

Genome-wide RNAi screen reveals host factors involved in *Brucella* infection

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie
vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultät
der Universität Basel

von

Shyan Huey Low
aus Kuala Lumpur, Malaysia

Basel, 2015

Originaldokument gespeichert auf dem Dokumentenserver
der Universität Basel
edoc.unibas.ch

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät
auf Antrag von

Prof. Dr. Christoph Dehio

Prof. Dr. Jean Pieters

Basel, den 24.06.2014

Prof. Dr. Jörg Schibler
Dekan

For my parents

Statement to my Thesis

This work was carried out in the group of Prof. Christoph Dehio in the Focal Area Infection Biology at the Biozentrum of the University of Basel, Switzerland.

My PhD thesis committee consisted of:

Prof. Dr. Christoph Dehio

Prof. Dr. Jean Pieters

Prof. Dr. Martin Spiess

My thesis is written in a cumulative format. It consists of an introduction section covering various aspects related to my work and is followed by the result sections that are composed of two published manuscripts, a manuscript in preparation and some unpublished results. Finally, I summarize the major findings of this thesis, providing suggestions for the next steps of this project.

Table of contents

1. Introduction.....	4
1.1 Intracellular compartments and trafficking pathways.....	4
1.1.1 Endoplasmic reticulum (ER) and Golgi apparatus.....	5
1.1.2 Exocytic pathway.....	6
1.1.3 Endocytic pathway.....	8
1.1.3.1 Mechanisms of uptake.....	9
1.1.3.2 Endosomal maturation.....	10
1.1.4 Golgi to ER trafficking.....	10
1.1.5 Trafficking pathways and pathogenesis.....	11
1.2 <i>Brucella</i> -host interactions.....	14
1.2.1 The genus <i>Brucella</i>	14
1.2.2 <i>Brucella</i> and different hosts cell types.....	15
1.2.3 <i>Brucella</i> intracellular trafficking.....	16
1.2.3.1 Adhesion and entry.....	17
1.2.3.2 Trafficking along the endocytic pathway and VirB type IV secretion system (T4SS).....	18
1.2.3.3 Survival in the replicative niche and egression.....	19
1.3 Systems biology.....	20
1.3.1 RNA interference (RNAi).....	21
1.3.1.1 RNAi mechanism.....	21
1.3.1.2 Small interfering RNA (siRNA), enzymatically generated siRNAs (esiRNA), short hairpin RNA (shRNA) and microRNA (miRNA).....	22
1.3.1.3 RNAi as a tool – pros and cons.....	23
1.3.2 Genome-wide RNAi screening to study systems-level host pathogen interaction.....	24
1.3.2.1 RNAi screening for host factors important for viral and bacterial pathogens.....	24
1.3.2.2 High content microscopy and multiparametric analysis in RNAi screening.....	25

1.4 References.....	27
2. Aim of thesis.....	45
3. Results	47
3.1 RESEARCH ARTICLE I	47
3.1.1 Summary.....	48
3.1.2 Manuscript.....	49
3.2 RESEARCH ARTICLE II.....	66
3.2.1 Summary.....	67
3.2.2 Manuscript.....	68
3.3 RESEARCH ARTICLE III (in preparation).....	87
3.3.1 Manuscript.....	112
Genome-wide siRNA screen in HeLa cells reveals host factors involved in <i>Brucella</i> infection	112
3.4 Unpublished results: Transforming-growth factor beta (TGF- β) signaling and <i>Brucella</i> infection	147
3.5 Unpublished results: The role of retromer complex in <i>Brucella</i> infection.....	182
3.6 Additional tools developed – understanding <i>Brucella</i> intracellular lifestyle by fluorescent microscopy and / or electron microscopy	198
4. General conclusions and outlook	226
4.1 Genome-wide siRNA screen reveals novel host signaling pathways involved in <i>Brucella</i> infection.....	226
4.2 Active TGF- β signaling increases <i>Brucella</i> entry in HeLa cells.....	228
4.3 Vps35, a retromer complex component is required for <i>Brucella</i> trafficking to its replicative niche	230
5. References	234
6. Acknowledgements	248
7. Curriculum vitae	251

INTRODUCTION

1. INTRODUCTION

1.1 Intracellular compartments and trafficking pathways

Eukaryotic cells contain membrane-bound intracellular compartments that carry out specialized functions, with communication between these compartments achieved via vesicular transport. Vesicular transport of proteins and lipids occurs via two major pathways: the exocytic pathway that carries material from the cytoplasm to the cell surface and the endocytic pathway that internalizes material from the environment into the cell (Figure 1). These two pathways are highly connected with disruption of one of the pathways commonly leading to a dysfunction of the other pathway (1, 2).

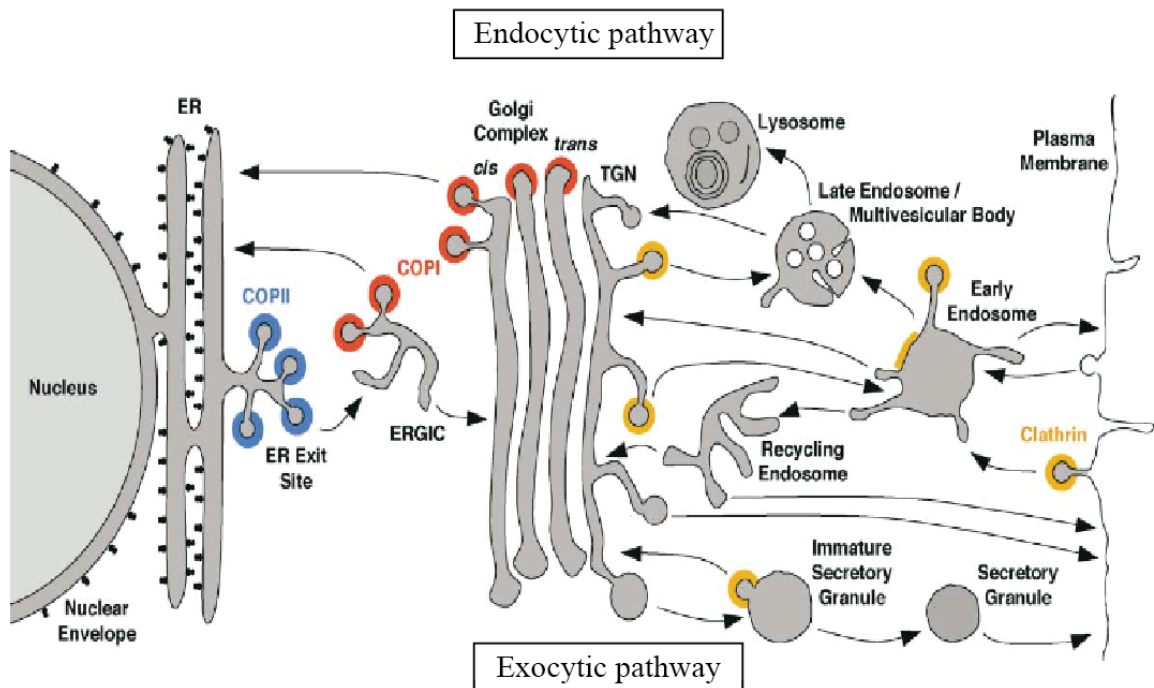


Figure 1 Intracellular trafficking pathway. Compartments of various intracellular pathways are depicted, covering the endocytic pathway, Golgi to ER transport as well as ER to Golgi transport. COPII (blue), COPI (red) and clathrin (orange) are indicated at their locations, with COPII labeling ER exit sites. Golgi is composed of *cis*-, medial- and *trans*-cisternae while only the rough ER that is associated with ribosomes is shown in this scheme. Picture is taken from (3) and adapted.

1.1.1 Endoplasmic reticulum (ER) and Golgi apparatus

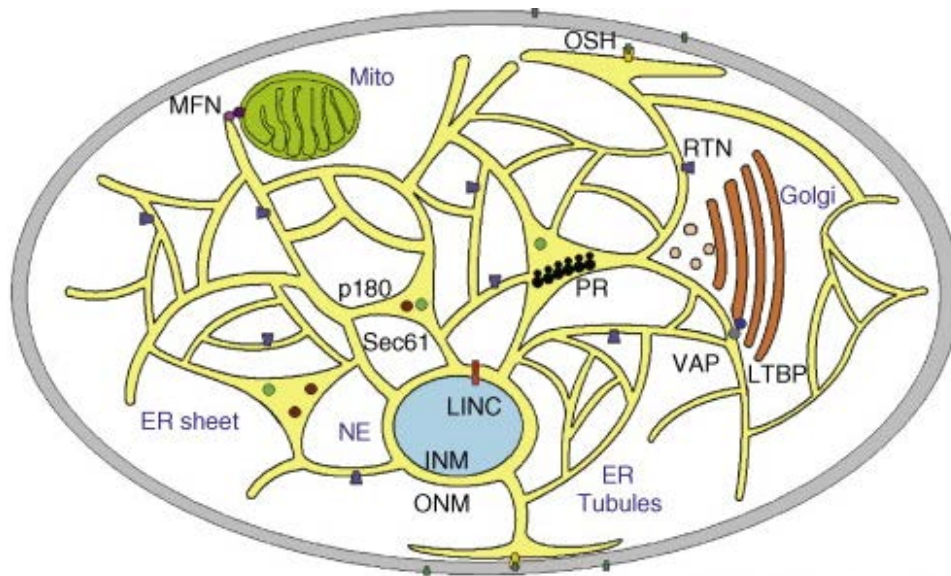


Figure 2 Scheme shows a cell with different ER subcompartments (ER sheets, ER tubules, ER exit site – not highlighted) and interaction with other compartments of the cell. Tubular ER is shaped by the reticulon (RTN) while sheet-like ER is shaped by polyribosomes (PR), and Sec61 translocon complex. The peripheral ER is connected to the mitochondria via mitofusin (MFN) proteins, with the Golgi through VAP proteins that interact with lipid transfer binding proteins (LTBP), and plasma membrane possibly through Osh proteins. The nucleus is represented in blue, with the inner membrane (INM) and outer membrane (ONM) of the nuclear envelope (NE) being linked by the LINC complex. Picture is taken from (4) and adapted.

The ER is a continuous membrane system that is comprised of the nuclear envelope as well as a peripheral network of tubules and sheets (5) (Figure 2). Its main function includes protein-synthesis, protein folding and modification, and the quality control of proteins before being exported to other compartments. The ER is equally involved in lipid synthesis, regulation of Ca^{2+} homeostasis and secretion (6). The ER can be classified into smooth ER (SER) and rough ER (RER). The sheet-like RER is associated with ribosomes that synthesize secretory and membrane proteins, while the SER is devoid of ribosomes and has a more tubular structure. The ER is closely associated with mitochondria, Golgi, endosomes, lysosomes, peroxisomes, and plasma membrane to allow transfer of proteins, lipids, and intracellular signals (Figure 2). Interaction of the ER with the cytoskeleton plays a key role in its dynamics and distribution (7).

Introduction

The Golgi apparatus serves as a platform connecting anterograde and retrograde trafficking (8). Most proteins that are synthesized in the ER are transported to the Golgi. The latter is a major site of glycosylation for many proteins and lipids, and also of carbohydrate synthesis (9). It also serves as a platform for binding of various signaling and sorting proteins (10). The Golgi is separated into *cis*-, medial- and *trans*-cisternae with the *cis*-side directly communicating with the ER while the *trans*-Golgi network (TGN) performs final steps of protein sorting before delivery to their final destination (11) (Figure 1).

1.1.2 Exocytic pathway

The exocytic or secretory pathway is involved in anterograde transport of cargo from the ER to the Golgi and finally to the plasma membrane (PM) (Figure 1). Cargoes of this pathway include soluble proteins to be secreted to the extracellular environment as well as membrane protein and lipid components of the PM. Proteins enter the ER during their translation via the pore of the Sec61 translocon (12). This transfer is mediated by the presence of signal sequences on the nascent protein and the signal recognition particle, a complex mediating the link between newly synthesized peptide and the translocon in the ER membrane. ER resident proteins bear a retention signal that defines their permanent localization in the ER, while proteins that leave the ER upon proper folding and assembly exit via regions called ER exit sites (ERES) (13) (Figure 1).

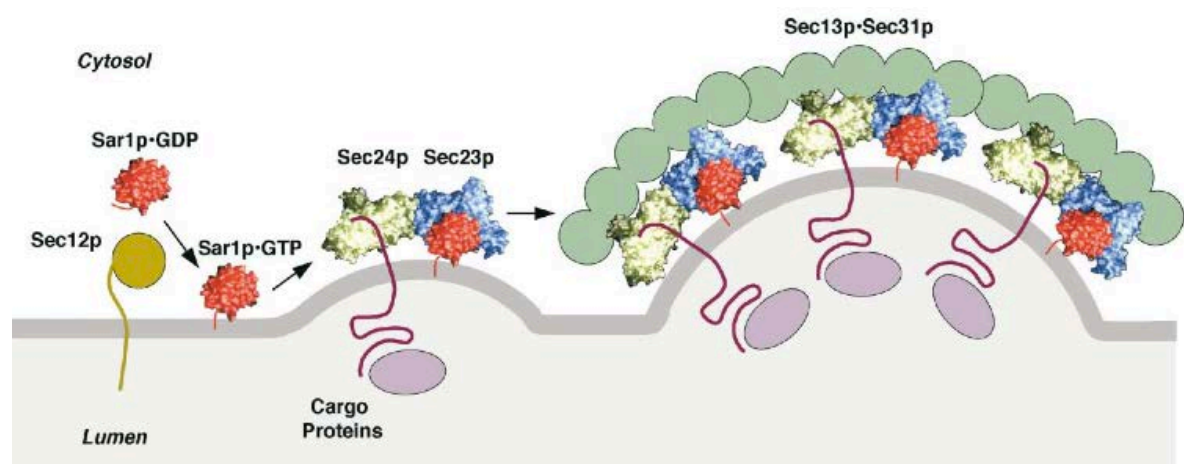


Figure 3 COPII coat assembly. Sec12 converts Sar1-GDP to Sar1-GTP. Sar1-GTP together with Sec23-Sec24 forms the pre-budding complex, with Sec24 involved in cargo recognition and Sec23

Introduction

binding to Sar1-GTP. Sec13-Sec31 complex then polymerize as the outer layer of the COP complex, leading to membrane deformation and eventually vesicle budding. Picture is taken from (3).

At ERES, COPII-coated vesicles are formed and mediate protein export (14). The COPII coat is composed of the small GTPase Sar1 and the protein subcomplexes Sec23-Sec24 and Sec13-Sec31. The formation of the COPII coat at ERES is believed to be initiated by Sec16 that localizes to ERES and forms a scaffold that recruits COPII subunits (15). Sec12 converts cytosolic Sar1-GDP to membrane bound Sar1-GTP. Sar1-GTP together with Sec23-Sec24 form the pre-budding complex, with Sec23 making direct contact with Sar1-GTP while Sec24 is involved in cargo recognition. Sec13-Sec31 subunits then polymerize as the outer layer of the COPII complex, leading to the deformation of the ER membrane needed to drive transport vesicle formation (16) (Figure 3).

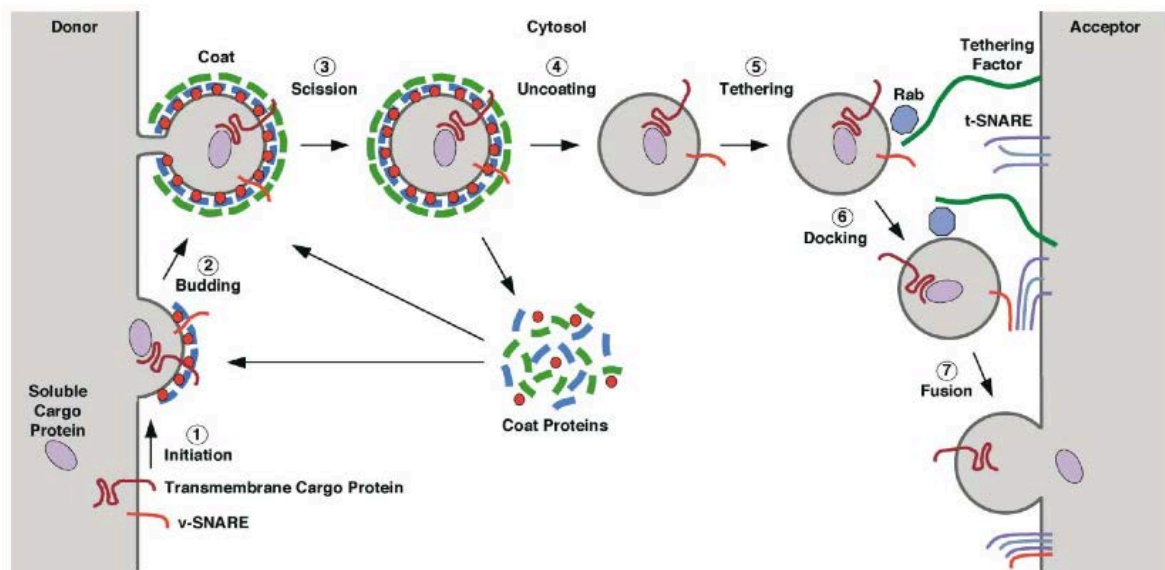


Figure 4 Budding and fusion. 1) Initiation of coat assembly involves recruitment of coat components (blue) by binding to membrane-associated GTPase (red) e.g. Sar1. Cargo proteins and SNAREs gather at the assembling coat. 2) Budding of the vesicle occurs upon assembly of the coat, in which coat proteins cause membrane curvature, leading to deformation of the membrane. 3) Scission occurs by direct action of the coat or accessory proteins. 4) Uncoating of the coat due to e.g. inactivation of small GTPase. Cytosolic coat proteins are recycled for additional rounds of vesicle budding. 5) Tethering occurs when uncoated vesicle moves to the acceptor compartment and is tethered by GTP-bound Rab protein and a tethering factor. 6) Docking via assembly of v- and t-SNAREs. 7) Fusion events are promoted by SNARE complex, allowing cargo to be transferred to the acceptor compartment and SNAREs to be recycled. Picture is taken from (3).

Introduction

Vesicles are then transported across to the *cis*-Golgi via the region between the ER and Golgi, the ER-Golgi intermediate compartment (ERGIC) (17, 18) (Figure 1). Hydrolysis by Sar1 destabilizes the COPII-vesicle coat, allowing fusion between vesicle and target membrane (19). Tethering factors, for example p115 and GM130, are recruited and tether vesicles to the acceptor membrane (20-22) while SNARE proteins, for example syntaxin-6 and Sec22B, mediate the membrane fusion of vesicles and target membrane (21, 23) (Figure 4). Early studies also suggest a role of Rab2 in anterograde transport, as an inactive form of Rab2 has a negative effect on transport of vesicles from the ER to the Golgi (24). To balance the anterograde transport of secretory cargo, organelle homeostasis requires retrieval of material. This recycling of membrane and protein components is accomplished via the retrograde trafficking pathway.

1.1.3 Endocytic pathway

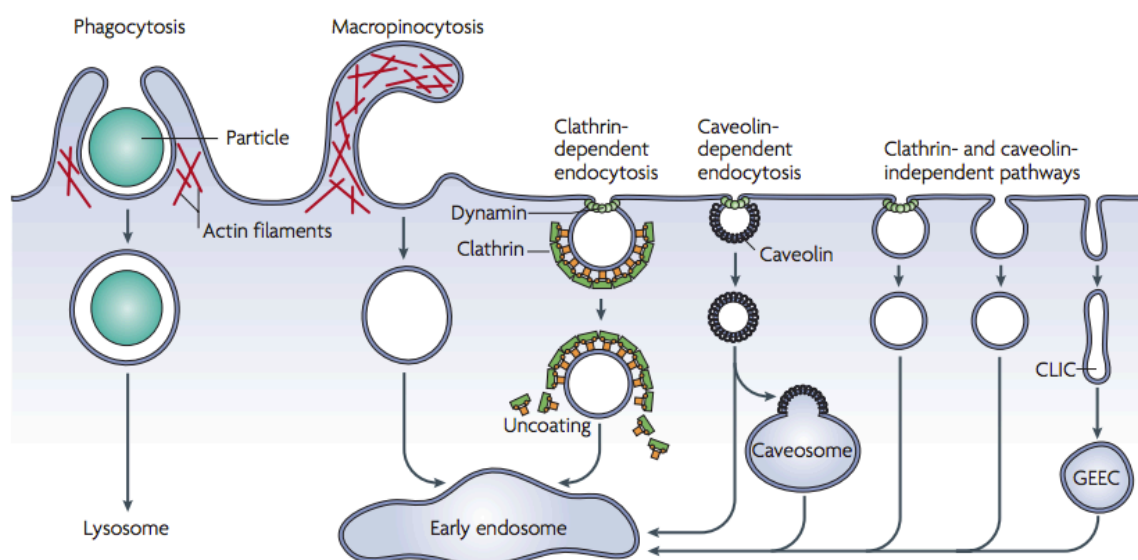


Figure 5 Mechanisms of uptake in the cell. Picture shows phagocytosis that involves taking up of large particles, fluid uptake via macropinocytosis, clathrin-dependent endocytosis through the formation of clathrin-coated pits and various clathrin-independent (CI) pathways. Phagocytosis and macropinocytosis are both triggered by actin-mediated remodeling of the PM. CI pathways include caveolin-dependent endocytosis as well as clathrin- and caveolin- independent pathways. Some pathways traffic through intermediate compartments e.g. caveosome or glycosyl phosphatidylinositol-anchored protein enriched early endosomal compartments (GEEC) before arriving at the early endosomal compartment. Dynamin is required in clathrin- and caveolin-dependent endocytosis as well as in certain clathrin- and caveolin- independent pathways. Picture is taken from (25).

1.1.3.1 Mechanisms of uptake

Cells are in constant contact with their environment and extracellular materials are taken up via different endocytic routes including phagocytosis, macropinocytosis, clathrin-dependent endocytosis and various clathrin-independent endocytic pathways (Figure 5). Actin cytoskeleton remodeling and regulation are important for the function of all these endocytic processes (25). Phagocytosis involves internalization of large particulate material and is initiated by binding of the particle to a cell surface receptor. This leads to reorganization of the PM and associated cytoskeletal elements and finally culminates in particle engulfment. Macropinocytosis is a non-selective mode of endocytosis that engulfs fluids and all associated solutes surrounding the cells. Remodeling of the cortical actin cytoskeleton during macropinocytosis causes membrane spreading and ruffling and involves Ras, Rac, Cdc42, and Rho.

Clathrin-coated pits are common entry points for cell surface receptors that bind ligands to be internalized, a process therefore also termed receptor-mediated endocytosis. Receptors, e.g. low-density lipoprotein receptor (LDLR), epidermal growth factor receptor (EGFR), transforming growth factor receptor (TGFBR), insulin receptor, and their respective ligands are internalized via this route (26-28). For the establishment of a clathrin-coated structure, Arf-GTP recruits specific phosphoinositides (PI) that favor the binding of clathrin adaptors to the membrane (29). Adaptor proteins also bind to cargo proteins by recognizing sorting signals found in their cytosolic domains (30, 31). Clathrin adaptors, for instance heterotetrameric AP2 complex (α -adaptin, β 2-adaptin, μ 2-chain, σ 2-chain), form complexes onto which the clathrin coat is subsequently assembled. Clathrin and the respective adaptor complex polymerize into cage-like structures and scission of the vesicle depends on accessory factors such as dynamins (31). Uncoating of the vesicle with the help of cytosolic chaperones Hsc70 and auxilin (32) then allows fusion of the vesicle with its target membrane.

One of the most prominent clathrin-independent (CI) endocytic pathways is caveolae / caveolin-mediated endocytosis that relies on dynamin and involves membrane fractions enriched in sphingolipids, cholesterol, signaling proteins, and glycosyl phosphatidylinositol-anchored proteins (GPI-APs) (33, 34). Nevertheless, there is also a plethora of clathrin and caveolae-independent endocytic mechanisms that are not well characterized. Since there are no adaptors reported for recruiting cargoes in CI

Introduction

endocytosis, cargoes are selected based on specific internalization signals with an example being ubiquitylation (35).

1.1.3.2 Endosomal maturation

Upon reaching the early endosome (EE), housekeeping receptors and certain proteins are recycled back to the PM, directly from EE via a Rab4-dependent mechanism or indirectly via recycling endosomes in a Rab11-dependent manner (36) (Figure 1). For proteins traveling from EE to the TGN, the retromer complex mediates sorting of endosomal cargo destined for the TGN (37). This pathway will be discussed in more detail in the introduction section to Results Part 3.5 in this thesis. Proteins destined for degradation travel from the EE to the late endosomes (LE) (Figure 1). Vacuolar-ATPase (v-ATPase), a multi-subunit proton pump acidifies EE and LE and the switch from early to late endosomes is driven by conversion from Rab5 to Rab7 (38, 39). Prior to degradation, proteins that need to be downregulated are sorted into luminal invaginations of the EE that pinch off as cargo-containing intraluminal vesicles (ILVs). EEs with ILVs form free multivesicular bodies (MVBs), eventually fusing with LE (Figure 1). This is mediated by ESCRT complex (ESCRT 0, I, II and III) that recruits the receptor to be downregulated into ILVs (40, 41). Upon reaching the LE stage, fusion with lysosomes forms endo-lysosomes that mature into lysosomes (42).

1.1.4 Golgi to ER trafficking

At the *cis*-Golgi, COPI mediates retrograde transport from Golgi to ERGIC and then to the ER (Figure 1). COPI vesicle formation begins by recruitment of GBF1, the Arf1 guanine nucleotide exchange factor (GEF). This process requires the presence of phosphatidylinositol-4-phosphate (PI4P) (43-45). Localization of GBF1 determines the location of small GTPase Arf1 activation. Upon activation of Arf1, it is recruited to the Golgi where it initiates binding of the heptameric coat complex, the coatomer. The tetrameric complex of β -COP, γ -COP, δ -COP, and ζ -COP constitutes the inner core of the coat, while the trimeric complex of α -COP, β' -COP, and ϵ -COP forms the outer layer of the coat and imposes membrane deformation (46, 47). Coatomer subunits α -COP, β' -COP, γ -COP and δ -COP recognize sorting motifs in the cytosolic domain of membrane cargoes and mediate incorporation of these cargoes into COPI vesicles (48) (Figure 6). Finally, ArfGAP 2/3 stimulate GTP hydrolysis by Arf1 (49-51), allowing release of Arf from the complex and subsequent coat dissociation. Rab1

Introduction

GTPase is involved in the retrograde transport of COPI vesicles from Golgi to the ER, with tethering complex syntaxin 18 and SNARE proteins allowing subsequent fusion with the target membrane to occur (52, 53).

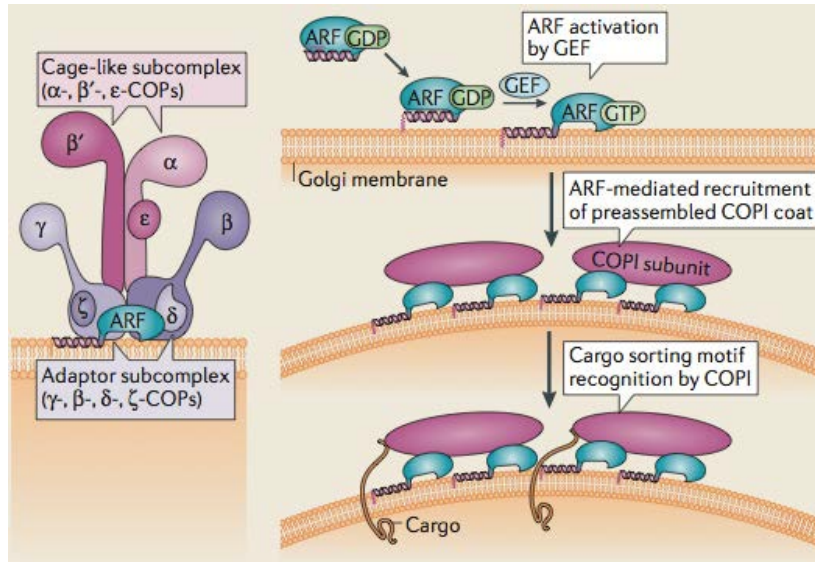


Figure 6 Heptameric COPI complex and action. ARF is activated by GEF after which it mediates recruitment of COPI coat. Cargo sorting motif is recognized by COPI coat that incorporate these cargoes into the vesicle. Picture is taken from (16).

The GAPDH interacts with Rab2 and has also been shown to be required for retrograde transport from Golgi to ER, with Rab2 modulating protein transport by recruiting GAPDH that is able to form an active complex with PKC_i/λ and COPI, localizes to vesicular and tubular clusters (VTC) between the ER and the Golgi (54). There is also the COPI-independent pathway that transports proteins from Golgi to the ER via a Rab6A-dependent pathway (55) or a Rab6-COPI independent (56) pathway, mainly utilized by toxins and functioning in parallel to the classical COPI retrograde pathway.

1.1.5 Trafficking pathways and pathogenesis

Pathogens invade host cells to escape the host immune's response and to take advantage of the nutrient sources available in the cell. Endocytic pathways are exploited to reach the cell interior where interaction with different compartments of the cell occurs. This promotes the subsequent arrival at their final destination. The

Introduction

interaction of pathogens with the host's intracellular pathway at different stages will be introduced briefly in this section.

Bacterial pathogens secrete toxins into the host cells to subvert the functions of the host. A few prominent examples include *Shigella dysenteriae* and enterohemorrhagic *Escherichia coli* that secrete Shiga toxin, and *Vibrio cholera* that secretes cholera toxin. These toxins contain two subunits A and B, with B subunit binding to specific glycolipids of the host cell (57) and A subunit disrupting protein synthesis via binding to the ribosome. Due to its toxicity, B subunit only of these toxins is commonly used in the field of biological research. Shiga toxin subunit B (StxB) is internalized by clathrin-dependent and CI endocytosis (58-60) and cholera toxin subunit B (CtxB) via a caveolae-independent route (61). Upon internalization, toxins are transported along the retrograde pathway via a Rab6A-dependent pathway (55). Escape of Stx from the early endocytic pathway to enter the retrograde pathway depends on clathrin (59), its adaptor epsinR (62) and the retromer complex (63). Upon binding of Stx to its receptor Gb3, the delta isoform of the protein kinase C (PKC δ) gets activated. This goes along with rapid phosphorylation of the clathrin heavy chain (CHC) that is regulated by spleen tyrosine kinase (Syk). These processes are important for transport of Stx from early endosomes to the Golgi (64, 65). The plant toxin ricin also enters the cell via clathrin-dependent and -independent pathways (66) and is retrogradely transported to the ER. However, transport of ricin to the ER is highly inefficient with only 5% of toxin arriving at the ER while the rest is recycled back to the cell surface or degraded in the lysosomes (67). Only a subset of host components is shared between the ricin and Shiga toxin for their transport to the ER (68).

Several bacterial pathogens such as *Mycobacteria tuberculosis* and *Brucella* enter the cell via phagocytosis in an unspecific uptake process together with extracellular fluid. Since only a subset of cells is able to perform phagocytosis, bacteria have also developed strategies to actively induce uptake into non-phagocytic cells. This generally occurs either via the trigger or the zipper mechanism (69). The trigger mechanism is used by *Salmonella enterica*, *Shigella flexneri*, and *Pseudomonas aeruginosa*. Bacteria bind to specialized lipid membrane microdomains that are enriched in cholesterol and sphingolipids, activating their type III secretion system that leads to translocation of effectors into the host cytosol (70-72)(Figure 7). These factors manipulate the host signaling and cytoskeleton organization in a way to facilitate and promote bacterial uptake. The zipper mechanism that is used by *Listeria*

Introduction

monocytogenes and *Yersinia pseudotuberculosis* instead engages specific receptors of the target cell, leading to moderate actin remodeling and less dramatic alteration of the host cell surface. In the case of *Listeria* infection, these surface molecules are E cadherin and Met (73, 74) (Figure 7).

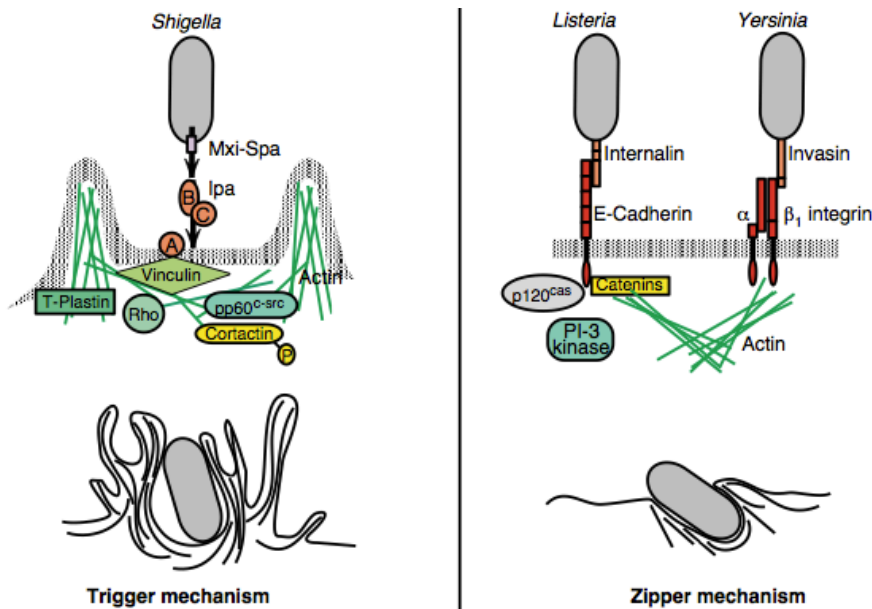


Figure 7 Mechanisms of bacterial invasion. Representation of trigger mechanism and zipper mechanisms that are used by *Shigella*, *Listeria*, and *Yersinia* respectively during invasion of host cells. Picture taken from (69) and adapted.

Clathrin, dynamin, and several other components of the endocytic machinery have been shown to colocalize with the bacterial entry site and are essential for invasion of *Listeria monocytogenes* (75). While clathrin is crucial for internalization of ‘zippering’ bacteria, it is not required for entry of ‘triggering’ bacteria (76). There are also pathogens that cannot be unambiguously assigned to one or the other uptake mechanism. In the case of *Bartonella henselae* for instance, the bacterium requires interactions with integrin β1 and activation of this receptor (77) – a characteristic of the zipper mechanism. However, *B. henselae* also translocates effectors that lead to bacterial aggregation, their engulfment and internalization by a unique actin-surrounded structure termed ‘invasome’ (78).

1.2 *Brucella*-host interactions

1.2.1 The genus *Brucella*

Brucella is a gram negative, facultative intracellular pathogen that belongs to the alpha-2 subdivision of *Proteobacteria* that includes other intracellular pathogens such as *Agrobacterium tumefaciens*, *Rickettsia*, and *Bartonella* species. It is a zoonotic pathogen that infects humans as incidental host. 11 species of *Brucella* have been described thus far (79, 80) (Table 1) with a wide range of reservoir hosts, of which 6 were shown to be pathogenic to humans. These include *Brucella melitensis* that infects goats, sheep, and camels as natural hosts, *Brucella suis* that infects pigs and *Brucella abortus* that causes bovine brucellosis. These three species are responsible for most of the reported infections in humans. *Brucella canis* (dogs), *Brucella ovis* (sheep and rams) and *Brucella neotomae* (desert wood rats) are of lower pathogenicity to humans (81).

Species	Host preference	Zoonotic potential ^a
<i>Brucella melitensis</i>	Sheep, goat (<i>Ovis</i> spp. and <i>Capra</i> spp.)	High
<i>Brucella abortus</i>	Cattle (<i>Bos taurus</i> and <i>Bos indicus</i>)	Moderate
<i>Brucella suis</i>	Pig (<i>Sus scrofa</i>)	Moderate
<i>Brucella canis</i>	Dog (<i>Canis lupus familiaris</i>)	Mild
<i>Brucella ceti</i>	Dolphin, porpoise, whale (Cetacea)	Mild
<i>Brucella pinnipedialis</i>	Seal (Pinnipedia)	Mild
<i>Brucella inopinata</i>	Unknown	Mild
<i>Brucella ovis</i>	Sheep (<i>Ovis</i> spp.)	No reported infections
<i>Brucella neotomae</i>	Desert woodrat (<i>Neotoma lepida</i>)	No reported infections
<i>Brucella microti</i>	Common vole (<i>Microtus arvalis</i>)	No reported infections
<i>Brucella</i> sp. (baboon isolate)	Baboon (<i>Papio</i> spp.)	No reported infections

^aBased on the number of human cases reported and depends on a combination of exposure to the pathogen and infectivity.

Table 1 *Brucella* species with their host preference and zoonotic potential. Table is taken from (81).

Brucella causes animal and human brucellosis, being the most important zoonotic bacterial pathogen with about 500,000 new human cases annually worldwide (82). *Brucella* is transmitted to humans via aerosols, direct contact with infected animals, or ingestion of contaminated food products while human-to-human transmission has not been reported. In animals, brucellosis leads to sterility, abortion or the birth of weak offspring due to the infection of the reproductive organs (81). In humans, *Brucella*

causes a febrile disease with relatively unspecific symptoms such as undulant fever and body aches (Malta fever). Without treatment, this can lead to a chronic infection of persistent bacteremia, endocarditis, or meningitis. There is currently no vaccine available for humans and treatment includes a combination of different antibiotics for a long period of time (83). Hence, *Brucella* causes significant economic losses and is a global health problem in endemic areas.

1.2.2 *Brucella* and different hosts cell types

In the animal or human hosts, *Brucella* enters mainly through the mucosa, wounds or the digestive tract. From the stomach, *Brucella* enters via Peyer's patches and M cell could be a route for bacteria to dissemination from the mucosal surface (84, 85). Upon entering the bloodstream and regional lymph nodes, *Brucella* is then able to spread systemically throughout the host via interaction with macrophages, dendritic cells (DCs), or neutrophils (86-91). Macrophages are the predominant cell type that is infected in both natural and human hosts. The ability to persist in this phagocytic cell enables *Brucella* to cause a chronic and long lasting infection. *Brucella* that can persist inside host cells is able to replicate intracellularly. This leads to large bacterial titers in infected organs such as the liver and spleen. It has also been shown that alveolar macrophages are a replicative niche and important for initial containment of bacteria in the lungs. Artificial reduction of alveolar macrophages results in an increase in infected pulmonary DCs and massive recruitment of TNF-alpha and inducible nitric oxide synthase (iNOS) producing DCs (92).

In addition to phagocytic cells, *Brucella* is able to infect various non-phagocytic cells. In pregnant ruminants, *Brucella* replicates within the rough endoplasmic reticulum (ER) of trophoblastic epithelial cells (88). Colonization of the reproductive organs causes abortion in these pregnant animals. *Brucella* also infects the mammary glands, endocardium, brain, joints, bones, and persistently colonizes the reticuloendothelial system (81).

In vitro studies of *Brucella* host-pathogen interaction are mostly performed with cultured murine, bovine, or human cells, including epithelial cell lines, macrophage cell lines, and trophoblastic cell lines. In macrophages, 90% of internalized *Brucella* is degraded soon after phagocytosis while a few bacteria manage to escape

Introduction

intracellular killing and proliferate. Even though activated macrophages are more efficient in killing *Brucella* (93-95), virulent wild type *Brucella* is still able to replicate at later time points in this system (96). To validate studies in *vitro*, there are also mouse experimental model of brucellosis available.

1.2.3 *Brucella* intracellular trafficking

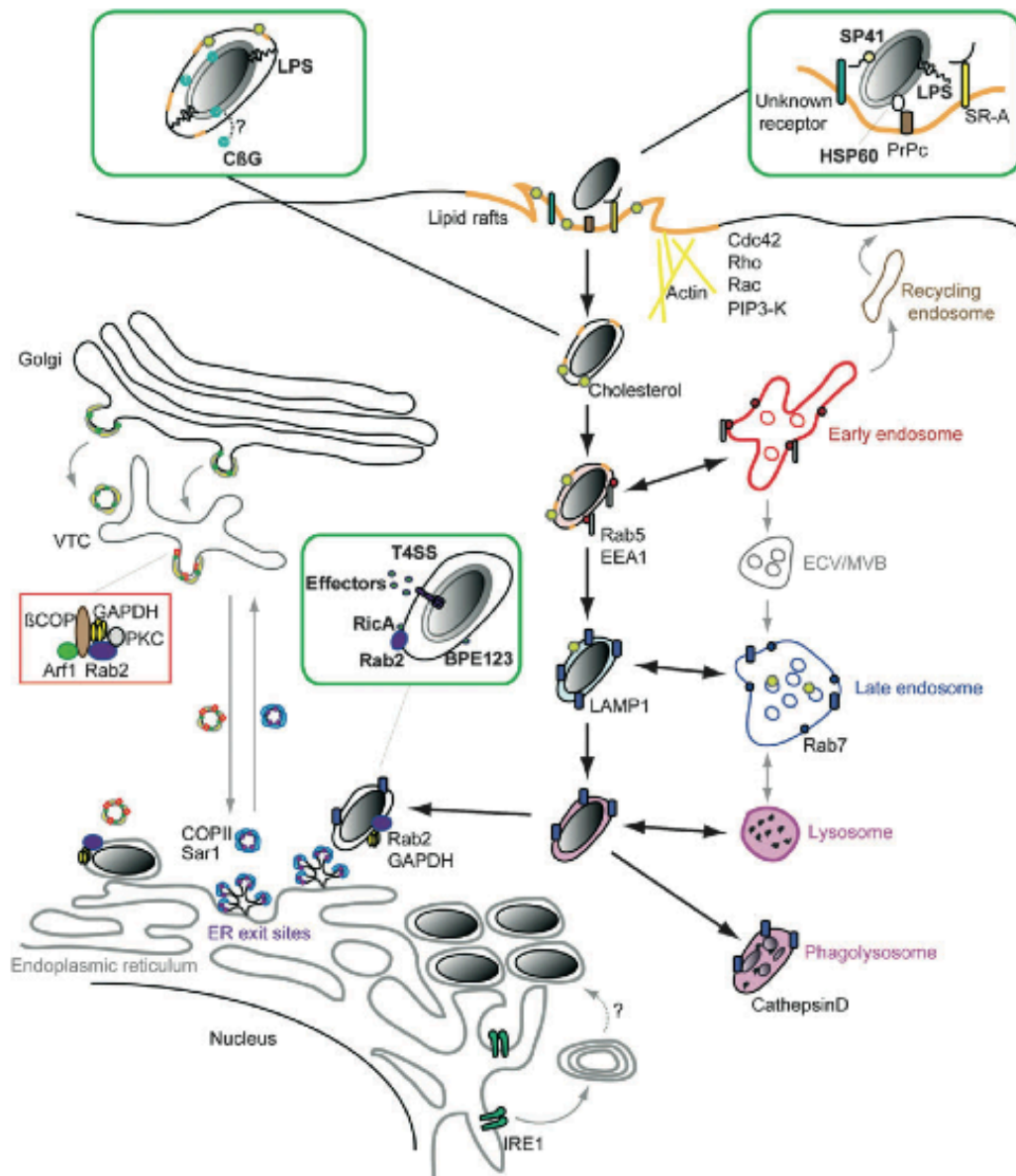


Figure 8 *Brucella* entry and intracellular trafficking in the host. MVB : multivesicular bodies, VTC : vesicular tubular clusters. Green box represents bacterial factors involved. Picture is taken from (86).

1.2.3.1 Adhesion and entry

In both macrophages and epithelial cells, adhesion of *Brucella* is mediated by interaction of surface protein 41 (SP41) with sialic acid residues present on eukaryotic receptors (97) (Figure 8). *Brucella* is also able to bind to fibronectin and vironectin (98). While the receptor for vironectin remains unknown, the large monomeric autotransporter BmaC was shown to promote binding to extracellular fibronectin in non-phagocytic cells (99). Additional bacterial factors involved in adhesion and internalization include the *efp* gene (100) and a pathogenicity island that harbors a bacterial immunoglobulin-like protein (101).

In non-phagocytic cells, *Brucella* enters the host via receptor-mediated phagocytosis (102, 103) through unknown receptors, a process that requires F-actin recruitment, activity of Rac and Rho, and direct activation of Cdc42 (104) (Figure 8). In trophoblast giant cells, entry depends on the surface protein Hsc70 and ezrin that interacts with Hsc70, tethering actin filaments to the PM (105).

In macrophages, studies have been done with both opsonized and non-opsonized *Brucella*. The uptake of non-opsonized *Brucella* requires lipid rafts (106-108). Bacteria are internalized due to membrane ruffling at the cell surface for a few minutes, a process that is dependent on phosphoinositide-3-kinase (PI3K) activity (109). Glycophosphatidylinositol (GPI) anchored proteins, GM1 ganglioside, and cholesterol are then selectively incorporated into the macropinosomes (106). Three macrophage receptors are implicated in *Brucella* uptake: class A scavenger receptor (SR-A) which interacts with LPS (110), Toll-like receptor 4 (TLR4), and potentially the cellular prion protein (PrPC) which interacts with Hsp60 of *Brucella abortus* (105, 109, 111) (Figure 8). The role of PrPC receptor on *Brucella* infection is still controversial as a separate study failed to show its involvement in entry in macrophages (112).

For opsonized *Brucella*, uptake is independent of lipid rafts (108) and depends on Fc receptors for IgG. Even though entry is strongly enhanced, opsonized *Brucella* is unable to replicate as efficiently as non-opsonized *Brucella* as they replicate in a vacuole that lacks ER markers (113). Therefore, different uptake mechanisms result in different trafficking routes or intracellular fates in macrophages.

Introduction

1.2.3.2 Trafficking along the endocytic pathway and VirB type IV secretion system (T4SS)

Upon internalization into phagocytic or non-phagocytic cells, *Brucella* containing vacuoles (BCVs) traffic along the endocytic pathway, interacting transiently with early endosomes containing Rab5, early endosomal antigen (EEA1), and transferrin receptor (TfR) (113-115) at 10 minutes post infection. Early BCVs are also positive for flotillin-1, a component of lipid raft (Figure 8). Cyclic beta-1,2-glucan present in the periplasm of *Brucella* modulates the organization of lipid rafts and is important for maturation of BCVs (116) (Figure 8). Afterwards, BCVs interact with the late endosomal markers Rab7, Rab7's effector Rab-interacting lysosomal protein (RILP), and Lamp1 (117)(Figure 8), and transiently with the autophagosomal marker monodansylcadaverin (115). Interaction with late endosomal markers is important since *Brucella* fails to replicate in an ER-like compartment in cells expressing dominant negative Rab7 (117).

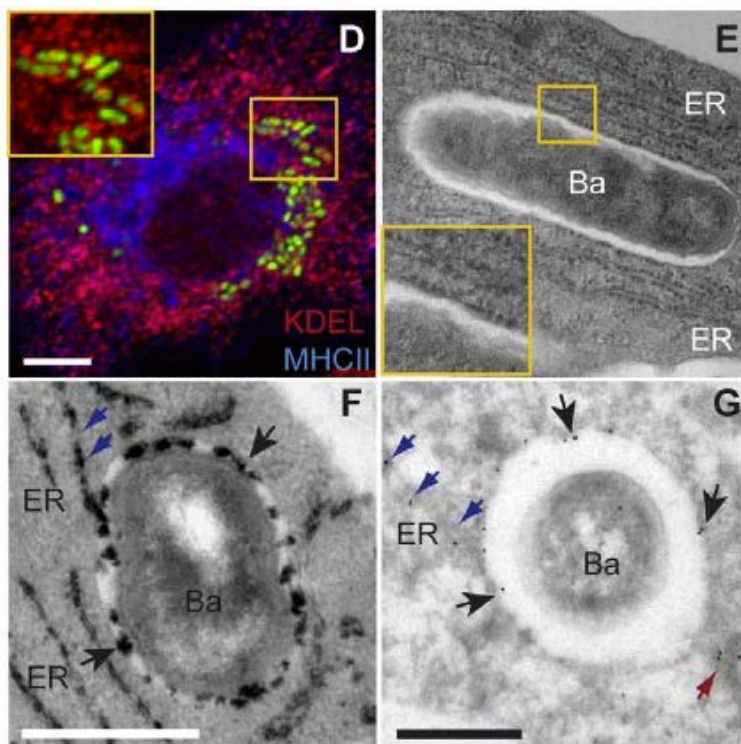


Figure 9 BCV and ER markers. D) confocal image of dendritic cells infected for 24 h with GFP-expressing *Brucella*, and labeled with MHC II (blue) and KDEL (red) antibodies. E) cytochemistry for glucose 6 phosphatase detection F) immunogold labeling with anti-calnexin antibody. Figure and figure legends are taken from (118)

Introduction

Early acidification of BCVs is crucial for the expression of the VirB type IV secretion system (T4SS) (119, 120). It is believed that secretion of still unknown effectors during early trafficking is important for *Brucella* to avoid fusion with lysosomes since *Brucella virB* mutants are degraded in lysosomes. *Brucella* that manages to divert from the endocytic pathway interacts with the secretory pathway and finally replicates in an endoplasmic reticulum (ER) derived replicative niche (Figure 9) (114) (Figure 8). During the first hours of infection, most BCVs are only Lamp1 positive while at 2-8 hours post infection (hpi), BCVs start to acquire ER markers such as calnexin in addition to Lamp1. BCVs then gradually lose Lamp1 and at 24 hpi most BCVs are negative for Lamp1 but retained the ER marker (114). Despite the acidification of the BCV and interaction with late endosomes, *Brucella* avoids cathepsin D, suggesting that they do not fuse with lysosomes (114, 121).

1.2.3.3 Survival in the replicative niche and egression

Several factors have been shown to be important for *Brucella* interaction and survival within its replicative niche. The small GTPase Sar1 regulates budding of transport vesicles from ER exit sites (ERES) to be transported to the Golgi. Inhibition of Sar1 activity results in disruption of ERES and blocks intracellular replication of *Brucella* by preventing its interaction with the ER (122). Therefore, *Brucella* initiates contact with the ER at ERES via interaction with Sar1 and the COPII complex (Figure 8). The VirB T4SS has also been shown to be important for sustained interaction of *Brucella* with the ER (114).

The small GTPase Rab2, a protein that is required for maturation of the ER-Golgi intermediate compartment (ERGIC) has been found through proteomics studies to be present on the BCV membrane. Inhibition of Rab2 prevents the fusion of BCVs with ER-derived vesicles and BCVs retain Lamp1 (123). The GADPH / COPI / PKC / Rab2 complex forms vesicular tubular clusters (VTCs) that control vesicular trafficking from Golgi to ER in the ERGIC (Figure 8). All members of this complex are required for intracellular replication of *Brucella* (123), suggesting that BCVs interact with VTCs and may intercept with the retrograde trafficking pathway. The *Brucella* effector RicA has been shown to interact with Rab2-GDP via a yeast two-hybrid screen (Figure 8). However, the role of this interaction has to be further studied (124). Taken together, components of both retrograde as well as anterograde vesicular trafficking were found to be involved in intracellular trafficking of *Brucella*.

