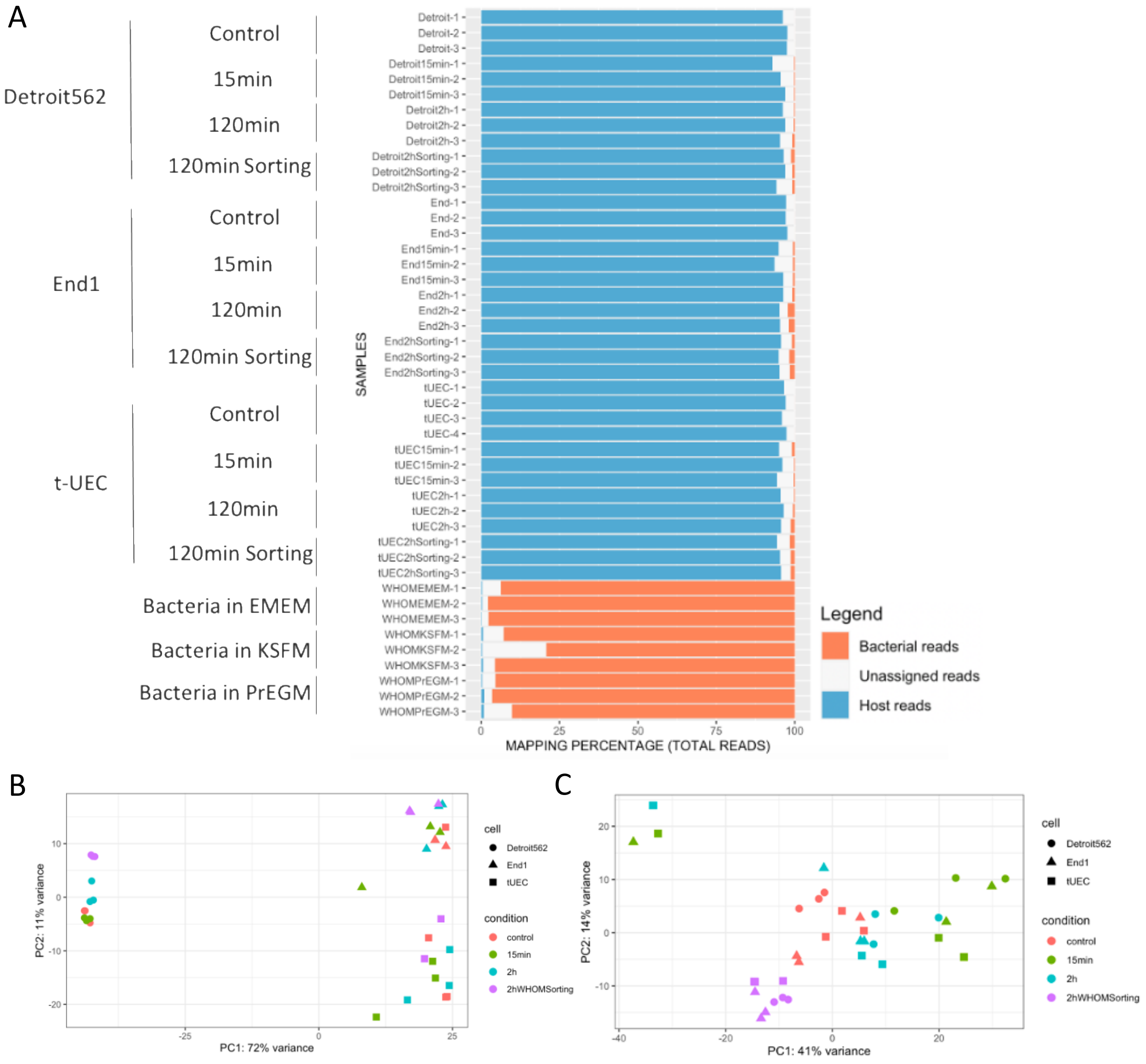


Figure 6 | Relative mapping percentages and PCA analysis of the dual RNA-Seq experiment. A) Relative mapping percentages respect the total number of sequenced reads for both the host (blue) and the pathogen (orange) in the different samples. **B-C)** PCA plot for the host (**B**) and the pathogen (**C**).

3.3 HOST TRANSCRIPTOMIC RESPONSE

3.3.1 HOST IMMEDIATE RESPONSE (15 MINUTES POST-INFECTION)

3.3.1.1 OROPHARYNGEAL CELLS DON'T CHANGE THEIR TRANSCRIPTOME. FOSB IS UP-REGULATED

Starting from the host cells, the transcriptomic picture upon 15 minutes of infection was able to capture the first events probably linked with the early interaction with the pathogen, subsequently leading to its internalization, and the beginning of a pro-inflammatory response sustained by the infected cells. Consistently with the image of fewer bacteria initially attaching to the oropharyngeal cell lines and to the host PCA picture poorly separating 15 mpi Detroit562 samples from the control ones, Detroit562 response upon 15 minutes was smaller compared to that one mounted by End1 and t-UEC cells. A total of 35 genes (27 up-regulated and 8 down-regulated) emerged as differentially expressed compared to the pre-infection controls (adjusted p-value ≤ 0.01) (Figure 7A). Among them, different genes involved in calcium-dependent and cytoskeletal signaling pathways, probably induced upon the first host-pathogen interaction, emerged as a small up-regulated cluster, including calpain 5 (*CAPN5*), the ATPase plasma membrane calcium transporting 4 (*ATP2B4*), the type I intermediate filament chain keratine 18 (*KRT18*) and the microtubule binding protein showing also Ca^{2+} /calmodulin-dependent kinase functions *DCLK1* (Figure 7A). Furthermore, *CLDN11*, encoding a claudin, was among the down-regulated genes, further suggesting physical rearrangements concerning cell adhesion, polarity and signal transduction (Figure 7A). Two important regulators for the host response started to be up-regulated already at 15 minutes, that are the stress induced, MAPK-regulated member of the ATF/cAMP-response-element-binding protein (CREB) family of transcription factors (TFs) *ATF3* and *NFKBIZ*, involved in the regulation of NF-kB transcription factor complexes (Figure 7A). Furthermore, we observed the up-regulation of *CD55*, the metallopeptidase *MMP1*, the phosphatases *DUSP1* and *DUSP5* and the interleukin 1 receptor *IL1RL1*, which signaling, probably in concert with *NFKBIZ*, activates the expression of inflammatory genes (Figure 7A). From the enrichment analysis on transcription factor binding sites (TFBSs) in the promoters of the resulting 35 DE genes, different motifs for TFs belonging to *FOS* and *JUN* gene families emerged (Figure 7B). *FOSB* was also among the significantly regulated genes, suggesting the idea that autoregulatory loops involving AP-1 regulators of the inflammatory response were starting to be established (Figure 7A-B). Interestingly, two members of the carcinoembryonic antigen gene family were up-regulated in a similar way by these cells, *CEACAM1* and *CEACAM5* (Figure 7A).

3.3.1.2 ENDOCERVICAL CELLS START TO MOUNT A PRO-INFLAMMATORY RESPONSE

Moving to the End1 cell line, an initial pro-inflammatory response upon 15 minutes of infection was a bit clearer (a total of 557 DE genes, adjusted p-value ≤ 0.01) (Figure 7C). Among the enriched KEGG pathways for the list of DE genes at 15 mpi compared to pre-infection state, we observed

- TNF signaling pathway (adjusted p-value 1.19e-05),
- IL-17 signaling pathway (adjusted p-value 2.81e-04),
- cytokine-cytokine receptor interaction (adjusted p-value 6.38e-03),
- NF-kappa B signaling pathway (adjusted p-value 8.13e-03) (Figure 7C).

Among the most up-regulated genes belonging to these enriched categories different chemokines (*CXCL5*, *CXCL8*, *CCL20*, *CXCL10*, *CXCL1*) already emerged, indicating that these cells already started to mount a chemotactic response to recruit immune cells to counteract the infection (Figure 7C). Also, some cytokines were up-regulated, such as the interleukin 15 (*IL15*) and the tumor necrosis factor (*TNF*), together with TNF-induced genes (*TNFAIP3*, *TNFSF15*) (Figure 7C). Furthermore, *BIRC3*, a factor modulating the inflammatory signaling and inhibiting apoptosis by intersecting TNF signaling pathway, also emerged as up-regulated upon 15 minutes of infection (Figure 7C).

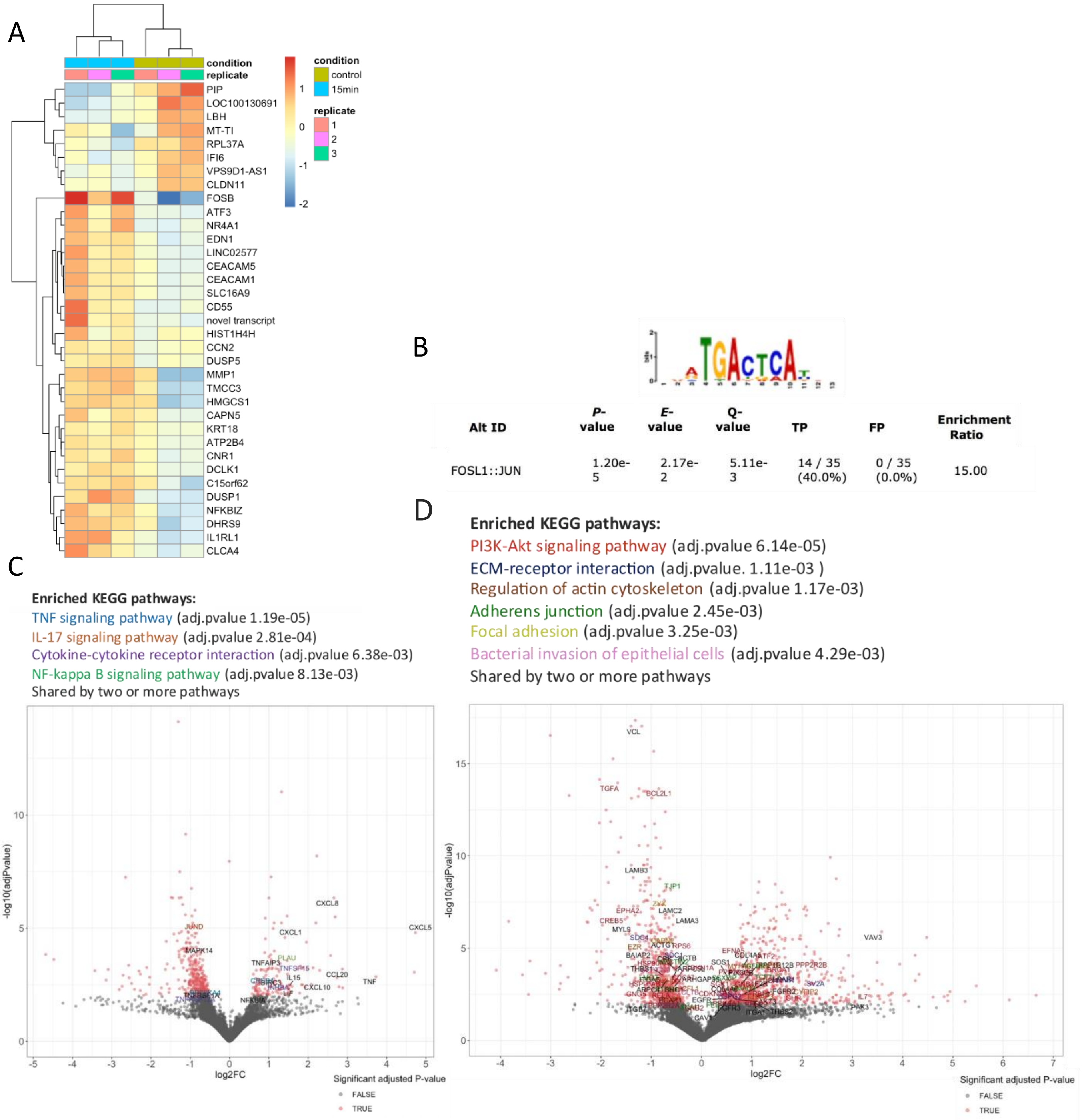
3.3.1.3 IN URETHRAL CELLS THE INITIAL INFECTION TRIGGERS MORPHOLOGICAL CHANGES BUT NO INFLAMMATORY SIGNATURES

We observed the biggest response by the t-UEC cell line (1538 DE genes, adjusted p-value ≤ 0.01) (Figure 7D). However, no inflammatory signatures emerged as significantly enriched inside the list of detected DE genes, probably indicating a delay or a lower efficiency in the response by these cells. On the contrary, among the enriched KEGG pathways we observed categories like

- PI3K-Akt signaling pathways (adjusted p-value 6.14e-05),

- ECM-receptor interaction (adjusted p-value 1.11e-03),
- regulation of actin cytoskeleton (adjusted p-value 1.17e-03),
- adherens junction (adjusted p-value 2.45e-03),
- focal adhesion (adjusted p-value 3.25e-03),
- bacterial invasion of epithelial cells (adjusted p-value 4.29e-03),

promoting these cells as the best candidates to focus on the regulations that accompany pathogen invasion (Figure 7D). Vinculin (*VCL*) and different laminins (*LAMB3*, *LAMC2*, *LAMA3*) emerged among the top significantly down-regulated genes belonging to these pathways, while the guanine nucleotide exchange factor (GEF) for Rho family GTPases *VAV3* was among the most up-regulated ones, indicating the physical rearrangements in cell-cell adhesion, cell-matrix interaction and cytoskeletal actin that cells undergo upon the interaction with the pathogen (Figure 7D). Other regulated genes that could be particularly involved in bacterial invasion included different actins and actin-related proteins (*ACTG1*, *ACTB*, *ARPC5L*, *ARPC4*), cortactin (*CTTN*), clathrin light chain B (*CLTB*) and caveolin 1 (*CAV1*), different adapter or scaffold proteins that could link tyrosine kinase-based signaling with cell adhesion and cytoskeleton (*CRK*, *SHC1*, *BCAR1*), the integrin subunit alpha 5 (*ITGA5*), the phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit beta (*PIK3CB*) and septin 9 (*SEPTIN9*) (Figure 7D).



Alt ID	P-value	E-value	Q-value	TP	FP	Enrichment Ratio
FOSL1::JUN	1.20e-5	2.17e-2	5.11e-3	14 / 35 (40.0%)	0 / 35 (0.0%)	15.00

Figure 7 | Host transcriptomics response upon 15 minutes of infection. A) Heatmap of mean-centered transformed read counts (normalized with respect to library size). Colors from yellow to red represent up-regulated genes, colors from light blue to blue represent down-regulated genes. All DE genes for Detroit562 cells at 15mpi are reported (adjusted p-value ≤ 0.01). **B)** Sequence logo and related statistics for the top significant result (mouse-related one excluded) of SEA motif enrichment analysis on the promoters of the whole set of DE genes at 15mpi for Detroit562 cells (motifs need to have an E-value of 10 or lower for significance). Alt ID = alternative name for the motif, TP = percentage of primary sequences matching the motif, FP = percentage of control sequences matching the motif, the Enrichment Ratio is given by ((TP+1)/(NPOS+1)) / ((FP+1)/(NNEG+1)), where NPOS is the number of primary sequences in the input and NNEG is the number of control sequences in the input. **C-D)** Volcano plots for End1 (**C**) and t-UEC (**D**) cells considering 15mpi time point versus uninfected controls. Dots for genes are colored accordingly to their significance, while gene names associated to some enriched KEGG pathways (reported on the top of the plot) among the emerging top ones are colored accordingly to the pathway. For the volcano plot in **D**, the first 7 significant DE genes, not falling inside the reported pathways, were removed just for visualization purposes.

3.3.2 HOST RESPONSE UPON 120 MINUTES OF INFECTION

3.3.2.1 A NF-kB AND JUN-DRIVEN INFLAMMATORY RESPONSE IS ESTABLISHED BY HOST CELLS UPON 120 MINUTES OF INFECTION

Upon 120 minutes of infection, the host responses were more comparable among the three cell lines. In total 318 genes (262 up-regulated, 56 down-regulated) for Detroit562 cells, 239 genes (201 up-regulated, 38 down-regulated) for End1 cells and 153 genes (150 up-regulated, 3 down-regulated) for t-UEC cells were called as significantly differentially expressed compared to pre-infection controls (adjusted p-value ≤ 0.01). GSEA analysis for all the three cell models clearly highlighted from the up-regulated genes an involvement of inflammatory processes and responses to bacteria, regulation of signaling pathways involved in defense and immune cells activation and recruitment. Indeed, among the top 20 significantly enriched GO terms with positive normalized enrichment score (NES) shared by all the cells we reported

. inflammatory response,
. response to bacterium,
. positive regulation of MAPK cascade,
. cellular response to lipid,
. taxis,
. response to toxic substance (Figure 9A).

For each of the three cell lines, upon building an entire human TF-target genes regulatory network, we mapped all the genes in the leading edge for these six enriched terms that showed regulations among them to depict the putative regulatory dynamics the cells were undergoing upon 120 minutes of infection (Supplementary Figure 4). *NFKB1*, *REL* and all their associated genes, including *BCL3* in the autoregulatory loop, as well as *JUN*, so all final regulatory effectors of the signaling cascade occurring in the host cells upon pathogen recognition, clearly emerged as hub genes in Detroit562 and End1 networks. They may be responsible for the up-regulation of many target genes involved in the inflammatory process (including chemokines, *TNF*, *IL1A*, *IL1B*, *IL6*), therefore suggesting a picture that upon 120 minutes of infection poses these cells in the middle of an acute inflammatory response to the invading bacteria (Supplementary Figure 1A-B). A really similar picture emerged from t-UEC network, even if upstream regulations seemed to be a bit over-represented compared to the other two cell models, suggesting again the idea that the infection process in these cells is slower (Supplementary Figure 1C). Interestingly, also some epigenetic factors, like *HDAC9* and *SIRT1*, seemed to participate to the response (Supplementary Figure 1A-B). An efficient host response to the infection is fundamental for the homeostasis maintenance, however an exaggerated one could favor pathological conditions, so it's interesting to note in the networks also the presence of many repressors and modulators of NF-kB and the inflammatory response, such as *TNFAIP3, NR4A1, NR4A2, NCOR2* (Supplementary Figure 1). Figure 9B reports how was similar the response, in terms of top 10 significantly differentially expressed genes, observed in Detroit562 and End1 at 120 mpi. Among the most significantly DE genes, we observed chemokines, cytokines and NF-kB-related factors, as well as inflammatory modulators (the aforementioned rapidly TNF-induced *TNFAIP3* and *ZC3H12A*), highlighting once again how much is important for the host to guarantee a transient nature of the inflammatory processes (Figure 9B).

3.3.2.2 t-UEC CELLS DOWN-REGULATE THEIR LYSOSOMAL GENES EXPRESSION

From the enrichment analysis on Cellular Compartment GO terms, only for t-UEC cells that internalized the bacteria the lysosome term emerged as significantly enriched in our data (adjusted p-value 0.01), even if a good portion of regulated genes also in case of the other two infections fell inside this category. Interestingly, a total of 153 lysosome-associated genes were regulated by t-UEC cells with the majority of them (101) being down-regulated (contrary to what is observed in the other two cell lines), as a possible consequence of gonococcal mechanisms activated in this particular infection to limit autophagy and to survive in the host cells (Ayala *et al.*, 1998).

3.3.2.3 AFTER 120 MINUTES OF INFECTION HOST CELLS RESPOND TO INTERNALIZED BACTERIA REGULATING CHEMOTACTIC FACTORS AND OTHER INFLAMMATORY GENES

To amplify the signal from cells that internalized the bacteria at 120 minutes, we sequenced also such population upon enrichment by FACS sorting. This gave us more sensitivity in detecting DE genes (in the contrast with uninfected controls) compared to the bulk population. It was particularly critical for t-UEC cells, for which the infection appeared to be less efficient from both phenotypic and transcriptomic characterizations. Indeed, the correlation of the $\log_2$ fold change for all the genes excluded the ones specifically regulated in sorted or unsorted samples (the latter being quite few, particularly for Detroit562 infection) was worse for t-UEC cells compared to End1 and especially Detroit562 ones, recapitulating the different situations observed by confocal microscopy and suggesting that a higher diversity separated the two populations in the case of t-UEC cell line (Figure 8). Finally, Figure 9C depicts the distribution of the mean of centered normalized counts for relevant inflammatory genes, divided in the different classes they belong to, resulting as differentially expressed (adjusted p-value ≤ 0.01, | $\log_2$ fold change | > 0.8) in the sorted samples of the three cell models compared to the relative controls. Among the most expressed categories of genes, we clearly saw chemokines, highlighting how the host, upon 120 minutes of infection, is dedicating a huge portion of its efforts in trying to recruit immune cells to clear the infecting pathogens (Figure 9C).

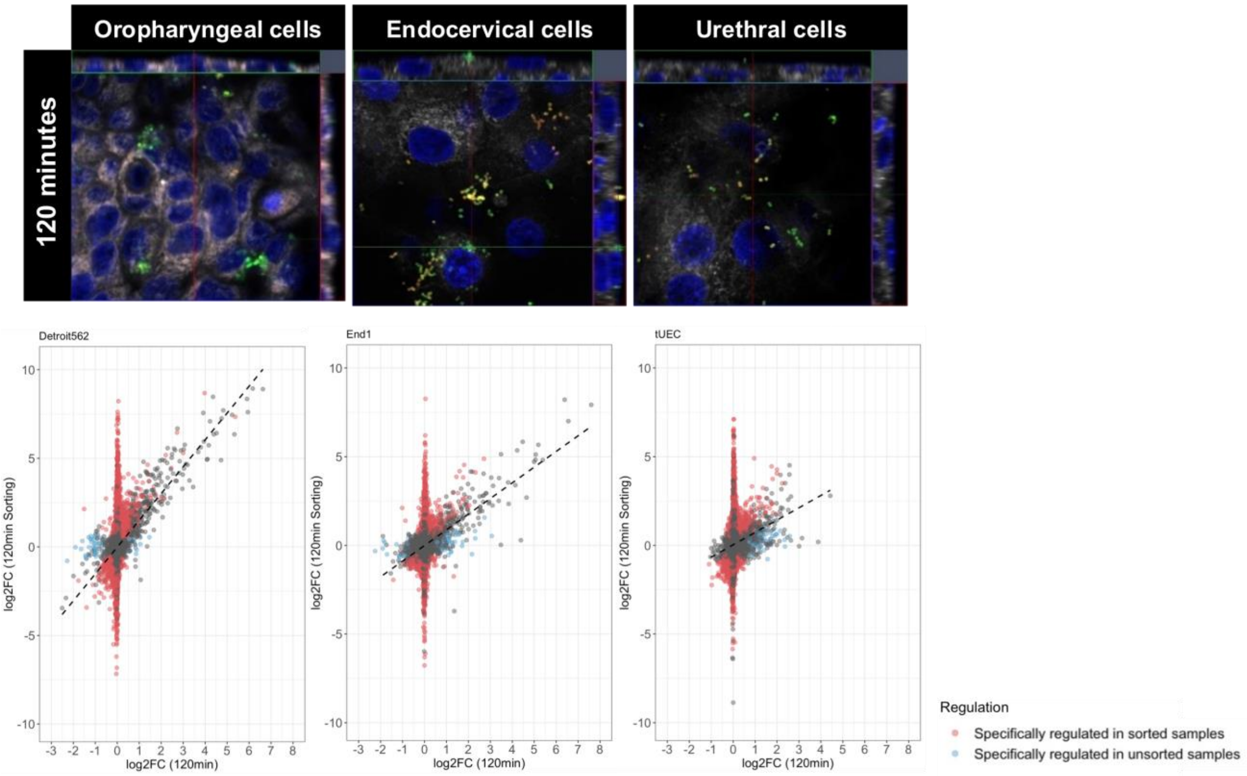


Figure 8 | Transcriptomics data for the host reflected the different infection situations observed in confocal microscopy at 120 mpi. On the top, confocal microscopy images for the three cell models upon 120 minutes of infection by *N. gonorrhoeae* WHO M, as reported in Figure 5A. On the bottom, plots comparing $\log_2$ fold changes for all the genes in unsorted and sorted samples at 120mpi for all the three cell models. Correlation straight lines are drawn on the basis of shared, significantly regulated or not genes only. Red dots indicate genes emerging as significantly regulated only in sorted samples, blue dots indicate genes emerging as significantly regulated only in unsorted samples, gray dots all the others. Adjusted p-value to call significant DE genes ≤ 0.01.

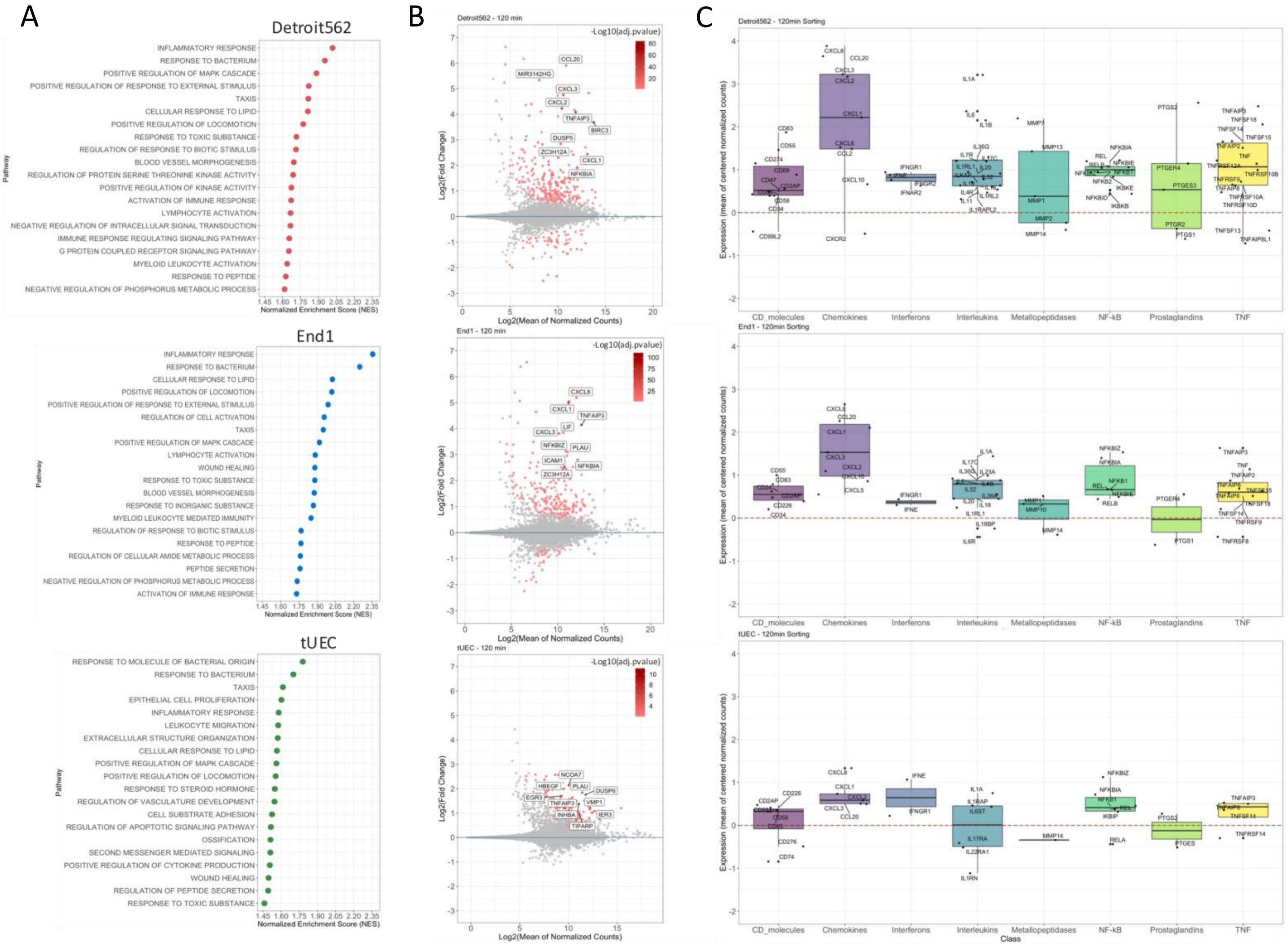


Figure 9 | Host transcriptomics overview in the three cell models upon 120 minutes of infection by *N. gonorrhoeae* WHO M. A) Top 20 significantly enriched GO biological processes with positive normalized enrichment score (NES) by GSEA analysis on DE data at 120mpi (no sorting) for the three cell models. Significant adjusted p-value to call DE genes ≤ 0.01, significant adjusted p-value for GSEA analysis ≤ 0.05. **B)** MA plots representing DE analysis results for the comparison between 120mpi unsorted samples and the uninfected ones in the three cell models. Red dots represent significantly DE genes, the intensity of the color is scaled accordingly to the significance. Significant adjusted p-value to call DE genes ≤ 0.01. **C)** Boxplots of mean-centered transformed and normalized gene counts for relevant genes emerging as differentially regulated at 120mpi compared to uninfected controls, considering sorted samples of the three cell models, grouped by their inflammatory-associated category. Genes were selected based on their significance from DE testing ≤ 0.01 and on an estimated |log$_2$ fold change| > 0.8.

3.4 PATHOGEN TRANSCRIPTOMIC RESPONSE

3.4.1 OVERVIEW OF PATHOGEN RESPONSE UPON INFECTION OF THE THREE CELL LINES

3.4.1.1 BACTERIAL TRANSCRIPTOMES SHOW BOTH PECULIAR AND COMMON PATTERNS DURING EARLY INFECTION OF THE THREE CELL MODELS

Moving to the pathogen, the picture concerning the number of detected DE genes (adjusted p-value ≤ 0.05) confirmed the aforementioned idea of strong transcriptional changes driven by the bacteria upon 15 minutes of infection compared to non-infecting controls (up to ~16% of the total genes in case of Detroit562 infection, i.e. 371 DE genes out of 2342 total genes), followed by a situation of higher adaption upon 120 minutes, but starting an intracellular lifestyle, as reported by the analyses on the sorted samples, required again significant transcriptional activities (Figure 10A). This trend was less evident only considering the infection of t-UEC cell line, where maybe the already observed delay or difficulty in the host response to the infection process could reflect also in a more delayed global transcriptome rewiring occurring in the pathogen (Figure 10A). On the contrary, consistently with the described image of a lower initial bacterial adhesion to Detroit562 cells, the highest signal upon 15 minutes emerged from the bacterial infection of this cell line (Figure 10A). Importantly, from the comparison of bacteria upon 30 minutes of adaptation in the three different cell media (pre-infection samples), strong differences emerged from bacteria adapted to the EMEM medium compared to bacteria in KSFM and PrEGM, making more difficult a direct comparison between pathogens infecting Detroit562 cells and those infecting the other two cell lines (Figure 10A). On the contrary, comparing bacteria adapted to PrEGM medium to bacteria adapted to KSFM one, just 12 genes were called as differentially expressed (7 up-regulated, 5 down-regulated; adjusted p-value ≤ 0.05), with nearly all of them being potentially implicated in iron uptake and metabolism, such as

. *fbpA* ($\log_2$ fold change 1.4),

. two bacterioferritin genes (C7R98_RS05060 and C7R98_RS05065, $\log_2$ fold changes -2.5 and -2.4, respectively),

. a transferrin-binding protein-like solute binding protein (C7R98_RS08240, $\log_2$ fold change 2.1),

. ExbB and ExbD (both with a $\log_2$ fold change around 1.8), probably because PrEGM medium was supplemented with transferrin. These genes were differentially expressed in a similar way also comparing bacteria adapted to the EMEM medium and those adapted to the KSFM one. Figure 10B depicts the distribution of all genes regulated during the infection of the three cell lines into the different KEGG pathway macro-groups. Importantly, since WHO M strain is not present into KEGG database, genes were first annotated by homology to the database, resulting in 62.6% of them successfully being annotated, therefore such analysis could in some extents suffer from the limited number of assigned genes. From this picture, clearly emerged the global transcriptional program that bacteria started already upon 15 minutes of infection. As expected, translation was one of the most up-regulated classes at any of the investigated infection stages (Figure 10B). Interestingly, it was also quite evident, in case of Detroit562 and t-UEC infections, the up-regulation of genes involved in replication and repair in bacteria that have been internalized (sorted samples), suggesting that the pathogen was starting to replicate and successfully survive in the host environment (Figure 10B). Curiously, this was not valid for End1 infection. A similar trend common to the three intracellular infections was also linked to the regulation of membrane transport pathways, probably linked to the nutrient acquisition needed for bacterial survival in the host environment (Figure 10B). Finally, a small interesting portion of genes related to antimicrobial resistance was also up-regulated, particularly during the intracellular stages (Figure 10B).

3.4.1.2 LOS GENES ARE UP-REGULATED AND mtrCDE ONES ARE DOWN-REGULATED BY WHO M INITIALLY INFECTING DETROIT562 CELL LINE

Supplementary Figure 5 reports the regulations of DE genes by bacteria at 15 mpi and 120 mpi (sorted samples) falling inside some interesting categories among the aforementioned KEGG ones. For what concerns Detroit562 infection, we observed considerable carbohydrate and amino acid metabolisms regulation already upon 15 minutes of infection. For instance, the glycolytic enzyme enolase *eno* and the ATP synthase *atpD* were induced, and particularly methionine (*metE*, *metF*) and leucine (*leuB*, *leuD*) biosynthetic pathways were up-regulated. At the same time point, we reported also the induction of different genes encoding important cofactors or secondary metabolites, such as *ribB* (riboflavin metabolism), a bacterioferritin (C7R98_RS05060, iron-storage protein), *hemC* (metabolism of porphyrin-containing compounds) and *coaBC* (coenzyme A biosynthesis). Interestingly, some genes for LOS were also up-regulated, indicating their possible involvement also in the initial phases of bacterial adhesion, such as the 2-dehydro-3-deoxyphosphooctonate aldolase *kdsA* and the *lpxD* gene, related to lipid A biosynthesis, while *mtrCDE* genes for the antimicrobial resistance pump were initially down-regulated. Upon 120 minutes of infection, carbohydrate and amino acid metabolisms continued to be strongly regulated (both up- and down-regulations). Particularly the up-regulation of ATP synthases *atpB* and *atpG*, together with the oxidoreductase *lpdA*, emerged about the former, while D-alanine, isoleucine and serine biosynthetic pathways (*alr*, *ilvA*, *serB*) appeared to be induced concerning the latter. Also, *secE* and *secY*, encoding components of the protein translocation machinery, were strongly up-regulated by internalized bacteria.

3.4.1.3 LIPID METABOLISM IS DOWN-REGULATED IN BACTERIA INTERNALIZED BY UROGENITAL CELL LINES

Upon 15 minutes of infection of End1 cell line, compared to the previous case, we observed more up-regulation concerning carbohydrate metabolism, with different genes involved in the tricarboxylic acid cycle being induced (*sucC*, *sucD*, *sdhA*, *gltA*), over the ATP synthases *atpG* and *atpD*. At 120 mpi, bacteria internalized by these cells again regulated different genes involved in carbohydrate and amino acid metabolisms (*fumC*, *pgl*, *sdhA*, *atpG*, *alr*, *ilvB*), while lipid metabolism was mainly down-regulated compared to the pre-infection controls. Also, *ubiD*, involved in ubiquinone biosynthesis, the aforementioned *coaBC*, and *hemN*, involved in the biosynthesis of porphyrin-containing compounds, emerged as up-regulated at this time point. Bacteria infecting t-UEC cell line at 15 mpi up-regulated some genes involved in amino acid biosynthesis, such as *metE*, *argB* and *luxS*, the last being also related to the synthesis of the autoinducer (AI-2), secreted by bacteria to communicate both cell density and metabolic potential of the environment, and some bacterioferritin genes (C7R98_RS05060 and C7R98_RS05065). Finally, upon 120 minutes of infection, bacteria significantly regulated carbohydrate, amino acid and cofactor and secondary metabolisms. Concerning the latter, among the other genes, we observed again the up-regulation of *ubiD*, *coaBC*, *hemN* and *ribA*. The lipid metabolism again appeared down-regulated, while *waaF* and *kdsA*, involved in LOS biosynthesis, interestingly emerged as up-regulated genes.

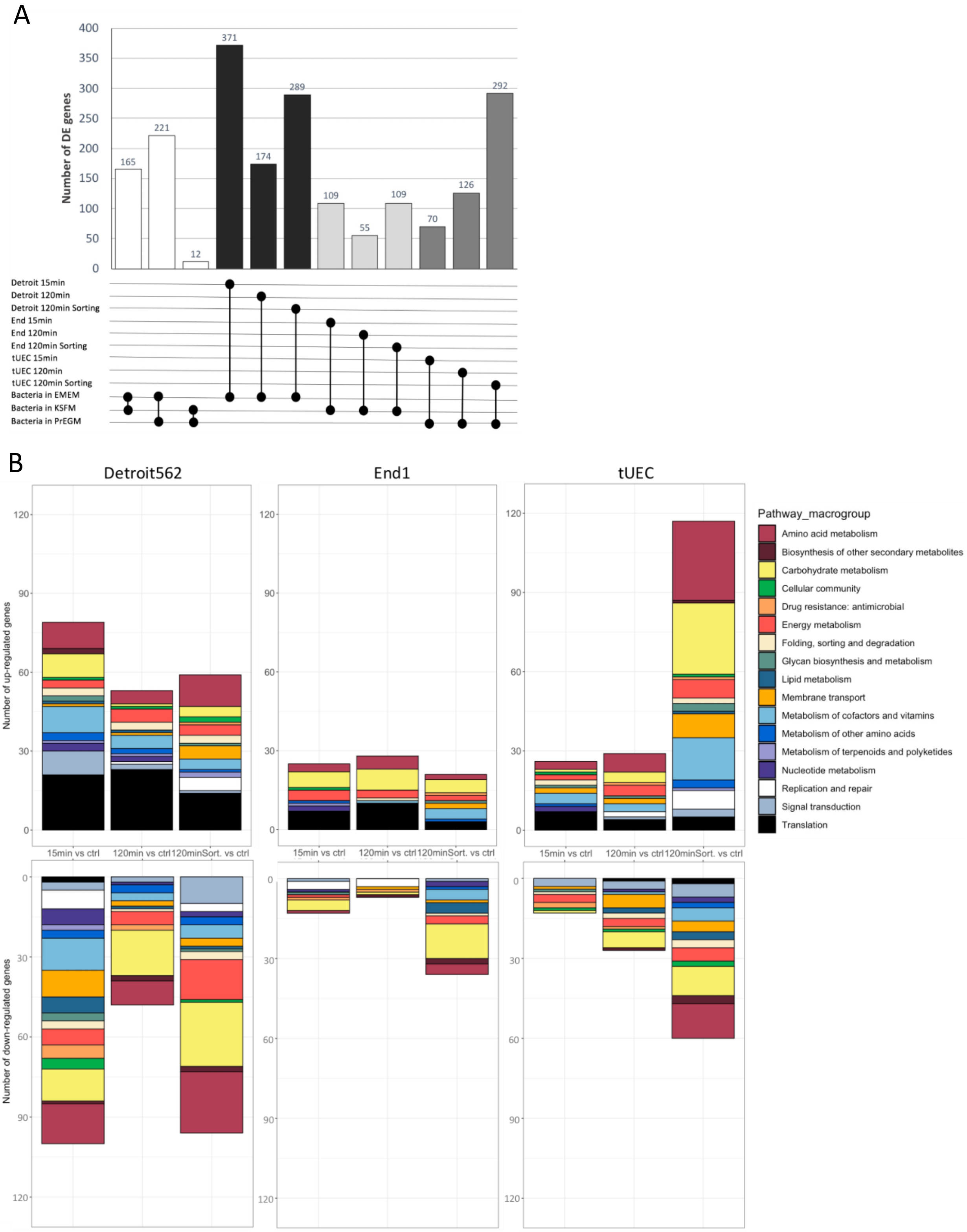


Figure 10 | Overview of *N. gonorrhoeae* WHO M transcriptome rewiring during early infection of the three cell models. A) Bar plot representing the number of pathogen genes (on a total of 2342 genes) emerging as significant differentially expressed genes from the different comparisons reported at the bottom. Significant adjusted p-value to call DE genes ≤ 0.05. **B)** Bar plots subdividing the sets of emerging pathogen DE genes (that match by orthology search the KEGG database) from the various comparisons in the different infection models in their relative pathway macro groups. Significant adjusted p-value to call DE genes ≤ 0.05.

3.4.2 *N. gonorrhoeae* WHO M REGULATION OF ADHESION AND INVASION GENES

3.4.2.1 PILINS AND OUTER MEMBRANE BETA-BARREL PROTEINS ARE DIFFERENTIALLY REGULATED BY WHO M INFECTING DETROIT562 CELLS

Upon 15 minutes of infection, bacteria infecting Detroit562 cells clearly up-regulated a set of genes that could favor adhesion, mainly genes encoding for pilins (such as type IV pilins and related genes, well-known factors required by *N. gonorrhoeae* to adhere to the host cells) and different outer membrane beta-barrel proteins showing around 75-85% amino acid sequence identity with the opacity protein opA54 in FA1090, as emerged by BLAST alignment (Figure 11A). On the contrary, upon 120 minutes of infection a more "stable" picture of no regulation was reported for these genes, that once the signal for internalized bacteria was enriched, further evolved in different down-regulations, regarding for instance type IV pilin proteins and pilin assembly or biogenesis proteins like PilM and PilD (Figure 11A). Great portions of genes were commonly regulated comparing all the different stages of Detroit562 infection (Figure 11B). Interestingly, a total of 8 genes were regulated in an opposite way and showed a strong anti-correlation in their fold changes comparing the situation at 15 minutes and that one in the internalized bacteria, so representing potential candidates for pathogen adaptation to the changing environment (Figure 11C). A clear example of that is depicted in Figure 11D, that shows how the type IV pilin protein C7R98_RS07960 was up-regulated and down-regulated by the pathogens in a nearly specular way upon 15 and 120 minutes of infection, respectively. Among the others, we observed different genes, showing up-regulation in the internalized bacteria, having a known role in cell survival for *N. gonorrhoeae* and/or other bacteria, such as the GTPase Era, the LPS export system protein LptC, the antimicrobial resistance efflux pump subunit MacA and the cell envelope homeostasis maintenance protein MlaC (Figure 11C).

3.4.2.2 A SIMILAR EXPRESSION TREND IS SHOWED BY BACTERIA INFECTING END1 CELL LINE. OUTER MEMBRANE BETA-BARREL PROTEINS ARE LOW EXPRESSED IN BACTERIA INFECTING t-UEC CELL LINE

The aforementioned outer membrane beta-barrel protein genes that were regulated in case of Detroit562 infection upon 15 minutes showed similar high expression and trend (in terms of variance-stabilized and normalized counts) also considering bacteria infecting End1 cells, however in this case they didn't emerged as differentially expressed due to the fact that they appeared as highly expressed already in bacteria adapted to the cell medium (Figure 12A).

This could also explain the smaller bacterial response characterizing the early infection of End1 cells and the higher rate of initial adhesion to them compared to Detroit562 ones. Strangely, this set of genes was characterized by a consistently low expression if considering bacteria infecting t-UEC cells and bacteria adapted to the PrEGM medium (Figure 12A). The majority of pilin genes showed lower levels of transcript abundances, but a trend of higher expression at 15 mpi emerged for all the three infections (Figure 12B). It was clear also for the infection of End1 and t-UEC cell lines the down-regulation of many genes putatively involved in adhesion by the internalized bacteria. For instance, bacteria internalized by End1 cells showed down-regulation of the outer membrane beta-barrel protein NspA ($\log_2$ fold change -2.6), two type IV pilin proteins (the aforementioned C7R98_RS07960 and C7R98_RS09370, $\log_2$ fold changes -1.9 and -1.4, respectively), a pilin (C7R98_RS13820, $\log_2$ fold change -0.9) and an outer membrane beta-barrel protein (C7R98_RS11645, $\log_2$ fold change -0.9). All these genes were down-regulated also in bacteria internalized by t-UEC cells ($\log_2$ fold changes -2.4, -1.9, -2.4, -1.3 and -1.9, respectively), in addition the bacteria inside these cells also down-regulated type 4 pilus assembly protein PilG ($\log_2$ fold change -1.1), pilus assembly protein PilM ($\log_2$ fold change -1.5), PilN domain-containing protein ($\log_2$ fold change -1.0) and type 4a pilus biogenesis protein PilO ($\log_2$ fold change -1.0).

3.4.2.3 porB IS AMONG THE MOST EXPRESSED BACTERIAL GENES

As expected, a gene annotated as a porin-encoding one (C7R98_RS11970) showing 100% amino acid sequence identity and coverage by BLAST search with *Neisseria gonorrhoeae* major outer membrane PorB.1B protein was reported among the top 20 most expressed genes in each condition and cell line infection (Supplementary Figure 6A). For this reason, it didn’t emerge as DE gene, however it was possible the same to see a trend of higher expression at 15 mpi, confirming the possible involvement also of this major porin in initial bacterial adhesion and invasion (Supplementary Figure 6B) (van Putten, Duensing and Carlson, 1998).

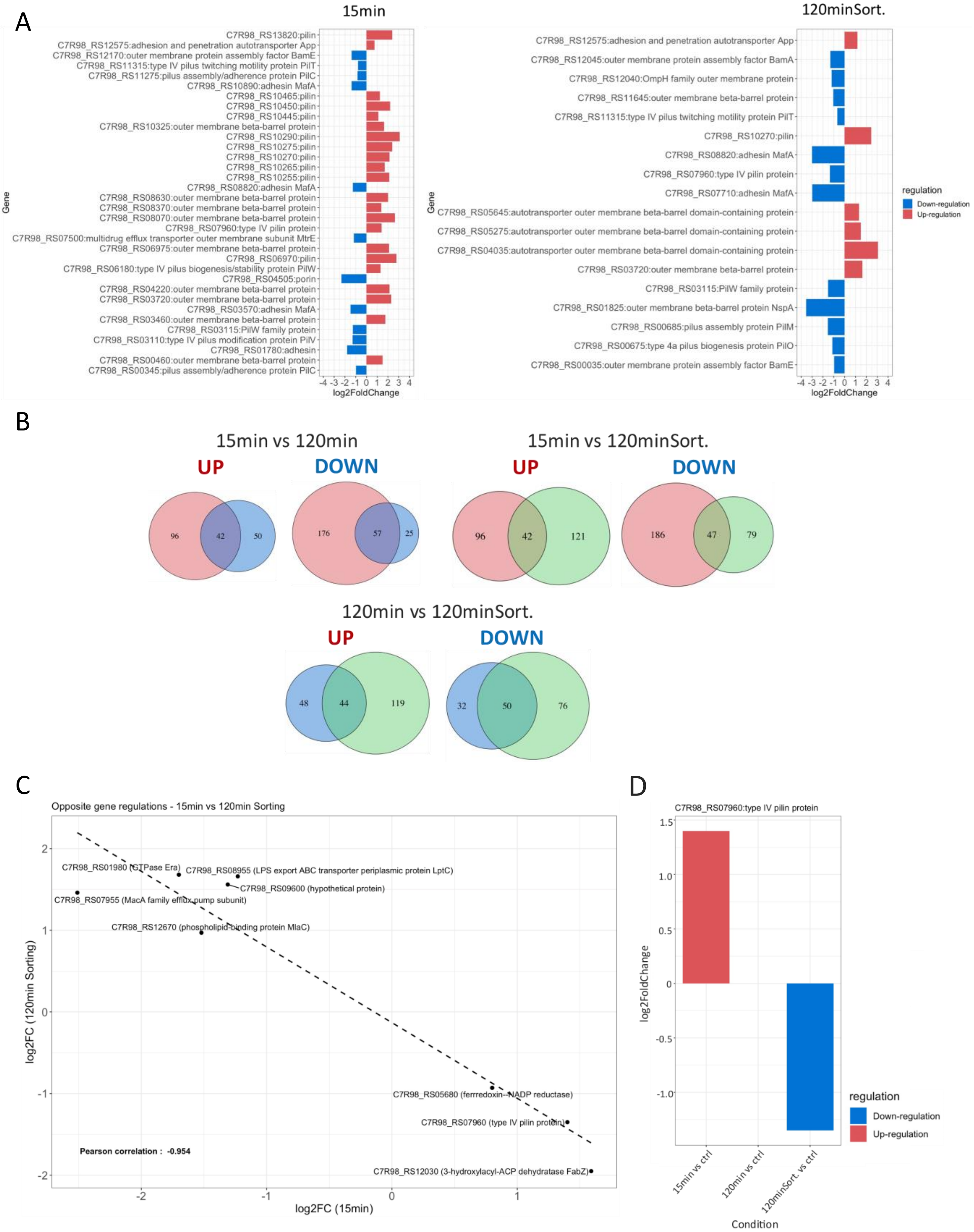


Figure 11 | Set of putative adhesion factors regulated by *N. gonorrhoeae* WHO M infecting Detroit562 cells. A) Bar plots reporting the estimated $\log_2$ fold change for significantly regulated genes putatively involved in bacterial adhesion to Detroit562 host cells, considering both 15mpi and 120mpi (internalized pathogens) comparisons with pre-infection controls. Significant adjusted p-value to call DE genes ≤ 0.05. **B)** Intersection between the sets of significantly DE genes (up- and down-regulated) emerging at the different time points of Detroit562 infection compared to pre-infection controls. Significant adjusted p-value to call DE genes ≤ 0.05. **C)** Correlation of estimated $\log_2$ fold change for significant DE genes showing an opposite regulation if considering 15mpi and 120mpi (sorting) comparisons with pre-infection controls for Detroit562 cell model. Significant adjusted p-value to call DE genes ≤ 0.05. **D)** Estimated $\log_2$ fold change of the indicated gene in the reported comparisons for Detroit562 cell model.

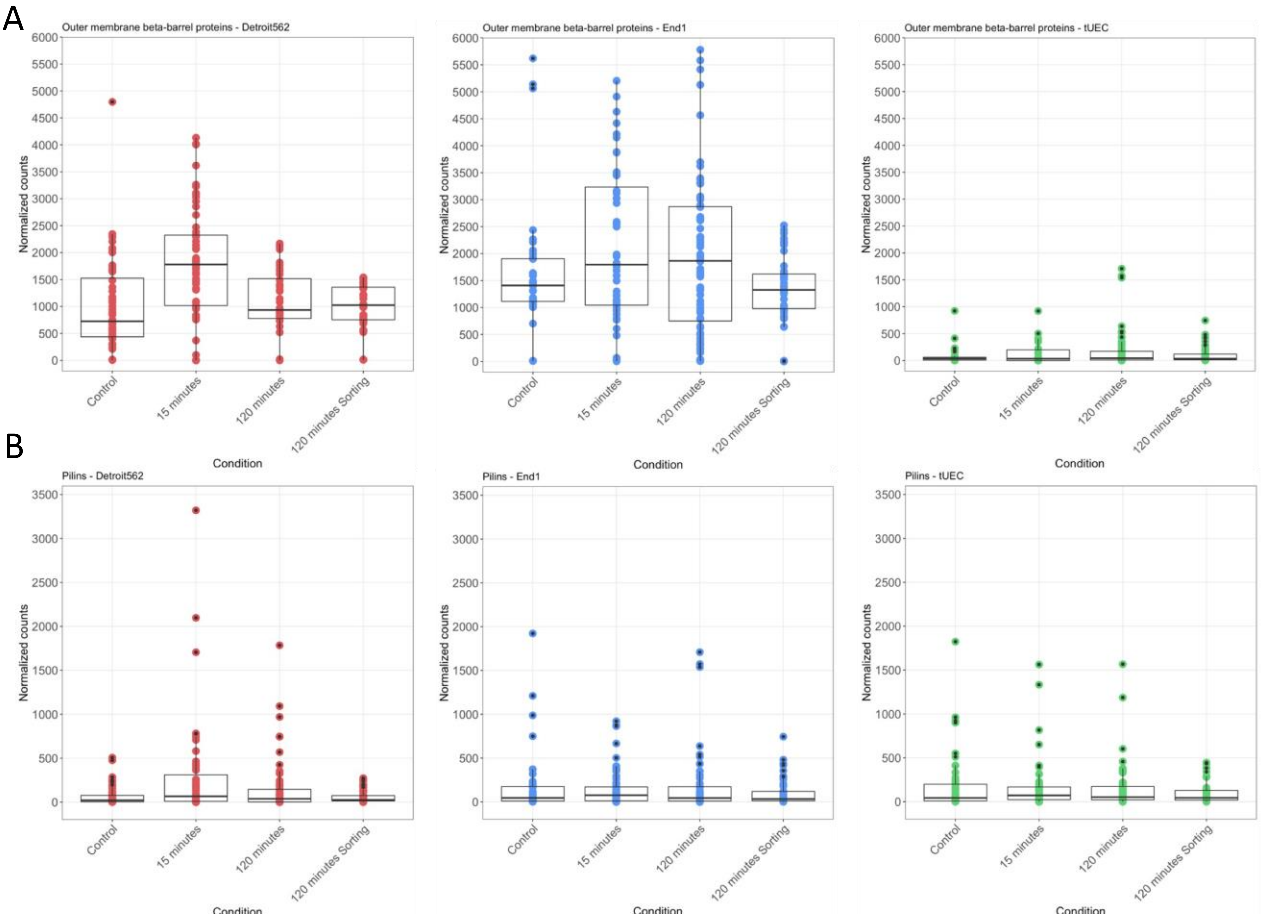


Figure 12 | Outer membrane beta-barrel proteins and pilins regulation by *N. gonorrhoeae* WHO M during early infection of the three cell models. A-B) Normalized counts (with respect to library size) plus a pseudocount of 0.5 related to all outer membrane beta-barrel protein (**A**) or pilin (**B**) annotated genes in *N. gonorrhoeae* WHO M genome for the three infection models at the different time points.

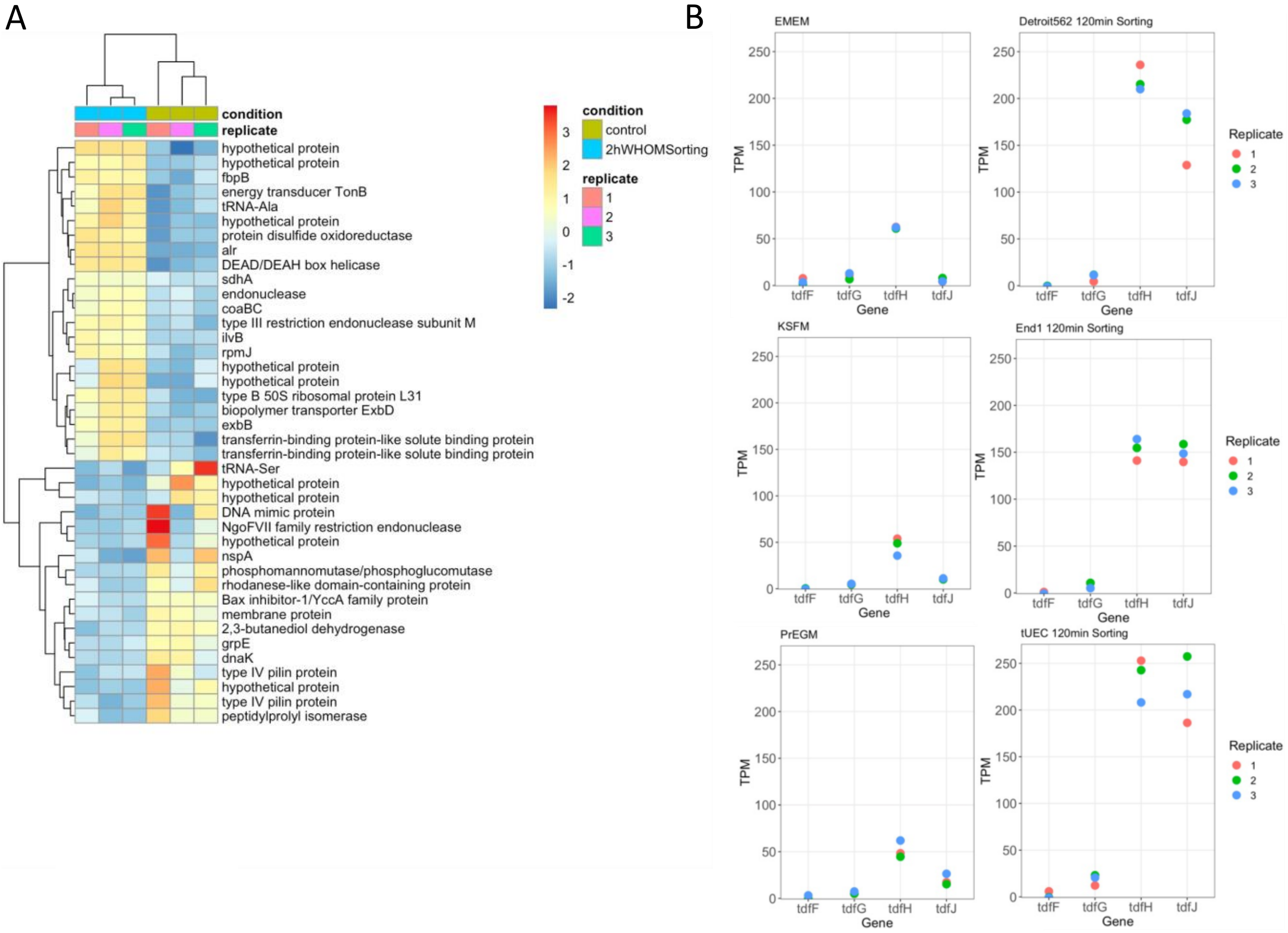
3.4.3 TRANSCRIPTOME REWIRING OCCURRING IN PATHOGENS INTERNALIZED BY HOST CELLS

3.4.3.1 IRON-, ZINC- AND AMINO ACID-RELATED TRANSPORTER GENES ARE REGULATED BY INTRACELLULAR BACTERIA

An iron-related response strongly emerged from bacteria internalized by End1 cells, while it was less evident in case of the other two infections. As expected, TonB ($\log_2$ fold change 2.2) and the related genes ExbB ($\log_2$ fold change 1.5) and ExbD ($\log_2$ fold change 1.6) were among the top differentially expressed genes when considering the sorted samples at 120 minutes compared to the controls for End1 infection, while they were not significantly regulated in case of the other two infections (Figure 13A). Furthermore, among the most significantly DE genes when considering End1 model we observed also *fbpB* ($\log_2$ fold change 1.9) and two transferrin-binding protein-like solute binding proteins (C7R98_RS08240 and C7R98_RS01990, $\log_2$ fold change 1.6 and 1.3, respectively) (Figure 13A). However, it is important to keep in mind that bacteria adapted to the media for Detroit562 and t-UEC cells seemed to differ for what concerns iron metabolism compared to those adapted to the medium for End1 cells, as mentioned before. Interestingly, in contrast with what was previously reported for the FA1090 strain infection of cervical cells, the putative transporter *tdfF* was not required by strain WHO M during the infection of endocervical cells (Figure 13B). On the contrary, other four annotated TonB-dependent receptors emerged as regulated in bacteria internalized by End1 cells: C7R98_RS01530 ($\log_2$ fold change 1.7) showing 99.87% amino acid sequence identity with *tdfJ*, C7R98_RS12500 ($\log_2$ fold change 1.7), C7R98_RS04210 ($\log_2$ fold change 1.0) showing 99.57% amino acid sequence identity with *tdfH* and C7R98_RS06380 ($\log_2$ fold change 1.1) (Figure 13B). A similar situation for the putative TonB-related transporters was reported also in case of Detroit562 and t-UEC intracellular stages (Figure 13B). Finally, in Figure 13C are highlighted other annotated transporters that emerged as DE genes at least in one infection model when comparing internalized bacteria with pre-infection controls. Many of them are not fully characterized, however they seemed to include especially amino acid-, sulfate-, zinc- and ammonium-related transporters (Figure 13C).

3.4.3.2 COMMON BACTERIAL GENES EMERGE AS DIFFERENTIALLY EXPRESSED IN SORTED SAMPLES COMPARED TO UNSORTED ONES

Since the majority of bacteria were not internalized upon 120 minutes of infection, especially considering End1 and t-UEC infections, we directly compared sorted samples and unsorted ones to extract bacterial DE genes that potentially could be relevant during the intracellular stage. The highest number of genes emerged from the End1 infection comparison, promoting this infection as the one with the most well-established diversity in the transcriptome – and probably, behavior – of the two aforementioned bacterial populations. In general, we observed a great set of common genes in all three cases, confirming the transcriptional similarity characterizing internalized bacteria in all the infection models already showed by PCA plot (Figure 6C). A signal in these bacteria concerning for instance down-regulation of some pilins (like C7R98_RS13820) or up-regulation of amino acid-related transporters or metabolic genes (like *ilvB*) and other relevant genes, like *lptC*, was obtained from all the infections (Supplementary table 2-3-4). Furthermore, the energy transducer TonB emerged among the top up-regulated genes in internalized bacteria during Detroit562 and, of course, End1 infections, while it was not reported among DE genes when considering the t-UEC one (Supplementary table 2-3-4).



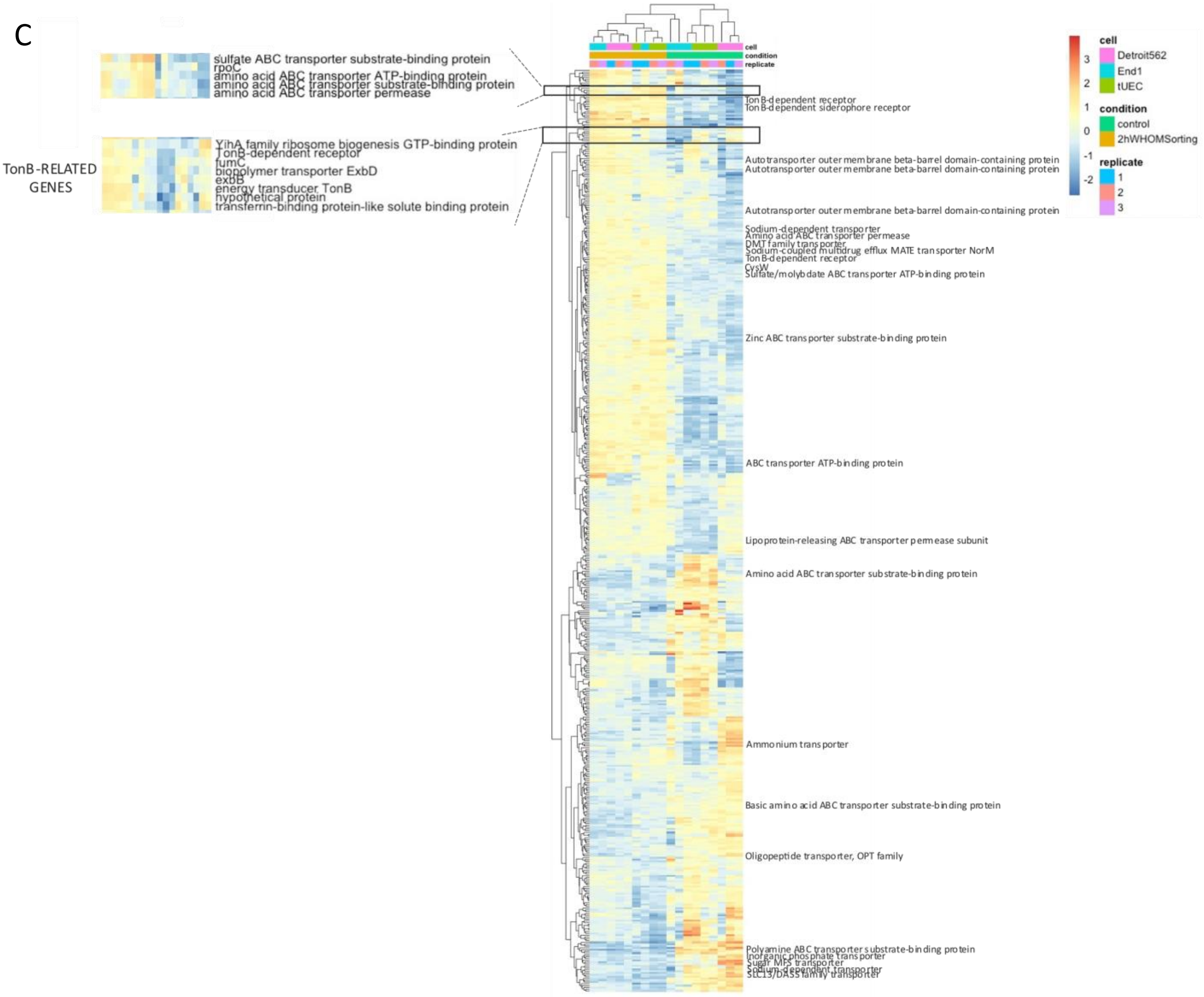


Figure 13 | Regulation of the expression of different transporters by internalized bacteria. A) Heatmap of mean-centered transformed read counts (normalized with respect to library size). Colors from yellow to red represent up-regulated genes, colors from light blue to blue represent down-regulated genes. Top 40 statistically significant DE genes for bacteria internalized by End1 cells compared to the pre-infection controls are reported (adjusted p-value ≤ 0.05). **B)** Transcripts per kilobase of length per million mapped reads (TPM) related to the four literature-described putative gonococcal TonB-dependent transporters for bacteria adapted to the three different cell media and bacteria internalized by the three different cell lines. **C)** Heatmap of mean-centered transformed read counts (normalized with respect to library size). Colors from yellow to red represent up-regulated genes, colors from light blue to blue represent down-regulated genes. All statistically significant DE genes for bacteria internalized by at least one cell line compared to the relative pre-infection controls are reported, but only annotated transporter genes are indicated (adjusted p-value ≤ 0.05).

3.4.4 *N. gonorrhoeae* WHO M CO-EXPRESSION NETWORK ANALYSIS

3.4.4.1 HUBS COULD REPRESENT IMPORTANT GENES FOR PATHOGENS VIRULENCE AND SURVIVAL

Since different hypothetical proteins were called as differentially expressed by our analysis, we also reconstructed from our data the first co-expression network concerning *Neisseria gonorrhoeae* WHO M, in order to try to assign a putative function to these completely uncharacterized genes. The resulting network comprised 1315 nodes (~56% of all the genes) and 1500 edges (Supplementary Figure 7A, Materials and Methods Section 2.13). A typical degree distribution described the network's nodes from the minimum observed degree 1 to the maximum observed degree 11 (Supplementary Figure 7B). First, we inspected the hub genes (node degree ≥ 6) from the network, since they could represent important genes for the virulence and survival of this strain during the infection process. As expected and reported from a previous work on the reference strain (FA1090) (McClure *et al.*, 2020), many of such genes were tRNA-related or encoding for ribosomal proteins (degree 6-10). However, also other interesting genes fell in such list, like

- the sialyltransferase *lst* (degree 6),
- the glycosyltransferase *pglE* (degree 6),
- the essential phosphoenolpyruvate synthase *ppsA* (degree 9),
- the central subunit of the protein translocation machinery *secY* (degree 8),
- genes involved in the energy metabolism such as *nuoL* (degree 6),
- genes involved in the stress response such as *groEL* and *lon* (degree 6-7),
- genes involved in transcription such as *rpoA*, *dksA* and *nusG* (degree 6-11),
- the glutamate dehydrogenase *gdhA* (degree 10),
- different genes encoding for amino acid transporters (degree 7-8).

3.4.4.2 THE NETWORK SHOWS DIFFERENT CO-EXPRESSED MODULES ENRICHED IN REGULATED GENES

Next, we detected a total of 28 co-expressed modules, characterized by genes showing higher connection between them than with the remaining nodes in the network, with a minimum of 12 nodes inside, and we tested for their enrichment in significant DE genes in any condition and infection models. Interesting co-expression modules enriched in regulated genes upon 15 minutes of infection were module 11 (enriched for DE genes upon 15 min of Detroit562 infection), being characterized by different outer membrane beta-barrel protein genes and amino acid ABC transporters permease or ATP-binding protein genes, module 17 (Detroit562 infection, but also enriched for DE genes at 120 mpi, no sorting, for both Detroit562 and End1 models), representing surface structure genes (examples are D-alanyl-D-alanine carboxypeptidase genes, peptidoglycan DD-metalloendopeptidase family protein genes, *mtrC*), module 27 (End1 infection), comprising different toxin-antitoxin system genes, and the small module 28 (Detroit562 infection), with few pilin genes (Supplementary Figure 4C). Some interesting modules for the regulations at 120 mpi were instead module 13 (t-UEC infection, both sorted and unsorted samples), having different immunity family protein genes and some toxin genes, module 25 (End1 infection, unsorted samples), that may be associated to a oxidative stress response (examples of genes are iron-sulfur cluster-binding protein genes, *lpdA*, *crgA*), and module 12 (End1 infection, sorted samples), being associated to iron metabolism (examples of genes are *fur*, transferrin-binding protein-like solute binding protein genes, *exbD*, TonB-dependent receptor, *tbpA*) (Supplementary Figure 4D).

3.4.4.3 THE NETWORK CAN HELP IN INFERRING FUNCTIONS FOR SOME TOP DE UNKNOWN GENES EMERGING FROM THE INFECTION TIME COURSES

Top significant DE hypothetical proteins falling in these interesting modules were for instance C7R98_RS05105 (regulated by bacteria infecting t-UEC cell line at 120 mpi, no sorting) in module 17, C7R98_RS03240 (regulated by bacteria infecting End1 and t-UEC cell lines at 120 mpi, no sorting) in module 27, and C7R98_RS02350 (regulated by bacteria infecting End1 cell line at 120 mpi, sorting) in module 12 (Figure 14A-B-C). Furthermore, different top significant DE hypothetical proteins fell inside module 13, so they may represent other molecules that bacteria use to defense themselves from other bacteria, like the ones encoded by different genes that are present in such module. These hypothetical proteins were C7R98_RS03920 (regulated by bacteria infecting End1 cell line at 120 mpi, sorting, and t-UEC cell line at 120 mpi, both unsorted and sorted samples), C7R98_RS05605 (regulated by bacteria infecting End1 cell line at 120 mpi, sorting, and t-UEC cell line at 120 mpi, both unsorted and sorted samples), C7R98_RS03925, C7R98_RS11505, C7R98_RS01805 (these three regulated by bacteria infecting t-UEC cell line at 120 mpi, no sorting) (Figure 14D). Also, two hypothetical proteins were inside a co-expressed cluster (module 1) emerging as enriched in DE genes at 120 mpi, sorting, during End1 infection, with a less clear association to functions but reporting some interesting genes, such as *secB*, *ftsN*, the ATP-dependent protease *clpX*, the guanylate kinase *gmK*, some *mafA* genes, the NAD biosynthesis gene *nadB* and the outer membrane protein assembly factor *bamA*. These two hypothetical proteins were C7R98_RS12230 (regulated by bacteria infecting Detroit562 and End1 cell lines at 120 mpi, sorting, and t-UEC cell line at 120 mpi, both unsorted and sorted samples) and C7R98_RS05650 (regulated by bacteria infecting Detroit562 and End1 cell lines at 120 mpi, sorting). Finally, another one, C7R98_RS13065 (regulated by bacteria infecting End1 cell line at 15 mpi), was observed in a quite small module (module 23) that didn't show a significant enrichment for DE genes, but presented different genes involved in amino acid metabolism (examples are *proB*, *argB*, *gdhA*, *glnA*, *ilvE* and *aroG*).

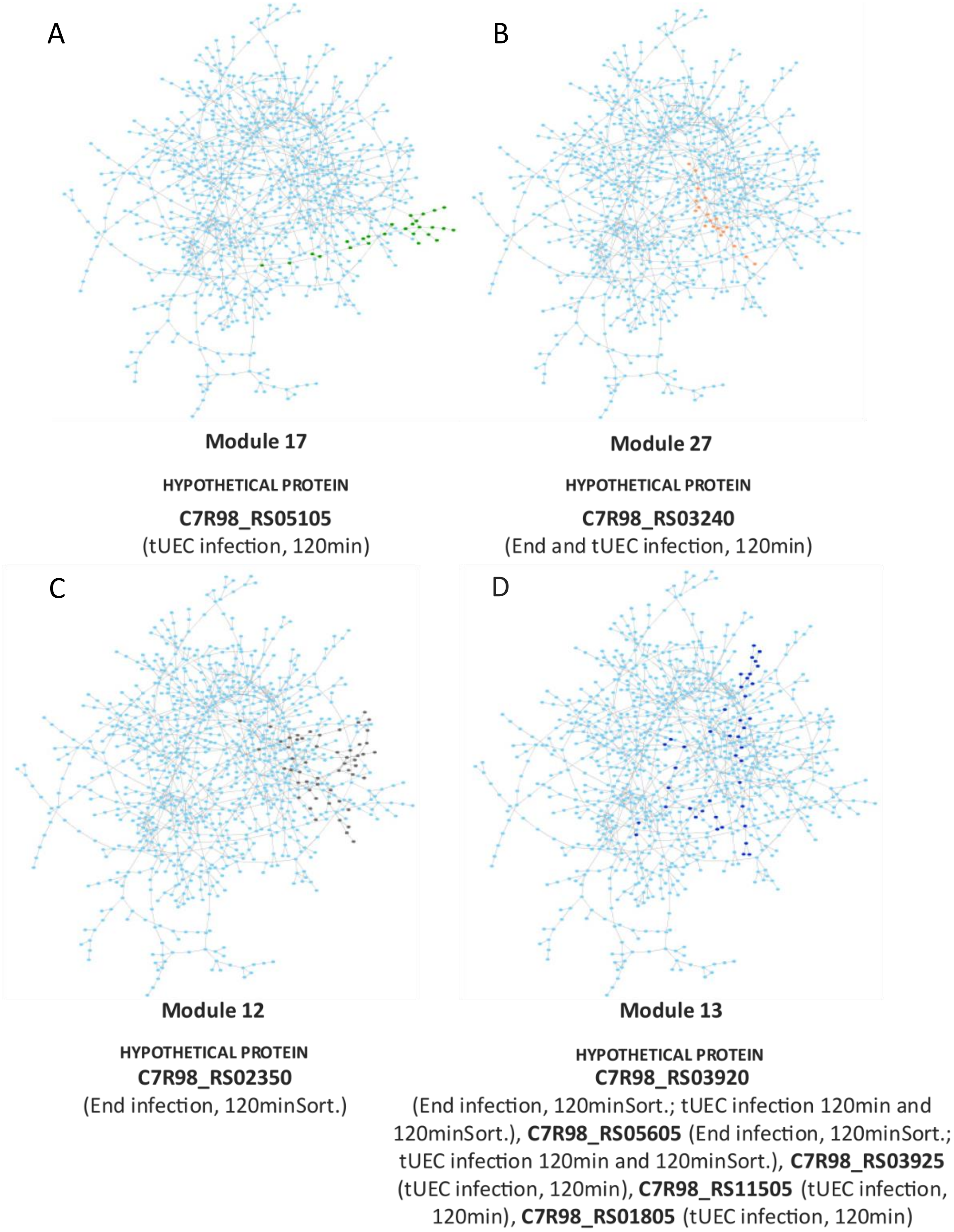
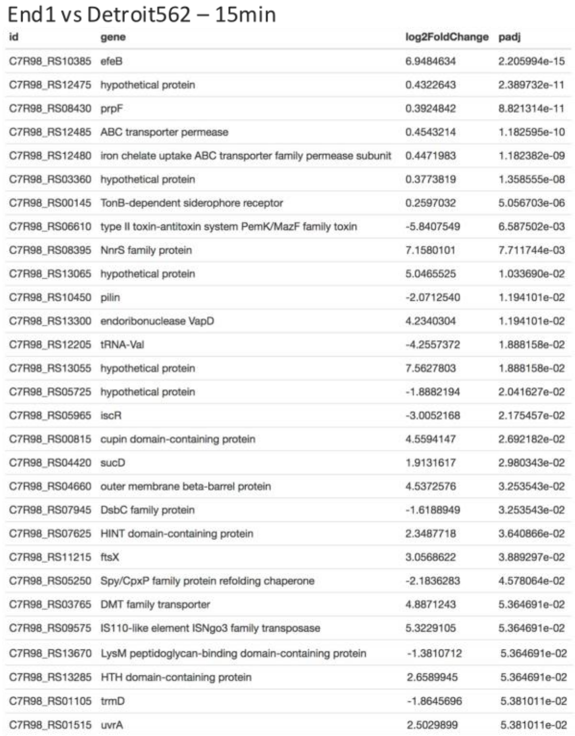

A
B
Module 17
HYPOTHETICAL PROTEIN
C7R98_RS05105
(tUEC infection, 120min)
Module 27
HYPOTHETICAL PROTEIN
C7R98_RS03240
(End and tUEC infection, 120min)
C
D
Module 12
HYPOTHETICAL PROTEIN
C7R98_RS02350
(End infection, 120minSort.)
Module 13
HYPOTHETICAL PROTEIN
C7R98_RS03920
(End infection, 120minSort.; tUEC infection 120min and 120minSort.), C7R98_RS05605 (End infection, 120minSort.; tUEC infection 120min and 120minSort.), C7R98_RS03925 (tUEC infection, 120min), C7R98_RS11505 (tUEC infection, 120min), C7R98_RS01805 (tUEC infection, 120min)

Figure 14 | Top DE hypothetical proteins falling inside bacterial co-expression modules enriched in genes regulated during the infection of the three cell models. A-D) *N. gonorrhoeae* WHO M gene co-expression network visualization by Cytoscape showing in color the co-expression module 17 (**A**), the co-expression module 27 (**B**), the co-expression module 12 (**C**) or the co-expression module 13 (**D**). At the bottom, hypothetical protein genes, emerging among the top 30 statistically significant DE genes in the indicated cell line infection and time point compared to the relative pre-infection controls, falling inside the reported co-expression modules.

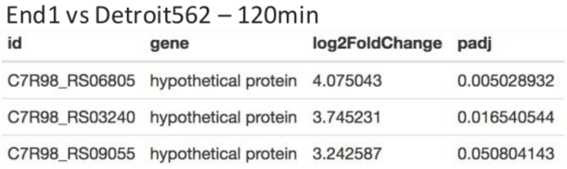
3.4.5 SOME BACTERIAL REGULATED GENES SHOW A DIFFERENT CONDITION EFFECT WHEN COMPARING UROGENITAL INFECTIONS WITH THE OROPHARYNGEAL ONE

Our design for GLM model fitting in case of pathogen samples allowed us to look for regulated genes showing a different condition effect upon infection of the different cell lines. A great portion of such signal was probably derived from the initial adaptation of bacteria to media that differed in nature and concentration of their different components, therefore for this analysis we focused only on the emerging genes that didn't appear as differentially expressed from the comparisons between the relative controls. We didn't observe genes showing such a different effect when comparing End1 and t-UEC infections, both at 15 mpi and at 120 mpi (sorted samples). On the contrary, different genes were called by the analysis if contrasting both End1 and t-UEC infection to Detroit562 one upon 15 minutes. For what concerns End1 infection compared to Detroit562 one, we observed a different condition effect for genes involved in iron uptake and metabolism such as *efeB*, an iron chelate uptake ABC transporter family permease subunit (C7R98_RS12480) and *iscR* (Table 3). Some of them showed a different condition effect also comparing the situation upon 15 minutes of infection in t-UEC cells to the one in Detroit562, together with other interesting genes, such as the peptidoglycan-related amidase *ampD*, a MacA family efflux pump subunit (C7R98_RS07955) and the resistance gene *farB*, the cell division gene *ftsL*, a factor H binding family protein (C7R98_RS00205) and, with low significance, the quorum sensing gene *luxS* (Table 3). On the contrary, upon 120 minutes of infection, internalized bacteria regulated a lower number of genes having a different condition effect in End1 and t-UEC infections compared to Detroit562 one. In the first case (End1 infection compared to Detroit562 infection), just three hypothetical proteins emerged, one of them (C7R98_RS03240) falling inside module 27 in our co-expression analysis (Table 3; Figure 14B). In the second case (t-UEC infection compared to Detroit562 infection), few genes emerged that could be associated to a more pronounced oxidative stress response in pathogen infecting this cell line compared to the situation in Detroit562 cells, like a NAD(P)-dependent oxidoreductase (C7R98_RS01330), *katA*, *lpdA* and *topA* (Table 3).



End1 vs Detroit562 – 15min

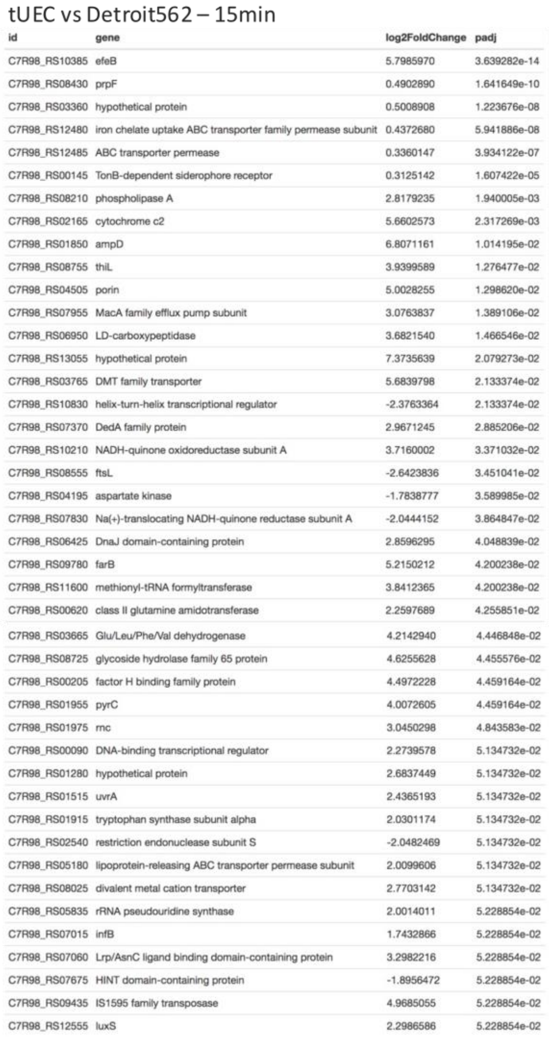
id	gene	log2FoldChange	padj
C7R98_RS10385	efeB	6.9484634	2.205994e-15
C7R98_RS12475	hypothetical protein	0.4322643	2.389732e-11
C7R98_RS08430	prpF	0.3924842	8.821314e-11
C7R98_RS12485	ABC transporter permease	0.4543214	1.182595e-10
C7R98_RS12480	iron chelate uptake ABC transporter family permease subunit	0.4471983	1.182382e-09
C7R98_RS03360	hypothetical protein	0.3773819	1.358555e-08
C7R98_RS00145	TonB-dependent siderophore receptor	0.2597032	5.056703e-06
C7R98_RS06610	type II toxin-antitoxin system PemK/MazF family toxin	-5.8407549	6.587502e-03
C7R98_RS08395	NnrS family protein	7.1580101	7.711744e-03
C7R98_RS13065	hypothetical protein	5.0465525	1.033690e-02
C7R98_RS10450	pilin	-2.0712540	1.194101e-02
C7R98_RS13300	endoribonuclease VapD	4.2340304	1.194101e-02
C7R98_RS12205	tRNA-Val	-4.2557372	1.888158e-02
C7R98_RS13055	hypothetical protein	7.5627803	1.888158e-02
C7R98_RS05725	hypothetical protein	-1.8882194	2.041627e-02
C7R98_RS05965	iscR	-3.0052168	2.175457e-02
C7R98_RS00815	cupin domain-containing protein	4.5594147	2.692182e-02
C7R98_RS04420	sucD	1.9131617	2.980343e-02
C7R98_RS04660	outer membrane beta-barrel protein	4.5372576	3.253543e-02
C7R98_RS07945	DsbC family protein	-1.6188949	3.253543e-02
C7R98_RS07625	HINT domain-containing protein	2.3487718	3.640866e-02
C7R98_RS11215	ftsX	3.0568622	3.889297e-02
C7R98_RS05250	Spy/CpxP family protein refolding chaperone	-2.1836283	4.578064e-02
C7R98_RS03765	DMT family transporter	4.8871243	5.364691e-02
C7R98_RS09575	IS110-like element ISNgo3 family transposase	5.3229105	5.364691e-02
C7R98_RS13670	LysM peptidoglycan-binding domain-containing protein	-1.3810712	5.364691e-02
C7R98_RS13285	HTH domain-containing protein	2.6589945	5.364691e-02
C7R98_RS01105	trmD	-1.8645696	5.381011e-02
C7R98_RS01515	uvrA	2.5029899	5.381011e-02



End1 vs Detroit562 – 120min

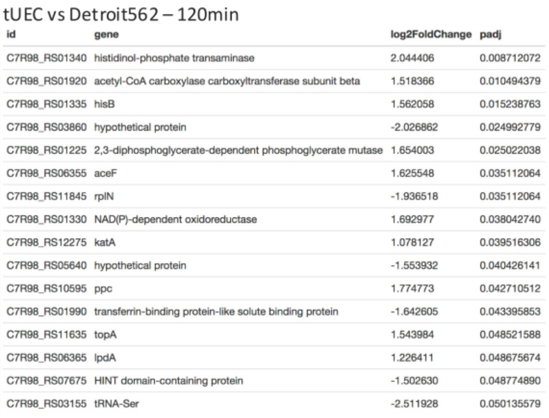
id	gene	log2FoldChange	padj
C7R98_RS06805	hypothetical protein	4.075043	0.005028932
C7R98_RS03240	hypothetical protein	3.745231	0.016540544
C7R98_RS09055	hypothetical protein	3.242587	0.050804143

Table 3 | Bacterial regulated genes showing a different condition effect during the infection of the three different cell lines. *N. gonorrhoeae* WHO M genes being characterized by a different condition effect comparing End1 infection versus Detroit562 infection upon 15 minutes, t-UEC infection versus Detroit562 infection upon 15 minutes, End1 infection versus Detroit562 infection upon 120 minutes (sorted samples) and t-UEC infection versus Detroit562 infection upon 120 minutes (sorted samples). Genes for the various comparisons were filtered for those not appearing as differentially expressed when comparing the relative pre-infection controls. Adjusted p-value for significance ≤ 0.05.



tUEC vs Detroit562 – 15min

id	gene	log2FoldChange	padj
C7R98_RS10385	efeB	5.7985970	3.639282e-14
C7R98_RS08430	prpF	0.4902890	1.641649e-10
C7R98_RS03360	hypothetical protein	0.5008908	1.223676e-08
C7R98_RS12480	iron chelate uptake ABC transporter family permease subunit	0.4372680	5.941886e-08
C7R98_RS12485	ABC transporter permease	0.3360147	3.934122e-07
C7R98_RS00145	TonB-dependent siderophore receptor	0.3125142	1.607422e-05
C7R98_RS08210	phospholipase A	2.8179235	1.940005e-03
C7R98_RS02165	cytochrome c2	5.6602573	2.317269e-03
C7R98_RS01850	ampD	6.8071161	1.014195e-02
C7R98_RS08755	thiL	3.9399589	1.276477e-02
C7R98_RS04505	porin	5.0028255	1.298620e-02
C7R98_RS07955	MacA family efflux pump subunit	3.0763837	1.389106e-02
C7R98_RS06950	LD-carboxypeptidase	3.6821540	1.466546e-02
C7R98_RS13055	hypothetical protein	7.3735639	2.079273e-02
C7R98_RS03765	DMT family transporter	5.6839798	2.133374e-02
C7R98_RS10830	helix-turn-helix transcriptional regulator	-2.3763364	2.133374e-02
C7R98_RS07370	DedA family protein	2.9671245	2.885206e-02
C7R98_RS10210	NADH-quinone oxidoreductase subunit A	3.7160002	3.371032e-02
C7R98_RS08555	ftsL	-2.6423836	3.451041e-02
C7R98_RS04195	aspartate kinase	-1.7838777	3.589985e-02
C7R98_RS07830	Na(+)-translocating NADH-quinone reductase subunit A	-2.0444152	3.864847e-02
C7R98_RS06425	DnaJ domain-containing protein	2.8596295	4.048839e-02
C7R98_RS09780	farB	5.2150212	4.200238e-02
C7R98_RS11600	methionyl-tRNA formyltransferase	3.8412365	4.200238e-02
C7R98_RS00620	class II glutamine amidotransferase	2.2597689	4.255851e-02
C7R98_RS03665	Glu/Leu/Phe/Val dehydrogenase	4.2142940	4.446848e-02
C7R98_RS08725	glycoside hydrolase family 65 protein	4.6255628	4.455576e-02
C7R98_RS00205	factor H binding family protein	4.4972228	4.459164e-02
C7R98_RS01955	pyrC	4.0072605	4.459164e-02
C7R98_RS01975	rnc	3.0450298	4.843583e-02
C7R98_RS00090	DNA-binding transcriptional regulator	2.2739578	5.134732e-02
C7R98_RS01280	hypothetical protein	2.6837449	5.134732e-02
C7R98_RS01515	uvrA	2.4365193	5.134732e-02
C7R98_RS01915	tryptophan synthase subunit alpha	2.0301174	5.134732e-02
C7R98_RS02540	restriction endonuclease subunit S	-2.0482469	5.134732e-02
C7R98_RS05180	lipoprotein-releasing ABC transporter permease subunit	2.0099606	5.134732e-02
C7R98_RS08025	divalent metal cation transporter	2.7703142	5.134732e-02
C7R98_RS05835	rRNA pseudouridine synthase	2.0014011	5.228854e-02
C7R98_RS07015	infB	1.7432866	5.228854e-02
C7R98_RS07060	Lrp/AsnC ligand binding domain-containing protein	3.2982216	5.228854e-02
C7R98_RS07675	HINT domain-containing protein	-1.8956472	5.228854e-02
C7R98_RS09435	IS1595 family transposase	4.9685055	5.228854e-02
C7R98_RS12555	luxS	2.2986586	5.228854e-02



tUEC vs Detroit562 – 120min

id	gene	log2FoldChange	padj
C7R98_RS01340	histidinol-phosphate transaminase	2.044406	0.008712072
C7R98_RS01920	acetyl-CoA carboxylase carboxyltransferase subunit beta	1.518366	0.010494379
C7R98_RS01335	hisB	1.562058	0.015238763
C7R98_RS03860	hypothetical protein	-2.026862	0.024992779
C7R98_RS01225	2,3-diphosphoglycerate-dependent phosphoglycerate mutase	1.654003	0.025022038
C7R98_RS06355	aceF	1.625548	0.035112064
C7R98_RS11845	rplN	-1.936518	0.035112064
C7R98_RS01330	NAD(P)-dependent oxidoreductase	1.692977	0.038042740
C7R98_RS12275	katA	1.078127	0.039516306
C7R98_RS05640	hypothetical protein	-1.553932	0.040426141
C7R98_RS10595	ppc	1.774773	0.042710512
C7R98_RS01990	transferrin-binding protein-like solute binding protein	-1.642605	0.043395853
C7R98_RS11635	topA	1.543984	0.048521588
C7R98_RS06365	lpdA	1.226411	0.048675674
C7R98_RS07675	HINT domain-containing protein	-1.502630	0.048774890
C7R98_RS03155	tRNA-Ser	-2.511928	0.050135579

3.5 HOST-PATHOGEN INTERACTIONS

3.5.1 THE NEW STRATEGY TO STUDY HOST-PATHOGEN INTERACTIONS APPLIED TO A PUBLICLY AVAILABLE DATA SET EXTRACTS KNOWN REGULATIONS NEEDED BY *Streptococcus pneumoniae* D39 TO START THE INTERACTION AND COLONIZATION OF HOST EPITHELIAL CELLS

3.5.1.1 REGULATIONS OF PNEUMOCOCCAL NEURAMINIDASES, COMPETENCE AND CAPSULE GENES EMERGE AS KEY ONES FOR THE PATHOGEN

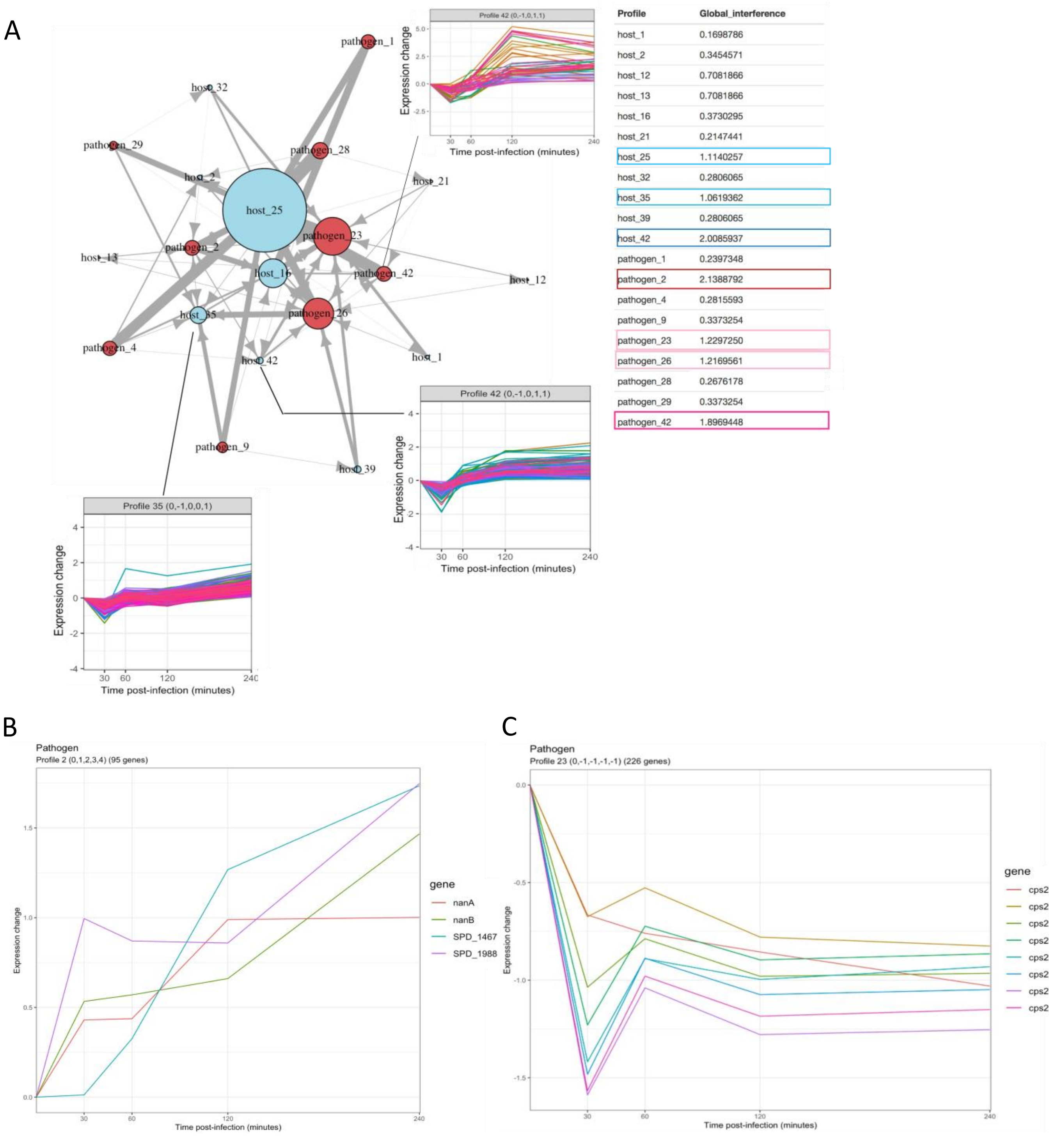
The transcriptomics characterization of the early interaction between a pathogen and its host is critical to identify genes that the bacteria require to start the infection and that can represent optimal candidates for the development of preventive or therapeutic strategies. The infection process originates from the interplay between an infecting agent and the host, fully captured at the transcriptional level by the dual RNA-Seq technique, but still poorly considered from an analytical point of view. We proposed a new strategy that takes in consideration the host-pathogen interaction in its whole complexity and try to extract those regulations that make it possible. This could be helpful in prioritizing relevant genes, especially when poorly characterized strains, such as *N. gonorrhoeae* WHO M one, are used. First of all, we tested our method on a publicly available dual RNA-Seq data set concerning, similarly to our study, an early infection, but investigating the pathogenesis of a well-studied bacterial strain, *Streptococcus pneumoniae* D39 (Aprianto *et al.*, 2016). Particularly, in this study, the authors provided a dual transcriptomics overview of the pneumococcal colonization of human alveolar epithelial cells (A549) at 30, 60, 120 and 240 minutes post-infection. During the infection process, the capsule of this pathogen is dynamically regulated to accompany the various stages. Particularly, bacteria undergo capsule shedding to favor the adhesion to the host cells. Therefore, the authors compared wild-type encapsulated bacteria with unencapsulated *Δcps2E* mutants, which adhere more readily to the epithelial cells. Considering the wild-type infection, we found a total of 11 significantly enriched temporal gene expression patterns or clusters for the host and 9 for the pathogen (from the 44 selected representative patterns) (Supplementary Figure 8). Applying our strategy, these clusters became the nodes of the network representing the transcriptional influences driving the early host-pathogen crosstalk that is depicted in Figure 13A. From the global interference calculation, pathogen profile 2 (0, 1, 2, 3, 4; 95 genes), showing an ever-increasing up-regulation along the time points, emerged as the node characterized by the highest value (Figure 15A and 15B). Top enriched KEGG pathways for the genes inside this cluster were represented by Other glycan degradation (adj. p-value 1.33e-03) and Two-component system (adj. p-value 3.24e-03). Basically, the genes responsible for these enrichments were the pneumococcal neuraminidases (mainly *nanA* and *nanB*), known factors required by this pathogen to start the interaction by cleaving terminal sialic acid residues from host glycoconjugates and exposing potential binding receptors, together with different competence genes (*comX1, comW, comA, comB, comE, comD, comC1*), that also could be involved, directly or indirectly, in regulating adhesion and colonization (Figure 15B). Other pathogen regulations emerging as basal ones for the early interaction were described by pathogen profile 23 (0, -1, -1, -1 -1; 226 genes), characterized by a quite constant down-regulation of many capsule genes (*cps2B, cps2G, cps2H, cps2I, cps2J, cps2K, cps2L, cps2P*) (Figure 15C), and pathogen profile 42 (0, -1, 0, 1, 1; 51 genes), showing a later up-regulation of other competence genes (such as *comEC, comM, cglA*) (Figure 15A).