Once *Brucella* reaches its replicative niche, it replicates extensively without disrupting host cell integrity. Apoptosis is inhibited in infected cells via down-regulation of gene expression in mitochondria that is normally responsible for apoptosis induction (125), and up-regulation of BCL2, a member of the anti-apoptotic pathway (126). Inositol-requiring enzyme (IRE1-alpha), a kinase that regulates host cell unfolded protein response is also crucial for *Brucella* replication in insect or mammalian cells (127) (Figure 8). However, its precise role in this process is unknown.

Spreading of bacteria from an infected cell to neighboring cells has not been investigated in detail, with the exception of a recent study that showed the involvement of autophagy initiation proteins. Autophagy initiation proteins ULK1, Beclin 1, ATG14L, and PI3K activity are required for conversion of the BCV to a compartment with autophagic features (aBCV). This conversion is independent of autophagy elongation proteins ATG5, ATG16L1, ATG4B, ATG7, and LC3B. aBCV then completes the intracellular life cycle of *Brucella* by facilitating its egress from the host, leading to cell-to-cell spreading (128).

1.3 Systems biology

Systems biology is an interdisciplinary biology-based approach that focuses on complex interactions within biological systems, using a more holistic approach compared to traditional reductionism strategies to study the properties of for example, a cell, tissue or organism as a system (129). In systems biology, the study of complex biological systems involves integration of various experimental and computational methods. With the advancement of technology and robotics, different ways of systematically perturbing the biological system (e.g. genetically or chemically) or acquiring biological information (e.g. transcriptomics, proteomics, metabolomics) could be performed in a high-throughput manner. Quantitative measurements in the large biological datasets are efficiently evaluated using high computing power. Such comprehensive data allows development of mechanistic, mathematical and computational models, further generating hypotheses for experimental validation. Due to the complex interplay between different components in a biological system, systems biology studies are much more informative than the reductionist approach,

yielding results that cannot be predicted when studying the individual components on their own (130).

1.3.1 RNA interference (RNAi)

RNAi is a natural RNA-dependent gene silencing process in which RNA molecules bind and destroy their complementary mRNAs, thereby inhibiting gene expression. This phenomenon was first described using the model organism *Caenorhabditis elegans* in 1998 where they found that introducing double stranded RNAs (dsRNAs) led to tenfold more effective silencing than the sense or anti-sense alone (131). Subsequently, RNAi was described as a potent anti-viral defense in plants (132) and later on this mechanism of gene silencing was also shown in organisms for example trypanosomes (133), flies (134), and vertebrates (135).

1.3.1.1 RNAi mechanism

RNAi pathway is normally initiated by an enzyme Dicer (136) that binds and cleaves long double stranded RNA molecules into short double stranded RNA fragments of around 20 nucleotides in length (137), with a 2-nucleotide overhang at the 3'-end. These shorter fragments are then separated into single stranded RNAs (ssRNAs), the so-called passenger and guide strands. The passenger strand is degraded while the guide strand gets incorporated into the RNA-induced silencing complex (RISC) (138). The incorporated guide strand base pairs with its complementary mRNA molecule and recruits RISC to the target mRNA. The cleavage of the target mRNA is then induced by Argonaute, the catalytic component of the RISC complex (138, 139). This process causes destruction of the mRNA, prevents protein production and is thus a gene silencing mechanism at a translational level (Figure 10).

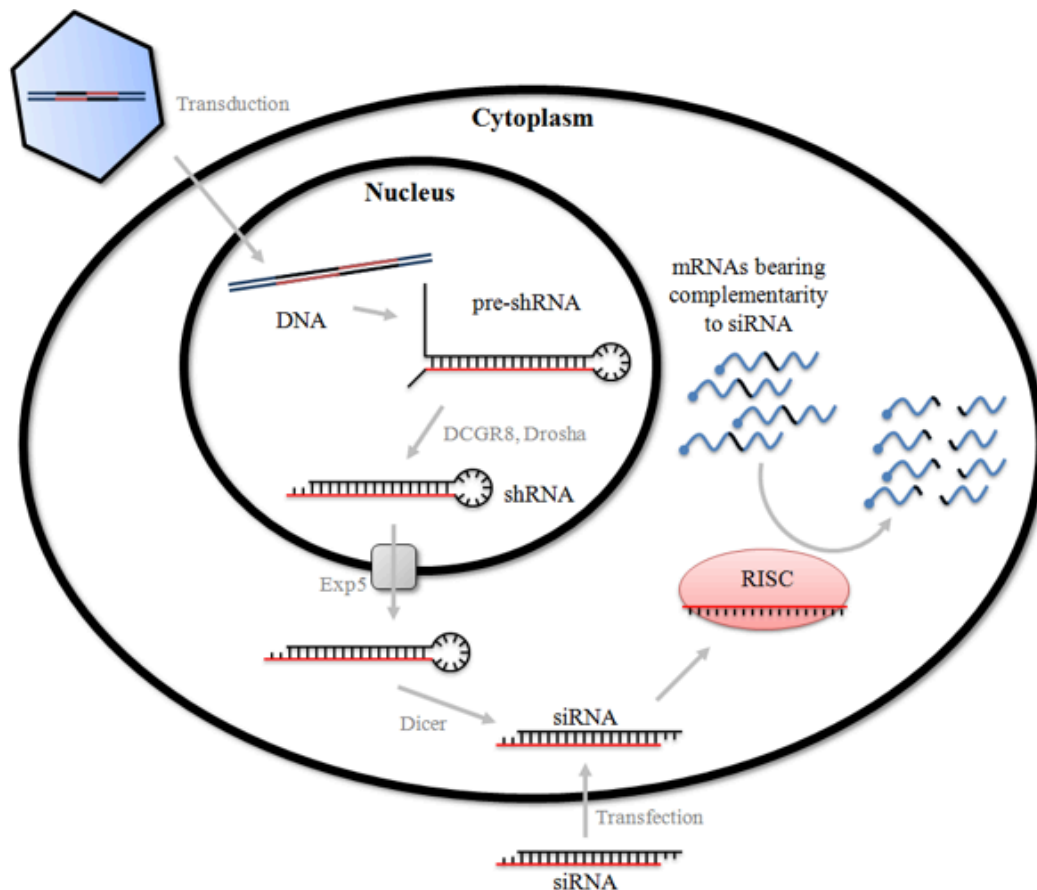


Figure 10 RNAi mechanism. shRNA is transduced into the host via viral vectors, after which it integrates with the host DNA. Expression in the nucleus allows shRNA to be processed by Drosha and exported by exportin-5 to the cytoplasm. There, it associates with Dicer and the loop sequence is removed, giving a product that is the same as siRNA introduced through transfection. Afterwards, it associates with the RISC complex and one of the RNA strands is removed. Next, it targets complementary RNA sequence, resulting in RNA degradation and gene silencing. Picture is taken from (140).

1.3.1.2 Small interfering RNA (siRNA), enzymatically generated siRNAs (esiRNA), short hairpin RNA (shRNA) and microRNA (miRNA)

In mammalian cells, there are several types of RNA molecules that are involved in RNAi: siRNA/esiRNA, shRNA, and microRNA. *In vivo*, mRNA transcripts can be regulated by basepairing with endogenous miRNA or siRNAs. Artificially, a similar effect could be produced with the addition of different exogenous small RNA species for example siRNA, esiRNA and shRNA. miRNA targets multiple mRNA of the host due to its ability to recognize target mRNA with only the 6-8 nucleotides (the seed region) at the 5' end of the miRNA(141).

siRNA is formed from shRNA or long dsRNA molecules by the Dicer enzyme (Figure 10). It is a short dsRNA molecule (21 bp in length) and can be introduced

directly into the RISC complex as described above. esiRNA instead are produced *in vitro* from long dsRNA which are digested into short dsRNAs by Dicer or RNase III. This produces a mixture of short RNA molecules targeting a gene of interest (142). Both types of RNA molecules are short-lived in target cells and hence relatively high doses in the nanomolar range are required. Generally, these molecules are delivered into the host cell by lipid-based transfection and are therefore not suitable for difficult to transfect cell lines e.g. primary cells or macrophages.

Alternative to transfection-based delivery, shRNA can be introduced into the host via viral or bacterial vectors (Figure 10). In the context of this work, only delivery via the lentiviral vector will be discussed. Lentiviruses containing a DNA construct that encodes for the shRNA are used to transduce the host, after which the DNA gets delivered into the nucleus and integrates in transcriptionally active sites within the host genome. Afterwards, shRNA is transcribed, resulting in pre-shRNA that is exported from the nucleus by Exportin 5 (143). Cytoplasmic pre-shRNA is then processed by Dicer to form siRNA molecules that are loaded into the RISC complex and follow the RNAi pathway as described above for siRNA-mediated gene silencing (Figure 10). Compared to siRNA, lentiviral delivery of shRNA and integration into the host allows lower dosage and stable, long lasting gene silencing. Transduction also allows introduction of the shRNA into a variety of cell types, including those that are not amenable to siRNA transfections (144, 145).

1.3.1.3 RNAi as a tool – pros and cons

In the field of biological research, RNAi is a very widely used tool to study the function of certain genes by reducing their expression. The RNAi technique is nowadays developed to a degree that can be used in *in vivo* model organisms (146), with major applications still being in cell culture setups (147). In cell cultures, exogenous or synthetic RNA is introduced as short RNA molecules (137) since longer dsRNA molecules are identified as foreign and induce mammalian interferon response (147-150). With the availability of genome-wide siRNA libraries targeting the mouse and human genomes, comprehensive studies of gene function can be performed. RNAi screens have been established in many systems e.g. *Caenorhabditis elegans*, *Drosophila* cultured cells and mammalian cell lines (151-153). Introducing RNAi into an inducible system also allows study of the gene of interest in a time-resolved manner.

Introduction

RNAi is a popular tool for loss-of function experiments due to its ease of use, efficiency, and relatively affordable cost compared to the knockout technology that only works for selected organisms and requires years for characterization of the mutant. It is also more advantageous than overexpression of dominant negative mutants that normally does not reflect the true endogenous function of the protein of interest and is difficult to be studied in a high-throughput manner. Due to its promising potential in treatment of viral infections, cancers and neurodegenerative diseases (154) and its ability to be delivered systemically in liposomal formulation into non-human primates (155), RNAi is a potential new class of drugs.

Despite all its benefits, RNAi technology has its limitations. Major concerns with RNAi include incomplete loss of function of the gene that could lead to a phenotype different from a knockout condition. Also, the non specific base-pairing of RNA oligos with mRNA molecules of a similar but not identical sequence may lead to undesired off-target effects (156, 157). Therefore, knockdowns are often done with many siRNAs to corroborate the observed phenotype and minimize the risk of following an off-target phenotype. In this respect, it has been shown that pooling of siRNAs is beneficial in rendering greater phenotypic penetrance compared to individual oligos (158, 159). In some cases, siRNA could also activate the interferon system of the cell (150), potentially affecting gene expression in a much broader scale. shRNA expression has also been reported to interfere with the endogenous microRNA pathways and causes non-specific fatality in mice (160). Therefore, with the current limitations of RNAi, such experiments always require validation with RNAi-independent methods and different methods are available to identify off target effects of this technology (157, 161). Titration experiments to obtain the minimal amount of siRNA needed for maximum efficiency is also useful to obtain the least off-target effects from this technology.

1.3.2 Genome-wide RNAi screening to study systems-level host pathogen interaction

1.3.2.1 RNAi screening for host factors important for viral and bacterial pathogens

RNAi-based genome-wide screens have been performed extensively to study bacterial or viral pathogen interactions with the host (162-177). These screens have uncovered

Introduction

host-signaling pathways that are hijacked by the pathogens during different stages of their intracellular cycle e.g. invasion (165), modulation of phagosomal maturation (166), phagosomal escape or release to the cytosol (162, 170, 172), and replication (162, 166, 170, 171, 173, 175). Many screens are done and compared between different pathogens (163, 164, 167-169) to distinguish general pathways from pathogen-specific pathways. In light of the increasing problem of antibiotic resistance developed by bacterial pathogens due to inappropriate usage of antibiotics, the understanding of host signaling pathways hijacked by pathogens becomes essential in revealing host factors that could be targeted as an alternative to the antibiotic regime. However, there is a common problem in the RNAi field studying host-pathogen interactions that there is very little overlap between the genes that are identified from each screen (178-181). One example is the four RNAi screens (174-177) that all sought to identify cellular genes important for HIV-1 infection or replication. Comparison of these four screens showed genes that are significantly similar in their effect on the HIV-1 infection process, even though not more than 3-6% of genes are shared between two screens and only three genes are identified in all three screens (179). Such low overlap is largely due to variable experimental procedures followed in different studies, including the difference in the choice of cell type and siRNA library, as well as the viral strain used. Furthermore, another factor includes the length of siRNA treatment and the exposure time of cells to the virus. The latter determines whether only the early or entire infectious cycle is covered by the assay (178, 179). Different analysis and hit selection procedures are also a source of variation between screens (178, 179). Analyses showed a much greater overlap when whole cellular processes and protein complexes that are populated by the identified host factors are compared between screens (179). Therefore, it is useful to perform functional analysis of RNAi data (182, 183) than having a focused study of individual genes. Efforts to standardize RNAi experiments, improve reagents and analysis methodologies might also allow better comparability between screens of different origin (178, 179).

1.3.2.2 High content microscopy and multiparametric analysis in RNAi screening

With the availability of automated microscopes, imaging of RNAi screens in a high-throughput manner became possible, which enables the acquisition of highly resolved spatial and temporal aspects of the investigated process (184, 185). Upon data acquisition, normalization steps to correct for experimental variations, e.g. plate-to-

Introduction

plate variations are executed before further analysis of the images. Using image processing tools e.g. CellProfiler, different objects in the pictures can be identified and segmented. Such objects are then used to extract different features e.g. bacterial colony size, cell shape, and actin texture (186). Various quantitative measurements could then be based on these features, allowing analysis on a cell population as well as single cell level. Computational methods that are available allow multidimensional data interpretation as well as supervised machine learning, automatically classifying different cellular and subcellular phenotypes (187-191).

1.4 References

1. Ward TH, Polishchuk RS, Caplan S, Hirschberg K, & Lippincott-Schwartz J (2001) Maintenance of Golgi structure and function depends on the integrity of ER export. *The Journal of cell biology* 155(4):557-570.
2. Brandizzi F, Snapp EL, Roberts AG, Lippincott-Schwartz J, & Hawes C (2002) Membrane protein transport between the endoplasmic reticulum and the Golgi in tobacco leaves is energy dependent but cytoskeleton independent: evidence from selective photobleaching. *The Plant cell* 14(6):1293-1309.
3. Bonifacino JS & Glick BS (2004) The mechanisms of vesicle budding and fusion. *Cell* 116(2):153-166.
4. English AR, Zurek N, & Voeltz GK (2009) Peripheral ER structure and function. *Current opinion in cell biology* 21(4):596-602.
5. Voeltz GK, Rolls MM, & Rapoport TA (2002) Structural organization of the endoplasmic reticulum. *EMBO reports* 3(10):944-950.
6. Baumann O & Walz B (2001) Endoplasmic reticulum of animal cells and its organization into structural and functional domains. *International review of cytology* 205:149-214.
7. Friedman JR & Voeltz GK (2011) The ER in 3D: a multifunctional dynamic membrane network. *Trends in cell biology* 21(12):709-717.
8. Altan-Bonnet N, Sougrat R, & Lippincott-Schwartz J (2004) Molecular basis for Golgi maintenance and biogenesis. *Current opinion in cell biology* 16(4):364-372.
9. Farquhar MG & Palade GE (1998) The Golgi apparatus: 100 years of progress and controversy. *Trends in cell biology* 8(1):2-10.
10. De Matteis MA & Morrow JS (2000) Spectrin tethers and mesh in the biosynthetic pathway. *Journal of cell science* 113 (Pt 13):2331-2343.
11. Klumperman J (2000) The growing Golgi: in search of its independence. *Nature cell biology* 2(12):E217-219.
12. Sanders SL, Whitfield KM, Vogel JP, Rose MD, & Schekman RW (1992) Sec61p and BiP directly facilitate polypeptide translocation into the ER. *Cell* 69(2):353-365.
13. Bannykh SI, Rowe T, & Balch WE (1996) The organization of endoplasmic reticulum export complexes. *The Journal of cell biology* 135(1):19-35.

Introduction

14. Barlowe C, *et al.* (1994) COPII: a membrane coat formed by Sec proteins that drive vesicle budding from the endoplasmic reticulum. *Cell* 77(6):895-907.
15. Miller EA & Barlowe C (2010) Regulation of coat assembly--sorting things out at the ER. *Current opinion in cell biology* 22(4):447-453.
16. Brandizzi F & Barlowe C (2013) Organization of the ER-Golgi interface for membrane traffic control. *Nature reviews. Molecular cell biology* 14(6):382-392.
17. Schweizer A, *et al.* (1990) Identification of an intermediate compartment involved in protein transport from endoplasmic reticulum to Golgi apparatus. *European journal of cell biology* 53(2):185-196.
18. Spang A (2009) On vesicle formation and tethering in the ER-Golgi shuttle. *Current opinion in cell biology* 21(4):531-536.
19. Oka T & Nakano A (1994) Inhibition of GTP hydrolysis by Sar1p causes accumulation of vesicles that are a functional intermediate of the ER-to-Golgi transport in yeast. *The Journal of cell biology* 124(4):425-434.
20. Moyer BD, Allan BB, & Balch WE (2001) Rab1 interaction with a GM130 effector complex regulates COPII vesicle cis--Golgi tethering. *Traffic* 2(4):268-276.
21. Rowe T, Dascher C, Bannykh S, Plutner H, & Balch WE (1998) Role of vesicle-associated syntaxin 5 in the assembly of pre-Golgi intermediates. *Science* 279(5351):696-700.
22. Shorter J, Beard MB, Seemann J, Dirac-Svejstrup AB, & Warren G (2002) Sequential tethering of Golgins and catalysis of SNAREpin assembly by the vesicle-tethering protein p115. *The Journal of cell biology* 157(1):45-62.
23. Hay JC, Chao DS, Kuo CS, & Scheller RH (1997) Protein interactions regulating vesicle transport between the endoplasmic reticulum and Golgi apparatus in mammalian cells. *Cell* 89(1):149-158.
24. Tisdale EJ, Bourne JR, Khosravi-Far R, Der CJ, & Balch WE (1992) GTP-binding mutants of rab1 and rab2 are potent inhibitors of vesicular transport from the endoplasmic reticulum to the Golgi complex. *The Journal of cell biology* 119(4):749-761.
25. Mayor S & Pagano RE (2007) Pathways of clathrin-independent endocytosis. *Nature reviews. Molecular cell biology* 8(8):603-612.

Introduction

26. Maxfield FR, Schlessinger J, Shechter Y, Pastan I, & Willingham MC (1978) Collection of insulin, EGF and alpha2-macroglobulin in the same patches on the surface of cultured fibroblasts and common internalization. *Cell* 14(4):805-810.
27. Anderson RG, Brown MS, & Goldstein JL (1977) Role of the coated endocytic vesicle in the uptake of receptor-bound low density lipoprotein in human fibroblasts. *Cell* 10(3):351-364.
28. Di Guglielmo GM, Le Roy C, Goodfellow AF, & Wrana JL (2003) Distinct endocytic pathways regulate TGF-beta receptor signalling and turnover. *Nature cell biology* 5(5):410-421.
29. Bonifacino JS & Lippincott-Schwartz J (2003) Coat proteins: shaping membrane transport. *Nature reviews. Molecular cell biology* 4(5):409-414.
30. Bonifacino JS & Traub LM (2003) Signals for sorting of transmembrane proteins to endosomes and lysosomes. *Annual review of biochemistry* 72:395-447.
31. Sever S (2002) Dynamin and endocytosis. *Current opinion in cell biology* 14(4):463-467.
32. Rothman JE & Schmid SL (1986) Enzymatic recycling of clathrin from coated vesicles. *Cell* 46(1):5-9.
33. Lemaitre G, *et al.* (2005) CD98, a novel marker of transient amplifying human keratinocytes. *Proteomics* 5(14):3637-3645.
34. Sprenger RR, *et al.* (2004) Comparative proteomics of human endothelial cell caveolae and rafts using two-dimensional gel electrophoresis and mass spectrometry. *Electrophoresis* 25(1):156-172.
35. Sigismund S, *et al.* (2005) Clathrin-independent endocytosis of ubiquitinated cargos. *Proceedings of the National Academy of Sciences of the United States of America* 102(8):2760-2765.
36. Galvez T, Gilleron J, Zerial M, & O'Sullivan GA (2012) SnapShot: Mammalian Rab proteins in endocytic trafficking. *Cell* 151(1):234-234 e232.
37. Seaman MN (2012) The retromer complex - endosomal protein recycling and beyond. *Journal of cell science* 125(Pt 20):4693-4702.
38. Poteryaev D, Datta S, Ackema K, Zerial M, & Spang A (2010) Identification of the switch in early-to-late endosome transition. *Cell* 141(3):497-508.

Introduction

39. Rink J, Ghigo E, Kalaidzidis Y, & Zerial M (2005) Rab conversion as a mechanism of progression from early to late endosomes. *Cell* 122(5):735-749.
40. Hurley JH & Stenmark H (2011) Molecular mechanisms of ubiquitin-dependent membrane traffic. *Annual review of biophysics* 40:119-142.
41. Henne WM, Buchkovich NJ, & Emr SD (2011) The ESCRT pathway. *Developmental cell* 21(1):77-91.
42. Huotari J & Helenius A (2011) Endosome maturation. *The EMBO journal* 30(17):3481-3500.
43. Peyroche A, Paris S, & Jackson CL (1996) Nucleotide exchange on ARF mediated by yeast Gea1 protein. *Nature* 384(6608):479-481.
44. Spang A, Herrmann JM, Hamamoto S, & Schekman R (2001) The ADP ribosylation factor-nucleotide exchange factors Gea1p and Gea2p have overlapping, but not redundant functions in retrograde transport from the Golgi to the endoplasmic reticulum. *Molecular biology of the cell* 12(4):1035-1045.
45. Dumaresq-Doiron K, Savard MF, Akam S, Costantino S, & Lefrancois S (2010) The phosphatidylinositol 4-kinase PI4KIIIalpha is required for the recruitment of GBF1 to Golgi membranes. *Journal of cell science* 123(Pt 13):2273-2280.
46. Yu X, Breitman M, & Goldberg J (2012) A structure-based mechanism for Arf1-dependent recruitment of coatamer to membranes. *Cell* 148(3):530-542.
47. Eugster A, Frigerio G, Dale M, & Duden R (2000) COP I domains required for coatamer integrity, and novel interactions with ARF and ARF-GAP. *The EMBO journal* 19(15):3905-3917.
48. Szul T & Sztul E (2011) COPII and COPI traffic at the ER-Golgi interface. *Physiology (Bethesda)* 26(5):348-364.
49. Lanoix J, *et al.* (2001) Sorting of Golgi resident proteins into different subpopulations of COPI vesicles: a role for ArfGAP1. *The Journal of cell biology* 155(7):1199-1212.
50. Rein U, Andag U, Duden R, Schmitt HD, & Spang A (2002) ARF-GAP-mediated interaction between the ER-Golgi v-SNAREs and the COPI coat. *The Journal of cell biology* 157(3):395-404.

Introduction

51. Schindler C, *et al.* (2009) The GAP domain and the SNARE, coatomer and cargo interaction region of the ArfGAP2/3 Glo3 are sufficient for Glo3 function. *Traffic* 10(9):1362-1375.
52. Kamena F, Diefenbacher M, Kilchert C, Schwarz H, & Spang A (2008) Ypt1p is essential for retrograde Golgi-ER transport and for Golgi maintenance in *S. cerevisiae*. *Journal of cell science* 121(Pt 8):1293-1302.
53. Zink S, Wenzel D, Wurm CA, & Schmitt HD (2009) A link between ER tethering and COP-I vesicle uncoating. *Developmental cell* 17(3):403-416.
54. Tisdale EJ, Kelly C, & Artalejo CR (2004) Glyceraldehyde-3-phosphate dehydrogenase interacts with Rab2 and plays an essential role in endoplasmic reticulum to Golgi transport exclusive of its glycolytic activity. *The Journal of biological chemistry* 279(52):54046-54052.
55. Girod A, *et al.* (1999) Evidence for a COP-I-independent transport route from the Golgi complex to the endoplasmic reticulum. *Nature cell biology* 1(7):423-430.
56. Chen A, AbuJarour RJ, & Draper RK (2003) Evidence that the transport of ricin to the cytoplasm is independent of both Rab6A and COPI. *Journal of cell science* 116(Pt 17):3503-3510.
57. Lindberg AA, *et al.* (1987) Identification of the carbohydrate receptor for Shiga toxin produced by *Shigella dysenteriae* type 1. *The Journal of biological chemistry* 262(4):1779-1785.
58. Sandvig K, Olsnes S, Brown JE, Petersen OW, & van Deurs B (1989) Endocytosis from coated pits of Shiga toxin: a glycolipid-binding protein from *Shigella dysenteriae* 1. *The Journal of cell biology* 108(4):1331-1343.
59. Lauvrak SU, Torgersen ML, & Sandvig K (2004) Efficient endosome-to-Golgi transport of Shiga toxin is dependent on dynamin and clathrin. *Journal of cell science* 117(Pt 11):2321-2331.
60. Nichols BJ, *et al.* (2001) Rapid cycling of lipid raft markers between the cell surface and Golgi complex. *The Journal of cell biology* 153(3):529-541.
61. Kirkham M & Parton RG (2005) Clathrin-independent endocytosis: new insights into caveolae and non-caveolar lipid raft carriers. *Biochimica et biophysica acta* 1745(3):273-286.
62. Saint-Pol A, *et al.* (2004) Clathrin adaptor epsinR is required for retrograde sorting on early endosomal membranes. *Developmental cell* 6(4):525-538.

Introduction

63. Bujny MV, Popoff V, Johannes L, & Cullen PJ (2007) The retromer component sorting nexin-1 is required for efficient retrograde transport of Shiga toxin from early endosome to the trans Golgi network. *Journal of cell science* 120(Pt 12):2010-2021.
64. Torgersen ML, Walchli S, Grimmer S, Skanland SS, & Sandvig K (2007) Protein kinase Cdelta is activated by Shiga toxin and regulates its transport. *The Journal of biological chemistry* 282(22):16317-16328.
65. Utskarpen A, *et al.* (2010) Shiga toxin increases formation of clathrin-coated pits through Syk kinase. *PloS one* 5(7):e10944.
66. Lord MJ, *et al.* (2003) Ricin. Mechanisms of cytotoxicity. *Toxicological reviews* 22(1):53-64.
67. van Deurs B, *et al.* (1988) Estimation of the amount of internalized ricin that reaches the trans-Golgi network. *The Journal of cell biology* 106(2):253-267.
68. Moreau D, *et al.* (2011) Genome-wide RNAi screens identify genes required for Ricin and PE intoxications. *Developmental cell* 21(2):231-244.
69. Cossart P & Sansonetti PJ (2004) Bacterial invasion: the paradigms of enteroinvasive pathogens. *Science* 304(5668):242-248.
70. Lafont F, Tran Van Nhieu G, Hanada K, Sansonetti P, & van der Goot FG (2002) Initial steps of Shigella infection depend on the cholesterol/sphingolipid raft-mediated CD44-IpaB interaction. *The EMBO journal* 21(17):4449-4457.
71. Shin JS, Gao Z, & Abraham SN (2000) Involvement of cellular caveolae in bacterial entry into mast cells. *Science* 289(5480):785-788.
72. van der Goot FG, Tran van Nhieu G, Allaoui A, Sansonetti P, & Lafont F (2004) Rafts can trigger contact-mediated secretion of bacterial effectors via a lipid-based mechanism. *The Journal of biological chemistry* 279(46):47792-47798.
73. Shen Y, Naujokas M, Park M, & Ireton K (2000) InIB-dependent internalization of Listeria is mediated by the Met receptor tyrosine kinase. *Cell* 103(3):501-510.
74. Mengaud J, Ohayon H, Gounon P, Mege RM, & Cossart P (1996) E-cadherin is the receptor for internalin, a surface protein required for entry of L. monocytogenes into epithelial cells. *Cell* 84(6):923-932.

Introduction

75. Veiga E & Cossart P (2005) Listeria hijacks the clathrin-dependent endocytic machinery to invade mammalian cells. *Nature cell biology* 7(9):894-900.
76. Veiga E, *et al.* (2007) Invasive and adherent bacterial pathogens co-Opt host clathrin for infection. *Cell host & microbe* 2(5):340-351.
77. Truttmann MC, *et al.* (2011) Bartonella henselae engages inside-out and outside-in signaling by integrin beta1 and talin1 during invasome-mediated bacterial uptake. *Journal of cell science* 124(Pt 21):3591-3602.
78. Dehio C, Meyer M, Berger J, Schwarz H, & Lanz C (1997) Interaction of Bartonella henselae with endothelial cells results in bacterial aggregation on the cell surface and the subsequent engulfment and internalisation of the bacterial aggregate by a unique structure, the invasome. *Journal of cell science* 110 (Pt 18):2141-2154.
79. Audic S, Lescot M, Claverie JM, & Scholz HC (2009) Brucella microti: the genome sequence of an emerging pathogen. *BMC genomics* 10:352.
80. Moreno E, Cloeckaert A, & Moriyon I (2002) Brucella evolution and taxonomy. *Veterinary microbiology* 90(1-4):209-227.
81. Atluri VL, Xavier MN, de Jong MF, den Hartigh AB, & Tsolis RM (2011) Interactions of the human pathogenic Brucella species with their hosts. *Annual review of microbiology* 65:523-541.
82. Pappas G, Papadimitriou P, Akritidis N, Christou L, & Tsianos EV (2006) The new global map of human brucellosis. *The Lancet infectious diseases* 6(2):91-99.
83. Ariza J, *et al.* (2007) Perspectives for the treatment of brucellosis in the 21st century: the Ioannina recommendations. *PLoS medicine* 4(12):e317.
84. Paixao TA, *et al.* (2009) Establishment of systemic Brucella melitensis infection through the digestive tract requires urease, the type IV secretion system, and lipopolysaccharide O antigen. *Infection and immunity* 77(10):4197-4208.
85. Ackermann MR, Cheville NF, & Deyoe BL (1988) Bovine ileal dome lymphoepithelial cells: endocytosis and transport of Brucella abortus strain 19. *Veterinary pathology* 25(1):28-35.
86. von Bargen K, Gorvel JP, & Salcedo SP (2012) Internal affairs: investigating the Brucella intracellular lifestyle. *FEMS microbiology reviews* 36(3):533-562.

Introduction

87. Billard E, Cazevielle C, Dornand J, & Gross A (2005) High susceptibility of human dendritic cells to invasion by the intracellular pathogens *Brucella suis*, *B. abortus*, and *B. melitensis*. *Infection and immunity* 73(12):8418-8424.
88. Anderson TD, Cheville NF, & Meador VP (1986) Pathogenesis of placentitis in the goat inoculated with *Brucella abortus*. II. Ultrastructural studies. *Veterinary pathology* 23(3):227-239.
89. Copin R, *et al.* (2012) In situ microscopy analysis reveals local innate immune response developed around *Brucella* infected cells in resistant and susceptible mice. *PLoS pathogens* 8(3):e1002575.
90. Meador VP, Deyoe BL, & Cheville NF (1989) Pathogenesis of *Brucella abortus* infection of the mammary gland and supramammary lymph node of the goat. *Veterinary pathology* 26(5):357-368.
91. Enright FM, Araya LN, Elzer PH, Rowe GE, & Winter AJ (1990) Comparative histopathology in BALB/c mice infected with virulent and attenuated strains of *Brucella abortus*. *Veterinary immunology and immunopathology* 26(2):171-182.
92. Archambaud C, *et al.* (2010) Contrasting roles of macrophages and dendritic cells in controlling initial pulmonary *Brucella* infection. *European journal of immunology* 40(12):3458-3471.
93. Jiang X & Baldwin CL (1993) Iron augments macrophage-mediated killing of *Brucella abortus* alone and in conjunction with interferon-gamma. *Cellular immunology* 148(2):397-407.
94. Eze MO, *et al.* (2000) Effects of opsonization and gamma interferon on growth of *Brucella melitensis* 16M in mouse peritoneal macrophages in vitro. *Infection and immunity* 68(1):257-263.
95. Sathiyaseelan J, Jiang X, & Baldwin CL (2000) Growth of *Brucella abortus* in macrophages from resistant and susceptible mouse strains. *Clinical and experimental immunology* 121(2):289-294.
96. Baldwin CL & Goenka R (2006) Host immune responses to the intracellular bacteria *Brucella*: does the bacteria instruct the host to facilitate chronic infection? *Critical reviews in immunology* 26(5):407-442.
97. Castaneda-Roldan EI, *et al.* (2006) Characterization of SP41, a surface protein of *Brucella* associated with adherence and invasion of host epithelial cells. *Cellular microbiology* 8(12):1877-1887.

Introduction

98. Castaneda-Roldan EI, *et al.* (2004) Adherence of Brucella to human epithelial cells and macrophages is mediated by sialic acid residues. *Cellular microbiology* 6(5):435-445.
99. Posadas DM, Ruiz-Ranwez V, Bonomi HR, Martin FA, & Zorreguieta A (2012) BmaC, a novel autotransporter of Brucella suis, is involved in bacterial adhesion to host cells. *Cellular microbiology* 14(6):965-982.
100. Iannino F, Ugalde JE, & Inon de Iannino N (2012) Brucella abortus efp gene is required for an efficient internalization in HeLa cells. *Microbial pathogenesis* 52(1):31-40.
101. Czibener C & Ugalde JE (2012) Identification of a unique gene cluster of Brucella spp. that mediates adhesion to host cells. *Microbes and infection / Institut Pasteur* 14(1):79-85.
102. Detilleux PG, Deyoe BL, & Cheville NF (1990) Entry and intracellular localization of Brucella spp. in Vero cells: fluorescence and electron microscopy. *Veterinary pathology* 27(5):317-328.
103. Detilleux PG, Deyoe BL, & Cheville NF (1991) Effect of endocytic and metabolic inhibitors on the internalization and intracellular growth of Brucella abortus in Vero cells. *American journal of veterinary research* 52(10):1658-1664.
104. Guzman-Verri C, *et al.* (2001) GTPases of the Rho subfamily are required for Brucella abortus internalization in nonprofessional phagocytes: direct activation of Cdc42. *The Journal of biological chemistry* 276(48):44435-44443.
105. Watanabe K, Tachibana M, Kim S, & Watarai M (2009) EEVD motif of heat shock cognate protein 70 contributes to bacterial uptake by trophoblast giant cells. *Journal of biomedical science* 16:113.
106. Watarai M, Makino S, Fujii Y, Okamoto K, & Shirahata T (2002) Modulation of Brucella-induced macropinocytosis by lipid rafts mediates intracellular replication. *Cellular microbiology* 4(6):341-355.
107. Watarai M, *et al.* (2002) Macrophage plasma membrane cholesterol contributes to Brucella abortus infection of mice. *Infection and immunity* 70(9):4818-4825.

Introduction

108. Naroeni A & Porte F (2002) Role of cholesterol and the ganglioside GM(1) in entry and short-term survival of *Brucella suis* in murine macrophages. *Infection and immunity* 70(3):1640-1644.
109. Pei J, Turse JE, & Ficht TA (2008) Evidence of *Brucella abortus* OPS dictating uptake and restricting NF-kappaB activation in murine macrophages. *Microbes and infection / Institut Pasteur* 10(6):582-590.
110. Kim S, *et al.* (2004) Lipid raft microdomains mediate class A scavenger receptor-dependent infection of *Brucella abortus*. *Microbial pathogenesis* 37(1):11-19.
111. Watarai M (2004) Interaction between *Brucella abortus* and cellular prion protein in lipid raft microdomains. *Microbes and infection / Institut Pasteur* 6(1):93-100.
112. Fontes P, *et al.* (2005) Absence of evidence for the participation of the macrophage cellular prion protein in infection with *Brucella suis*. *Infection and immunity* 73(10):6229-6236.
113. Bellaire BH, Roop RM, 2nd, & Cardelli JA (2005) Opsonized virulent *Brucella abortus* replicates within nonacidic, endoplasmic reticulum-negative, LAMP-1-positive phagosomes in human monocytes. *Infection and immunity* 73(6):3702-3713.
114. Celli J, *et al.* (2003) *Brucella* evades macrophage killing via VirB-dependent sustained interactions with the endoplasmic reticulum. *The Journal of experimental medicine* 198(4):545-556.
115. Pizarro-Cerda J, *et al.* (1998) *Brucella abortus* transits through the autophagic pathway and replicates in the endoplasmic reticulum of nonprofessional phagocytes. *Infection and immunity* 66(12):5711-5724.
116. Arellano-Reynoso B, *et al.* (2005) Cyclic beta-1,2-glucan is a *Brucella* virulence factor required for intracellular survival. *Nature immunology* 6(6):618-625.
117. Starr T, Ng TW, Wehrly TD, Knodler LA, & Celli J (2008) *Brucella* intracellular replication requires trafficking through the late endosomal/lysosomal compartment. *Traffic* 9(5):678-694.
118. Salcedo SP, *et al.* (2008) *Brucella* control of dendritic cell maturation is dependent on the TIR-containing protein Btp1. *PLoS pathogens* 4(2):e21.

Introduction

119. Porte F, Liautard JP, & Kohler S (1999) Early acidification of phagosomes containing *Brucella suis* is essential for intracellular survival in murine macrophages. *Infection and immunity* 67(8):4041-4047.
120. Boschirolini ML, *et al.* (2002) The *Brucella suis* virB operon is induced intracellularly in macrophages. *Proceedings of the National Academy of Sciences of the United States of America* 99(3):1544-1549.
121. Pizarro-Cerda J, Moreno E, Sanguedolce V, Mege JL, & Gorvel JP (1998) Virulent *Brucella abortus* prevents lysosome fusion and is distributed within autophagosome-like compartments. *Infection and immunity* 66(5):2387-2392.
122. Celli J, Salcedo SP, & Gorvel JP (2005) *Brucella* coopts the small GTPase Sar1 for intracellular replication. *Proceedings of the National Academy of Sciences of the United States of America* 102(5):1673-1678.
123. Fugier E, *et al.* (2009) The glyceraldehyde-3-phosphate dehydrogenase and the small GTPase Rab 2 are crucial for *Brucella* replication. *PLoS pathogens* 5(6):e1000487.
124. de Barsey M, *et al.* (2011) Identification of a *Brucella* spp. secreted effector specifically interacting with human small GTPase Rab2. *Cellular microbiology* 13(7):1044-1058.
125. He Y, *et al.* (2006) *Brucella melitensis* triggers time-dependent modulation of apoptosis and down-regulation of mitochondrion-associated gene expression in mouse macrophages. *Infection and immunity* 74(9):5035-5046.
126. Gross A, Terraza A, Ouahrani-Bettache S, Liautard JP, & Dornand J (2000) In vitro *Brucella suis* infection prevents the programmed cell death of human monocytic cells. *Infection and immunity* 68(1):342-351.
127. Qin QM, *et al.* (2008) RNAi screen of endoplasmic reticulum-associated host factors reveals a role for IRE1alpha in supporting *Brucella* replication. *PLoS pathogens* 4(7):e1000110.
128. Starr T, *et al.* (2012) Selective subversion of autophagy complexes facilitates completion of the *Brucella* intracellular cycle. *Cell host & microbe* 11(1):33-45.
129. wikipedia (http://en.wikipedia.org/wiki/Systems_biology).
130. Ahn AC, Tewari M, Poon CS, & Phillips RS (2006) The limits of reductionism in medicine: could systems biology offer an alternative? *PLoS medicine* 3(6):e208.

Introduction

131. Fire A, *et al.* (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391(6669):806-811.
132. Stram Y & Kuzntzova L (2006) Inhibition of viruses by RNA interference. *Virus genes* 32(3):299-306.
133. Ngo H, Tschudi C, Gull K, & Ullu E (1998) Double-stranded RNA induces mRNA degradation in *Trypanosoma brucei*. *Proceedings of the National Academy of Sciences of the United States of America* 95(25):14687-14692.
134. Kennerdell JR & Carthew RW (1998) Use of dsRNA-mediated genetic interference to demonstrate that *frizzled* and *frizzled 2* act in the wingless pathway. *Cell* 95(7):1017-1026.
135. Wianny F & Zernicka-Goetz M (2000) Specific interference with gene function by double-stranded RNA in early mouse development. *Nature cell biology* 2(2):70-75.
136. Bernstein E, Caudy AA, Hammond SM, & Hannon GJ (2001) Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* 409(6818):363-366.
137. Zamore PD, Tuschl T, Sharp PA, & Bartel DP (2000) RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell* 101(1):25-33.
138. Hammond SM, Bernstein E, Beach D, & Hannon GJ (2000) An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells. *Nature* 404(6775):293-296.
139. Tang G (2005) siRNA and miRNA: an insight into RISCs. *Trends in biochemical sciences* 30(2):106-114.
140. O'Keefe EP (2013) siRNAs and shRNAs: Tools for Protein Knockdown by Gene silencing. *MATER METHODS 2013: 3:197* / <http://www.labome.com/method/siRNAs-and-shRNA-Tools-for-Protein-Knockdown-by-Gene-Silencing.html>.
141. Lewis BP, Burge CB, & Bartel DP (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* 120(1):15-20.
142. Yang D, *et al.* (2002) Short RNA duplexes produced by hydrolysis with *Escherichia coli* RNase III mediate effective RNA interference in mammalian

Introduction

- cells. *Proceedings of the National Academy of Sciences of the United States of America* 99(15):9942-9947.
143. Yi R, Doehle BP, Qin Y, Macara IG, & Cullen BR (2005) Overexpression of exportin 5 enhances RNA interference mediated by short hairpin RNAs and microRNAs. *RNA* 11(2):220-226.
 144. Rao DD, Vorhies JS, Senzer N, & Nemunaitis J (2009) siRNA vs. shRNA: similarities and differences. *Advanced drug delivery reviews* 61(9):746-759.
 145. Lord CJ, Martin SA, & Ashworth A (2009) RNA interference screening demystified. *Journal of clinical pathology* 62(3):195-200.
 146. Snove O, Jr. & Rossi JJ (2006) Expressing short hairpin RNAs in vivo. *Nature methods* 3(9):689-695.
 147. Elbashir SM, *et al.* (2001) Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* 411(6836):494-498.
 148. Reynolds A, *et al.* (2006) Induction of the interferon response by siRNA is cell type- and duplex length-dependent. *RNA* 12(6):988-993.
 149. Bridge AJ, Pebernard S, Ducraux A, Nicoulaz AL, & Iggo R (2003) Induction of an interferon response by RNAi vectors in mammalian cells. *Nature genetics* 34(3):263-264.
 150. Sledz CA, Holko M, de Veer MJ, Silverman RH, & Williams BR (2003) Activation of the interferon system by short-interfering RNAs. *Nature cell biology* 5(9):834-839.
 151. Echeverri CJ & Perrimon N (2006) High-throughput RNAi screening in cultured cells: a user's guide. *Nature reviews. Genetics* 7(5):373-384.
 152. Moffat J & Sabatini DM (2006) Building mammalian signalling pathways with RNAi screens. *Nature reviews. Molecular cell biology* 7(3):177-187.
 153. Poulin G, Nandakumar R, & Ahringer J (2004) Genome-wide RNAi screens in *Caenorhabditis elegans*: impact on cancer research. *Oncogene* 23(51):8340-8345.
 154. Pushparaj PN & Melendez AJ (2006) Short interfering RNA (siRNA) as a novel therapeutic. *Clinical and experimental pharmacology & physiology* 33(5-6):504-510.
 155. Zimmermann TS, *et al.* (2006) RNAi-mediated gene silencing in non-human primates. *Nature* 441(7089):111-114.

Introduction

156. Jackson AL, *et al.* (2006) Widespread siRNA "off-target" transcript silencing mediated by seed region sequence complementarity. *RNA* 12(7):1179-1187.
157. Franceschini A, *et al.* (2014) Specific inhibition of diverse pathogens in human cells by synthetic microRNA-like oligonucleotides inferred from RNAi screens. *Proceedings of the National Academy of Sciences of the United States of America* 111(12):4548-4553.
158. Pauli Rämö AD, Cécile Arrieumerlou, Niko Beerenwinkel, Houchaima Ben-Tekaya, Bettina Cardel, Alain Casanova, Raquel Conde-Alvarez, Pascale Cossart, Gábor Csúcs, Simone Eicher, Mario Emmenlauer, Urs Greber, Wolf-Dietrich Hardt, Ari Helenius, Christoph Kasper, Andreas Kaufmann, Saskia Kreibich, Andreas Kühbacher, Peter Kunszt, Shyan Huey Low, Jason Mercer, Daria Mudrak, Simone Muntwiler, Lucas Pelkmans, Javier Pizarro-Cerdá, Michael Podvinec, Eva Pujadas, Bernd Rinn, Vincent Rouilly, Fabian Schmich, Juliane Siebourg-Polster, Berend Snijder, Michael Stebler, Gabriel Studer, Ewa Szczurek, Matthias Truttmann, Christian von Mering, Andreas Vonderheit, Artur Yakimovich, Peter Bühlmann & Christoph Dehio (submitted manuscript) Parallel Mixed Model for Simultaneous Analysis of Multiple Large-Scale RNAi Screens.
159. Parsons BD, Schindler A, Evans DH, & Foley E (2009) A direct phenotypic comparison of siRNA pools and multiple individual duplexes in a functional assay. *PloS one* 4(12):e8471.
160. Grimm D, *et al.* (2006) Fatality in mice due to oversaturation of cellular microRNA/short hairpin RNA pathways. *Nature* 441(7092):537-541.
161. Marine S, Bahl A, Ferrer M, & Buehler E (2012) Common seed analysis to identify off-target effects in siRNA screens. *Journal of biomolecular screening* 17(3):370-378.
162. Akimana C, Al-Khodori S, & Abu Kwaik Y (2010) Host factors required for modulation of phagosome biogenesis and proliferation of *Francisella tularensis* within the cytosol. *PloS one* 5(6):e11025.
163. Derre I, Pypaert M, Dautry-Varsat A, & Agaisse H (2007) RNAi screen in *Drosophila* cells reveals the involvement of the Tom complex in *Chlamydia* infection. *PLoS pathogens* 3(10):1446-1458.

Introduction

164. Kumar D, *et al.* (2010) Genome-wide analysis of the host intracellular network that regulates survival of *Mycobacterium tuberculosis*. *Cell* 140(5):731-743.
165. Misselwitz B, *et al.* (2011) RNAi screen of *Salmonella* invasion shows role of COPI in membrane targeting of cholesterol and Cdc42. *Molecular systems biology* 7:474.
166. McDonough JA, *et al.* (2013) Host pathways important for *Coxiella burnetii* infection revealed by genome-wide RNA interference screening. *mBio* 4(1):e00606-00612.
167. Philips JA, Rubin EJ, & Perrimon N (2005) *Drosophila* RNAi screen reveals CD36 family member required for mycobacterial infection. *Science* 309(5738):1251-1253.
168. Agaisse H, *et al.* (2005) Genome-wide RNAi screen for host factors required for intracellular bacterial infection. *Science* 309(5738):1248-1251.
169. Krishnan MN, *et al.* (2008) RNA interference screen for human genes associated with West Nile virus infection. *Nature* 455(7210):242-245.
170. Mercer J, *et al.* (2012) RNAi screening reveals proteasome- and Cullin3-dependent stages in vaccinia virus infection. *Cell reports* 2(4):1036-1047.
171. Konig R, *et al.* (2010) Human host factors required for influenza virus replication. *Nature* 463(7282):813-817.
172. Su WC, *et al.* (2013) Pooled RNAi screen identifies ubiquitin ligase Itch as crucial for influenza A virus release from the endosome during virus entry. *Proceedings of the National Academy of Sciences of the United States of America* 110(43):17516-17521.
173. Sivan G, *et al.* (2013) Human genome-wide RNAi screen reveals a role for nuclear pore proteins in poxvirus morphogenesis. *Proceedings of the National Academy of Sciences of the United States of America* 110(9):3519-3524.
174. Yeung ML, Houzet L, Yedavalli VS, & Jeang KT (2009) A genome-wide short hairpin RNA screening of jurkat T-cells for human proteins contributing to productive HIV-1 replication. *The Journal of biological chemistry* 284(29):19463-19473.
175. Zhou H, *et al.* (2008) Genome-scale RNAi screen for host factors required for HIV replication. *Cell host & microbe* 4(5):495-504.

Introduction

176. König R, *et al.* (2008) Global analysis of host-pathogen interactions that regulate early-stage HIV-1 replication. *Cell* 135(1):49-60.
177. Brass AL, *et al.* (2008) Identification of host proteins required for HIV infection through a functional genomic screen. *Science* 319(5865):921-926.
178. Pache L, König R, & Chanda SK (2011) Identifying HIV-1 host cell factors by genome-scale RNAi screening. *Methods* 53(1):3-12.
179. Bushman FD, *et al.* (2009) Host cell factors in HIV replication: meta-analysis of genome-wide studies. *PLoS pathogens* 5(5):e1000437.
180. Eicher SC & Dehio C (2013) Systems-level analysis of host-pathogen interaction using RNA interference. *New biotechnology* 30(3):308-313.
181. Hao L, *et al.* (2013) Limited agreement of independent RNAi screens for virus-required host genes owes more to false-negative than false-positive factors. *PLoS computational biology* 9(9):e1003235.
182. Kampmann M, Bassik MC, & Weissman JS (2013) Integrated platform for genome-wide screening and construction of high-density genetic interaction maps in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America* 110(25):E2317-2326.
183. Bassik MC, *et al.* (2013) A systematic mammalian genetic interaction map reveals pathways underlying ricin susceptibility. *Cell* 152(4):909-922.
184. Pepperkok R & Ellenberg J (2006) High-throughput fluorescence microscopy for systems biology. *Nature reviews. Molecular cell biology* 7(9):690-696.
185. Conrad C & Gerlich DW (2010) Automated microscopy for high-content RNAi screening. *The Journal of cell biology* 188(4):453-461.
186. Carpenter AE, *et al.* (2006) CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome biology* 7(10):R100.
187. Ramo P, Sacher R, Snijder B, Begemann B, & Pelkmans L (2009) CellClassifier: supervised learning of cellular phenotypes. *Bioinformatics* 25(22):3028-3030.
188. Walter T, *et al.* (2010) Automatic identification and clustering of chromosome phenotypes in a genome wide RNAi screen by time-lapse imaging. *Journal of structural biology* 170(1):1-9.
189. Glory E & Murphy RF (2007) Automated subcellular location determination and high-throughput microscopy. *Developmental cell* 12(1):7-16.