3.5.1.2 ACTIN FILAMENT-BASED AND GLYCOSAMINOGLYCAN CATABOLIC PROCESSES EMERGE AS BASAL REGULATED PATHWAYS FOR THE HOST

For the host, the profiles emerging for being characterized by the highest values of global interference were the profile 42 (0, -1, 0, 1, 1; 118 genes), enriched in genes involved in intracellular signal transduction, cell adhesion regulation, stress resistance and actin organization, profile 25 (0, 0, 0, 0, 1; 992 genes), linked to cell adhesion and being also the node with the highest total node strength in the network, and profile 35 (0, -1, 0, 0, 1; 252 genes), related to glycosaminoglycan catabolic process and actin filament-based process. So, all profiles probably describing morphological events in the host cells triggered by the first encounters with the pathogen (Figure 15A). Furthermore, host profile 35 was also the profile showing the highest decrease in betweenness centrality upon pathogen node 2 removal, suggesting in some way a link between the glycans dynamics described by this profile and the sialidase activity of *S. pneumoniae* neuraminidases regulated following the pattern indicated by pathogen profile 2.

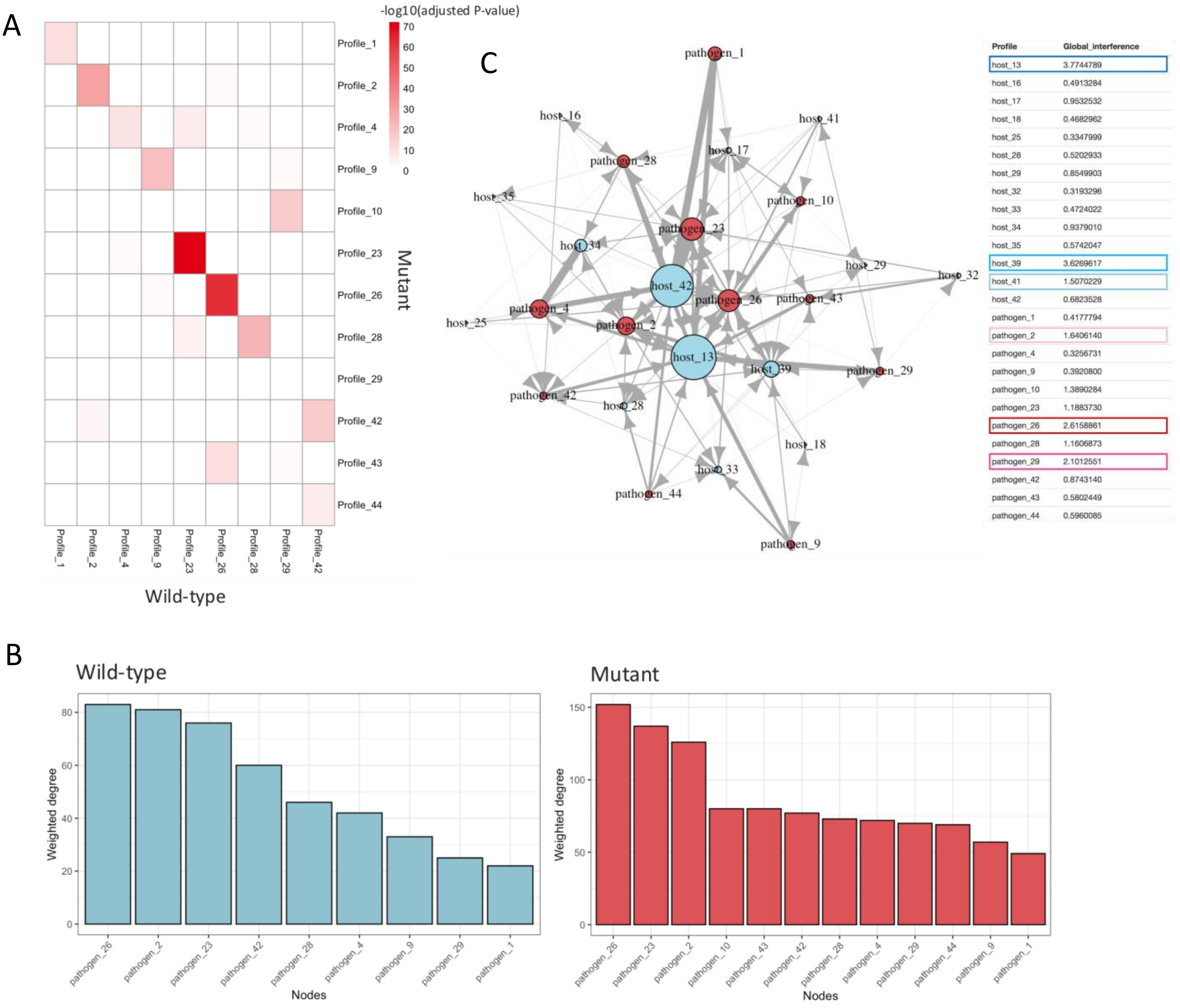
3.5.1.3 COMMON REGULATIONS SHOWING THE SAME INTERPLAY WITH THE HOST RESPONSE CHARACTERIZE INFECTING WILD-TYPE AND MUTANT PATHOGENS

The analysis of the data from the mutant infection reported more enriched expression patterns for both the host (14) and the pathogen (12) and a more intertwined network (Supplementary Figure 9; Figure 16C), in line with the idea that the mutation can accelerate the infection process and the crosstalk between the two organisms. The highest differences of course were about the involvement of the host and its response, while all the significantly enriched expression patterns that wild-type bacteria showed before emerged again for the mutant ones. As expected, for bacteria there was a significant overlap between the genes in the wild-type clusters and those in the corresponding ones in the mutant, with the only exception of wild-type profile 29 (Figure 16A). Interestingly, the same profiles, that were also among the pathogen profiles with the highest value of global interference in the wild-type analysis, emerged as hubs in both the pathogen projections originating from the bipartite graphs of the two conditions (wild-type and mutant infection), suggesting common underlying transcriptional events sharing the same interplay with the host response (Figure 16B). However, the centrality in terms of global interference in case of the mutant infection was gained by a metabolic crosstalk, supporting the idea that, because of the mutation, we were not really looking anymore at the early stages of the infection. Indeed, the profiles characterized by the highest value of global interference were host profiles (profiles 13 and 39) mainly involved in cellular metabolic process and nitrogen compound metabolic process (profile 13), and TORC1 signaling and leucine sensing (profile 39), while for the pathogen the emerging profiles (profiles 26, that started to emerge also from the wild-type analysis, and 29) were linked to amino acid, sugars and secondary metabolisms, as well as antimicrobial compounds biosynthesis (Figure 16C). Interestingly, host profile 13 showed dependence from pathogen profile 10, for which among the top enriched KEGG pathways we reported PTS system, sugars metabolism and arginine biosynthesis.



Profile	Global_interference
host_1	0.1698786
host_2	0.3454571
host_12	0.7081866
host_13	0.7081866
host_16	0.3730295
host_21	0.2147441
host_25	1.1140257
host_32	0.2806065
host_35	1.0619362
host_39	0.2806065
host_42	2.0085937
pathogen_1	0.2397348
pathogen_2	2.1388792
pathogen_4	0.2815593
pathogen_9	0.3373254
pathogen_23	1.2297250
pathogen_26	1.2169561
pathogen_28	0.2676178
pathogen_29	0.3373254
pathogen_42	1.8969448

Figure 15 | Study of host-pathogen interactions on a publicly available dual RNA-Seq data set concerning host epithelial cells infected by wild-type *Streptococcus pneumoniae* D39. A) Emerging network of transcriptional influences occurring between host and pathogen significant temporal patterns of gene expression, with some highlighted profiles and the global interference value associated to each node. In the network, host profiles are represented in blue, pathogen ones in red. Edge's size is proportional to the edge weight, based on -ln(Bonferroni-adjusted p-value) of the interaction from the permutation test. Node's size is proportional to the total node strength (sum of the weights on in-coming and out-coming edges). Adjusted p-value to define an edge from the permutation test ≤ 0.05. In the highlighted profiles, each line corresponds to a gene falling inside that regulation pattern. **B)** Temporal expression pattern for bacterial genes associated to the top KEGG pathway (Other glycan degradation) emerging as significantly enriched for pathogen profile 2. **C)** Temporal expression pattern for bacterial capsule genes falling inside pathogen profile 23.



Profile	Global_interference
host_13	3.7744789
host_16	0.4913284
host_17	0.9532532
host_18	0.4682962
host_25	0.3347999
host_28	0.5202933
host_29	0.8549903
host_32	0.3193296
host_33	0.4724022
host_34	0.9379010
host_35	0.5742047
host_39	3.6269617
host_41	1.5070229
host_42	0.6823528
pathogen_1	0.4177794
pathogen_2	1.6406140
pathogen_4	0.3256731
pathogen_9	0.3920800
pathogen_10	1.3890284
pathogen_23	1.1883730
pathogen_26	2.6158861
pathogen_28	1.1606873
pathogen_29	2.1012551
pathogen_42	0.8743140
pathogen_43	0.5802449
pathogen_44	0.5960085

Figure 16 | Comparison of the studies of host-pathogen interactions on the publicly available dual RNA-Seq data set concerning host epithelial cells infected by both wild-type and mutant *Streptococcus pneumoniae* D39. A) Heatmap reporting the -log10(adjusted p-value) from the hypergeometric test evaluating the intersection between genes in pathogen significant profiles emerging from the analysis on the wild-type data and those from the analysis on the mutant one. **B)** Bar plots reporting the weighted degree for all pathogen profiles in the pathogen projections from the bipartite graph describing wild type data and the one in **C** describing the mutant data. **C)** Emerging network of transcriptional influences occurring between host and pathogen significant temporal patterns of gene expression when considering the mutant infection, with the global interference value associated to each node. In the network, host profiles are represented in blue, pathogen ones in red. Edge's size is proportional to the edge weight, based on -ln(Bonferroni-adjusted p-value) of the interaction from the permutation test. Node's size is proportional to the total node strength (sum of the weights on in-coming and out-coming edges). Adjusted p-value to define an edge from the permutation test ≤ 0.05.

3.5.2 A SIMILAR SIGNAL CONCERNING KEY HOST AND PATHOGEN REGULATIONS DURING EARLY INFECTION BY *N. gonorrhoeae* WHO M EMERGE FROM ALL THE INFECTION MODELS

3.5.2.1 KEY REGULATIONS FOR THE HOST IN DETROIT562 EARLY INFECTION CONCERN CYTOSKELETON REARRANGEMENTS AND VESICULAR DYNAMICS

The clustering for Detroit562 infection data returned 12 statistically significant temporal expression patterns for the host (between 72 and 637 genes per cluster) and 8 statistically significant temporal expression patterns for the pathogen (between 69 and 148 genes per cluster) out of 48 possible ones (Supplementary Figure 10). A host stress response leading to a pro-inflammatory condition already begun upon 15 minutes of infection, as reported by profile 43 (0, 1, 4; 278 genes), whose genes showed enrichment for transduction pathways like JNK cascade (adj. p-value 2.94e-5) and stress-activated MAPK cascade (adj. p-value 3.13e-4), and profile 40 (0, 1, 3; 82 genes), showing among the top significantly enriched GO terms epiboly involved in wound healing (adj. p-value 0.000523) and arachidonic acid metabolic process (adj. p-value 0.000737) (Supplementary Figure 10A). Upon 120 minutes, such condition in cells that internalized the bacteria strongly evolved, as reported by a clear up-regulation of genes in cluster 30 (0, 0, 1; 635 genes) related to cytokine-mediated signaling pathways (adj. p-value 0.000000165), response to cytokine (adj. p-value 0.000000231) and regulation of signaling receptor activity (adj. p-value 0.000000951), and in the partially-overlapping cluster 39 (0, 0, 3; 637 genes) associated to terms like response to cytokines (adj. p-value 1.85e-9), regulation of signaling receptor activity (adj. p-value 1.98e-6), immune response (adj. p-value 2.27e-5), regulation of response to stress (adj. p-value 8.55e-5), defense response (adj. p-value 2.76e-4) and NIK/NF-kappaB signaling (adj. p-value 4.27e-4) (Supplementary Figure 10A). From the host picture, a strong effort that infected cells put into translational control upon 120 minutes of infection was also clear, indeed we reported two clusters, profile 29 (0, -1, 1; 146 genes) and profile 34 (0, -1, 2; 540 genes), being implicated in ncRNA processing (adj. p-value 2.43e-3) and positive regulation of translation (adj. p-value 2.61e-3), and in ribosome biogenesis (adj. p-value 1.43e- 6), translation (adj. p-value 4.33e- 6) and co-translational protein targeting to membrane (adj. p-value 8.03e- 6), respectively (Supplementary Figure 10A). Furthermore, we also observed a strong downregulation of genes involved in positive regulation of nuclear cell cycle DNA replication (adj. p-value 6.63e-4), when looking at the enriched profile 9 (0, -1, -3; 72 genes) (Supplementary Figure 10A). Applying our new strategy, we obtained the well-intertwined graph represented in Figure 17A. The profiles emerging with the highest value of global interference were host profiles 19 (0, 0, -1; 545 genes) and 20 (0, 1, -1; 151 genes) (Figure 17A-B). The first one was strongly related to cytoskeleton organization (adj. p-value 5.00e-9) (examples of genes included *KRT13*, *KRT19*, *CAPZB*, *WASF2*, *DCTN1*, *ACTN4*, *BIN1*, *NCKIPSD*, *DAG1*, *ELMO1*, *LIMK1*, *SORBS3*, *KIF18B*, *KIF24*, *TLN1*, *CKAP2*, *PDPK1*, *PIP5K1C…*), probably describing the host cytoskeletal rearrangements occurring upon the interaction with the pathogen, and the second one to lipid metabolic process (adj. p-value 8.50e-5) and positive regulation of phosphatidylinositol 3-kinase activity (adj. p-value 8.17e-4) (examples of genes included *PIK3C2B*, *FGFR3*, *VAV3*, *ATG9A*, *ATG9B*, *ATP13A2*, *NAGA*, *KIT*, *PSAPL1*, *ST3GAL5…*), so to an intracellular vesicular trafficking that could accompany bacterial internalization (Figure 17B-C).

3.5.2.2 KEY REGULATIONS FOR THE PATHOGEN IN DETROIT562 EARLY INFECTION CONCERN BOTH GENES BELONGING TO KNOWN PATHOGENETIC CLASSES AND UNCHARACTERIZED ONES

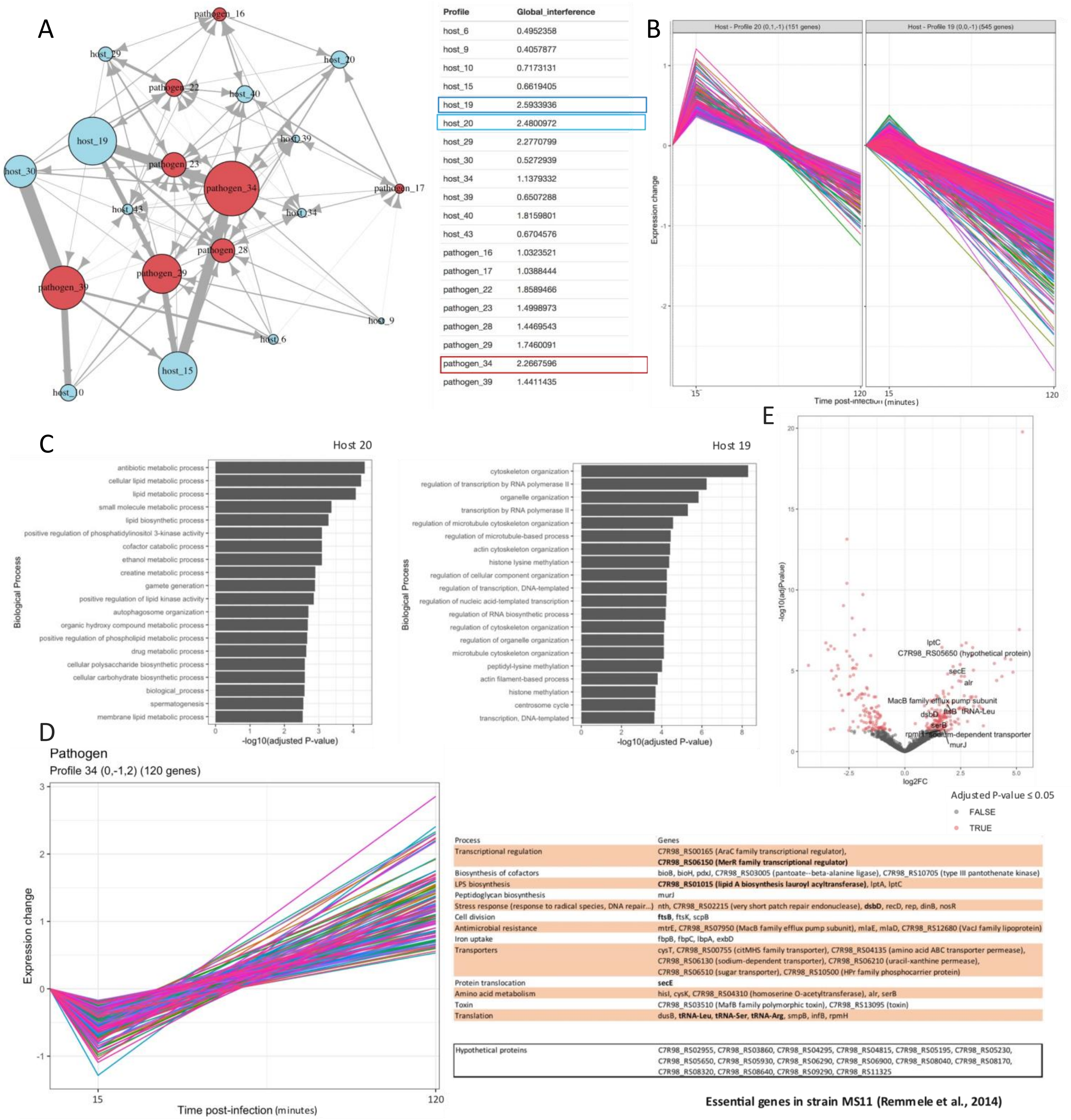
For what concerns the pathogen, the profile showing the highest value of global interference was pathogen profile 34 (0, -1, 2; 120 genes), reporting the regulation of different genes involved in known pathogenetic classes or in bacterial survival to the changing environment (Figure 17D). For example, we observed some genes involved in LOS or peptidoglycan biosynthesis, such as C7R98_RS01015 (lipid A biosynthesis lauroyl acyltransferase), *lptA*, *lptC* and *murJ*, and in antimicrobial resistance, such as *mtrE*, C7R98_RS07950 (MacB family efflux pump subunit) and the membrane asymmetry maintenance proteins *mlaE* and *mlaD*. Furthermore, we reported the regulation of different genes involved in stress response, so in the response to radical species or in DNA repair (*nth*, *dsbD*, *recD*, *rep*, *dinB*, *nosR*), and in cell division (*ftsB*, *ftsK*, *scpB*). Importantly, also genes involved in iron uptake (*fbpB*, *fbpC*, *lbpA*, *exbD*), protein translocation (*secE*) and coding many transporters fell inside this cluster. Finally, a total of 16 hypothetical proteins were also co-regulated with the aforementioned genes, one of which (C7R98_RS05650) being also among the top significantly up-regulated genes when considering internalized bacteria compared to non-infecting bacteria (Figure 17D-E). Importantly, the genes encoding for pilins and various outer membrane beta-barrel proteins that we observed as differentially expressed upon 15 minutes of infection (Figure 11A), because of their different extent of up-regulation at 15 minutes and down- or no regulation at 120 minutes, didn’t emerged as homogenously regulated in a unique cluster, but divided in multiple ones. Particularly, many pilins fell inside

- cluster 21 (0, 2, 1; 31 genes),
- cluster 25 (0, 1, 0; 27 genes),
- cluster 27 (0, 3, 0; 25 genes) and
- cluster 33 (0, 3, 1; 23 genes),

while many outer membrane beta-barrel proteins fell inside

- cluster 32 (0, 2, 1; 21 genes) and
- cluster 38 (0, 3, 2; 37 genes).

Unfortunately, all these clusters collected very few genes and didn’t pass the significance threshold of the permutation test, therefore they were not among the pathogen significant profiles that were tested for host-pathogen interactions and used to build the graph. So, from the pathogen point of view, our network was probably capturing as basal regulations the ones characterizing a step ahead, probably linked to essential genes that the pathogen regulates to survive in an intracellular context.



Profile	Global_interference
host_6	0.4952358
host_9	0.4057877
host_10	0.7173131
host_15	0.6619405
host_19	2.5933936
host_20	2.4800972
host_29	2.2770799
host_30	0.5272939
host_34	1.1379332
host_39	0.6507288
host_40	1.8159801
host_43	0.6704576
pathogen_16	1.0323521
pathogen_17	1.0388444
pathogen_22	1.8589466
pathogen_23	1.4998973
pathogen_28	1.4469543
pathogen_29	1.7460091
pathogen_34	2.2667596
pathogen_39	1.4411435

Process	Genes
Transcriptional regulation	C7R98_RS00165 (AraC family transcriptional regulator), **C7R98_RS06150 (MerR family transcriptional regulator)**
Biosynthesis of cofactors	bioB, bioH, pdxJ, C7R98_RS03005 (pantoate--beta-alanine ligase), C7R98_RS10705 (type III pantothenate kinase)
LPS biosynthesis	**C7R98_RS01015 (lipid A biosynthesis lauroyl acyltransferase)**, lptA, lptC
Peptidoglycan biosynthesis	murJ
Stress response (response to radical species, DNA repair...)	nth, C7R98_RS02215 (very short patch repair endonuclease), **dsbD**, recD, rep, dinB, nosR
Cell division	**ftsB**, ftsK, scpB
Antimicrobial resistance	mtrE, C7R98_RS07950 (MacB family efflux pump subunit), mlaE, mlaD, C7R98_RS12680 (VacJ family lipoprotein)
Iron uptake	fbpB, fbpC, lbpA, exbD
Transporters	cysT, C7R98_RS00755 (citMHS family transporter), C7R98_RS04135 (amino acid ABC transporter permease), C7R98_RS06130 (sodium-dependent transporter), C7R98_RS06210 (uracil-xanthine permease), C7R98_RS06510 (sugar transporter), C7R98_RS10500 (HPr family phosphocarrier protein)
Protein translocation	**secE**
Amino acid metabolism	hisI, cysK, C7R98_RS04310 (homoserine O-acetyltransferase), alr, serB
Toxin	C7R98_RS03510 (MafB family polymorphic toxin), C7R98_RS13095 (toxin)
Translation	dusB, **tRNA-Leu**, **tRNA-Ser**, **tRNA-Arg**, smpB, infB, rpmH

Hypothetical proteins	C7R98_RS02955, C7R98_RS03860, C7R98_RS04295, C7R98_RS04815, C7R98_RS05195, C7R98_RS05230, C7R98_RS05650, C7R98_RS05930, C7R98_RS06290, C7R98_RS06900, C7R98_RS08040, C7R98_RS08170, C7R98_RS08320, C7R98_RS08640, C7R98_RS09290, C7R98_RS11325

Essential genes in strain MS11 (Remmele et al., 2014)

Figure 17 | Study of host-pathogen interactions on the data about Detroit562 early infection by *N. gonorrhoeae* WHO M. A) Emerging network of transcriptional influences occurring between host and pathogen significant temporal patterns of gene expression when considering Detroit562 infection, with the global interference value associated to each node. In the network, host profiles are represented in blue, pathogen ones in red. Edge's size is proportional to the edge weight, based on -ln(Bonferroni-adjusted p-value) of the interaction from the permutation test. Node's size is proportional to the total node strength (sum of the weights on in-coming and out-coming edges). Adjusted p-value to define an edge from the permutation test ≤ 0.05. **B)** Temporal expression patterns describing host profiles 20 and 19, each line corresponds to a gene falling inside that regulation pattern. **C)** Hypergeometric tests on GO biological processes for the genes falling inside host profiles 20 and 19. Adjusted p-value to call the significance ≤ 0.05. **D)** Temporal expression pattern describing pathogen profile 34 and some associated interesting genes grouped by pathways or categories. Each line in the profile corresponds to a gene falling inside the regulation pattern. **E)** Volcano plot from DE analysis on Detroit562 infection comparing the 120mpi time point (internalized bacteria) versus pre-infection controls. Dots for genes are colored accordingly to their significance from DE analysis. Gene reported in **D** and also emerging from DE test are indicated in the volcano plot. Adjusted p-value to call DE genes ≤ 0.05.

3.5.2.3 KEY REGULATIONS FOR THE HOST IN END1 EARLY INFECTION CONCERN CYTOSKELETAL DYNAMICS AND STRESS RESPONSES

The data about the infection of End1 cell line returned 12 statistically significant clusters for the host (between 74 and 284 genes per profile) and only 4 statistically significant clusters for the pathogen (between 56 and 75 genes per profile) out of 48 possible ones (Supplementary Figure 11). Contrary to what described for Detroit562, as already deduced from differential expression analyses, looking at enriched profiles 40 (0, 1, 3; 89 genes) and 43 (0, 1, 4; 181 genes) a host defense response was already evident upon 15 minutes of infection. Associated biological processes to the first mentioned profile comprehended cytokine-mediated signaling pathway (adj. p-value 0.00000000826), NIK/NF-kappaB signaling (adj. p-value 0.0000000248), regulation of immune system process (adj. p-value 0.0000000353) and response to bacterium (adj. p-value 0.000000463), while regulation of CD4-positive, alpha-beta T cell activation (adj. p-value 1.17e-4), negative regulation of interleukin-1 secretion (adj. p-value 1.86e-4) and interleukin-6 production (adj. p-value 7.71e-4) were among the top significantly enriched processes related to the second one. Also, some profiles characterized by an up-regulation upon 120 minutes of genes involved in nucleic acid and nitrogen compound metabolic processes emerged. In Figure 18A is represented the resulting graph from the application of our method. Again, the host profiles characterized by having the highest values of global interference were profiles 19 (0, 0, -1; 144 genes) and 20 (0, 1, -1; 165 genes), the first one mainly associated to cell division (adj. p-value 0.00000000835) and cytoskeleton (adj. p-value 0.00000201) (examples of genes included *KANK4*, *TCHH*, *KIF20A*, *KIF18B*, *ADD2*, *MSN*, *ACTR1A*…), the second one to stress response and regulation of MAP kinase activity (adj. p-value 1.21e-2) (examples of genes included *MAP3K12*, *MAP4K2*, *DUSP18*, *RGS3*, *TP73*, *GRHL3*…) (Figure 18A-B).

3.5.2.4 KEY REGULATIONS FOR THE PATHOGEN IN END1 EARLY INFECTION CONCERN INVASION AND MEMBRANE DYNAMICS GENES

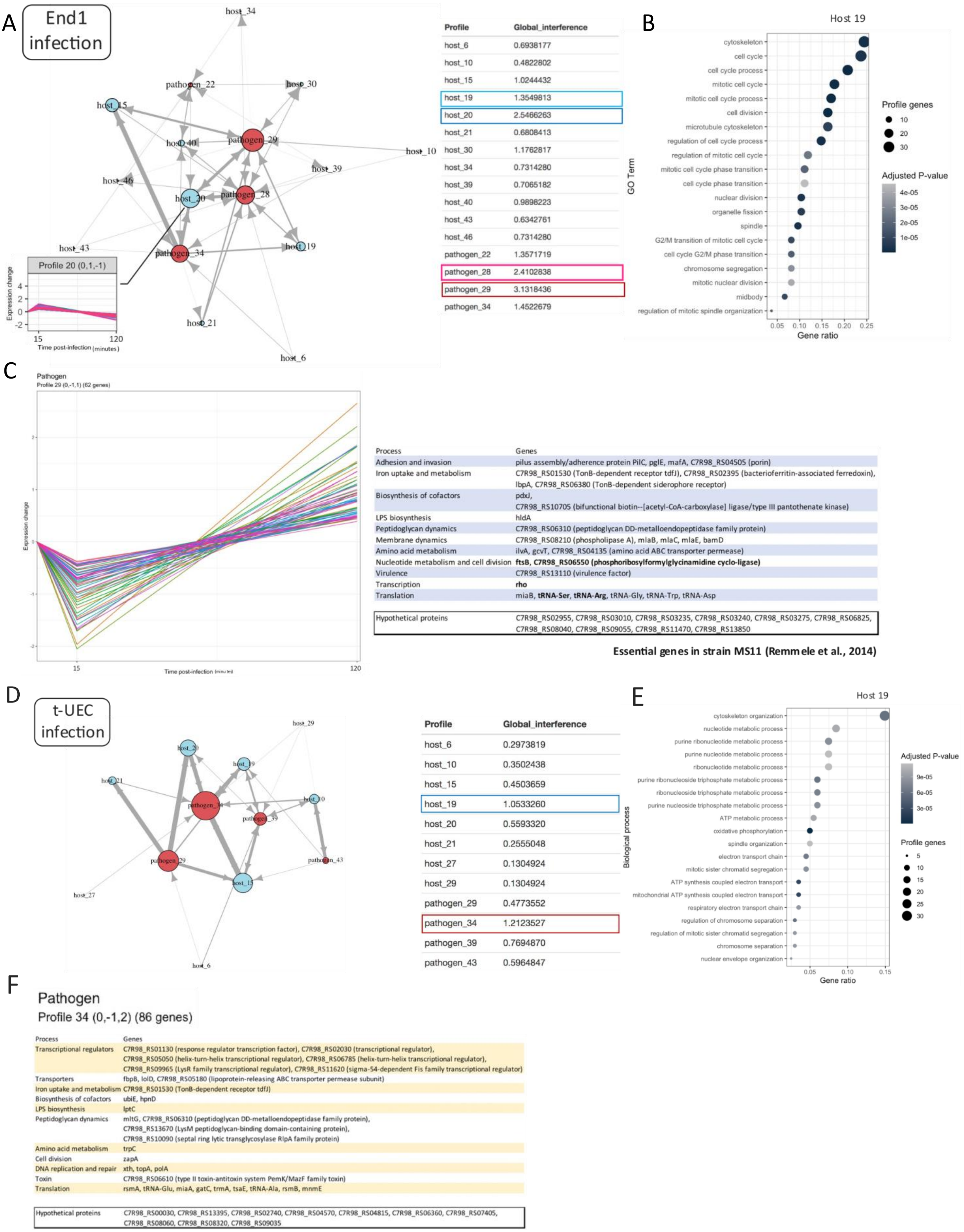
Considering the pathogen, the emerging profile in terms of global interference value was pathogen profile 29 (0, -1, 1; 62 genes) (Figure 18A-C). Inside this profile fell some genes involved in pathogen adhesion and invasion (*pilC*, *pglE*, *mafA*, the porin C7R98_RS04505) and we observed different genes involved in iron uptake and metabolism, such as *tdfJ*, a bacterioferritin-associated ferredoxin (C7R98_RS02395), *lbpA* and a TonB-dependent siderophore receptor (C7R98_RS06380). Again, this profile reported the up-regulation in intracellular bacteria of genes involved in nucleotide metabolism and cell division (*ftsB* and the phosphoribosylformylglycinamidine cyclo-ligase C7R98_RS06550) and in membrane dynamics, like the phospholipase A C7R98_RS08210, *mlaB, mlaC, mlaE* and *bamD,* as well as in peptidoglycan or LOS biosynthesis (the peptidoglycan DD-metalloendopeptidase family protein C7R98_RS06310 and *hldA*). Interestingly, also *rho* and an unknown virulence factor (C7R98_RS13110) were inside this profile. Finally, a total of 10 hypothetical proteins were regulated following this pattern (Figure 18C).

3.5.2.5 KEY REGULATIONS FOR THE HOST IN t-UEC EARLY INFECTION CONCERN CYTOSKELETON ORGANIZATION

Clustering the data about t-UEC infection we observed 9 statistically significant clusters for the host (between 150 and 399 genes per profile) and again only 4 statistically significant clusters for the pathogen (between 49 and 86 genes per profile) out of 48 possible ones (Supplementary Figure 12). For the host, we reported some profiles linked to a down-regulation of genes involved in the mitotic process for cells that internalized the bacteria, such as profile 20 (0, 1, -1; 289 genes), for which we obtained among the top significantly enriched biological processes mitotic chromosome condensation (adj. p-value 0.00214), and profile 10 (0, 0, -3; 215 genes), for which we obtained spindle organization (adj. p-value 9.34e-5). Others were particularly linked to post-transcriptional and post-translational processes, such as profile 29 (0, -1, 1; 199 genes), reporting an enrichment in RNA processing (adj. p-value 1.24e-9) and in ribonucleoprotein complex subunit organization (adj. p-value 3.59e-8), and profile 34 (0, -1, 2; 300 genes), reporting an enrichment in protein targeting to ER (adj. p-value 3.69e-9) and protein targeting to membrane (adj. p-value 6.40e-9). Importantly, no statistically significant profiles linked to an up-regulation of pro-inflammatory genes clearly emerged. Only profile 27 (0, 3, 0; 291 genes) reported upon 15 minutes an up-regulation of genes involved in positive regulation of B cell activation (adj. p-value 3.14e-4), such as *IL7*, *TLR4* and *NOD2*, while profile 6 (0, -1, -4; 150 genes) reported a growing down-regulation of genes involved in the response to tumor necrosis factor (adj. p-value 1.49e-3). So, this confirmed the idea that a strong defense response by these cells was not mounted, at least in the considered time window. From the application of our strategy, it was also clear that a strong interaction between the two organisms was not already present after 120 minutes of infection, since we didn't obtain a well-interconnected graph (Figure 18D). However, again the host profile characterized by the highest value of global interference was the host profile 19 (0, 0, -1; 223 genes), showing an association to oxidative phosphorylation (adj. p-value 6.58e-7) and cytoskeleton organization (adj. p-value 5.98e-5) (Figure 18D-E). Examples of genes included in this profile were *ACTN4*, *BIN1*, *LIMK1*, *KRT5*, *NCKAP5L*, *KIF18B*, *CORO7* and *TUBGCP6*.

3.5.2.6 KEY REGULATIONS FOR THE PATHOGEN IN t-UEC EARLY INFECTION CONCERN DIFFERENT TRANSCRIPTIONAL REGULATORS

From the pathogen's point of view, the emerging profile in terms of global interference was profile 34 (0, -1, 2; 86 genes), characterized by different transcriptional regulators, such as C7R98_RS01130 (response regulator transcription factor), C7R98_RS02030 (transcriptional regulator), C7R98_RS05050 and C7R98_RS06785 (helix-turn-helix transcriptional regulators), C7R98_RS09965 (LysR family transcriptional regulator) and C7R98_RS11620 (sigma-54-dependent Fis family transcriptional regulator), some transporters (*fbpB*, *lolD*, the lipoprotein-releasing ABC transporter permease subunit gene C7R98_RS05180), the TonB-dependent receptor *tdfJ*, different genes involved in LOS biosynthesis or peptidoglycan dynamics (*lptC*, *mltG*, the peptidoglycan DD-metalloendopeptidase family protein gene C7R98_RS06310, the LysM peptidoglycan-binding domain-containing protein gene C7R98_RS13670, the septal ring lytic transglycosylase RlpA family protein gene C7R98_RS10090) and genes involved in DNA replication and repair (*xth*, *topA* and *polA*) (Figure 18F). Finally, also in this case different hypothetical proteins (10 in total) fell inside the profile (Figure 18F).



Profile	Global_interference
host_6	0.6938177
host_10	0.4822802
host_15	1.0244432
host_19	1.3549813
host_20	2.5466263
host_21	0.6808413
host_30	1.1762817
host_34	0.7314280
host_39	0.7065182
host_40	0.9898223
host_43	0.6342761
host_46	0.7314280
pathogen_22	1.3571719
pathogen_28	2.4102838
pathogen_29	3.1318436
pathogen_34	1.4522679

Process	Genes
Adhesion and invasion	pilus assembly/adherence protein PilC, pglE, mafA, C7R98_RS04505 (porin)
Iron uptake and metabolism	C7R98_RS01530 (TonB-dependent receptor tdfJ), C7R98_RS02395 (bacterioferritin-associated ferredoxin), lbpA, C7R98_RS06380 (TonB-dependent siderophore receptor)
Biosynthesis of cofactors	pdxJ, C7R98_RS10705 (bifunctional biotin--[acetyl-CoA-carboxylase] ligase/type III pantothenate kinase)
LPS biosynthesis	hldA
Peptidoglycan dynamics	C7R98_RS06310 (peptidoglycan DD-metalloendopeptidase family protein)
Membrane dynamics	C7R98_RS08210 (phospholipase A), mlaB, mlaC, mlaE, bamD
Amino acid metabolism	ilvA, gcvT, C7R98_RS04135 (amino acid ABC transporter permease)
Nucleotide metabolism and cell division	**ftsB**, **C7R98_RS06550 (phosphoribosylformylglycinamidine cyclo-ligase)**
Virulence	C7R98_RS13110 (virulence factor)
Transcription	**rho**
Translation	miaB, **tRNA-Ser**, **tRNA-Arg**, tRNA-Gly, tRNA-Trp, tRNA-Asp
Hypothetical proteins	C7R98_RS02955, C7R98_RS03010, C7R98_RS03235, C7R98_RS03240, C7R98_RS03275, C7R98_RS06825, C7R98_RS08040, C7R98_RS09055, C7R98_RS11470, C7R98_RS13850

Essential genes in strain MS11 (Remmele et al., 2014)

Profile	Global_interference
host_6	0.2973819
host_10	0.3502438
host_15	0.4503659
host_19	1.0533260
host_20	0.5593320
host_21	0.2555048
host_27	0.1304924
host_29	0.1304924
pathogen_29	0.4773552
pathogen_34	1.2123527
pathogen_39	0.7694870
pathogen_43	0.5964847

Pathogen

Profile 34 (0,-1,2) (86 genes)

Process	Genes
Transcriptional regulators	C7R98_RS01130 (response regulator transcription factor), C7R98_RS02030 (transcriptional regulator), C7R98_RS05050 (helix-turn-helix transcriptional regulator), C7R98_RS06785 (helix-turn-helix transcriptional regulator), C7R98_RS09965 (LysR family transcriptional regulator), C7R98_RS11620 (sigma-54-dependent Fis family transcriptional regulator)
Transporters	fbpB, lolD, C7R98_RS05180 (lipoprotein-releasing ABC transporter permease subunit)
Iron uptake and metabolism	C7R98_RS01530 (TonB-dependent receptor tdfJ)
Biosynthesis of cofactors	ubiE, hpnD
LPS biosynthesis	lptC
Peptidoglycan dynamics	mltG, C7R98_RS06310 (peptidoglycan DD-metalloendopeptidase family protein), C7R98_RS13670 (LysM peptidoglycan-binding domain-containing protein), C7R98_RS10090 (septal ring lytic transglycosylase RlpA family protein)
Amino acid metabolism	trpC
Cell division	zapA
DNA replication and repair	xth, topA, polA
Toxin	C7R98_RS06610 (type II toxin-antitoxin system PemK/MazF family toxin)
Translation	rsmA, tRNA-Glu, miaA, gatC, trmA, tsaE, tRNA-Ala, rsmB, mnmE
Hypothetical proteins	C7R98_RS00030, C7R98_RS13395, C7R98_RS02740, C7R98_RS04570, C7R98_RS04815, C7R98_RS06360, C7R98_RS07405, C7R98_RS08060, C7R98_RS08320, C7R98_RS09035

Figure 18 | Study of host-pathogen interactions on the data about End1 and t-UEC early infection by *N. gonorrhoeae* WHO M. A-D) Emerging networks of transcriptional influences occurring between host and pathogen significant temporal patterns of gene expression when considering End1 infection (**A**) and t-UEC infection (**D**), with the global interference value associated to each node. In the network, host profiles are represented in blue, pathogen ones in red. Edge's size is proportional to the edge weight, based on -ln(Bonferroni-adjusted p-value) of the interaction from the permutation test. Node's size is proportional to the total node strength (sum of the weights on in-coming and out-coming edges). Adjusted p-value to define an edge from the permutation test ≤ 0.05. **B-E)** Hypergeometric test on GO terms (**B**) or biological process (**E**) for genes falling inside host profile 19 in End1 infection network (**B**) or in t-UEC one (**E**). Dot's size for enriched terms is scaled based on the number of profile genes falling inside them, while dot's color is based on the adjusted p-value from the hypergeometric test. Enriched terms are ranked based on the ratio between the number of profile genes falling inside that term and the total number of genes in the profile. Adjusted p-value to call the significance ≤ 0.05. **C)** Temporal expression pattern describing pathogen profile 29 in the End1 infection network and some associated interesting genes grouped by pathways or categories. Each line in the profile corresponds to a gene falling inside the regulation pattern. **E)** Some interesting genes associated to pathogen profile 34 in the t-UEC infection network grouped by pathways or categories.