Introduction

190. Jones TR, *et al.* (2008) CellProfiler Analyst: data exploration and analysis software for complex image-based screens. *BMC bioinformatics* 9:482.
191. Gerlich D, Beaudouin J, Gebhard M, Ellenberg J, & Eils R (2001) Four-dimensional imaging and quantitative reconstruction to analyse complex spatiotemporal processes in live cells. *Nature cell biology* 3(9):852-855.

AIM OF THESIS

2. AIM OF THESIS

Started in July 2010, the aim of my thesis was to systematically identify host factors involved in *Brucella* entry, trafficking and replication and to perform follow-up studies on interesting pathways that are involved in *Brucella* infection. To this end, I performed a high-throughput, microscopy based genome-wide siRNA screen in HeLa cells. A few signaling pathways were selected for validation with siRNA-independent methods and to understand the molecular function during infection. Being a part of the InfectX consortium that consists of five bacterial pathogens and three viral pathogens performing the same genome-wide screens, the aim of the project was also to compare our results between pathogens and to identify shared or unique hits. Moreover, I developed tools consisting of stable cell lines expressing various fluorescently labeled cellular compartmental markers. The aim was to use these cell lines to understand the intracellular trafficking pathway of *Brucella* and to dissect the step that is affected upon knockdown of specific genes of interest. It was also aimed to use these cell lines to study *Brucella* trafficking with fluorescence microscopy, live cell imaging and electron microscopy studies.

RESULTS

3. RESULTS

3.1 RESEARCH ARTICLE I

Specific inhibition of diverse pathogens in human cells by synthetic microRNA-like oligonucleotides inferred from RNAi screens.

Andreas Franceschini*, Roger Meier*, Alain Casanova*, Saskia Kreibich*, Neha Daga, Daniel Andritschke, Sabrina Dilling, Pauli Ramo, Mario Emmelauer, Andreas Kaufmann, Raquel Conde-Alvarez, Shyan Huey Low, Lucas Pelkmans, Ari Helenius, Wolf-Dietrich Hardt, Christoph Dehio, and Christian von Mering

PNAS, March 25 2014, vol. 111 no 12

*These authors contributed equally to this manuscript

Statement of my own contribution

I contributed to this manuscript by performing the genome-wide siRNA screen as well as kinome screens for pathogen *Brucella*. The data from these screens were used for analysis in this paper that led to the identification of seed regions responsible for off-target effects in the screen.

3.1.1 Summary

High-throughput RNA interference (RNAi) screens are often used to study at a systems level host factors that are involved in a certain process. However, high levels of false positive in siRNA screens are often associated with sequence-dependent off-target effects (1). This is mainly due to the ‘seed’ region of 21-nucleotide siRNA (nucleotide 2-8), which is sufficient to recognize its target, even though there is low complementarity in the rest of the sequence (2). The binding of the oligo to multiple transcripts results in perturbation of many genes simultaneously, similar to the effect of a microRNA. In this study, genome-wide siRNA screens from three pathogens (*Brucella*, *Salmonella* and *Uukuniemi*) as well as kinome screens from *Brucella* and *Salmonella* were used to study seed-mediated off-target effects. It was shown that majority of the siRNA phenotype is off-target dictated, with relatively less correlation to the on-target effect. Quantitative analysis allowed prediction of seeds that block or increase infection. The effect of these seeds could be confirmed with independent experiments using custom ordered sequences, with a mutation at the seed completely abolishing the phenotype. Furthermore, seed sequences with no matching on-target sequence were still able to reproduce the predicted phenotype. Therefore, this suggests that RNAi screens are off-target driven. All together, this study provides a possible way to predict seed reagents that have an effect on the phenotype of interest, allowing us to identify and address off-target effects that are present in RNAi screens.

References

1. Jackson AL, Linsley PS (2010) Recognizing and avoiding siRNA off-target effects for target identification and therapeutic application. *Nat Rev Drug Discov* 9: 57-67.
2. Birmingham A, Anderson EM, Reynolds A, Ilsley-Tyree D, Leake D, et al. (2006) 3' UTR seed matches, but not overall identity, are associated with RNAi off-targets. *Nat Methods* 3: 199-204.

3.1.2 Manuscript



Specific inhibition of diverse pathogens in human cells by synthetic microRNA-like oligonucleotides inferred from RNAi screens

Andrea Franceschini ^{a,1}, Roger Meier ^{b,1,2}, Alain Casanova ^{c,1}, Saskia Kreibich ^{d,1}, Neha Daga ^a, Daniel Andritschke ^d, Sabrina Dilling ^d, Pauli Rämö ^c, Mario Emmenlauer ^c, Andreas Kaufmann ^c, Raquel Conde-Álvarez ^{c,3}, Shyan Huey Low ^c, Lucas Pelkmans ^e, Ari Helenius ^{b,4}, Wolf-Dietrich Hardt ^d, Christoph Dehio ^c, and Christian von Mering ^{a,4}

^aInstitute of Molecular Life Sciences and Swiss Institute of Bioinformatics, University of Zurich, CH-8057 Zurich, Switzerland; ^bInstitute of Biochemistry, Eidgenössische Technische Hochschule Zurich, CH-8093 Zurich, Switzerland; ^cBiozentrum, University of Basel, CH-4056 Basel, Switzerland; ^dInstitute of Microbiology, Eidgenössische Technische Hochschule Zurich, CH-8093 Zurich, Switzerland; and ^eInstitute of Molecular Life Sciences, University of Zurich, CH-8057 Zurich, Switzerland

Contributed by Ari Helenius, February 7, 2014 (sent for review November 26, 2013)

Systematic genetic perturbation screening in human cells remains technically challenging. Typically, large libraries of chemically synthesized siRNA oligonucleotides are used, each designed to degrade a specific cellular mRNA via the RNA interference (RNAi) mechanism. Here, we report on data from three genome-wide siRNA screens, conducted to uncover host factors required for infection of human cells by two bacterial and one viral pathogen. We find that the majority of phenotypic effects of siRNAs are unrelated to the intended “on-target” mechanism, defined by full complementarity of the 21-nt siRNA sequence to a target mRNA. Instead, phenotypes are largely dictated by “off-target” effects resulting from partial complementarity of siRNAs to multiple mRNAs via the “seed” region (i.e., nucleotides 2–8), reminiscent of the way specificity is determined for endogenous microRNAs. Quantitative analysis enabled the prediction of seeds that strongly and specifically block infection, independent of the intended on-target effect. This prediction was confirmed experimentally by designing oligos that do not have any on-target sequence match at all, yet can strongly reproduce the predicted phenotypes. Our results suggest that published RNAi screens have primarily, and unintentionally, screened the sequence space of microRNA seeds instead of the intended on-target space of protein-coding genes. This helps to explain why previously published RNAi screens have exhibited relatively little overlap. Our analysis suggests a possible way of identifying “seed reagents” for controlling phenotypes of interest and establishes a general strategy for extracting valuable untapped information from past and future RNAi screens.

high-throughput RNAi screening | antimicrobials

High-throughput, genome-wide perturbation screening is a powerful tool for uncovering novel genes and pathways responsible for phenotypes or functions of interest (1). In many model organisms, systematic collections of deletion or knockout strains have been established, enabling well-controlled and efficient screening experiments. In contrast, when working with human cells, the technical possibilities for gene perturbations are much more limited. Although promising technologies for targeted genome editing in human cells have been introduced recently (2–5), these are at present too cumbersome for routine, genome-wide screening.

Nevertheless, systematic genetic screening directly in human cells is highly desirable: for example, when working with infectious human pathogens. Pathogens are often fast-evolving and locked in a molecular “arms race” with their hosts; thus, their interactions with cellular genes are often host-specific and must be screened in the native host species. For systematically perturbing human genes, the most widely used method is RNA interference (RNAi), which involves the use of commercial

libraries of synthetic small interfering RNA (siRNA) molecules (6). A number of pioneering RNAi screens for host factors required by human pathogens have already been conducted (7–15), and many other human phenotypes have been screened as well (16). Although these screens have revealed numerous seminal insights into the molecular processes under study, they have also highlighted recurring (and poorly understood) problems with respect to the reliability and specificity of RNAi reagents used in high throughput. Among the initial hits from the primary screens, a high prevalence of false positives is often observed, forcing researchers to allocate significant resources to validation and follow-up studies of each candidate gene. Furthermore, the overlap between independently published screens can be frustratingly low—as exemplified by the three initial HIV screens that showed hardly any significant overlap in a metaanalysis (17).

Apart from false positives generated by statistical noise or by nonspecific toxicity of the RNAi reagents, the most problematic sources of false positives are thought to be the sequence-

Significance

Pathogens can enter into human cells using a variety of specific mechanisms, often hitchhiking on naturally existing transport pathways. To uncover parts of the host machinery that are required for entry, scientists conduct infection screens in cultured cells. In these screens, human genes are systematically inactivated by short RNA oligos, designed to bind and inactivate mRNA molecules. Here, we show that many of these oligos additionally bind unintended mRNA targets as well, and that this effect overall dominates and complicates such screens. Focusing on the strong “off-target” signal, we design novel oligos that no longer bind any one gene specifically but nevertheless strongly and reproducibly block pathogen entry pointing to pathogen/host interactions at a higher-order, pathway level.

Author contributions: A.F., A.H., W.-D.H., C.D., and C.v.M. designed research; R.M., A.C., S.K., D.A., S.D., R.C.-Á., and S.H.L. performed research; A.K. contributed new reagents/analytic tools; A.F., N.D., P.R., M.E., L.P., and C.v.M. analyzed data; and A.H., W.-D.H., C.D., and C.v.M. wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

¹A.F., R.M., A.C., and S.K. contributed equally to this work.

²Present address: Light Microscopy and Screening Center, Eidgenössische Technische Hochschule Zurich, CH-8093 Zurich, Switzerland.

³Present address: Department of Microbiology and Parasitology, University of Navarra, 31008 Pamplona, Spain.

⁴To whom correspondence may be addressed. E-mail: ari.helenius@bc.biol.ethz.ch or mering@imls.usz.ch.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1402353111/-/DCSupplemental.

dependent, so-called “off-target” effects (18). These are problematic because they can be highly reproducible and will thus not be canceled out automatically over multiple replicates of the same perturbation. Sequence-specific off-target effects may originate from partial complementarity of the siRNA oligos to unintended, noncognate cellular mRNA targets; such mRNAs are bound by the siRNAs and subsequently perturbed in terms of their stability and/or protein translation rate. At least some of these off-target effects are presumably mediated by the cellular microRNA-processing machinery, which mistakes transfected siRNA oligos for endogenous microRNAs, loading them onto the RNA-induced silencing complex and scanning for mRNAs with suitable binding sites. Consistent with this hypothesis, it has been observed that sequence-dependent off-target effects of siRNAs are primarily controlled and initiated by the “seed” region of their sequence (nucleotide positions 2–8), similar to what is the case for microRNAs (6, 19, 20). Matches to any given seed sequence typically occur in several hundred different human transcripts, suggesting that each off-target event can potentially perturb tens or hundreds of genes simultaneously. A number of studies have analyzed RNAi datasets for experimental evidence of seed-mediated off-target effects (19–25), using both global gene-expression readouts as well as defined, single-gene readouts that have been the subject of screens. These studies reported that “seed effects” can indeed be visible in the raw data and that they can explain some of the unexpected or apparent false-positive findings.

Here, we comprehensively quantify the prevalence of seed effects in screens that address two important classes of phenotypes: cellular infection by pathogens and cellular survival and proliferation. Such complex phenotype/gene associations are the

central aim of genome-wide RNAi screening. We address this issue in the context of three pathogen-infection screens, which have been conducted in different laboratories, working with three distinct pathogens. We analyze both the infection phenotypes as well as the cellular proliferation phenotypes of these screens, assuming them to be good representatives of complex molecular processes involving many putative “hit” genes.

We find that seed-mediated phenotypes are dominating in all three screens, to an extent that they threaten to camouflage on-target phenotypes for all but the most clear-cut, strongest on-target gene effects. In a systematic approach, we took advantage of the strength of the observed seed effects to quantitatively characterize the potential space of microRNA-like regulation of pathogen entry/replication. We show that novel siRNA oligo sequences can be designed that replicate the seed effect and that strongly and specifically control the pathogen’s ability to infect cells. In addition to consequences for screen design and analysis, we are discussing possible implications for therapeutic applications and for the role of microRNAs in the evolution of resistance toward pathogen infection.

Results

We analyzed raw data from genome-wide RNAi infection screens for two invasive bacterial pathogens (*Brucella abortus*, *Salmonella typhimurium*) and one virus (*Uukuniemi virus*, an enveloped RNA virus of the *Bunyaviridae* family) (26). All three screens were conducted using HeLa cells. Here, we are focusing on the sequences of the individual siRNA oligos and how they relate to the observed phenotypes (Fig. 1). For each of the three different pathogens, the same commercially available, genome-wide, deconvoluted siRNA library was used. For the two

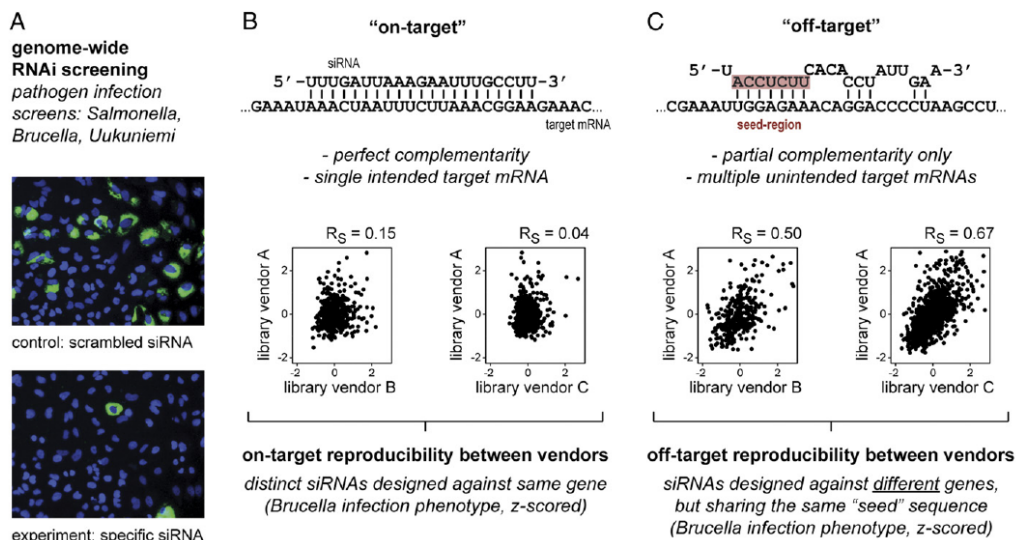


Fig. 1. Off-target effects in RNAi pathogen infection screens. (A) Experimental setup. HeLa cells were screened for host factors required for pathogen entry. Microscopy images from two separate wells of a typical perturbation experiment are shown (DAPI-stained HeLa cell nuclei in blue; successful pathogen infection in green from *B. abortus* expressing GFP). All three pathogens were screened using a genome-wide library (Qiagen), and *Brucella* and *Salmonella* additionally with two kinome-wide libraries (Ambion, Dharmacon). (B) Intended on-target mechanism of siRNA action. Below, in the correlation plots, each data point represents one gene, whereby the infection phenotypes (infection index) were averaged over all of the oligos designed for a given gene by a given library vendor. (C) Unintended off-target mechanism of siRNA actions. Here, each data point represents one seed sequence, with phenotypes averaged over all oligos that happen to contain that seed sequence in a given library. For all plots in B and C, pairs of oligos that happened to share the same seed sequence and the same on-target gene (in any of the three libraries) were excluded. Note that intervendor comparisons are based on the subset of genes screened with all three libraries (i.e., the kinome subset). Both correlations in B and C are highly significant ($P \leq 10^{-20}$).

bacterial pathogens, we complemented the genome-wide screens with additional library screening focusing on the set of kinases and kinase-related genes in the human genome, using siRNA libraries from two other commercial vendors. All three libraries typically consisted of our distinct siRNA oligos per human gene, transfected and measured separately. The infection readouts and other cellular phenotypes were assessed by automated microscopy, followed by standardized image-processing procedures (see Materials and Methods for a brief summary). The analysis procedure included state-of-the-art normalization and image-correction steps, and all phenotypes were z score-normalized before further analysis. Apart from the infection phenotype, we also systematically assessed the number of cells observed in each well; this latter phenotype reflects the net sum of perturbation effects on cell proliferation and survival and constitutes a second, independent readout that should yield largely equivalent results in all three screens.

First, we observed that the overall consistency of “on-target” effects appeared to be surprisingly low: when comparing the results of distinct oligos designed to target the exact same gene, the phenotypes were virtually uncorrelated (Fig. 1 and Fig. S1). This was the case both when comparing different oligos from the same library and when comparing across the libraries from three different commercial siRNA vendors. Even when averaging over all oligos of a given gene in a given library, rank correlations across libraries were often below 0.1 and never exceeded 0.2, both for the infection phenotype as well as for the cell-number phenotype (Fig. 1 and Fig. S1).

We next observed the oligos from different vendors again, but this time not based on their designated on-targets (full 21-nt complementarity), but instead based on their presumed off-targets (by grouping them according to the sequences of their heptameric seed regions at nucleotide positions 2-8) (Fig. 1). If phenotypes were attributable to the on-target (not the off-target)

mechanism, this second test should not yield any correlation—note that all pairs of oligos that happened to share both the seed region and the designated on-target were excluded.

Strikingly, however, we here observed much higher correlations for all pairwise comparisons of library vendors (Fig. 1 and Fig. S1). Correlations were highly significant, both for the case of the infection phenotypes as well as for the cell-number phenotypes. In 12 out of 12 comparisons, such “off-target correlations” were significantly greater than the on-target correlations, usually by a factor of five or more (Fig. S1). In our view, this suggests that (i) the lack of correlation in the first test was not attributable to improper screen execution, image processing, or normalizations, (ii) most of the siRNA oligos do result in nonrandom phenotypes, and (iii) for all three commercial library vendors, the average siRNA oligo is predominantly and reproducibly acting via the off-target mechanism.

We next aggregated the entire genome-wide screening data based on shared seed sequences (Fig. 2 and Dataset S1). Of the theoretically possible “space” of 16,384 heptamer seeds, 64% are represented in the genome-wide library, many by dozens of different siRNA oligos. Among the subset of seeds represented 10 times or more, we observe that roughly one third result in statistically significant infection phenotypes (by extension, this fraction would likely apply also to nonobservable seeds that happened to be insufficiently covered by the library). The statistical strength of this signal is high, with seed effects reaching P values of 10^{-12} , even after correcting for multiple testing (Dataset S1). We observe that the seed signal is strictly position-dependent with respect to the siRNA nucleotide sequence as hardly any statistical signal remained when the seed was assumed at the “wrong” position (Fig. 2). Moreover, our analysis also confirms that there seem to be no off-target signals stemming from the opposite (“passenger”) strand of the double-stranded siRNA molecules (Fig. S2).

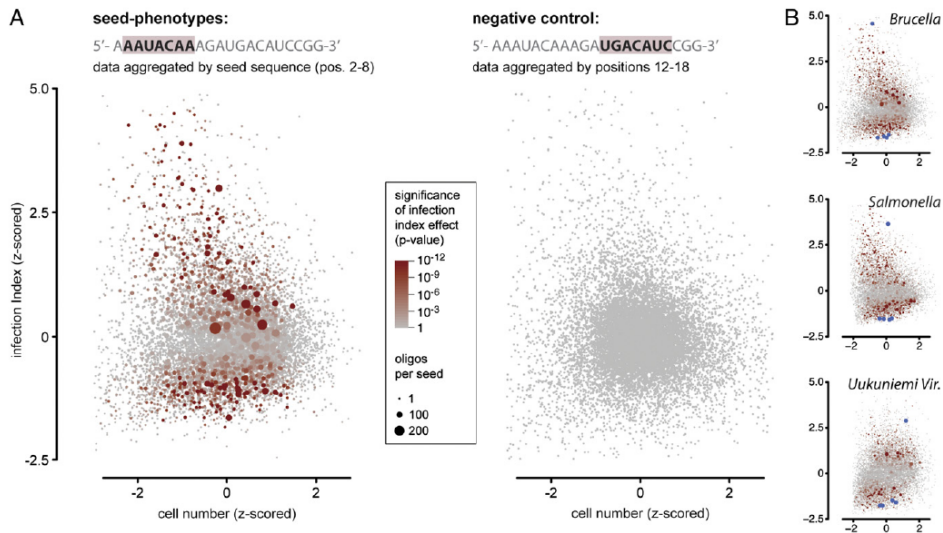


Fig. 2. Genome-wide screening data aggregated by shared seed sequences. (A) Visualization of the entire genome-wide data of the infection screen for *B. abortus*, aggregated by the seed sequences found in the various siRNA oligos. Each data point represents one heptameric seed sequence, showing the averaged phenotypes over all siRNA oligos that happen to share that seed. The color code indicates the statistical significance of the observed infection phenotypes. For the negative control, data were plotted in exactly the same way, but the position of the seed in each siRNA oligo was incorrectly assumed to be at positions 12–18. (B) Visualizations for all three pathogens screened here; blue dots mark the seeds that have been selected for experimental follow-up.

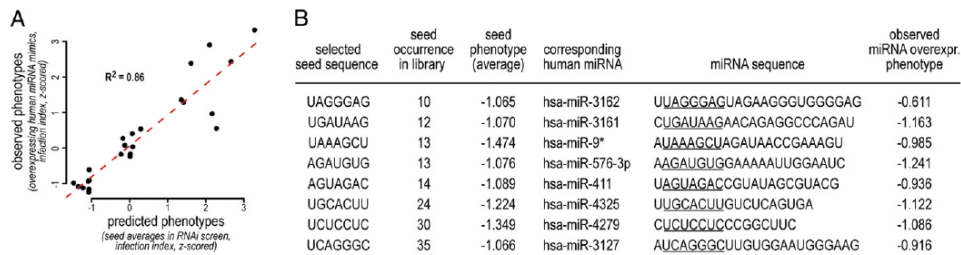


Fig. 4. Human miRNA overexpression phenotypes. (A) Based on the *B. abortus* genome-wide siRNA screen, specific seeds were selected that happened to occur also in known, endogenous human miRNAs. Eight of these seeds were predicted to reduce infection, eight were predicted to enhance infection, and eight were predicted to be neutral. To be selected, seeds had to be represented at least 10 times in the siRNA library and had to correspond to a single known human miRNA only. The figure shows the infection outcomes of transfecting these known miRNAs (as molecular mimics), compared with their predicted phenotypes as inferred from the seed analysis. (B) Tabulated details of the eight human miRNAs that were predicted, and confirmed, to block infection.

conferring pronounced toxic side effects on the host cell and without targeting any one gene specifically by design.

Discussion

For complex genome-wide RNAi screens, our analysis suggests that seed-mediated off-target effects can dominate the phenotypic readouts and may present a serious problem for properly inferring the intended on-target effects. Considering that genome-wide screens have the additional statistical problem of massive multiple testing, it becomes evident that ad hoc gene lists of “best hit” candidate genes can be severely contaminated by seed-mediated off-target effects. Indeed, for the three screens described here, we determined that, in a typical list of candidate hit genes, much of the phenotypic effect comes from oligos with “active” off-target seeds—there are roughly twofold more such oligos among top-scoring genes than expected by chance (i.e., comparing with a random selection of genes of the same size from the same screen) (Fig. S6). Therefore, a sizable fraction of candidate-gene hits are probably false positives (with respect to the intended on-target effect). Nevertheless, for about half of the phenotypes/screens, significant overlaps between the libraries are detectable (Fig. S1) (see Fig. S9), and these screens will typically lead to confident, true positive hits upon rescreening and further validation.

We find that seed effects are also present in published large-scale RNAi datasets that have been corrected for indirect effects occurring through changes in a single cells microenvironment (27, 28) (“population context”) (Fig. S7). This observation

indicates that seed effects likely act directly on the molecular machinery underlying pathogen infection inside single cells, and not via population context only. In our hands, the phenotypic variance introduced by the seed effect is clearly larger than the variance observed across multiple biological or technical replicates of the same perturbation. Thus, it seems advisable to repeat RNAi measurements using as many different oligo sequences as possible, aiming to average out seed effects, rather than conducting multiple biological replicates of the very same oligos. Furthermore, to systematically learn and correct for seed effects from the data itself is difficult, as most seeds are not represented well enough in genome-wide libraries to learn their phenotypic mean and variance reliably. A possible strategy for the future would be to redesign genome-wide libraries to use a deliberately restricted set of seeds (which should still be on the order of hundreds of seeds—but these seeds would be designed to be represented frequently enough in the library to learn and correct for their effects). To pool distinct oligos intended for the same gene may also be a strategy although we clearly observed significant seed effects in pooled libraries as well (Fig. S8).

In principle, it should be possible to use the known sequences of human mRNAs (particularly their 3' UTR sections) to predict where the various siRNA oligos might bind to mRNAs and how, cumulatively, this might bring about the observed phenotypes. Two software pipelines dedicated to this task have been published already, GESS (“Genome-Wide Enrichment of Seed Sequence Matches”) (25) and Haystack (21). However, at least for the phenotypes screened here, both approaches failed to enrich

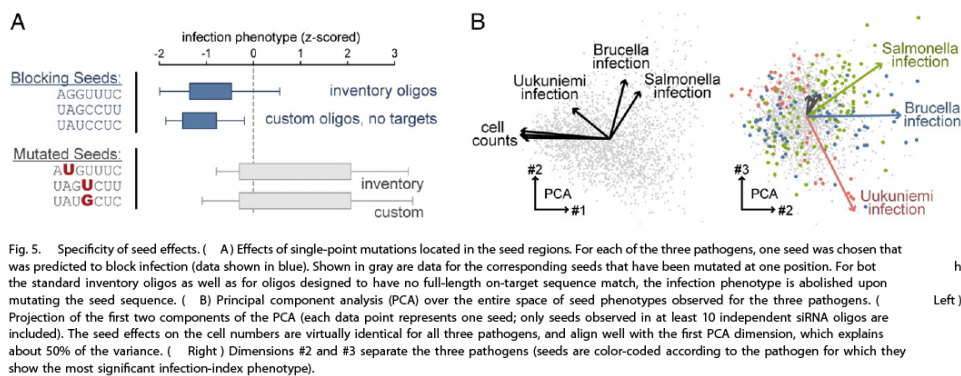


Fig. 5. Specificity of seed effects. (A) Effects of single-point mutations located in the seed regions. For each of the three pathogens, one seed was chosen that was predicted to block infection (data shown in blue). Shown in gray are data for the corresponding seeds that have been mutated at one position. For both the standard inventory oligos as well as for oligos designed to have no full-length on-target sequence match, the infection phenotype is abolished upon mutating the seed sequence. (B) Principal component analysis (PCA) over the entire space of seed phenotypes observed for the three pathogens. (Left) Projection of the first two components of the PCA (each data point represents one seed; only seeds observed in at least 10 independent siRNA oligos are included). The seed effects on the cell numbers are virtually identical for all three pathogens, and align well with the first PCA dimension, which explains about 50% of the variance. (Right) Dimensions #2 and #3 separate the three pathogens (seeds are color-coded according to the pathogen for which they show the most significant infection-index phenotype).

for “causal,” on-target genes, as judged by their inability to improve interlibrary correlations (Fig. S9). In a similar vein, for those active seeds that happen to coincide with known, endogenous human miRNAs, it might be possible to explain some of their off-target effects by searching for predicted targets of those known miRNAs among the top hit lists of the primary screens. However, upon testing three different miRNA target-prediction algorithms (29–31), we did not observe any significant overlap between primary hits and predicted miRNA targets (Fig. S10).

On the positive side, it has become evident that each genome-wide screen represents a powerful interrogation of the sequence space of natural and synthetic miRNA seeds. Natural miRNAs often act as endogenous regulators of entire pathways and processes (as opposed to regulating individual genes only). If we assume that synthetic miRNA seeds can mimic their natural counterparts mechanistically (e.g., with respect to regulating susceptibility to infectious agents), then genome-wide siRNA screens provide a potent tool to assess whether and how host organisms might evolve pathogen resistance by creating new miRNAs. In many cases, it might take only a very small number of mutations to change an existing miRNA into one that is effective against a new pathogen. Experimentally, any strategy for screening the space of miRNA seeds might quickly yield potent therapeutics or laboratory reagents for many processes of interest. Perhaps the most important conclusion of our analysis, however, is that raw “oligo-by-oligo” phenotypic data of genome-wide RNAi screens clearly merit a second look and can yield interesting new insights—provided they are made available to researchers worldwide (32).

Materials and Methods

For the genome-wide infections screens, HeLa cells were grown in 384-well microtiter plates and reverse-transfected with siRNAs 72 h before infections. Pathogens were added, and their cell entry was assessed after a specified

incubation time, using pathogen-specific single-cell readouts in high-throughput automated microscopy imaging of each well. Incubation times were as follows: 4 h for *S. typhimurium*, 44 h for *B. abortus*, and 20 h for Uukuniemi Virus. The detailed experimental methods for each pathogen assay will be published elsewhere. For the data analysis, microscopy images were scaled, corrected for shading, segmented into objects using CellProfiler, and quantitative features were extracted for each cell (up to 200 features per cell). Nuclei and cell bodies were recognized based on DAPI and Actin stainings, respectively. Extracted quantitative features included intensity, texture and shape. Pathogen-specific procedures were then used to discriminate infected from uninfected cells, using Decision Trees with user-provided thresholds on selected single-cell features such as GFP intensity. The phenotypes in each well were normalized first by plate-wise Z-scoring, then by experiment-wide Z-scoring, followed by population regression (Lowess), to control for systematic dependencies between cell-number, -density, and infection rate. Well-by-well resolved, library-wide phenotypes for the three pathogens and the three libraries are available in [Datasets S2–S4](#). The nucleotide sequences of all library siRNA oligos were kindly provided by the commercial vendors. The statistical significance of seed-mediated off-target effects was assessed by aggregating all oligos containing a given seed and comparing the distribution of their phenotypes with the background distribution of phenotypes from the entire screen, using two-sided Kolmogorov-Smirnov tests. Correction for multiple testing was according to Benjamini and Hochberg (33). Human miRNA overexpression experiments were conducted using Dharmacon miRIDIAN microRNA mimics, in the same cell line as the primary screens.

ACKNOWLEDGMENTS. We acknowledge support from Grant 51RT0_126008 (“InfectX”) from SystemsX.ch, the Swiss Initiative for Systems Biology. From the Swiss National Science Foundation, C.D. acknowledges Grant 31003A-132979, W.-D.H. acknowledges Grant 310030-132997/1, and C.v.M. acknowledges Grant 31003A-135688. Work by W.-D.H. and D.A. was supported by a grant from the Vontobel Foundation, work by R.M. and A.H. by an Advanced Grant from the European Research Council, and work by A.C. and S.H.L. by the International PhD Program “Fellowships for Excellence” of Biozentrum Basel.

- Nagy A, Perrimon N, Sandmeyer S, Plasterk R (2003) Tailoring the genome: The power of genetic approaches. *Nat Genet* 33(Suppl):276–284.
- Cong L, et al. (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science* 339(6121):819–823.
- Mali P, et al. (2013) RNA-guided human genome engineering via Cas9. *Science* 339(6121):823–826.
- Miller JC, et al. (2011) A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol* 29(2):143–148.
- Zhang F, et al. (2011) Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nat Biotechnol* 29(2):149–153.
- Mohr SE, Perrimon N (2012) RNAi screening: New approaches, understandings, and organisms. *Wiley Interdiscip Rev RNA* 3(2):145–158.
- Brass AL, et al. (2008) Identification of host proteins required for HIV infection through a functional genomic screen. *Science* 319(5865):921–926.
- Brass AL, et al. (2009) The IFITM proteins mediate cellular resistance to influenza A H1N1 virus, West Nile virus, and dengue virus. *Cell* 139(7):1243–1254.
- Clemente R, Sisman E, Aza-Blanc P, de la Torre JC (2010) Identification of host factors involved in borna disease virus cell entry through a small interfering RNA functional genetic screen. *J Virol* 84(7):3562–3575.
- Karlas A, et al. (2010) Genome-wide RNAi screen identifies human host factors crucial for influenza virus replication. *Nature* 463(7282):818–822.
- Krishnan MN, et al. (2008) RNA interference screen for human genes associated with West Nile virus infection. *Nature* 455(7210):242–245.
- Mercer J, et al. (2012) RNAi screening reveals proteasome- and Cullin3-dependent stages in vaccinia virus infection. *Cell Rep* 2(4):1036–1047.
- Misselwitz B, et al. (2011) RNAi screen of Salmonella invasion shows role of COPII in membrane targeting of cholesterol and Cdc42. *Mol Syst Biol* 7:474.
- Tai AW, et al. (2009) A functional genomic screen identifies cellular cofactors of hepatitis C virus replication. *Cell Host Microbe* 5(3):298–307.
- Zhou H, et al. (2008) Genome-scale RNAi screen for host factors required for HIV replication. *Cell Host Microbe* 4(5):495–504.
- Schmidt EE, et al. (2013) GenomeRNAi: A database for cell-based and in vivo RNAi phenotypes, 2013 update. *Nucleic Acids Res* 41(Database issue):D1021–D1026.
- Bushman FD, et al. (2009) Host cell factors in HIV replication: Meta-analysis of genome-wide studies. *PLoS Pathog* 5(5):e1000437.
- Jackson AL, Linsley PS (2010) Recognizing and avoiding siRNA off-target effects for target identification and therapeutic application. *Nat Rev Drug Discov* 9(1):57–67.
- Jackson AL, et al. (2006) Widespread siRNA “off-target” transcript silencing mediated by seed region sequence complementarity. *RNA* 12(7):1179–1187.
- Birmingham A, et al. (2006) 3′ UTR seed matches, but not overall identity, are associated with RNAi off-targets. *Nat Methods* 3(3):199–204.
- Buehler E, et al. (2012) siRNA off-target effects in genome-wide screens identify signaling pathway members. *Sci Rep* 2:428.
- Jackson AL, et al. (2003) Expression profiling reveals off-target gene regulation by RNAi. *Nat Biotechnol* 21(6):635–637.
- Marine S, Bahl A, Ferrer M, Buehler E (2012) Common seed analysis to identify off-target effects in siRNA screens. *J Biomol Screen* 17(3):370–378.
- Schultz N, et al. (2011) Off-target effects dominate a large-scale RNAi screen for modulators of the TGF- β pathway and reveal microRNA regulation of TGF β 2. *Silence* 2:3.
- Sigoillot FD, et al. (2012) A bioinformatics method identifies prominent off-targeted transcripts in RNAi screens. *Nat Methods* 9(4):363–366.
- Schmaljohn C, Nichol S (2007) Bunyaviridae. *Virology*, ed Fields KD (Lippincott, Williams & Wilkins, Philadelphia), pp 1741–1788.
- Snijder B, et al. (2009) Population context determines cell-to-cell variability in endocytosis and virus infection. *Nature* 461(7263):520–523.
- Snijder B, et al. (2012) Single-cell analysis of population context advances RNAi screening at multiple levels. *Mol Syst Biol* 8(1):579.
- Garcia DM, et al. (2011) Weak seed-pairing stability and high target-site abundance decrease the proficiency of si-6 and other microRNAs. *Nat Struct Mol Biol* 18(10):1139–1146.
- Vejnar CE, Zdobnov EM (2012) MiRmap: Comprehensive prediction of microRNA target repression strength. *Nucleic Acids Res* 40(22):11673–11683.
- Hsu JB, et al. (2011) miRTar: An integrated system for identifying miRNA-target interactions in human. *BMC Bioinformatics* 12:300.
- Shamu CE, Wiemann S, Boutros M (2012) On target: A public repository for large-scale RNAi experiments. *Nat Cell Biol* 14(2):115.
- Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc B* 57(1):289–300.

Supporting Information

Franceschini et al. 10.1073/pnas.1402353111

PNAS

a correlations between siRNA libraries from different vendors

screen / phenotype	Vendor A vs. Vendor B		Vendor A vs. Vendor C		Vendor B vs. Vendor C	
	on-target correlation	off-target correlation	on-target correlation	off-target correlation	on-target correlation	off-target correlation
<i>Brucella</i> : infection index phenotype	0.148*	0.496***	0.040	0.668***	-0.027	0.564***
<i>Brucella</i> : cell number phenotype	0.146*	0.377**	0.112	0.509***	0.167*	0.544***
<i>Salmonella</i> : infection index phenotype	0.026	0.311**	0.044	0.252***	0.041	0.119**
<i>Salmonella</i> : cell number phenotype	0.076	0.211*	0.069	0.420***	0.046	0.329***

b correlations within siRNA libraries from different vendors

Vendor A	on-target correlation	off-target correlation	off-target (downsampled)
<i>Brucella</i> : infection index phenotype	0.092	0.592***	0.585***
<i>Brucella</i> : cell number phenotype	0.074	0.444**	0.448**
<i>Salmonella</i> : infection index phenotype	0.095	0.627***	0.629***
<i>Salmonella</i> : cell number phenotype	0.099	0.478**	0.479**

Vendor B	on-target correlation	off-target correlation	off-target (downsampled)
<i>Brucella</i> : infection index phenotype	0.031	0.574***	0.574***
<i>Brucella</i> : cell number phenotype	0.139*	0.572***	0.573***
<i>Salmonella</i> : infection index phenotype	0.141*	0.451***	0.454***
<i>Salmonella</i> : cell number phenotype	0.099	0.331**	0.333**

Vendor C	on-target correlation	off-target correlation	off-target (downsampled)
<i>Brucella</i> : infection index phenotype	0.071***	0.750***	0.649***
<i>Brucella</i> : cell number phenotype	0.394***	0.646***	0.538***
<i>Salmonella</i> : infection index phenotype	0.021	0.786***	0.690***
<i>Salmonella</i> : cell number phenotype	0.108***	0.679***	0.578***
<i>Uukun. Virus</i> : infection index phenotype	0.235***	0.727***	0.623***
<i>Uukun. Virus</i> : cell number phenotype	0.122**	0.724***	0.629***

* p <= 10⁻³ ** p <= 10⁻⁶ *** p <= 10⁻¹²

Fig. S1. Reproducibility between distinct siRNA oligos. (A) reproducibility between distinct library vendors. This table is based on the set of genes that are shared among all three siRNA libraries used here (i.e., several hundred kinases and kinase-related genes). The "on-target" correlations are defined as correlations of observed phenotypes per gene: i.e., after averaging over all siRNA oligos for a given gene in a given library. In contrast, "off-target" correlations are defined as correlations of observed phenotypes per seed: i.e., after averaging over all siRNA oligos in a given library that share a given seed sequence. All correlation values shown are Spearman correlations (rank-based). (B) Reproducibility within the libraries of distinct vendors. For these correlations, each library has been randomly split into a set "1" and a set "2," such that the distinct oligos for each given gene were randomly assigned to sets 1 and 2. The correlations shown are those between the two sets. In addition, the third column shows correlations after downsampling each seed to cover at most four oligos each (to put the coverage of individual seeds on an equivalent footing with the coverage of individual genes). Vendor A, Ambion, Silencer Select Human Kinome V4, 3 siRNAs per gene, 710 kinases and kinase-related genes, 2,130 siRNAs total, purchased in 2011. Vendor B, Dharmacon, Human ON-TARGETplus, 4 siRNAs per gene, 715 kinases and kinase-related genes, 2,860 siRNAs total, purchased in 2012. Vendor C, Qiagen, HP GenomeWide, 4 siRNAs per gene, genome-wide library, 22,402 genes, 91,800 siRNAs total, purchased in 2005.

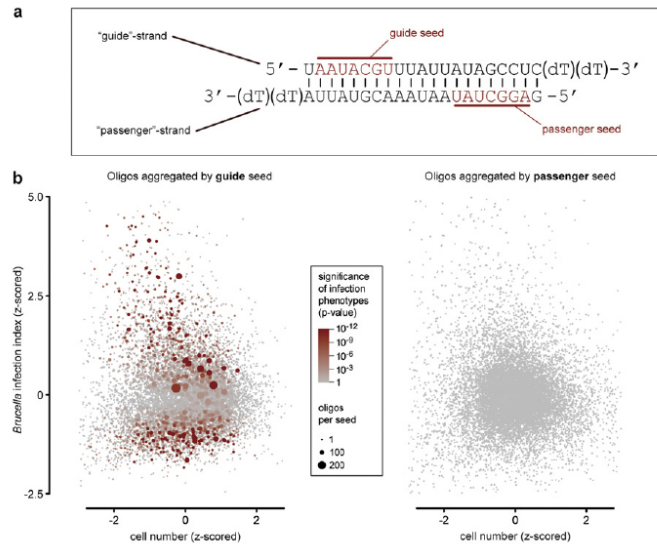


Fig. S2. Guide strands versus passenger strands. (A) Assuming a double-stranded design of the siRNA reagents, there are two potential active seed sequences (i.e., after the strands have separated). (B) siRNA oligo phenotypes aggregated per seed, as in Fig. 2. Only the guide seed shows strongly significant phenotypes, after correcting for multiple testing.

<p>Controls</p>	<p>random control oligos (64 inventory oligos from genome-wide library, randomly chosen)</p> <p>AUAUGUAACGAAUAUUUCUG AUCAACACUAUACACCCUCUG UAACACUGAAAGGAUACUCA UAACCAUUGGGUUAUUUCGG UAACUCUUCUGAAAGUCGGG UAUUGGCAAAUGAAGGCG UACACCUUUGAGGUCACCCUG UACAUCCAUUAUAUGUCUUG UACCUUAUAUAUCUGGCUU UAGAAGUUUCUCCAGGCG UAGAUAUAGACUUAUCCAGU UAGCUGGUCGUAUCUCCGA UAGGAGUAACACUCCACAG UAGGUAGCAAGAGUCCGGC UAGGUUCAGCUACUCCGG UAUAACUUAUUCUACCCAG</p> <p>predicted neutral seeds (seeds predicted to be infection-neutral, for all three pathogens; 4 seeds with 4 inventory oligos each)</p> <p>AUUGACUUGAAUAUUCCAG UUUGACUUAGGAUCCUGCUG UUUGACUUGAAUCCAGAA UUUGACUUGUAUCCUAGCAG</p>	<p>UUAUGGAAUACUUAUUUCUGA UUAUUCUCCAGUAGCAACCCUG UUAUCCAUUUGAAGAGCCCGU UUAUCCUUUGCUUAUAUUUCU UUAUGAUGCCAUUAAGCCGG UCGAUCCUGAUGAAACCCUG UCUAUGGCGUUAAGCACCUG UGUAUAUAGAAUUAUAUGUAG UUAUUCUUGAGGUUAACGCAU UUAUCCUUUAUUUCUCCG UUAUUCACAUUAGUCUUGCAC UUAUUGCAUUCUGUCUGCU UUAUUGUUAUAUAGCAGCUG UUAUUAUACGUAUUAUGCCU UUAUUAUAUAUUAUUCUUCU</p> <p>infection-blocking seeds (library inventory oligos)</p> <p>AUAUCCUCCACCUCCACCAUA UUAUCCUCCAGAAUCUGUAU UUAUCCUCCAGUACACACUGGA UUAUCCUCCAGAAUUCGUCG UUAUCCUCCAGAAUUAUAGGGA UUAUCCUCCAGUACCAUUAUU UUAUCCUCCAGUUAUUAUUCU UUAUCCUCCAGUUAAGCCU</p> <p>infection-enhancing seeds (library inventory oligos)</p> <p>UUCCUUAACAGCCACCAAGUA UUCCUUAACAGCGUUAUCUA</p>	<p>UUUUGAAUCCAAUUCUCCCA UUCAGAUAGCUGACCUUGGUU UUCCUUUUGUACACGCAUUGU UUCUGUAUAAAGUCGUAUGGA UUCUUGUCCAUUGUCUCUUG UUGAAGAACAGGAUGGUCCCA UUGAAGAAGGUUAUAUGUAG UUGAUAUAGCGUUCACCGAG UUGACUUAAGGCCAACCUCU UUGAUAUUUGUCUGAUUGU UUCCUUUUGAUGUAGCCUG UUGCUCUUAUACUACGCCUG UUGACUCUUAUCCGAAAGUG UUGCUCUUAUAUUCUUGAG UUGGUAAGGACUUAUUGGG UUGAUAUAGCAUUCUUCU</p> <p>infection-blocking seeds (custom-designed oligos)</p> <p>UUAUCCUCCAGGUUAUCCAAU UUAUCCUCCAGUACUCCUUCU UUAUCCUCCAGUUAUCCUUCU UUAUCCUCCAGUUAUCCUUCU</p> <p>infection-enhancing seeds (custom-designed oligos)</p> <p>UUCCUUAACAGGGGAUUAUAG UUCCUUAACAGGCAUUAUCCU</p>	<p>UUGUGGUCUGAUCUGACCCUG UUGUUCUCUAAAGAUACUCUGU UUGUUUAAGGUAACACUCUGUA UUUAACAGUUAACACUCUCUG UUUAACAUCAACACUCUGUU UUUAACUAUAUUAUAUAAG UUUAUAUUGUAUAUAUUCGUA UUUAUAUUGGAUAUAUUGCU UUUAUUGGAAUACACCCUG UUUAUUCUGCAUAUAUAGU UUUAUUCUCCAAUUAUAGCUC UUUAUUCUGCUUAUAUAGU UUUAUUGCAAUAUUAUAGU UUUCUUCUUAUAUUAUAGCUC UUUGCAUAGCAGUAGCCGGG</p> <p>UUUGCUGAAGCUGGGUCCUG UUUGCUGAGGGCAAGUCCAG UUUGCUGAUUAUUAUAGCUC UUUGCUGAUUGAAUAGCUC</p> <p>UUAGUAACAUAUCCAUAGAG UUAGUAACAUAUCCAUAGAG UUAGUAACAUAUCCAUAGAG UUAGUAACAUAUCCAUAGAG</p> <p>UUAGUUUAGGGACUGGCCUGA UUAGUUUAGGAAACAGAGAG UUAGUUUAGGCUUAUAGUAG UUAGUUUAGGCUUAUAGCACA</p> <p>UUAGUUUAGGCUUAUAGUAG UUAGUUUAGGCUUAUAGCACA</p> <p>UUAGUUUAGGCUUAUAGUAG UUAGUUUAGGCUUAUAGCACA</p> <p>UUAGUUUAGGCUUAUAGUAG UUAGUUUAGGCUUAUAGCACA</p> <p>UUAGUUUAGGCUUAUAGUAG UUAGUUUAGGCUUAUAGCACA</p> <p>UUAGUUUAGGCUUAUAGUAG UUAGUUUAGGCUUAUAGCACA</p>
<p>Brucella</p>	<p>infection-blocking seeds (library inventory oligos)</p> <p>AUAUCCUCCACCUCCACCAUA UUAUCCUCCAGAAUCUGUAU UUAUCCUCCAGUACACACUGGA UUAUCCUCCAGAAUUCGUCG UUAUCCUCCAGAAUUAUAGGGA UUAUCCUCCAGUACCAUUAUU UUAUCCUCCAGUUAUUAUUCU UUAUCCUCCAGUUAAGCCU</p> <p>infection-enhancing seeds (library inventory oligos)</p> <p>UUCCUUAACAGCCACCAAGUA UUCCUUAACAGCGUUAUCUA</p>	<p>UUGAGUACACUUAUUCUUCUG UUGAGUACCUUUAUUAUUCUA UUGAGUACUUAUUAUUAUUCUA UUGAGUACUUCUUCUUCGCGG UUGCUCUUAUAUUAUUCUCCGA UUGCUCUUAUAUUAUUAUUCG UUGCUCUUAUAUUAUUAUUCG UUGCUCUUAUAUUAUUAUUCG</p> <p>infection-blocking seeds (custom-designed oligos)</p> <p>UUAUCCUCCAGGUUAUCCAAU UUAUCCUCCAGUACUCCUUCU UUAUCCUCCAGUUAUCCUUCU UUAUCCUCCAGUUAUCCUUCU</p> <p>infection-enhancing seeds (custom-designed oligos)</p> <p>UUCCUUAACAGGGGAUUAUAG UUCCUUAACAGGCAUUAUCCU</p>	<p>UUAUCCUCCACCUCCACCAUA UUAUCCUCCAGAAUCUGUAU UUAUCCUCCAGUACACACUGGA UUAUCCUCCAGAAUUCGUCG UUAUCCUCCAGAAUUAUAGGGA UUAUCCUCCAGUACCAUUAUU UUAUCCUCCAGUUAUUAUUCU UUAUCCUCCAGUUAAGCCU</p> <p>infection-blocking seeds (library inventory oligos)</p> <p>UUAGUCUUUUUAUUAAGACCAG UUAGUCUUUAUUAAGCAGUUGU UUAGUCUUUAUUAUUAUUAUUCU UUAGUCUUUUUUUUAUUAUUAU UUAGUCUUUUUUUUAUUAUUAU UUAGUCUUUUUUUUAUUAUUAU UUAGUCUUUUUUUUAUUAUUAU UUAGUCUUUUUUUUAUUAUUAU</p> <p>infection-enhancing seeds (library inventory oligos)</p> <p>UUUGGAGUUAUUAUUAUUAUUCU UUUGGAGUUAUUAUUAUUAUUCG UUUGGAGUUAUUAUUAUUAUUCG UUUGGAGUUAUUAUUAUUAUUCG</p> <p>infection-blocking seeds (custom-designed oligos)</p> <p>UUAUCCUCCAGGUUAUCCAAU UUAUCCUCCAGUACUCCUUCU UUAUCCUCCAGUUAUCCUUCU UUAUCCUCCAGUUAUCCUUCU</p> <p>infection-enhancing seeds (custom-designed oligos)</p> <p>UUCCUUAACAGGGGAUUAUAG UUCCUUAACAGGCAUUAUCCU</p>	<p>UUGAGUAACAUAUCCAUAGAG UUGAGUAACAUAUCCAUAGAG UUGAGUAACAUAUCCAUAGAG UUGAGUAACAUAUCCAUAGAG</p> <p>UUAGUUUAGGGACUGGCCUGA UUAGUUUAGGAAACAGAGAG UUAGUUUAGGCUUAUAGUAG UUAGUUUAGGCUUAUAGCACA</p>
<p>Salmonella</p>	<p>infection-blocking seeds (library inventory oligos)</p> <p>AUAGUUUGCCUUGUUCGCGUG UUAGUUUGCUAUAUUAAGCAGCGU UUAGUUUGCUAUAUUAAGCAGCGU UUAGUUUGGAAACGACAUUGCAU UACAAUUCAAUAUUAUUAUUAU UACAAUUCCAAUAUUAUUAUUAU UACAAUUCCAAUAUUAUUAUUAU UACAAUUCCAAUAUUAUUAUUAU</p> <p>infection-enhancing seeds (library inventory oligos)</p> <p>UUUGGAGUUAUUAUUAUUAUUCU UUUGGAGUUAUUAUUAUUAUUCG UUUGGAGUUAUUAUUAUUAUUCU UUUGGAGUUAUUAUUAUUAUUCG</p>	<p>UUAGUCUUUUUAUUAAGACCAG UUAGUCUUUAUUAAGCAGUUGU UUAGUCUUUAUUAUUAUUAUUCU UUAGUCUUUUUUUUAUUAUUAU UUAGUCUUUUUUUUAUUAUUAU UUAGUCUUUUUUUUAUUAUUAU UUAGUCUUUUUUUUAUUAUUAU UUAGUCUUUUUUUUAUUAUUAU</p> <p>infection-blocking seeds (library inventory oligos)</p> <p>UUAGUUUAGGCUUAUAGUAG UUAGUUUAGGCUUAUAGCACA UUAGUUUAGGCUUAUAGUAG UUAGUUUAGGCUUAUAGCACA</p> <p>infection-enhancing seeds (library inventory oligos)</p> <p>UUUGGAGUUAUUAUUAUUAUUCU UUUGGAGUUAUUAUUAUUAUUCG UUUGGAGUUAUUAUUAUUAUUCU UUUGGAGUUAUUAUUAUUAUUCG</p>	<p>UUAUCCUCCAGGUUAUCCAAU UUAUCCUCCAGUACUCCUUCU UUAUCCUCCAGUUAUCCUUCU UUAUCCUCCAGUUAUCCUUCU</p> <p>infection-blocking seeds (custom-designed oligos)</p> <p>UUAUCCUCCAGGUUAUCCAAU UUAUCCUCCAGUACUCCUUCU UUAUCCUCCAGUUAUCCUUCU UUAUCCUCCAGUUAUCCUUCU</p> <p>infection-enhancing seeds (custom-designed oligos)</p> <p>UUCCUUAACAGGGGAUUAUAG UUCCUUAACAGGCAUUAUCCU</p>	<p>UUAGUUUAGGGACUGGCCUGA UUAGUUUAGGAAACAGAGAG UUAGUUUAGGCUUAUAGUAG UUAGUUUAGGCUUAUAGCACA</p>
<p>Uukuniemi Vir.</p>	<p>infection-blocking seeds (library inventory oligos)</p> <p>AUAGAGAAUAUUAUUAUUAUUAU AUAGAGAAUAUUAUUAUUAUUAU UUAGAGAAUAUUAUUAUUAUUAU UUAGAGAAUAUUAUUAUUAUUAU UAGGUUUUUAUUAUUAUUAUUAU UAGGUUUUUAUUAUUAUUAUUAU UAGGUUUUUAUUAUUAUUAUUAU UAGGUUUUUAUUAUUAUUAUUAU</p> <p>infection-enhancing seeds (library inventory oligos)</p> <p>UUUAGUUUAUUAUUAUUAUUAU UUUAGUUUAUUAUUAUUAUUAU UUUAGUUUAUUAUUAUUAUUAU UUUAGUUUAUUAUUAUUAUUAU</p>	<p>UUAGUUUAGGCUUAUAGUAG UUAGUUUAGGCUUAUAGCACA UUAGUUUAGGCUUAUAGUAG UUAGUUUAGGCUUAUAGCACA</p> <p>infection-blocking seeds (library inventory oligos)</p> <p>UUAGUUUAGGCUUAUAGUAG UUAGUUUAGGCUUAUAGCACA UUAGUUUAGGCUUAUAGUAG UUAGUUUAGGCUUAUAGCACA</p> <p>infection-enhancing seeds (library inventory oligos)</p> <p>UUUAGUUUAUUAUUAUUAUUAU UUUAGUUUAUUAUUAUUAUUAU UUUAGUUUAUUAUUAUUAUUAU UUUAGUUUAUUAUUAUUAUUAU</p>	<p>UUAGUUUAGGCUUAUAGUAG UUAGUUUAGGCUUAUAGCACA UUAGUUUAGGCUUAUAGUAG UUAGUUUAGGCUUAUAGCACA</p> <p>infection-blocking seeds (custom-designed oligos)</p> <p>UUAGUUUAGGCUUAUAGUAG UUAGUUUAGGCUUAUAGCACA UUAGUUUAGGCUUAUAGUAG UUAGUUUAGGCUUAUAGCACA</p> <p>infection-enhancing seeds (custom-designed oligos)</p> <p>UUUAGUUUAUUAUUAUUAUUAU UUUAGUUUAUUAUUAUUAUUAU UUUAGUUUAUUAUUAUUAUUAU UUUAGUUUAUUAUUAUUAUUAU</p>	<p>UUAGUUUAGGCUUAUAGUAG UUAGUUUAGGCUUAUAGCACA UUAGUUUAGGCUUAUAGUAG UUAGUUUAGGCUUAUAGCACA</p> <p>UUAGUUUAGGCUUAUAGUAG UUAGUUUAGGCUUAUAGCACA UUAGUUUAGGCUUAUAGUAG UUAGUUUAGGCUUAUAGCACA</p> <p>UUAGUUUAGGCUUAUAGUAG UUAGUUUAGGCUUAUAGCACA UUAGUUUAGGCUUAUAGUAG UUAGUUUAGGCUUAUAGCACA</p> <p>UUAGUUUAGGCUUAUAGUAG UUAGUUUAGGCUUAUAGCACA UUAGUUUAGGCUUAUAGUAG UUAGUUUAGGCUUAUAGCACA</p>

Fig. S3. Detailed siRNA oligo sequences for Fig. 3. In Fig. 3, the phenotypes of inventory siRNA oligos of the genome-wide library are compared against phenotypes of custom-designed oligos lacking a full sequence match to any known gene (i.e., lacking an on-target component by design). Here, the full set of oligos in this experiment is shown. Sequences are denoted in the directionality of the actual, active siRNA molecule (" guide strand "); the position of the seed sequences is at residues 2 to 8.