CHAPTER 4: DISCUSSION

During the late 1980s, campaigns for prevention and care of sexually transmitted infections spread in many industrialized countries. As a consequence, decreases in rates of gonococcal infections were observed. However, about ten years later, new increases started to be reported in the same countries. In 2016, the incidence of gonorrhea was estimated by WHO to reach 86.9 million global cases (global prevalence 0.9%) among adults being 15-49 years old (Rowley *et al.*, 2019). The highest incidences were observed in low-income countries and poorer communities, clearly linked to differences in the access to information regarding this kind of diseases, as well as to availability, accessibility and quality of health-care services. On the contrary, a reported increase in case rates in the past 10 years in countries with more solid health systems can be in part explained by changes in sexual behaviour in the era of antiretroviral treatments for HIV infection, increased connectivity due to the diffusion of dating apps and larger sexual networks. In absence of an effective vaccine, prevention basically relies on promoting safe sexual behaviours and reducing the embarrassment usually associated to these pathologies, which impedes timely diagnosis and treatment, thereby increasing transmission. Single-dose injectable ceftriaxone plus oral azithromycin is the usual recommended first-line treatment. However, an extensive resistance towards antimicrobials has been developed by *N. gonorrhoeae* through the accumulation of many resistance determinants that don't seem to reduce the biological fitness. For instance, epistatic interactions across almost the entire *mtrD* gene, and between mosaic *mtrD* and mosaic *mtr* promoter regions, were found to increase resistance to azithromycin (Wadsworth *et al.*, 2018). Mutations of *mtrR* gene, encoding a transcriptional repressor of the *mtrCDE* efflux pump, were associated to increased gonococcal resistance to different antimicrobial factors and to fitness benefit in a mouse infection model (Warner *et al.*, 2007). Mutations in *gyrA*, or in *gyrA* and *parC*, were showed to confer fluoroquinolones resistance to this pathogen (Kunz *et al.*, 2012). Furthermore, over the *mtrCDE* pump, other efflux pump systems involved in removing toxic molecules and antimicrobials have been identified in *N. gonorrhoeae*: MacA-MacB-MtrE, NorM, FarA-FarB-MtrE and MtrF (Lee and Shafer, 1999). Therefore, to avoid future dramatic scenarios and to control gonorrhea, the development of novel therapeutic strategies or gonococcal vaccines is crucial.

Characterizing gonococcal infections will help the discovery phase for new drugs, but the pathogenesis of such bacterium is not easy to be studied. Upon centuries of co-evolution with its human host, *N. gonorrhoeae* is today a strict, host-adapted human pathogen. It shows sensitivity to many environmental factors, therefore it is not able to survive for long outside the human host and it's also difficult to culture. *In vitro* bacterial cultures under different experimental conditions provided us the majority of information concerning growth and nutrients requirements associated to this pathogen, as well as insights on its arsenal of surface-exposed molecules. In order to study gonococcal interactions with the host, the human challenge model is actually the most relevant existing model, but it's also really limited, because of small cohorts per study, the need for treatment as soon as symptoms start to develop and the possibility to be applied only to men (since women have higher risk of complications) (Cohen and Cannon, 1999). Animal models can be useful to study colonization and immune response in a host, but *N. gonorrhoeae* is restricted to humans and its bacterial proteins have evolved highly specific interactions with human molecules, so also mouse models are of limited value (Jerse *et al.*, 2011). Therefore, in order to study attachment to and internalization by the host, cell culture models are mainly used. Because primary cultures are usually difficult to isolate and maintain and show high heterogeneity, for many studies immortalized transformed human cell lines represented the primary choice. Human endocervical and urethral cell lines (End1 and t-UEC, respectively) have been immortalized by expression of human papillomavirus E6E7 oncogenes, showing high similarity in expression and phenotype with the respective primary cultures or tissue of origin and becoming reference *in vitro* models to study host genital mucosa response to gonococcal infection (Rheinwald and Anderson, no date; Harvey, Post and Apicella, 2002).

More recently, also Detroit562 pharyngeal carcinoma cell monolayer model has been proposed to assess the rate of *N. meningitidis* and *N. gonorrhoeae* attachment to and invasion of epithelial cells. Both End1 and t-UEC cells were tested by enzyme-linked immunosorbent assay (ELISA) for their cytokine response upon challenges with gonococci, demonstrating the former up-regulation of chemokines and cytokines (IL-8 and IL-6) upon 4 hours of infection and of IL-1 after 8 hours of infection (Fichorova *et al.*, 2001), and the latter higher expression of both IL-6 and IL-8 upon interaction with the pathogen (Harvey, Post and Apicella, 2002). No similar information is available for Detroit562 model in literature. Our comparative dual RNA-Seq study for the first time investigated from a global transcriptional point of view host cytokine and pro-inflammatory responsiveness of all these three different cell culture models to early gonococcal infection, and simultaneously allowed the characterization of *N. gonorrhoeae* transcriptome upon adhesion and invasion of the three epithelia, representative of the main anatomical sites infected by this pathogen. Notably, we provided the first early infection-related transcriptional overview for WHO M strain from the recently published WHO panel of *Neisseria gonorrhoeae* strains validated as representative of the species.

Both from a phenotypical and host-pathogen coupled transcriptional point of view, the three infection models showed different dynamics of early infection. This could be also relevant to guide the setup of future experiments that utilize similar experimental settings. Adhesion of *N. gonorrhoeae* WHO M to Detroit562 cell line appeared to be slower or more difficult compared to the other two systems, probably explaining why pathogens infecting this cell line at 15 mpi more clearly up-regulated a set of pilins and outer membrane beta-barrel proteins having high homology with Opa proteins. However, it is important to keep in mind that both pilins and Opa proteins show different levels of regulation over the transcriptional one, in particular recombination for the firsts and phase variation for the seconds, that could explain the less evident pattern of differential expression at 15 mpi emerging in the other two cases (especially in case of t-UEC infection) and surely make difficult to correctly interpret these signals. Concomitantly, Detroit562 DE response upon 15 minutes of infection was quite small, demonstrating the presence of a still early pro-inflammatory program in these cells at this time point. From clustering analyses, important transduction pathways like JNK and MAPK cascades emerged as up-regulated at the same time point, while DE analysis highlighted increased transcript levels for relevant regulators, such as the MAPK-regulated *ATF3* and the NF-kB complex-regulating *NFKBIZ*. Indeed, it has been demonstrated that adherence, rather than invasion, is already responsible for the initial induction of NF-kB complexes in the host (Naumann *et al.*, 1997). Furthermore, upon 15 minutes of infection these cells up-regulated genes from the CEACAM family of molecules (*CEACAM1* and *CEACAM5*) as a possible consequence of pathogen induction mechanisms and that could be involved in the interaction with the regulated gonococcal outer membrane proteins. Both from clustering and enrichment analyses, a NF-kB-related and chemotactic response was quite clear already upon 15 minutes of infection for End1 cell line, while t-UEC cells at this time point were mainly involved in physical rearrangements, probably accompanying pathogen invasion. Indeed, different actin and actin-related proteins, clathrin and caveolin genes, together with signalling molecules like *PIK3CB*, emerged as regulated by these cells. A comparative investigation of bacterial strategies mounted following adhesion to and early intracellular persistence into the different cell culture models was mainly hindered in our study by the initial, and essential to answer our questions, adaptation of bacteria to the three cell media differing in their nutrients composition and concentrations, especially considering bacteria ready to infect Detroit562 cell line compared to the ones prepared for the infection of End1 and t-UEC cell lines. Nevertheless, some differences and interesting common patterns emerged, already upon 15 minutes of infection. Almost all the best-known genes involved in antimicrobial resistance by this pathogen were found as down-regulated in the initial phases of attachment to Detroit562 cells, including *mtrCDE* genes, the penicillin-binding protein 1A, the broad-spectrum beta-lactamase TEM-1, *macA* and *farB*, while *macA* emerged as up-regulated once the bacteria were internalized by these host cells, together with *secE* and *secY* protein export system genes. Similarly, some of these genes, such as *mtrC* and the penicillin-binding protein 2, were initially down-regulated by bacteria attaching to t-UEC cells. Therefore, this data suggested this strain requirement for such genes in later phases of epithelia colonization and invasion, rather than in the initial ones.

luxS is a gene involved in the biosynthesis of a quorum-sensing molecule, the autoinducer-2, that has been reported to have a role in virulence in *Escherichia coli* and other bacteria (DeLisa *et al.*, 2001; Sperandio, Torres and Kaper, 2002). A study showed that a *Neisseria meningitidis ΔluxS* mutant was attenuated for bacteremic infection in a rat model (Winzer *et al.*, 2002), however the role of this gene in quorum sensing for *N. meningitidis* still lacks evidences, both in culture and in contact with epithelial cells, therefore up to now it is just considered a metabolic by-product (Dove *et al.*, 2003). Similarly, a lack of knowledge is reported also for what concerns *N. gonorrhoeae*. Interestingly, from our analyses, *luxS* appeared to be up-regulated only by bacteria infecting t-UEC cell line at 15 mpi, a situation in our system where many bacteria were observed to be attached to the epithelium but still waiting for successfully invading the host, in which the induction of this gene could suggest associated bacterial cell-to-cell communication functions.

Upon 120 minutes of infection, Detroit562 host cells were more homogeneous in their response, with a higher number of cells that internalized the bacteria (detected in clear clusters inside the host cells from confocal microscopy) and, as a consequence, higher transcriptional similarity emerging from unsorted and sorted samples. End1 infection appeared to be more spread in the epithelium, with different bacterial cells detected as extracellular. Finally, t-UEC cells were characterized by lower extents of infection. Enrichment analyses on the set of up-regulated genes clearly highlighted for all the three cell models the activation of pro-inflammatory and defense pathways, but examining both the top statistically significant DE genes and the regulations of genes belonging to know pro-inflammatory categories a similar program emerged from Detroit562 and End1 cells, while t-UEC ones appeared to be in delay or less efficient in their response. Different chemokines and some inflammatory modulators (such as the TNF-induced *TNFAIP3* and *ZC3H12A*) particularly emerged among the top statistically regulated genes and/or the most up-regulated ones for the first two models. Furthermore, *ICAM-1* emerged among the top statistically significant up-regulated genes for End1 cells, and it was also up-regulated by Detroit562 ones ($\log_2$ fold change 1.5, adjusted p-value 2.70e-09). ICAM-1, an adhesion molecule mediating inflammatory responses, has been reported as being clustered in the cortical plaques that are formed in host cancer epithelial cells at the site of pilus-mediated gonococci adhesion and entry (Jarvis, Li and V. Swanson, 1999; Merz, Enns and So, 1999b), and markedly up-regulated in immortalized End1 cell line by both piliated and non-piliated gonococci at 4h, 8h and 24h post-infection (Fichorova *et al.*, 2001).

Among the most up-regulated cytokines expressed upon 120 minutes of infection by Detroit562 cells that internalized the bacteria, our study reported the potent pro-inflammatory cytokines *TNF*, *IL-6*, *IL-17C* and different members of the interleukin-1 cytokine family, such as *IL-1A*, *IL-1B*, *IL-36G* and the interleukin 1 receptors *IL1RL1* and *IL1RL2*. Similarly, for End1 cells at the same time point we reported the high up-regulation of *TNF*, *IL-1A*, *IL-17C*, *IL-36G*, *IL-6*, together with the important infection-related cytokine *IL-23A*. Finally, t-UEC cells started to up-regulate *IL-1A*, the interleukin-1 receptor accessory protein *IL1RAP* and the interleukin-6 receptor subunit beta *IL6ST*. All the three cell lines over-expressed mainly *CXCL8* (or *IL-8*), *CCL20*, *CXCL3*, *CXCL2* and *CXCL1* chemokines. Among interferons, our study promoted *IFNE* among the most up-regulated one in the three cell culture models. By mapping emerging genes into the human regulatory network we highlighted the centrality that AP-1 and especially NF-kB TFs may have in regulating the expression of almost all these defense factors after *N. gonorrhoeae* infection, as it has been already suggested by previous studies (Naumann *et al.*, 1997). Furthermore, together with such a strong pro-inflammatory response, our work also highlighted the importance, already during the early phases of gonococcal interaction, of modulatory pathways activated by the host to finely tune the response and avoid cell damages, as it was particularly suggested by the high increase in transcript abundance in all the three models of modulators like the aforementioned *TNFAIP3* and the complement system regulator *CD55*. Different studies proposed the regulation of host cell death by gonococcus infection as a mechanism used by this pathogen to promote its survival and proliferation in the host epithelium. In particular, the anti-apoptotic genes *BFL-1, COX-2, c-IAP-2* and *MCL-1* were found to be up-regulated in infected urethral cells compared to uninfected ones (Binnicker, Williams and Apicella, 2003, 2004). From our work, the strongly statistically significant up-regulation of *BIRC3*, an anti-apoptotic gene that act by binding the tumor necrosis factor receptor-associated factors TRAF1 and TRAF2, emerged by all the three cell lines upon 120 minutes of infection ($\log_2$ fold change in cells that internalized the bacteria 5.4, 4.2 and 1.0 for Detroit562, End1 and t-UEC, respectively), probably as a mechanism exploited by the pathogen to ensure the host survival in a condition of TNF over-expression. *MCL1* gene, that has been demonstrated to have a prominent role in tumor cell death evasion (Deng *et al.*, 2007; Touzeau *et al.*, 2016; Gong *et al.*, 2016), was also up-regulated at the same time point by Detroit562 cancer cells that internalized the bacteria ($\log_2$ fold change 1.0). Finally, negative modulators of the intrinsic apoptotis pathway also showed higher transcript abundances in both Detroit562 and End1 invaded cells at 120 mpi, such as *BCL2A1* for both the cell lines ($\log_2$ fold change 8.5 and 4.0, respectively) and *BCL2* for Detroit562 cells ($\log_2$ fold change 1.4). Interestingly, different lysosomal genes, including *LAMP1*, were reported as downregulated only by t-UEC cells invaded by bacteria. Some studies demonstrated that gonococci, in later phases of infection, could degrade LAMP1 protein by IgA1 protease activity and, indirectly, also lower the amount of other lysosomal components to disturb the autophagic pathway and ultimately promote intracellular survival (Kim *et al.*, 2019; Ayala *et al.*, 1998). Therefore, what we observed could reflect at the transcriptional level a similar condition. Such survival strategies were associated only to piliated strains with Opas phase switched off, as the only consequence of pilum-mediated attachment to epithelial cells (Kim *et al.*, 2019), similarly to what we described from a transcriptional point of view in case of our t-UEC infection model. This could explain why such behaviour seemed to be valid only for the infection of these cells.

From the other side, pathogen transcriptomes upon internalization by the different cell lines strongly clustered in PCA analysis, and all were characterized by the consistent down-regulation of pilins and other factors putatively involved in the initial adhesion to the host. Curiously, genes involved in lipid biosynthesis appeared to be consistently down-regulated by bacteria being inside genital mucosa cell lines (End1 and t-UEC), in particular *fabFGZ* in case of both the infections and *fabI* in case of the End1 infection. Some of them also appeared as up-regulated by bacteria adapted to the EMEM medium compared to the ones adapted to the KSFM or PrEGM one, so a similar expression shift could be valid also for bacteria infecting Detroit562 cells. This trend could reflect a situation in which the pathogen is activating mechanisms for exogenous fatty acids incorporation once it is inside the host (Yao *et al.*, 2016).

As expected, once bacteria were successfully internalized by all the different host cells, they started to up-regulate a series of membrane transporters in order to sequester nutrients from the intracellular environment and to survive inside the host. Incomplete biosynthetic capabilities by this pathogen, as it is the case for amino acid biosynthesis, probably derive by the fact that important nutrients are readily obtained from the host. Indeed, the gene co-expression network derived from our data set confirmed the centrality of many amino acid transporter genes also for our strain. Furthermore, different amino acid transporters emerged as up-regulated by internalized bacteria in our systems. Other emerging transporters were sulfate-, zinc- and of course iron-related. Iron acquisition is crucial for many bacterial pathogens, including *N. gonorrhoeae*, and an extensive literature exist on this topic. Our transcriptomic profiling confirmed the importance of TonB and TonB-related genes regulation also for WHO M strain ability to successfully infect host epithelium, especially the endocervical one. However, contrary to previous reports regarding the essentiality for the reference strain FA1090 of the putative TonB-dependent transporter *tdfF* (Hagen and Cornelissen, 2006), WHO M during early phases of infection seemed to depend more on *tdfH* and especially *tdfJ* ones for evading the nutritional immunity by the host. Recent findings involve these two transporters particularly in zinc sequestration. Indeed, tdfH has been reported to be able to bind human calprotectin, a human protein sequestering metals including zinc and manganese, and subsequently incorporate zinc into the cell (Stork *et al.*, 2013; Jean *et al.*, 2016; Kammerman *et al.*, 2020). Similarly, tdfJ was described as contributing to *N. gonorrhoeae* growth in Zn-restricted conditions, and subsequently demonstrated to bind human S100A7, that like calprotectin is involved in metal sequestration, helping in zinc internalization inside the bacterial cell (Stork *et al.*, 2010; Jean *et al.*, 2016; Maurakis *et al.*, 2019). Importantly, S100A7 is enriched in human epithelial tissues, including those of the oral and genital mucosa, and a meningococcal mutant unable to produce the tdfJ homolog ZnuD was impaired for survival in epithelial cells (Kumar, Sannigrahi and Tzeng, 2012).

Dual RNA-Seq is a recent technique that showed in the last years a fast-growing trend, as demonstrated by the great number of its applications appearing in literature. Infection mechanisms cannot be comprehensively elucidated without taking in consideration both the counterparts involved in the process. Coupled with the high resolution and sensitivity reached by RNA-Seq, nowadays dual transcriptomics allows the deep and simultaneous snapshot of both host and bacteria transcriptomes during the different phases of the interaction, therefore helping in elucidating the molecular crosstalk that is behind the whole process. However, in addition to profile the respective one-side transcriptional status, such data in their entirety should in principle contain also the information related to the interaction occurring between the two organisms. Therefore, we developed a new analytical strategy to ask this data for such a hidden potential and to test for possible transcriptional reciprocal influences developed by the system during the examined infection period. In this way, key regulations sustaining the whole "architecture" of the host-pathogen transcriptional interaction, so gene modulations that are fundamental to define an infection condition by both the actors, could be extrapolated. Theoretically speaking, the problem is quite difficult to approach, since it implies to look at two interacting complex systems – host and pathogen cells – that start a complex relationship, basically "co-evolving" along time. However, the major drawback is mainly represented by the paucity in analysed time points that, due to different reasons, usually characterizes these experiments and weakens the statistical reliability of any exploratory approach. Inspired by the work of Ernst et al. (Ernst, Nau and Bar-Joseph, 2005a), our permutation-based method helped us to mine the whole complexity of host-pathogen interactions naturally emerging from these experiments, limiting the issue of having few time points. From our strategy, it was not possible to infer direct causality – i.e. an ordered series of expression patterns each one being caused by a previous one and influencing the occurrence of another one in the system – but node removal tests allowed to consider the whole complexity of resulting networks and somehow to rank or link relevant regulations by the two organisms. By first testing the strategy on a publicly-available dual RNA-Seq data set of early infection (Aprianto *et al.*, 2016), our method demonstrated to correctly capture the most relevant patterns of gene expression that are known to be at the basis of the initial adhesion and colonization of epithelial cells by the well-studied *Streptococcus pneumoniae* strain D39. The analysis of the data set about the infection by the mutants confirmed our strategy ability to highlight the most relevant regulations that permit the observed interaction at different infection stages or conditions, even if many underlying host-pathogen transcriptional influences could be basically the same. Altogether, these results prompted us to perform this kind of analysis also on our new data set, for which it could be particularly useful since we aimed to study infection by a still poorly characterized, non-reference bacterial strain, for which interpretation and prioritization of relevant genes only from classical DE analyses is more challenging. Similarly also to what we observed by applying the method to the aforementioned data set, the application to the new *N. gonorrhoeae* WHO M dual RNA-Seq data confirmed for all the three infection models the induction of cytoskeletal rearrangements in the infected cells as the basal event for the host during early interaction (Grassmé, Ireland and van Putten, 1996b). Due to the fact that the majority of putative or known bacterial adhesion factors implicated in the early phases of infection were not significantly or uniformly regulated along the time points, few patterns were selected as significantly enriched by the clustering strategy for the pathogen and none among them described inductions occurring at the beginning of the interaction. Therefore, the emerging profiles by our strategy for the bacterium were mainly linked to important genes for pathogens survival once they already invaded host cells – i.e., genes mainly involved in nutrients uptake, stress response and cell division. Interestingly, many genes of the maintenance of lipid asymmetry (Mla) six-component system, MlaA-F, fell inside these profiles in both the well-established Detroit562 and End1 infections. In Gram-negative bacteria, this system participates in the retrograde transport of phospholipids from the outer leaflet of the outer membrane to the inner membrane, ensuring the asymmetry structure of the cell envelope and its barrier function to prevent the entry of toxic lipophilic molecules (Baarda *et al.*, 2019). Therefore, even if little is known about the regulation of this system, our data suggested its centrality among the antimicrobial resistance strategies mounted by this strain invading host cells. Furthermore, *mla* mutants have been associated to a loss in virulence in *Escherichia coli* (Malinverni and Silhavy, 2009) and *Shigella flexneri*, where these genes have been reported to be required for the escape from host double membrane in the process that allow the bacterial intercellular spread into adjacent cells (Hong *et al.*, 1998). The majority of the components of this system (i.e., *mlaBCDE*) were particularly observed in the pathogen profile emerging from End1 infection, that upon 120 minutes showed bacteria spread over all the epithelium, so this may suggest a similar role of this system also in *N. gonorrhoeae* WHO M. Several other genes involved in external cell structures (both peptidoglycan and LOS) were also reported in these profiles, indicating the importance of structural dynamics in invading pathogens. In general, our new strategy suffered from some limitations, many of them particularly arising in the analysis of the new data set:

i) Influences between temporal patterns of gene expression occurring in the same organism were not considered by our method. The assumption was that the major driver for transcriptional regulation in both the organisms was represented by the interaction with the partner;
ii) In the analysis of the gonococcus data set, the number of time points was the minimum (just three, considering also the pre-infection experiment), leading to a really low number of permutations that could be computed and to more overlapping profiles;
iii) Relevant genes need to uniformly generate a co-expressed cluster of regulated genes in order to be selected and tested for their relevance in the interaction with the partner.

Our new strategy could be useful, in combination with standard DE analyses, to prioritize genes for further validation procedures and experiments. In our case, the new method also highlighted different sets of unknown genes that could be better characterized particularly for their relevance in the infection process.

In conclusion, we performed a first comparative dual RNA-Seq study to characterize three different *in vitro* infection models of human epithelial cells infected by *N. gonorrhoeae* during the early phases of the process. This kind of work highlighted common patterns, as well as differences, in the infection dynamics both from a phenotypical point of view and from a transcriptional one, providing detailed transcriptomics insights on the early interaction between the two organisms from both perspectives. Such information could be useful to guide future experiments that use these established *in vitro* models to study gonococcal pathogenesis and to sustain knowledge and observations regarding putative targets to be used in vaccine or therapeutic formulations. Further studies should be directed to validate specific mechanisms that occur in such systems emerging from our global transcriptional overview, and of course to assess the importance of selected factors for the infection process established by this pathogen and/or their ability to trigger a defensive response in the host. Finally, we also proposed a new analytical strategy to mine the complexity of host-pathogen interaction and to ask for relevant information in short time series dual RNA-Seq data. The major limitations in developing such new approaches are intrinsic to the nature of these experiments. We attempt to speculate that, with the advances in protocols and methods and with further reduction in the costs, in future this data could be less limited and their full potential more easily extrapolated. Here, we posed the first “stone” for the development of new integrative methods that could be added in the analysis workflow used to characterize host-pathogen interaction by dual transcriptomics.

BIBLIOGRAPHY

Albrecht, M. *et al.* (2010) ‘Deep sequencing-based discovery of the Chlamydia trachomatis transcriptome’, *Nucleic Acids Research*, 38(3), pp. 868–877. doi:10.1093/nar/gkp1032.

Albrecht, M. *et al.* (2011) ‘The transcriptional landscape of Chlamydia pneumoniae’, *Genome Biology*, 12(10), p. R98. doi:10.1186/gb-2011-12-10-r98.

Aprianto, R. *et al.* (2016) ‘Time-resolved dual RNA-seq reveals extensive rewiring of lung epithelial and pneumococcal transcriptomes during early infection’, *Genome Biology*, 17(1), p. 198. doi:10.1186/s13059-016-1054-5.

Aukema, K.G. *et al.* (2005) ‘Functional Dissection of a Conserved Motif within the Pilus Retraction Protein PilT’, *Journal of Bacteriology*, 187(2), pp. 611–618. doi:10.1128/JB.187.2.611-618.2005.

Avraham, R. *et al.* (2015) ‘Pathogen Cell-to-Cell Variability Drives Heterogeneity in Host Immune Responses’, *Cell*, 162(6), pp. 1309–1321. doi:10.1016/j.cell.2015.08.027.

Ayala, P. *et al.* (1998) ‘Infection of Epithelial Cells by Pathogenic Neisseriae Reduces the Levels of Multiple Lysosomal Constituents’, *Infection and Immunity*. Edited by P.J. Sansonetti, 66(10), pp. 5001–5007. doi:10.1128/IAI.66.10.5001-5007.1998.

Baarda, B.I. *et al.* (2019) ‘Neisseria gonorrhoeae MlaA influences gonococcal virulence and membrane vesicle production’, *PLOS Pathogens*. Edited by C. Tang, 15(3), p. e1007385. doi:10.1371/journal.ppat.1007385.

Baddal, B. *et al.* (2015) ‘Dual RNA-seq of Nontypeable Haemophilus influenzae and Host Cell Transcriptomes Reveals Novel Insights into Host-Pathogen Cross Talk’, *mBio*. Edited by S.J. Projan, 6(6), pp. e01765-15. doi:10.1128/mBio.01765-15.

Bailey, T.L. and Grant, C.E. (2021) *SEA: Simple Enrichment Analysis of motifs*. preprint. Bioinformatics. doi:10.1101/2021.08.23.457422.

Barabàsi, A.L. (2015) 'Network Science', Cambridge University Press.

Belcher, C.E. *et al.* (2000) ‘The transcriptional responses of respiratory epithelial cells to Bordetella pertussis reveal host defensive and pathogen counter-defensive strategies’, *Proceedings of the National Academy of Sciences*, 97(25), pp. 13847–13852. doi:10.1073/pnas.230262797.

Belland, R.J. *et al.* (2003) ‘Genomic transcriptional profiling of the developmental cycle of *Chlamydia trachomatis*’, *Proceedings of the National Academy of Sciences*, 100(14), pp. 8478–8483. doi:10.1073/pnas.1331135100.

Bernstein, K.T. *et al.* (2009) ‘*Chlamydia trachomatis* and *Neisseria gonorrhoeae* Transmission from the Oropharynx to the Urethra among Men Who Have Sex with Men’, *Clinical Infectious Diseases*, 49(12), pp. 1793–1797. doi:10.1086/648427.

Binnicker, M.J., Williams, R.D. and Apicella, M.A. (2003) ‘Infection of human urethral epithelium with Neisseria gonorrhoeae elicits an upregulation of host anti-apoptotic factors and protects cells from staurosporine-induced apoptosis’, *Cellular Microbiology*, 5(8), pp. 549–560. doi:10.1046/j.1462-5822.2003.00300.x.

Binnicker, M.J., Williams, R.D. and Apicella, M.A. (2004) ‘Gonococcal Porin IB Activates NF-κB in Human Urethral Epithelium and Increases the Expression of Host Antiapoptotic Factors’, *Infection and Immunity*, 72(11), pp. 6408–6417. doi:10.1128/IAI.72.11.6408-6417.2004.

Birrell, J.M. *et al.* (2019) ‘Characteristics and Impact of Disseminated Gonococcal Infection in the “Top End” of Australia’, *The American Journal of Tropical Medicine and Hygiene*, 101(4), pp. 753–760. doi:10.4269/ajtmh.19-0288.

Borgatti, S.P. (no date) '2-mode Concepts in Social Network Analysis', *Encyclopedia of Complexity and System Science*.

Bray, N.L. *et al.* (no date) ‘Near-optimal RNA-Seq quantification’, p. 21.

Cabili, M.N. *et al.* (2011) ‘Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses’, *Genes & Development*, 25(18), pp. 1915–1927. doi:10.1101/gad.17446611.

Cassat, J.E. and Skaar, E.P. (2013) ‘Iron in Infection and Immunity’, *Cell Host & Microbe*, 13(5), pp. 509–519. doi:10.1016/j.chom.2013.04.010.

Cohen, M.S. and Cannon, J.G. (1999) ‘Human Experimentation with *Neisseria gonorrhoeae:* Progress and Goals’, *The Journal of Infectious Diseases*, 179(s2), pp. S375–S379. doi:10.1086/513847.

Cohen, P. *et al.* (2000) ‘Monitoring Cellular Responses to Listeria monocytogenes with Oligonucleotide Arrays’, *Journal of Biological Chemistry*, 275(15), pp. 11181–11190. doi:10.1074/jbc.275.15.11181.

Dehio, M. *et al.* (1998) ‘Vitronectin-dependent invasion of epithelial cells by *Neisseria gonorrhoeae* involves α_{v} integrin receptors’, *FEBS Letters*, 424(1–2), pp. 84–88. doi:10.1016/S0014-5793(98)00144-6.

DeLisa, M.P. *et al.* (2001) ‘DNA Microarray-Based Identification of Genes Controlled by Autoinducer 2-Stimulated Quorum Sensing in *Escherichia coli*’, *Journal of Bacteriology*, 183(18), pp. 5239–5247. doi:10.1128/JB.183.18.5239-5247.2001.

Deng, J. *et al.* (2007) ‘BH3 Profiling Identifies Three Distinct Classes of Apoptotic Blocks to Predict Response to ABT-737 and Conventional Chemotherapeutic Agents’, *Cancer Cell*, 12(2), pp. 171–185. doi:10.1016/j.ccr.2007.07.001.

Djebali, S. *et al.* (2012) ‘Landscape of transcription in human cells’, *Nature*, 489(7414), pp. 101–108. doi:10.1038/nature11233.

Dobin, A. *et al.* (2013) ‘STAR: ultrafast universal RNA-seq aligner’, *Bioinformatics*, 29(1), pp. 15–21. doi:10.1093/bioinformatics/bts635.

Dove, J.E. *et al.* (2003) ‘Production of the signalling molecule, autoinducer-2, by Neisseria meningitidis: lack of evidence for a concerted transcriptional response’, *Microbiology*, 149(7), pp. 1859–1869. doi:10.1099/mic.0.26185-0.

Du, Y., Lenz, J. and Arvidson, C.G. (2005) ‘Global Gene Expression and the Role of Sigma Factors in *Neisseria gonorrhoeae* in Interactions with Epithelial Cells’, *Infection and Immunity*, 73(8), pp. 4834–4845. doi:10.1128/IAI.73.8.4834-4845.2005.

Eckmann, L. *et al.* (2000) ‘Analysis by High Density cDNA Arrays of Altered Gene Expression in Human Intestinal Epithelial Cells in Response to Infection with the Invasive Enteric BacteriaSalmonella’, *Journal of Biological Chemistry*, 275(19), pp. 14084–14094. doi:10.1074/jbc.275.19.14084.

Edwards, J.L. and Apicella, M.A. (2004) ‘The Molecular Mechanisms Used by *Neisseria gonorrhoeae* To Initiate Infection Differ between Men and Women’, *Clinical Microbiology Reviews*, 17(4), pp. 965–981. doi:10.1128/CMR.17.4.965-981.2004.

Eriksson, S. *et al.* (2003) ‘Unravelling the biology of macrophage infection by gene expression profiling of intracellular *Salmonella enterica*’, *Molecular Microbiology*, 47(1), pp. 103–118. doi:10.1046/j.1365-2958.2003.03313.x.

Ernst, J. and Bar-Joseph, Z. (2006) ‘[No title found]’, *BMC Bioinformatics*, 7(1), p. 191. doi:10.1186/1471-2105-7-191.

Ernst, J., Nau, G.J. and Bar-Joseph, Z. (2005a) ‘Clustering short time series gene expression data’, *Bioinformatics*, 21(Suppl 1), pp. i159–i168. doi:10.1093/bioinformatics/bti1022.

Ernst, J., Nau, G.J. and Bar-Joseph, Z. (2005b) ‘Clustering short time series gene expression data’, *Bioinformatics*, 21(Suppl 1), pp. i159–i168. doi:10.1093/bioinformatics/bti1022.

Fichorova, R.N. *et al.* (2001) ‘Distinct Proinflammatory Host Responses to *Neisseria gonorrhoeae* Infection in Immortalized Human Cervical and Vaginal Epithelial Cells’, *Infection and Immunity*. Edited by E.I. Tuomanen, 69(9), pp. 5840–5848. doi:10.1128/IAI.69.9.5840-5848.2001.

Fichorova, R.N. and Anderson, D.J. (1999) ‘Differential Expression of Immunobiological Mediators by Immortalized Human Cervical and Vaginal Epithelial Cells1’, *Biology of Reproduction*, 60(2), pp. 508–514. doi:10.1095/biolreprod60.2.508.

Goh, K.-I. *et al.* (2007) ‘The human disease network’, *Proceedings of the National Academy of Sciences*, 104(21), pp. 8685–8690. doi:10.1073/pnas.0701361104.

Gómez-Duarte, O.G. *et al.* (1997) ‘Binding of vitronectin to opa-expressing Neisseria gonorrhoeae mediates invasion of HeLa cells’, *Infection and Immunity*, 65(9), pp. 3857–3866. doi:10.1128/iai.65.9.3857-3866.1997.

Gong, J.-N. *et al.* (2016) ‘Hierarchy for targeting prosurvival BCL2 family proteins in multiple myeloma: pivotal role of MCL1’, *Blood*, 128(14), pp. 1834–1844. doi:10.1182/blood-2016-03-704908.

Grassmé, H.U., Ireland, R.M. and van Putten, J.P. (1996a) ‘Gonococcal opacity protein promotes bacterial entry-associated rearrangements of the epithelial cell actin cytoskeleton’, *Infection and Immunity*, 64(5), pp. 1621–1630. doi:10.1128/iai.64.5.1621-1630.1996.

Grassmé, H.U., Ireland, R.M. and van Putten, J.P. (1996b) ‘Gonococcal opacity protein promotes bacterial entry-associated rearrangements of the epithelial cell actin cytoskeleton’, *Infection and Immunity*, 64(5), pp. 1621–1630. doi:10.1128/iai.64.5.1621-1630.1996.

Griesenauer, B. *et al.* (2019) ‘Determination of an Interaction Network between an Extracellular Bacterial Pathogen and the Human Host’, *mBio*. Edited by J. Vogel, 10(3), pp. e01193-19, /mbio/10/3/mBio.01193-19.atom. doi:10.1128/mBio.01193-19.

Griffiss, J.M. *et al.* (1999) ‘*Neisseria gonorrhoeae* Coordinately Uses Pili and Opa To Activate HEC-1-B Cell Microvilli, Which Causes Engulfment of the Gonococci’, *Infection and Immunity*. Edited by D.L. Burns, 67(7), pp. 3469–3480. doi:10.1128/IAI.67.7.3469-3480.1999.

Guthke, R. *et al.* (2005) ‘Dynamic network reconstruction from gene expression data applied to immune response during bacterial infection’, *Bioinformatics*, 21(8), pp. 1626–1634. doi:10.1093/bioinformatics/bti226.

Hagen, T.A. and Cornelissen, C.N. (2006) ‘Neisseria gonorrhoeae requires expression of TonB and the putative transporter TdfF to replicate within cervical epithelial cells’, *Molecular Microbiology*, 62(4), pp. 1144–1157. doi:10.1111/j.1365-2958.2006.05429.x.

Han, H. *et al.* (2018) ‘TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions’, *Nucleic Acids Research*, 46(D1), pp. D380–D386. doi:10.1093/nar/gkx1013.

Hardcastle, T.J. and Kelly, K.A. (2010) ‘baySeq: Empirical Bayesian methods for identifying differential expression in sequence count data’, *BMC Bioinformatics*, 11(1), p. 422. doi:10.1186/1471-2105-11-422.

Harvey, H.A. *et al.* (1997) ‘Ultrastructural Analysis of Primary Human Urethral Epithelial Cell Cultures Infected with Neisseria gonorrhoeae’, *INFECT. IMMUN.*, 65, p. 8.

Harvey, H.A., Post, D.M.B. and Apicella, M.A. (2002) ‘Immortalization of Human Urethral Epithelial Cells: a Model for the Study of the Pathogenesis of and the Inflammatory Cytokine Response to *Neisseria gonorrhoeae* Infection’, *Infection and Immunity*, 70(10), pp. 5808–5815. doi:10.1128/IAI.70.10.5808-5815.2002.

Hautefort, I. *et al.* (2008) ‘During infection of epithelial cells Salmonella enterica serovar Typhimurium undergoes a time-dependent transcriptional adaptation that results in simultaneous expression of three type 3 secretion systems’, *Cellular Microbiology*, 10(4), pp. 958–984. doi:10.1111/j.1462-5822.2007.01099.x.

Higashi, D.L. *et al.* (2007) ‘Dynamics of *Neisseria gonorrhoeae* Attachment: Microcolony Development, Cortical Plaque Formation, and Cytoprotection’, *Infection and Immunity*, 75(10), pp. 4743–4753. doi:10.1128/IAI.00687-07.

Hong, M. *et al.* (1998) ‘Identification of Two *Shigella flexneri* Chromosomal Loci Involved in Intercellular Spreading’, *Infection and Immunity*. Edited by J.T. Barbieri, 66(10), pp. 4700–4710. doi:10.1128/IAI.66.10.4700-4710.1998.

Humphrys, M.S. *et al.* (2013) ‘Simultaneous Transcriptional Profiling of Bacteria and Their Host Cells’, *PLoS ONE*. Edited by K. Ramsey, 8(12), p. e80597. doi:10.1371/journal.pone.0080597.

Ichikawa, J.K. *et al.* (2000) ‘Interaction of Pseudomonas aeruginosa with epithelial cells: Identification of differentially regulated genes by expression microarray analysis of human cDNAs’, *Proceedings of the National Academy of Sciences*, 97(17), pp. 9659–9664. doi:10.1073/pnas.160140297.

International Human Genome Sequencing Consortium (2004) ‘Finishing the euchromatic sequence of the human genome’, *Nature*, 431(7011), pp. 931–945. doi:10.1038/nature03001.

Isabella, V.M. and Clark, V.L. (2011) ‘Deep sequencing-based analysis of the anaerobic stimulon in Neisseria gonorrhoeae’, *BMC Genomics*, 12(1), p. 51. doi:10.1186/1471-2164-12-51.

Jackson, L.A. *et al.* (2013) ‘Control of RNA Stability by NrrF, an Iron-Regulated Small RNA in Neisseria gonorrhoeae’, *Journal of Bacteriology*, 195(22), pp. 5166–5173. doi:10.1128/JB.00839-13.

Jarvis, G.A., Li, J. and V. Swanson, K. (1999) ‘Invasion of Human Mucosal Epithelial Cells by *Neisseria gonorrhoeae* Upregulates Expression of Intercellular Adhesion Molecule 1 (ICAM-1)’, *Infection and Immunity*. Edited by J.R. McGhee, 67(3), pp. 1149–1156. doi:10.1128/IAI.67.3.1149-1156.1999.

Jean, S. *et al.* (2016) ‘Neisseria gonorrhoeae Evades Calprotectin-Mediated Nutritional Immunity and Survives Neutrophil Extracellular Traps by Production of TdfH’, *Infection and Immunity*. Edited by S.M. Payne, 84(10), pp. 2982–2994. doi:10.1128/IAI.00319-16.

Jenner, R.G. and Young, R.A. (2005) ‘Insights into host responses against pathogens from transcriptional profiling’, *Nature Reviews Microbiology*, 3(4), pp. 281–294. doi:10.1038/nrmicro1126.

Jerse, A.E. (1999) ‘Experimental Gonococcal Genital Tract Infection and Opacity Protein Expression in Estradiol-Treated Mice’, *INFECT. IMMUN.*, 67, p. 10.

Jerse, A.E. *et al.* (2011) ‘Estradiol-Treated Female Mice as Surrogate Hosts for Neisseria gonorrhoeae Genital Tract Infections’, *Frontiers in Microbiology*, 2. doi:10.3389/fmicb.2011.00107.

Kallstrom, H. *et al.* (2001) ‘Attachment of Neisseria gonorrhoeae to the cellular pilus receptor CD46: identification of domains important for bacterial adherence’, *Cellular Microbiology*, 3(3), pp. 133–143. doi:10.1046/j.1462-5822.2001.00095.x.

Kammerman, M.T. *et al.* (2020) ‘Molecular Insight into TdfH-Mediated Zinc Piracy from Human Calprotectin by Neisseria gonorrhoeae’, *mBio*. Edited by N.E. Freitag, 11(3). doi:10.1128/mBio.00949-20.

Kibble E.A., Sarkar-Tyson M., Coombs G.W., Kahler C.M. (2019) 'The Detroit 562 Pharyngeal Immortalized Cell Line Model for the Assessment of Infectivity of Pathogenic *Neisseria* sp.', Seib K., Peak I. (eds) Neisseria meningitidis. Methods in Molecular Biology, vol 1969. Humana Press, New York, NY.

Kim, D., Langmead, B. and Salzberg, S.L. (2015) ‘HISAT: a fast spliced aligner with low memory requirements’, *Nature Methods*, 12(4), pp. 357–360. doi:10.1038/nmeth.3317.

Kim, W.J. *et al.* (2019) ‘Neisseria gonorrhoeae evades autophagic killing by downregulating CD46-cyt1 and remodeling lysosomes’, *PLOS Pathogens*. Edited by S.R. Blanke, 15(2), p. e1007495. doi:10.1371/journal.ppat.1007495.

Kumar, P., Sannigrahi, S. and Tzeng, Y.-L. (2012) ‘The Neisseria meningitidis ZnuD Zinc Receptor Contributes to Interactions with Epithelial Cells and Supports Heme Utilization when Expressed in Escherichia coli’, *Infection and Immunity*. Edited by J.N. Weiser, 80(2), pp. 657–667. doi:10.1128/IAI.05208-11.

Kumar, R. *et al.* (2010) ‘RIdeseearnchtairftiiccleation of novel non-coding small RNAs from Streptococcus pneumoniae TIGR4 using high-resolution genome tiling arrays’, p. 19.

Kunz, A.N. *et al.* (2012) ‘Impact of Fluoroquinolone Resistance Mutations on Gonococcal Fitness and In Vivo Selection for Compensatory Mutations’, *Journal of Infectious Diseases*, 205(12), pp. 1821–1829. doi:10.1093/infdis/jis277.

Landgraf, P. *et al.* (2007) ‘A Mammalian microRNA Expression Atlas Based on Small RNA Library Sequencing’, *Cell*, 129(7), pp. 1401–1414. doi:10.1016/j.cell.2007.04.040.

Langmead, B. *et al.* (2009) ‘Ultrafast and memory-efficient alignment of short DNA sequences to the human genome’, *Genome Biology*, 10(3), p. R25. doi:10.1186/gb-2009-10-3-r25.

Laskos, L. *et al.* (2004) ‘The RpoH-Mediated Stress Response in *Neisseria gonorrhoeae* Is Regulated at the Level of Activity’, *Journal of Bacteriology*, 186(24), pp. 8443–8452. doi:10.1128/JB.186.24.8443-8452.2004.

Lee, E.H. and Shafer, W.M. (1999) ‘The farAB-encoded efflux pump mediates resistance of gonococci to long-chained antibacterial fatty acids’, *Molecular Microbiology*, 33(4), pp. 839–845. doi:10.1046/j.1365-2958.1999.01530.x.

Lee, H.J. *et al.* (2018) ‘Integrated pathogen load and dual transcriptome analysis of systemic host-pathogen interactions in severe malaria’, *Science Translational Medicine*, 10(447), p. eaar3619. doi:10.1126/scitranslmed.aar3619.

Li, R. *et al.* (2009) ‘SOAP2: an improved ultrafast tool for short read alignment’, *Bioinformatics*, 25(15), pp. 1966–1967. doi:10.1093/bioinformatics/btp336.

Linde, J. *et al.* (2010) ‘Regulatory network modelling of iron acquisition by a fungal pathogen in contact with epithelial cells’, *BMC Systems Biology*, 4(1), p. 148. doi:10.1186/1752-0509-4-148.

Little, J. W. (2006) Gonorrhea: update. *Oral Surg., Oral Med., Oral Pathol., Oral Radiol., Endodont.* 101, 137–143.

Love, M.I., Huber, W. and Anders, S. (2014) ‘Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2’, *Genome Biology*, 15(12), p. 550. doi:10.1186/s13059-014-0550-8.