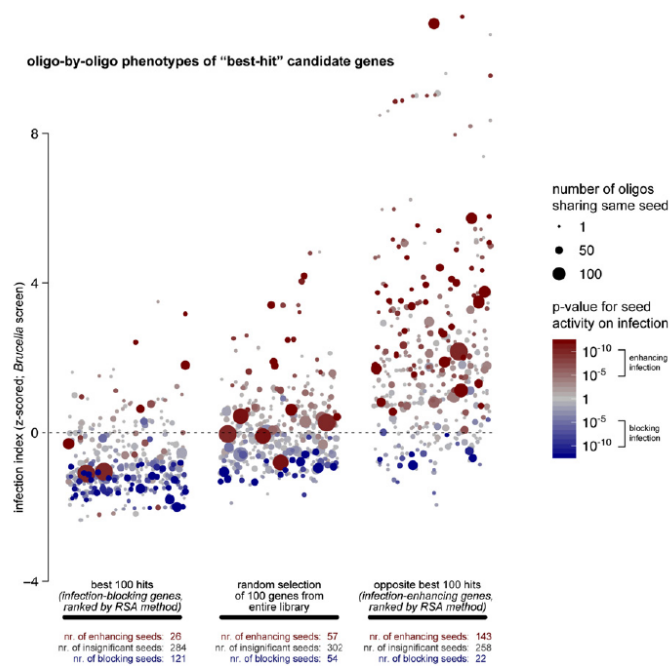


Fig. S6. Influence of the seed effect on "best hit" lists. For the genome-wide *Brucella* screen, all genes were ranked according to phenotypic strength and consistency by the widely used "Redundant siRNA Activity" ("RSA") method (1), and the best hits are shown here (together with a random set for comparison). Each data point in the figure corresponds to one siRNA oligo (all oligos designed for a given gene are vertically aligned at the same x-axis position). Genes are ranked according to the RSA scores, with the best hit (strongest infection reduction) on the far left. The position on the y axis corresponds to the observed phenotype of each oligo in the genome-wide screen. The size of each dot indicates how many other oligos share the same seed sequence, and the color indicates whether that seed sequence alone causes a phenotype. Note how the top 100 hits blocking infection are enriched in "blue" seeds, and the opposite top 100 hits enhancing infection are enriched in "red" seeds.

1. König R, et al. (2007) A probability-based approach for the analysis of large-scale RNAi screens. *Nat Methods* 4(10):847–849.

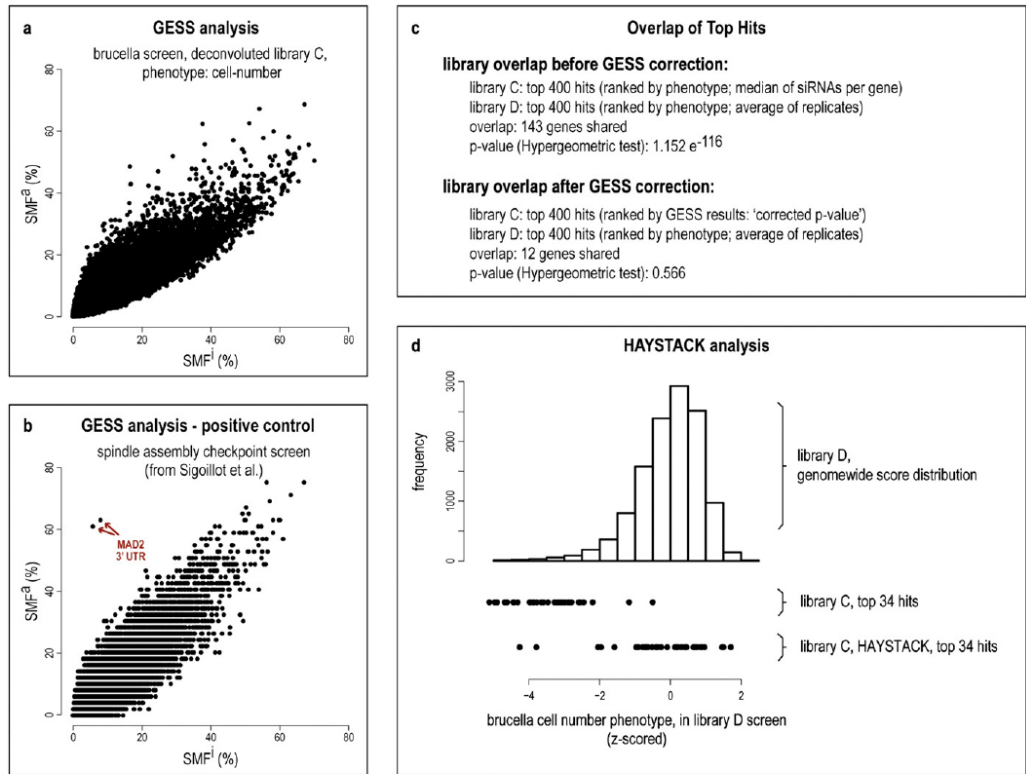


Fig. S9. GESS and Haystack analysis of off-target effects. The software packages GESS ("Genome-wide Enrichment of Seed Sequence Matches") (1) and Haystack (2) have been demonstrated to be capable of selecting true positive hits from genome-wide RNAi screens, at least in some instances. They were tested here on the infection and cell-growth phenotypes reported in the Results section, but they did not show detectable power to enrich for true positive hits. This lack of power could be due to the nature of the phenotypes tested—our phenotypes may be "complex" in the sense that they presumably rely on the involvement of many individual genes. This would result in a very large number of potential off-target effects working additively, perhaps overwhelming the algorithms (other explanations may also exist of course). As a representative screen, in this figure, we show the results for the cell-growth phenotype in Brucella because this phenotype is one of those with the strongest on-target effects as demonstrated by interlibrary overlap testing. (A) Typical GESS result plot showing the ratio of active vs. inactive seeds [Seed-Match Frequency (SMF)^a vs. SMFⁱ] matching to the sequences of human mRNA 3' UTRs. Each dot in the plot corresponds to one UTR of a human protein-coding mRNA. As recommended by the GESS authors, the top 10% of seeds in the screen were considered as "active" seeds, and the bottom 10% of seeds were considered as "inactive." No overt outliers outside of the distribution are seen (compare also with the plot in B below). Nevertheless, there are some UTRs with moderately high active/inactive ratios, and the top 400 of these are further tested in C. (B) As a positive control for the GESS procedure, the same plot was prepared for the screen data from the original GESS publication. Indeed, one clear outlier is seen, corresponding to a gene whose on-target effect on the phenotype can explain many of the off-target effects seen in that screen. (C) To test the relevance of UTRs reported by GESS to be high-ranking in terms of active/inactive ratio in our screen, their overlap to an uncorrected, independent library was tested. Here, library D is a pooled library, meaning that it consists of sets of several siRNAs for each gene (Dharmacon ON-TARGETplus SmartPool, genome-wide, purchased in 2010). Before GESS analysis, libraries C and D do show a limited but highly significant overlap among the genes reported to be relevant for the phenotype tested. This overlap demonstrates that there is a measurable degree of on-target information in both screens. After GESS analysis, the highly ranked genes of library C no longer show any significant enrichment among the top genes of library D. (D) In the case of the HAYSTACK software, similar tests have been executed—however, because HAYSTACK reports only a more limited number of genes as having an on-target signal, the overlap with library D is shown by comparing the score distributions. Before HAYSTACK correction, the top 34 genes of the library C screen are seen left-shifted in the plot: i.e., they also score strongly in the library D screen. However, the 34 genes reported by HAYSTACK instead center around the neutral phenotype in the library D screen (z-score 0).

1. Sigollot FD, et al. (2012) A bioinformatics method identifies prominent off-targeted transcripts in RNAi screens. Nat Methods 9(4):363–366.
 2. Buehler E, et al. (2012) siRNA off-target effects in genome-wide screens identify signaling pathway members. Scientific Reports 2:428.

	seed sequence with significant infection phenotype	average infection phenotype (z-scored)	human endogenous miRNA sharing same seed sequence	TargetScan			miRmap			miRTar		
				number of predicted target genes	overlap with primary screens	expected overlap by chance	number of predicted target genes	overlap with primary screens	expected overlap by chance	number of predicted target genes	overlap with primary screens	expected overlap by chance
Brucella Infection Reduction	UAIACAG	-1.33	hsa-miR-7f2-3p	29	3	1.1	23	0	0.9	2	0	0.1
	UCUCCUC	-1.35	hsa-miR-4279	279	9	10.8	383	8	14.9	-	-	-
	UGCACUU	-1.22	hsa-miR-4325	243	3	9.4	205	3	8.0	-	-	-
	UAGUGGU	-1.15	hsa-miR-4490	209	4	8.1	158	7	6.1	-	-	-
	UUCUUGA	-1.31	hsa-miR-544a	224	6	8.7	280	9	10.9	-	-	-
	UCUCUCU	-1.18	hsa-miR-6071	252	5	11.3	451	9	17.5	-	-	-
	AAAGCAC	-1.33	hsa-miR-6835-3p	112	4	4.3	-	-	-	-	-	-
UAAAGCU	-1.47	hsa-miR-9-3p	46	0	1.8	35	1	1.4	2	0	0.1	
Salmonella Infection Reduction	UUGUUCU	-1.30	hsa-miR-133a-3p	364	18	14.1	286	8	11.1	51	1	2.0
	AAGGCCU	-1.32	hsa-miR-3910	156	3	6.1	373	12	14.5	-	-	-
	UCCUUCU	-1.23	hsa-miR-502-5p	208	3	8.1	156	3	6.1	-	-	-
	UCUCCAA	-1.27	hsa-miR-515-5p	183	9	7.1	340	11	13.2	-	-	-
	UGGUCAC	-1.20	hsa-miR-5687-5p	236	15	9.2	232	7	9.0	-	-	-
Brucella Infection Enhancement	UGAAGCU	2.14	hsa-miR-127-5p	101	5	3.9	293	12	11.3	-	-	-
	UGCAGAC	1.64	hsa-miR-1273b-3p	167	5	6.5	-	-	-	-	-	-
	UAUUGCU	2.28	hsa-miR-155-5p	91	2	3.5	51	1	2.0	829	34	32.1
	UGAAUUG	2.10	hsa-miR-203a	1	0	0.0	76	1	2.9	23	3	0.9
	UGUGUCU	2.16	hsa-miR-218-5p	202	8	7.8	316	14	12.2	88	4	3.4
	UGAGUGA	1.61	hsa-miR-3136-5p	134	1	5.2	102	8	3.9	-	-	-
	UACUUCU	1.36	hsa-miR-3682-5p	60	1	2.3	35	0	1.4	-	-	-
	ACAGGAC	1.74	hsa-miR-4700-3p	236	12	9.1	405	11	15.7	-	-	-
	UUCUGUC	1.43	hsa-miR-4743-3p	44	1	1.7	96	1	3.7	-	-	-
	UGUGGAG	3.63	hsa-miR-4752	152	9	5.9	156	4	6.0	-	-	-
	UUCUCAU	3.31	hsa-miR-4776-3p	239	15	9.2	228	8	8.8	-	-	-
	UGUACAU	2.28	hsa-miR-463-5p	91	2	3.5	78	6	3.0	-	-	-
	AUUUGGU	3.64	hsa-miR-5002-5p	45	6	1.7	115	3	4.5	-	-	-
	UUGGAUU	1.55	hsa-miR-5004-3p	41	0	1.6	70	1	2.7	-	-	-
	UUCUUGU	1.41	hsa-miR-578	11	0	0.4	40	4	1.5	-	-	-
	ACAGUGU	3.63	hsa-miR-624-3p	182	11	7.0	212	7	8.2	1	0	0.0
	UGUUGAA	3.28	hsa-miR-653-5p	73	0	2.8	-	-	-	-	-	-
	UGUGGAA	2.20	hsa-miR-683-5p	175	11	6.8	-	-	-	-	-	-
	UGGUCUC	2.06	hsa-miR-6843-3p	198	2	7.7	-	-	-	-	-	-
	UCCUUCU	1.79	hsa-miR-6868-3p	158	8	6.1	-	-	-	-	-	-
UCUUCUC	2.57	hsa-miR-6868-3p	149	2	5.8	-	-	-	-	-	-	
UUCUUGU	1.71	hsa-miR-7150-5p	167	0	6.5	-	-	-	-	-	-	
UGAAGCC	1.69	hsa-miR-7843-3p	211	7	8.2	-	-	-	-	-	-	
UGACUGU	2.66	hsa-miR-943	138	2	5.3	141	8	5.5	1	0	0.0	
Salmonella Infection Enhancement	AAUUGAG	1.64	hsa-miR-1252-3p	9	0	0.3	-	-	-	-	-	-
	UGAAUUG	2.98	hsa-miR-203a	1	0	0.0	76	1	2.9	23	3	0.9
	UGUGUCU	1.39	hsa-miR-218-5p	202	8	7.8	316	14	12.2	88	4	3.4
	AAAGCUG	2.24	hsa-miR-320a	39	0	1.5	206	7	8.0	482	28	18.7
	UAUAAAG	3.16	hsa-miR-340-5p	-	-	-	5	0	0.2	123	1	4.8
	UGGAGAA	2.03	hsa-miR-4531	133	5	5.1	362	7	14.0	-	-	-
	AGAGACG	2.43	hsa-miR-4773	92	5	3.6	216	8	8.4	-	-	-
	UCUAGAU	2.10	hsa-miR-4777-5p	63	3	2.4	16	1	0.6	-	-	-
	UAUACAC	1.50	hsa-miR-4789-5p	53	2	2.1	26	0	1.0	-	-	-
	AAAGUAC	2.33	hsa-miR-548k	24	0	0.9	65	2	2.5	-	-	-
	UGUADUC	2.46	hsa-miR-675-3p	141	8	5.5	48	3	1.9	-	-	-
	UCUGGAA	1.71	hsa-miR-6833-5p	175	11	6.8	-	-	-	-	-	-

Fig. S10. Human microRNAs corresponding to strongly significant seed sequences. Active seeds that correspond to human miRNAs were selected, and we tested whether any of the predicted targets of those human miRNAs were overlapping with top hits in the primary genome-wide screens. The latter are derived from the union of two genome-wide screens in Brucella (one using Qiagen unpooled and the other Dharmacon pooled libraries), ranking both screens by median infection phenotypes and forming the union of the top 400 hits from each screen. For TargetScan, the best predicted target site for a gene had to surpass a score cutoff of -0.33 to be included. For miRmap, the best target site of a given target gene had to surpass a score of -0.23. For miRTa considered all targets annotated on the miRTarBase website.

Dataset S1. Complete seed phenotypes

[Dataset S1](#)

Each row in this Excel file corresponds to one seven-mer seed sequence. The various columns in the dataset indicate the average phenotypes of all RNAi oligos containing that seed sequence, in all three genome-wide screens. Significance is given as P values, after correction for multiple testing.

Dataset S2. Genome-wide, well-by-well –resolved primary screening results for Brucella

[Dataset S2](#)

The full primary screening results, after image processing and normalization. The infection phenotypes as well as the cell-number phenotypes are reported as z-scores. The siRNA libraries have been remapped to the human genome, meaning that some genes may have a different number of siRNAs assigned to them than what they had assigned in the original library annotation by the vendors. The gene identities and the siRNA sequences outside of the seed region have been anonymized. The files allow one to reproduce all figures and conclusions in the paper. The data are in a standard tab-separated flatfile format.



Dataset S3. Genome-wide, well-by-well –resolved primary screening results for Salmonella

[Dataset S3](#)

The full primary screening results, after image processing and normalization. The infection phenotypes as well as the cell-number phenotypes are reported as z-scores. The siRNA libraries have been remapped to the human genome, meaning that some genes may have a different number of siRNAs assigned to them than what they had assigned in the original library annotation by the vendors. The gene identities and the siRNA sequences outside of the seed region have been anonymized. The files allow one to reproduce all figures and conclusions in the paper. The data are in a standard tab-separated flatfile format.

Dataset S4. Genome-wide, well-by-well –resolved primary screening results for Uukuniemi

[Dataset S4](#)

The full primary screening results, after image processing and normalization. The infection phenotypes as well as the cell-number phenotypes are reported as z-scores. The siRNA libraries have been remapped to the human genome, meaning that some genes may have a different number of siRNAs assigned to them than what they had assigned in the original library annotation by the vendors. The gene identities and the siRNA sequences outside of the seed region have been anonymized. The files allow one to reproduce all figures and conclusions in the paper. The data are in a standard tab-separated flatfile format.

3.2 RESEARCH ARTICLE II

Simultaneous analysis of large-scale RNAi screens for pathogen entry

Pauli Rämö^{†,1}, Anna Drewek^{†,2}, Cécile Arrieumerlou¹, Niko Beerenwinkel^{4,5}, Houchaima Ben-Tekaya¹, Bettina Cardel³, Alain Casanova¹, Raquel Conde-Alvarez⁷, Pascale Cossart⁶, Gábor Csúcs¹¹, Simone Eicher¹, Mario Emmenlauer¹, Urs Greber³, Wolf-Dietrich Hardt⁹, Ari Helenius⁸, Christoph Kasper¹, Andreas Kaufmann¹¹, Saskia Kreibich⁹, Andreas Kühbacher⁶, Peter Kunszt¹³, Shyan Huey Low¹, Jason Mercer⁸, Daria Mudrak³, Simone Muntwiler¹, Lucas Pelkmans³, Javier Pizarro-Cerdá⁶, Michael Podvinec¹², Eva Pujadas¹², Bernd Rinn^{4,5}, Vincent Rouilly¹², Fabian Schmich⁴, Juliane Siebourg-Polster⁴, Berend Snijder³, Michael Stebler¹¹, Gabriel Studer¹, Ewa Szczurek^{4,5}, Matthias Truttmann¹, Christian von Mering³, Andreas Vonderheit¹⁰, Artur Yakimovich³, Peter Bühlmann² & Christoph Dehio^{*,1}

Statement of my own contribution

I contributed to this manuscript by performing the siRNA screens for pathogen *Brucella*. Analysis of the data from these screens contributes to the findings of this paper.

3.2.1 Summary

High-throughput RNAi screens are often performed in this era to knockdown host factors at a systems level, allowing the study of biological process of interest. However, not many of the siRNA screens published thus far were compared. Also, there is generally poor reproducibility between similar screens using siRNA libraries from different vendors, suggesting strong off-target effects in the RNAi screening field [1]. With this, a statistical model named Parallel Mixed Model (PMM) was developed that allows analysis of multiple RNAi screens, performed with a shared library. Eight different pathogen screens and four different commercially available libraries were used for the analysis. Finally, it was shown that PMM is able to improve the statistical power and hit identification compared to other methods. This suggests the advantage of having parallel screening. PMM allows incorporation of RNAi weights according to the quality of the siRNA. Furthermore, PMM also estimates a sharedness score, allowing identification of unique or common genes between the pathogens. All in all, PMM promises a statistical model for better analysis and comparison between high-throughput RNAi screens.

Reference

1. Pache L, Konig R, Chanda SK (2011) Identifying HIV-1 host cell factors by genome-scale RNAi screening. *Methods* 53: 3-12.

3.2.2 Manuscript

Rämö et al. BMC Genomics 2014, 15:1162
<http://www.biomedcentral.com/1471-2164/15/1162>



RESEARCH ARTICLE

Open Access

Simultaneous analysis of large-scale RNAi screens for pathogen entry

Pauli Rämö¹, Anna Drewek¹, Cécile Arriemertou³, Niko Beerenwinkel⁵, Houchaima Ben-Tekaya⁶, Bettina Cardé⁷, Alain Casanova⁸, Raquel Conde-Alvarez⁹, Pascale Cossart¹⁰, Gábor Csúcs¹¹, Simone Eicher¹², Mario Emmenlaue¹³, Urs Greber¹⁴, Wolf-Dietrich Hardt¹⁵, Ari Helenius¹⁶, Christoph Kasper¹⁷, Andreas Kaufman¹⁸, Saskia Kreibich¹⁹, Andreas Kühbach²⁰, Peter Kunszt²¹, Shyan Huey Low²², Jason Merce²³, Daria Mudra²⁴, Simone Muntwiler²⁵, Lucas Pelkman²⁶, Javier Pizarro-Cerdá²⁷, Michael Podvinec²⁸, Eva Pujadas²⁹, Bernd Rinn³⁰, Vincent Rouilly³¹, Fabian Schmicke³², Juliane Siebourg-Polster³³, Berend Snijder³⁴, Michael Stehle³⁵, Gabriel Studer³⁶, Ewa Szczurek³⁷, Matthias Truttmann³⁸, Christian von Merin³⁹, Andreas Vonderheide⁴⁰, Artur Yakimovich⁴¹, Peter Bühlmann⁴² and Christoph Dehid^{*}

Abstract

Background: Large-scale RNAi screening has become an important technology for identifying genes involved in biological processes of interest. However, the quality of large-scale RNAi screening is often deteriorated by off-target effects. In order to find statistically significant effector genes for pathogen entry, we systematically analyzed entry pathways in human host cells for eight pathogens using image-based kinome-wide siRNA screens with siRNAs from three vendors. We propose a Parallel Mixed Model (PMM) approach that simultaneously analyzes several non-identical screens performed with the same RNAi libraries.

Results: We show that PMM gains statistical power for hit detection due to parallel screening. PMM allows incorporating siRNA weights that can be assigned according to available information on RNAi quality. Moreover, PMM is able to estimate a sharedness score that can be used to focus follow-up efforts on generic or specific gene regulators. By fitting a PMM model to our data, we found several novel hit genes for most of the pathogens studied.

Conclusions: Our results show parallel RNAi screening can improve results of individual screens. This is currently particularly interesting when large-scale parallel datasets are becoming more and more publicly available. Our comprehensive siRNA dataset provides a public, freely available source for further statistical and biological analyses in the high-content, high-throughput siRNA screening field.

Keywords: High-throughput high-content RNAi screening, pathogen entry, Linear mixed model, Hit detection

Background

Large-scale RNAi screening is a widely used technology to knock-down expressions of genes and study their protein function in a biological process of interest [1-5]. In several published studies in the field of infection biology, cells perturbed with siRNAs were exposed to pathogens and differences in phenotypic outcomes were measured in order to identify the genes involved in successful infection or to

develop functional models of host signaling and trafficking pathways [6-9].

RNAi libraries are mostly sold in formats containing enough material for numerous large-scale screens. Therefore, several large-scale siRNA screens are typically performed using the same libraries within a unit such as a university or company in order to optimize material costs and to simplify plate handling. However, parallel screens are typically performed and analyzed separately without common protocols or analysis pipelines. Therefore, comparing results between the screens is challenging. Co-operative efforts, such as assays using common key parameters for imaging and data analyses, could enable to

* Correspondence: christoph.dehid@unibas.ch

¹Equal contributors

²Focal Area Infection Biology, Biozentrum, University of Basel, Klingelbarstrasse 70, CH-4056 Basel, Switzerland

Full list of author information is available at the end of the article



© 2014 Rämö et al.; licensee BioMed Central. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

produce more comparable results and gain parallel information for each individual screen. In the field of RNAi screening, there has been progress in relation to the standardization of data publication formats, in particular in the context of the "Minimum Information About an RNAi Experiment" (MIARE, <http://miare.sourceforge.net>) and GenomeRNAi [10] efforts. However, the provided metadata information and data analysis approaches are often diverse so that data comparison between the screens from different laboratories is very difficult.

Poor reproducibility rates of large-scale RNAi screens are a common concern. They are mostly caused by strong off-target effects from particular siRNAs [11-16]. Strategies have been proposed to alleviate the confounding effects of RNAi screens, including experimental [17,18] and statistical approaches [9,19-22]. In this study, we aim to use the parallel screening structure in order to gain statistical power for hit selection in large-scale RNAi screens. We generated high-content siRNA datasets that are uniquely comprehensive in terms of the siRNA libraries and various pathogens used. We employed highly unified protocols for parallel screens and common data analysis pipelines to allow a direct comparison between the readouts of different pathogen screens. In addition to obtain a list of hits for individual pathogens, our aim was to discover shared mechanisms between pathogens. To this purpose, we propose a new statistical method: the Parallel Mixed Model (PMM). Our approach simultaneously takes into account the knock-down effects of several non-identical screens performed in parallel with the same RNAi libraries. Additionally, the PMM provides a local False Discovery Rate (FDR) for every gene, resulting in a probability estimate that a gene is a false positive. We will show that the model improves statistical power thanks to parallel screening and that it yields stable hits, novel to the studied pathogens, without compromising the detection of unique hits for any given single screen.

Results and discussion

High-content siRNA screening

Our InfectX consortium, consisting of eleven research groups, generated kinome-wide siRNA screens for five bacterial pathogens (*Bartonella henselae*, *Brucella abortus*, *Listeria monocytogenes*, *Salmonella typhimurium*, and *Shigella flexneri*) and three viruses (*Adenovirus*, *Rhinovirus*, and *Vaccinia virus*) and systematically analyzed the biological pathways leading to successful infection in human host cells (Figure 1). This choice of bacterial and viral pathogens covered a wide variety of mechanism to invade host cells. *B. henselae* for example, invades host cells by invasome structures [23], the pathogens *S. typhimurium* and *S. flexneri* use the trigger mechanism, while *L. monocytogenes* uses the zipper mechanism [24]. *Adenovirus* and *Rhinovirus* enter

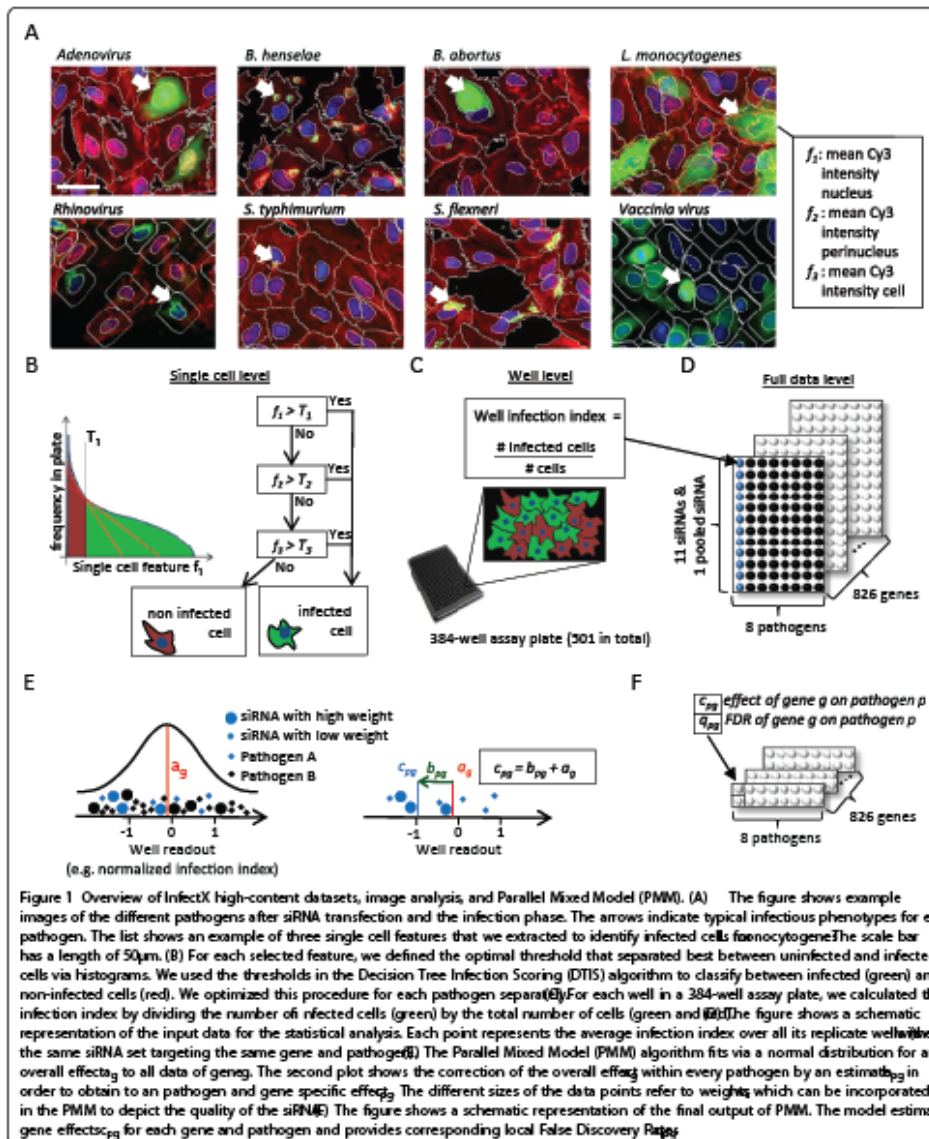
by a dynamin and clathrin dependent pathway [25] and *Vaccinia virus* by macropinocytosis [26].

We conducted the screens in a highly parallel manner under one common protocol for all eight pathogens. We carried out all screens in the same HeLa ATCC-CCL-2 cell line and with the same reagent batches of shared providers. The set of 826 targeted genes comprised almost the whole kinome, plus selected kinome-associated genes, and we targeted each gene by a total of eleven independent siRNAs coming from three manufactures: Ambion (Silencer Select) with 3 siRNAs per gene, Qiagen (Human Kinase siRNA Set V4.1) with 4 siRNAs per gene and Dharmacon (Human ON-TARGETplus) with 4 siRNAs per gene. Additionally, we performed screens where we targeted each kinase with a pool of the four Dharmacon siRNAs (Human ON-TARGETplus SMARTpool). However, not all of the 826 genes have a full set of 11 siRNAs and 1 siRNA pool available. Depending on the pathogen and library, we independently repeated the screens one to six times as replicates (see Additional file 1: Table S1). To obtain an optimal dynamic range of infectivity, we chose the pathogen dose and entry time to be pathogen specific (see Additional file 1: Table S2). We fixed and stained the cells using DAPI or Hoechst to detect nuclei, fluorescent labeled phalloidin to detect actin filaments and the cell body, and a pathogen specific marker to detect infected cells. In a final step, we imaged the screens using microscopes of the same brand. Thus, we only permitted deviations from the common protocols when the infection assay required it.

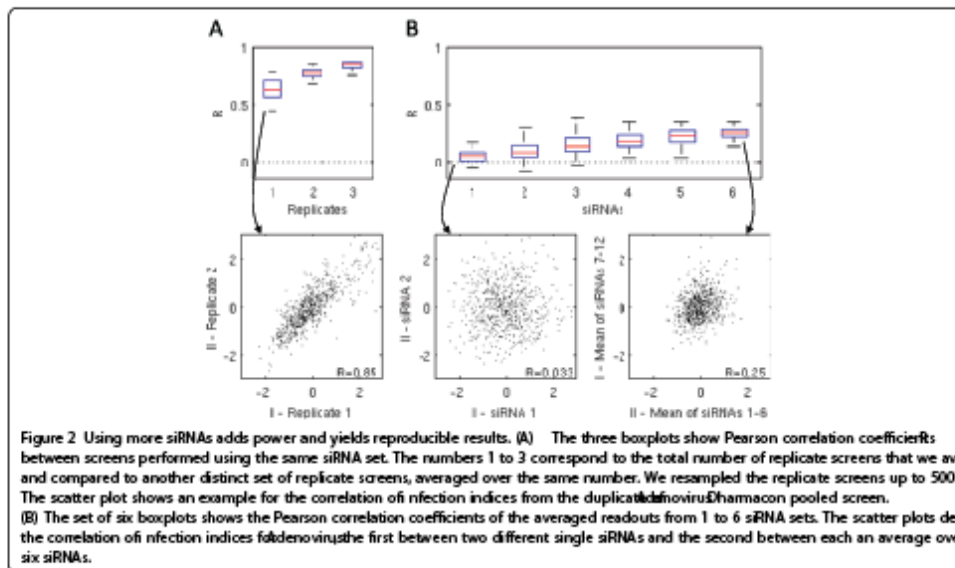
We separately optimized image analysis for each pathogen and established for each pathogen a list of image features that described the phenotypes of infected cells. For example, for *S. flexneri*, we chose as one feature the RFP intensity of the extracted bacteria objects and fol: monocytogenes the mean Cy3 intensity of the cell (see Figure 1A and Additional file 1). In the next step, we classified the cells in each well as infected or uninfected with a Decision Tree Infection Scoring (DTIS) algorithm (see Additional file 1) and obtained a rate of infection per well (infection index) (Figures 1B-C). Besides assay-specific readouts the image analysis also provided several assay-independent readouts (e.g. cell number). We alleviated possible batch effects, dependencies to the population context, and further experimental confounders by data normalization (see Additional file 1) [27-32]. We performed all analyses presented in this paper with the normalized infection index readout, unless otherwise stated.

Data reproducibility

Our data confirmed the reported [20] low reproducibility rates of siRNA data originating from different siRNAs targeting the same genes. The normalized infection indices of two different siRNA sets targeting the same genes



showed a Pearson correlation coefficient R between 0 and 0.2 depending on the screens (Figure 2B). Adding independent siRNAs to the screen yielded an increase in the correlation coefficients, but the correlation still stayed at an unsatisfactory level, even with six separate siRNAs targeting each gene R was between 0.1-0.4 in



averaged and separate sets of six independent siRNAs). In contrast, replicate screens (screens performed using the same protocols and siRNA set, but performed at a different time) were reproducible R was between 0.5-0.9) (Figure 2A). For practical reasons, assuming a similar assay quality as ours, performing screens in duplicates seems sufficient since having more replicates does not improve the data to a great extent (Figure 2A). On the other hand, performing screens at least in duplicates is necessary for quality control and performing only single screens is therefore not recommendable. The cell number readouts (see Additional file 1: Figure S4) showed qualitatively similar results for data reproducibility. In summary, the main error source in our siRNA screening was the bias caused by varying specificity of siRNAs and not by technical variability of the screens.

Parallel Mixed Model (PMM)

Assuming that the sources of variability between different siRNAs targeting the same gene are statistically independent, we can benefit from the fact that the true signal is enhanced by using more siRNAs targeting the individual genes [17] (Figure 2B). In order to increase the statistical power of individual siRNA screens, we performed screens with 11 siRNAs (and one pool of siRNAs) targeting each gene. Moreover, when using the parallel structure in the data and combining data points from all pathogen screens together, we reached $8 \times 12 = 96$ data points for every gene

(averaging over the replicate screens). We propose the Parallel Mixed Model (PMM) as a suitable approach to model the distribution of the siRNA readouts using all data together, including all available siRNAs and pathogen screens.

PMM is composed of a linear mixed model and an assessment of the local False Discovery Rate (FDR) (Figure 1E-F). The linear mixed model is an extension of the ordinary linear model by random effects [33]. In particular, random effects are not determined by fixed coefficients, but by Gaussian distributions. Therefore, we can incorporate the variation among the siRNAs in form of random effects and estimate all effects for different pathogens simultaneously. To be more precise, the linear mixed model consists of a fixed effect μ_p for pathogen p and two random effects a_g for gene g and b_{pg} as a correction term for gene g within pathogen p :

$$y_{pgs} = \mu_p + a_g + b_{pg} + \epsilon_{pgs}$$

where y_{pgs} denotes the readout (for example the normalized infection index of a well) of pathogen p and gene g knocked-down with siRNA s and ϵ_{pgs} denotes the unobserved error term. We fitted the linear mixed model by using the "lmer" function from the "lme4" R-package [34]. The sum of two random effects a_g and b_{pg} describes the total effect of the siRNAs within pathogen p . We

define the estimated effect c_{pg} for gene g within pathogen p as

$$c_{pg} = \frac{1}{a} \log \left(\frac{b}{b_{pg}} \right)$$

A positive estimated c_{pg} effect means that the infection level was enhanced if the corresponding gene is knocked down. A negative effect means that the infection level was reduced. To distinguish hit genes, PMM provides as second step an estimate α_{pg} of the local False Discovery Rate (FDR). We computed the local False Discovery Rate using the approach presented in [59] and the "locfdr" function in the R-package of the same name [35]. We assigned the local False Discovery Rate to every gene and interpreted it as the probability describing how likely the corresponding gene is a false discovery (see Methods for more details). The PMM method is published as "PMM" R-package on the InfectX data-access page.

As a first verification for the increase in power by simultaneously using the parallel screening structure, we resampled datasets, each consisting of a fixed number of siRNAs and pathogens, and fitted the PMM, respectively the Moderated T-Test (MTT) [36] for the case of one pathogen (see Methods for details). We evaluated the mean and variation (i.e. stability) of the ranks in the ordered lists of genes based on their estimated c_{pg} values

over 1000 resampling runs for MET (a known effector gene for *L. monocytogenes* [37]), MTOR (a role of MTOR in the infection pathways of several pathogens has already been established [6,15,38]) and a non-hit ALK as control (Figure 3). The results showed, in particular in the case of MTOR, that the rank and its stability improved by simultaneously using more siRNAs and pathogens. In the case of MET the use of parallel screens did not cause an increase in statistical power, since MET was a hit for *L. monocytogenes* only. However, for MET there was no reduction of statistical power either. These examples already indicated that the parallel screening structure and PMM can be used to more reliably discover expected effector genes even in the case where only a fraction of effector genes is shared between the screens.

Analysis of siRNA libraries

PMM allows the assignment of weights to each siRNA (see Methods). With weighting, we can assign more power to siRNAs that are estimated to have little off-target effects and strong knock-down efficiencies. Within this study, we weighted siRNAs according to the reproducibility in terms of correlation of their corresponding library to other libraries (Figure 4A). There are several potential other ways how weights could be determined. However, we did not follow them further within the context of this paper.

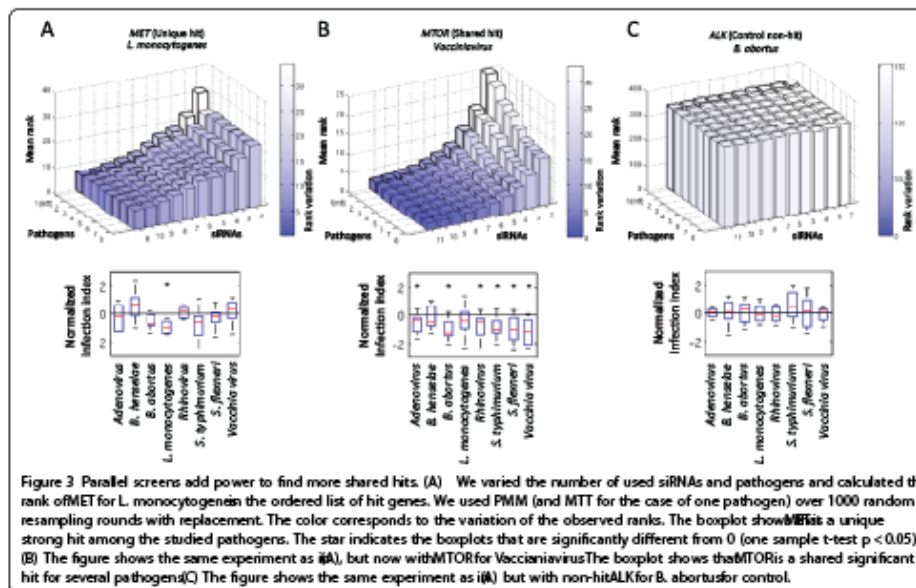
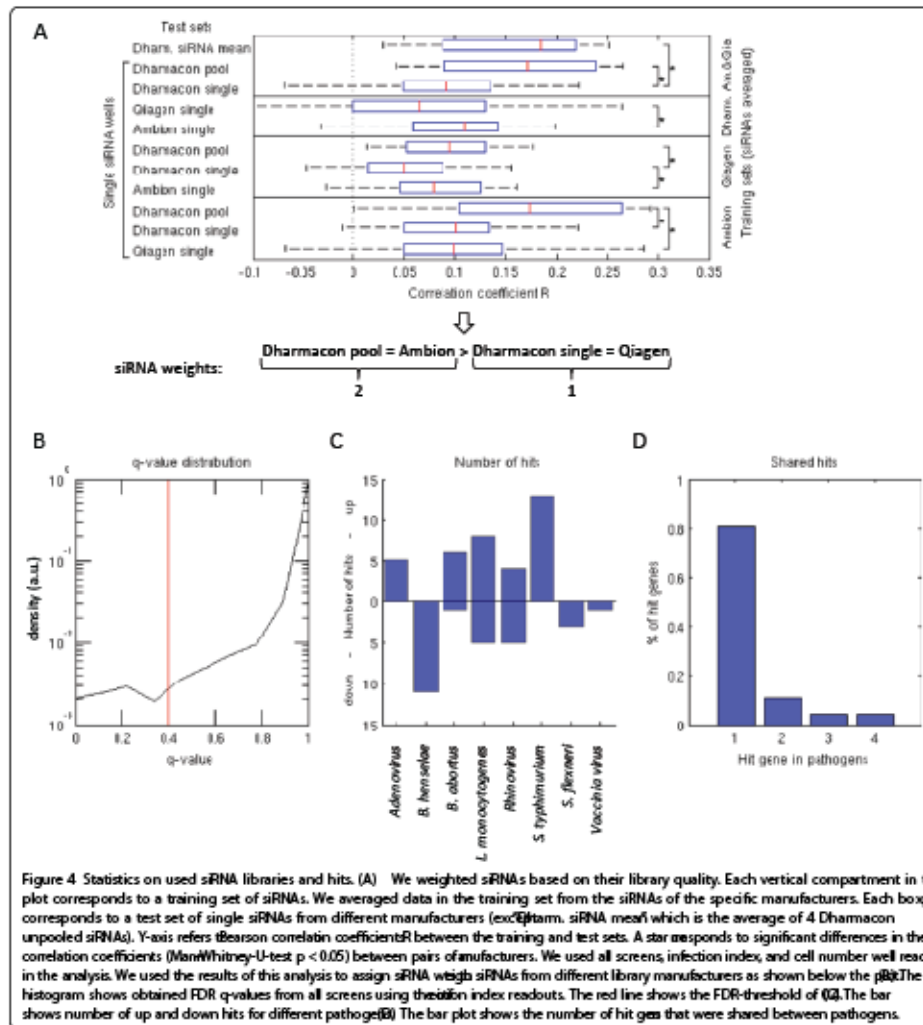


Figure 3 Parallel screens add power to find more shared hits. (A) We varied the number of used siRNAs and pathogens and calculated the rank of MET for *L. monocytogenes* the ordered list of hit genes. We used PMM (and MTT for the case of one pathogen) over 1000 random resampling rounds with replacement. The color corresponds to the variation of the observed ranks. The boxplot shows that MET is a unique strong hit among the studied pathogens. The star indicates the boxplots that are significantly different from 0 (one sample t-test $p < 0.05$). (B) The figure shows the same experiment as (A), but now with MTOR for Vaccinia virus. The boxplot shows that MTOR is a shared significant hit for several pathogens. (C) The figure shows the same experiment as (A), but with non-hit ALK for *B. abortus* for control.



We cross-validated different libraries to each other by fixing one library manufacturer (training set) at a time (Figure 4A). We averaged phenotypic readouts from siRNAs targeting the same gene in the training set in order to obtain reference gene readouts. In this analysis we used both infection index and cell number readouts. We then compared single siRNA readouts of the remaining two library manufacturers (test set) to the reference gene

readouts. The Pearson correlation coefficients of the test sets enable to quantify which of the two test manufacturers produces more reproducible results. By repeating the procedure for all manufacturers as the training set we could order the manufacturers in terms of their reproducibility performance. Our results based on phenotypic readouts showed that the pooled Dharmacon library performed the best. The pooled library was

followed by the unpooled libraries of Ambion, Dharmacon, and Qiagen in this order. However, there were no statistically significant differences (Wilcoxon rank-sum test $p < 0.05$) between Dharmacon pooled and Ambion single, and Dharmacon single and Qiagen single siRNA data reproducibility. In addition, the data showed that the averaged single siRNAs of Dharmacon performed at most as

good as the single pooled siRNA consisting of the same siRNAs. This indicated that for most screening purposes, it is more practical to use the pooled library instead of several unpooled libraries. This result of better performance of pooled libraries compared to averaged single siRNA libraries is in contradiction with what has been reported in [19]. However, good quality single siRNA libraries (such as

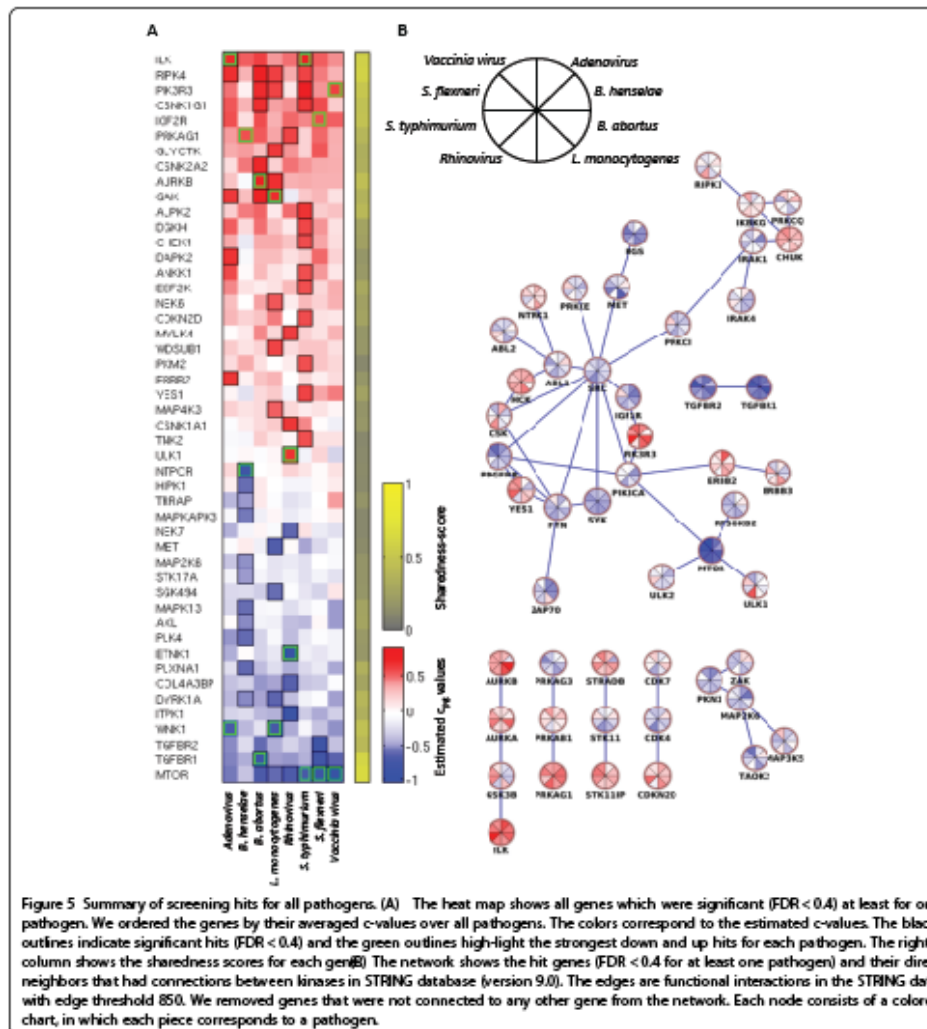


Figure 5 Summary of screening hits for all pathogens. (A) The heat map shows all genes which were significant (FDR < 0.4) at least for one pathogen. We ordered the genes by their averaged c-values over all pathogens. The colors correspond to the estimated c-values. The black outlines indicate significant hits (FDR < 0.4) and the green outlines highlight the strongest down and up hits for each pathogen. The rightmost column shows the sharedness scores for each gene (B) The network shows the hit genes (FDR < 0.4 for at least one pathogen) and their direct neighbors that had connections between kinases in STRING database (version 9.0). The edges are functional interactions in the STRING database with edge threshold 850. We removed genes that were not connected to any other gene from the network. Each node consists of a colored pie chart, in which each piece corresponds to a pathogen.

Ambion Silencer Select) performed nearly as well as pooled libraries of less good single siRNAs (in our case Dharmacon SMARTpool). Following the results of the library analysis, we assigned a higher weight to Dharmacon Pooled and Ambion libraries (weight 2) than to the unpooled libraries Dharmacon and Qiagen (weight 1). PMM benefitted from the assigned library weights. The residual standard error of the linear mixed model reduced from 0.87 to 0.83.

Sharedness of detected significant genes

By fitting PMM to our data, we found a left tailed local False Discovery Rate distribution, ending with a set of 50 different genes that reached the threshold of 0.4 (Figure 4B, Figure 5A). We selected threshold 0.4 as a reasonable hit threshold for this study since the difference was small compared to the set of hits with the commonly used threshold 0.2 and 40% false-positive rate was still acceptable in biological follow-up studies for us. The number of up and down hits varied between the pathogens (Figure 4C). Using FDR threshold 0.4, 80% of hits were unique and 20% of hits were shared between two or more studied pathogens (Figure 4D). This provided a rough estimate that about 20% of genes gained statistical power from the parallel analysis using the PMM with our data. To quantify the hits according to their level of being shared between screens independently from the FDR-threshold, we developed the following sharedness score s_g

$$s_g = \frac{1}{2} \left(1 - \text{Mean}_p(q_{pg}) \right) \left(\frac{x}{P} \right)^{\frac{1}{A}}$$

Here P is the total number of pathogens (8 in our case). The sharedness score is a combination of two quantities. The first part defines the shift away from 1 and the second part describes how many pathogens support this shift (proportion of $q_{pg} < 1$). The score returns a value between 0 and 1 for each gene. Score 0 indicates that a gene is not shared among the pathogens and score 1 indicates that the gene is significant among all pathogens (Figure 5A). Since the sharedness score takes only the strength of a gene and not the directionality into account, a gene can be also highly shared if it inhibits in one pathogen and enhances the infection by another pathogen. Therefore, a gene shared between pathogens should be interpreted as being involved in the entry of these pathogens.

Result comparison to existing hit ranking methods

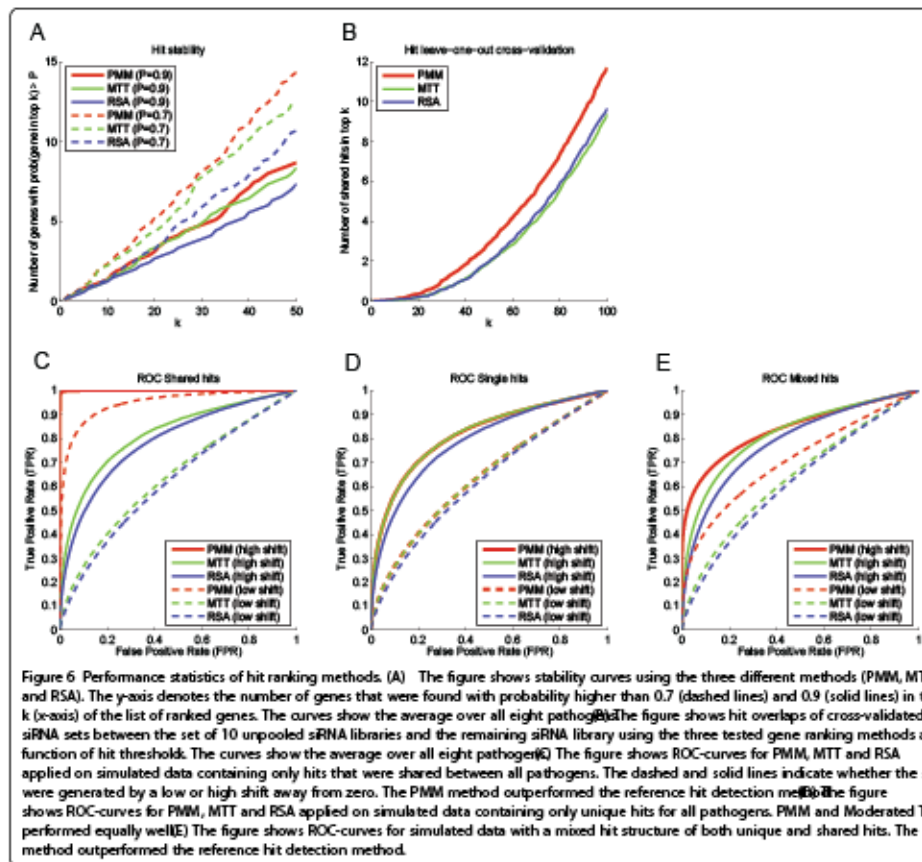
In order to validate the PMM approach and its results we compared it to other existing hit ranking methods and performed different kind of statistical tests. As reference methods we selected the Moderated T-Test (MTT) [36] and Redundant siRNA Activity (RSA) [39] which

are commonly used in high-throughput RNAi screening. We could not apply other widely used hit ranking methods, such as Strictly Standardized Mean Difference (SSMD) [40] or percent inhibition [29] since many of our pathogen screens did not have effective and reliable positive and negative control wells.

As a first test, we analyzed the stability of the obtained gene rankings with respect to the estimated q_{pg} values [30,41]. We resampled with replacement 1000 datasets (12 siRNAs randomly selected with replacement for each gene) and calculated the number of genes that appear with high probability ($\text{prob} > 0.9$ and $\text{prob} > 0.7$) in the top of the ordered lists of genes based on their estimated q_{pg} values (see Methods for details). This measure of stability showed similar results for PMM and the reference methods MTT and RSA (Figure 6A).

To mimic primary and validation screening setup and to study hit reproducibilities of the gene ranking methods we performed a leave-one-out cross-validation experiment. We used the siRNAs of unpooled libraries (11 in total) and left one siRNA set at a time away. We ran PMM, MTT, and RSA on the data sets consisting of 10 individual siRNAs and compared the resulting gene ranking to the ranked gene list of the remaining siRNA set. The averaged hit overlaps over all pathogens as a function of hit threshold k are illustrated in Figure 6B. PMM performed the best indicating that the hits found by PMM are more reproducible by an independent siRNA screen than the hits found by the other methods.

In order to further estimate the hit-calling performance for different methods we performed data simulation with a-priori known hit structure. Data simulation was required since reliable ground truth hits are not generally available for the real biological systems. We simulated data by generating 1000 Gaussian distributed screens for each pathogen with four siRNAs. We selected four siRNAs since it makes up a realistic screening approach. We incorporated hits in each simulated screen by randomly selecting 10% of the genes and shifting them away from zero. We distinguished between three types of simulated data. In the first case the hits were different for each pathogen (unique hits only) and in the second case all hits were shared between the pathogens. The third case is probably the most realistic scenario containing both unique and shared hits to a varying degree (see Methods). We then applied PMM, MTT, and RSA to the simulated data and evaluated them by Receiver Operating Characteristic (ROC)-curves (with false positive versus true positive rates plotted for each FDR-threshold; Figure 6C-E). The results showed that PMM performed the best especially in the case of shared hits. For the case of unique hits PMM and MTT exhibited about the same performance while RSA performed the worst. As expected, with a higher shift of the hit genes the ROC curves got better for all methods.



We also studied how simultaneous modeling affects the ranking of genes in individual screens using PMM. We performed a test where we selected a pathogen and created two datasets. The first dataset was the full data without any changes and the second dataset had the original data for the selected pathogen and randomized data for the 7 other pathogens. We then compared the gene rankings obtained by PMM performed using both datasets for the selected pathogen. The results for *L. monocytogenes* are illustrated in Figure 7A (see Additional file 1: Figure S6 for all the other pathogens). The correlation graph shows that the addition of parallel screens had only a mild effect on the overall gene ranking. However, when considering the number of significant genes (FDR < 0.4), PMM mainly added genes to the list of significant genes (7 novel

significant genes for *L. monocytogenes* and only few genes (1 for *L. monocytogenes*) were dropped off the list. In general, we concluded that using parallelism added novel significant genes while losing almost none. Moreover, the few lost hit genes had high FDR values, just slightly below the selected threshold FDR < 0.4.

In the next step we analyzed the differences between the resulting gene rankings of the tested methods. Differences in gene rankings between PMM and other hit ranking methods were not very strong (see Figure 7B for MTT compared to PMM for *L. monocytogenes* and Additional file 1: Figure S7 for all other cases). Genes that had a high sharedness score and had an effect on the screen of interest (in particular MTOR and TGFBR1/2 for *L. monocytogenes*) gained statistical power from the simultaneous analysis and

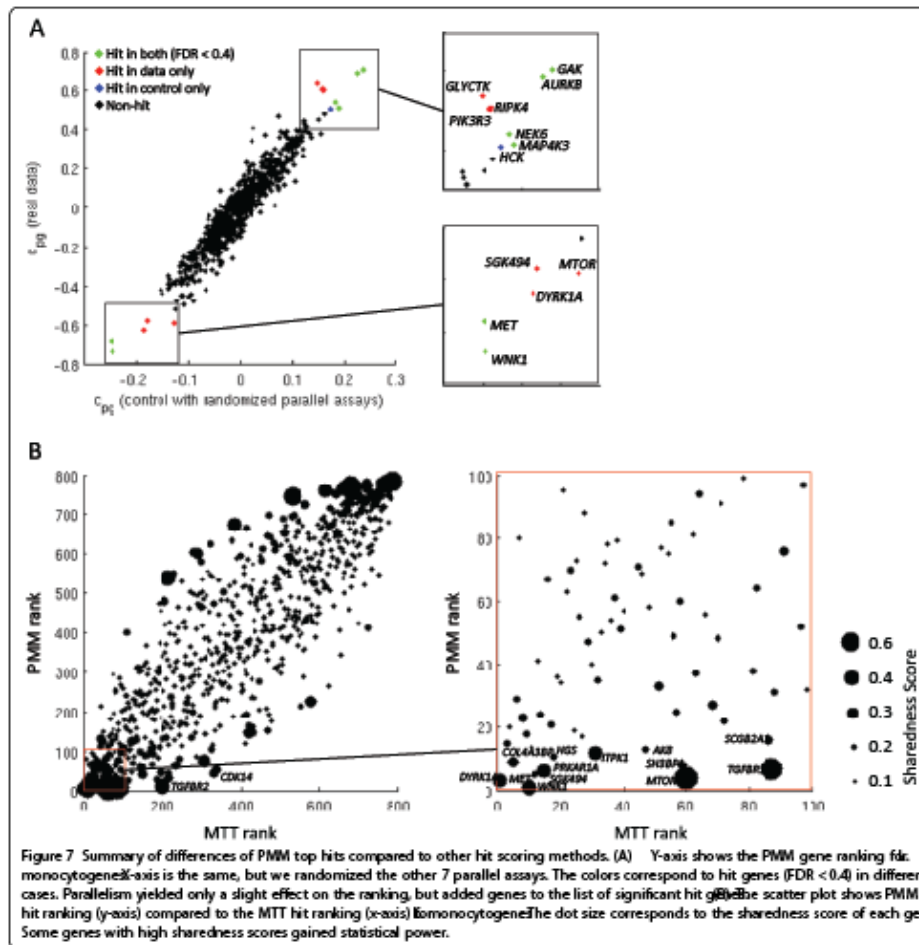


Figure 7 Summary of differences of PMM top hits compared to other hit scoring methods. (A) Y-axis shows the PMM gene ranking for monocyogenesis. X-axis is the same, but we randomized the other 7 parallel assays. The colors correspond to hit genes (FDR < 0.4) in different cases. Parallelism yielded only a slight effect on the ranking, but added genes to the list of significant hit genes. (B) Scatter plot shows PMM hit ranking (y-axis) compared to the MTT hit ranking (x-axis) for monocyogenesis. The dot size corresponds to the sharedness score of each gene. Some genes with high sharedness scores gained statistical power.

were pushed up in the gene ranking. On the other hand, we observed that PMM detected several genes with low sharedness scores, indicating that unique hits were not neglected.

In order to evaluate the biological relevance of observed hits, we calculated pathway enrichment scores separately for each pathogen by the Gene Set Enrichment Analysis (GSEA) algorithm [42] using as input the results from the three hit ranking algorithms PMM, MTT, and RSA (see Additional file 1). We selected all pathways that were significant (GSEA pathway enrichment FDR score < 0.2) for at least one pathogen and

method pair. We used the ranking of infection indices as the input for GSEA and focused on hits that reduce infection levels. By assuming that most pathways in the used database are biologically valid, we would expect that better hit detection methods give a higher number of enriched pathways than less powerful hit detection methods. However, we only screened kinases and the applicable pathways are limited to those that are highly enriched in phosphorylation events and it may be that some pathogens do not show strong enrichments within this set of pathways. Moreover, differences in pathway

enrichments between methods may have occurred because they treated missing values differently. Therefore, the enrichment results should be evaluated with caution. Additional file 1: Figure S9 illustrates the observed significant pathways. The number of enriched (GSEA FDR < 0.2) pathways for each method was an indication that PMM detected biologically more relevant hit genes than the other methods.

Biological inquiry on detected significant genes

The performed screens yield several interesting hits of which most are novel to the corresponding pathogen (Figure 5A, see Additional file 1: Figure S10 for cell number hits). Many of the strongest hits, including MTOR, TGFBR1/2 for negative hits and ILK for positive hits, were shared between most of the studied pathogens. This was also illustrated by the sharedness scores of detected hit genes. Many of the strongest shared hits were related to SRC, MTOR, or CDK related pathways. Although SRC and CDK4 were not part of the hit lists ($q_{adj} < 0.4$) for any of the pathogens, they exhibited consistent semi-strong effect for most pathogens. A network analysis of hit genes showed that several of the shared hits can be described as "network hubs" that are involved in many cellular processes and highly connected to other genes (including MTOR and SRC) (Figure 5B) [43]. MTOR is a mammalian target of rapamycin, serine/threonine protein kinase that regulates cell growth, cell proliferation, cell motility, cell survival, protein synthesis, and transcription. The involvement of MTOR in Adenovirus, Poliovirus, Enterovirus 71, Coxsackievirus, Vaccinia virus and other pathogens has already been established [6,9,15]. Our data also reproduced the established role of MTOR during *S. typhimurium* infections, since *S. typhimurium* depends on a reactivation of MTOR during its course of infection in order to escape autophagy [38]. Interestingly, TGFBR1 and TGFBR2 came up both as strong hits for many pathogens. TGFBR1 and TGFBR2 proteins must heterodimerize to form a functional TGF-beta receptor at the plasma membrane. Their similar strong infection reducing knock-down phenotypes, seen in most independent pathogen screens, indicated the validity of these hits and suggested a broad, yet poorly understood, function of this membrane protein for various pathogens. In particular, there are suggestions [44] that the TGF-beta pathway might be important for *B. abortus* infection since in chronic brucellosis patients there is increased TGF-beta production and this could aid infection by depressing lymphocyte functions. In addition, our study confirmed the role of DYRK family members (in particular DYRK1A) as they have been identified to be general regulators for several viruses in Snijder, Sacher et al. [9].

Despite the overall similarity of infection patterns between pathogens, most pathogens also contained hits that

were specific for the pathogen (for example MET for *L. monocytogenes*, NTPCR for *B. henselae* and ETK1 and ULK1 for Rhinovirus). Some of the hit genes have previously been found to be effectors, for example MET for *L. monocytogenes*. *L. monocytogenes* enters host cells by triggering signaling cascades activated through interaction of bacterial internalin A (InIA) or InIB with the adherens junction protein E-cadherin or the hepatocyte growth factor receptor MET [37] respectively. Since E-cadherin is not expressed in HeLa cells, which were used for our siRNA screens, the INLB / MET pathway is the only route of entry in this cellular system. In fact, MET [45] was one of the strongest hits for *L. monocytogenes*. The exact roles of most hit genes of all pathogens are largely unknown, but several hit genes create interesting hypotheses for follow-up. For example, it was proposed based on micro-RNA analysis of infected macrophages, that AMPK might be a target gene that promotes intracellular survival during *B. abortus* infection [46]. PIK3R3 (p55-gamma; Phosphatidylinositol 3-Kinase 55 KD a Regulatory Subunit Gamma) a semi-strong hit for several pathogens in our data was identified as a hit in an RNAi screen of *Drosophila* S2 cells, in agreement with the importance of PI3K during *B. abortus* infection [47]. PIK3CA probably plays a role also in *B. henselae* infection through actin modulation. PIK3CA levels influence RHOA and RAC1, which are involved in actin dynamics [48]. Furthermore, PIK3CA is involved in PIP3 production, which is a signaling molecule and has recently been shown to be related to the formation of dynamic F-actin-related structures [49]. ULK1 (unc-51 like autophagy activating kinase 1) plays an important role in autophagy as well as Hepatitis C virus infection. Therefore, ULK1 has a possible link to Rhinovirus induced autophagy. COL4A3BP is possibly linked to Rhinovirus entry through ceramide-enriched membrane platforms [50] since COL4A3BP specifically phosphorylates the N-terminal region of the non-collagenous domain of the alpha 3 chain of type IV collagen, known as the Goodpasture antigen, also involved in ceramide intracellular transport (from ER to Golgi).

Conclusions

We produced a uniquely wide high-content siRNA dataset, in terms of used siRNA libraries (11 single siRNAs and one pool) and eight different pathogens. Our highly unified protocols and common image analysis as well as similar data analysis pipelines enabled a direct comparison between the phenotypic readouts of the different pathogen screens. The unified structures of the datasets also aided discovering shared mechanisms between the studied pathogens.

Using our novel statistical approach PMM we detected several interesting and new hits from our kinome-wide pathogen screens. The hits will require further follow-up

work in order to understand the exact biological mechanisms of the genes. In addition, we discovered shared effector genes between the studied pathogens including MTOR, TGFBR1 and TGFBR2 that were strong hits for almost all studied pathogens. In particular, the obtained sharedness scores indicated whether a hit gene has a very specific function for a single pathogen or a more generic cellular function that is shared between many pathogens and thus gave us the first indications of the genes' roles. Pharmaceutically oriented follow-up studies could take advantage of this concept. For example, if we were interested in general regulators we could focus on genes with high sharedness scores. On the other hand, regulators that have a very specific effect and a low sharedness score could probably have fewer side effects.

We showed that the reliability of hit scoring in individual RNAi screens improved by using PMM that takes advantages of the parallelism in RNAi screening. PMM can, in principle, be applied to any kind of parallel RNAi screens almost independently of the underlying biology or field of application as long as the readouts of the screens are measured on the same scale. We can often obtain this by applying Z-Scoring or similar normalization methods to the well readouts. The difference to other approaches aiming at the comparison of independent parallel RNAi screens is that PMM takes simultaneously all screening data into account. For example, for the comparison of insect and human data in [51] the hit lists were derived by separate statistics on each screen. By taking all data into the analysis the statistical power can be increased. Based on our results, we expect that the more similar the parallel screens are in the sense of biological focus or protocols, the more statistical power can be gained from the simultaneous analysis. Even a slight overlap between the underlying biological pathways of the parallel screens can improve the hit detection in individual screens without compromising the detection of unique hits for any individual screens. Provided that the large-scale RNAi screening community reaches standardized data publication and sharing standards through projects such as MIARE and GenomeRNAi, the PMM approach could be expanded to include the vast number of different RNAi screens performed in different laboratories worldwide that used the same siRNA libraries. In principle and as a vision, this opens up great opportunities for simultaneous statistical approaches such as PMM. Every new screen could potentially gain statistical power by using the public resources. In addition, PMM can potentially be used to gain power for secondary validation screens. Such validation screens are typically performed with several independent siRNAs targeting the same gene under various conditions and PMM would be directly applicable. A beneficial feature of PMM is the possibility to assign weights to the siRNAs. The weights can incorporate a-priori information about

the performance of individual siRNAs and their phenotypical readout. This concept of weighting can be expanded over what we presented in this paper. In particular, statistical and bioinformatics analyses on seed sequence induced off-target effects could potentially be used as basis for weights. Naturally any additional high-throughput data, such as proteomics analyses on cells under siRNA perturbations, or genomic analyses on specific cell lines, could be used to assign realistic siRNA weights to improve hit scoring.

We aimed to take a step forward in determining minimal requirements for image-based RNAi screening data publication. All the raw images, library metadata, single cell measurements, and well measurements are publicly available through our openBIS based publication portal. In addition, we provide easy-to-access data aggregates in standardized tabular formats with all the necessary metadata information. Our uniquely wide datasets provide a large resource for infection biologists, image analysts, and statisticians for future research.

Methods

Wet-lab protocols

Cell culturing conditions

HeLa CCL-2 (ATCC) cells were maintained at 37°C and 5% CO₂ in Dulbecco Modified Eagle Medium (DMEM, Invitrogen) supplemented with 10% inactivated FCS (Invitrogen).

siRNA reverse transfection

RNA interference directed against human kinases and kinase-associated genes (826 genes in total) was achieved using commercially available siRNA libraries. All experiments were conducted in a 384-well plate format. In addition to screening plates, control plates were included in each screen. All plates contained general siRNA controls for transfection efficiency and toxicity (e.g. Kif11), as well as, control siRNAs for infection effects of each pathogen assayed. However, for most of the pathogens in this study, reliable and well established positive control siRNAs (reducing or enhancing infection levels) were not available prior to screening. In addition, negative controls such as MOCK (no siRNA) and SCRAMBLED (non-targeting siRNA) were added to every plate.

In each experiment, 25 µl of RNAiMAX/DMEM (0.1 µl/24.9 µl) mixture was added to each well of the screening plates containing 1.6 pmol siRNA diluted in 5 µl RNase-free ddH₂O. Screening plates were thereafter incubated at room temperature (RT) for 1 h. Following incubation, a pathogen assay-specific number of HeLa CCL-2 cells (see Additional file 1: Table S1) were added per well in a volume of 50 µl DMEM/16% FCS, resulting in a final FCS concentration of 10%. Adenovirus screens contained 6.7%

final FCS). Plates were incubated at 37°C and 5% CO₂ for 72 h prior to infection.

Fixation and staining

After infection cells were fixed using paraformaldehyde (PFA). Cells were stained for DNA, F-actin and infection specific markers. Screening plates were sealed prior to imaging.

Adenovirus-specific protocol

All liquid handling stages of infection, fixation, and immunofluorescence staining were performed on the automated pipetting system Well Mate (Thermo Scientific Matrix) and washer Hydrospeed (Tecan). For infection screens recombinant Ad₂ΔE3B-eGFP (short Adenovirus) was utilized as described before [52,53]. Adenovirus was added to cells at a multiplicity of infection (moi) of 0.1 in 10 μl of an infection media/FBS (DMEM supplemented with L-glutamine, 10% FBS, 1% Pen/Strep, Invitrogen). Screening plates were incubated at 37°C for 16 h, and cells were fixed by adding 21 μl of 16% PFA directly to the cells in culture media for 45 min at RT or long-term storage at 4°C. Cells were washed 2 times with PBS/25 mM NH₄Cl, permeabilized with 25 μl 0.1% Triton X-100 (Pharmacie-biothek). After 2 washes with PBS the samples were incubated at RT for 1 h with 25 μl staining solution (PBS) containing DAPI (1 μg/ml, Sigma-Aldrich) and DY-647-phalloidin (1 U/ml, Dyomics), washed 2 times with PBS and stored until imaging in 50 μl PBS/Na₃.

Bartonella henselae -specific protocol

Bacterial strain SEB0109: Bartonella henselae ATCC49882^T ΔbepG containing plasmid pCD353 [54] for IPTG-inducible expression of GFP. Culturing conditions: bacteria were grown on Columbia base agar (CBA) plates supplemented with 5% defibrinated sheep blood (Oxoid) and 50 μg/ml kanamycin. Bacteria were incubated at 35°C in 5% CO₂ for 72 h before re-streaking them on fresh CBA and further growth for 48 h. Infection: siRNA-transfected cells were washed once with M199 (Invitrogen)/10% FCS using a plate washer (ELx50-16, BioTek). Cells were infected with B. henselae at an MOI of 400 in 50 μl of M199/10% FCS and 0.5 mM IPTG (Applchem) and were incubated at 35°C in 5% CO₂ for 30 h. Fixation at RT: using a Multidrop 384 (Thermo Scientific) cells were washed with 50 μl of PBS, fixed in 20 μl of 3.7% PFA for 10 min, and washed once more with 50 μl of PBS. Staining on a Biomek liquid handling platform: fixed cells were washed twice with 25 μl of PBS and blocked in PBS/0.2% BSA for 10 min. Extracellular bacteria were labeled with a rabbit serum 2037 against B. henselae [23] and a secondary antibody goat anti rabbit Alexa Fluor 647 (Jackson Immuno) in PBS/0.2% BSA. Antibodies were incubated for 30 min each and both incubations were

followed by two washings with 25 μl of PBS. Cells were then permeabilized with 20 μl of 0.1% Triton X-100 (Sigma) for 10 min and afterwards washed twice with 25 μl of PBS, followed by the addition of 20 μl of staining solution (PBS containing 1.5 U/ml DY-547-Phalloidin (Dyomics) and 1 μg/ml DAPI (Roche)). After 30 min of incubation in the staining solution, cells were washed twice with 25 μl PBS, followed by a final addition of 50 μl of PBS.

Brucella abortus -specific protocol

Brucella abortus 2308 pJC43 (aphT::GFP) [55] were grown in tryptic soy broth (TSB) medium containing 50 μg/ml kanamycin for 20 h at 37°C and shaking (100 rpm) to an OD of 0.8-1.1. 50 μl of DMEM/10% containing bacteria was added per well to obtain a final moi of 10000 using a cell plate washer (ELx50-16, BioTek). Plates were then centrifuged at 400 g for 20 min at 4°C to synchronize bacterial entry. After 4 h incubation at 37°C and 5% CO₂, extracellular bacteria were killed by exchanging the infection medium by 50 μl medium supplemented with 10% FCS and 100 μg/ml gentamicin (Sigma). After a total infection time of 44 h cells were fixed with 3.7% PFA for 20 min at RT with the cell plate washer. Staining was performed using a Biomek liquid handling platform. Cells were washed twice with PBS and permeabilized with 0.1% Triton X (Sigma) for 10 min. Then, cells were washed twice with PBS, followed by addition of 20 μl of staining solution which includes DAPI (1 μg/ml, Roche) and DY-547-phalloidin (1.5 U/ml, Dyomics) in 0.5% BSA in PBS. Cells were incubated with staining solution for 30 min at RT, washed twice with PBS, followed by final addition of 50 μl PBS.

Listeria monocytogenes -specific protocol

After washing an overnight culture of L. monocytogenes EGDe.Prfa*GFP three times with PBS, bacteria were diluted in DMEM supplemented with 1% FCS. Cells were infected at a moi of 25 in 30 μl infection medium per well. After centrifugation at 1000 rpm for 5 min and incubation for 1 h at 37°C in 5% CO₂ to allow the bacteria to enter, extracellular bacteria were killed by exchanging the infection medium by 30 μl DMEM supplemented with 10% FCS and 40 μg/ml gentamicin (Gibco). Both medium exchange steps were carried out with a plate washer (ELx50-16, BioTek). After additional 4 h at 37°C in a 5% CO₂ atmosphere, cells were fixed for 15 min at RT by adding 30 μl of 8% PFA in PBS to each well using a multidrop 384 device (Thermo Electron Corporation). PFA was removed by four washes with 500 μl PBS per well using the Power Washer 384 (Tecan). Fixed cells were stained for nuclei, actin and bacterially secreted nLC. First, cells were incubated for 30 min with 10 μl/well of primary staining solution (0.2% saponin, PBS) containing

rabbit derived anti-InlC serum (1:250). After four washes with 40 μ l PBS per well cells were stained with 10 μ l/well of the secondary staining solution (0.2% saponin, PBS) containing Alexa Fluor-546 coupled anti-rabbit antibody (1:250, Invitrogen), DAPI (0.7 μ g/ml, Roche), and DY-647-Phalloidin (2 U/ml, Dyomics). After four washes with 40 μ l PBS per well, the cells were kept in 40 μ l PBS per well. The staining procedure was carried out with a Tecan freedom evo robot.

Rhinovirus-specific protocol

All liquid handling stages of infection, fixation, and immunofluorescence staining were performed on the automated pipetting system Well Mate (Thermo Scientific Matrix) and washer Hydrospeed (Tecan). For infection assays with human Rhinovirus serotype 1a (HRV1a) were carried out as described, except that the anti-VP2 antibody Mab 16/7 was used for staining of the infected cells as described earlier [56-58]. Rhinovirus at a moi of 8 was added to cells in 20 μ l of an infection media/BSA (DMEM supplemented with GlutaMAX, 30 mM MgCl₂ and 0.2% BSA, Invitrogen). Screening plates were incubated for 7 h at 37°C, and cells were fixed by adding 33 μ l of 16% PFA directly to the culture medium. Fixation was either for 30 min at RT or long term storage at 4°C. Cells were washed twice with PBS/25 mM NH₄Cl, permeabilized with 50 μ l 0.2% Triton X-100 (Sigma-Aldrich) followed by 3 PBS washes and blocking with PBS containing 1% BSA (Fraction V, Sigma-Aldrich). Fixed and permeabilized cells were incubated at RT for 1 h with diluted mabR16-7 antibody (0.45 μ g/ml) in PBS/1% BSA. Cells were washed 3 times with PBS and incubated with 25 μ l secondary staining solution (PBS/1% BSA) containing Alexa Fluor 488 secondary antibody (1 μ g/ml, Invitrogen), DAPI (1 μ g/ml, Sigma-Aldrich), and DY-647-phalloidin (0.2 U/ml, Dyomics). Cells were washed twice with PBS after 2 h of incubation in secondary staining solution and stored in 50 μ l PBS/NaN₃.

Salmonella typhimurium-specific protocol

All liquid handling stages of infection, fixation, and immunofluorescence staining were performed on a liquid handling robot (BioTek; EL406). For infection the *S. typhimurium* strain S.Tm^{SopE-pM975} was used. This strain is a single effector strain, only expressing SopE out of the main four SPI-1 encoded effectors (SipA, SopB, SopE2 and SopE). Additionally this strain harbors a plasmid (pM975) that expresses GFP under the control of a SPI2 (*ssaG*)-dependent promoter. The bacterial solution was prepared by cultivating a 12 h culture in 0.3 M LB medium containing 50 μ g/ml streptomycin and 50 μ g/ml ampicillin. Afterwards a 4 h subculture (1:20 diluted from the 12 h culture) was cultivated in 0.3 M LB medium containing 50 μ g/ml streptomycin, which

reached an OD_{600nm} = 1.0 after the respective 4 h of incubation time. To perform the infection, 16 μ l of diluted *S. typhimurium* (moi = 80) were added to the HeLa cells. After 20 min of incubation at 37°C and 5% CO₂, the *S. typhimurium*-containing media was replaced by 60 μ l DMEM/10% FCS containing 50 μ g/ μ l streptomycin and 400 μ g/ μ l gentamicin to kill all remaining extracellular bacteria. After additional 3 h 40 min incubation at 37°C and 5% CO₂ cells were fixed by adding 35 μ l 4% PFA, 4% sucrose in PBS for 20 min at RT. The fixation solution was removed by adding 60 μ l PBS containing 400 μ g/ml gentamicin. Cells were permeabilized for 5 min with 40 μ l 0.1% Triton X-100 (Sigma-Aldrich). Afterwards 24 μ l of staining solution containing DAPI (1:1000, Sigma-Aldrich) and DY-547-phalloidin (1.2 U/ml, Dyomics) was added (prepared in blocking buffer consisting of 4% BSA and 4% Sucrose in PBS). After 1 h of incubation at RT, cells were washed three times with PBS followed by the addition of 60 μ l PBS containing 400 μ g/ml gentamicin.

Shigella flexneri-specific protocol

S. flexneri M90T Δ virG pCK100 (PuhPT:dsRed) were harvested in exponential growth phase and coated with 0.005% poly-L-lysine (Sigma-Aldrich). Afterwards, bacteria were washed with PBS and resuspended in assay medium (DMEM, 2 mM L-Glutamine, 10 mM HEPES). 20 μ l of bacterial suspension was added to each well with a final moi of 15. Plates were then centrifuged for 1 min at 37°C and incubated at 37°C and 5% CO₂. After 30 min of infection, 75 μ l were aspirated from each well and monensin (Sigma) and gentamicin (Gibco) were added to a final concentration of 66.7 μ M and 66.7 μ g/ml, respectively. After a total infection time of 3.5 h, cells were fixed in 4% PFA for 10 min. Liquid handling was performed using the Multidrop 384 (Thermo Scientific) for dispensation steps and a plate washer (ELx50-16, BioTek) for aspiration steps. For immunofluorescent staining, cells were washed with PBS using the Power Washer 384 (Tecan). Subsequently, cells were incubated with a mouse anti-human IL-8 antibody (1:300, BD Biosciences) in staining solution (0.2% saponin in PBS) for 2 h at RT. After washing the cells with PBS, Hoechst (5 μ g/ml, Invitrogen), DY-495-phalloidin (1.2 U/ml, Dyomics) and Alexa Fluor 647-coupled goat anti-mouse IgG (1:400, Invitrogen) were added and incubated for 1 h at RT. The staining procedure was performed using the Biomek NXP Laboratory Automation Workstation (Beckman Coulter).

Vaccinia virus-specific protocol

All liquid handling stages of infection, fixation, and immunofluorescence staining were performed on a liquid handling robot (BioTek, EL406). For infection assays a recombinant WR VACV, WR E EGFP/L mCherry, was utilized. For infection, media was aspirated from the

RNAi-transfected cell plates and replaced with 40 μ l of virus solution per well (moi = 0.125). Screening plates were incubated for 1 h at 37°C to allow for infection, after which virus-containing media was removed and replaced with 40 μ l DMEM/10% FCS. 8 h after infection 40 μ l of DMEM/10%FCS containing 20 μ M cytosine arabinoside (AraC) was added to all wells to prevent virus DNA replication in secondary infected cells. 24 h after infection cells were fixed by the addition of 20 μ l 18% PFA for 30 min followed by two PBS washes of 80 μ l. For immunofluorescence staining of EGFP, cells were incubated for 2 h in 30 μ l primary staining solution (0.5% Triton X-100, 0.5% BSA, PBS) per well, containing anti-GFP antibody (1:1000). Cells were washed twice in 80 μ l PBS, followed by the addition of 30 μ l secondary staining solution (0.5% BSA, PBS) containing Alexa Fluor 488 secondary antibody (1:1000), Hoechst (1:10000), and DY-647-phalloidin (1:1200, Dyomics). Cells were washed twice with 80 μ l PBS after 1 h incubation in secondary staining solution followed by the addition of 80 μ l H₂O.

Microscopy

Microscopy was performed with Molecular Devices ImageXpress microscopes. We used the MetaXpress plate acquisition wizard with no gain, 12 bit dynamic range, 9 sites per well in a 3x3 grid with no spacing and no overlap and laser-based focusing. Channels were assay specific (see Additional file 1: Table S2). Robotic plate handling was used to load and unload plates (Thermo Scientific). The objective was a 10X S Fluor with 0.45NA. The Site Autofocus was set to "All Sites" and the initial well for finding the sample was set to "First well acquired". Z-Offset for Focus was selected manually and "AutoExpose" was used to get a good exposure time. Manual correction of the exposure time was applied to ensure a wide dynamic range with low overexposure, when necessary.

Statistical analyses

Image analysis and data normalization

Image analysis and data normalization was based on modified CellProfiler [28] workflows. Please refer to Additional file 1 for detailed description of computational infrastructure, image analysis, and data normalization.

Parallel Mixed Model (PMM)

We denote the readout of siRNA *s* silencing gene *g* for a pathogen *p* as y_{pgs} . The linear mixed model of PMM is defined as the following linear model

$$y_{pgs} = \mu_p + a_g + b_{pg} + \epsilon_{pgs}$$

$$a_g \sim N(0, \sigma_a^2); b_{pg} \sim N(0, \sigma_b^2); \epsilon_{pgs} \sim N(0, \sigma_\epsilon^2);$$

where μ_p is the fixed effect for pathogen *p* (typically close to 0 because of data Z-scoring) a_g is the gene effect

overall pathogen, b_{pg} is the gene effect within a pathogen and ϵ_{pgs} denotes the error term. The parameters are estimated by maximizing the restricted maximum likelihood using the Newton-Raphson algorithm [33]. We used the implemented version in the "lmer" function from the "lme4" R-package [34]. This implementation allows also the use of weights, which are incorporated by a weighted maximum likelihood formulation. The weights are constant values where each constant corresponds to exactly one data point. For our data, each weight is associated with a single readout of an independent siRNA. The size of the weight indicates the precision of the information contained in the associated readout. The assumptions of the linear mixed model are fulfilled (see Additional file 1: Figure S11).

Local false discovery rate (q) estimation in PMM

The observed distribution of the estimated \hat{c}_{pg} is a mixture of the null f_0 and the non-null distribution f_1 . The null distribution describes the distribution of all genes that are no-hits. The non-null distribution corresponds to the genes that are hits, having either a positive or negative effect. The two distributions are assumed to differ only in the mean. The non-null distribution is shifted by a factor θ away from zero. With this we define the local false discovery rate as

$$f_0 = \frac{\pi_0 f_0(\hat{c}_{pg})}{\pi_0 f_0(\hat{c}_{pg}) + \pi_1 f_1(\hat{c}_{pg})}$$

$$f_0 \sim N(0; \sigma_0^2); f_1 \sim N(\theta; \sigma_1^2)$$

where π_0 = proportion of true hits and $\pi_1 = 1 - \pi_0$ [59]. The three quantities needed for the estimation of the false discovery rate, are estimated separately by using Maximum Likelihood, Poisson regression, and moment estimation. The estimation procedure is implemented in the function "locfdr" from the "locfdr" R-package [35].

Data resampling to show that parallel screens add power

We chose gene *g* and pathogen *p* for which we wanted to show the increase in power by simultaneously using the parallel screening structure. In our case, we repeated the analysis for three different cases, consisting of a unique hit (g: MET, p: L. monocytogenes); a shared hit (g: MTOR, p: Vaccinia virus) and a non-hit (g: ALK, p: B. abortus). Each time we resampled data for a fixed number of siRNAs ($n_s = 2, \dots, 11$) and a fixed number of pathogens ($n_p = 2, \dots, 8$) from the full dataset. In detail, we chose randomly $(n_p - 1)$ pathogens and added additionally pathogen *p*. In the next step, we sampled n_s siRNA sets from the full available set of siRNAs for every gene within all sampled pathogens. We applied PMM on the sampled data and we reported the rank of gene *g* within pathogen *p*. This was repeated

1000 times for each combination of n_p and n_g . As a last step we calculated for each combination the mean and variance of the rank for genes within pathogen p . For the resampling we omitted genes that have less than 6 siRNA sets, in order to have a good resampling basis. Moreover, we applied the same procedure for the case of $n_p = 1$ using MTT.

Stability analysis

We resampled with replacement 1000 datasets from the full data, taking for each gene the same number of siRNAs as in the full dataset. For each resampled dataset, PMM, MTT and RSA were applied and the corresponding ranking saved. For PMM the ranking was done according to the absolute value of the estimated c_{pg} effects, for MTT we used the absolute values of the estimated mean and for RSA the ranking based on the $\log(p)$ values. We took absolute values to take into account down and up hits simultaneously. From the 1000 rankings we calculated the number of genes that appear with high probability ($\text{prob} > 0.9$ and $\text{prob} > 0.7$ in the top k ($k = 1, \dots, 50$) of the ranking.

Hit overlaps examined by cross-validation

For the hit cross-validation analysis we only used data coming from the siRNAs of all unpooled libraries (11 in total). In each run, we ran PMM, MTT, and RSA on a subset of the data consisting of 10 individual siRNAs and used the remaining siRNA set as test set. For PMM we ranked the results according to the absolute value of the estimated c_{pg} effects, for MTT we did ranking with respect to the absolute values of the estimated mean, for RSA we based the ranking based on the $\log(p)$ values and for the test set we ordered the genes by the absolute value of infection score. We counted the number of genes that appeared in top k ($k = 1, \dots, 100$) in both the training and test sets. We determined the counts separately for each pathogen and averaged them in the end.

Data simulation and ROC-curves

We simulated data by generating 1000 normally distributed screens (mean = 0, std = 0.5) for eight pathogens, taking 4 siRNAs each. Hits were incorporated in the simulated screens by randomly selecting about 10% of the genes (80 out of 826) and shifting them away from zero. The shift was determined by a uniformly distributed random variable. We used the interval [0.2,0.3] as parameter for the uniform distribution for "low shift" and the interval [0.4,0.5] for "high shift". We distinguished between three cases: In the first case the hits were different for each pathogen (80 unique hits per pathogen), in the second case all hits were shared between the pathogens (same 80 hits for all pathogens) and in the third case we generated mixed hits (20 unique

hits, 20 hits shared between two pathogens, 20 hits shared between four pathogens and 20 hits shared between all eight pathogens). PMM, MTT, and RSA were applied to the simulated data and the ranking was saved. For PMM the results were ranked according to the absolute value of the estimated c_{pg} effects, for MTT the ranking was done with respect to the absolute values of the estimated mean and for RSA the ranking based on the $\log(p)$ values. For every ranking list we counted the number of true positives, true negatives, false positives and false negatives in the top k ($k = 1, \dots, 826$) and computed the true positive rate ($\text{TPR} = \text{FP}/(\text{FP} + \text{TN})$) and the false positive rate ($\text{FPR} = \text{FP}/(\text{FP} + \text{TN})$).

Influence of parallelism

For selected pathogen p we generated 1000 new datasets by fixing the data of p and randomizing the data of the other 7 pathogens. We applied PMM to each dataset and saved the resulting ranking of p . In the next step we aggregated the 1000 rankings by taking the average over the c_{pg} scores. We compared the averaged scores to the gene rankings obtained by PMM performed using the original dataset. We independently performed the study for each pathogen.

Availability of supporting data

The data sets supporting the results of this article are available on the InfectX openBIS data publication portal, that is located at <http://www.infectx.ch/dataaccess/>. The visitor username is "rdgr2014" and the corresponding password is "IXPubReview". The R-package PMM and related documentation is also available on this page.

Additional file

Additional file 1: Supplementary Information. The additional data file 1 contains supporting information and further analysis results.

Abbreviations

FDR: False discovery rate; GSEA: Gene set enrichment analysis; MTT: Moderated T-test; PMM: Parallel mixed model; ROC: Receiver operating characteristic; RSA: Redundant siRNA Activity.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

We PR, AD, CA, NB, PC, ME, UG, W-DH, AH, CvM, LP, and CD planned experiments and analysis pipeline. HB-T, BC, RC, SE, CK, SK, AK, SHL, JM, DM, SM, JP-C, MT, and AY devised biological assays. GC, AK, SM, MS and AV plated siRNA libraries. HB-T, AC, RC, SE, CK, SK, AK, SHL, JM, DM, and SM performed siRNA screening. PR, ME, PK, EP, MP, BR, VR, BS, and LP contributed to the establishment of the computational analysis pipeline. ME, CK, MT, and AV planned image acquisition. ME with support of GS and EP performed image analysis and data management. PR with support of ME, NB, AD, FS, JS, ES, and CvM normalized image-derived data. PR and AD with support of NB, PB, FS, JS, and ES performed statistical analysis. AD and PB developed the PMM method. The manuscript was written by PR, AD, and CD. All authors read and approved the final manuscript.

Acknowledgements

We acknowledge support by grants 51RT0_126008 and 51RTP0_151029 for the Research and Technology Development (RTD) projects InfectX and TargetInfectX, respectively, in the frame of SystemsX.ch, the Swiss Initiative for Systems Biology. C.D. acknowledges grant 31003A-132079 from the Swiss National Science Foundation (SNSF) and ERC advanced grant 340330. P.C. would like to acknowledge the following grants: ERC advanced grant 233348, ANR grant ME-2009-SignRupVac, Fondation La Roch and Fondation Jeantat. A.K. is a recipient of a scholarship from the Pasteur-Paris University International Doctoral Program/Institut Carnot 10. Maladies Infectieuses. E.S. was supported by the ETH Zurich Postdoctoral Fellowship Program and the Marie Curie Actions for People COFUND program (grant No. FEL-13 121) (to E.S. and M.B.). F.S. has been supported by SystemsX.ch, the Swiss initiative in systems biology, under IPHD grant No. 2009/025 (to M.B.). Work of S.H.L. and A.C. was supported by the International PhD Program Fellowships for Excellence of the Biozentrum. The work on cell and virus and Rhinovirus screening infrastructure was supported by grant SystemsX iPhD, grant Novartis: Stiftung für Medizinisch-Biologische Forschung and University of Zurich: Bauten und Investitionen. Work in the Graber lab was funded by a grant from the SNSF (31003A_141222/1) and an iPhD fellowship to partly support A.Y. (SXP.HIO_142001/1). We acknowledge support from Dr. A. Jurgé in the development of the rhinovirus infection assay. We thank Dr. Jean Pierre Gorvel (Centre d'immunologie de Marseille-Luminy, Marseille, France) for kindly providing B. abortus GFP strain. We wish to thank Rainer Pöhlmann (Biozentrum, University of Basel, Switzerland) and Konstantin Arnold (Biozentrum, University of Basel and Swiss Institute of Bioinformatics), as well as the [BC]2 Basel Computational Biology Center for support and provision of high-performance computing resources. We acknowledge openBIS / CISD and Ramakrishnan Chandrasekhar, who implemented many extensions and modifications in screening openBIS.

Author details

¹Focal Area Infection Biology, Biozentrum, University of Basel, Klingelstrasse 70, CH-4056 Basel, Switzerland; ²Berlin Institute for Statistics, ETH Zurich, Zurich, Switzerland; ³Institute of Molecular Life Sciences, University of Zurich, Zurich, Switzerland; ⁴Department of Biosystems Science and Engineering, ETH Zurich, Zurich, Switzerland; ⁵Swiss Institute of Bioinformatics, Basel, Switzerland; ⁶Unité des Interactions Bactéries Cellules; INSERM, U604; INRA, USC2020, Paris, France; ⁷Institute for Tropical Health and Departamento de Microbiología y Parasitología, Universidad de Navarra, Pamplona, Spain; ⁸Institute of Biochemistry, ETH Zurich, Zurich, Switzerland; ⁹Institute of Molecular Biology, Mainz, Germany; ¹⁰Light Microscopy and Screening Center, ETH Zurich, Zurich, Switzerland; ¹¹Research IT, Biozentrum, University of Basel, Basel, Switzerland; ¹²Institut Cochin, INSERM U1016, CNRS 8104, Université Paris Descartes, Paris, France; ¹³SyBIT, SystemsX.ch, Zurich, Switzerland.