Lovegrove, F.E. *et al.* (2006) ‘Simultaneous host and parasite expression profiling identifies tissue-specific transcriptional programs associated with susceptibility or resistance to experimental cerebral malaria’, *BMC Genomics*, 7(1), p. 295. doi:10.1186/1471-2164-7-295.

Malinverni, J.C. and Silhavy, T.J. (2009) ‘An ABC transport system that maintains lipid asymmetry in the Gram-negative outer membrane’, *Proceedings of the National Academy of Sciences*, 106(19), pp. 8009–8014. doi:10.1073/pnas.0903229106.

Mandlik, A. *et al.* (2011) ‘RNA-Seq-Based Monitoring of Infection-Linked Changes in Vibrio cholerae Gene Expression’, *Cell Host & Microbe*, 10(2), pp. 165–174. doi:10.1016/j.chom.2011.07.007.

Mandrell, R.E. (1992) ‘Further antigenic similarities of Neisseria gonorrhoeae lipooligosaccharides and human glycosphingolipids’, Infect. Immun., 60(7): 3017–3020.

Margulies, M. *et al.* (2005) ‘Genome sequencing in microfabricated high-density picolitre reactors’, *Nature*, 437(7057), pp. 376–380. doi:10.1038/nature03959.

Marsh, J.W. *et al.* (2017) ‘Bioinformatic analysis of bacteria and host cell dual RNA-sequencing experiments’, *Briefings in Bioinformatics* [Preprint]. doi:10.1093/bib/bbx043.

Maurakis, S. *et al.* (2019) ‘The novel interaction between Neisseria gonorrhoeae TdfJ and human S100A7 allows gonococci to subvert host zinc restriction’, *PLOS Pathogens*. Edited by C. Tang, 15(8), p. e1007937. doi:10.1371/journal.ppat.1007937.

Mäurer, A.P. *et al.* (2007) ‘Gene Expression Profiles of Chlamydophila pneumoniae during the Developmental Cycle and Iron Depletion–Mediated Persistence’, *PLoS Pathogens*. Edited by J.N. Engel, 3(6), p. e83. doi:10.1371/journal.ppat.0030083.

Mavromatis, C. (Harris) *et al.* (2015) ‘The co-transcriptome of uropathogenic *E scherichia coli* -infected mouse macrophages reveals new insights into host–pathogen interactions’, *Cellular Microbiology*, 17(5), pp. 730–746. doi:10.1111/cmi.12397.

McClure, R. *et al.* (2015) ‘The Gonococcal Transcriptome during Infection of the Lower Genital Tract in Women’, *PLOS ONE*. Edited by T. Rudel, 10(8), p. e0133982. doi:10.1371/journal.pone.0133982.

McClure, R. *et al.* (2020) ‘Global Network Analysis of Neisseria gonorrhoeae Identifies Coordination between Pathways, Processes, and Regulators Expressed during Human Infection’, *mSystems*. Edited by N. Chia, 5(1). doi:10.1128/mSystems.00729-19.

McClure, R., Tjaden, B. and Genco, C. (2014) ‘Identification of sRNAs expressed by the human pathogen Neisseria gonorrhoeae under disparate growth conditions’, *Frontiers in Microbiology*, 5. doi:10.3389/fmicb.2014.00456.

Mercante, A.D. *et al.* (2012) ‘MpeR Regulates the *mtr* Efflux Locus in Neisseria gonorrhoeae and Modulates Antimicrobial Resistance by an Iron-Responsive Mechanism’, *Antimicrobial Agents and Chemotherapy*, 56(3), pp. 1491–1501. doi:10.1128/AAC.06112-11.

Merrell, D.S. *et al.* (2002) ‘Host-induced epidemic spread of the cholera bacterium’, *Nature*, 417(6889), pp. 642–645. doi:10.1038/nature00778.

Merz, A.J., Enns, C.A. and So, M. (1999a) ‘Type IV pili of pathogenic Neisseriae elicit cortical plaque formation in epithelial cells’, *Molecular Microbiology*, 32(6), pp. 1316–1332. doi:10.1046/j.1365-2958.1999.01459.x.

Merz, A.J., Enns, C.A. and So, M. (1999b) ‘Type IV pili of pathogenic Neisseriae elicit cortical plaque formation in epithelial cells’, *Molecular Microbiology*, 32(6), pp. 1316–1332. doi:10.1046/j.1365-2958.1999.01459.x.

Merz, A.J. and So, M. (1997) ‘Attachment of piliated, Opa- and Opc- gonococci and meningococci to epithelial cells elicits cortical actin rearrangements and clustering of tyrosine-phosphorylated proteins’, *Infection and Immunity*, 65(10), pp. 4341–4349. doi:10.1128/iai.65.10.4341-4349.1997.

Mortazavi, A. *et al.* (2008) ‘Mapping and quantifying mammalian transcriptomes by RNA-Seq’, *Nature Methods*, 5(7), pp. 621–628. doi:10.1038/nmeth.1226.

Motley, S.T. *et al.* (2004) ‘Simultaneous analysis of host and pathogen interactions during an in vivo infection reveals local induction of host acute phase response proteins, a novel bacterial stress response, and evidence of a host-imposed metal ion limited environment: Simultaneous analysis of host-pathogen interactions in vivo’, *Cellular Microbiology*, 6(9), pp. 849–865. doi:10.1111/j.1462-5822.2004.00407.x.

Nagalakshmi, U. *et al.* (2008) ‘The Transcriptional Landscape of the Yeast Genome Defined by RNA Sequencing’, *Science*, 320(5881), pp. 1344–1349. doi:10.1126/science.1158441.

Naumann, M. (1997) ‘*Neisseria gonorrhoeae* Epithelial Cell Interaction Leads to the Activation of the Transcription Factors Nuclear Factor κB and Activator Protein 1 and the Induction of Inflammatory Cytokines’, J. Exp. Med., 186(2): 247–258.

Nudel, K. *et al.* (2018) ‘Transcriptome Analysis of Neisseria gonorrhoeae during Natural Infection Reveals Differential Expression of Antibiotic Resistance Determinants between Men and Women’, *mSphere*. Edited by S.E.F. D’Orazio, 3(3). doi:10.1128/mSphereDirect.00312-18.

Olive, A.J. and Sassetti, C.M. (2016) ‘Metabolic crosstalk between host and pathogen: sensing, adapting and competing’, *Nature Reviews Microbiology*, 14(4), pp. 221–234. doi:10.1038/nrmicro.2016.12.

Oosthuizen, J.L. *et al.* (2011) ‘Dual Organism Transcriptomics of Airway Epithelial Cells Interacting with Conidia of Aspergillus fumigatus’, *PLoS ONE*. Edited by M. Rojas, 6(5), p. e20527. doi:10.1371/journal.pone.0020527.

Patro, R. *et al.* (2017) ‘Salmon provides fast and bias-aware quantification of transcript expression’, *Nature Methods*, 14(4), pp. 417–419. doi:10.1038/nmeth.4197.

Perez, N. *et al.* (2009) ‘A Genome-Wide Analysis of Small Regulatory RNAs in the Human Pathogen Group A Streptococcus’, *PLoS ONE*. Edited by R.K. Aziz, 4(11), p. e7668. doi:10.1371/journal.pone.0007668.

Pérez-Losada, M. *et al.* (2015) ‘Dual Transcriptomic Profiling of Host and Microbiota during Health and Disease in Pediatric Asthma’, *PLOS ONE*. Edited by B.A. Wilson, 10(6), p. e0131819. doi:10.1371/journal.pone.0131819.

Pireddu, L., Leo, S. and Zanetti, G. (2011) ‘SEAL: a distributed short read mapping and duplicate removal tool’, *Bioinformatics*, 27(15), pp. 2159–2160. doi:10.1093/bioinformatics/btr325.

Pisu, D. *et al.* (2020) ‘Dual RNA-Seq of Mtb-Infected Macrophages In Vivo Reveals Ontologically Distinct Host-Pathogen Interactions’, *Cell Reports*, 30(2), pp. 335-350.e4. doi:10.1016/j.celrep.2019.12.033.

van Putten, J.P. (1993) ‘Phase variation of lipopolysaccharide directs interconversion of invasive and immuno-resistant phenotypes of Neisseria gonorrhoeae.’, *The EMBO Journal*, 12(11), pp. 4043–4051. doi:10.1002/j.1460-2075.1993.tb06088.x.

van Putten, J.P.M., Duensing, T.D. and Carlson, J. (1998) ‘Gonococcal Invasion of Epithelial Cells Driven by P.IA, a Bacterial Ion Channel with GTP Binding Properties’, *Journal of Experimental Medicine*, 188(5), pp. 941–952. doi:10.1084/jem.188.5.941.

Quillin, S.J. and Seifert, H.S. (2018) ‘Neisseria gonorrhoeae host adaptation and pathogenesis’, *Nature Reviews Microbiology*, 16(4), pp. 226–240. doi:10.1038/nrmicro.2017.169.

‘Redefined Nomenclature for Members of the Carcinoembryonic Antigen Family’ (1999) *Experimental Cell Research*, 252(2), pp. 243–249. doi:10.1006/excr.1999.4610.

Remmele, C.W. *et al.* (2014) ‘Transcriptional landscape and essential genes of Neisseria gonorrhoeae’, *Nucleic Acids Research*, 42(16), pp. 10579–10595. doi:10.1093/nar/gku762.

Revel, A.T., Talaat, A.M. and Norgard, M.V. (2002) ‘DNA microarray analysis of differential gene expression in Borrelia burgdorferi, the Lyme disease spirochete’, *Proceedings of the National Academy of Sciences*, 99(3), pp. 1562–1567. doi:10.1073/pnas.032667699.

Rheinwald, J.G. and Anderson, D.J. (no date) ‘Generation of Papillomavirus-lmmortalized Cell Lines from Normal Human Ectocervical, Endocervical, and Vaginal Epithelium That Maintain Expression of Tissue-Specific Differentiation Proteins’’, p. 9.

Rienksma, R.A. *et al.* (2015) ‘Comprehensive insights into transcriptional adaptation of intracellular mycobacteria by microbe-enriched dual RNA sequencing’, *BMC Genomics*, 16(1), p. 34. doi:10.1186/s12864-014-1197-2.

Risso, D. *et al.* (2014) ‘Normalization of RNA-seq data using factor analysis of control genes or samples’, *Nature Biotechnology*, 32(9), pp. 896–902. doi:10.1038/nbt.2931.

Robinson, M.D., McCarthy, D.J. and Smyth, G.K. (2010) ‘edgeR: a Bioconductor package for differential expression analysis of digital gene expression data’, *Bioinformatics*, 26(1), pp. 139–140. doi:10.1093/bioinformatics/btp616.

Rosenberger, C.M. *et al.* (2000) ‘*Salmonella typhimurium* Infection and Lipopolysaccharide Stimulation Induce Similar Changes in Macrophage Gene Expression’, *The Journal of Immunology*, 164(11), pp. 5894–5904. doi:10.4049/jimmunol.164.11.5894.

Rowley, J. *et al.* (2019) ‘Chlamydia, gonorrhoea, trichomoniasis and syphilis: global prevalence and incidence estimates, 2016’, *Bulletin of the World Health Organization*, 97(8), pp. 548-562P. doi:10.2471/BLT.18.228486.

Sandstrom, I. (1987) Etiology and diagnosis of neonatal conjunctivitis. *Acta Paediatr. Scand.* 76, 221–227.

Scardoni, G. *et al.* (2014) ‘Node Interference and Robustness: Performing Virtual Knock-Out Experiments on Biological Networks: The Case of Leukocyte Integrin Activation Network’, *PLoS ONE*. Edited by P. Holme, 9(2), p. e88938. doi:10.1371/journal.pone.0088938.

Schulte, L.N. *et al.* (2011) ‘Analysis of the host microRNA response to *Salmonella* uncovers the control of major cytokines by the *let-7* family: MicroRNA and bacterial infection’, *The EMBO Journal*, 30(10), pp. 1977–1989. doi:10.1038/emboj.2011.94.

Seelbinder, B. *et al.* (2020) ‘Triple RNA-Seq Reveals Synergy in a Human Virus-Fungus Co-infection Model’, *Cell Reports*, 33(7), p. 108389. doi:10.1016/j.celrep.2020.108389.

Sergushichev, A.A. (2016) ‘An algorithm for fast preranked gene set enrichment analysis using cumulative statistic calculation’, p. 9.

Shafer, W.M. and Ohneck, E.A. (2011) ‘Taking the Gonococcus-Human Relationship to a Whole New Level: Implications for the Coevolution of Microbes and Humans’, *mBio*, 2(3). doi:10.1128/mBio.00067-11.

Sharma, C.M. *et al.* (2010) ‘The primary transcriptome of the major human pathogen Helicobacter pylori’, *Nature*, 464(7286), pp. 250–255. doi:10.1038/nature08756.

Shell, D.M. *et al.* (2002) ‘The *Neisseria* Lipooligosaccharide-Specific α-2,3-Sialyltransferase Is a Surface-Exposed Outer Membrane Protein’, *Infection and Immunity*, 70(7), pp. 3744–3751. doi:10.1128/IAI.70.7.3744-3751.2002.

Smyth, G.K. (2005) ‘limma: Linear Models for Microarray Data’, in Gentleman, R. et al. (eds) *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*. New York: Springer-Verlag (Statistics for Biology and Health), pp. 397–420. doi:10.1007/0-387-29362-0_23.

Sperandio, V., Torres, A.G. and Kaper, J.B. (2002) ‘Quorum sensing *Escherichia coli* regulators B and C (QseBC): a novel two-component regulatory system involved in the regulation of flagella and motility by quorum sensing in *E. coli*: QseBC regulates flagella and motility in *E. coli*’, *Molecular Microbiology*, 43(3), pp. 809–821. doi:10.1046/j.1365-2958.2002.02803.x.

Srikhanta, Y.N. *et al.* (2009) ‘Phasevarions Mediate Random Switching of Gene Expression in Pathogenic Neisseria’, *PLoS Pathogens*. Edited by H.S. Seifert, 5(4), p. e1000400. doi:10.1371/journal.ppat.1000400.

Stephens, M. (2016) ‘False discovery rates: a new deal’, *Biostatistics*, p. kxw041. doi:10.1093/biostatistics/kxw041.

Stork, M. *et al.* (2010) 'An Outer Membrane Receptor of Neisseria meningitidis Involved in Zinc Acquisition with Vaccine Potential', *PLoS Pathogens*. Edited by H.S. Seifert, 6(7), p. e1000969. doi:10.1371/journal.ppat.1000969.

Stork, M. *et al.* (2013) 'Zinc Piracy as a Mechanism of Neisseria meningitidis for Evasion of Nutritional Immunity', *PLoS Pathogens*. Edited by X. Nassif, 9(10), p. e1003733. doi:10.1371/journal.ppat.1003733.

Subramanian, A. *et al.* (2005) 'Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles', *Proceedings of the National Academy of Sciences*, 102(43), pp. 15545–15550. doi:10.1073/pnas.0506580102.

Tierney, L. *et al.* (2012) 'An Interspecies Regulatory Network Inferred from Simultaneous RNA-seq of Candida albicans Invading Innate Immune Cells', *Frontiers in Microbiology*, 3. doi:10.3389/fmicb.2012.00085.

Touzeau, C. *et al.* (2016) 'BH3 profiling identifies heterogeneous dependency on Bcl-2 family members in multiple myeloma and predicts sensitivity to BH3 mimetics', *Leukemia*, 30(3), pp. 761–764. doi:10.1038/leu.2015.184.

Trapnell, C. *et al.* (2013) 'Differential analysis of gene regulation at transcript resolution with RNA-seq', *Nature Biotechnology*, 31(1), pp. 46–53. doi:10.1038/nbt.2450.

Trapnell, C., Pachter, L. and Salzberg, S.L. (2009) 'TopHat: discovering splice junctions with RNA-Seq', *Bioinformatics*, 25(9), pp. 1105–1111. doi:10.1093/bioinformatics/btp120.

Unemo, M. *et al.* (2016) 'The novel 2016 WHO *Neisseria gonorrhoeae* reference strains for global quality assurance of laboratory investigations: phenotypic, genetic and reference genome characterization', *Journal of Antimicrobial Chemotherapy*, 71(11), pp. 3096–3108. doi:10.1093/jac/dkw288.

Unemo, M. *et al.* (2019) 'Gonorrhoea', *Nature Reviews Disease Primers*, 5(1), p. 79. doi:10.1038/s41572-019-0128-6.

Unemo, M. and Shafer, W.M. (2014) 'Antimicrobial Resistance in Neisseria gonorrhoeae in the 21st Century: Past, Evolution, and Future', *Clinical Microbiology Reviews*, 27(3), pp. 587–613. doi:10.1128/CMR.00010-14.

Van Putten, J.P.M., Duensing, T.D. and Cole, R.L. (1998) 'Entry of OpaA $^+$ gonococci into HEp-2 cells requires concerted action of glycosaminoglycans, fibronectin and integrin receptors', *Molecular Microbiology*, 29(1), pp. 369–379. doi:10.1046/j.1365-2958.1998.00951.x.

Vannucci, F.A., Foster, D.N. and Gebhart, C.J. (2013) 'Laser microdissection coupled with RNA-seq analysis of porcine enterocytes infected with an obligate intracellular pathogen (Lawsonia intracellularis)', *BMC Genomics*, 14(1), p. 421. doi:10.1186/1471-2164-14-421.

Wadsworth, C.B. *et al.* (2018) 'Azithromycin Resistance through Interspecific Acquisition of an Epistasis-Dependent Efflux Pump Component and Transcriptional Regulator in

Neisseria gonorrhoeae', *mBio*. Edited by W. Shafer and M.S. Gilmore, 9(4). doi:10.1128/mBio.01419-18.

Warner, D.M. *et al.* (2007) 'Regulation of the MtrC-MtrD-MtrE Efflux-Pump System Modulates the In Vivo Fitness of *Neisseria gonorrhoeae*', *The Journal of Infectious Diseases*, 196(12), pp. 1804–1812. doi:10.1086/522964.

Weinstock, H. and Workowski, K.A. (2009) 'Pharyngeal Gonorrhea: An Important Reservoir of Infection?', *Clinical Infectious Diseases*, 49(12), pp. 1798–1800. doi:10.1086/648428.

Westermann, A.J. *et al.* (2016) 'Dual RNA-seq unveils noncoding RNA functions in host–pathogen interactions', *Nature*, 529(7587), pp. 496–501. doi:10.1038/nature16547.

Westermann, A.J., Barquist, L. and Vogel, J. (2017) 'Resolving host–pathogen interactions by dual RNA-seq', *PLOS Pathogens*. Edited by J.B. Bliska, 13(2), p. e1006033. doi:10.1371/journal.ppat.1006033.

Williams, C.R. *et al.* (2016) 'Trimming of sequence reads alters RNA-Seq gene expression estimates', *BMC Bioinformatics*, 17(1), p. 103. doi:10.1186/s12859-016-0956-2.

Winzer, K. *et al.* (2002) 'Role of *Neisseria meningitidis luxS* in Cell-to-Cell Signaling and Bacteremic Infection', *Infection and Immunity*, 70(4), pp. 2245–2248. doi:10.1128/IAI.70.4.2245-2248.2002.

Xiong, M. *et al.* (2016) 'Analysis of the sex ratio of reported gonorrhoea incidence in Shenzhen, China', *BMJ Open*, 6(3), p. e009629. doi:10.1136/bmjopen-2015-009629.

Yamasaki, R. *et al.* (1999) 'Structural and Immunochemical Characterization of aNeisseria gonorrhoeae Epitope Defined by a Monoclonal Antibody 2C7; the Antibody Recognizes a Conserved Epitope on Specific Lipo-oligosaccharides in Spite of the Presence of Human Carbohydrate Epitopes', *Journal of Biological Chemistry*, 274(51), pp. 36550–36558. doi:10.1074/jbc.274.51.36550.

Yao, J. *et al.* (2016) 'Activation of Exogenous Fatty Acids to Acyl-Acyl Carrier Protein Cannot Bypass FabI Inhibition in Neisseria', *Journal of Biological Chemistry*, 291(1), pp. 171–181. doi:10.1074/jbc.M115.699462.

Yevshin, I. *et al.* (2019) 'GTRD: a database on gene transcription regulation—2019 update', *Nucleic Acids Research*, 47(D1), pp. D100–D105. doi:10.1093/nar/gky1128.

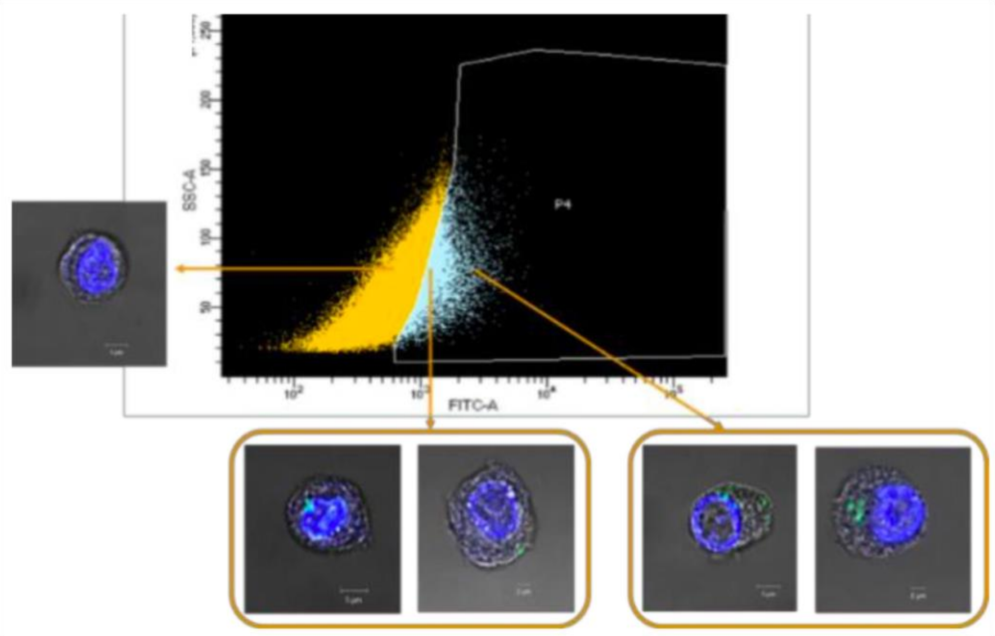
Yu, C. *et al.* (2016) 'Characterization of the Neisseria gonorrhoeae Iron and Fur Regulatory Network', *Journal of Bacteriology*. Edited by V.J. DiRita, 198(16), pp. 2180–2191. doi:10.1128/JB.00166-16.

Zheng, X. *et al.* (2011) 'Identification of Genes and Genomic Islands Correlated with High Pathogenicity in Streptococcus suis Using Whole Genome Tilling Microarrays', *PLoS ONE*. Edited by D. Dean, 6(3), p. e17987. doi:10.1371/journal.pone.0017987.

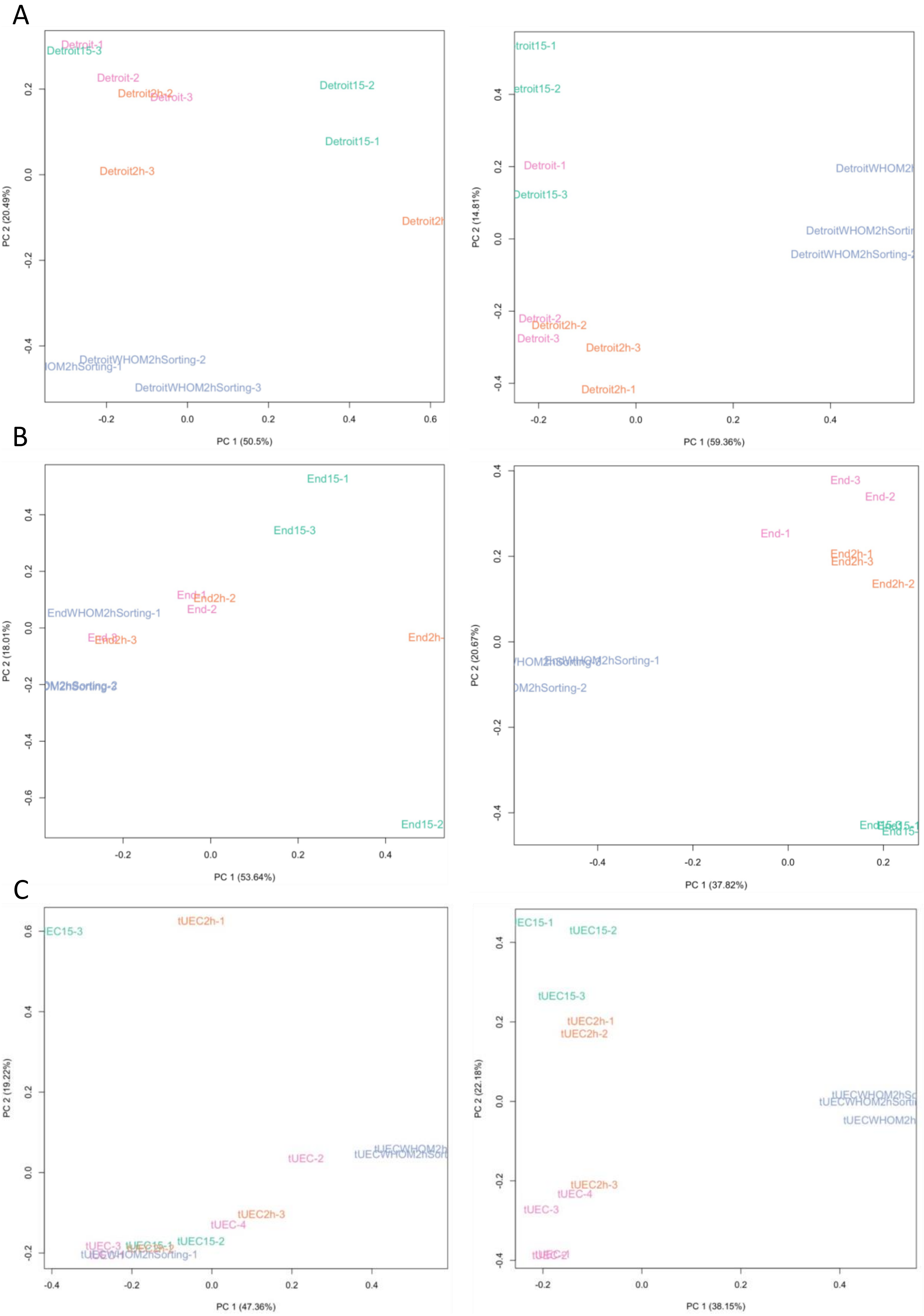
Zhu, A., Ibrahim, J.G. and Love, M.I. (2019) ‘Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences’, *Bioinformatics*. Edited by O. Stegle, 35(12), pp. 2084–2092. doi:10.1093/bioinformatics/bty895.

Zhu, S. *et al.* (2013) ‘A quantitative model of transcriptional differentiation driving host-pathogen interactions’, *Briefings in Bioinformatics*, 14(6), pp. 713–723. doi:10.1093/bib/bbs047.

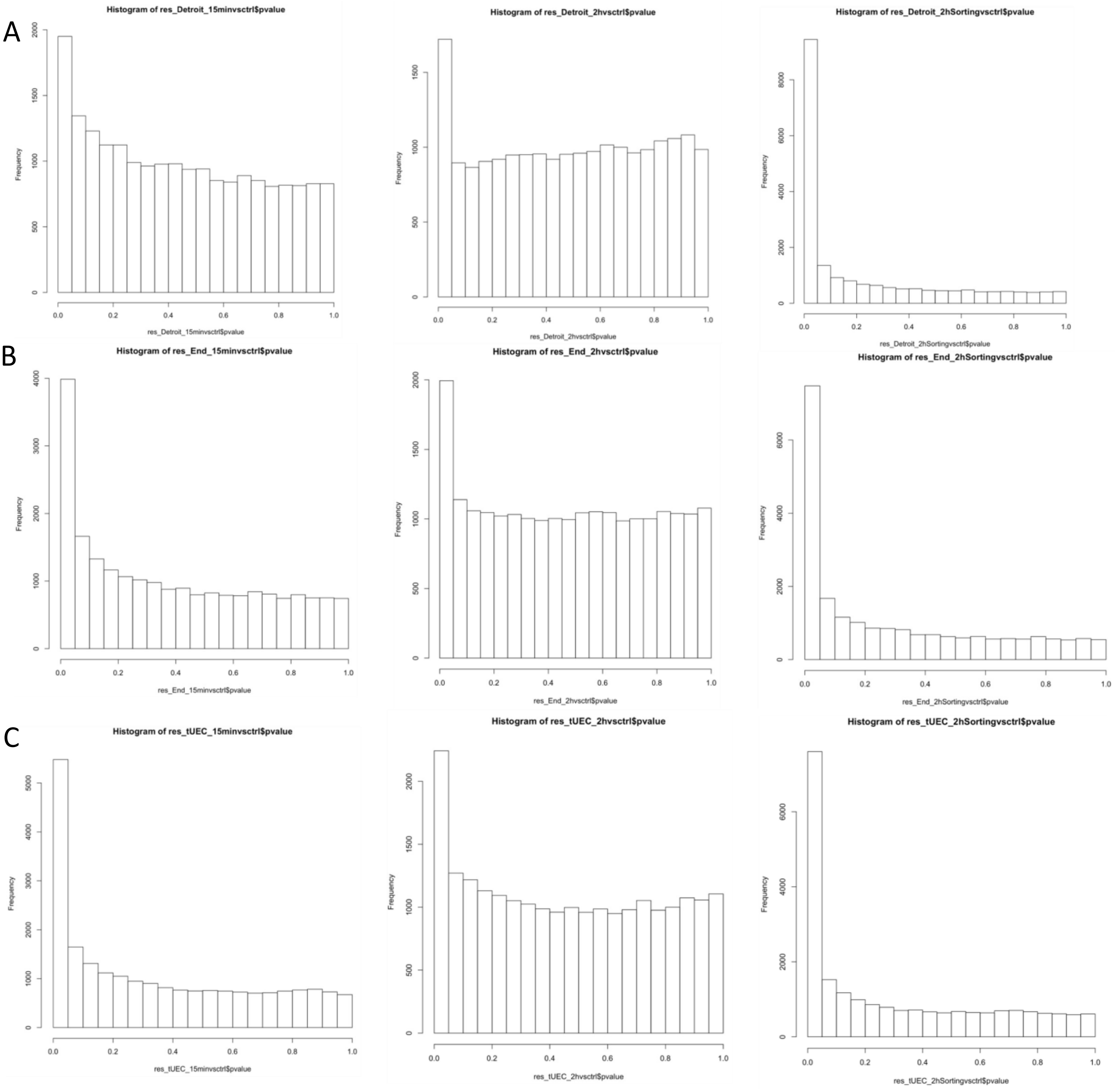
SUPPLEMENTARY DATA

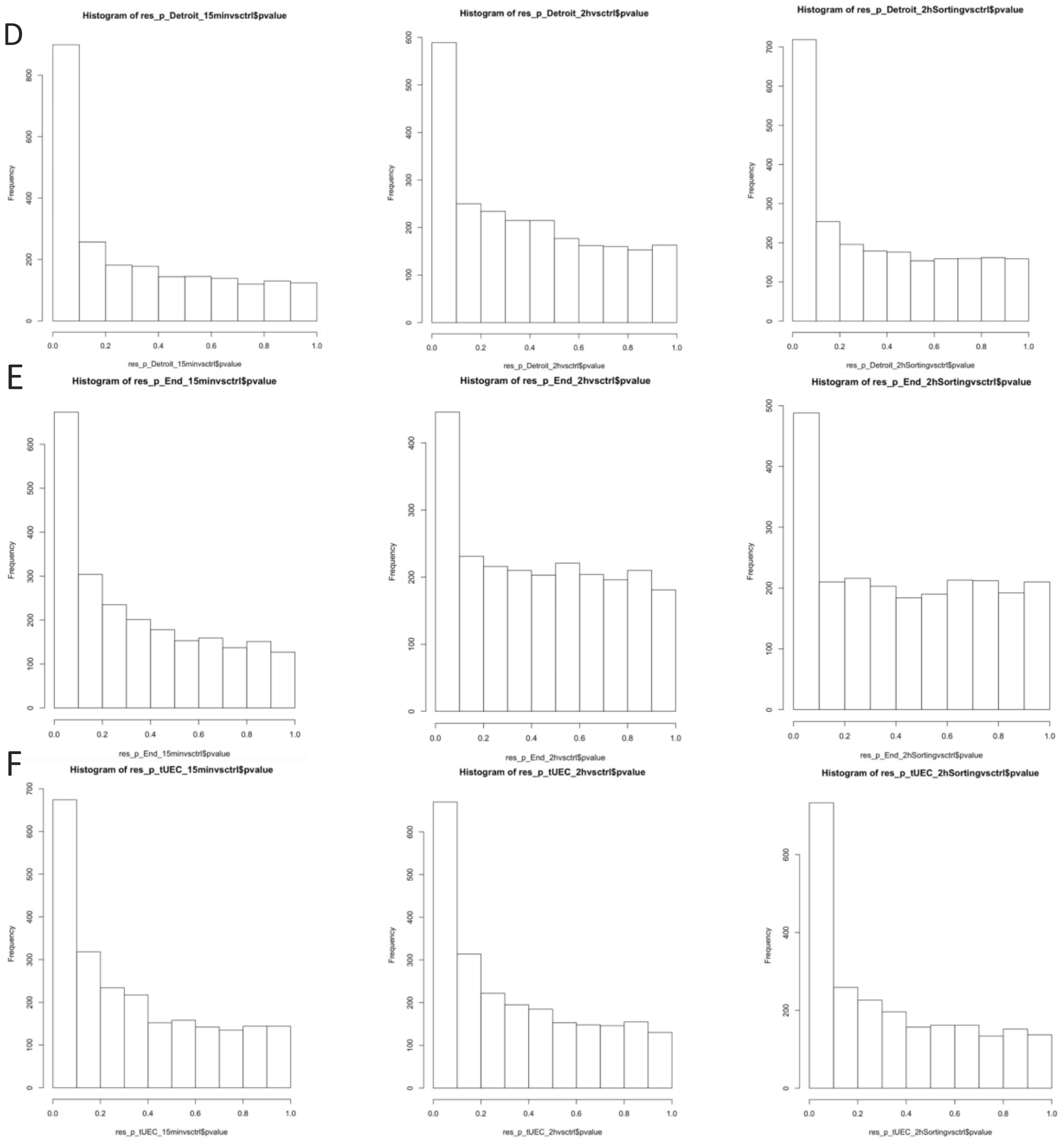


Supplementary Figure 1 | Confocal microscopy image of sorted cells. Representative image of End1 cells isolated by bacterial fluorescence upon 120 minutes of infection. Cells isolated through FACSAria II device from different areas of the dot plot were analyzed by confocal microscopy to verify the accuracy of the sorting. Cells from gate P4 (*Neisseria gonorrhoeae*-infected cells gate) showed the presence of intracellular bacteria (blue: nuclei; green: bacteria; bright field: eukaryotic cells).

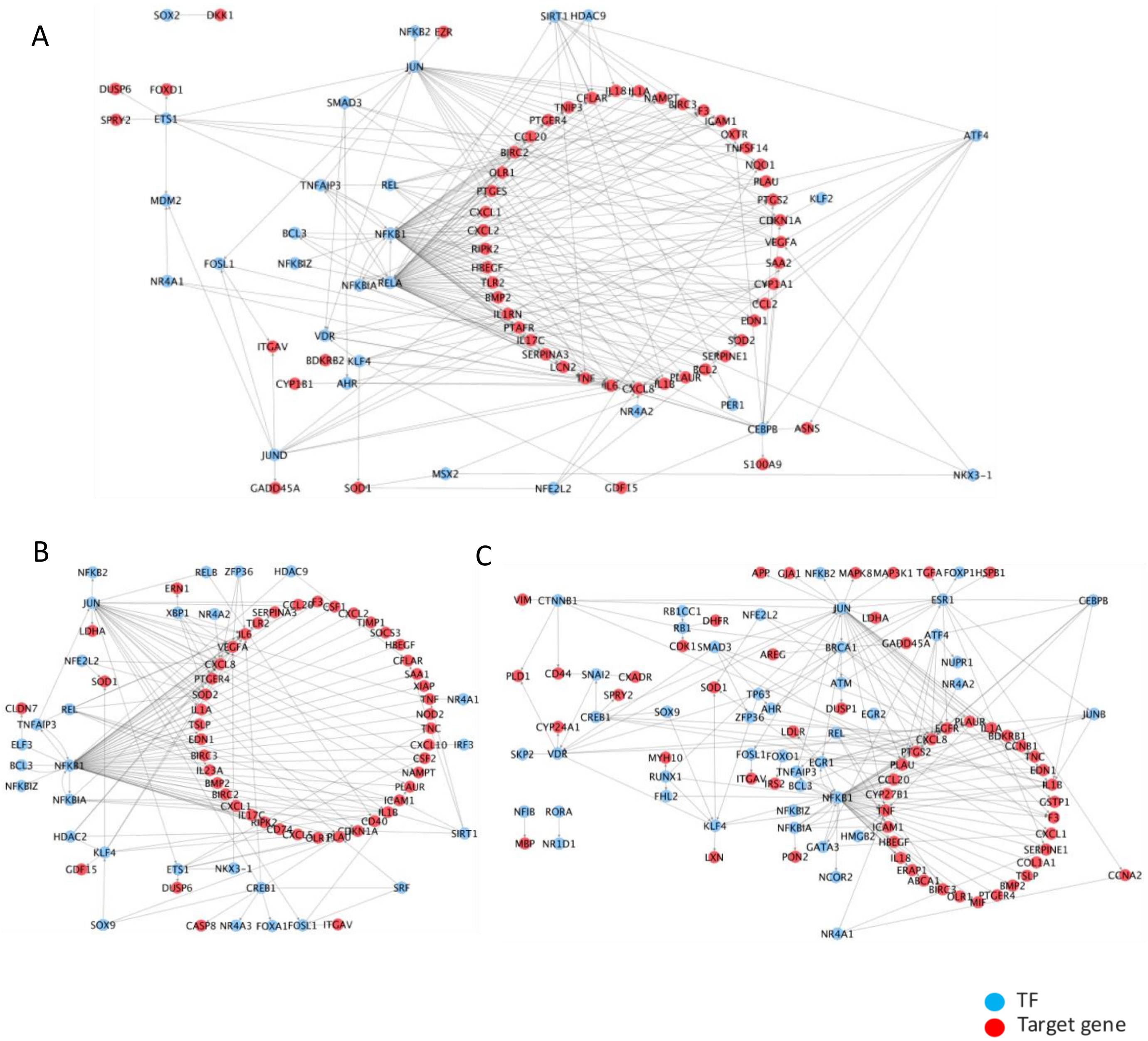


Supplementary Figure 2 | PCA analysis on host data before and after RUV normalization. A-C) PCA plots representing host samples before (on the left) and after (on the right) applying RUV normalization for Detroit562 (**A**), End1 (**B**) and t-UEC (**C**) infections.

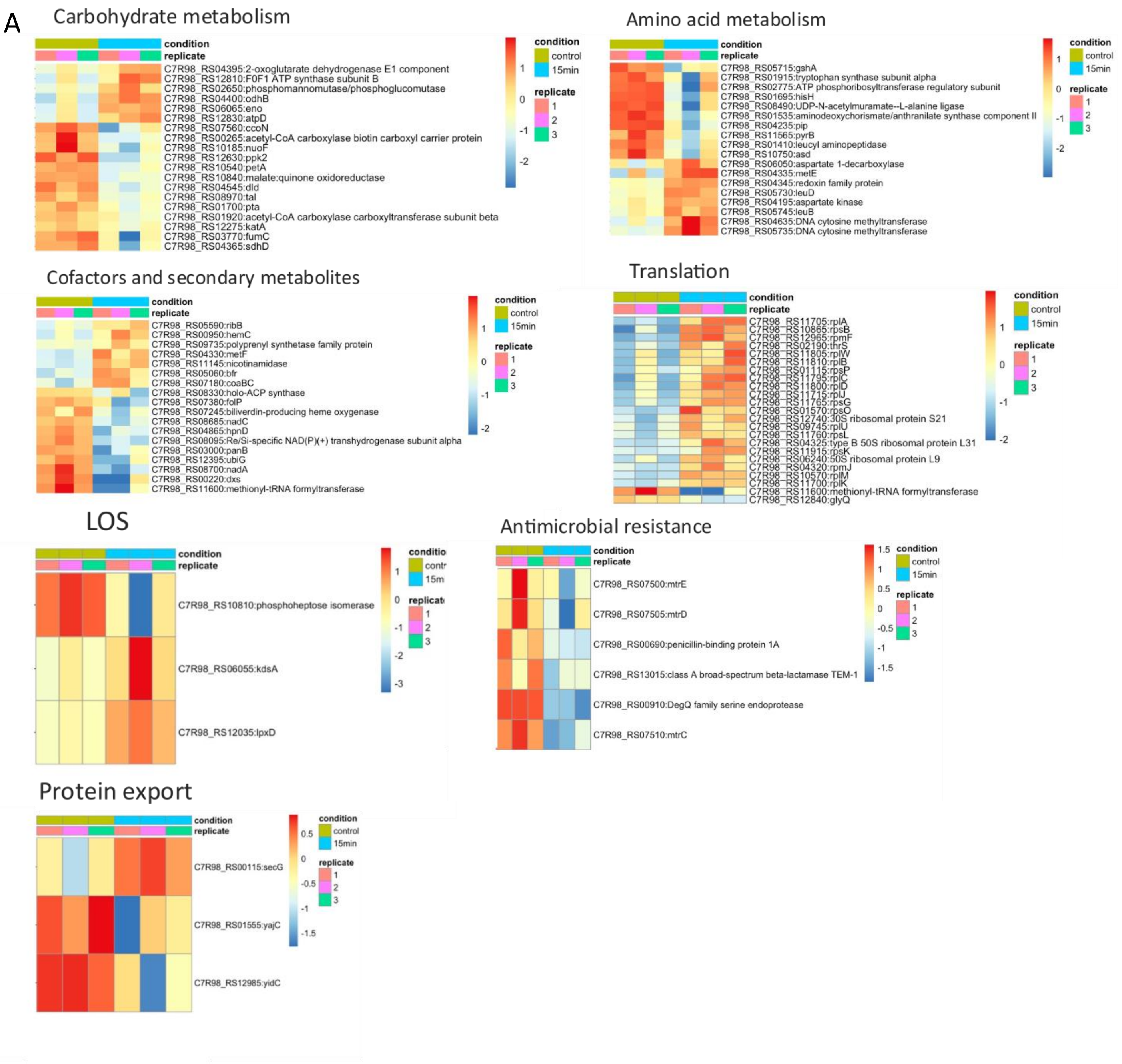

A
Histogram of res_Detroit_15minvsctrl$pvalue
Frequency
res_Detroit_15minvsctrl$pvalue
Histogram of res_Detroit_2hvsctrl$pvalue
Frequency
res_Detroit_2hvsctrl$pvalue
Histogram of res_Detroit_2hSortingvsctrl$pvalue
Frequency
res_Detroit_2hSortingvsctrl$pvalue
B
Histogram of res_End_15minvsctrl$pvalue
Frequency
res_End_15minvsctrl$pvalue
Histogram of res_End_2hvsctrl$pvalue
Frequency
res_End_2hvsctrl$pvalue
Histogram of res_End_2hSortingvsctrl$pvalue
Frequency
res_End_2hSortingvsctrl$pvalue
C
Histogram of res_tUEC_15minvsctrl$pvalue
Frequency
res_tUEC_15minvsctrl$pvalue
Histogram of res_tUEC_2hvsctrl$pvalue
Frequency
res_tUEC_2hvsctrl$pvalue
Histogram of res_tUEC_2hSortingvsctrl$pvalue
Frequency
res_tUEC_2hSortingvsctrl$pvalue
0.0
0.2
0.4
0.6
0.8
1.0

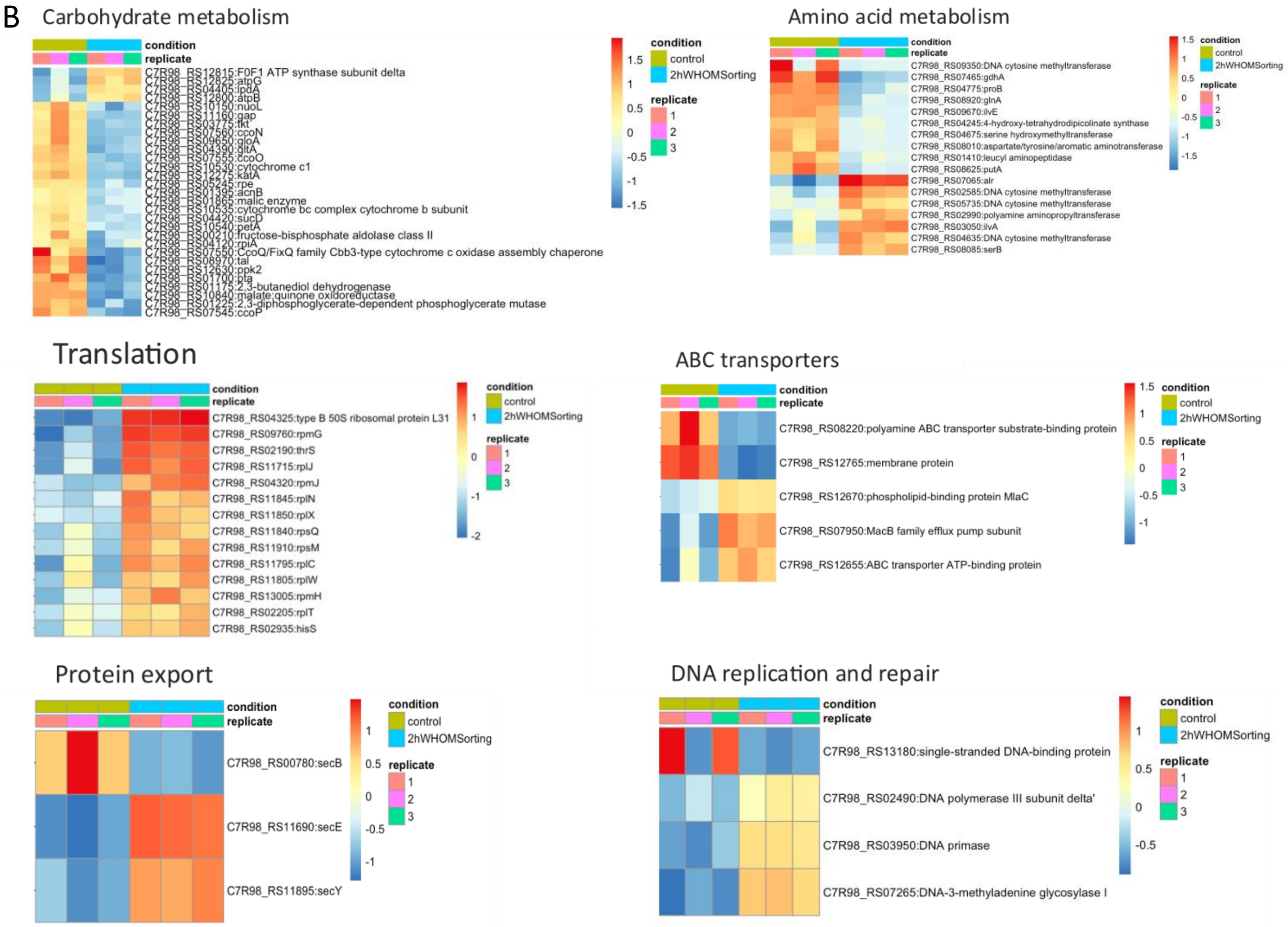


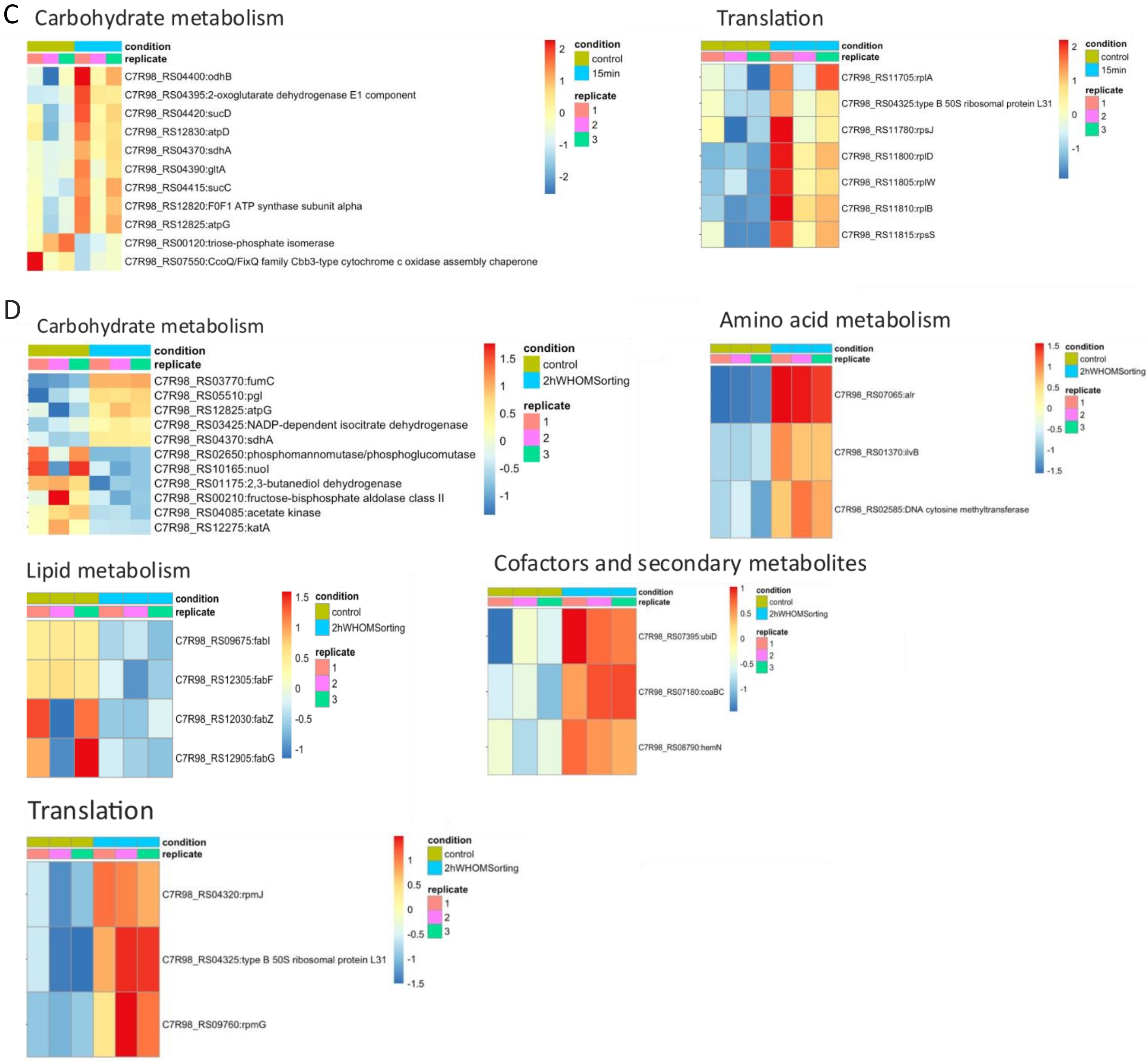
Supplementary Figure 3 | Histograms of p-values from DE analysis. A-F) Histograms of p-values from DE testing at each time point compared to controls for Detroit562 cells (**A**), End1 cells (**B**), t-UEC cells (**C**), pathogens infecting Detroit562 cells (**D**), pathogens infecting End1 cells (**E**) and pathogens infecting t-UEC cells (**F**). For each host cell line or pathogen infection, from left to right, the considered time points are 15mpi, 120mpi and 120mpi upon sorting.