Received: 11 July 2014 Accepted: 12 December 2014
Published: 22 December 2014

References

- Conrad C, Gerlich DW: Automated microscopy for high-content RNAi screening. *J Cell Biol* 2010, 188(4):453-461.
- Mohr S, Bakal C, Patmon N: Genomic screening with RNAi: results and challenges. *Annu Rev Biochem* 2010, 79:37-64.
- Mohr SE, Patmon N: RNAi screening: new approaches, understandings, and organisms. *Wiley Interdiscip Rev RNA* 2012, 3(2):145-158.
- Simpson KJ, Davis GM, Boag P: Comparative high-throughput RNAi screening methodologies in *C. elegans* and mammalian cells. *Nat Biotechnol* 2012, 29(4):459-470.
- Ebeshir SM, Harborth J, Landeckel W, Yalcin A, Weber K, Tuschke P:plexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* 2001, 411(6836):494-498.
- Mercor J, Snijder B, Sacher R, Burkard C, Black CK, Stahlberg H, Palmans L, Halenius A: RNAi screening reveals proteasome- and cullin-dependent stages in vaccinia virus infection. *Cell Rep* 2012, 2(4):1036-1047.
- Startz S, Shaw ML: Uncovering the global host cell requirements for influenza virus replication via RNAi screening. *Microbes Infect* 2011, 13(5):516-525.
- Misselwitz B, Dilling S, Vonaesch P, Sacher R, Snijder B, Schlumberger M, Rout S, Stark M, von Meining C, Palmans L, Harck: VMDNAi screen of *Salmonella* invasion shows role of COPI in membrane targeting of cholesterol and Cdc42. *Mol Syst Biol* 2011, 7:474.
- Snijder B, Sacher R, Ramo P, Libarali P, Manck K, Wolfrum N, Burchard L, Scott CC, Verheije MH, Mercor J, Moos S, Heger T, Thomsen K, Jurgé A, Lamparter D, Balistreri G, Schellhaas M, De Haan CA, Marjomaki V, Hyytiä T, Rottler PJM, Sodak B, Marsh M, Guanaberg J, Amara A, Graber U, Halenius A, Palmans L: Single-cell analysis of population context advances RNAi screening at multiple levels. *Mol Syst Biol* 2012, 8:570.
- Schmidt EE, Palz O, Buhlmann S, Karr G, Horn T, Boutros M: MinomRNAi: a database for cell-based and in vivo RNAi phenotypes, 2013 update. *Nucleic Acids Res* 2013, 41(Database issue):D1034-D1036.
- Pache L, König R, Chanda S: Identifying HIV-1 host cell factors by genome-scale RNAi screening. *Methods* 2011, 53(1):3-12.
- Sigoillot FD, King RW: Wiggle and validation: keys to success in RNAi screening. *ACS Chem Biol* 2011, 6(1):42-60.
- Bushman FD, Malani N, Fernandes J, Orso I, Cagney G, Diamond TL, Zhou H, Hazuda DJ, Espaseth AS, König R, Bandyopadhyay S, Ideker T, Goff SP, Krogan NJ, Frank AD, Young JA, Chanda S: Host cell factors in HIV replication: meta-analysis of genome-wide studies. *PLoS Pathog* 2009, 5(5):e1000437.
- Hayman-Gene S, Pache L, Chanda SK, Rosen: Functional genomic and high-content screening for target discovery and deconvolution. *Expert Opin Drug Discov* 2012, 7(10):955-968.
- Sivan G, Martin SE, Myers TG, Buehler E, Szymczyk KH, Ormanoglu P, Moss B: Human genome-wide RNAi screen reveals a role for nuclear pore proteins in poxvirus morphogenesis. *Proc Natl Acad Sci U S A* 2013, 110(9):3519-3524.
- Buehler E, Chen YC, Martin SE: A bench-level control for sequence specific siRNA off-target effects. *PLoS One* 2012, 7(12):e41942.
- Basik MC, Kampmann M, Lebbink RJ, Wang S, Hain MY, Poser I, Weibezahn J, Horback MA, Chan S, Mann M, Hyman AA, Laproust EM, McManus MT, Weissman JS: A systematic mammalian genetic interaction map reveals pathways underlying ricin susceptibility. *Cell* 2013, 152(4):909-922.
- Kittler R, Sundararath V, Haninger AK, Slabicki M, Theis M, Putz G, Franke K, Caldarella A, Grabner H, Kozak K, Wagner J, Rees E, Korn B, Frenzel C, Sachse C, Sonnichsen B, Guo J, Schalter J, Burchard J, Linsley PS, Jackson AL, Habermann B, Buchholz F: Genome-wide resources of endoribonuclease-prepared short interfering RNAs for specific loss-of-function studies. *Nat Methods* 2007, 4(4):332-344.
- Collinet C, Stoter M, Bradshaw CR, Samusik N, Rink JC, Kanski D, Habermann B, Buchholz F, Henschel R, Mueller MS, Nagal WE, Fava E, Kalaidzidis Y, Zerial M: Systems survey of endocytosis by multiparametric image analysis. *Nature* 2010, 464(7286):243-249.
- Marina S, Bahl A, Famer M, Buehler E: Common seed analysis to identify off-target effects in siRNA screens. *J Biomol Screen* 2012, 17(3):370-378.
- Buehler E, Khan AA, Marina S, Rajaram M, Bahl A, Burchard J, Famer M: siRNA off-target effects in genome-wide screens identify signaling pathway members. *Sci Rep* 2012, 2:428.
- Sigoillot FD, Lyman S, Huckins JF, Adamson B, Chung E, Quattrocchi B, King RW: A bioinformatics method identifies prominent off-targeted transcripts in RNAi screens. *Nat Methods* 2012, 9(4):363-366.
- Dahio C, Meyer M, Berger J, Schwarz H, Lanz: Interaction of *Bartonella henselae* with endothelial cells results in bacterial aggregation on the cell surface and the subsequent engulfment and internalisation of the bacterial aggregate by a unique structure, the invasome. *J Cell Sci* 2007, 119(Pt 18):2141-2154.
- Cossart P, Sansonetti P: Bacterial invasion: the paradigms of enterovasive pathogens. *Science* 2004, 304(5668):242-248.
- Meier O, Bouckle K, Hammar SV, Keller S, Scidwili RP, Hammi S, Graber UF: Adenovirus triggers macropinosytosis and endosomal leakage together with its clathrin-mediated uptake. *J Cell Biol* 2002, 158(8):1119-1131.
- Mercor J, Halenius A: Vaccinia virus uses macropinosytosis and apoptotic mimicry to enter host cells. *Science* 2008, 320(5875):531-535.
- Malo N, Hanley JA, Gerquozzi S, Palotier J, Nadori: Statistical practice in high-throughput screening data analysis. *Nat Biotechnol* 2006, 24(2):162-175.
- Carpenter AE, Jones TR, Lamprecht MR, Clarke C, Kang H, Friman O, Guertin DA, Chang JH, Lindquist RA, Moffat J, Golland P, Sabatini DM: iDRI: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol* 2006, 7(10):R100.
- Birmingham A, Salfors LM, Forster T, Wrobel D, Kennedy CJ, Shanks E, Santoyo-Lopez J, Dunican DJ, Long A, Kalleher D, Smith Q, Bejarsbergen RL, Ghazal P,

- Shamu CE: Statistical methods for analysis of high-throughput RNA interference screens. *Nat Methods* 2009, 6(8):569-575.
30. Siebourg J, Mendes G, Misselwitz B, Hardt WD, Beerenwinkel: Stability of gene rankings from RNAi screens. *Bioinformatics* 2012, 28(12):1612-1618.
 31. Srijder B, Paikmans I: Origins of regulated cell-to-cell variability. *Nat Rev Mol Cell Biol* 2011, 12(2):119-125.
 32. Rouilly V, Pujadas E, Hulst B, Balazs C, Kunstz P, Podvino: BRAINZ: automated analysis and data handling for RNAi screens. *Stud Health Technol Inform* 2012, 175:205-213.
 33. Pinheiro JC, Bates DM: *Mixed-Effects Models in S and S-PLUS*. New York: Springer; 2000.
 34. Bates D, Maechler M, Bolker B: *lme4: Linear Mixed-Effects Models Using Eigen and Eigen*. 2013. <http://cran.r-project.org/package=lme4>.
 35. Efron B, Tibshirani RJ, Narasimhan B, Taqqu FS: *Computationally Intensive Methods for Data Analysis*. 2011. <http://cran.r-project.org/package=locfdr>.
 36. Smyth GK: Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* 2004, 3:Article3.
 37. Pizarro-Cada J, Kuhbacher A, Cossart B: Entry of *Listeria monocytogenes* in Mammalian Epithelial Cells: an updated view. *Cold Spring Harbor Perspect Med* 2012, 2:11.
 38. Tattoli I, Philippot DJ, Girardin SE: The bacterial and cellular determinants controlling the recruitment of mTOR to the Salmonella-containing vacuole. *Biol Open* 2012, 1(12):1215-1225.
 39. König R, Chiang CY, Tu BP, Yan SF, Dalesio PD, Romero A, Bergauer T, Orth A, Krueger U, Zhou Y, Chanda SK: probability-based approach for the analysis of large-scale RNAi screens. *Nat Methods* 2007, 4(10):847-849.
 40. Zhang XD, Ferrer M, Espaseth AS, Marina SD, Stec EM, Crickower MA, Holder DJ, Hayes JF, Studovic B: The use of strictly standardized mean difference for hit selection in primary RNA interference high-throughput screening experiments. *J Biomol Screen* 2007, 12(4):497-509.
 41. Meinshausen N, Bühlmann P: Stability selection. *J R Stat Soc Ser B (Statistical Methodology)* 2010, 72(4):417-473.
 42. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP: Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005, 102(43):15545-15550.
 43. Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P, Doerks T, Stark M, Müller J, Bork P, Janssen U, von Mering: The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res* 2011, 39(Database issue):D561-D568.
 44. Elhafi MG, Al-Hokail A: Transforming growth factor beta production correlates with depressed lymphocytes function in humans with chronic brucellosis. *Microbes Infect* 2006, 11(14-15):1089-1096.
 45. Shan Y, Naujokas K, Park M, Irton M: B2B-dependent internalization of *Listeria* is mediated by the Met receptor tyrosine kinase. *Cell* 2000, 103(3):501-510.
 46. Zhang K, Chen DS, Wu YQ, Xu XJ, Zhang H, Chen CF, Chen HC, Liu ZF: MicroRNA expression profile in RAW264.7 cells in response to *Brucella melitensis* infection. *Int J Biol Sci* 2012, 8(7):1013-1022.
 47. Qin QM, Pei J, Ancona V, Shaw BD, Ficht TA, de Figueiredo: RNAi screen of endoplasmic reticulum-associated host factors reveals a role for IRE1alpha in supporting *Brucella* replication. *PLoS Pathog* 2006, 4(7):e1000110.
 48. Cain RJ, Vanhaesebroeck B, Ridley AD: The PI3K p110alpha isoform regulates endothelial adherens junctions via Pyk2 and Rac1. *J Cell Biol* 2010, 188(6):863-876.
 49. Nakamoto T, Nakata T: Optogenetic control of PIP3: PIP3 is sufficient to induce the actin-based active part of growth cones and is regulated via endocytosis. *PLoS One* 2013, 8(8):e70861.
 50. Grassie H, Rishle A, Wilker B, Gulbins E: Binoviruses infect human epithelial cells via ceramide-enriched membrane platforms. *J Biol Chem* 2005, 280(28):26256-26262.
 51. Sessions OM, Barrows NJ, Souza-Neto JA, Robinson TJ, Hershay CL, Rodgers MA, Ramirez JL, Dimopoulos G, Yang PL, Pearson JL, Garcia-Blanco MA: Discovery of insect and human dengue virus host factors. *Nature* 2009, 458(7241):1047-1050.
 52. Suomalainen M, Luisoni S, Boucka K, Bianchi S, Engal DA, Graber AF: Direct and versatile assay measuring membrane penetration of adenovirus in single cells. *J Virol* 2013, 87(22):12367-12379.
 53. Yakimovich A, Gumpert H, Burckhardt CJ, Lutschg VA, Jurgait A, Szalzarini IF, Graber UF: Cell-free transmission of human adenovirus by passive mass transfer in cell culture simulated in a computer model. *J Virol* 2012, 86(8):10123-10137.
 54. Dehio M, Knöke A, Lanz C, Dehio C: Construction of versatile high-level expression vectors for *Bartonella henselae* and the use of green fluorescent protein as a new expression marker. *Gene* 1998, 215(2):223-229.
 55. Cali J, Salgado SP, Gonval JB: Brucella coopts the small GTPase Sar1 for intracellular replication. *Proc Natl Acad Sci U S A* 2005, 102(5):1673-1678.
 56. Jurgait A, Moessa S, Roulin P, Dorsch A, Lotzsch M, Lee WM, Graber AF: RNA replication-center assay for high content image-based quantifications of human rhinovirus and coxsackievirus infections. *Virol J* 2010, 7:264.
 57. Jurgait A, McDowell R, Moessa S, Meldrum E, Schwendener R, Graber UF: Niclosamide is a proton carrier and targets acid: endosomes with broad antiviral effects. *PLoS Pathog* 2012, 10:8.
 58. Mosser AG, Brockman-Schneider R, Aminava S, Burchell L, Sedgwick JB, Busse WW, Gem JE: Similar frequency of rhinovirus-infectible cells in upper and lower airway epithelium. *J Infect Dis* 2002, 185(6):734-743.
 59. Efron B: *Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction*. Cambridge: Cambridge University Press; 2010.

doi:10.1186/1471-2164-15-1162

Cite this article as: Râmô et al: Simultaneous analysis of large-scale RNAi screens for pathogen entry. *BMC Genomics* 2014 15:1162.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

Supplementary information

Table of contents

InfectX.....	2
Assays performed.....	2
Differences in protocols.....	2
Quality control.....	3
Computational infrastructure.....	4
openBIS.....	4
Cluster computing and data storage.....	4
Image analysis.....	4
Object detection.....	4
Feature extraction.....	5
Infection detection and measurement.....	5
Data preprocessing and normalization.....	8
Prediction of siRNA on-target genes.....	8
Z-Scoring.....	8
Dependency of infection index to population context.....	8
Cell number readout reproduces conclusions data reproducibility.....	9
Data fulfills requirements for Moderated T-Test.....	9
Study of parallelism.....	9
Reference methods (RSA and MTT).....	9
Comparing the results of PMM, MTT and RSA.....	10
Comparative GSEA results.....	10
PMM results for cell number readout.....	10
Data fulfills requirements for PMM.....	10
Supplementary references.....	11
Figure Legends.....	13
Supplementary Figures.....	15

InfectX

InfectX is a Switzerland-based consortium consisting of eleven research groups in the framework of SystemsX.ch, the Swiss initiative for systems biology. Its focus is on comparatively studying the infection processes of bacterial and viral pathogens and developing computational methods for image analysis and statistical modeling. InfectX has a strong emphasis on developing and using unified wet-lab and analysis protocols and workflows. Supplementary Figure 1 shows the outline of the analysis workflow.

Assays performed

We screened all studied pathogens with four different siRNA libraries (Dharmacon pooled, Dharmacon unpooled, Ambion unpooled, and Qiagen unpooled). Supplementary Table I reports the number of replicates of different pathogen assays. The value in parenthesis indicates the number of assays that needed to be removed because of different problems (e.g. transfection did not work or an older protocol was used that gave a weak fluorescent signal). We used the Dharmacon pooled library for assay optimization and several of them needed to be removed. In all analyses we averaged data from replicates in order to have only one value per library and gene.

Supplementary Table: Replicates performed for different pathogens and libraries

Pathogen	Dharmacon 1 pooled siRNA	Ambion 3 unpooled siRNAs	Qiagen 4 unpooled siRNAs	Dharmacon 4 unpooled siRNAs
Adenovirus	7 (+1)	2	1	1
B. abortus	7 (+1)	2	1	2
B. henselae	4 (+1)	3 (+1)	1 (+1)	1
L. monocytogenes	4	2 (+2)	1	1
Rhinovirus	6 (+2)	2	1	1
S. flexneri	4 (+2)	2	1	1
S. typhimurium	3 (+4)	3	1	1
Vaccinia virus	1 (+1)	2	1	1

Differences in protocols

We designed all pathogen specific protocols so that the resulting phenotypes (infection index and cell number) would end up as similar as possible between all the screens. We aimed to reach a cell number for wells where there would be some empty space between the cell colonies (meaning roughly 1500 cells the imaged area of a well). In addition, we aimed to reach infectivity close to 30-50% in a typical well. This would enable the reliable detection of both down and up hits in the infection index readout. However, due to biological issues, sometimes we could not reach these goals, for example for B. abortus the typical infection rate remained low (typically close to 5%) and for B. henselae the infection rate was high (typically close to 90%). The different infection biology of different pathogens required adaptations to the protocols in order to aim to the above mentioned goals. Supplementary Table II summarizes the main differences in protocols.

SupplementaryTable II: Table of pathogens and main differences in protocols.

Pathogen	Strain	Seeded cell number/well	Multiplicity of infection (MOI)	Pathogen entry time (pi: primary infection; si: secondary infection)	Total infection time	DNA stain	Actin stain	Pathogen detection	Additional stain
Adenovirus	Ad2_ΔE3B-eGFP	700	0.1	16 h	16 h	DAPI	Cy3: DY-647-phalloidin	GFP	
<i>B. henselae</i>	<i>B. henselae</i> ΔbepG	300	400	30 h	24 h	DAPI	RFP: DY-347-Phalloidin	GFP + Actin	Cy3: Alexa 647
<i>B. abortus</i>	<i>B. abortus</i> 2308 pUC43 (aphT::GFP)	500	10000	4 h	44 h	DAPI	RFP: DY-347-phalloidin	GFP	-
<i>L. monocytogenes</i>	EGD-e.PrfA ⁺ GFP	600	25	1 h	3 h	DAPI	Cy3: DY-647-phalloidin	GFP	0.2% saponin, PBS containing Alexa Fluor-647 coupled anti-rabbit antibody, Cy3: InC
Rhinovirus	HRV1a	1000	8	7 h	7 h	DAPI	Cy3: DY-647-phalloidin	GFP	Anti-VP2 antibody Mab 16/7 followed by anti-mouse IgG coupled to Alexa Fluor 488
<i>S. typhimurium</i>	S.TmSopE_pM975, SL1344, sopE2sipAsop8 (S. TmSopE pM975)	550	80	20 min	4 h	DAPI	RFP: DY-347-Phalloidin	GFP	-
<i>S. flexneri</i>	<i>S. flexneri</i> M90T ΔvirG pCK100	600	15	30 min	3.5 h	Hoechst	GFP: DY-485-Phalloidin	DsRed	Anti-mouse IgG coupled to Alexa Fluor 647, Cy3: IL-8
Vaccinia virus	WR E EGFP/L mCherry	600	0.125	1 h (pi), 8 h (si)	24 h	Hoechst	Cy3: DY-647-phalloidin	GFP, RFP	Anti-GFP antibody followed by Alexa Fluor 488

Quality control

Quality control is an umbrella term that covers many different approaches and methods to guarantee “correctness” and “unbiasedness” of the data. Since a large-scale high-content screening setup includes many different kind of steps (wet-lab, imaging, data storage, image analysis and statistics - just to name a few), systematic quality control is essential. The main goal is to control known problem types and tackle them by appropriate approaches in each step. In addition, some a-priori unknown problems can be discovered by clever quality control and automated analysis. However, it is hard to be aware of all possible error sources and therefore we may not be able to check the data against all of them. For complex high-throughput and high-content data the number of possible known error types is usually very large. Therefore, we created standard operating procedures (SOPs) for quality control in order to systematically go through them and ensure that our

data does not contain any of these known problem types. InfectX defined one common Quality Control SOP for all pathogens. We also clearly defined the responsible person for each task.

Computational infrastructure

We have identified data analysis and computational infrastructure to be key aspects of attaining reproducible high-content screening results. Small variations in analysis procedures or the software environment can lead to differences in numerical results or file formats. These incompatibilities are hard to spot at the time they arise, but might significantly increase the effort or even invalidate comparable data analysis from varying sources. InfectX has made considerable efforts to ensure that the data is analyzed in a reproducible and accessible way. In more detail, key aspects of our data analysis are: all results are annotated with the methods and settings that were used in their creation new methods undergo regression testing before being applied to any dataset. Result datasets are automatically shared through our openBIS data portal with all members of the consortium.

openBIS

Data management and data sharing was performed using the openBIS biology information system [1]. To this end, openBIS has been extended to support screening metadata like the siRNA library, and screening results like images, well-based readouts and object-based features. For example, one extension includes the visualization of images together with their analysis results.

iBRAIN2: workflow and process management for screening data

In response to the data analysis requirements of InfectX, an open-source workflow management solution called iBRAIN2 was developed [2]. (<http://ibrain2.sourceforge.net/>). iBRAIN2 can employ openBIS as a data management solution and enables the parallel analysis of HCS datasets on high-performance computing (HPC) clusters. It provides resilient and flexible workflow management capabilities in order to face the number and complexity of the analysis steps performed on these typically large datasets. The modular design of this solution allowed the definition of sequential analysis steps to be performed on acquired data. These workflows are not instance-specific and success/failure criteria can be defined for each sub-step. They can therefore be part of SOPs and simplify quality control during data analysis.

Cluster computing and data storage

Data analysis was performed on a Linux-based computer cluster of heterogeneous multicore nodes running Sun Grid Engine on CentOS Linux. The nodes are based on x86_64 architecture with 2GB RAM per process. About 100.000 CPU hours were used in data analysis. Data storage is performed on an NFS-mounted IBM SONAS storage system.

Image analysis

Object detection

The following steps are common for all pathogens. Images were first scaled so that pixel intensities of a full plate are in the 0 to 1 range. Images were then corrected for shading (flat field correction, vignetting correction) by applying a shading model to the image pixels. Shading-corrected images were stored in floating points to reduce the loss of information. For bacterial pathogens *B. henselae*, *B. abortus*, and *S. flexneri*, the pathogen signal in the DAPI channel (referring to both DAPI and Hoechst stainings) was removed to increase the quality of the nucleus segmentation. The pathogen signal was removed by subtracting a linear transformation of the GFP channel (referring to pathogen specific infection channel) from the DAPI channel. After the pathogen signal reduction, DAPI images

were stored in double precision to reduce loss of information. On the corrected images, object detection was performed using CellProfiler [3]. First, nucleus objects labeled “Nuclei” were segmented in the DAPI channel using OTSU’s method (CellProfiler module IdentifyPrimAutomatic). Second, a peri-nuclear ring object labeled “PeriNuclei” was constructed by extending the nucleus object by eight pixels and removing the nuclear area from the so extended nuclear area (CellProfiler modules ExpandOrShrink and IdentifyTertiary). Third, a cell body object labeled “Cells” was segmented in the Actin channel using the “Propagation” method around the nucleus object (CellProfiler module IdentifySecondaryInformed). Fourth, a non-actin based cell body object labeled “VoronoiCells” was constructed by extending the nucleus object by twenty-five pixels (CellProfiler module ExpandOrShrink).

Feature extraction

The following steps were common for all pathogens. On the segmented objects, measurements were performed using CellProfiler. On all four segmented objects (Nuclei, PeriNuclei, Cells, VoronoiCells) shape measurements were extracted. Intensity and texture measurements were extracted with respect to all available channels (DAPI, Actin, Pathogen, and pathogen-specific channel where applicable). The neighborhood relationship was measured for cell body objects that are within a two-pixel distance of each other. All measurement result files of CellProfiler were stored in the openBIS database alongside the original images. For ease of access, the data was refactored so that only one class of measurements is contained in a single unique file with the same internal structure as the original CellProfiler result file.

Infection detection and measurement

The approach for detection of infected cells in images was pathogen specific (details for each pathogen are listed below). The result of the infection detection was for all pathogen assays a cellular phenotype that indicates infection of an individual cell on a binary level (the cell is infected vs. the cell is not infected). In addition, some of the infectious phenotypes indicate the level of infection for each cell. The binary infection phenotype allows us to define the infection index readout for all pathogens. The infection index is defined as: number of infected cell / total number of cells in the well. The infection index is the main readout for all pathogens. Most of the pathogen screens also include additional stains for other pathogen infection related phenotypes (secondary readouts). These additional phenotypes are not discussed or analyzed in detail in this paper. We used several different algorithms to detect the binary infection phenotypes (i.e. infection scoring). We found that for quality control it is important to use different methods of infection scoring. Deviations in the infection score readouts point to possible problems and agreements between the readouts cross-validate the various infection scoring methods. For all pathogens, with the exception of *B. henselae*, we used at least two of the following infection scoring algorithms for infection detection. For *B. henselae* we applied a separate algorithm because of its special infection phenotype in form of invasomes (see *B. henselae* infection detection and measurement).

Decision Tree Infection Scoring (DTIS)

We selected a small number of image analysis single cell features that were most sensitive to the infection phenotype (typically from two up to five features). The N features are evaluated in a decision tree, which is a complete binary tree with N levels and 2^N nodes. Each node is evaluated by applying a threshold to the corresponding feature. During traversal of the tree, if the feature exceeds the threshold, evaluation continues with the one child, and if the feature does not exceed the threshold, evaluation continues with the other child. Nodes of the lowest level connect to one of the

two distinct end states “infected” and “uninfected”. The connection of the nodes to children and the choice of features are performed once by an expert and remain static for all plates of a pathogen. The choice of the decision tree thresholds is affected by plate-specific parameters like quality of the staining, cell vitality and microscope illumination, and must be adjusted on a plate-by-plate basis. We supply a table that lists for each plate the used features and their corresponding thresholds.

SVM infection scoring

We used CellClassifier [4] and supervised machine learning using a Support Vector Machine based binary classifier [5] to separate infected cells from non-infected cells. Most pathogens show a clear binary infection phenotype for cells (for example *S. typhimurium* or Vaccinia virus). For these screens, the supervised machine learning infection scoring was relatively straight-forward and the Support Vector Machine based classifier typically produced high quality results (classification accuracy >99%). However, some pathogens have relatively continuous infections levels (for example *L. monocytogenes*). For these screens, supervised binary classification was not optimal and subjective evaluation (e.g. to decide which cell is infected enough to be classified as infected) was required during the training phase. For all pathogens 3 to 5 features were manually selected and the features were plate-wise Z-Scored prior to applying SVM learning.

Segmentation based infection scoring

For some pathogens we can apply image-based segmentation of pathogen objects in CellProfiler to detect pathogen colonies or single pathogens in the cell. Therefore, we used a segmentation method based on the OTSU method or on wavelets. Pathogen object segmentation leads directly to a binary infection scoring for each cell. A cell is defined as “infected” if a pathogen object overlaps mostly with this cell. This definition ensures that no pathogen object is considered belonging to more than one cell, even though if it overlaps with more than one cell.

Adenovirus infection detection and measurement

Cells were infected with a replication competent Ad2_GFP_dE3B (in short Adenovirus) [6]. Infected cells were scored by their dispersed GFP signal of variable intensity across the cell body, most prominently in the nuclear area. Strength of the signal was strictly dependent on the amount of virus added to the cells, ranging from very strong intensity (high infection) to background intensity (no or very low infection). To quantify Adenovirus infection, GFP intensity was measured in the objects Nuclei, PeriNuclei, Cells and VoronoiCells using CellProfiler module MeasureObjectIntensity

B. henselae infection detection and measurement

For *B. henselae* screens, infection of a cell is defined by the appearance of pathogen-induced invasome structure, an Actin surrounded membrane structures containing a large bacterial aggregate internalized as a whole in the cell body. Invasome object detection was performed by applying template matching to the Actin channel using templates of an idealized invasomes of varying size. A segmentation of the invasome was achieved by determining the sphere of maximum actin intensity surrounding the candidate location. On the segmented object intensity measurements in the Actin and GFP channel were extracted using CellProfiler module MeasureObjectIntensity. Shape measurements were extracted using module MeasureObjectAreaShape. The invasome detection algorithm was very sensitive to invasome candidates, but also detects false positive invasome structures. We trained classifier based on a Support Vector Machine to separate true invasomes from false positive invasomes. We found the most descriptive feature for true invasomes to be the GFP

channel intensity, which indicates presence of a bacterial cluster. The classification was based on the information in the GFP channel. After the true invasomes were detected, a cell was classified as infected if the cell has one or more true invasomes assigned [7].

B. abortus infection detection and measurement

B. abortus infection appears as large micro colonies across the cell body. To quantify infection for **B. abortus**, GFP intensity was measured in the objects Nuclei, PeriNuclei, Cells and VoronoiCells using CellProfiler module MeasureObjectIntensity.

L. monocytogenes infection detection and measurement

L. monocytogenes infection appears as dispersed Cy3 signal of varying intensity across the cell body. The strength of the signal is dependent on the amount of bacteria in the cell, ranging from very strong intensity (high infection) to background intensity (very low to no infection). The majority of signal is accumulated in the perinucleus. To quantify infection for **L. monocytogenes** Cy3 intensity was measured in the objects Nuclei, PeriNuclei, Cells and VoronoiCells using CellProfiler module MeasureObjectIntensity.

Rhinovirus infection detection and measurement

Cells were infected with the strain HRV1A as described [8]. The infection phenotype were small cytoplasmic clusters of viral replication sites of varying size detected by the monoclonal antibody 16-7, and a secondary anti-mouse IgG conjugated to Alexa488. Wavelet-based object detection was used to segment the replication objects, and infection was measured by determining the fluorescence intensity using the CellProfiler module MeasureObjectIntensity.

S. flexneri infection detection and measurement

S. flexneri infection appears in form of an accumulation of micro colonies which is often localized in the vicinity of the nucleus (perinucleus). The accumulation is formed because we use a non-motile mutant of **S. flexneri** ($\Delta virG$), which is not able to move by actin-based motility. Once intracellular, the bacteria replicate themselves resulting in the formation of micro colonies. A cell can contain one or several micro colonies. Extracellular bacteria are not visible as only intracellular bacteria express the fluorescent marker. Segmentation using OTSU's method was used to segment bacteria objects. To quantify infection for **S. flexneri**, RFP intensity was measured in the bacteria objects using CellProfiler module MeasureObjectIntensity.

S. typhimurium infection detection and measurement

S. typhimurium infection appears as small GFP dots in the cell body. Wavelet-based object detection was used to segment virus objects. To quantify infection for **S. typhimurium**, GFP intensity was measured in the virus objects using CellProfiler module MeasureObjectIntensity.

Vaccinia virus infection detection and measurement

Vaccinia virus infection appears as dispersed GFP signal of varying intensity across the cell body. To quantify infection for Vaccinia virus, GFP intensity was measured in the objects Nuclei, PeriNuclei, Cells and VoronoiCells using CellProfiler module MeasureObjectIntensity.

Data preprocessing and normalization

Prediction of siRNA on -target genes

A target gene for a specific siRNA is defined as a gene, which exhibits perfect complementarity within its coding region to this siRNA. This is not necessarily a 1:1 relation, i.e. siRNAs can potentially have multiple target genes. In order to identify these target relations, siRNA sequences were searched against genomic transcript sequences from RefSeq (release hg19, downloaded 17.07.2012) and ENSEMBL (release GRCh37.67, downloaded 20.07.2012) using BLAST version 2.2.27. The BLAST parameter `word_size` was set to 7. Transcript matches shorter than siRNA sequence length, as well as matches with gaps were removed. Finally, transcript IDs were translated to gene IDs and the unique set of target gene ID(s) considering both genomic data sources were reported for each siRNA.

Z-Scoring

Individual plates of screens cannot always be handled identically in the wet-lab. For this reason, we often observe differences in the readout levels of single plates or in plate batches. There are several approaches in the literature to correct for these differences [9]. Negative controls (MOCK and SCRAMBLED) sometimes show non-typical phenotypes (such as relatively high cell number) and good positive controls were not always available for all pathogens before primary screening. Therefore, we chose non-control based data normalization methods. We used Z-Scoring to normalize variations between plates as:

$$x_{new} = \frac{x_{old} - \mu}{\sigma}$$

Here μ is the mean of all siRNA well readouts in the plate, σ is the standard deviation of all siRNA well readouts in the plate, x_{old} is the raw well readout and x_{new} is the normalized well readout. The non-control based normalization assumes that all genes are randomly distributed among all plates and that there are relatively few positive phenotype genes in the whole screen. After the plate Z-scoring we also Z-scored the whole screen in order to generate comparable screens. We used the above mentioned method with the mean and standard deviation of the whole screen.

For Z-Scoring, we need to assume that the data is approximately Gaussian distributed and that only a relatively small number of data points are outliers in each plate. Supplementary Figure 2 shows a histogram of the data and QQ-plot for the AdenovirusDharmacon pooled screen for an example plate. The infection indices are nearly Gaussian distributed. Only a minority of data points in the plate are outliers. Therefore, the assumptions are fulfilled in a good approximation. The results for all the other plates and pathogens are qualitatively similar.

Dependency of infection index to population context

Infection phenotypes can depend on the population context, such as the total cell number [10-12]. Supplementary Figure 3 shows examples that for some screens there is a slight dependency on the cell number (for example *B. henselae*, *L. monocytogenes* and Vaccinia virus), but for some pathogens (*B. abortus*) the dependency is not visible. We conjecture that the HeLa ATCC cell line shows lesser dependency of phenotypes to population context than some other cell lines reported in the literature.

To reduce the bias caused by the correlation we applied non-parametric regression correction with the Lowess-method [13]. To normalize the Z-Scored infection index, we use a sliding window of size 200 to go through the ranked cell number readouts. For each window we calculate the mean m and

standard deviation s . With those values we then Z-Score the infection index of the well in the center of the window. The Lowess-normalized value is

$$x_{new} = \frac{x_{old} - m}{s}.$$

This method also normalizes the possible biases in standard deviation.

Cell number readout reproduces conclusions data reproducibility

Figure 2 shows data correlations of replicate screens and screens performed with different siRNAs targeting the same genes using the infection index readout. Supplementary Figure 4 shows the results of the same analysis using the cell number readout. Identical qualitative conclusions as for the infection index readout can also be drawn from the cell number readout results.

Data fulfills requirements for Moderated T-Test

The Moderated T-Test (MTT) assumes that the sample standard deviations of the siRNAs within a gene are Chi-squared distributed. This assumption is tested by plotting the observed quantiles of the sample standard deviation versus a Chi-Squared distribution (Supplementary Figure 5a). The points show a straight line for all eight pathogens, indicating that our data satisfies these assumptions. MTT is followed by Storey's multiple testing correction. This approach assumes that the p-values have a flat distribution with a possible peak at the low end. The histograms illustrate that our p-values fulfill this assumption for all pathogens (Supplementary Figure 5b).

Study of parallelism

In the main text we presented a study how parallelism affects the ranking of genes in individual screens when using PMM. In the main Figure 6 we only used *L. monocytogenes* as an example. Supplementary Figure 6 shows the same results for all the pathogens. In all cases we mostly gain hits by including parallel screens into the simultaneous analysis.

Reference methods (RSA and MTT)

As a reference method, we used Moderated T-Test (MTT) [14]. It tests whether the observed distribution of a sample (in our case the collection of readouts of one gene) has a mean equal to 0. In contrast to the one-sample t-test, the test statistic includes as prior information the different variances of the siRNAs within the genes. Therefore, it assumes that the standard deviation of the test samples are chi-squared distributed. We performed MTT using the R implementation presented in [15]. Our data satisfies the method assumptions (see Supplementary Figure 5a). The obtained p-values cannot be directly used in large-scale screening because of the problems caused by multiple testing [16-19]. In recent years, several methods have been proposed to control the significance levels with respect to the False Discovery Rate (FDR) and corresponding q-values [16, 17]. We used the method in Storey and Tibshirani (2003). We refer to their paper for the full description and R implementation of the method. The Storey multiple testing correction assumes that the distribution of p-values is flat, with a possible peak at the lower end [17]. Supplementary Figure 5b shows the histogram of the p-values for all pathogens.

The Redundant SiRNA Analysis (RSA) ranks all siRNAs targeting a given gene over all siRNAs in the screens. It assigns the p-values for each gene based on a hypergeometric distribution that indicates whether the distribution of ranks of this gene is shifted significantly towards low ranks [12]. RSA was run using the R-package "RSA" release 1.2 [12] with parameters: $l=1.5$ and $u=1$, where l refers to the

threshold where a single siRNA readout is considered to be true positive at the low end and U refers to the threshold where a single siRNA readout is considered to be true positive at the high end.

Comparing the results of PMM, MTT and RSA

We compared the results of PMM, MTT and RSA by scatterplots of the gene ranks originating from these methods. In the main text we presented an example of MTT compared to PMM for *L. monocytogenes* (Figure 6b). Supplementary Figure 7 shows the comparisons between all method pairs for all pathogens. The correlations between the three different methods are relatively high. In particular, the top hits are most similar and genes are only slightly shuffled when using different hit ranking methods. Using the same setting as for the calculation of the ROC curves, we also compared missed rates between PMM; MTT and RSA. The results show that the false negative rate of our model is not higher than with the other commonly used methods (see Supplementary Figure 8).

Comparative GSEA results

In order to evaluate the biological relevance of found hits, we calculated pathway enrichment scores separately for each pathogen by the Gene Set Enrichment Analysis (GSEA) algorithm using as input the results from the three hit ranking algorithms PMM, MTT, and RSA. Gene Set Enrichment Analysis (GSEA) was run using the Java-package "gsea2-2.1.0.jar" and the curated canonical pathways "c2.cp.v4.0.entrez.gmt" [20]. The following settings were used for the parameters: collapse was set to false, mode to Max_probe, norm to meandiv, nperm to 1000, scoring_scheme to classic, include_only_symbols to true, make_sets to true, set_max to 500 and set_min to 7. We limited ourselves to pathways that had at least 7 kinases within the pathway in order to avoid bias towards too small pathways. We decided to use the "classic" approach (instead of the recommended "weighted" approach) in order to keep different hit scoring methods (PMM, MTT, and RSA) comparable. We selected all the pathways that were significant (GSEA pathway enrichment FDR score < 0.2) for any pathogen and method pair. The results are illustrated in Supplementary Figure 9. The heatmap shows GSEA pathway enrichment scores for all pathogens using as input the ranked lists of infection index down hits detected by PMM, MTT, and RSA. PMM found more significant pathways than the other methods for most pathogens.

PMM results for cell number readout

We fitted PMM also for the cell number readout. The fitted PMM yielded for all pathogens the same C_{pg} scores (the random effects b_{pg} were estimated to 0) and the same Q_{pg} for all genes. Therefore, we obtain the same significant genes for all pathogens (Supplementary Figure 10). The results reflect the fact that cell number is a pathogen independent readout.

Data fulfills requirements for PMM

PMM assumes that the model residuals ϵ_{pgs} , as well as the random coefficients a_g and b_{pg} are normally distributed. Supplementary Figure 11 shows diagnostic plots for the fitted PMM. The QQ-plot shows that the normal distribution for residuals is only approximately satisfied (Supplementary Figure 11a). There are outliers in the residuals with respect to positive and negative infection indices (marked with red points). To check whether the estimation of the PMM is affected by these outliers, we refitted the PMM without the red marked points. The resulting hit lists of the PMM with and without the red points are almost identical. Moreover, the residuals are randomly distributed around zero for each gene within a pathogen (Supplement Figure 11b-c). This indicates that there is no systematic error in the estimation of the PMM. Therefore, the PMM fit is reliable concerning the

residuals assumptions. The two other QQ-plots confirm the assumption of normal distribution for the random coefficients (Supplementary Figure 11d–e).

Supplementary references

1. Bauch A, Adamczyk I, Buczek P, Elmer FJ, Enimanev K, Glyzewski P, Kohler M, Pylak T, Quandt A, Ramakrishnan C et al: openBIS: a flexible framework for managing and analyzing complex data in biology research. *BMC Bioinformatics* 2011, 12:468.
2. Rouilly V, Pujadas E, Hullar B, Balazs C, Kunszt P, Podvinec M: iBRAIN2: automated analysis and data handling for RNAi screens. *Stud Health Technol Inform* 2012, 175:205-213.
3. Carpenter AE, Jones TR, Lamprecht MR, Clarke C, Kang IH, Friman O, Guertin DA, Chang JH, Lindquist RA, Moffat J et al: CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol* 2006, 7(10):R100.
4. Ramo P, Sacher R, Snijder B, Begemann B, Pelkmans L: CellClassifier: supervised learning of cellular phenotypes. *Bioinformatics* 2009, 25(22):3028-3030.
5. Hlava VF: Pattern recognition Toolbox for Matlab. Czech Technical University in Prague; 2000.
6. Yakimovich A, Gumpert H, Burckhardt CJ, Lutschg VA, Jurgeit A, Sbalzarini IF, Greber UF: Cell-free transmission of human adenovirus by passive mass transfer in cell culture simulated in a computer model. *J Virol* 2012, 86(18):10123-10137.
7. Truttmann MC, Guye P, Dehio C: BID-F1 and BID-F2 domains of Bartonella henselae effector protein BepF trigger together with BepC the formation of invasome structures. *PLoS one* 2011, 6(10):e25106.
8. Jurgeit A, Moese S, Roulin P, Dorsch A, Lotzerich M, Lee WM, Greber UF: An RNA replication-center assay for high content image-based quantifications of human rhinovirus and coxsackievirus infections. *Virology* 2010, 7:264.
9. Birmingham A, Selfors LM, Forster T, Wrobel D, Kennedy CJ, Shanks E, Santoyo-Lopez J, Dunican DJ, Long A, Kelleher D et al: Statistical methods for analysis of high-throughput RNA interference screens. *Nat Methods* 2009, 6(8):569-575.
10. Snijder B, Sacher R, Ramo P, Liberali P, Mench K, Wolfrum N, Burleigh L, Scott CC, Verheije MH, Mercer J et al: Single-cell analysis of population context advances RNAi screening at multiple levels. *Mol Syst Biol* 2012, 8:579.
11. Knapp B, Rebhan I, Kumar A, Matula P, Kiani NA, Binder M, Erfle H, Rohr K, Eils R, Bartenschlager R et al: Normalizing for individual cell population context in the analysis of high-content cellular screens. *BMC Bioinformatics* 2011, 12:485.
12. König R, Chiang CY, Tu BP, Yan SF, DeJesus PD, Romero A, Bergauer T, Orth A, Krueger U, Zhou Y et al: A probability-based approach for the analysis of large-scale RNAi screens. *Nat Methods* 2007, 4(10):847-849.
13. Yang YH, Dudoit S, Luu P, Lin DM, Peng V, Ngai J, Speed TP: Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res* 2002, 30(4):e15.
14. Smyth GK: Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* 2004, 3:Article3.
15. Smyth GK: Limma: linear models for microarray data. In: *Bioinformatics and Computational Biology Solutions using R and Bioconductor* Edited by R. Gentleman VC, S. Dudoit, R. Irizarry, W. Huber. New York: Springer; 2005: 397-420.
16. Storey JD: A direct approach to false discovery rates. *Journal of the Royal Statistical Society Series B-Statistical Methodology* 2002, 64:479-498.
17. Storey JD, Tibshirani R: Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A* 2003, 100(16):9440-9445.
18. Prummer M: Hypothesis Testing in High-Throughput Screening for Drug Discovery *J Biomol Screen* 2012, 17(4):519-529.

19. Benjamini Y, Hochberg y.: Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B*1995, 57(1):289–300.
20. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES et al: Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*2005, 102(43):15545-15550.

Figure Legends

Supplementary Figure 1.

InfectX data analysis workflow. We screened 11 single siRNA libraries (4 siRNAs from Dharmacon, 4 siRNAs from Qiagen, and 3 siRNAs from Ambion) and one pooled library from Dharmacon with 8 pathogens. We performed imaging with Molecular Devices ImageXPRESS microscopes with 10x magnification with 9 sites per well on 3-4 channels depending on the assay. Image analysis consisted of image shading correction, object segmentation (nuclei, perinuclei, cells, and Voronoi cells), feature extraction (typically 200 features per cell), and infection scoring (with up to three algorithms). We normalized well-based data with plate Z-scoring, population regression (Lowess), and experiment Z-scoring. Technical data aggregation steps were followed by hit detection (PMM, MTT, and RSA) and False Discovery Rate (FDR) analysis. We performed several comparative analyses (method comparison, GSEA enrichment analysis, and STRING network analysis). Data is publicly shared using the openBIS database. All data including raw images, single cell data, assay metadata, and well data are fully accessible through openBIS web GUI and several programming interfaces (APIs).

Supplementary Figure 2.

Data fulfills assumptions for Z-Scoring. (a) Histogram and (b) QQ-plot of non-normalized infection indices from plate 3 of the Adenovirus Dharmacon Pooled screen. The plots show a distribution with slightly fatter tails than Gaussian.

Supplementary Figure 3.

(a) Scatter plots showing dependencies of Z-Scored infection indices to cell number. Red lines correspond to the smoothed Lowess average estimate and green lines to the +/- standard deviation estimates of the dependency. The example data are from the Dharmacon pooled siRNA libraries.

Supplementary Figure 4.

Using more siRNAs adds power to yield reproducible results. (a) The three boxplots show the Pearson correlation coefficient R between screens performed using the same siRNA set. The numbers 1 to 3 correspond to the number of replicate screens that are averaged and compared to another distinct set of replicate screens, averaged over the same number. The replicate screens were resampled 500 times. The scatter plot shows an example for the correlation of cell numbers (CN) from a duplicate of Adenovirus Dharmacon pooled screen. (b) The set of six boxplots show the Pearson correlation coefficients of the averaged readouts from 1 to 6 siRNA sets. The scatter plots depict the correlation of cell numbers for Adenovirus, the first between two different single siRNAs and the second between each an average over six siRNAs.

Supplementary Figure 5.

(a) QQ-plot comparing the observed quantiles of the sample standard deviations of the siRNAs within each gene to a Chi-squared distribution. Different colors represent different pathogen assays. (b) Each line shows the distributions of p-values originating from MTT using as input the infection index readout for one pathogen.

Supplementary Figure 6.

The y-axis shows the estimated cpG scores for the pathogen indicated in the title. The x-axis shows cpG scores originating from a refitted PMM based on data where we randomized the other 7 parallel assays. The colors correspond to hit genes ($FDR < 0.4$) in different cases: green is a hit in both cases,

red is a hit in the fit based on the original data, and blue is a hit in the fit based on the randomized data.

Supplementary Figure 7.

(a) The y-axis shows the rank of a gene given by PMM and the x-axis the rank defined by MTT. The dot size corresponds to the sharedness score of each gene. The results of each pathogen are plotted separately in each plot. (b) Comparison of the ranks resulting from PMM and RSA for all eight pathogens. (c) Comparison of the ranks resulting from RSA and MTT for all eight pathogens.

Supplementary Figure 8.

The figure shows DET-curves for PMM, MTT and RSA applied on simulated data for three different scenarios (containing only hits that were shared between all pathogens, unique hits for all pathogens and mixed hit structure of both unique and shared hits). The dashed and solid lines indicate whether the shifts were generated by a low or high shift away from zero.

Supplementary Figure 9

GSEA pathway enrichment results for PMM, MTT, and RSA hit ranking methods. The numbers give the number of highly significant pathways (GSEA FDR < 0.2) for each hit detection method and pathogen. The significant pathways are high-lighted with a red square.

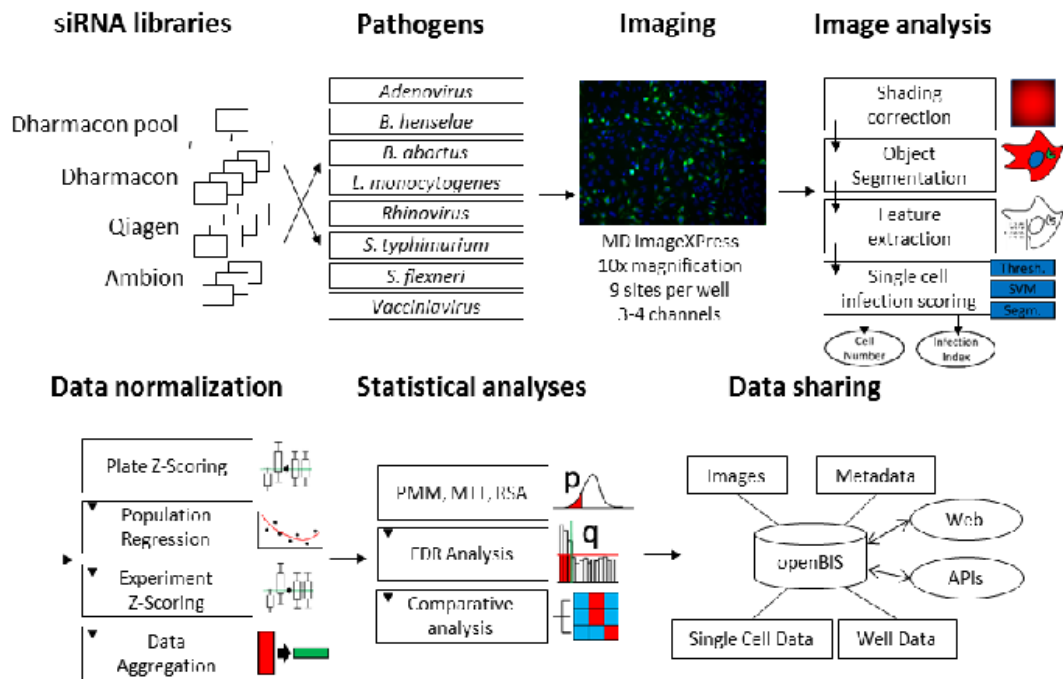
Supplementary Figure 10

Summary of screening hits for the cell count readout. The heat map shows all genes which were significant (FDR < 0.4) at least for one pathogen. The black outlines indicate significant genes (all the genes were significant for all pathogens) and the green outlines indicate the strongest hit. The colors correspond to the estimated cpg values.

Supplementary Figure 11.

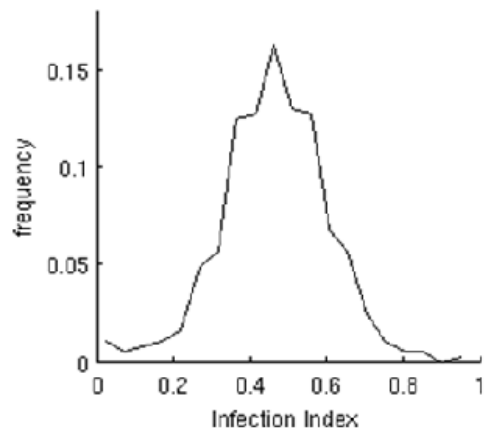
(a) QQ-plot comparing the observed quantiles of residuals from the PMM to the theoretical quantiles of a normal distribution. (b) Scatterplot of residuals of all combinations of genes and pathogens. (c) "Zoomed-in-version" of the scatterplot in (b) at both ends. The plot shows a random scatter around 0, indicating that there are no systematic errors in the estimation of PMM. (d) QQ-plot comparing the observed quantiles of the gene random effects (ag) to the theoretical quantiles of a normal distribution. (e) QQ-plot of the gene random effect within a pathogen (bpg).