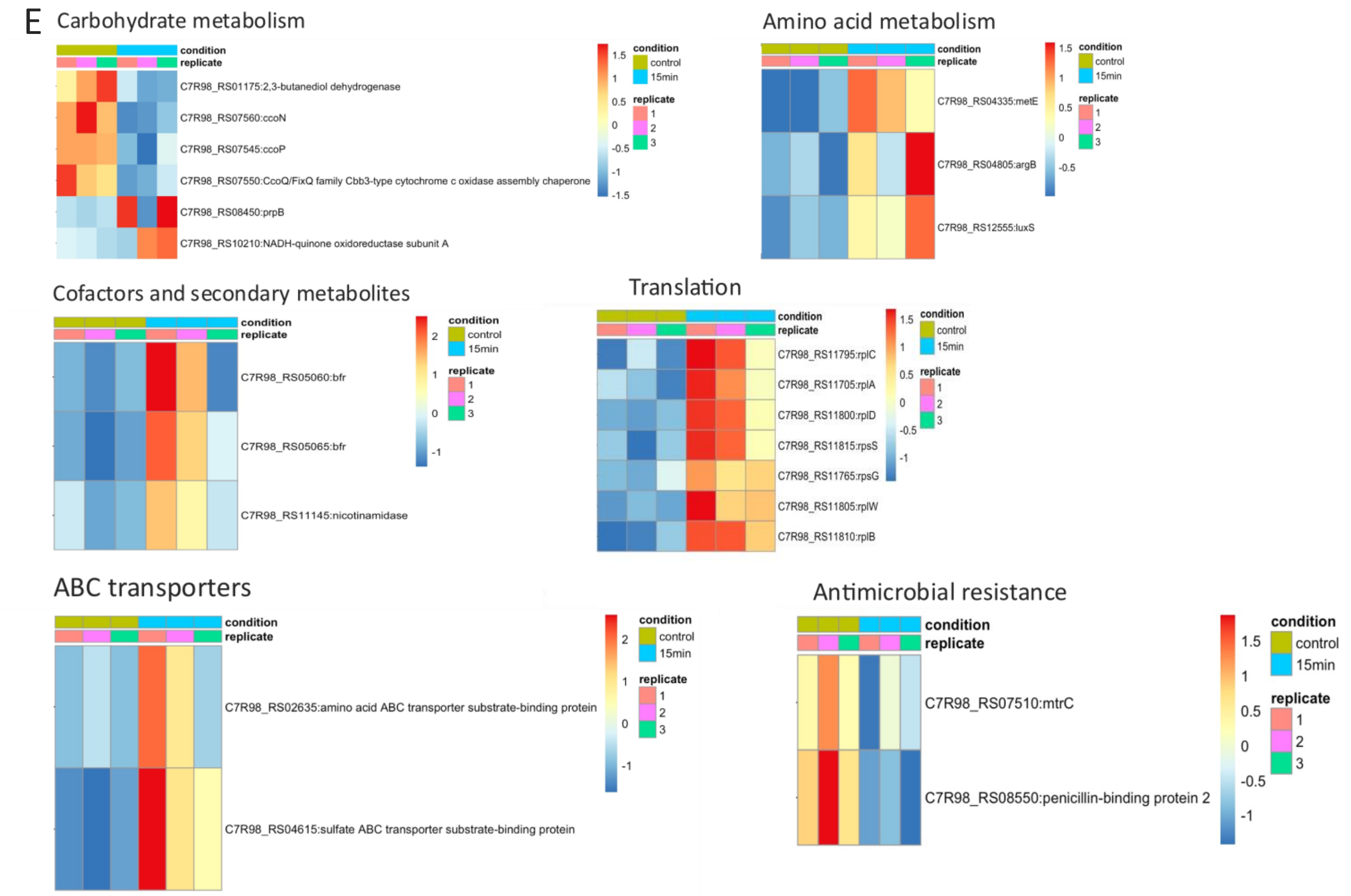


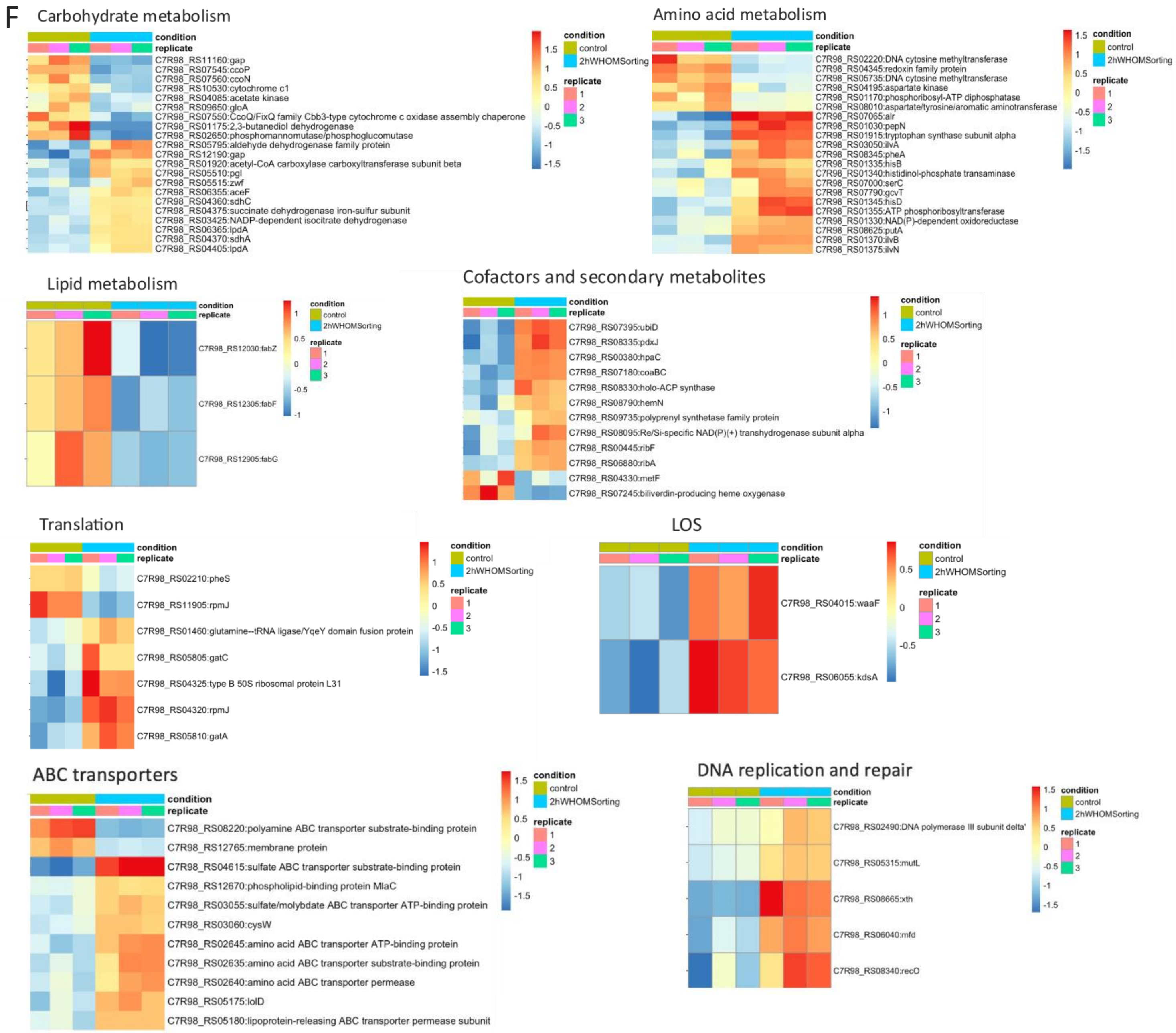
Supplementary Figure 4 | Regulatory networks behind the host response at 120mpi. A-C) TF-target genes regulatory subnetwork for genes in the leading edge of the top significantly enriched GO biological processes with positive normalized enrichment score (NES) shared by all the cells in GSEA upon 120 minutes of infection. Regulatory connections among such genes are reported for Detroit562 cells (**A**), End1 cells (**B**) and t-UEC cells (**C**). Genes reported as TF in TRRUST version 2 database are colored in blue, the ones reported as target genes are colored in red. Visualization is performed through Cytoscape.


A
Carbohydrate metabolism
condition
replicate
C7R98_RS04395:2-oxoglutarate dehydrogenase E1 component
C7R98_RS12810:F0F1 ATP synthase subunit B
C7R98_RS02650:phosphomannomutase/phosphoglucomutase
C7R98_RS04400:odhB
C7R98_RS06065:eno
C7R98_RS12830:atpD
C7R98_RS07560:ccoN
C7R98_RS00265:acetyl-CoA carboxylase biotin carboxyl carrier protein
C7R98_RS10185:nuoF
C7R98_RS12630:ppk2
C7R98_RS10540:petA
C7R98_RS10840:malate:quinone oxidoreductase
C7R98_RS04545:dld
C7R98_RS08970:tal
C7R98_RS01700:pta
C7R98_RS01920:acetyl-CoA carboxylase carboxyltransferase subunit beta
C7R98_RS12275:katA
C7R98_RS03770:fumC
C7R98_RS04365:sdhD
1
0
-1
-2
condition
control
15min
replicate
1
2
3
Amino acid metabolism
condition
replicate
C7R98_RS05715:gshA
C7R98_RS01915:tryptophan synthase subunit alpha
C7R98_RS02775:ATP phosphoribosyltransferase regulatory subunit
C7R98_RS01695:hisH
C7R98_RS08490:UDP-N-acetylmuramate--L-alanine ligase
C7R98_RS01535:aminodeoxychorismate/anthranilate synthase component II
C7R98_RS04235:pip
C7R98_RS11565:pyrB
C7R98_RS01410:leucyl aminopeptidase
C7R98_RS10750:asd
C7R98_RS06050:aspartate 1-decarboxylase
C7R98_RS04335:metE
C7R98_RS04345:redoxin family protein
C7R98_RS05730:leuD
C7R98_RS04195:aspartate kinase
C7R98_RS05745:leuB
C7R98_RS04635:DNA cytosine methyltransferase
C7R98_RS05735:DNA cytosine methyltransferase
1
0
-1
-2
condition
control
15min
replicate
1
2
3
Cofactors and secondary metabolites
condition
replicate
C7R98_RS05590:ribB
C7R98_RS00950:hemC
C7R98_RS09735:polyprenyl synthetase family protein
C7R98_RS04330:metF
C7R98_RS11145:nicotinamidase
C7R98_RS05060:bfr
C7R98_RS07180:coaBC
C7R98_RS08330:holo-ACP synthase
C7R98_RS07380:folP
C7R98_RS07245:biliverdin-producing heme oxygenase
C7R98_RS08685:nadC
C7R98_RS04865:hpnD
C7R98_RS08095:Re/Si-specific NAD(P)(+) transhydrogenase subunit alpha
C7R98_RS03000:panB
C7R98_RS12395:ubiG
C7R98_RS08700:nadA
C7R98_RS00220:dxs
C7R98_RS11600:methionyl-tRNA formyltransferase
1
0
-1
-2
condition
control
15min
replicate
1
2
3
Translation
condition
replicate
C7R98_RS11705:rplA
C7R98_RS10865:rpsB
C7R98_RS12965:rpmF
C7R98_RS02190:thrS
C7R98_RS11805:rplW
C7R98_RS11810:rplB
C7R98_RS01115:rpsP
C7R98_RS11795:rplC
C7R98_RS11800:rplD
C7R98_RS11715:rplJ
C7R98_RS11765:rpsG
C7R98_RS01570:rpsO
C7R98_RS12740:30S ribosomal protein S21
C7R98_RS09745:rplU
C7R98_RS11760:rpsL
C7R98_RS04325:type B 50S ribosomal protein L31
C7R98_RS11915:rpsK
C7R98_RS06240:50S ribosomal protein L9
C7R98_RS04320:rpmJ
C7R98_RS10570:rplM
C7R98_RS11700:rplK
C7R98_RS11600:methionyl-tRNA formyltransferase
C7R98_RS12840:glyQ
1
0
-1
-2
condition
control
15min
replicate
1
2
3
LOS
condition
replicate
C7R98_RS10810:phosphoheptose isomerase
C7R98_RS06055:kdsA
C7R98_RS12035:lpxD
1
0
-1
-2
-3
conditio
contr
15m
replicat
1
2
3
Antimicrobial resistance
condition
replicate
C7R98_RS07500:mtrE
C7R98_RS07505:mtrD
C7R98_RS00690:penicillin-binding protein 1A
C7R98_RS13015:class A broad-spectrum beta-lactamase TEM-1
C7R98_RS00910:DegQ family serine endoprotease
C7R98_RS07510:mtrC
1.5
1
0.5
0
-0.5
-1
-1.5
condition
control
15min
replicate
1
2
3
Protein export
condition
replicate
C7R98_RS00115:secG
C7R98_RS01555:yajC
C7R98_RS12985:yidC
0.5
0
-0.5
-1
-1.5
condition
control
15min
replicate
1
2
3

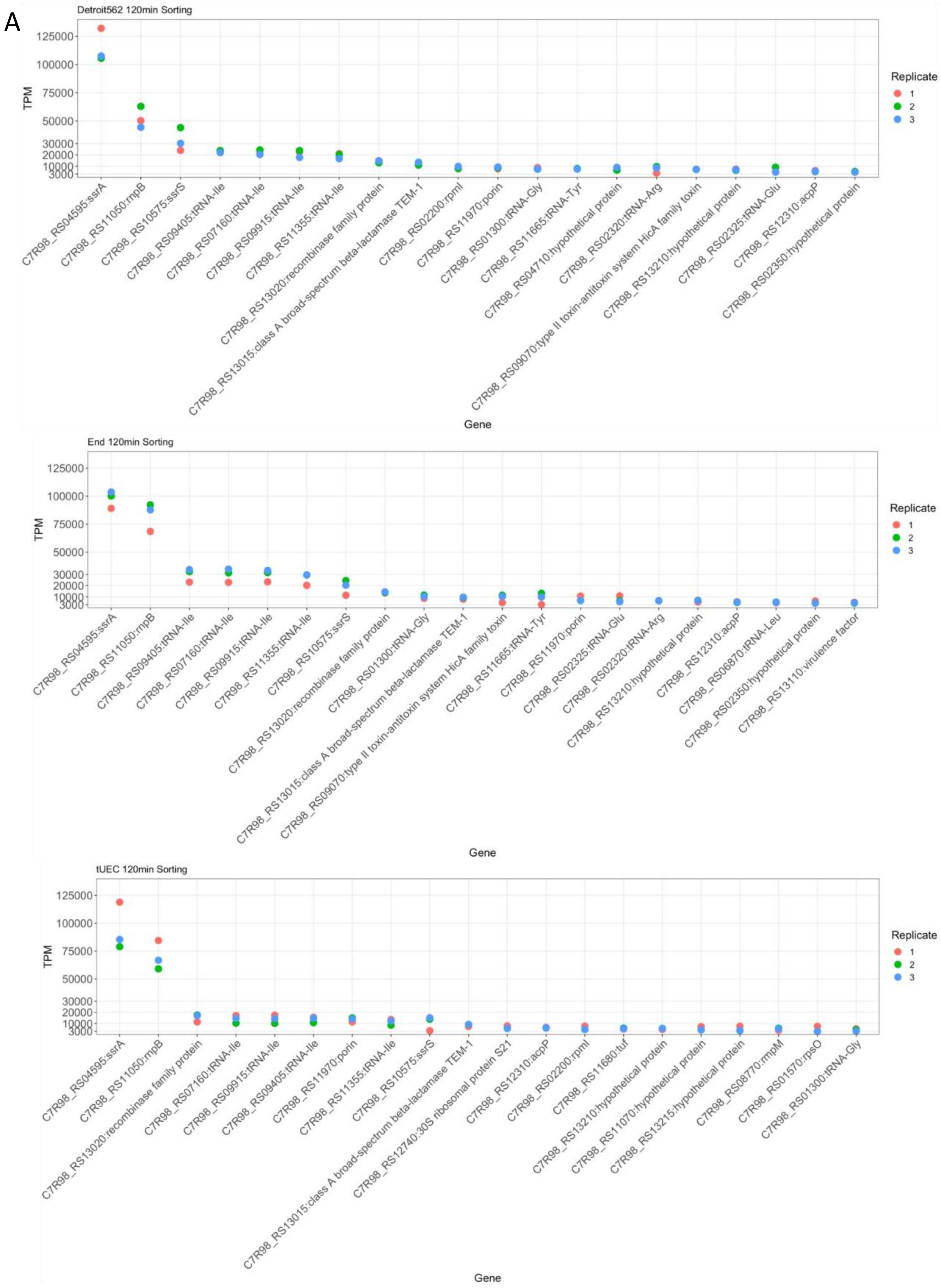

B
Carbohydrate metabolism
condition
replicate
C7R98_RS12815:F0F1 ATP synthase subunit delta
C7R98_RS12825:atpG
C7R98_RS04405:lpdA
C7R98_RS12800:atpB
C7R98_RS10150:nuoL
C7R98_RS11160:gap
C7R98_RS03775:tkt
C7R98_RS07560:ccoN
C7R98_RS09650:gloA
C7R98_RS04390:gltA
C7R98_RS07555:ccoO
C7R98_RS10530:cytochrome c1
C7R98_RS12275:katA
C7R98_RS05245:rpe
C7R98_RS01395:acnB
C7R98_RS01865:malic enzyme
C7R98_RS10535:cytochrome bc complex cytochrome b subunit
C7R98_RS04420:sucD
C7R98_RS10540:petA
C7R98_RS00210:fructose-bisphosphate aldolase class II
C7R98_RS04120:rpiA
C7R98_RS07550:CcoQ/FixQ family Cbb3-type cytochrome c oxidase assembly chaperone
C7R98_RS08970:tal
C7R98_RS12630:ppk2
C7R98_RS01700:pta
C7R98_RS01175:2,3-butanediol dehydrogenase
C7R98_RS10840:malate:quinone oxidoreductase
C7R98_RS01225:2,3-diphosphoglycerate-dependent phosphoglycerate mutase
C7R98_RS07545:ccoP
1.5
1
0.5
0
-0.5
-1
-1.5
condition
control
2hWHOMSorting
replicate
1
2
3
Amino acid metabolism
condition
replicate
C7R98_RS09350:DNA cytosine methyltransferase
C7R98_RS07465:gdhA
C7R98_RS04775:proB
C7R98_RS08920:glnA
C7R98_RS09670:ilvE
C7R98_RS04245:4-hydroxy-tetrahydrodipicolinate synthase
C7R98_RS04675:serine hydroxymethyltransferase
C7R98_RS08010:aspartate/tyrosine/aromatic aminotransferase
C7R98_RS01410:leucyl aminopeptidase
C7R98_RS08625:putA
C7R98_RS07065:alr
C7R98_RS02585:DNA cytosine methyltransferase
C7R98_RS05735:DNA cytosine methyltransferase
C7R98_RS02990:polyamine aminopropyltransferase
C7R98_RS03050:ilvA
C7R98_RS04635:DNA cytosine methyltransferase
C7R98_RS08085:serB
1.5
1
0.5
0
-0.5
-1
-1.5
condition
control
2hWHOMSorting
replicate
1
2
3
Translation
condition
replicate
C7R98_RS04325:type B 50S ribosomal protein L31
C7R98_RS09760:rpmG
C7R98_RS02190:thrS
C7R98_RS11715:rplJ
C7R98_RS04320:rpmJ
C7R98_RS11845:rplN
C7R98_RS11850:rplX
C7R98_RS11840:rpsQ
C7R98_RS11910:rpsM
C7R98_RS11795:rplC
C7R98_RS11805:rplW
C7R98_RS13005:rpmH
C7R98_RS02205:rplT
C7R98_RS02935:hisS
1
0
-1
-2
condition
control
2hWHOMSorting
replicate
1
2
3
ABC transporters
condition
replicate
C7R98_RS08220:polyamine ABC transporter substrate-binding protein
C7R98_RS12765:membrane protein
C7R98_RS12670:phospholipid-binding protein MlaC
C7R98_RS07950:MacB family efflux pump subunit
C7R98_RS12655:ABC transporter ATP-binding protein
1.5
1
0.5
0
-0.5
-1
condition
control
2hWHOMSorting
replicate
1
2
3
Protein export
condition
replicate
C7R98_RS00780:secB
C7R98_RS11690:secE
C7R98_RS11895:secY
1
0.5
0
-0.5
-1
condition
control
2hWHOMSorting
replicate
1
2
3
DNA replication and repair
condition
replicate
C7R98_RS13180:single-stranded DNA-binding protein
C7R98_RS02490:DNA polymerase III subunit delta'
C7R98_RS03950:DNA primase
C7R98_RS07265:DNA-3-methyladenine glycosylase I
1
0.5
0
-0.5
condition
control
2hWHOMSorting
replicate
1
2
3

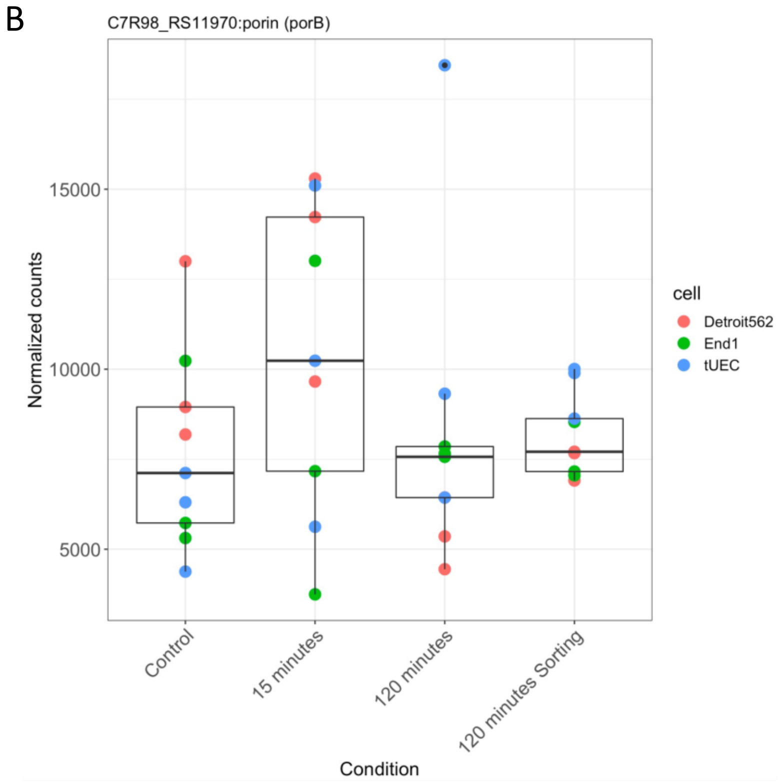
C
Carbohydrate metabolism
condition
replicate
C7R98_RS04400:odhB
C7R98_RS04395:2-oxoglutarate dehydrogenase E1 component
C7R98_RS04420:sucD
C7R98_RS12830:atpD
C7R98_RS04370:sdhA
C7R98_RS04390:gltA
C7R98_RS04415:sucC
C7R98_RS12820:F0F1 ATP synthase subunit alpha
C7R98_RS12825:atpG
C7R98_RS00120:triose-phosphate isomerase
C7R98_RS07550:CcoQ/FixQ family Cbb3-type cytochrome c oxidase assembly chaperone
2
1
0
-1
-2
condition
control
15min
replicate
1
2
3
Translation
condition
replicate
C7R98_RS11705:rplA
C7R98_RS04325:type B 50S ribosomal protein L31
C7R98_RS11780:rpsJ
C7R98_RS11800:rplD
C7R98_RS11805:rplW
C7R98_RS11810:rplB
C7R98_RS11815:rpsS
2
1
0
-1
condition
control
15min
replicate
1
2
3
D
Carbohydrate metabolism
condition
replicate
C7R98_RS03770:fumC
C7R98_RS05510:pgl
C7R98_RS12825:atpG
C7R98_RS03425:NADP-dependent isocitrate dehydrogenase
C7R98_RS04370:sdhA
C7R98_RS02650:phosphomannomutase/phosphoglucomutase
C7R98_RS10165:nuoI
C7R98_RS01175:2,3-butanediol dehydrogenase
C7R98_RS00210:fructose-bisphosphate aldolase class II
C7R98_RS04085:acetate kinase
C7R98_RS12275:katA
1.5
1
0.5
0
-0.5
-1
condition
control
2hWHOMSorting
replicate
1
2
3
Amino acid metabolism
condition
replicate
C7R98_RS07065:alr
C7R98_RS01370:ilvB
C7R98_RS02585:DNA cytosine methyltransferase
1.5
1
0.5
0
-0.5
-1
-1.5
condition
control
2hWHOMSorting
replicate
1
2
3
Lipid metabolism
condition
replicate
C7R98_RS09675:fabI
C7R98_RS12305:fabF
C7R98_RS12030:fabZ
C7R98_RS12905:fabG
1.5
1
0.5
0
-0.5
-1
condition
control
2hWHOMSorting
replicate
1
2
3
Cofactors and secondary metabolites
condition
replicate
C7R98_RS07395:ubiD
C7R98_RS07180:coaBC
C7R98_RS08790:hemN
1
0.5
0
-0.5
-1
condition
control
2hWHOMSorting
replicate
1
2
3
Translation
condition
replicate
C7R98_RS04320:rpmJ
C7R98_RS04325:type B 50S ribosomal protein L31
C7R98_RS09760:rpmG
1
0.5
0
-0.5
-1
-1.5
condition
control
2hWHOMSorting
replicate
1
2
3


E
Carbohydrate metabolism
condition
replicate
C7R98_RS01175:2,3-butanediol dehydrogenase
C7R98_RS07560:ccoN
C7R98_RS07545:ccoP
C7R98_RS07550:CcoQ/FixQ family Cbb3-type cytochrome c oxidase assembly chaperone
C7R98_RS08450:prpB
C7R98_RS10210:NADH-quinone oxidoreductase subunit A
1.5
1
0.5
0
-0.5
-1
-1.5
condition
control
15min
replicate
1
2
3
Amino acid metabolism
condition
replicate
C7R98_RS04335:metE
C7R98_RS04805:argB
C7R98_RS12555:luxS
1.5
1
0.5
0
-0.5
condition
control
15min
replicate
1
2
3
Cofactors and secondary metabolites
condition
replicate
C7R98_RS05060:bfr
C7R98_RS05065:bfr
C7R98_RS11145:nicotinamidase
2
1
0
-1
condition
control
15min
replicate
1
2
3
Translation
condition
replicate
C7R98_RS11795:rplC
C7R98_RS11705:rplA
C7R98_RS11800:rplD
C7R98_RS11815:rpsS
C7R98_RS11765:rpsG
C7R98_RS11805:rplW
C7R98_RS11810:rplB
1.5
1
0.5
0
-0.5
-1
condition
control
15min
replicate
1
2
3
ABC transporters
condition
replicate
C7R98_RS02635:amino acid ABC transporter substrate-binding protein
C7R98_RS04615:sulfate ABC transporter substrate-binding protein
2
1
0
-1
condition
control
15min
replicate
1
2
3
Antimicrobial resistance
condition
replicate
C7R98_RS07510:mtrC
C7R98_RS08550:penicillin-binding protein 2
1.5
1
0.5
0
-0.5
-1
condition
control
15min
replicate
1
2
3

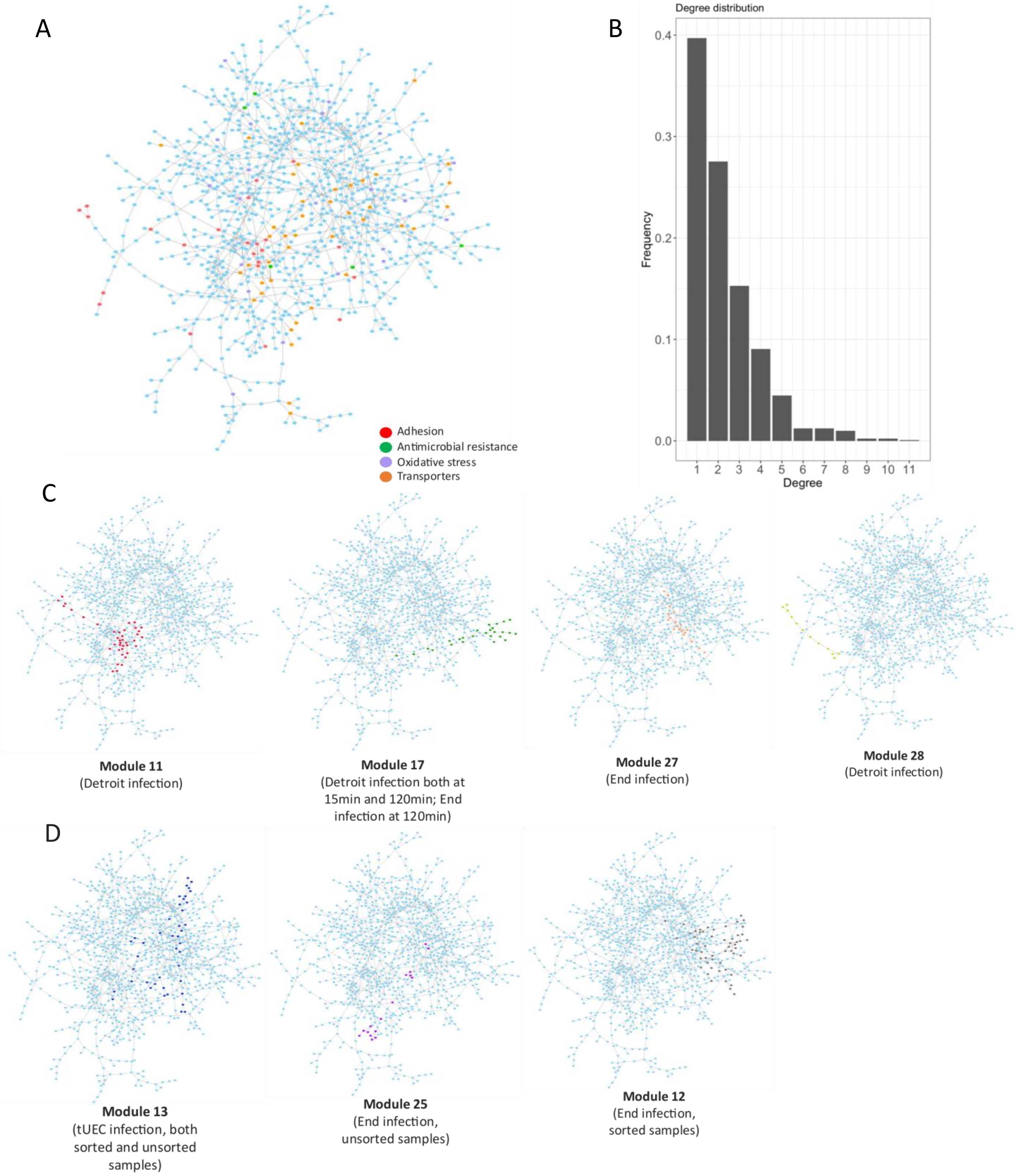


Supplementary Figure 5 | Gonococcus transcriptional reprogramming during the infection time course in the three cell models. A-F) Regulations of bacterial DE genes falling by orthology inside some interesting KEGG categories upon 15 min of Detroit562 infection (**A**), 120 min of Detroit562 infection (sorted samples) (**B**), 15 min of End1 infection (**C**), 120 min of End1 infection (sorted samples) (**D**), 15 min of t-UEC infection (**E**) and 120 min of t-UEC infection (sorted samples) (**F**). Heatmaps report mean-centered transformed read counts (normalized with respect to library size). Colors from yellow to red represent up-regulated genes, colors from light blue to blue represent down-regulated genes. Adjusted p-value to call DE genes ≤ 0.05.


A
Detroit562 120min Sorting
TPM
125000
100000
75000
50000
30000
20000
10000
3000
Replicate
1
2
3
C7R98_RS04595:ssrA
C7R98_RS11050:rnpB
C7R98_RS10575:ssrS
C7R98_RS09405:tRNA-Ile
C7R98_RS07160:tRNA-Ile
C7R98_RS09915:tRNA-Ile
C7R98_RS11355:tRNA-Ile
C7R98_RS13020:recombinase family protein
C7R98_RS13015:class A broad-spectrum beta-lactamase TEM-1
C7R98_RS02200:rpmI
C7R98_RS11970:porin
C7R98_RS01300:tRNA-Gly
C7R98_RS11665:tRNA-Tyr
C7R98_RS04710:hypothetical protein
C7R98_RS02320:tRNA-Arg
C7R98_RS09070:type II toxin-antitoxin system HicA family toxin
C7R98_RS13210:hypothetical protein
C7R98_RS02325:tRNA-Glu
C7R98_RS12310:acpP
C7R98_RS02350:hypothetical protein
Gene
End 120min Sorting
TPM
125000
100000
75000
50000
30000
20000
10000
3000
Replicate
1
2
3
C7R98_RS04595:ssrA
C7R98_RS11050:rnpB
C7R98_RS09405:tRNA-Ile
C7R98_RS07160:tRNA-Ile
C7R98_RS09915:tRNA-Ile
C7R98_RS11355:tRNA-Ile
C7R98_RS10575:ssrS
C7R98_RS13020:recombinase family protein
C7R98_RS01300:tRNA-Gly
C7R98_RS13015:class A broad-spectrum beta-lactamase TEM-1
C7R98_RS09070:type II toxin-antitoxin system HicA family toxin
C7R98_RS11665:tRNA-Tyr
C7R98_RS11970:porin
C7R98_RS02325:tRNA-Glu
C7R98_RS02320:tRNA-Arg
C7R98_RS13210:hypothetical protein
C7R98_RS12310:acpP
C7R98_RS06870:tRNA-Leu
C7R98_RS02350:hypothetical protein
C7R98_RS13110:virulence factor
Gene
tUEC 120min Sorting
TPM
125000
100000
75000
50000
30000
20000
10000
3000
Replicate
1
2
3
C7R98_RS04595:ssrA
C7R98_RS11050:rnpB
C7R98_RS13020:recombinase family protein
C7R98_RS07160:tRNA-Ile
C7R98_RS09915:tRNA-Ile
C7R98_RS09405:tRNA-Ile
C7R98_RS11970:porin
C7R98_RS11355:tRNA-Ile
C7R98_RS10575:ssrS
C7R98_RS13015:class A broad-spectrum beta-lactamase TEM-1
C7R98_RS12740:30S ribosomal protein S21
C7R98_RS12310:acpP
C7R98_RS02200:rpmI
C7R98_RS11680:tuf
C7R98_RS13210:hypothetical protein
C7R98_RS11070:hypothetical protein
C7R98_RS13215:hypothetical protein
C7R98_RS08770:rpmM
C7R98_RS01570:rpsO
C7R98_RS01300:tRNA-Gly
Gene



Supplementary Figure 6 | *N. gonorrhoeae* WHO M PorB.1B gene expression and regulation during early infection of the three cell models. A) Transcripts per kilobase of length per million mapped reads (TPM) of the top 20 most expressed genes (average TPM between replicates) by internalized *N. gonorrhoeae* WHO M bacterial cells in all the three infection models. **B)** Normalized counts (with respect to library size) plus a pseudocount of 0.5 related to porB.1B gene in *N. gonorrhoeae* WHO M during early infection of the three infection models.