Supplementary Figure 1

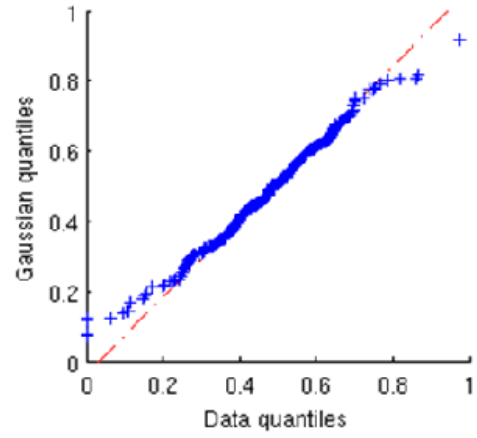


Supplementary Figure 2

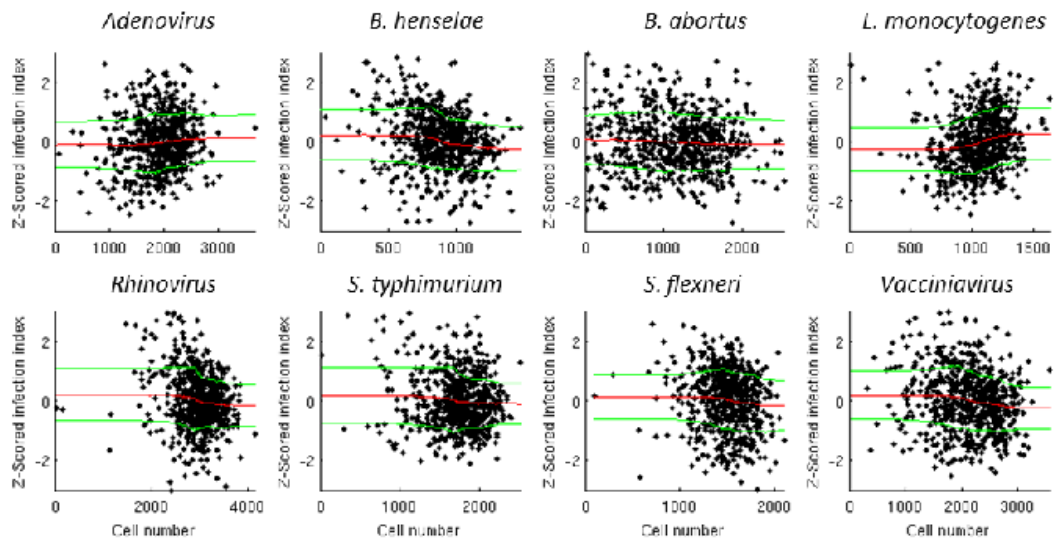
a) *Adenovirus* Dharmacon pooled, plate 3



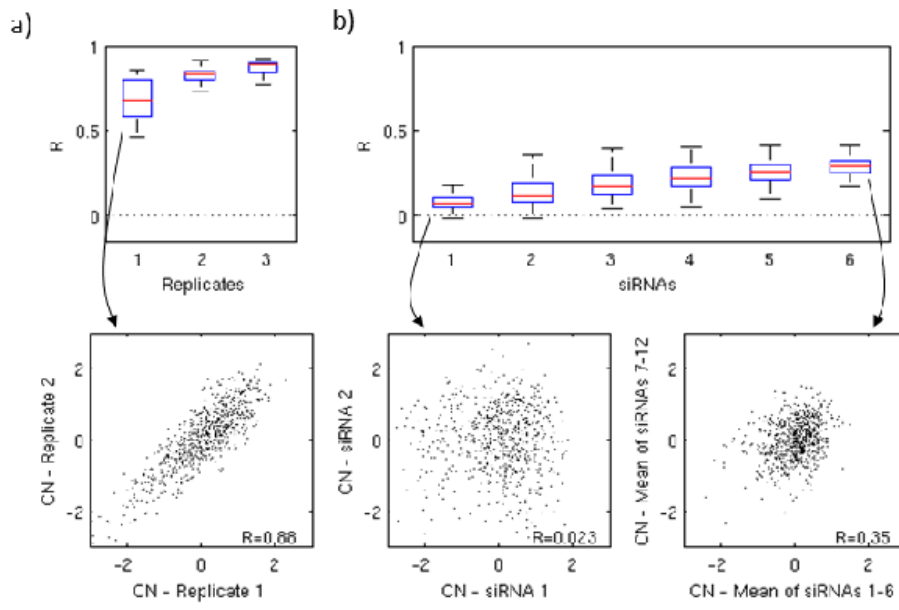
b) *Adenovirus* Dharmacon pooled, plate 3



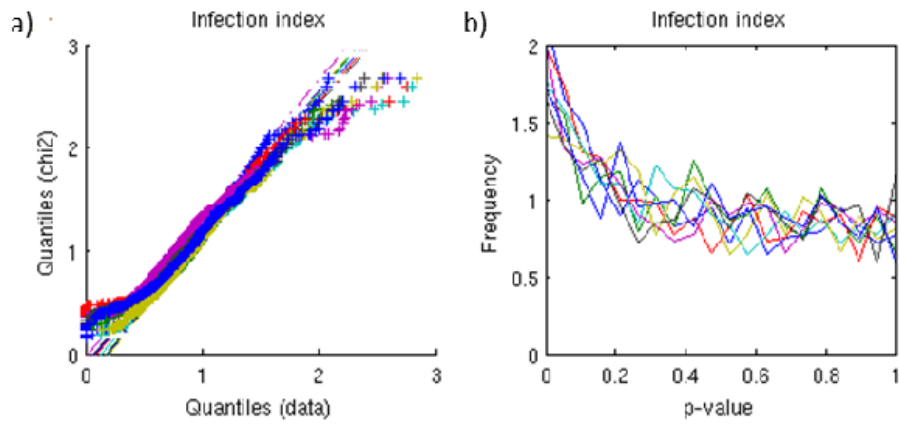
Supplementary Figure 3



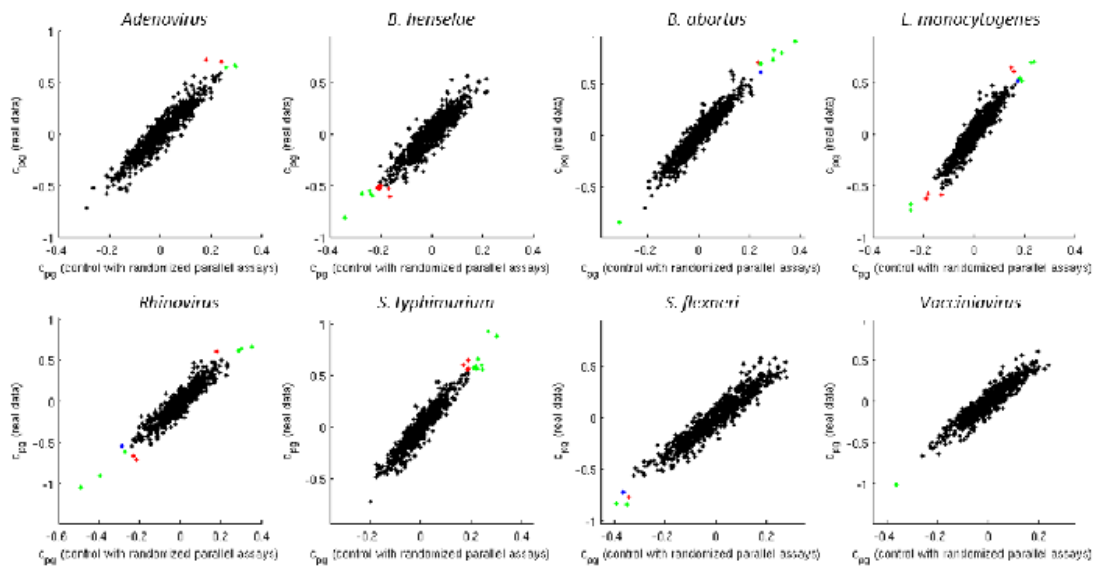
Supplementary Figure 4



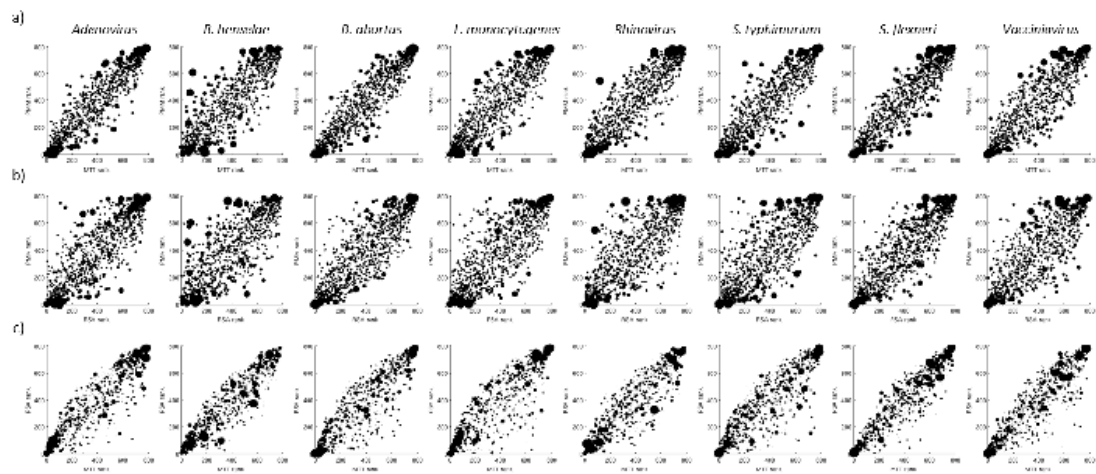
Supplementary Figure 5



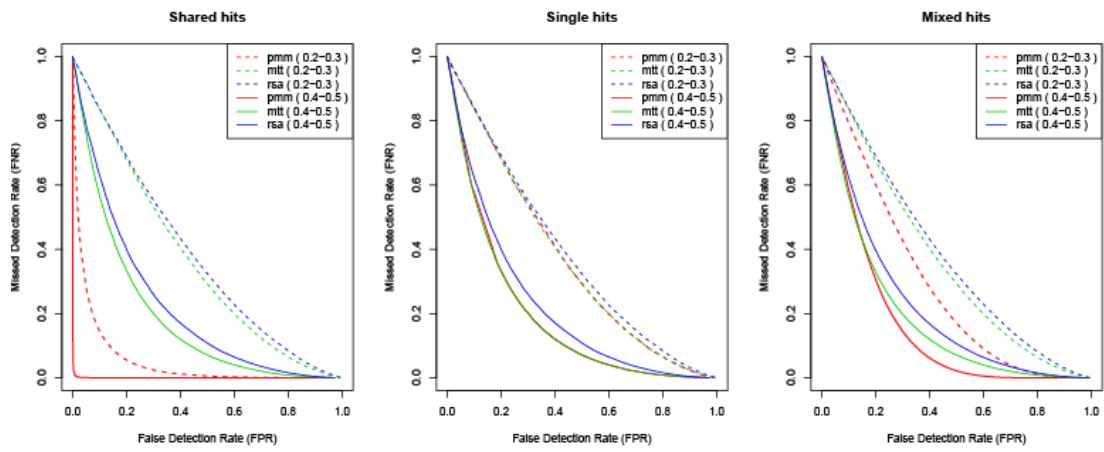
Supplementary Figure 6



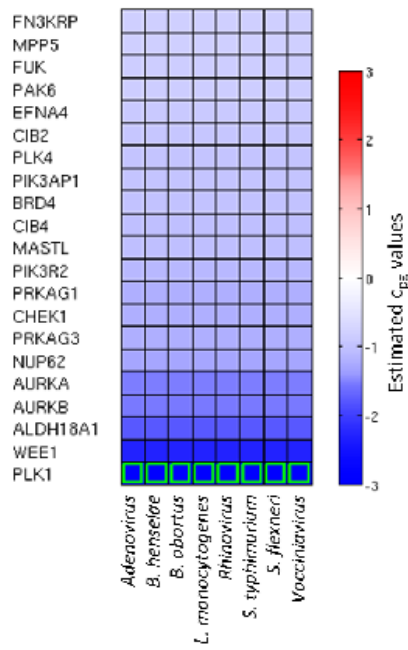
Supplementary Figure 7



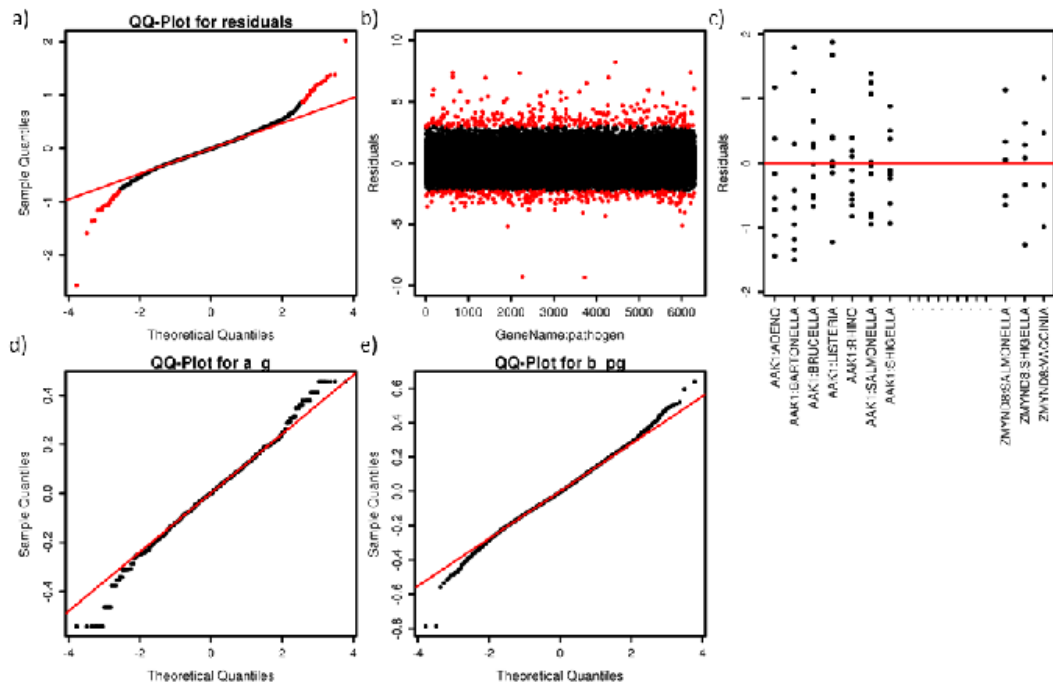
Supplementary Figure 8



Supplementary Figure 10



Supplementary Figure 11



3.3 RESEARCH ARTICLE III (in preparation)

Genome-wide siRNA screen in HeLa cells reveals host factors involved in *Brucella* infection

Shyan H. Low^{1*}, A. Casanova^{1*}, R. Conde-Alvarez³, M. Emmenlauer¹, P. Rämö¹, H.B.Tekaya¹, S. Muntwiler¹, S. Salcedo², J.P. Gorvel² and C. Dehio¹

* These authors contributed equally to this work

Manuscript in preparation

Statement of own contribution

Together with Dr. Raquel Conde-Alvarez, we developed the high-throughput microscopy based RNAi assay used in our siRNA screens. Screening with Qiagen druggable library was performed together with Dr. Raquel Conde-Alvarez and all other screens were performed together with Alain Casanova. Image analysis was performed by Mario Emmenlauer and with the help of Dr. Pauli Rämö, we were able to perform statistical analysis and gene interaction analysis on our screening data. Data normalization was done by Dr. Pauli Rämö. All figures were produced by me (except Figure 4 and Supplementary Figure 3 which were produced by Alain Casanova). Supplementary Figure 1 was generated using experimental data from both Alain Casanova and myself. Representation of the decision tree infection scoring in Figure 1 is provided by Mario Emmenlauer. Figure 4 represents results from the entry assay that was developed by Alain Casanova to study the early steps of *Brucella* infection. The manuscript was written by me and Alain Casanova.

3.3.1 Manuscript

Genome-wide siRNA screen in HeLa cells reveals host factors involved in *Brucella* infection

Shyan H. Low^{1*}, A. Casanova^{1*}, R. Conde-Alvarez³, M. Emmenlauer¹, P. Rämö¹, H.B.Tekaya¹, S. Muntwiler¹, S. Salcedo², J.P. Gorvel² and C.Dehio¹

¹ Focal Area Infection Biology, Biozentrum, University of Basel, Basel, Switzerland

² Centre d'Immunologie de Marseille-Luminy, INSERM-CNRS, France

³ Department of Microbiology, University of Navarra, Navarra, Spain

* These authors contributed equally to this work

Corresponding author: Prof. Christoph Dehio
Focal Area Infection Biology
Biozentrum, University of Basel
Klingelbergstrasse 50/70
CH-4056 Basel, Switzerland
Tel : +41-61-267-2140
Fax : +41-61-267-2118
E-mail : christoph.dehio@unibas.ch

Abstract

Brucella is an intracellular zoonotic pathogen that causes animal and human brucellosis worldwide. As natural hosts, *Brucella* infects various animal species including cows, goats, and pigs, causing abortion and birth of weak offspring. Humans are infected as incidental hosts, causing a febrile disease known as Malta fever that can develop into chronic infections with more severe symptoms such as endocarditis or meningitis. Therefore, brucellosis is a significant threat to the economy and general health in endemic areas. *Brucella* is able to invade phagocytic and non-phagocytic cells and replicates in an intracellular compartment known as the *Brucella*-containing vacuole (BCV). Following entry into a host cell, the BCV traffics along the endocytic pathway and despite interacting with endo-lysosomal compartments, degradation in these compartments are avoided. At later stages of infection when intracellular proliferation occurs, the BCV is found in close association with the endoplasmic reticulum (ER). Despite many advances in the field, the molecular mechanisms on how *Brucella* enters cells, avoids lysosomal degradation, and finally establishes an intracellular niche remain largely unknown. To study *Brucella* entry and replication in human cells, we performed a genome-wide, high-throughput microscopy-based RNA interference (RNAi) screen in HeLa cells. This allowed us to unravel host signaling pathways involved in *Brucella* infection, which includes actin-remodeling pathway, transforming growth factor (TGF- β) or fibroblast growth factor (FGF) signaling, ER-Golgi bidirectional transport, and some components of the endocytic pathway. To dissect the stage of infection that is regulated by these signaling pathways, a high-throughput entry assay was developed to study early stages of *Brucella* infection. We showed that TGF- β and FGF signaling pathways are involved in *Brucella* entry into non-phagocytic cells. Furthermore, we identified a novel host factor, Vps35 that is a component of the retromer complex involved in endosome to Golgi transport, to be involved in a post-entry process during *Brucella* infection.

Introduction

Brucella is a facultative intracellular zoonotic pathogen that infects humans as incidental host. *Brucella* causes animal and human brucellosis, with about 500,000 new cases of human brucellosis annually worldwide (1). This places *Brucella* as the most important zoonotic bacterial pathogen with *Brucella melitensis*, *Brucella abortus* and *Brucella suis* being the most common species that have been reported to cause human infections (2). Transmission of *Brucella* occurs via direct contact with infected livestock, ingestion of contaminated food products, or aerosol inhalation. Direct human-to-human transmission has not been reported thus far. In animal brucellosis, the infection of reproductive organs causes abortion or the birth of weak offspring. Human brucellosis on the contrary is associated with a febrile disease commonly known as Malta fever. Without treatment, *Brucella* can cause a chronic infection in various organs and lead to more severe symptoms such as endocarditis or meningitis (2). There is currently no effective vaccination for humans and even a complex antibiotic treatment for a prolonged duration is not able to completely protect against relapses (3). Therefore, *Brucella* remains a significant threat to the economy as well as public health in endemic areas.

Brucella infects phagocytic as well as non-phagocytic cells where bacteria replicate and persist inside the host. Bacteria adhere to the host cell surface via interaction with sialic acid residues that are present on eukaryotic receptors or bind to fibronectin and vitronectin (4, 5). Internalization then requires actin remodeling via activity of Rac, Rho, and direct activation of Cdc42 (6). Upon internalization, *Brucella* is contained within a vacuole termed *Brucella* containing vacuole (BCV) that interacts with the early endosomal markers Rab5, early endosomal antigen (EEA1), transferrin receptor (TfR), as well as flotillin-1, a component of lipid rafts (7-10). Next, BCV interacts with the late endosomal markers Rab7, Rab7's effector Rab interacting lysosomal protein (RILP), Lamp1, and transiently with autophagosomal markers (9, 11). Acidification of the BCV upon reaching a late endosomal compartment serves as a trigger for the expression of the VirB type IV secretion system (T4SS) (12, 13). The T4SS is believed to secrete yet unknown effectors that are essential for *Brucella* to avoid fusion with lysosomes, since VirB mutants are degraded in phagolysosomes (7, 14). *Brucella* that manage to divert from the endocytic pathway are then able to interact with the ER at ER exit sites (ERES) via interaction with the small GTPase

Sar1 and the COPII complex (7, 15). In addition, the small GTPase Rab2 was found to interact with the *Brucella* effector RicA. Rab2 controls vesicular-trafficking from Golgi to ER in the ER-Golgi intermediate compartment (ERGIC) and is required for fusion of BCVs with ER-derived vesicles and intracellular replication of *Brucella* (16, 17). This indicates that anterograde as well as retrograde trafficking components are required during infection. The VirB T4SS has also been shown to be important for a sustained interaction with the ER (7). Once in its replicative niche, *Brucella* requires the host factor inositol-requiring enzyme (IRE1-alpha) that regulates host cell unfolded protein response (18). To complete the infectious cycle, autophagy initiation proteins were recently shown to be vital for *Brucella* egression and cell-to-cell spreading (19).

Despite various efforts to understand the interaction of *Brucella* with host cells, relatively few host factors are known and many open questions remain. It is still unclear whether *Brucella* exploits host cell receptors to invade non-phagocytic cells and the role of T4SS effectors or their interacting partners at various stages of the intracellular life cycle of *Brucella* remains an open question. It is known that *Brucella* diverts from the endocytic pathway to reach an ER-derived compartment. However, the details of this process are still unclear. Furthermore, host factors that are needed for the maintenance of the BCV in its replicative niche are still largely unexplored. To understand in a systems level the host factors that are involved in *Brucella* entry and replication, we performed a genome-wide RNA interference (RNAi) screen targeting the human genome in HeLa cells. This revealed novel host pathways involved in *Brucella* infection. Efforts were taken to separate the identified components into different stages of the *Brucella* intracellular life cycle, allowing us to unravel Vps35, a novel host factor which is a component of the retromer complex to be involved in a post-entry step of the infection cycle.

Results

A high-throughput microscopy-based RNA interference (RNAi) assay for *Brucella* infection of human cells

To identify host factors involved in *Brucella* infection, we established a high-throughput microscopy-based RNAi assay in HeLa cells. In order to obtain optimal infection rates, infections were performed with an increasing multiplicity of infection (MOI) of GFP-expressing *Brucella abortus* 2308 and various time lengths of bacterial entry. As seen in Supplementary Figure 1, increasing the MOI as well as the time allowed for bacterial entry results in a corresponding increase in infection, with no saturation up to MOI 20000. For our screen, we used MOI 10000 of bacteria with 4 h of entry since this allows a sufficient infection rate (~2-10%) without reaching saturation of the system. The infection was allowed to continue for a total of 44 h. In unperturbed cells, this allows *Brucella* to traffic to an ER-derived compartment and intracellular replication leads to formation of a micro-colony (Figure 1B). The intracellular trafficking of bacteria under these conditions or with a lower MOI of 1000 as shown in Supplementary Figure 2 was similar and consistent with previous studies (7). *Brucella abortus* $\Delta virB9$ mutant and *Brucella abortus* acquire Lamp1 markers at 6 hpi. At 24 hpi, *Brucella abortus* $\Delta virB9$ mutant remains in Lamp1 containing endo-lysosomal compartment while *Brucella abortus* is excluded from this compartment (Supplementary Figure 2i and 2ii).

Figure 1A shows a summary of the experimental workflow used in our siRNA screens. Reverse siRNA transfection was performed for 72 h in HeLa cells after which cells were infected with GFP-expressing *Brucella abortus* 2308. After 4 h of infection, cells were washed with medium containing gentamicin to kill extracellular bacteria. Cells were fixed at 44 h post infection (hpi) and stained with DAPI and phalloidin-547 for nuclei and F-actin, respectively. Automated fluorescence imaging was performed and images were subjected to shading correction to correct for non-uniform illumination from the microscopes before image analysis was performed with CellProfiler (20). This allows objects such as the nucleus, perinucleus (8 pixels or 5.16 μm wide zone surrounding the nucleus), or cell body to be identified (Figure 1A and B) and features for example pathogen intensity (GFP intensity) to be extracted from each object. Decision tree based infection scoring was then performed using the extracted features (Figure 1B), giving single cell infection scores that could be used to

determine a well-based infection rate. In short, if the mean GFP intensity of any of the defined objects of a cell (nucleus, perinucleus, or cell body) exceeds a given threshold, the cell is considered infected. The thresholds are set in a way that only cells that contain proliferating bacteria will be detected as infected. Finally, to account for plate-to-plate variations, plate normalization was performed using Z-scoring as described in the materials and methods section.

Genome-wide siRNA screen for host factors involved in *Brucella* entry and replication in HeLa cells

To study on a systems level the interaction of *Brucella* with human host factors, primary genome-wide screens were performed in HeLa cells. Three replicates of the Dharmacon ON-TARGETplus SMART pool library and one replicate of the Qiagen Human Whole Genome siRNA Set HP GenomeWide siRNA library, both targeting the human genome, were screened. The Dharmacon library contains a single pool of four siRNA sequences targeting each gene while the Qiagen library comprises four individual siRNAs for each target.

To confirm the quality of our primary genome-wide screens, we compared the results from the independent replicates of the Dharmacon library as well as positive and negative controls present in all plates. As shown in Supplementary Figure 4i, we obtained good correlation (Pearson Correlation Coefficient $R = 0.5-0.7$) in both normalized infection index as well as normalized cell number between the independent replicates of our Dharmacon pooled library. Furthermore, positive controls from both Dharmacon and Qiagen libraries such as siRNAs targeting Rac1, Cdc42, and ATP6V1A that are known host factors for *Brucella* (6, 12) showed an expected reduction of infection (Supplementary Figure 3). Mock (transfection reagent only) and scrambled siRNA without specific host targets, both showed no effect on *Brucella* infection and cell number. The transfection controls Kif11 (Dharmacon) and AllStarsDeath (Qiagen) that are toxic to cells resulted in a strong reduction in cell number upon knockdown (Supplementary Figure 3). Altogether, this shows that with our experimental workflow we are able to obtain reproducible data between independent replicates and identify host factors that are involved in *Brucella* infection. To account for the well-known confounding off-target effects in siRNA screening, we performed statistical analysis of the primary screening data with the Redundant siRNA Analysis (RSA) algorithm (21). This analysis was performed separately on the

up and down hits of the screen. Details of the analysis can be found in the materials and methods section. RSA allows ranking of all siRNAs from different libraries targeting a given gene over all siRNAs in the screen. Genes targeted by different siRNAs that show a similar effect on *Brucella* infection are shifted towards a higher rank with a lower and more significant P-value while non-consistent effects from different siRNAs designed against the same target gene obtain a higher, less-significant P-value. This reduces the number of false positives caused by strong off-target effects of single siRNAs and favors genes with a reproducible effect from different siRNAs, indicating an on-target phenotype.

Using datasets obtained from screening with Dharmacon and Qiagen libraries as input for RSA analysis, we were able to rank genes according to their P-values. Figure 1C summarizes the general workflow that was used for analyzing the screening data and selection of genes for further validation. Some of the genes that were in the top ranks of our RSA analysis were selected for validation with additional siRNAs. To further prioritize the genes that are present in the RSA list of the primary screen, we performed gene ontology (GO) enrichment studies using the DAVID functional annotation database (22). As seen in Figure 2A, genes that appear in our top 200 RSA ranks for reducing *Brucella* infection upon knockdown shows GO enrichment terms of retrograde vesicle mediated transport from Golgi to ER, regulator of cellular component size, enzyme linked receptor protein signaling pathway, intracellular protein transport, regulation of actin filament polymerization, and phosphorylation. Figure 2B shows enrichment terms for genes that are in the top 200 RSA ranks for increasing *Brucella* infection upon knockdown. It includes the terms RNA processing, cell cycle, microtubule-based process and cytoskeleton organization. Genes present in the enriched pathways were also included for validation even if there were no strong phenotype shown in the primary screens. This screen was performed in the framework of InfectX, a consortium that aims at identifying the human infectome of several viral and bacterial pathogens. Therefore, we also included genes in our secondary screens that were selected by other pathogen groups that performed the same genome-wide screens. This strategy ensures sufficient negative controls on each screening plate that is needed for plate normalization.

Genome-wide RNAi screen reveals novel pathways involved in *Brucella abortus* infection

Secondary screens for all selected targets were performed with up to 3 siRNAs from the Ambion Silencer and Ambion Silencer Select unpooled libraries each as well as one esiRNA from the Sigma MISSION library. As shown in Supplementary Figure 4ii, we were able to obtain high correlation between the independent replicates of the secondary screens. Finally, RSA analysis was performed separately for both up and down hits, with the combination of all data from primary and secondary screens (Supplementary Table 1 - attached CD). Figure 3 represents the high confidence STRING database (23) interaction between top ranking genes from the RSA analysis that reduce and increase *Brucella* infection upon knockdown. We were able to confirm components that are known to be crucial for *Brucella* infection, e.g. subunits of the v-ATPase complex, Rab7A, Rac1, and Cdc42 (6, 11, 12). Furthermore, multiple components of pathways involved in TGF- β or FGF signaling, actin remodeling, endosome to Golgi transport, endocytic route, ER-Golgi bidirectional transport, proteasomal degradation, and clathrin-mediated endocytosis were found in our top ranking gene lists suggesting a role of these signaling pathways in *Brucella* infection.

Entry assay identifies Vps35, a component of the retromer complex, as a host factor involved in a post-entry process

To dissect the process regulated by the identified genes during *Brucella* infection, we developed an assay to study early steps of infection in a high-throughput format. The entry assay is based on the infection of HeLa cells with *Brucella abortus* that express GFP under a tetracycline inducible system and dsRed from a constitutive promoter. Since induction of GFP expression was performed in parallel to gentamicin addition to the medium, only intracellular bacteria were able to express GFP while extracellular bacteria were killed by gentamicin in the medium. As shown in Figure 4A, all *Brucella* expressed dsRed and only intracellular bacteria induced GFP expression. Bacteria were allowed to enter cells for 4 h, after which GFP expression was induced for another 4 h. This gave a sufficient signal above background to distinguish extracellular from intracellular bacteria using automated image analysis (data not shown). Image analysis was performed with a CellProfiler pipeline that detects the nucleus of cells (DAPI-stained) as well as single bacteria based on the

GFP signal. A voronoi cell body is calculated by extension of the nucleus by 25 pixels (16.125 μm) and decision tree based infection scoring separates infected from uninfected cells. An infected cell is defined by the presence of at least one bacterium of sufficient size and GFP intensity that overlaps with the voronoi cell body.

As shown in Figure 4B, most of the genes tested showed a direct correlation in infection between the entry assay and endpoint assay (indicated by the line of linear regression). This was the case for components from the TGF- β signaling, endocytic pathway, Golgi to ER transport, or the actin-remodeling pathway, indicating their role during *Brucella* entry into HeLa cells. This suggests that the reduced infection that is seen with the endpoint assay upon knockdown of these components is likely due to a perturbed entry of *Brucella* into HeLa cells. The negative controls RLUC (Renilla luciferase), AllStars and scrambled siRNAs, all with no specific host targets did not show an effect upon knockdown in the entry assay. Interestingly, Vps35 showed no effect on *Brucella* entry upon knockdown even though there was a significant reduction in infection rate with the endpoint assay. This suggests that Vps35 that is a key component of the retromer complex is involved in a post-entry step during *Brucella* infection in HeLa cells.

Discussion

Studies that have been performed thus far to understand *Brucella* interaction with the host were mainly hypothesis driven, includes small-scale RNAi screens in *Drosophila* S2 cell, or proteomics studies to identify host components of the BCV (7, 11, 15, 16, 18, 19). To identify at a systems level host factors involved in *Brucella* infection, we performed a genome-wide RNAi screen in HeLa cells. We identified novel host factors covering different signaling pathways, being able to separate some of our hits with an entry assay at early steps of infection. Furthermore, we found a component of the retromer complex to be involved in a post-entry process during *Brucella* infection. With the data from our primary screen, we could indeed identify enriched pathways of biological processes that were expected to be required for *Brucella* infection. These included the pathways involved in regulation of actin cytoskeleton or vesicular trafficking among others. We then validated a selected number of interesting candidate genes by seven additional independent siRNAs in a secondary screen. This validation strategy is based on a small study comprising the human kinases that was

performed within the InfectX consortium. In this study, it was shown that testing a sufficiently large number of different siRNA sequences for each gene is able to account for the differences in knockdown strength and specificity of individual siRNAs (24).

On the final data, we performed RSA analysis for genes that reduce or increase *Brucella* infection using results from both primary and secondary screens. For genes that increase *Brucella* infection, datasets with less than 500 cells were removed before RSA analysis. Since the infection rate positively correlates with the MOI (Supplementary Figure 1), we would expect that knockdowns that negatively affect cell number would increase infection. Therefore, due to the higher possibility of false positives in our up hits caused by cell number effects from siRNA toxicity, we only considered the top 200 genes that increased *Brucella* infection. For genes that reduced *Brucella* infection the top 400 were taken for our final pathway analysis with the STRING database. As shown in Figure 3, many of our top ranked up and down hits interact within the high confidence STRING database interaction network (23). The most prominent clusters include components of signaling pathways of actin-remodeling, TGF- β or FGF signaling, endosome to Golgi transport, endocytic pathway, ER-Golgi bidirectional transport, or clathrin coated pit components. Some of the individual components of these signaling pathways are known to be important for *Brucella* infection. Rab7A is needed for trafficking to the replicative niche (11), subunits of the v-ATPase complex for acidification of the BCV which serves as a signal for the expression of T4SS (12, 13), Rac1 and Cdc42 are involved in internalization into non-phagocytic cells (6), COPB subunit of the COPI complex was implicated in *Brucella* replication (16), and Sec61 has been shown to localize to the BCV during replication (9). This confirms the ability of our screen to identify essential hits needed during various stages of the *Brucella* infection cycle in the host. Next, we performed an entry assay to separate our hits into functional stages during the intracellular life cycle of *Brucella* (Figure 4). Components from actin-remodeling (Rac1, Cdc42, CYFIP1, NCKAP1, ACTR3), TGF- β signaling (TGFB1, TGFB2, Smad4), endocytic pathway (Rab7A), and Golgi to ER transport (COPG) show decreased *Brucella* entry and subsequent decrease in the formation of an intracellular micro-colony upon knockdown. Interestingly, Vps35 was the only component tested that is not involved in *Brucella* entry. Vps35 is a component of the retromer complex that regulates endosome to Golgi transport (25). Furthermore, other components of

the retromer complex including Vps26a and to a lesser extent Vps29 also showed reduction of *Brucella* infection upon knockdown. Since it is still unclear how *Brucella* traffics from an endocytic compartment to its ER-derived replicative niche, the retromer complex could provide a possible route via transient interaction with the Golgi. Alternatively, the retromer complex could be involved in the establishment or maintenance of the replicative niche. USP6NL, a Rab GTPase-activating protein (GAP) that is involved in Shiga toxin transport from endosomes to the trans-Golgi network by regulating Rab43 (26) led to an increase in *Brucella* infection upon knockdown. It is thus tempting to speculate that the regulation of Shiga toxin transport by USP6NL might be needed by *Brucella* in a similar manner. Taken together, these findings are in line with the notion that endosome to Golgi transport is required during the intracellular lifecycle of *Brucella* infection.

Studies with drug inhibitors have shown the importance of Rac1 and Cdc42 in *Brucella* infection (6) and the role of these factors was confirmed in our RNAi screens. In addition, we identified additional components of the actin-remodeling pathway that have not been described previously. As expected, knockdown of RACGAP1 that reduces levels of active Rac1 led to an increase in *Brucella* infection while ARHGEF9 that is an activator of Cdc42 decreased infection upon knockdown. Upstream or downstream components of Rac1 such as the WAVE complex (NCKAP1, CYFIP1, Abl1), the Arp2/3 complex (ArpC2, ArpC3, ACTR3, ACTR2), or kinases (PTK2B, CRK) that are involved in the formation of branched actin networks, lamellipodia and membrane ruffling are all shown to be required for *Brucella* entry into host cells. In addition, Cdc42 and its interacting partner TRIP10 or TNK2 that are involved in filopodia formation are also down hits in our screen. This further confirms the role of actin-remodeling networks regulated by Cdc42 or Rac1 during *Brucella* infection with an extension of host factors in this signaling process being identified.

We further identified anterograde as well as retrograde trafficking to be required for *Brucella* infection. Components of the COPI complex (COPG, COPB2, COPA, COPZ1, ARCN1 (COPD)) reduced *Brucella* infection upon knockdown. This is consistent with a previous study that found COPB depletion to reduce *Brucella* replication (16). In this same study, the authors reported that *Brucella* replication was affected by prolonged treatment with Brefeldin A that causes a redistribution of the Golgi to the ER, suggesting that Golgi to ER trafficking is important for *Brucella*

replication (16). However, studies carried out by other research groups showed that COPI-dependent transport is not required during *Brucella* infection (7, 15). In these studies, BCVs were shown in close vicinity of COPII components labeled with Sec31 antibody but not with COPI components labeled with anti- β -COP antibody as seen with immunofluorescence studies. Also, dominant negative ARF1 that regulates Golgi to ER transport did not affect bacterial replication in HeLa cells, similar to their previous results with Brefeldin A treated cells. These controversies might be explained by the different experimental settings in these studies. In the case of siRNA treatment over a prolonged period, many intracellular trafficking routes including pathways outside the ER - Golgi network might be affected. This could explain the results of the entry assay that showed an involvement of CPG1 in bacterial entry. In support of this hypothesis, a functional COPI complex is required during *Salmonella typhimurium* invasion in maintaining cholesterol, sphingolipids, Rac1, and Cdc42 at the plasma membrane (27). It is conceivable that *Brucella* entry follows a similar route as described for *Salmonella*, requiring COP components for membrane ruffling (27) and should be investigated with further studies. We also identified components of the COPII complex (Sec24 and Sec13) to reduce *Brucella* infection upon depletion. This is consistent with previous studies showing the importance of the COPII complex and ERES for the interaction of *Brucella* with the ER and subsequent replication (15, 16). Taken together, these findings show the importance of an intact bidirectional vesicular trafficking between the ER and Golgi for successful *Brucella* infection. However, the exact molecular details of the individual components have to be confirmed and likely involve additional cellular components that regulate other pathways, including the composition of the plasma membrane.

Members of the TGF- β and FGF signaling promote *Brucella* infection. TGF- β signaling components (TGFBR2, TGFBR1, TGFB1, TGFB2, Smad2) all led to a decrease in *Brucella* infection upon knockdown. The fact that both subunits of the heteromeric receptor complex, TGFBR1 and TGFBR2, showed an effect on *Brucella* infection upon individual knockdown validates the role of this receptor complex during infection. It has previously been reported that patients with brucellosis show higher levels of TGF- β 1 in their sera that is correlated with depressed function of T cell responses (28). B cells were also shown to produce TGF- β at early stages of infection with *Brucella* in mice (29). Therefore, this suggests a role of *Brucella* during infection in terms of immunosuppression of the host. However, since our RNAi

screen was performed in HeLa cells, it could be that TGF- β signaling has another non-immunological role during host cell infection, specifically during early steps as suggested by the results of the entry assay. It has been reported that *Trypanosoma cruzi* requires active TGF- β signaling during its invasion of mammalian epithelial cells, with cells having a dysfunctional intracellular cascade being deficient in parasite invasion (30). Components of the FGF signaling pathway (FGFR1 and FGF10) were also shown to reduce *Brucella* infection upon knockdown. It has been reported that FGF2 enhances *Chlamydia trachomatis* binding and uptake into non-phagocytic cells in a heparin sulfate proteoglycan dependent manner. The pathogen additionally stimulates production of FGF2 that enhances subsequent rounds of infection (31). Therefore, it would be interesting to investigate the roles of TGF- β or FGF signaling during *Brucella* invasion in HeLa in comparison to other cell types of interest such as trophoblastic cells or immune cells.

Among the host factors that restrict *Brucella* infection, we identified components of clathrin-coated pits. Depletion of several factors (AP2S1, CLTC, AP2A1, CLTA, EPS15L1, EPN1) caused an increase in *Brucella* infection in our screen. Our data suggests that *Brucella* prefers to enter the host via a clathrin-independent pathway in HeLa cells. This is in contrast to a recent publication by Lee *et al.* that showed reduction in *Brucella* infection upon siRNA treatment against CLTC or with inhibitor experiments using clathrin inhibitor, chlorpromazine (32). It remains unclear whether this controversy is due to differences in the experimental setup or the use of different bacterial strains. siRNA treatment experiments performed by Lee *et al.* focused on bacterial entry while our screen is an endpoint infection assay. The effect of clathrin component depletion towards bacterial entry into HeLa cells and its subsequent intracellular fate remains to be investigated in our studies. Validation with siRNA-independent approaches could also aid in addressing this controversy.

In summary, we are able to identify novel signaling pathways involved in *Brucella* infection. Most of the hits identified are involved in *Brucella* entry, with one example being the TGF- β signaling pathway that is required for entry into HeLa cells. Additionally, we were also able to identify a novel host factor, Vps35 that is involved in a post-entry process, probably in *Brucella* trafficking to its replication niche. In this study, we found and presented host factors with known biological functions that clustered in well-described pathways. Nevertheless, many additional genes were

identified to play a role in *Brucella* infection, providing a valuable resource for future discoveries.

Materials and methods

Wet lab procedures

Materials

RNAimax (Invitrogen, 13778-150); Dulbecco Modified Eagle Medium (DMEM) (Sigma, D5796); HeLa (human cervical carcinoma epithelial cell line, ATCC, CCL-2); Fetal Calf Serum (FCS) (Gibco, 10270): heat inactivated at 56°C for 30 min before use; tryptic soy broth (TSB) (Fluka, 22092); kanamycin sulfate (Sigma-Aldrich, 60615); gentamicin (Sigma, G1397); Triton-x-100, sigma-ultra (Sigma-Aldrich, T9284); DAPI (Roche, 10236276001); phalloidin-547 (Dyomics, 547PI-33); albumin from bovine serum (BSA) (Sigma, A9647); paraformaldehyde (Sigma, P6148); phosphate buffered saline (PBS) (Gibco, 20012).

Cloning of pAC42.08 for entry assay

pJC44 (11) was digested with EcoRI followed by generation of blunt ends with Klenov enzyme and subsequent digestion with SalI. TetR-GFP was amplified from pNF106 (unpublished) using primer prAC090 (TTTTTGAATTCTGGCAATTCCGACGTCTAAGAAACC) and prAC092 (TTTTTGTCGACTTTGTCCTACTCAGGAGAGCGTTC). Following digestion with SalI, the TetR-GFP product was ligated to the digested pJC44 vector. This generated a plasmid that constitutively expressed dsRed and a tetracycline-inducible GFP. The plasmid was then transferred into *Brucella abortus* 2308 by conjugation.

siRNA reverse transfection

Genome-wide screens were performed with Dharmacon ON-TARGETplus SMART pool and Qiagen Human Whole Genome siRNA Set HP GenomeWide (QU) siRNA libraries. For the validation screens Ambion Silencer, Ambion Silencer Select and Sigma MISSION esiRNA libraries were used. All experiments were conducted in a 384 well plate format. All plates contained general siRNA controls for transfection efficiency and toxicity (e.g. Kif11) as well as positive controls (e.g. Cdc42, Rac1) that are known to have an effect on *Brucella* infection (6). In addition, negative controls

such as mock (transfection reagent only) and scrambled (non-targeting siRNA) were added to each plate.

The following specifications apply to all siRNA screens except the QU siRNA library where specifications are given in brackets. 25 μ l (QU: 15 μ l) of RNAiMAX in DMEM without FCS (1:250 dilution) was added to each well containing 1.6 pmol siRNA (QU: 1 pmol) or 15 ng esiRNA. Screening plates were then incubated at room temperature (RT) for 1 h. Following incubation, 500 HeLa cells were added per well in a volume of 50 μ l (QU: 30 μ l) DMEM/16% FCS, resulting in a final concentration of 10% FCS. Plates were incubated at 37°C and 5% CO₂ for 72 h prior to infection.

Infection

Brucella abortus 2308 pJC43 (aphT::GFP)(15) were grown in TSB medium containing 50 μ g/ml kanamycin for 20 h at 37°C and shaking (100 rpm) to an OD of 0.8- 1.1. 50 μ l of DMEM/10% FCS containing bacteria was added per well to obtain a final MOI of 10000. Plates were then centrifuged at 400 g for 20 min at 4°C to synchronize bacterial entry. After 4 h incubation at 37°C and 5% CO₂, extracellular bacteria were killed by exchanging the infection medium by 50 μ l DMEM/10% FCS supplemented with 100 μ g/ml gentamicin. After a total infection time of 44 h, cells were fixed with 3.7% PFA for 20 min at RT.

For the entry assay, *Brucella abortus* 2308 pAC042.08 was used as described above. GFP expression was induced for 4 h by the addition of anhydrotetracycline (100 ng/ml) during the gentamicin killing of extracellular bacteria as described above.

Staining

Cells were washed twice with PBS and permeabilized with 0.1% Triton-x-100 for 10 min. After washing twice with PBS, 20 μ l of staining solution that contains DAPI (1 μ g/ml) and DY-547-phalloidin (1.5 U/ml) in 0.5% BSA/PBS was added to cells. For the entry assay, cells were not stained with DY-547-phalloidin. Cells were then incubated with the staining solution for 30 min at RT, washed twice with PBS, followed by final addition of 50 μ l PBS.

For Lamp1 colocalization experiment, HeLa cells on coverslips were permeabilized with 0.1% Triton-x-100 for 10 min at RT, washed with PBS before incubated with 0.5% BSA/PBS for 30 min at RT. Afterwards, cells were labeled for Lamp1 using mouse

monoclonal anti-Lamp1 [H4A3] antibody (1:100) and secondary antibody Alexa Fluor 546 Goat Anti-mouse IgG (1:100).

Imaging with high-throughput microscopy

Microscopy was performed with Molecular Devices ImageXpress microscopes. MetaXpress plate acquisition wizard with no gain, 12 bit dynamic range, 9 sites per well in a 3x3 grid was used with no spacing and no overlap and laser-based focusing. DAPI channel was used for imaging nucleus, GFP for bacteria, and RFP for F-actin or dsRed of bacteria in the entry assay. Robotic plate handling was used to load and unload plates (Thermo Scientific). The objective was a 10X S Fluor with 0.45NA. The Site Autofocus was set to “All Sites” and the initial well for finding the sample was set to “First well acquired”. Z-Offset for Focus was selected manually and manual correction of the exposure time was applied to ensure a wide dynamic range with low overexposure.

Image analysis

Object detection

Images were first scaled that pixel intensities of a full plate are in the 0 to 1 range. Images were then corrected for shading (flat field correction, vignetting correction) by applying a shading model to the image pixels. Shading-corrected images were stored in floating points to reduce the loss of information. Pathogen signal in the DAPI channel was removed to increase the quality of the nucleus segmentation. The pathogen signal was removed by subtracting a linear transformation of the GFP channel from the DAPI channel. After the pathogen signal reduction, DAPI images were stored in double precision to reduce loss of information. On the corrected images, object detection was performed using CellProfiler (20). Firstly, nucleus objects labeled “Nuclei” were segmented in the DAPI channel using OTSU’s method (CellProfiler module IdentifyPrimAutomatic). Secondly, peri-nuclear ring object labeled “PeriNuclei” was constructed by extending the nucleus object by eight pixels and removing the nuclear area from the extended nuclear area (CellProfiler modules ExpandOrShrink and IdentifyTertiary). Thirdly, a cell body object labeled “Cells” was segmented in the Actin channel using the “Propagation” method around the nucleus object (CellProfiler module IdentifySecondaryInformed). Finally, a non-actin

based cell body object labeled “VoronoiCells” was constructed by extending the nucleus object by twenty-five pixels (CellProfiler module ExpandOrShrink).

For the entry assay, the cell body was not stained with a fluorescent marker and only a voronoi cell body is used. Intracellular bacteria are detected using the GFP signal.

Feature extraction

On the segmented objects, measurements were performed using CellProfiler. On all segmented objects (Nuclei, PeriNuclei, Cells, VoronoiCells, Bacteria), shape measurements were extracted. Intensity and texture measurements were extracted with respect to all available channels (DAPI, Actin, Pathogen). All measurement result files of CellProfiler were stored in the openBIS database alongside the original images.

Infection scoring

Infection detection and measurement

Infection detection was done on a binary level (infected vs. non infected) that allows the infection index to be defined. The infection index is the number of infected cells / total number of cells in the well.

Decision Tree Infection Scoring (DTIS)

We selected a number of image analysis single cell features that were most sensitive to the infection phenotype. The N features are evaluated in a decision tree, which is a complete binary tree with N levels and 2^N nodes. Each node is evaluated by applying a threshold to the corresponding feature. During traversal of the tree, if the feature exceeds the threshold, evaluation continues with the one child, and if the feature does not exceed the threshold, evaluation continues with the other child. Nodes of the lowest level connect to one of the two distinct end states “infected” and “uninfected”. The connection of the nodes to children and the choice of features are performed once by an expert and remain static for all plates. The choice of the decision tree thresholds is affected by plate-specific parameters like quality of the staining, cell vitality and microscope illumination, and must be adjusted on a plate-by-plate basis. To quantify *Brucella abortus* infection for the endpoint assay, GFP

intensity was measured in the objects Nuclei, PeriNuclei, Cells and VoronoiCells using CellProfiler module MeasureObjectIntensity.

Segmentation based infection scoring for entry assay

Segmentation of pathogen objects in CellProfiler was used to detect pathogen colonies or single pathogens in the cell. This segmentation method was based on the OTSU method or wavelets. A cell is defined as “infected” if a pathogen object of at least 2 pixels and GFP intensity above the threshold overlaps with a voronoi cell body.

Data normalization

Z-Scoring

Several approaches have been described in the literature to correct the differences from wet lab procedures for plate batches (33). Negative controls (mock and scrambled siRNAs) sometimes show non-typical phenotypes (such as relatively high cell number) and good positive controls were not available for all primary siRNA screens. Therefore, non-control based data normalization methods were chosen for primary and secondary screens. Z-Scoring was used to normalize variations between plates as:

$$x_{new} = \frac{x_{old} - \mu}{\sigma}$$

Here, μ is the mean of all siRNA well readouts in the plate, σ is the standard deviation of all siRNA well readouts in the plate, x_{old} is the raw well readout and x_{new} is the normalized well readout. The non-control based normalization assumes that all genes are randomly distributed among all plates and that there are relatively few positive phenotype genes in the whole screen. For the entry assay, data were normalized to mock wells since the assumptions for Z-Scoring do not apply for assays that mainly contain hit genes.