Module 11
(Detroit infection)

Module 17
(Detroit infection both at 15min and 120min; End infection at 120min)

Module 27
(End infection)

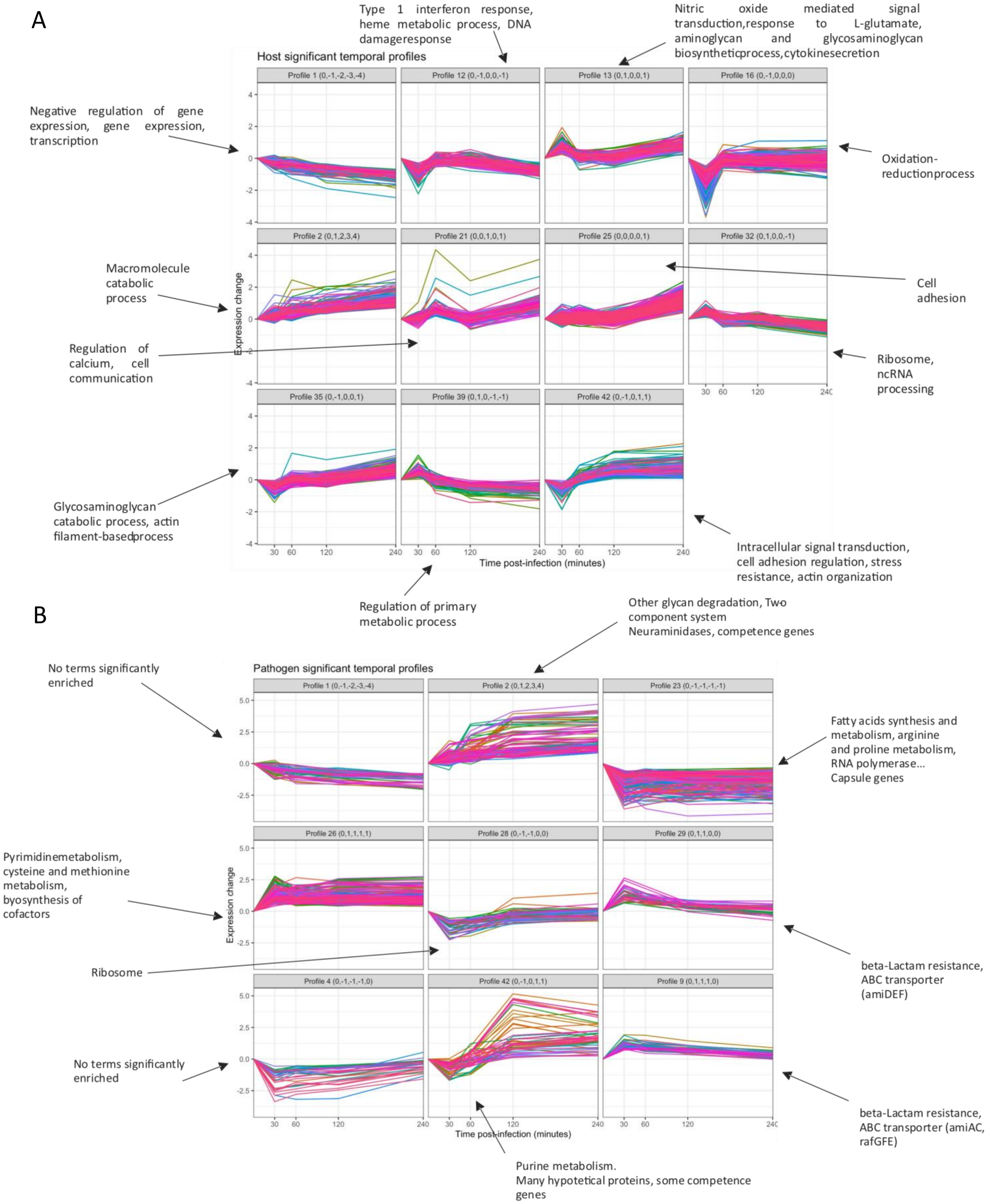
Module 28
(Detroit infection)

Module 13
(tUEC infection, both sorted and unsorted samples)

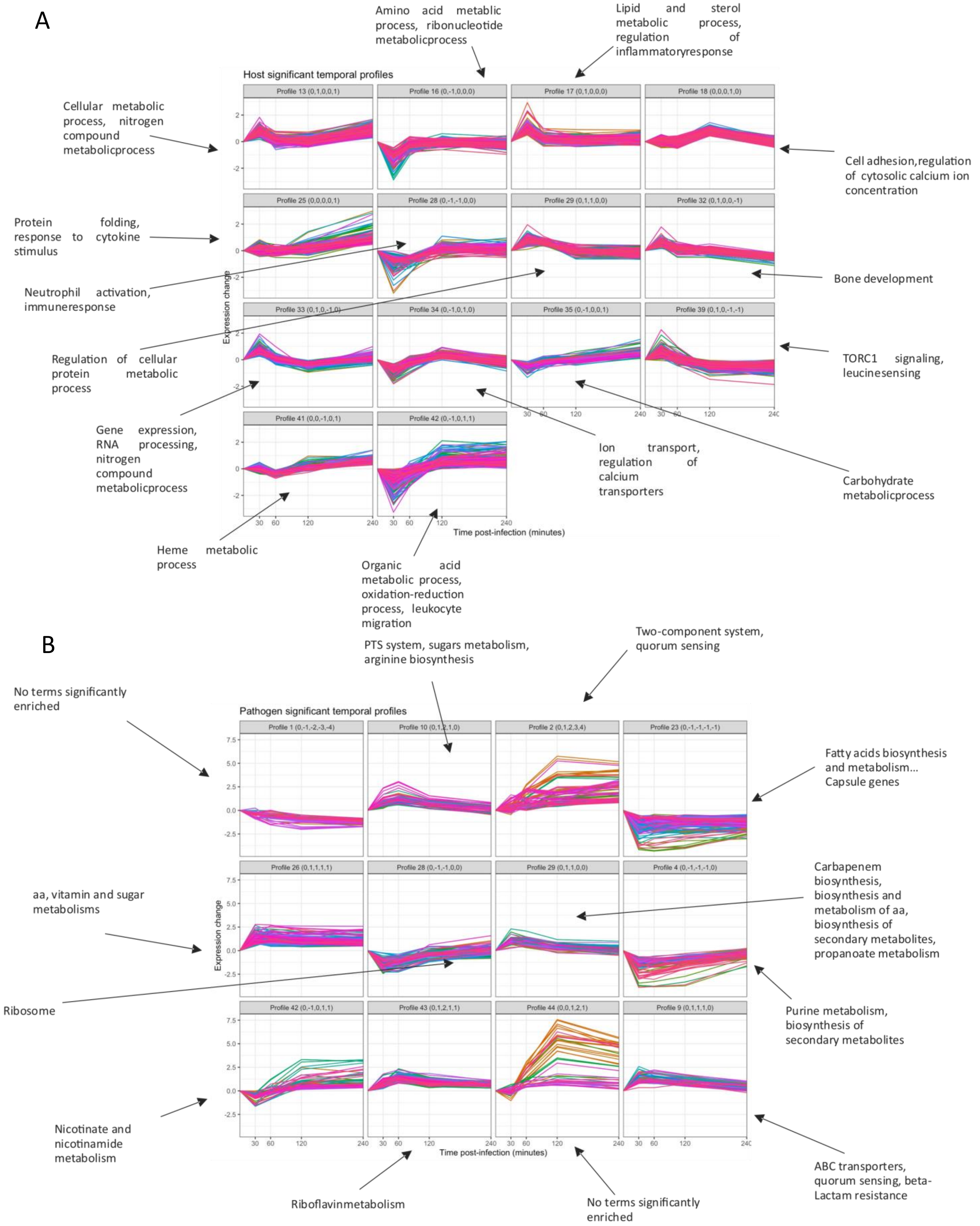
Module 25
(End infection, unsorted samples)

Module 12
(End infection, sorted samples)

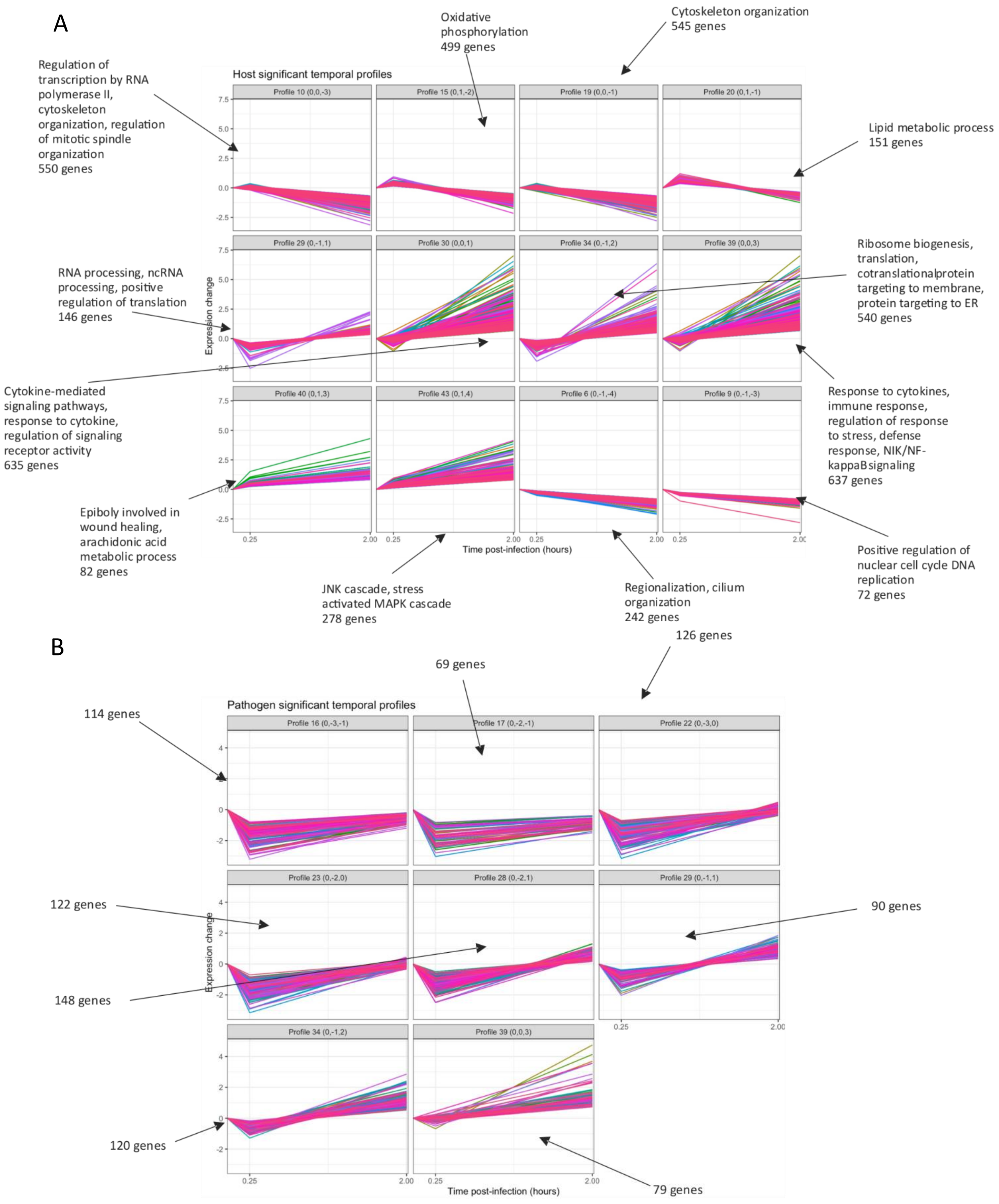
Supplementary Figure 7 | *N. gonorrhoeae* WHO M co-expression network analysis. A) Selected inferred *N. gonorrhoeae* WHO M co-expression network by context likelihood of relatedness (Z-score cutoff 8.418426) with highlighted the known adhesion, antimicrobial resistance, oxidative stress response or transporter genes (Cytoscape visualization). **B)** Degree distribution for the co-expression network reported in **A**. **C-D)** *N. gonorrhoeae* WHO M gene co-expression network visualization by Cytoscape showing in color different co-expression modules enriched in regulated genes upon 15 minutes of infection (**C**) or 120 minutes of infection (**D**) compared to pre-infection controls in the respective cell models indicated at the bottom.



Supplementary Figure 8 | Host and pathogen significant temporal expression patterns emerging from the analysis of the publicly available dual RNA-Seq data set of *S. pneumoniae* infecting epithelial cells (wild type infection data). A-B) Significant temporal expression patterns (adj. p-value ≤ 0.05) emerging for the host (**A**) and the pathogen (**B**) from the initial set of 44 model profiles (number of time points n=5, amount of change a gene can exhibit between successive time points c=1) considering the wild-type bacterial infection. Each line in a profile is a gene being associated to that pattern. Arrows indicate some top significantly enriched GO biological processes (**A**) or KEGG pathways (**B**) resulting from the hypergeometric test on the profile genes.

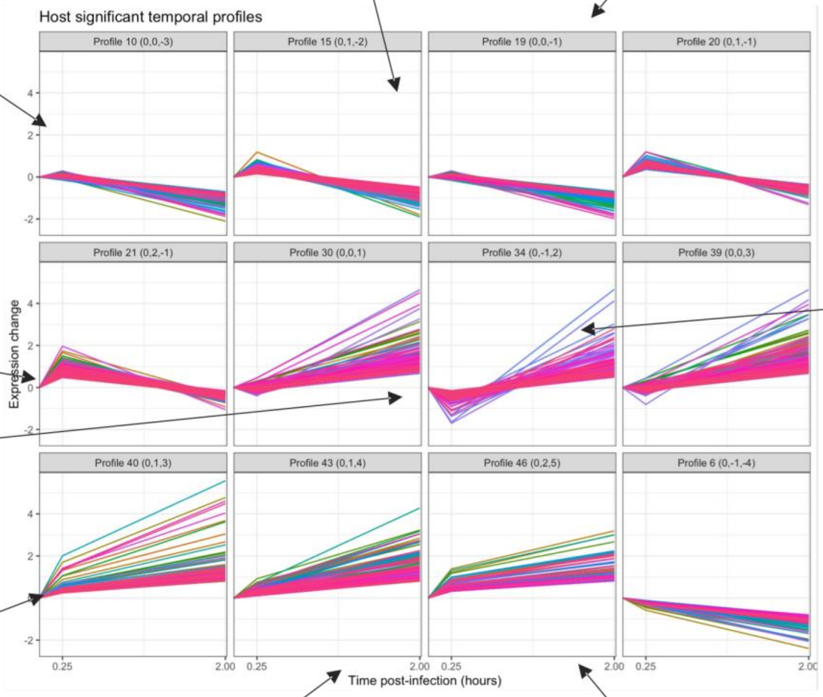


Supplementary Figure 9 | Host and pathogen significant temporal expression patterns emerging from the analysis of the publicly available dual RNA-Seq data set of *S. pneumoniae* infecting epithelial cells (mutants infection data). A-B) Significant temporal expression patterns (adj. p-value ≤ 0.05) emerging for the host (**A**) and the pathogen (**B**) from the initial set of 44 model profiles (number of time points n=5, amount of change a gene can exhibit between successive time points c=1) considering the infection by mutants. Each line in a profile is a gene being associated to that pattern. Arrows indicate some top significantly enriched GO biological processes (**A**) or KEGG pathways (**B**) resulting from the hypergeometric test on the profile genes.



Supplementary Figure 10 | Host and pathogen significant temporal expression patterns emerging from the analysis of the dual RNA-Seq data set of *N. gonorrhoeae* WHO M infecting Detroit562 cell model. A-B) Significant temporal expression patterns (adj. p-value ≤ 0.05) emerging for the host (**A**) and the pathogen (**B**) from the initial set of 48 model profiles (number of time points n=3, amount of change a gene can exhibit between successive time points c=3) considering Detroit562 cell infection. Each line in a profile is a gene being associated to that pattern. Arrows indicate some top significantly enriched GO biological processes (hypergeometric test) and the number of genes (**A**) or the number of genes (**B**) related to each profile.

A



Cell cycle process, cell division
160 genes

Microtubule-based process, mitotic cell cycle process
246 genes

Cytoskeleton, cell division
144 genes

Regulation of MAP kinase activity
165 genes

Vesicle fusion, leukocyte differentiation, calcium ion transmembrane import into cytosol
203 genes

RNA processing, nucleobase-containing compound metabolic process
284 genes

Nucleic acid metabolic process
253 genes

Cellular nitrogen compound metabolic process
246 genes

Cytokine-mediated signalingpathway, NIK/NF-kappaB signaling, regulation of immune system process, response to bacterium
89 genes

Regulation of CD4positive, alpha-beta T cell activation, negative regulation of interleukin1 secretion, interleukin6 production
181 genes

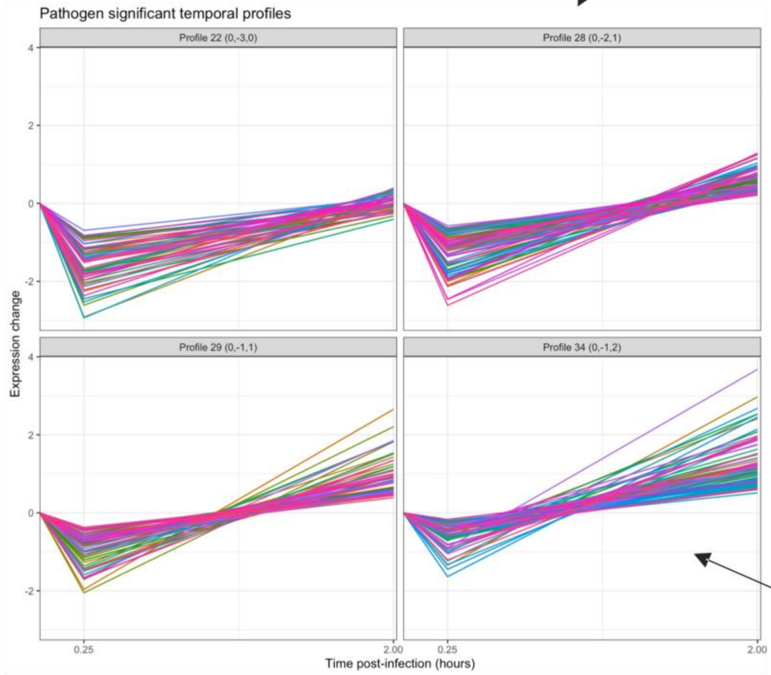
Regulation of cell proliferation, T cell differentiation involved in immune response
74 genes

Protein-lipid complex assembly, nucleus organization, cathecholamine secretion and transport, regulation of transcription
112 genes

B

56 genes

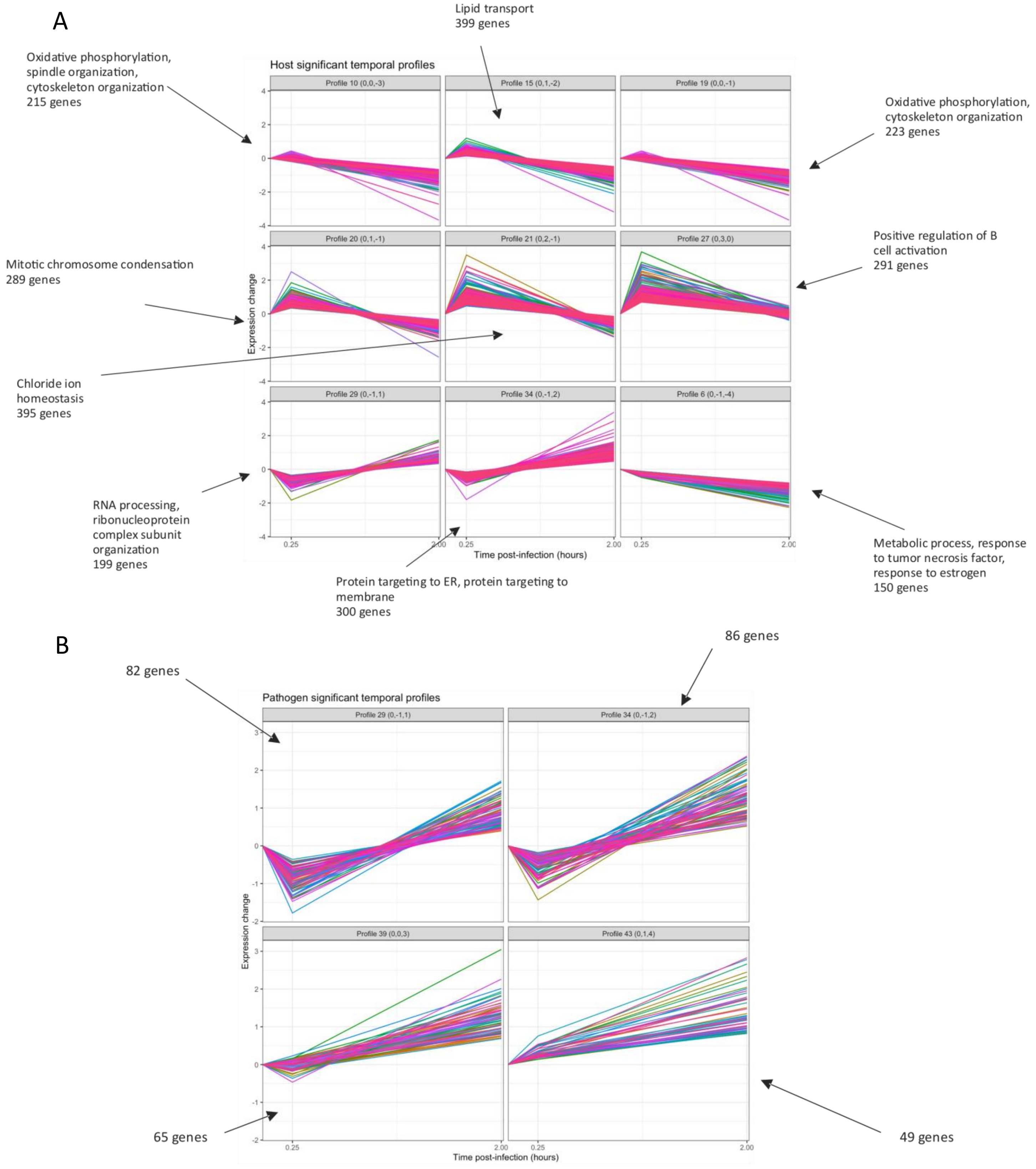
75 genes



62 genes

72 genes

Supplementary Figure 11 | Host and pathogen significant temporal expression patterns emerging from the analysis of the dual RNA-Seq data set of *N. gonorrhoeae* WHO M infecting End1 cell model. A-B) Significant temporal expression patterns (adj. p-value ≤ 0.05) emerging for the host (**A**) and the pathogen (**B**) from the initial set of 48 model profiles (number of time points n=3, amount of change a gene can exhibit between successive time points c=3) considering End1 cell infection. Each line in a profile is a gene being associated to that pattern. Arrows indicate some top significantly enriched GO biological processes (hypergeometric test) and the number of genes (**A**) or the number of genes (**B**) related to each profile.



Lipid transport
399 genes

Oxidative phosphorylation,
spindle organization,
cytoskeleton organization
215 genes

Oxidative phosphorylation,
cytoskeleton organization
223 genes

Positive regulation of B
cell activation
291 genes

Mitotic chromosome condensation
289 genes

Chloride ion
homeostasis
395 genes

RNA processing,
ribonucleoprotein
complex subunit
organization
199 genes

Protein targeting to ER, protein targeting to
membrane
300 genes

Metabolic process, response
to tumor necrosis factor,
response to estrogen
150 genes

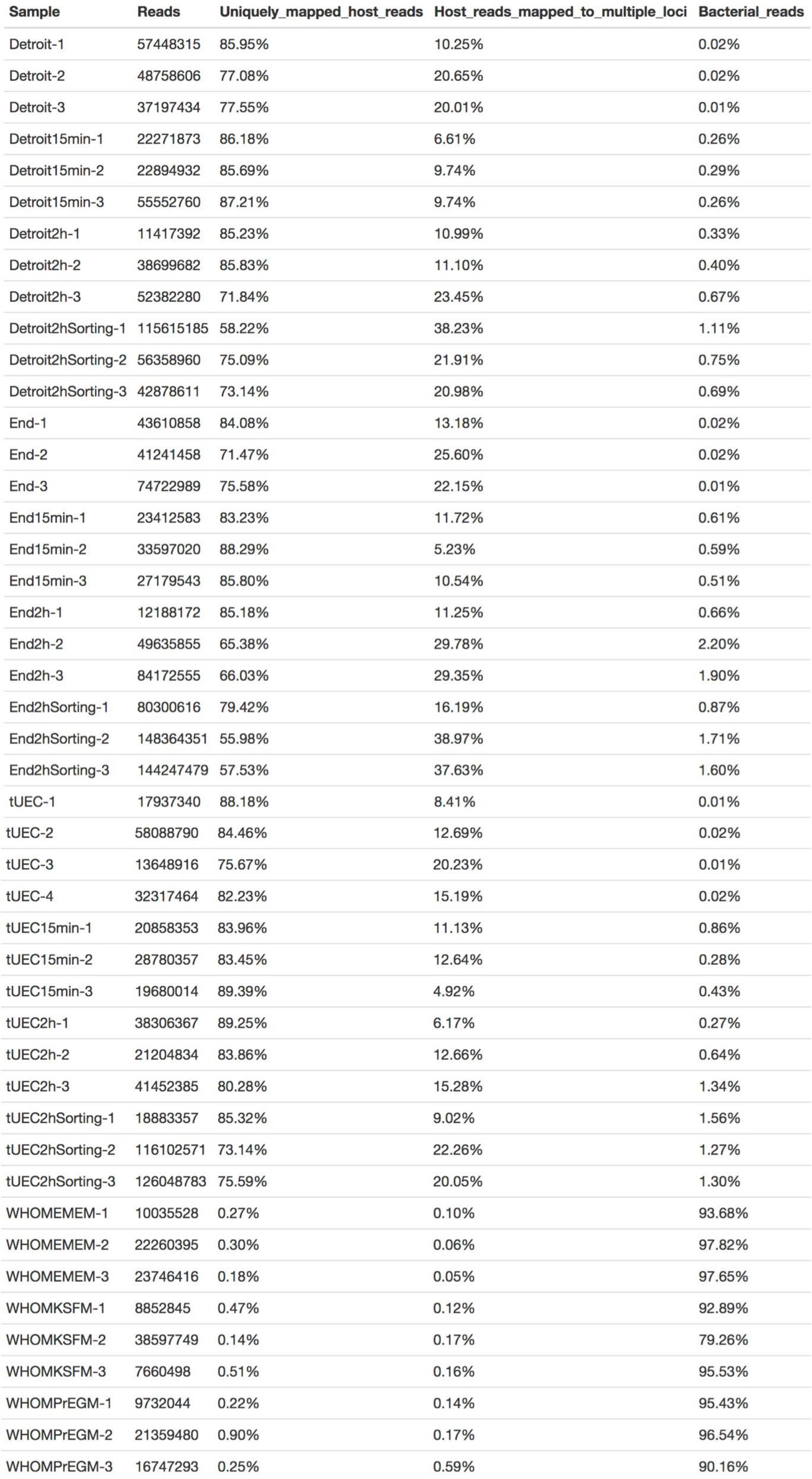
86 genes

82 genes

65 genes

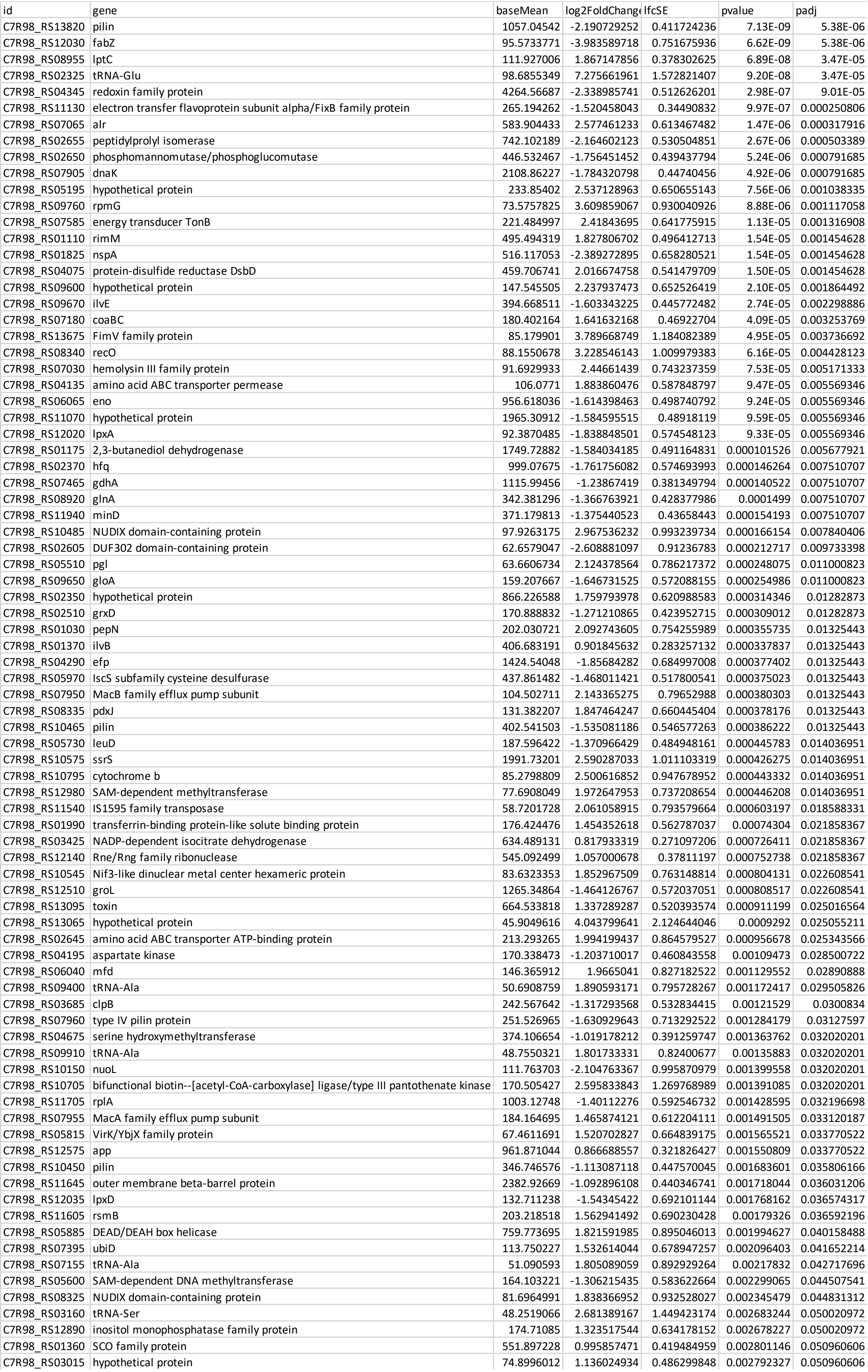
49 genes

Supplementary Figure 12 | Host and pathogen significant temporal expression patterns emerging from the analysis of the dual RNA-Seq data set of *N. gonorrhoeae* WHO M infecting t-UEC cell model. A-B) Significant temporal expression patterns (adj. p-value ≤ 0.05) emerging for the host (**A**) and the pathogen (**B**) from the initial set of 48 model profiles (number of time points n=3, amount of change a gene can exhibit between successive time points c=3) considering t-UEC cell infection. Each line in a profile is a gene being associated to that pattern. Arrows indicate some top significantly enriched GO biological processes (hypergeometric test) and the number of genes (**A**) or the number of genes (**B**) related to each profile.



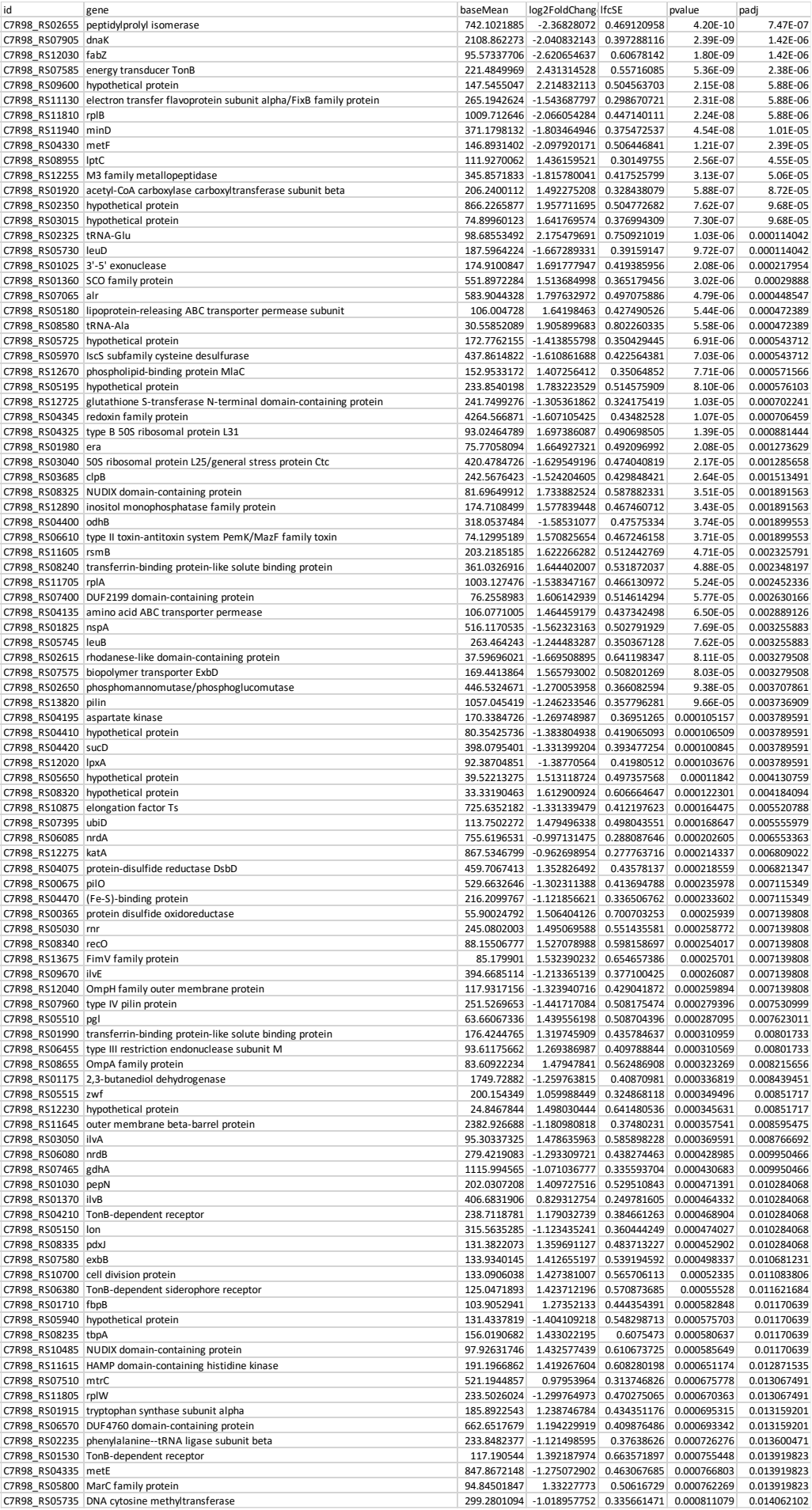
Sample	Reads	Uniquely_mapped_host_reads	Host_reads_mapped_to_multiple_loci	Bacterial_reads
Detroit-1	57448315	85.95%	10.25%	0.02%
Detroit-2	48758606	77.08%	20.65%	0.02%
Detroit-3	37197434	77.55%	20.01%	0.01%
Detroit15min-1	22271873	86.18%	6.61%	0.26%
Detroit15min-2	22894932	85.69%	9.74%	0.29%
Detroit15min-3	55552760	87.21%	9.74%	0.26%
Detroit2h-1	11417392	85.23%	10.99%	0.33%
Detroit2h-2	38699682	85.83%	11.10%	0.40%
Detroit2h-3	52382280	71.84%	23.45%	0.67%
Detroit2hSorting-1	115615185	58.22%	38.23%	1.11%
Detroit2hSorting-2	56358960	75.09%	21.91%	0.75%
Detroit2hSorting-3	42878611	73.14%	20.98%	0.69%
End-1	43610858	84.08%	13.18%	0.02%
End-2	41241458	71.47%	25.60%	0.02%
End-3	74722989	75.58%	22.15%	0.01%
End15min-1	23412583	83.23%	11.72%	0.61%
End15min-2	33597020	88.29%	5.23%	0.59%
End15min-3	27179543	85.80%	10.54%	0.51%
End2h-1	12188172	85.18%	11.25%	0.66%
End2h-2	49635855	65.38%	29.78%	2.20%
End2h-3	84172555	66.03%	29.35%	1.90%
End2hSorting-1	80300616	79.42%	16.19%	0.87%
End2hSorting-2	148364351	55.98%	38.97%	1.71%
End2hSorting-3	144247479	57.53%	37.63%	1.60%
tUEC-1	17937340	88.18%	8.41%	0.01%
tUEC-2	58088790	84.46%	12.69%	0.02%
tUEC-3	13648916	75.67%	20.23%	0.01%
tUEC-4	32317464	82.23%	15.19%	0.02%
tUEC15min-1	20858353	83.96%	11.13%	0.86%
tUEC15min-2	28780357	83.45%	12.64%	0.28%
tUEC15min-3	19680014	89.39%	4.92%	0.43%
tUEC2h-1	38306367	89.25%	6.17%	0.27%
tUEC2h-2	21204834	83.86%	12.66%	0.64%
tUEC2h-3	41452385	80.28%	15.28%	1.34%
tUEC2hSorting-1	18883357	85.32%	9.02%	1.56%
tUEC2hSorting-2	116102571	73.14%	22.26%	1.27%
tUEC2hSorting-3	126048783	75.59%	20.05%	1.30%
WHOMEMEM-1	10035528	0.27%	0.10%	93.68%
WHOMEMEM-2	22260395	0.30%	0.06%	97.82%
WHOMEMEM-3	23746416	0.18%	0.05%	97.65%
WHOMKSFM-1	8852845	0.47%	0.12%	92.89%
WHOMKSFM-2	38597749	0.14%	0.17%	79.26%
WHOMKSFM-3	7660498	0.51%	0.16%	95.53%
WHOMPrEGM-1	9732044	0.22%	0.14%	95.43%
WHOMPrEGM-2	21359480	0.90%	0.17%	96.54%
WHOMPrEGM-3	16747293	0.25%	0.59%	90.16%

Supplementary Table 1 | Table reporting the total number of sequenced reads, the percentage of reads uniquely mapping to the host genome, the percentage of host reads mapping to multiple loci and the percentage of bacterial reads for each library.



id	gene	baseMean	log2FoldChange	lfcSE	pvalue	padj
C7R98_RS13820	pilin	1057.04542	-2.190729252	0.411724236	7.13E-09	5.38E-06
C7R98_RS12030	fabZ	95.5733771	-3.983589718	0.751675936	6.62E-09	5.38E-06
C7R98_RS08955	lptC	111.927006	1.867147856	0.378302625	6.89E-08	3.47E-05
C7R98_RS02325	tRNA-Glu	98.6855349	7.275661961	1.572821407	9.20E-08	3.47E-05
C7R98_RS04345	redoxin family protein	4264.56687	-2.338985741	0.512626201	2.98E-07	9.01E-05
C7R98_RS11130	electron transfer flavoprotein subunit alpha/FixB family protein	265.194262	-1.520458043	0.34490832	9.97E-07	0.000250806
C7R98_RS07065	alr	583.904433	2.577461233	0.613467482	1.47E-06	0.000317916
C7R98_RS02655	peptidylprolyl isomerase	742.102189	-2.164602123	0.530504851	2.67E-06	0.000503389
C7R98_RS02650	phosphomannomutase/phosphoglucomutase	446.532467	-1.756451452	0.439437794	5.24E-06	0.000791685
C7R98_RS07905	dnaK	2108.86227	-1.784320798	0.44740456	4.92E-06	0.000791685
C7R98_RS05195	hypothetical protein	233.85402	2.537128963	0.650655143	7.56E-06	0.001038335
C7R98_RS09760	rpmG	73.5757825	3.609859067	0.930040926	8.88E-06	0.001117058
C7R98_RS07585	energy transducer TonB	221.484997	2.41843695	0.641775915	1.13E-05	0.001316908
C7R98_RS01110	rimM	495.494319	1.827806702	0.496412713	1.54E-05	0.001454628
C7R98_RS01825	nspA	516.117053	-2.389272895	0.658280521	1.54E-05	0.001454628
C7R98_RS04075	protein-disulfide reductase DsbD	459.706741	2.016674758	0.541479709	1.50E-05	0.001454628
C7R98_RS09600	hypothetical protein	147.545505	2.237937473	0.652526419	2.10E-05	0.001864492
C7R98_RS09670	ilvE	394.668511	-1.603343225	0.445772482	2.74E-05	0.002298886
C7R98_RS07180	coaBC	180.402164	1.641632168	0.46922704	4.09E-05	0.003253769
C7R98_RS13675	FimV family protein	85.179901	3.789668749	1.184082389	4.95E-05	0.003736692
C7R98_RS08340	recO	88.1550678	3.228546143	1.009979383	6.16E-05	0.004428123
C7R98_RS07030	hemolysin III family protein	91.6929933	2.44661439	0.743237359	7.53E-05	0.005171333
C7R98_RS04135	amino acid ABC transporter permease	106.0771	1.883860476	0.587848797	9.47E-05	0.005569346
C7R98_RS06065	eno	956.618036	-1.614398463	0.498740792	9.24E-05	0.005569346
C7R98_RS11070	hypothetical protein	1965.30912	-1.584595515	0.48918119	9.59E-05	0.005569346
C7R98_RS12020	lpxA	92.3870485	-1.838848501	0.574548123	9.33E-05	0.005569346
C7R98_RS01175	2,3-butanediol dehydrogenase	1749.72882	-1.584034185	0.491164831	0.000101526	0.005677921
C7R98_RS02370	hfq	999.07675	-1.761756082	0.574693993	0.000146264	0.007510707
C7R98_RS07465	gdhA	1115.99456	-1.23867419	0.381349794	0.000140522	0.007510707
C7R98_RS08920	glnA	342.381296	-1.366763921	0.428377986	0.0001499	0.007510707
C7R98_RS11940	minD	371.179813	-1.375440523	0.43658443	0.000154193	0.007510707
C7R98_RS10485	NUDIX domain-containing protein	97.9263175	2.967536232	0.993239734	0.000166154	0.007840406
C7R98_RS02605	DUF302 domain-containing protein	62.6579047	-2.608881097	0.91236783	0.000212717	0.009733398
C7R98_RS05510	pgl	63.6606734	2.124378564	0.786217372	0.000248075	0.011000823
C7R98_RS09650	gloA	159.207667	-1.646731525	0.572088155	0.000254986	0.011000823
C7R98_RS02350	hypothetical protein	866.226588	1.759793978	0.620988583	0.000314346	0.01282873
C7R98_RS02510	grxD	170.888832	-1.271210865	0.423952715	0.000309012	0.01282873
C7R98_RS01030	pepN	202.030721	2.092743605	0.754255989	0.000355735	0.01325443
C7R98_RS01370	ilvB	406.683191	0.901845632	0.283257132	0.000337837	0.01325443
C7R98_RS04290	efp	1424.54048	-1.85684282	0.684997008	0.000377402	0.01325443
C7R98_RS05970	IscS subfamily cysteine desulfurase	437.861482	-1.468011421	0.517800541	0.000375023	0.01325443
C7R98_RS07950	MacB family efflux pump subunit	104.502711	2.143365275	0.79652988	0.000380303	0.01325443
C7R98_RS08335	pdxJ	131.382207	1.847464247	0.660445404	0.000378176	0.01325443
C7R98_RS10465	pilin	402.541503	-1.535081186	0.546577263	0.000386222	0.01325443
C7R98_RS05730	leuD	187.596422	-1.370966429	0.484948161	0.000445783	0.014036951
C7R98_RS10575	ssrS	1991.73201	2.590287033	1.011103319	0.000426275	0.014036951
C7R98_RS10795	cytochrome b	85.2798809	2.500616852	0.947678952	0.000443332	0.014036951
C7R98_RS12980	SAM-dependent methyltransferase	77.6908049	1.972647953	0.737208654	0.000446208	0.014036951
C7R98_RS11540	IS1595 family transposase	58.7201728	2.061058915	0.793579664	0.000603197	0.018588331
C7R98_RS01990	transferrin-binding protein-like solute binding protein	176.424476	1.454352618	0.562787037	0.00074304	0.021858367
C7R98_RS03425	NADP-dependent isocitrate dehydrogenase	634.489131	0.817933319	0.271097206	0.000726411	0.021858367
C7R98_RS12140	Rne/Rng family ribonuclease	545.092499	1.057000678	0.37811197	0.000752738	0.021858367
C7R98_RS10545	Nif3-like dinuclear metal center hexameric protein	83.6323353	1.852967509	0.763148814	0.000804131	0.022608541
C7R98_RS12510	groL	1265.34864	-1.464126767	0.572037051	0.000808517	0.022608541
C7R98_RS13095	toxin	664.533818	1.337289287	0.520393574	0.000911199	0.025016564
C7R98_RS13065	hypothetical protein	45.9049616	4.043799641	2.124644046	0.0009292	0.025055211
C7R98_RS02645	amino acid ABC transporter ATP-binding protein	213.293265	1.994199437	0.864579527	0.000956678	0.025343566
C7R98_RS04195	aspartate kinase	170.338473	-1.203710017	0.460843558	0.00109473	0.028500722
C7R98_RS06040	mfd	146.365912	1.9665041	0.827182522	0.001129552	0.02890888
C7R98_RS09400	tRNA-Ala	50.6908759	1.890593171	0.795728267	0.001172417	0.029505826
C7R98_RS03685	clpB	242.567642	-1.317293568	0.532834415	0.00121529	0.0300834
C7R98_RS07960	type IV pilin protein	251.526965	-1.630929643	0.713292522	0.001284179	0.03127597
C7R98_RS04675	serine hydroxymethyltransferase	374.106654	-1.019178212	0.391259747	0.001363762	0.032020201
C7R98_RS09910	tRNA-Ala	48.7550321	1.801733331	0.82400677	0.00135883	0.032020201
C7R98_RS10150	nuoL	111.763703	-2.104763367	0.995870979	0.001399558	0.032020201
C7R98_RS10705	bifunctional biotin--[acetyl-CoA-carboxylase] ligase/type III pantothenate kinase	170.505427	2.595833843	1.269768989	0.001391085	0.032020201
C7R98_RS11705	rplA	1003.12748	-1.40112276	0.592546732	0.001428595	0.032196698
C7R98_RS07955	MacA family efflux pump subunit	184.164695	1.465874121	0.612204111	0.001491505	0.033120187
C7R98_RS05815	VirK/YbjX family protein	67.4611691	1.520702827	0.664839175	0.001565521	0.033770522
C7R98_RS12575	app	961.871044	0.866688557	0.321826427	0.001550809	0.033770522
C7R98_RS10450	pilin	346.746576	-1.113087118	0.447570045	0.001683601	0.035806166
C7R98_RS11645	outer membrane beta-barrel protein	2382.92669	-1.092896108	0.440346741	0.001718044	0.036031206
C7R98_RS12035	lpxD	132.711238	-1.54345422	0.692101144	0.001768162	0.036574317
C7R98_RS11605	rsmB	203.218518	1.562941492	0.690230428	0.00179326	0.036592196
C7R98_RS05885	DEAD/DEAH box helicase	759.773695	1.821591985	0.895046013	0.001994627	0.040158488
C7R98_RS07395	ubiD	113.750227	1.532614044	0.678947257	0.002096403	0.041652214
C7R98_RS07155	tRNA-Ala	51.090593	1.805089059	0.892929264	0.00217832	0.042717696
C7R98_RS05600	SAM-dependent DNA methyltransferase	164.103221	-1.306215435	0.583622664	0.002299065	0.044507541
C7R98_RS08325	NUDIX domain-containing protein	81.6964991	1.838366952	0.932528027	0.002345479	0.044831312
C7R98_RS03160	tRNA-Ser	48.2519066	2.681389167	1.449423174	0.002683244	0.050020972
C7R98_RS12890	inositol monophosphatase family protein	174.71085	1.323517544	0.634178152	0.002678227	0.050020972
C7R98_RS01360	SCO family protein	551.897228	0.995857471	0.419484959	0.002801146	0.050960606
C7R98_RS03015	hypothetical protein	74.8996012	1.136024934	0.486299848	0.002792327	0.050960606

Supplementary Table 2 | List of DE genes emerging by comparing bacterial data from sorted samples at 120 mpi and bacterial data from unsorted samples at 120 mpi for Detroit562 cell model (adj. p-value ≤ 0.05).



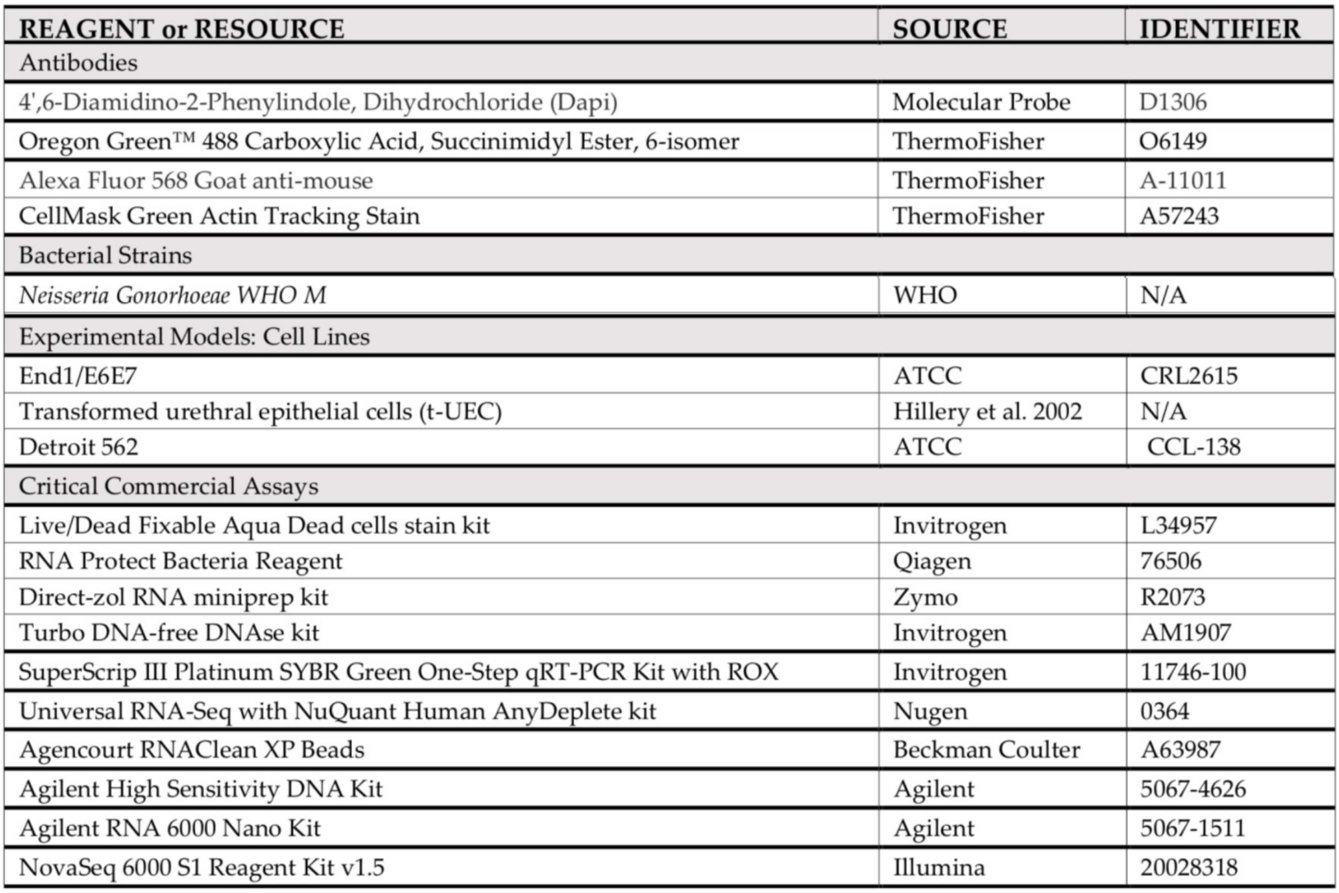
id	gene	baseMean	log2FoldChang	lfcSE	pvalue	padj
C7R98_RS02655	peptidylprolyl isomerase	742.1021885	-2.36828072	0.469120958	4.20E-10	7.47E-07
C7R98_RS07905	dnaK	2108.862273	-2.040832143	0.397288116	2.39E-09	1.42E-06
C7R98_RS12030	fabZ	95.57337706	-2.620654637	0.60678142	1.80E-09	1.42E-06
C7R98_RS07585	energy transducer TonB	221.4849969	2.431314528	0.55716085	5.36E-09	2.38E-06
C7R98_RS09600	hypothetical protein	147.5455047	2.214832113	0.504563703	2.15E-08	5.88E-06
C7R98_RS11130	electron transfer flavoprotein subunit alpha/FixB family protein	265.1942624	-1.543687797	0.298670721	2.31E-08	5.88E-06
C7R98_RS11810	rplB	1009.712646	-2.066054284	0.447140111	2.24E-08	5.88E-06
C7R98_RS11940	minD	371.1798132	-1.803464946	0.375472537	4.54E-08	1.01E-05
C7R98_RS04330	metF	146.8931402	-2.097920171	0.506446841	1.21E-07	2.39E-05
C7R98_RS08955	lptC	111.9270062	1.436159521	0.30149755	2.56E-07	4.55E-05
C7R98_RS12255	M3 family metallopeptidase	345.8571833	-1.815780041	0.417525799	3.13E-07	5.06E-05
C7R98_RS01920	acetyl-CoA carboxylase carboxyltransferase subunit beta	206.2400112	1.492275208	0.328438079	5.88E-07	8.72E-05
C7R98_RS02350	hypothetical protein	866.2265877	1.957711695	0.504772682	7.62E-07	9.68E-05
C7R98_RS03015	hypothetical protein	74.89960123	1.641769574	0.376994309	7.30E-07	9.68E-05
C7R98_RS02325	tRNA-Glu	98.68553492	2.175479691	0.750921019	1.03E-06	0.000114042
C7R98_RS05730	leuD	187.5964224	-1.667289331	0.39159147	9.72E-07	0.000114042
C7R98_RS01025	3'-5' exonuclease	174.9100847	1.691777947	0.419385956	2.08E-06	0.000217954
C7R98_RS01360	SCO family protein	551.8972284	1.513684998	0.365179456	3.02E-06	0.00029888
C7R98_RS07065	alr	583.9044328	1.797632972	0.497075886	4.79E-06	0.000448547
C7R98_RS05180	lipoprotein-releasing ABC transporter permease subunit	106.004728	1.64198463	0.427490526	5.44E-06	0.000472389
C7R98_RS08580	tRNA-Ala	30.55852089	1.905899683	0.802260335	5.58E-06	0.000472389
C7R98_RS05725	hypothetical protein	172.7762155	-1.413855798	0.350429445	6.91E-06	0.000543712
C7R98_RS05970	IscS subfamily cysteine desulfurase	437.8614822	-1.610861688	0.422564381	7.03E-06	0.000543712
C7R98_RS12670	phospholipid-binding protein MlaC	152.9533172	1.407256412	0.35064852	7.71E-06	0.000571566
C7R98_RS05195	hypothetical protein	233.8540198	1.783223529	0.514575909	8.10E-06	0.000576103
C7R98_RS12725	glutathione S-transferase N-terminal domain-containing protein	241.7499276	-1.305361862	0.324175419	1.03E-05	0.000702241
C7R98_RS04345	redoxin family protein	4264.566871	-1.607105425	0.43482528	1.07E-05	0.000706459
C7R98_RS04325	type B 50S ribosomal protein L31	93.02464789	1.697386087	0.490698505	1.39E-05	0.000881444
C7R98_RS01980	era	75.77058094	1.664927321	0.492096992	2.08E-05	0.001273629
C7R98_RS03040	50S ribosomal protein L25/general stress protein Ctc	420.4784726	-1.629549196	0.474040819	2.17E-05	0.001285658
C7R98_RS03685	clpB	242.5676423	-1.524204605	0.429848421	2.64E-05	0.001513491
C7R98_RS08325	NUDIX domain-containing protein	81.69649912	1.733882524	0.587882331	3.51E-05	0.001891563
C7R98_RS12890	inositol monophosphatase family protein	174.7108499	1.577839448	0.467460712	3.43E-05	0.001891563
C7R98_RS04400	odhB	318.0537484	-1.58531077	0.47575334	3.74E-05	0.001899553
C7R98_RS06610	type II toxin-antitoxin system PemK/MazF family toxin	74.12995189	1.570825654	0.467246158	3.71E-05	0.001899553
C7R98_RS11605	rsmB	203.2185185	1.622266282	0.512442769	4.71E-05	0.002325791
C7R98_RS08240	transferrin-binding protein-like solute binding protein	361.0326916	1.644402007	0.531872037	4.88E-05	0.002348197
C7R98_RS11705	rplA	1003.127476	-1.538347167	0.466130972	5.24E-05	0.002452336
C7R98_RS07400	DUF2199 domain-containing protein	76.2558983	1.606142939	0.514614294	5.77E-05	0.002630166
C7R98_RS04135	amino acid ABC transporter permease	106.0771005	1.464459179	0.437342498	6.50E-05	0.002889126
C7R98_RS01825	nspA	516.1170535	-1.562323163	0.502791929	7.69E-05	0.003255883
C7R98_RS05745	leuB	263.464243	-1.244483287	0.350367128	7.62E-05	0.003255883
C7R98_RS02615	rhodanese-like domain-containing protein	37.59696021	-1.669508895	0.641198347	8.11E-05	0.003279508
C7R98_RS07575	biopolymer transporter ExbD	169.4413864	1.565793002	0.508201269	8.03E-05	0.003279508
C7R98_RS02650	phosphomannomutase/phosphoglucomutase	446.5324671	-1.270053958	0.366082594	9.38E-05	0.003707861
C7R98_RS13820	pilin	1057.045419	-1.246233546	0.357796281	9.66E-05	0.003736909
C7R98_RS04195	aspartate kinase	170.3384726	-1.269748987	0.36951265	0.000105157	0.003789591
C7R98_RS04410	hypothetical protein	80.35425736	-1.383804938	0.419065093	0.000106509	0.003789591
C7R98_RS04420	sucD	398.0795401	-1.331399204	0.393477254	0.000100845	0.003789591
C7R98_RS12020	lpxA	92.38704851	-1.38770564	0.41980512	0.000103676	0.003789591
C7R98_RS05650	hypothetical protein	39.52213275	1.513118724	0.497357568	0.00011842	0.004130759
C7R98_RS08320	hypothetical protein	33.33190463	1.612900924	0.606664647	0.000122301	0.004184094
C7R98_RS10875	elongation factor Ts	725.6352182	-1.331339479	0.412197623	0.000164475	0.005520788
C7R98_RS07395	ubiD	113.7502272	1.479496338	0.498043551	0.000168647	0.005555979
C7R98_RS06085	nrdA	755.6196531	-0.997131475	0.288087646	0.000202605	0.006553363
C7R98_RS12275	katA	867.5346799	-0.962698954	0.277763716	0.000214337	0.006809022
C7R98_RS04075	protein-disulfide reductase DsbD	459.7067413	1.352826492	0.43578137	0.000218559	0.006821347
C7R98_RS00675	pilO	529.6632646	-1.302311388	0.413694788	0.000235978	0.007115349
C7R98_RS04470	(Fe-S)-binding protein	216.2099767	-1.121856621	0.336506762	0.000233602	0.007115349
C7R98_RS00365	protein disulfide oxidoreductase	55.90024792	1.506404126	0.700703253	0.00025939	0.007139808
C7R98_RS05030	rnr	245.0802003	1.495069588	0.551435581	0.000258772	0.007139808
C7R98_RS08340	recO	88.15506777	1.527078988	0.598158697	0.000254017	0.007139808
C7R98_RS13675	FimV family protein	85.179901	1.532390232	0.654657386	0.00025701	0.007139808
C7R98_RS09670	ilvE	394.6685114	-1.213365139	0.377100425	0.00026087	0.007139808
C7R98_RS12040	OmpH family outer membrane protein	117.9317156	-1.323940716	0.429041872	0.000259894	0.007139808
C7R98_RS07960	type IV pilin protein	251.5269653	-1.441717084	0.508175474	0.000279396	0.007530999
C7R98_RS05510	pgl	63.66067336	1.439556198	0.508704396	0.000287095	0.007623011
C7R98_RS01990	transferrin-binding protein-like solute binding protein	176.4244765	1.319745909	0.435784637	0.000310959	0.00801733
C7R98_RS06455	type III restriction endonuclease subunit M	93.61175662	1.269386987	0.409788844	0.000310569	0.00801733
C7R98_RS08655	OmpA family protein	83.60922234	1.47947841	0.562486908	0.000323269	0.008215656
C7R98_RS01175	2,3-butanediol dehydrogenase	1749.72882	-1.259763815	0.40870981	0.000336819	0.008439451
C7R98_RS05515	zwf	200.154349	1.059988449	0.324868118	0.000349496	0.00851717
C7R98_RS12230	hypothetical protein	24.8467844	1.498030444	0.641480536	0.000345631	0.00851717
C7R98_RS11645	outer membrane beta-barrel protein	2382.926688	-1.180980818	0.37480231	0.000357541	0.008595475
C7R98_RS03050	ilvA	95.30337325	1.478635963	0.585898228	0.000369591	0.008766692
C7R98_RS06080	nrdB	279.4219083	-1.293309721	0.438274463	0.000428985	0.009950466
C7R98_RS07465	gdhA	1115.994565	-1.071036777	0.335593704	0.000430683	0.009950466
C7R98_RS01030	pepN	202.0307208	1.409727516	0.529510843	0.000471391	0.010284068
C7R98_RS01370	ilvB	406.6831906	0.829312754	0.249781605	0.000464332	0.010284068
C7R98_RS04210	TonB-dependent receptor	238.7118781	1.179032739	0.384661263	0.000468904	0.010284068
C7R98_RS05150	lon	315.5635285	-1.123435241	0.360444249	0.000474027	0.010284068
C7R98_RS08335	pdxJ	131.3822073	1.359691127	0.483713227	0.000452902	0.010284068
C7R98_RS07580	exbB	133.9340145	1.412655197	0.539194592	0.000498337	0.010681231
C7R98_RS10700	cell division protein	133.0906038	1.427381007	0.565706113	0.00052335	0.011083806
C7R98_RS06380	TonB-dependent siderophore receptor	125.0471893	1.423712196	0.570873685	0.00055528	0.011621684
C7R98_RS01710	fbpB	103.9052941	1.27352133	0.444354391	0.000582848	0.01170639
C7R98_RS05940	hypothetical protein	131.4337819	-1.404109218	0.548298713	0.000575703	0.01170639
C7R98_RS08235	tbpA	156.0190682	1.433022195	0.6075473	0.000580637	0.01170639
C7R98_RS10485	NUDIX domain-containing protein	97.92631746	1.432577439	0.610673725	0.000585649	0.01170639
C7R98_RS11615	HAMP domain-containing histidine kinase	191.1966862	1.419267604	0.608280198	0.000651174	0.012871535
C7R98_RS07510	mtrC	521.1944857	0.97953964	0.313746826	0.000675778	0.013067491
C7R98_RS11805	rplW	233.5026024	-1.299764973	0.470275065	0.000670363	0.013067491
C7R98_RS01915	tryptophan synthase subunit alpha	185.8922543	1.238746784	0.434351176	0.000695315	0.013159201
C7R98_RS06570	DUF4760 domain-containing protein	662.6517679	1.194229919	0.409876486	0.000693342	0.013159201
C7R98_RS02235	phenylalanine--tRNA ligase subunit beta	233.8482377	-1.121498595	0.37638626	0.000726276	0.013600471
C7R98_RS01530	TonB-dependent receptor	117.190544	1.392187974	0.663571897	0.000755448	0.013919823
C7R98_RS04335	metE	847.8672148	-1.275072902	0.463067685	0.000766803	0.013919823
C7R98_RS05800	MarC family protein	94.84501847	1.33227773	0.50616729	0.000762269	0.013919823
C7R98_RS05735	DNA cytosine methyltransferase	299.2801094	-1.018957752	0.335661471	0.000811079	0.014062102

C7R98_RS06070	ftsB	33.99114663	1.374679414	0.557326994	0.000787496	0.014062102
C7R98_RS08010	aspartate/tyrosine/aromatic aminotransferase	232.1875827	-0.980759892	0.319288875	0.000793026	0.014062102
C7R98_RS08555	ftsL	64.86117928	1.224333906	0.435218489	0.000814163	0.014062102
C7R98_RS12160	hypothetical protein	369.2294736	1.31759134	0.498017124	0.000799304	0.014062102
C7R98_RS12510	groL	1265.348643	-1.246400567	0.450502987	0.000838785	0.01434806
C7R98_RS05885	DEAD/DEAH box helicase	759.7736951	1.376173584	0.582567394	0.000872377	0.014780563
C7R98_RS12765	membrane protein	533.9278849	-0.92050217	0.298927327	0.000883271	0.014823951
C7R98_RS04700	YdcH family protein	777.64538	-1.351168727	0.54971553	0.00092055	0.015305223
C7R98_RS03095	dnaB	255.98441	0.94732935	0.31170701	0.000948953	0.015631363
C7R98_RS10865	rpsB	894.991351	-1.219472065	0.442766256	0.000964799	0.015746574
C7R98_RS11935	minC	255.957612	-1.139768607	0.398760363	0.000974276	0.015756701
C7R98_RS07075	PFL family protein	141.0952755	-1.183088511	0.423510397	0.001000834	0.01603474
C7R98_RS12980	SAM-dependent methyltransferase	77.69080494	1.287325991	0.494252561	0.001009495	0.01603474
C7R98_RS05230	hypothetical protein	43.55823763	1.325855568	0.536288345	0.001046239	0.016471326
C7R98_RS08205	DUF502 domain-containing protein	111.4269481	1.271106424	0.486717436	0.001075197	0.016680678
C7R98_RS11770	fusA	2523.885638	-1.152778419	0.410876858	0.00107829	0.016680678
C7R98_RS03000	panB	81.36715964	1.211301888	0.451294728	0.001184536	0.018166293
C7R98_RS00210	fructose-bisphosphate aldolase class II	277.4834655	-1.191094198	0.439978848	0.001212285	0.018276744
C7R98_RS07015	infB	406.1986323	1.121890155	0.400270463	0.001204436	0.018276744
C7R98_RS07030	hemolysin III family protein	91.69299325	1.292897988	0.523594852	0.001262527	0.018742669
C7R98_RS09910	tRNA-Ala	48.75503212	1.272647944	0.504450923	0.001274797	0.018742669
C7R98_RS11800	rplD	743.2220427	-1.212147518	0.456617495	0.001266036	0.018742669
C7R98_RS11135	electron transfer flavoprotein subunit beta/FixA family protein	633.632515	-1.048117808	0.366883087	0.001290177	0.018813325
C7R98_RS07000	serC	105.3922683	1.17787308	0.437121745	0.001315426	0.018966751
C7R98_RS08150	23S rRNA methyltransferase	874.258644	0.869006242	0.291028545	0.001322022	0.018966751
C7R98_RS01260	DMT family transporter	53.83485554	1.291295992	0.535107667	0.001392144	0.019555395
C7R98_RS02370	hfq	999.0767496	-1.196481605	0.45422538	0.001416106	0.019555395
C7R98_RS04290	efp	1424.540477	-1.261996189	0.507628774	0.001439998	0.019555395
C7R98_RS04415	sucC	506.5292449	-1.060228762	0.376542811	0.001403024	0.019555395
C7R98_RS05870	trxA	166.8199066	-1.240849278	0.487314472	0.001418427	0.019555395
C7R98_RS09755	ubiM	354.6227475	-0.983326479	0.341746005	0.001438419	0.019555395
C7R98_RS10080	trmL	74.95443163	1.322143063	0.593227095	0.001392282	0.019555395
C7R98_RS07350	2-hydroxyacid dehydrogenase	206.0236168	1.143429685	0.427543279	0.001570566	0.021166948
C7R98_RS03420	hypothetical protein	261.7291886	-1.198453636	0.464282959	0.00158564	0.02120943
C7R98_RS11230	murA	344.026228	0.977427032	0.343532518	0.001605468	0.021314381
C7R98_RS08215	rpsT	719.7482314	-1.224326516	0.490528053	0.001695136	0.022204747
C7R98_RS09760	rpmG	73.57578246	1.298199893	0.59915386	0.001697496	0.022204747
C7R98_RS04920	acetyl-CoA carboxylase carboxyltransferase subunit alpha	219.1024918	1.041307415	0.378106977	0.001761609	0.022875204
C7R98_RS01480	mnmA	164.843451	1.034442086	0.375146099	0.001779047	0.022934234
C7R98_RS07685	hypothetical protein	128.2955095	1.20392104	0.479578268	0.00180543	0.023106901
C7R98_RS06205	DNA translocase FtsK	202.3357035	1.063618483	0.39488385	0.001945575	0.024697321
C7R98_RS08755	thiL	96.30586621	1.267724263	0.562442701	0.00195746	0.024697321
C7R98_RS12035	lpxD	132.711238	-1.204953961	0.491219693	0.00201702	0.02526957
C7R98_RS03960	DUF721 domain-containing protein	61.51648899	1.252989125	0.547883831	0.002059695	0.025623758
C7R98_RS11820	rplV	559.8934894	-1.180963694	0.475557578	0.002110852	0.026077818
C7R98_RS12310	acpP	1838.506486	-1.010593407	0.372562158	0.002152956	0.026414546
C7R98_RS11945	minE	112.7502871	-0.983531515	0.360971683	0.002236794	0.027255185
C7R98_RS01855	mltG	62.73663342	1.183446001	0.484127783	0.00226018	0.027352785
C7R98_RS12655	ABC transporter ATP-binding protein	139.5676793	1.197571921	0.497719544	0.002281596	0.027425397
C7R98_RS08985	dusA	136.6785684	1.237753672	0.548343738	0.002333096	0.02785623
C7R98_RS09220	basic amino acid ABC transporter substrate-binding protein	268.1922335	-0.762560081	0.266460609	0.002360784	0.027998903
C7R98_RS01995	lbpA	124.4050155	1.253671698	0.625718472	0.002380082	0.028005694
C7R98_RS05795	aldehyde dehydrogenase family protein	71.21152553	1.246842548	0.572530245	0.002392842	0.028005694
C7R98_RS04460	iron-sulfur cluster-binding protein	281.2901987	-1.039478165	0.394330081	0.002436404	0.028329166
C7R98_RS02645	amino acid ABC transporter ATP-binding protein	213.293265	1.213147796	0.546273588	0.002807314	0.032429949
C7R98_RS02490	DNA polymerase III subunit delta'	188.9366769	0.876366645	0.32112383	0.002838525	0.032578943
C7R98_RS00445	ribF	116.6218638	1.141170097	0.473564405	0.002906694	0.033147496
C7R98_RS01365	tRNA-Leu	36.64428802	1.18527944	0.685884417	0.002984838	0.033821822
C7R98_RS12755	methionine ABC transporter ATP-binding protein	80.86941979	1.218030634	0.575178505	0.003011541	0.033908424
C7R98_RS08920	glnA	342.3812961	-0.954445301	0.362099549	0.003079693	0.034242342
C7R98_RS11760	rpsL	1411.431886	-1.055130676	0.417631588	0.003064405	0.034242342
C7R98_RS06065	eno	956.6180357	-1.04101062	0.410184206	0.003116398	0.034435227
C7R98_RS11160	gap	287.7195573	-1.167030308	0.511104026	0.003269663	0.035905743
C7R98_RS11815	rpsS	418.6563465	-1.130290697	0.478163718	0.003335345	0.036402323
C7R98_RS07155	tRNA-Ala	51.09059302	1.179942393	0.532461806	0.003389073	0.036763172
C7R98_RS01665	pglE	577.7176528	1.110514321	0.464522228	0.003423315	0.036774673
C7R98_RS01975	rnc	75.87471074	1.160380175	0.511215831	0.003451681	0.036774673
C7R98_RS05130	Mth938-like domain-containing protein	54.30947868	1.076287245	0.4401021	0.003452148	0.036774673
C7R98_RS01300	tRNA-Gly	305.010353	1.164081667	0.521387167	0.003591751	0.038034081
C7R98_RS02825	restriction endonuclease subunit S	100.70769	-1.072674731	0.441716856	0.003634564	0.038259695
C7R98_RS02510	grxD	170.8888317	-0.919926896	0.353529506	0.003674824	0.038455957
C7R98_RS03765	DMT family transporter	62.08298029	1.186886617	0.570838008	0.003770924	0.038874666
C7R98_RS04760	DNA translocase FtsK	236.1932612	1.137254531	0.498761479	0.003780392	0.038874666
C7R98_RS11855	rplE	731.7526604	-1.144258653	0.504868829	0.003747281	0.038874666
C7R98_RS02345	ubiE	67.84830204	0.994330065	0.3967215	0.003880875	0.039678604
C7R98_RS00865	DUF1543 domain-containing protein	43.20502693	1.18531165	0.583941962	0.003926222	0.039831726
C7R98_RS04505	porin	66.91839991	1.167359601	0.661134859	0.003963022	0.039831726
C7R98_RS09400	tRNA-Ala	50.69087591	1.128296681	0.495351228	0.003952583	0.039831726
C7R98_RS00345	pilus assembly/adherence protein PilC	1700.030222	0.661029991	0.24137196	0.004006521	0.040042699
C7R98_RS05505	glucokinase	74.11720664	1.102831067	0.476311169	0.004109548	0.040842939
C7R98_RS03010	hypothetical protein	244.2882745	1.104808021	0.478917662	0.004143555	0.040952135
C7R98_RS12610	hupB	59.67039189	1.029692687	0.740803417	0.004304503	0.042307799
C7R98_RS02195	infC	580.4996207	-1.128331007	0.508442171	0.004363769	0.04265464
C7R98_RS11070	hypothetical protein	1965.309124	-0.995033919	0.405528076	0.004411001	0.042688379
C7R98_RS11865	rpsH	399.0509061	-1.123273606	0.504519411	0.004415212	0.042688379
C7R98_RS03800	nitronate monooxygenase	177.3378188	1.098190669	0.485149054	0.004620873	0.04443531
C7R98_RS02340	DUF971 domain-containing protein	62.24341705	1.113123454	0.503801868	0.004756962	0.045284033
C7R98_RS11610	DUF4390 domain-containing protein	94.32993234	1.150110611	0.556931835	0.004760042	0.045284033
C7R98_RS12805	atpE	233.0854159	-1.100144229	0.491783448	0.004807512	0.045492359
C7R98_RS11125	YigZ family protein	118.9536472	-0.947112078	0.384236325	0.004911779	0.046233091
C7R98_RS03085	superoxide dismutase	64.2626381	1.000140606	0.417236927	0.004970784	0.046542236
C7R98_RS10575	ssrS	1991.732007	1.150331827	0.622398068	0.005068694	0.046965979
C7R98_RS10900	SMI1/KNR4 family protein	109.4745645	-1.136163936	0.546067935	0.005068841	0.046965979
C7R98_RS11295	GIY-YIG nuclease family protein	81.91458834	1.150042103	0.618614694	0.005106491	0.047069675
C7R98_RS08345	pheA	155.225651	1.134397939	0.547085591	0.005157841	0.04729793
C7R98_RS11620	sigma-54-dependent Fis family transcriptional regulator	143.1502763	1.144210512	0.573458972	0.005209606	0.047527639
C7R98_RS03920	hypothetical protein	176.2209725	-1.095248752	0.508381573	0.005639841	0.050930339
C7R98_RS11915	rpsK	736.9335749	-1.089955144	0.502032786	0.00562491	0.050930339
C7R98_RS07250	paraquat-inducible protein A	82.64416154	1.136462116	0.603814846	0.005682088	0.051052702
C7R98_RS10145	NgoFVII family restriction endonuclease	298.5202108	-1.119309216	0.546038636	0.005746271	0.051369926
C7R98_RS06425	DnaJ domain-containing protein	57.43137513	1.083565436	0.500005151	0.00582488	0.051812311
C7R98_RS12180	fur	419.9432349	0.928854428	0.384904439	0.005864664	0.051906649
C7R98_RS10705	bifunctional biotin--[acetyl-CoA-carboxylase] ligase/type III pantothenate kinase	170.5054266	1.113029233	0.653342814	0.005993416	0.052783603

Supplementary Table 3 | List of DE genes emerging by comparing bacterial data from sorted samples at 120 mpi and bacterial data from unsorted samples at 120 mpi for End1 cell model (adj. p-value ≤ 0.05).

id	gene	baseMean	log2FoldChang	lfcSE	pvalue	padj
C7R98_RS01370	ilvB	406.6831906	1.354548092	0.260733902	6.85E-09	1.16E-05
C7R98_RS11130	electron transfer flavoprotein subunit alpha/FixB family protein	265.1942624	-1.32234443	0.298451353	3.01E-07	0.000253953
C7R98_RS03015	hypothetical protein	74.89960123	1.495992052	0.375258745	6.51E-07	0.000366694
C7R98_RS01920	acetyl-CoA carboxylase carboxyltransferase subunit beta	206.2400112	1.337929909	0.326540191	1.10E-06	0.000370194
C7R98_RS08955	lptC	111.9270062	1.298220296	0.312348532	1.06E-06	0.000370194
C7R98_RS09440	P-II family nitrogen regulator	211.6320461	-1.50777035	0.449635403	3.08E-06	0.000866046
C7R98_RS06490	methyltransferase regulatory domain-containing protein	178.3554523	1.352365701	0.374217958	5.49E-06	0.001325746
C7R98_RS05840	OFA family MFS transporter	82.48416589	-1.45488241	0.522054169	6.55E-06	0.00138384
C7R98_RS00745	NAD(P)H-dependent oxidoreductase	1415.692145	-1.34620949	0.381539993	7.39E-06	0.001387768
C7R98_RS11160	gap	287.7195573	-1.44346651	0.486265396	8.94E-06	0.001510484
C7R98_RS02020	ftsN	189.8643172	1.176408377	0.339195727	2.40E-05	0.003565404
C7R98_RS04330	metF	146.8931402	-1.34910701	0.456398996	2.53E-05	0.003565404
C7R98_RS08320	hypothetical protein	33.33190463	1.235482003	0.557069435	5.07E-05	0.006176339
C7R98_RS08330	holo-ACP synthase	63.520691	1.248432252	0.411100052	5.12E-05	0.006176339
C7R98_RS04540	hypothetical protein	152.5713496	-1.21758368	0.405069246	6.83E-05	0.007690472
C7R98_RS05180	lipoprotein-releasing ABC transporter permease subunit	106.004728	1.216759608	0.40885761	7.34E-05	0.007743309
C7R98_RS01825	nspA	516.1170535	-1.24841249	0.467258408	8.93E-05	0.008043739
C7R98_RS04325	type B 50S ribosomal protein L31	93.02464789	1.243434983	0.460532234	9.17E-05	0.008043739
C7R98_RS05385	hypothetical protein	1114.807723	-1.23497791	0.443551892	8.98E-05	0.008043739
C7R98_RS07575	biopolymer transporter ExbD	169.4413864	1.245472718	0.489287298	9.52E-05	0.008043739
C7R98_RS02370	hfq	999.0767496	-1.21288339	0.435997126	0.00010858	0.008335951
C7R98_RS02650	phosphomannomutase/phosphoglucomutase	446.5324671	-1.11223989	0.359127215	0.00010449	0.008335951
C7R98_RS02520	upp	114.1285908	-1.21658076	0.449141979	0.00011637	0.0085454
C7R98_RS01520	phosphatidylserine decarboxylase	109.7225969	-1.15784727	0.395834169	0.00012299	0.008655676
C7R98_RS04345	redoxin family protein	4264.566871	-1.1746761	0.412310055	0.00012905	0.008718341
C7R98_RS00345	pilus assembly/adherence protein PilC	1700.030222	0.806795239	0.239495084	0.00017044	0.008989106
C7R98_RS02030	transcriptional regulator	59.18814051	1.169904907	0.434035379	0.00017563	0.008989106
C7R98_RS04370	sdhA	352.2860309	0.875723633	0.265596582	0.0001681	0.008989106
C7R98_RS04700	YdcH family protein	777.64538	-1.20462964	0.506680373	0.00014146	0.008989106
C7R98_RS05975	hypothetical protein	64.366709	-1.18390566	0.523644608	0.00015483	0.008989106
C7R98_RS11605	rsmB	203.2185185	1.196953103	0.471917076	0.00016203	0.008989106
C7R98_RS12735	GatB/YqeY domain-containing protein	85.60024414	-1.13500961	0.398428311	0.00016969	0.008989106
C7R98_RS13225	conjugal transfer protein TraL	161.2769894	-1.20683117	0.470421338	0.00014458	0.008989106
C7R98_RS08110	NAD(P)H-dependent oxidoreductase	2091.443163	-1.16328281	0.514298352	0.00021	0.010432064
C7R98_RS12725	glutathione S-transferase N-terminal domain-containing protein	241.7499276	-0.97916528	0.320533154	0.00023378	0.011281481
C7R98_RS04290	efp	1424.540477	-1.15986595	0.476175135	0.00024519	0.011503599
C7R98_RS06295	carbonic anhydrase family protein	163.456456	-1.07956029	0.383366591	0.00025998	0.011867808
C7R98_RS09220	basic amino acid ABC transporter substrate-binding protein	268.1922335	-0.84123231	0.265700183	0.00029775	0.013234085
C7R98_RS05815	VirK/YbjX family protein	67.46116907	1.109521532	0.444780986	0.00037607	0.01628667
C7R98_RS01335	hisB	304.1640419	0.984704003	0.344260647	0.0003947	0.016666284
C7R98_RS08555	ftsL	64.86117928	1.090439463	0.436146549	0.00043638	0.017976928
C7R98_RS05970	IscS subfamily cysteine desulfurase	437.8614822	-1.04881574	0.400604885	0.0004835	0.019346095
C7R98_RS08335	pdxJ	131.3822073	1.089556969	0.455140725	0.00049253	0.019346095
C7R98_RS09755	ubiM	354.6227475	-0.94830654	0.335525378	0.00051903	0.019923701
C7R98_RS08215	rpsT	719.7482314	-1.07855197	0.462099623	0.00056693	0.021278917
C7R98_RS01175	2,3-butanediol dehydrogenase	1749.72882	-1.01050421	0.3930023	0.0006781	0.024898018
C7R98_RS02140	rpoH	1966.247959	-0.95812613	0.358950576	0.0007377	0.026510227
C7R98_RS02655	peptidylprolyl isomerase	742.1021885	-0.99959059	0.411446422	0.00094576	0.032172647
C7R98_RS08580	tRNA-Ala	30.55852089	0.801882558	0.591309821	0.00097127	0.032172647
C7R98_RS13820	pilin	1057.045419	-0.92266094	0.347866182	0.00092489	0.032172647
C7R98_RS11585	LysM peptidoglycan-binding domain-containing protein	844.1923667	0.749846317	0.258523961	0.00097147	0.032172647
C7R98_RS03010	hypothetical protein	244.2882745	1.016946113	0.454767756	0.00104755	0.034025147
C7R98_RS11145	nicotinamidase	222.7621076	-0.91788773	0.352916379	0.00107403	0.034227096
C7R98_RS10570	rplM	2405.024794	-0.9821226	0.416326551	0.00119313	0.037318428
C7R98_RS03420	hypothetical protein	261.7291886	-0.99539492	0.438919511	0.00121837	0.037414883
C7R98_RS04335	metE	847.8672148	-0.98853942	0.435576634	0.00128292	0.038186847
C7R98_RS05745	leuB	263.464243	-0.88331517	0.340090448	0.00128872	0.038186847
C7R98_RS04210	TonB-dependent receptor	238.7118781	0.918508907	0.37240105	0.00145037	0.042235705
C7R98_RS08150	23S rRNA methyltransferase	874.258644	0.77567992	0.28587222	0.00152051	0.043527784
C7R98_RS11610	DUF4390 domain-containing protein	94.32993234	0.964028312	0.50848889	0.00160862	0.045282755
C7R98_RS11935	minC	255.957612	-0.92059276	0.383664946	0.00164333	0.045501398
C7R98_RS11705	rplA	1003.127476	-0.96109898	0.433061316	0.00167354	0.045590583
C7R98_RS03885	transposase	34.35910262	0.950533919	0.515769824	0.00172359	0.046208501
C7R98_RS03000	panB	81.36715964	0.960311175	0.440424904	0.00175284	0.046258614
C7R98_RS06410	hypothetical protein	58.11375974	-0.95533938	0.49413743	0.00188835	0.048897882
C7R98_RS12140	Rne/Rng family ribonuclease	545.0924992	0.809019668	0.313945626	0.00191075	0.048897882
C7R98_RS08345	pheA	155.225651	0.942610942	0.499591315	0.00207446	0.052295056

Supplementary Table 4 | List of DE genes emerging by comparing bacterial data from sorted samples at 120 mpi and bacterial data from unsorted samples at 120 mpi for t-UEC cell model (adj. p-value ≤ 0.05).



REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
4',6-Diamidino-2-Phenylindole, Dihydrochloride (Dapi)	Molecular Probe	D1306
Oregon Green™ 488 Carboxylic Acid, Succinimidyl Ester, 6-isomer	ThermoFisher	O6149
Alexa Fluor 568 Goat anti-mouse	ThermoFisher	A-11011
CellMask Green Actin Tracking Stain	ThermoFisher	A57243
Bacterial Strains		
Neisseria Gonorhoeae WHO M	WHO	N/A
Experimental Models: Cell Lines		
End1/E6E7	ATCC	CRL2615
Transformed urethral epithelial cells (t-UEC)	Hillery et al. 2002	N/A
Detroit 562	ATCC	CCL-138
Critical Commercial Assays		
Live/Dead Fixable Aqua Dead cells stain kit	Invitrogen	L34957
RNA Protect Bacteria Reagent	Qiagen	76506
Direct-zol RNA miniprep kit	Zymo	R2073
Turbo DNA-free DNAse kit	Invitrogen	AM1907
SuperScrip III Platinum SYBR Green One-Step qRT-PCR Kit with ROX	Invitrogen	11746-100
Universal RNA-Seq with NuQuant Human AnyDeplete kit	Nugen	0364
Agencourt RNAClean XP Beads	Beckman Coulter	A63987
Agilent High Sensitivity DNA Kit	Agilent	5067-4626
Agilent RNA 6000 Nano Kit	Agilent	5067-1511
NovaSeq 6000 S1 Reagent Kit v1.5	Illumina	20028318

Supplementary Table 5 | Key source table.

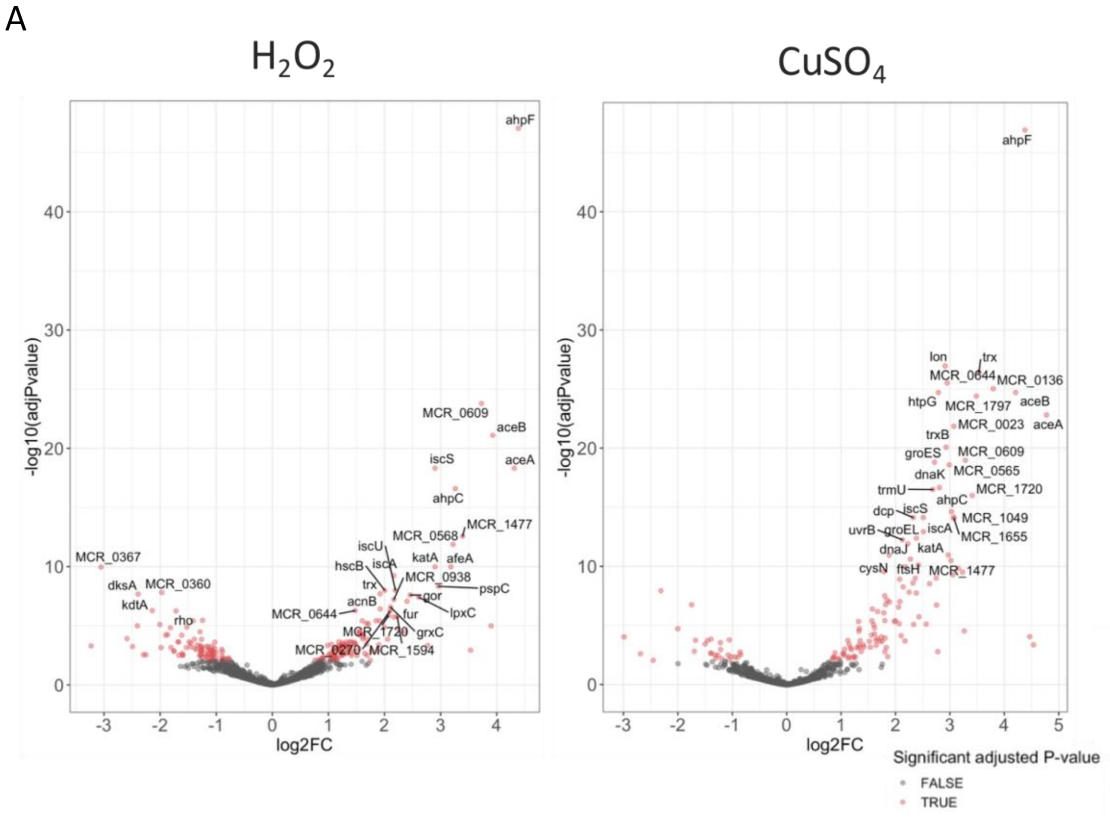
APPENDIX

Publications:

Accepted article

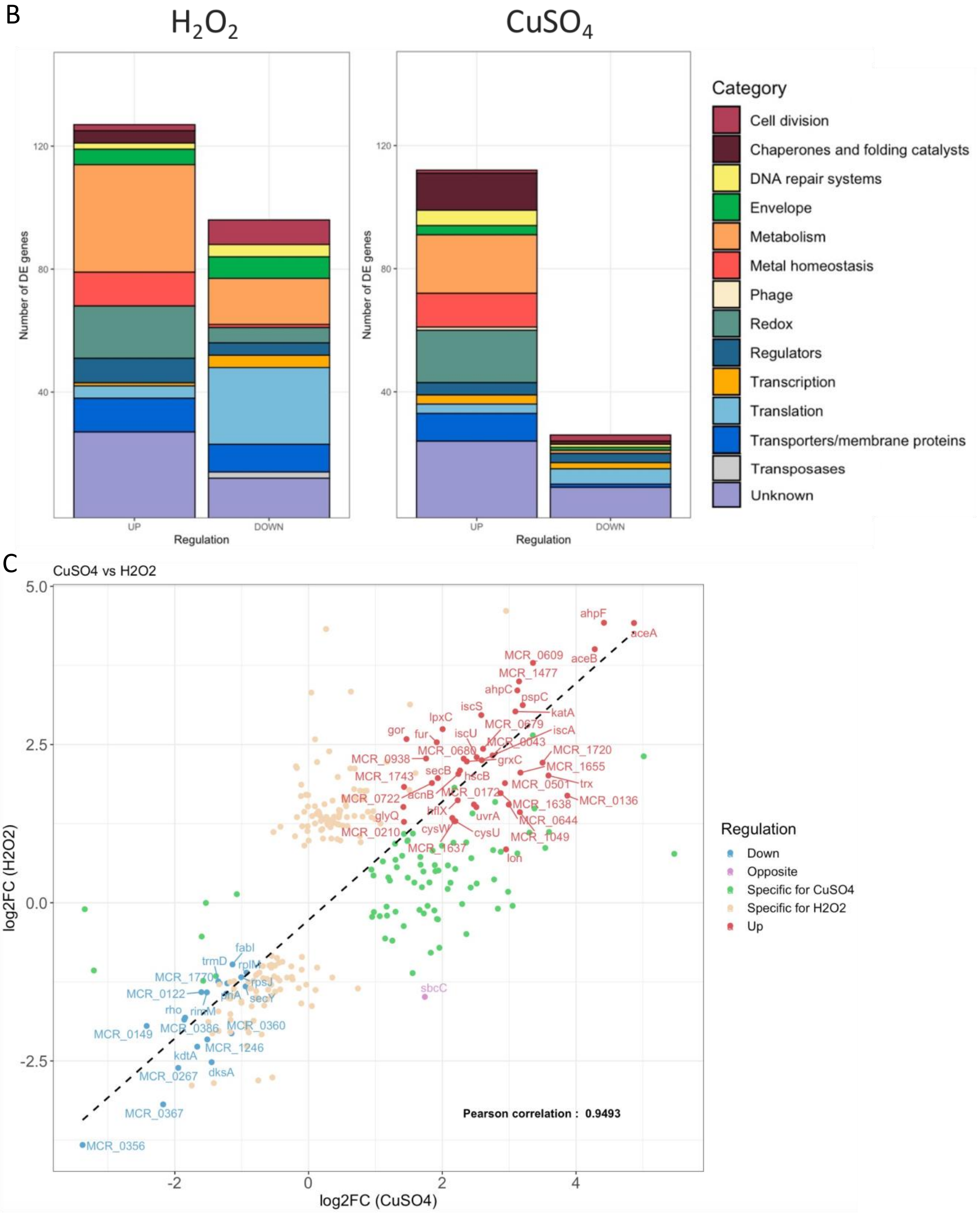
"*Moraxella catarrhalis* evades neutrophil oxidative stress responses providing a safer niche for NTHi", Nicchi S., Giusti F., Carello S., Tavarini S., Frigimelica E., Ferlenghi I., Rossi Paccani S., Merola M., Delany I., Scarlato V., Maione D., Brettoni C., *iScience*

I also collaborated in an RNA-Seq study on *Moraxella catarrhalis* (Mcat) BBH18 response to oxidative stress generated by sublethal concentrations of H_2O_2 and $CuSO_4$ (manuscript in preparation) (Appendix Figure).



N° of DE genes: 225 (128 UP, 97 DOWN)

N° of DE genes: 140 (113 UP, 27 DOWN)



Appendix Figure | Representative figure about the study on Mcat BBH18 global transcriptional responses to H_2O_2 and $CuSO_4$ and their overlap. A) Volcano plot from DE analysis comparing exponentially growing bacteria exposed to sublethal levels of H_2O_2 (on the left) or $CuSO_4$ (on the right) with unexposed ones. Dots for genes are colored accordingly to their significance. Names for the top 30 significantly DE genes are reported. On the bottom, the number of emerging DE genes is indicated. Adjusted p-value to call DE genes ≤ 0.01. **B)** Up- and down-regulated genes from the same comparisons in **A** classified accordingly to 13 functional categories. **C)** Plot comparing $\log_2$ fold changes for all the regulated genes emerging from at least one of the comparisons in **A**. The correlation straight line is drawn on the basis of commonly up- or down-regulated genes only.

ACKNOWLEDGEMENTS

This work was conducted at GSK Vaccines, Siena, in the Data Science and Computational Vaccinology (DSCV) group in collaboration with Pre-clinical laboratories of the site. First of all, I would like to thank my supervisors, Alessandro Brozzi in GSK and Michele Caselle at the University of Turin, for their constant support and guidance during my PhD. I'm very thankful to all (past and present) members of my group in GSK for hosting me during these three years and giving me the opportunity to participate in team events and discussions that helped me to grow. A special thank of course to the pre-clinical lab people that collaborated with me and that made this work possible, in particular Chiara Limongi, Silvia Guidotti and Stefano Censini. Among pre-clinical lab members, I would like to thank also Sonia Nicchi, for passionately collaborating with me in these years on the works related to *Moraxella catarrhalis*, even if they do not constitute the subject of my thesis. A big thank-you also to Isabel Delany, Rosanna Leuzzi, Mariagrazia Pizza and Cecilia Brettoni for participating in scientific discussions and giving me advices during my work. Finally, I would like to thank the whole GSK PhD Academy, PhD students and friends that I met in the past three years: I truly felt at home, and their "invisible" support was essential for my work as well. Of course, I'm always thankful to my family and old friends for constantly being by my side.

TRANSPARENCY STATEMENT

Stefano Carello is a PhD student at the University of Turin, with a scholarship supported by GlaxoSmithKline Biologicals SA. Alessandro Brozzi (Senior Data Scientist) is the PhD supervisor and mentor at GSK, Michele Caselle (Full Professor) is the University PhD supervisor.