Statistical analyses

Redundant SiRNA Analysis (RSA)

The Redundant SiRNA Analysis (RSA) ranks all siRNAs targeting a given gene over all siRNAs in the screens. It assigns the p-values for each gene based on a hypergeometric distribution that indicates whether the distribution of ranks of this

gene is shifted significantly towards low ranks (21). RSA was run using the R-package “RSA” release 1.3 (21) with parameters: $l=-1.5$ and $u=1$, where l refers to the threshold where a single siRNA readout is considered to be true positive at the low end and u refers to the threshold where a single siRNA readout is considered to be true positive at the high end.

References

1. Pappas G, Papadimitriou P, Akritidis N, Christou L, & Tsianos EV (2006) The new global map of human brucellosis. *The Lancet infectious diseases* 6(2):91-99.
2. Atluri VL, Xavier MN, de Jong MF, den Hartigh AB, & Tsolis RM (2011) Interactions of the human pathogenic *Brucella* species with their hosts. *Annual review of microbiology* 65:523-541.
3. Ariza J, *et al.* (2007) Perspectives for the treatment of brucellosis in the 21st century: the Ioannina recommendations. *PLoS medicine* 4(12):e317.
4. Castaneda-Roldan EI, *et al.* (2004) Adherence of *Brucella* to human epithelial cells and macrophages is mediated by sialic acid residues. *Cellular microbiology* 6(5):435-445.
5. Castaneda-Roldan EI, *et al.* (2006) Characterization of SP41, a surface protein of *Brucella* associated with adherence and invasion of host epithelial cells. *Cellular microbiology* 8(12):1877-1887.
6. Guzman-Verri C, *et al.* (2001) GTPases of the Rho subfamily are required for *Brucella abortus* internalization in nonprofessional phagocytes: direct activation of Cdc42. *The Journal of biological chemistry* 276(48):44435-44443.
7. Celli J, *et al.* (2003) *Brucella* evades macrophage killing via VirB-dependent sustained interactions with the endoplasmic reticulum. *The Journal of experimental medicine* 198(4):545-556.
8. Bellaire BH, Roop RM, 2nd, & Cardelli JA (2005) Opsonized virulent *Brucella abortus* replicates within nonacidic, endoplasmic reticulum-negative, LAMP-1-positive phagosomes in human monocytes. *Infection and immunity* 73(6):3702-3713.

Results: Research Article III

9. Pizarro-Cerda J, *et al.* (1998) Brucella abortus transits through the autophagic pathway and replicates in the endoplasmic reticulum of nonprofessional phagocytes. *Infection and immunity* 66(12):5711-5724.
10. Arellano-Reynoso B, *et al.* (2005) Cyclic beta-1,2-glucan is a Brucella virulence factor required for intracellular survival. *Nature immunology* 6(6):618-625.
11. Starr T, Ng TW, Wehrly TD, Knodler LA, & Celli J (2008) Brucella intracellular replication requires trafficking through the late endosomal/lysosomal compartment. *Traffic* 9(5):678-694.
12. Porte F, Liautard JP, & Kohler S (1999) Early acidification of phagosomes containing Brucella suis is essential for intracellular survival in murine macrophages. *Infection and immunity* 67(8):4041-4047.
13. Boschirolini ML, *et al.* (2002) The Brucella suis virB operon is induced intracellularly in macrophages. *Proceedings of the National Academy of Sciences of the United States of America* 99(3):1544-1549.
14. Comerci DJ, Martinez-Lorenzo MJ, Sieira R, Gorvel JP, & Ugalde RA (2001) Essential role of the VirB machinery in the maturation of the Brucella abortus-containing vacuole. *Cellular microbiology* 3(3):159-168.
15. Celli J, Salcedo SP, & Gorvel JP (2005) Brucella coopts the small GTPase Sar1 for intracellular replication. *Proceedings of the National Academy of Sciences of the United States of America* 102(5):1673-1678.
16. Fugier E, *et al.* (2009) The glyceraldehyde-3-phosphate dehydrogenase and the small GTPase Rab 2 are crucial for Brucella replication. *PLoS pathogens* 5(6):e1000487.
17. de Barsy M, *et al.* (2011) Identification of a Brucella spp. secreted effector specifically interacting with human small GTPase Rab2. *Cellular microbiology* 13(7):1044-1058.
18. Qin QM, *et al.* (2008) RNAi screen of endoplasmic reticulum-associated host factors reveals a role for IRE1alpha in supporting Brucella replication. *PLoS pathogens* 4(7):e1000110.
19. Starr T, *et al.* (2012) Selective subversion of autophagy complexes facilitates completion of the Brucella intracellular cycle. *Cell host & microbe* 11(1):33-45.

20. Carpenter AE, *et al.* (2006) CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome biology* 7(10):R100.
21. Konig R, *et al.* (2007) A probability-based approach for the analysis of large-scale RNAi screens. *Nature methods* 4(10):847-849.
22. Huang da W, *et al.* (2007) DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. *Nucleic acids research* 35(Web Server issue):W169-175.
23. Franceschini A, *et al.* (2013) STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic acids research* 41(Database issue):D808-815.
24. Ramo P, Drewek, A et al (submitted manuscript) Parallel Mixed Model for Simultaneous Analysis of Multiple Large-Scale RNAi Screens.
25. Wassmer T, *et al.* (2009) The retromer coat complex coordinates endosomal sorting and dynein-mediated transport, with carrier recognition by the trans-Golgi network. *Developmental cell* 17(1):110-122.
26. Fuchs E, *et al.* (2007) Specific Rab GTPase-activating proteins define the Shiga toxin and epidermal growth factor uptake pathways. *The Journal of cell biology* 177(6):1133-1143.
27. Misselwitz B, *et al.* (2011) RNAi screen of Salmonella invasion shows role of COPI in membrane targeting of cholesterol and Cdc42. *Molecular systems biology* 7:474.
28. Elfaki MG & Al-Hokail AA (2009) Transforming growth factor beta production correlates with depressed lymphocytes function in humans with chronic brucellosis. *Microbes and infection / Institut Pasteur* 11(14-15):1089-1096.
29. Goenka R, Parent MA, Elzer PH, & Baldwin CL (2011) B cell-deficient mice display markedly enhanced resistance to the intracellular bacterium *Brucella abortus*. *The Journal of infectious diseases* 203(8):1136-1146.
30. Ming M, Ewen ME, & Pereira ME (1995) Trypanosome invasion of mammalian cells requires activation of the TGF beta signaling pathway. *Cell* 82(2):287-296.
31. Kim JH, Jiang S, Elwell CA, & Engel JN (2011) *Chlamydia trachomatis* co-opts the FGF2 signaling pathway to enhance infection. *PLoS pathogens* 7(10):e1002285.

32. Lee JJ, *et al.* (2013) Interplay between clathrin and Rab5 controls the early phagocytic trafficking and intracellular survival of *Brucella abortus* within HeLa cells. *The Journal of biological chemistry* 288(39):28049-28057.
33. Birmingham A, *et al.* (2009) Statistical methods for analysis of high-throughput RNA interference screens. *Nature methods* 6(8):569-575.

Figure legends

Figure 1. Experimental workflow of high-throughput microscopy-based RNAi screen. A) Diagram illustrates the general workflow of our RNAi screen including wetlab procedures followed by image acquisition and analysis, infection scoring, and data normalization. B) Image on the left represents HeLa cells infected with GFP-expressing *Brucella abortus* with scale bar 50 μm . Segmentation of the cell body (white) as well as the nucleus surrounded by a perinucleus (light green) is shown in the middle. On the right, a graphical illustration shows the Decision tree based infection scoring. Decision tree infection scoring is performed using features that are extracted from identified objects, e.g. GFP intensity (pathogen intensity) in nucleus (f1), perinucleus (f2) and cell body (f3). A cell is considered infected if either one of these features exceeds a threshold that is manually determined. C) Workflow of gene selection for validation by additional siRNAs and analysis of the full data set.

Figure 2. Gene ontology (GO) enrichment terms of primary genome-wide RNAi screen using DAVID functional annotation database. GO terms that represent biological processes are shown. Bar graph shows $-\text{LogP}$ values of enrichment terms for top 200 RSA lists of A) genes that reduced *Brucella* infection upon siRNA knockdown or B) genes that increased *Brucella* infection upon siRNA knockdown. RSA was performed using individual siRNAs from Qiagen unpooled library combined with the average of three replicates of the Dharmacon pooled library. GO terms that cover at least five components of our list and with a P value lower than 0.05 are shown. A higher $-\text{LogP}$ value indicates a higher significance for the GO term shown.

Figure 3. Genome-wide RNAi screen reveals pathways involved in *Brucella* infection. Diagram represents host factors that are involved in *Brucella* infection. RSA was performed by combination of primary and secondary screening data. Individual siRNAs from the Qiagen library and the averages of independent replicates of the Dharmacon, Ambion, and Sigma libraries were used as input. To identify targets that increase *Brucella* infection upon knockdown, siRNA experiments with numbers less than 500 cells were removed before RSA analysis. To illustrate the interaction network within our hit lists, RSA top 400 genes that reduced *Brucella* infection and RSA top 200 genes that increased *Brucella* infection were added to the STRING database. The edges between genes indicate high-confidence (>0.9) STRING database interactions and only genes that contain at least one interacting partner are shown. Genes that reduced *Brucella* infection upon knockdown are surrounded by a blue outline, while a red outline indicates genes that increased infection. Nodes are colored based on their functional pathways.

Figure 4. Entry assay identifies the retromer complex component Vps35 as a host factor involved in post-entry process of *Brucella* infection.

A) Images in the upper row represent HeLa cells infected with *Brucella abortus* expressing GFP under a tetracycline inducible system and dsRed under a constitutive promoter. Cells were infected for 4 h followed by induction of GFP in intracellular bacteria for 4 h. Nuclei are stained with DAPI (blue). Scale bar represents 50 μm . Lower row shows CellProfiler based object segmentation of the nuclei (in white) as well as GFP positive bacteria (in pink). A voronoi cell body is calculated by extension of the nucleus by 25 pixels (in white). Decision tree infection scoring is used to separate infected (1) from uninfected (2) cells. Cells are considered infected if at least one segmented bacterial object with sufficient size and GFP signal overlays with the voronoi cell body. B) Scatter plot shows infection rates of the entry assay versus the endpoint assay, normalized to the mock dataset. Each point corresponds to the average of three replicates using a single siRNA or esiRNA for the targets indicated. With the exception of Vps35 (red), a direct correlation between the infection rates of both assays was observed as indicated by the linear of regression (dotted line).

Supplementary Figure 1. Infection rate increases in a MOI-dependent manner in HeLa cells. Bar graph represents infection index dependent on the MOI and the time

of infection prior to gentamicin treatment to kill extracellular bacteria. Data are normalized to dataset MOI 10000 and entry time of 4 h, the experimental condition used in our RNAi screens. Each dataset shows the mean \pm STDEV of three independent experiments.

Supplementary Figure 2. *Brucella abortus* Δ virB9 mutant interacts with the endo-lysosomal compartment at 24 hpi, while most *Brucella abortus* avoids this compartment, with low or high MOI of bacteria. i) Images represent infection of HeLa cells with i) GFP-expressing *Brucella abortus* Δ virB9 mutant or ii) GFP-expressing *Brucella abortus* at 6 hpi or 24 hpi, with MOI 1000 or MOI 10000. Samples were stained with Lamp1 antibody and images were taken with the 60x objective and FEI MORE with TIRF microscope. Image in stacks were deconvolved with HUVGENs remote manager and one represented slice around the middle of a stack is shown. Scale bar represents 10 μ m.

Supplementary Figure 3. Control plots for *Brucella* infection and siRNA transfection of genome-wide screens. A) Z-score normalized infection index of control siRNAs for *Brucella* infection used in the Dharmacon pooled (DP) and Qiagen unpooled (QU) genome-wide siRNA libraries. Whiskers and outliers of boxplot are calculated with the Tukey method. B) Cell number of siRNA transfection controls designed to kill transfected cells. Transfection with Kif11 or AllStarsDeath resulted in a median of 19 and 5 cells per well, respectively. Whiskers and outliers of boxplot are calculated with the Tukey method. C) Images represent HeLa cells infected with GFP-expressing *Brucella abortus* and stained with DAPI (blue). Scale bar represents 50 μ m.

Supplementary Figure 4. RNAi screen correlation plots. Graph shows i) Pearson correlation coefficient (R) of normalized infection index or normalized cell number between independent replicates of our primary screen with Dharmacon pooled library or ii) correlation of normalized infection index between independent replicates of individual siRNAs within the Ambion unpooled library.

Supplementary Table 1 (in attached CD). RSA list for genes that reduced (excel sheet 1) or increased (excel sheet 2) *Brucella* infection. Table shows list of genes that are ranked according to their LogP value. Score represents Z score normalized infection index from siRNA screens and count cells normalized represents Z score normalized cell number. Columns highlighted in yellow represent output from RSA analysis. Datasets from the primary screens (Dharmacon pooled and Qiagen unpooled libraries), secondary screens (Ambion and Sigma libraries) and kinome screens (Ambion library) were taken as input for RSA analysis, with independent replicates being averaged before analysis was performed. For some of the genes, only one replicate of the secondary screen is available as indicated in the column ‘experiment’. For genes that increase *Brucella* infection, score (Z score normalized Infection index) was inverted (e.g. 1 to -1) before RSA analysis was performed.

Acknowledgements

We would like to thank Swiss SystemsX.ch (InfectX) for funding. Shyan Huey Low and Alain Casanova were stipend holders of the ‘Fellowship of Excellence’ International PhD program.

Results: Research Article III

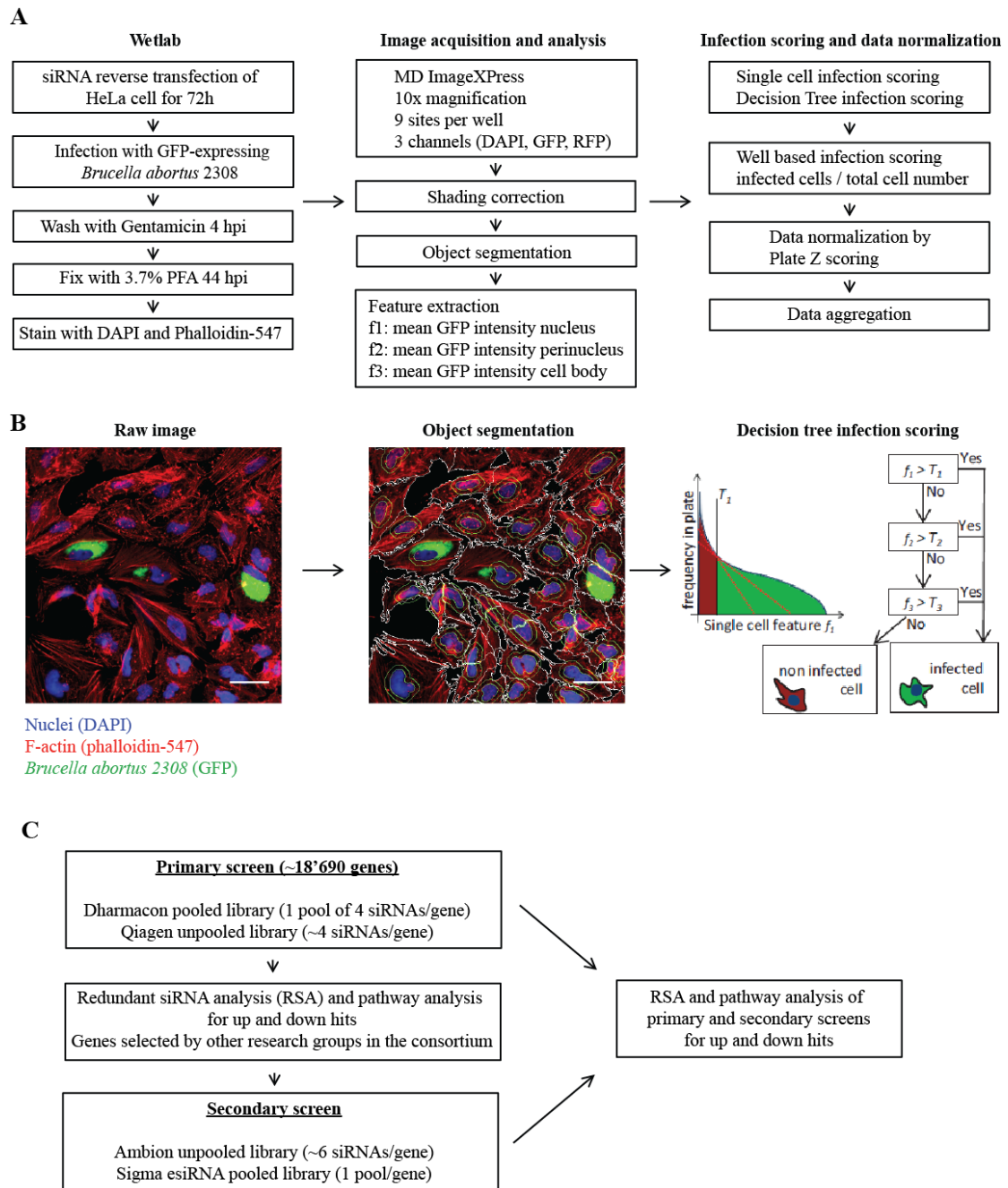
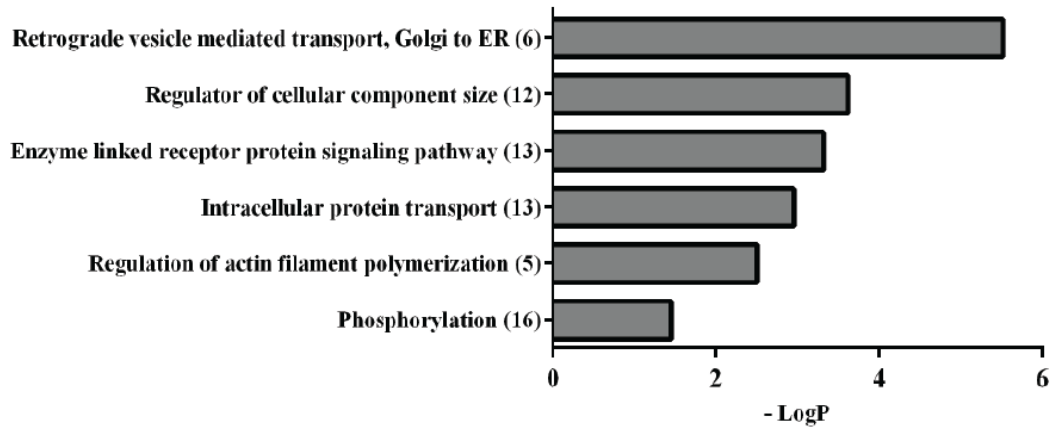


Figure 1

A



B

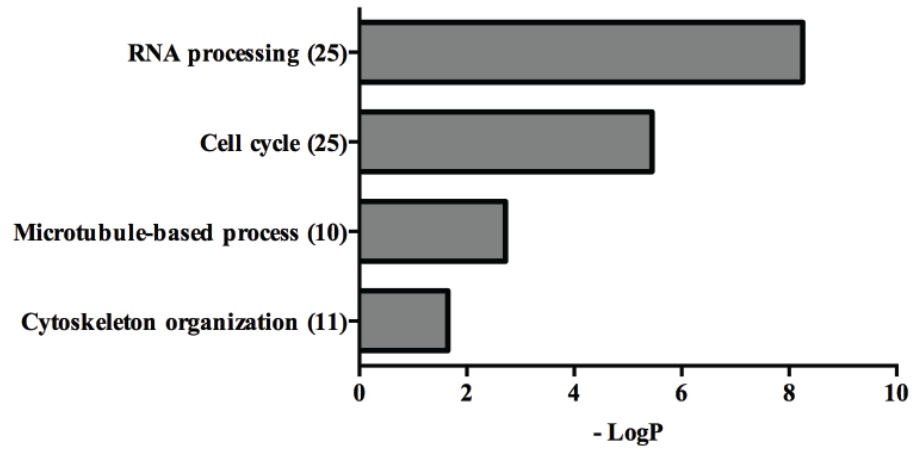


Figure 2

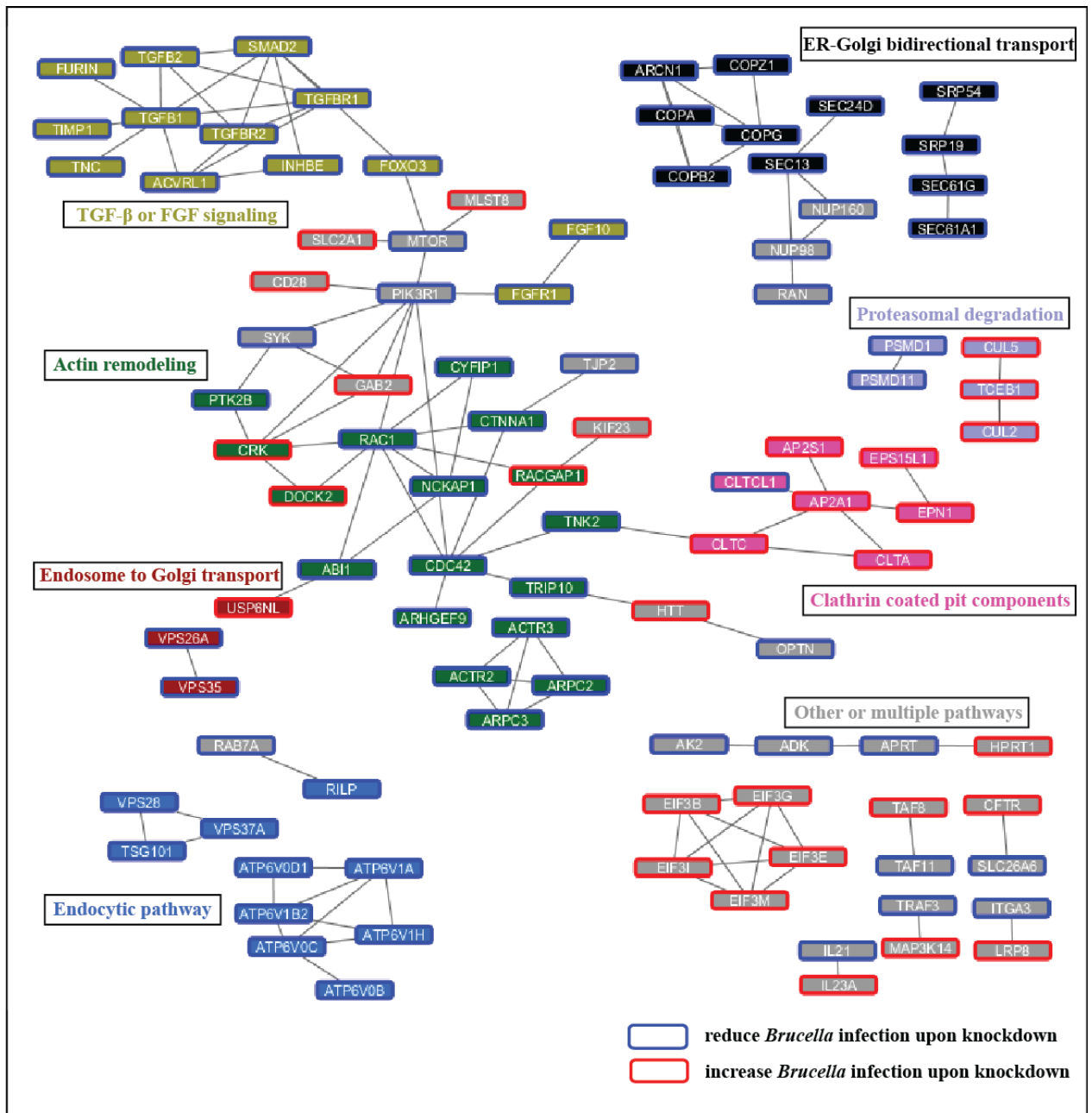
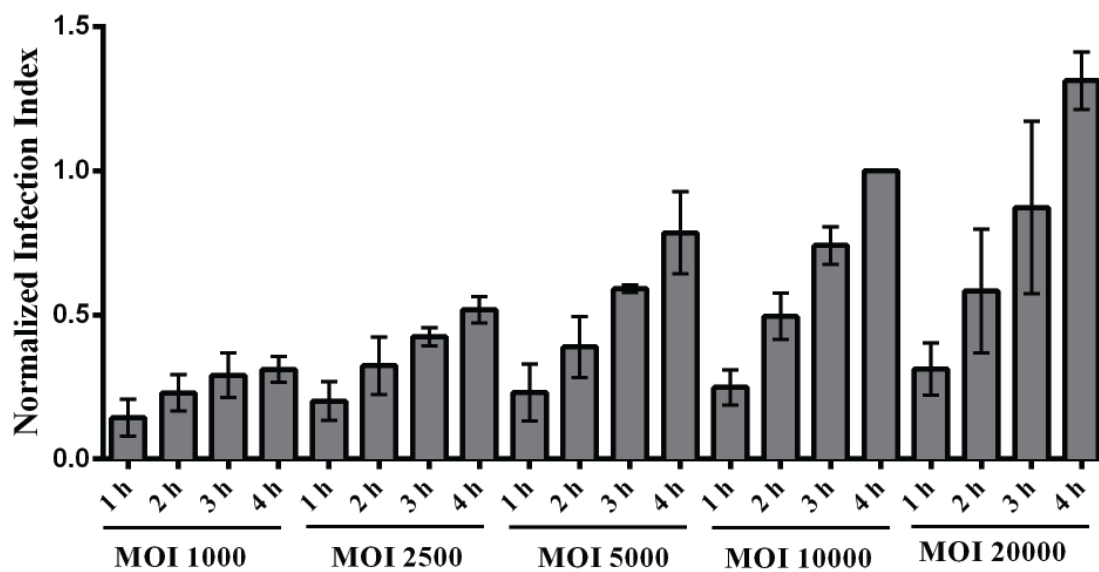
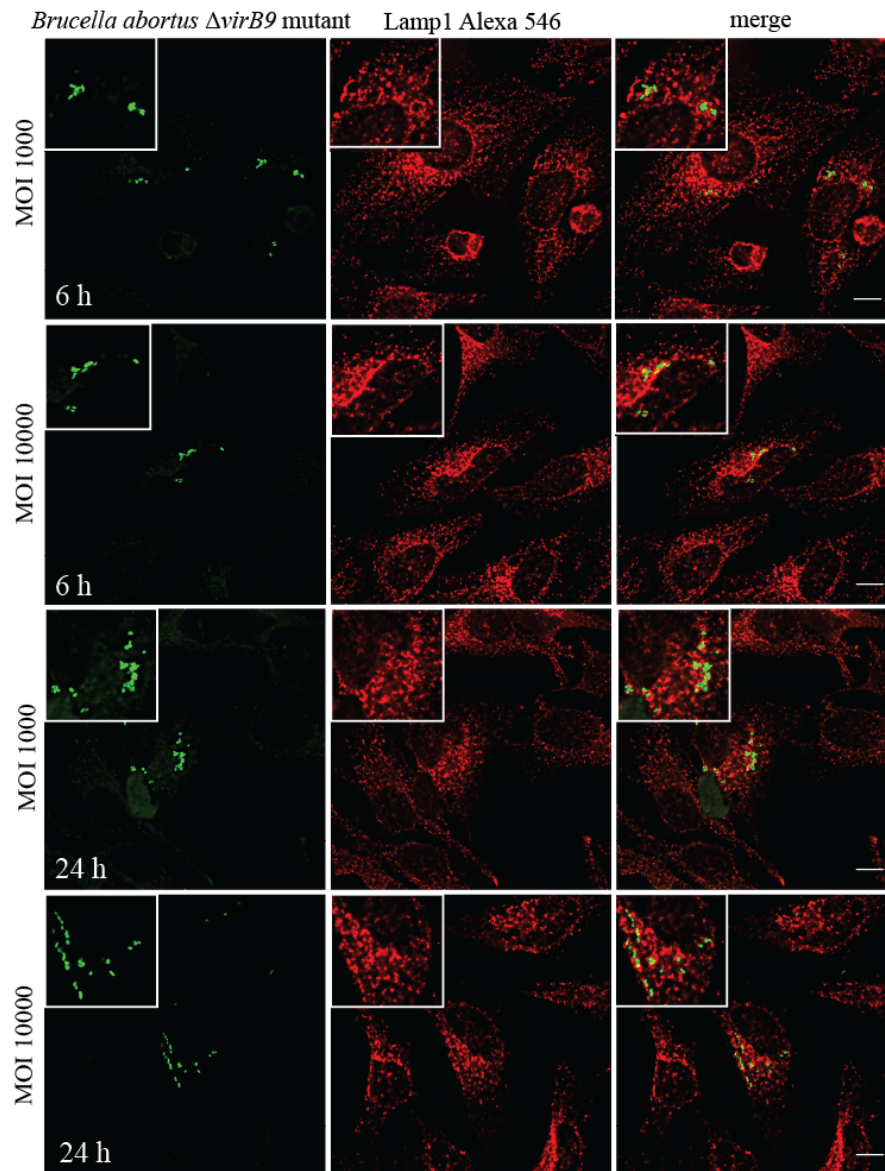


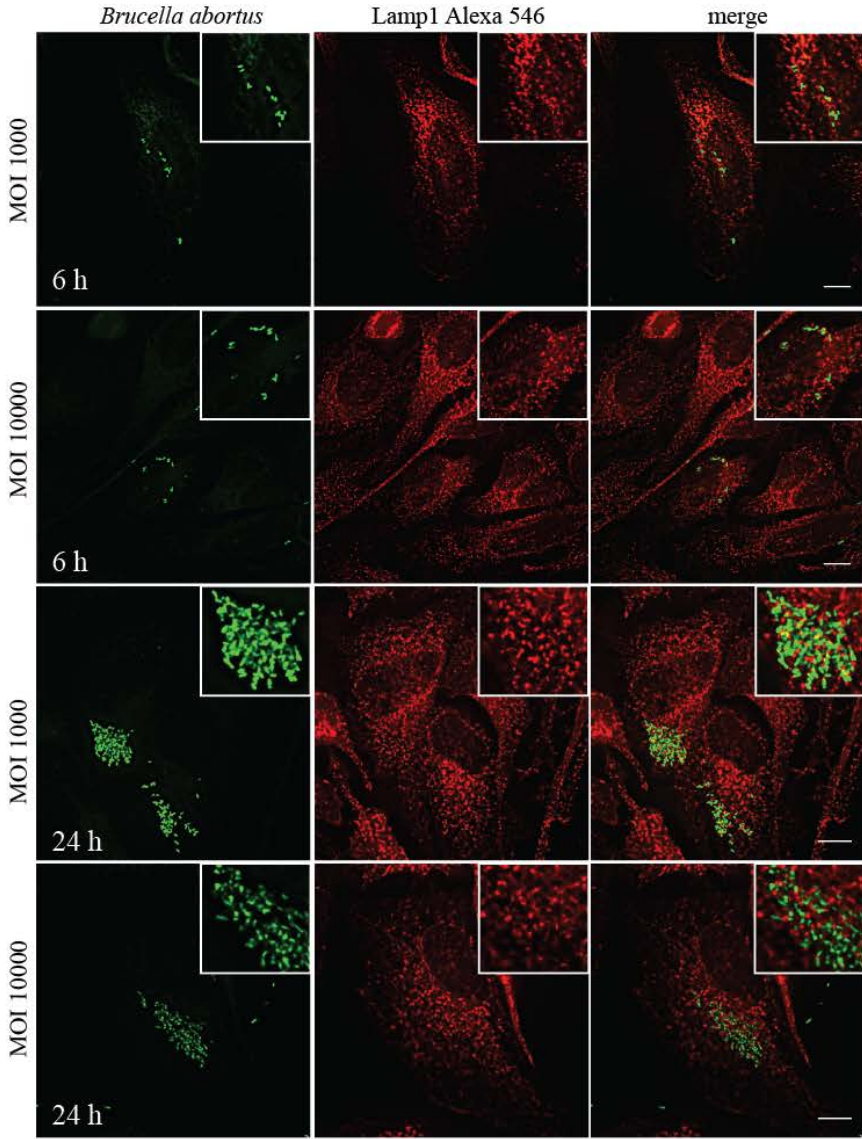
Figure 3



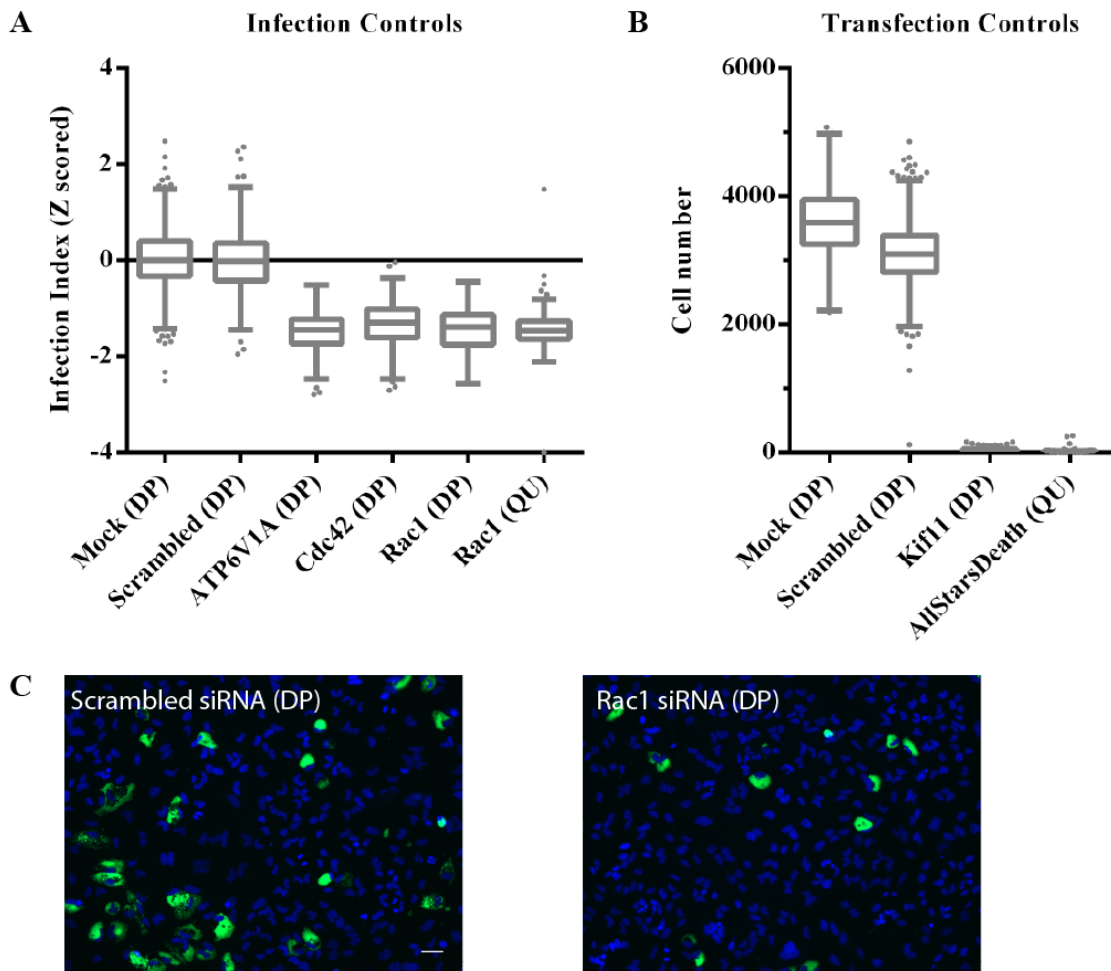
Supplementary Figure 1



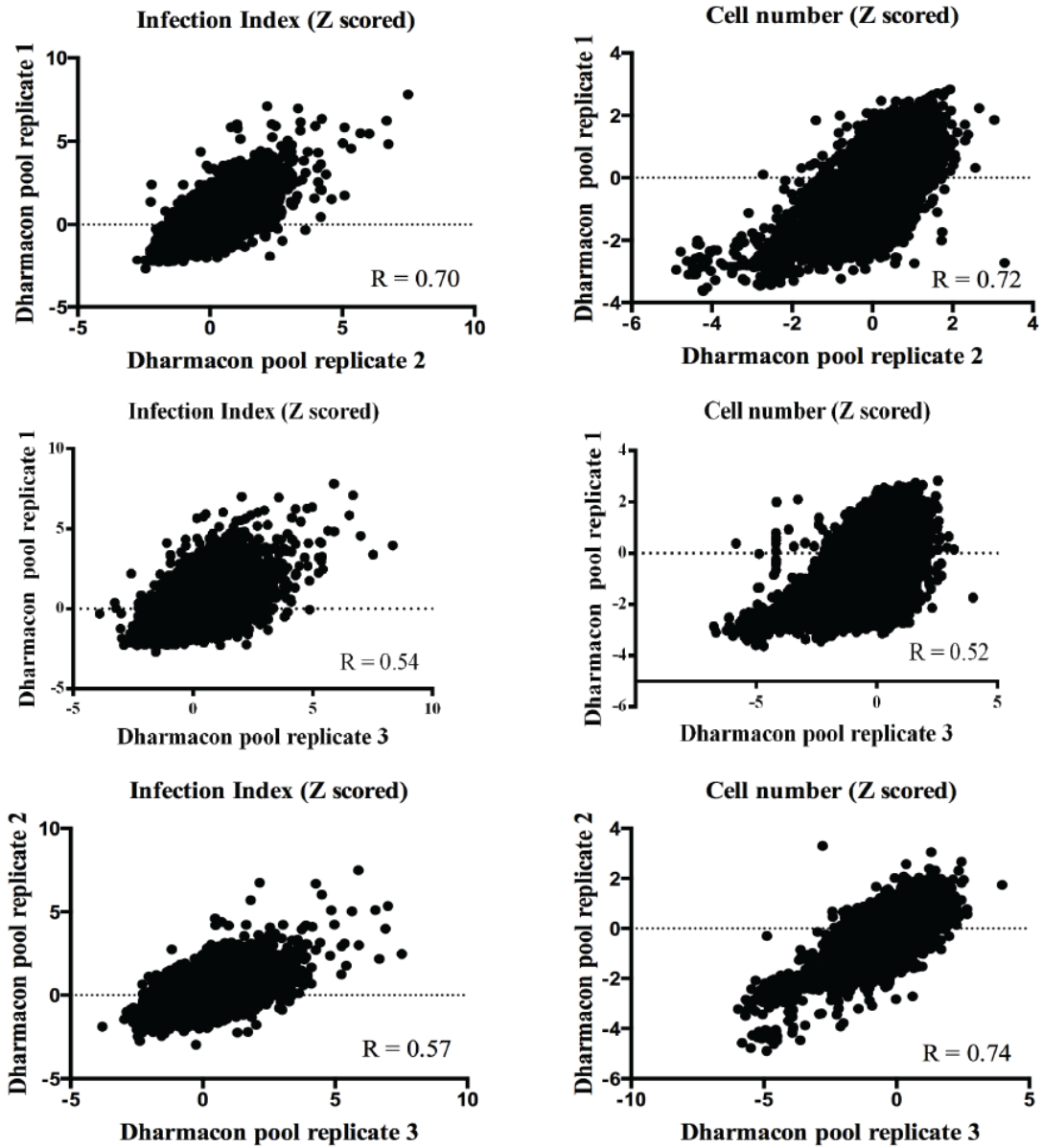
Supplementary Figure 2i



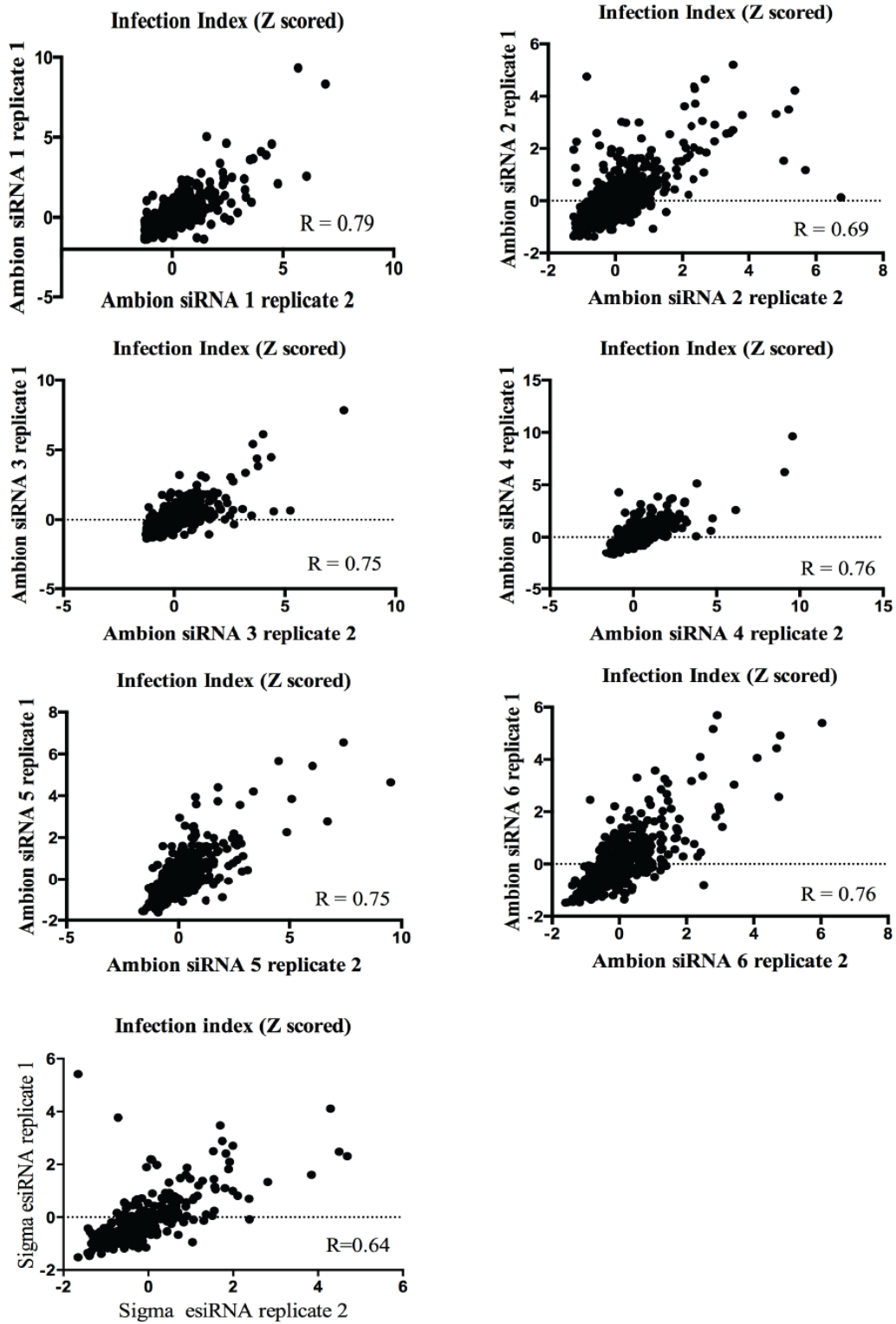
Supplementary Figure 2ii



Supplementary Figure 3



Supplementary Figure 4i



Supplementary Figure 4ii

3.4 Unpublished results: Transforming-growth factor beta (TGF- β) signaling and *Brucella* infection

Introduction

TGF- β signaling regulates various processes including proliferation, apoptosis, migration in differentiated cells through the control of the cytoskeletal machinery (3, 4), levels of cell adhesion protein receptors (5-8), endocytosis (9) as well as tumor suppression and progression (3). Being a powerful tumor suppressor, dysfunction of this pathway leads to a plethora of diseases such as cancer and tissue fibrosis (3). Therefore, tight regulation of this pathway at different levels is highly important (10). TGF- β signaling is initiated by binding of the ligand, TGF- β , to serine / threonine protein kinase type II TGF- β receptor (TGFBR2) on the cell membrane. There are three isoforms of TGF- β in mammals, TGF- β 1, TGF- β 2 and TGF- β 3, and association to TGFBR2 can be direct or mediated via the type III TGF- β receptor (TGFBR3). This leads to the formation of a heteromeric complex between TGFBR2 and type I TGF- β receptor (TGFBR1), in which TGFBR2 phosphorylates and activates TGFBR1 (11, 12). Activated TGFBR1 recruits and phosphorylates receptor-regulated Smad (R-Smad) proteins, Smad2/3, which then form a heterocomplex with the common mediator-Smad (Co-Smad), Smad4. The Smad complexes are then translocated into the nucleus to regulate transcription of target genes (13) (Figure 1). Effects of TGF- β on transcription can be negative or positive depending on the targeted gene and cellular context. Transcription factors, histone readers or modifiers as well as chromatin remodelers that bind to activated Smad proteins determine which genes are targeted (3).

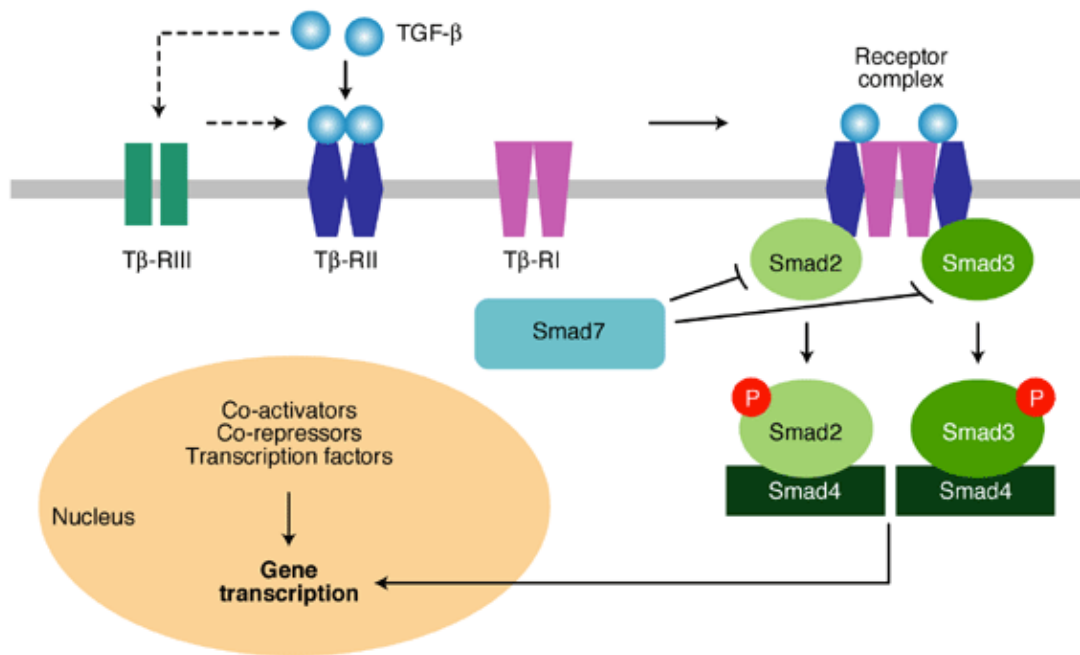


Figure 1 TGF- β signaling pathway. Picture taken from (14). Upon binding to ligand TGF- β , T β -RII (TGFBR2) forms complex with T β -RI (TGFBR1), phosphorylating Smad2 and Smad3. Interaction between T β -RII and T β -RI could also occur via T β -RIII (TGFBR3). Binding of phosphorylated Smad2/3 to Smad4 translocates them to the nucleus to activate gene transcription. Smad7 blocks Smad2/3 phosphorylation.

TGFBRs are constitutively internalized in a clathrin-dependent or lipid raft-dependent manner. Clathrin-dependent endocytosis brings TGFBRs to EEA1 or Rab5-positive endosomes. In these compartments, ortholog of the Smad anchor for receptor activation (SARA) and endofin molecules promote TGF- β signaling (15). However, in lipid rafts, caveolin-1 interacts with TGFBR1 and promotes its degradation in a Smad7-dependent manner and thus downregulates the signaling pathway (16). To this end, the inhibitory Smads (I-Smads), such as Smad7, interact with activated TGFBRs and recruit E3 ligases Smurf1 and Smurf2 that induces ubiquitination and degradation of the receptors (17-19). Smad proteins are also targeted for proteasome-mediated degradation via E3 ligases (20). In addition, TGFBRs can be degraded via the lysosomal pathway (21, 22). Apart from degradation, signaling can also be downregulated by dephosphorylation of TGFBR1 via the protein phosphatase 1 (PP1) and protein phosphatase PP2A, serving as a negative feedback mechanism (23, 24) for this pathway.

Results: TGF- β signaling and *Brucella* infection

As TGF- β signaling regulates a plethora of cellular signaling events such as gene transcription and immune response, it also plays a versatile role in the context of pathogenicity. Being an anti-inflammatory cytokine, it has been shown to promote survival of *Trypanosoma cruzi* in macrophages and *Leishmania* in mice (25, 26). In patients with chronic brucellosis, there was a correlation between increased TGF- β production and depressed function of T cell responses (27). B cells were found to produce TGF- β in murine brucellosis at early stages of infection, with B cell knockout mice being able to clear *Brucella abortus* very efficiently (28). Furthermore, TGF- β signaling is also reported to play a role apart from immune response during infections. Activation of TGFBRs was shown to be important for the entry and replication of *Trypanosoma cruzi* in mammalian cells (29, 30).

During infection, an increase in active TGF- β levels is often observed. This includes infections with bacterial pathogens like *Mycobacterium*, *Brucella* or *Chlamydia* (27, 31, 32), viral pathogens e.g. cytomegalovirus or hepatitis B virus (33, 34) and protozoans such as *Trypanosoma cruzi* or *Leishmania* (35, 36). Some pathogens were reported to activate TGF- β signaling by activation of latent TGF- β . For example, neuraminidase glycoprotein of Influenza A and B viruses that has been shown to directly activate latent TGF- β *in vitro* leads to the induction of host cells apoptosis (37). Furthermore, *Trypanosoma cruzi* and *Leishmania* have also been shown to activate latent TGF- β *in vitro* (38, 39).

In our kinome and genome-wide siRNA screens, TGF- β signaling components including surface receptors, ligands and downstream signaling components have been shown to be important for *Brucella* infection (Research Article II and Research Article III). Entry assay also identifies some of the TGF- β signaling components to be involved in *Brucella* entry in HeLa cells. Interestingly, TGF- β signaling pathway was also found to affect infection rates of most of the pathogens in the InfectX consortium (Research Article II), suggesting a general role that is widely used by different pathogens. In this results part, siRNA-independent experiments validated the importance of TGF- β signaling during *Brucella* infection in HeLa cells. We were able to show that active TGF- β signaling and the presence of kinase active receptors are important factors during *Brucella* infection in HeLa cells.

Results

Knockdown of TGF- β signaling components reduces *Brucella* infection in HeLa cells

From the kinome and genome-wide screens (Research Article II and Research Article III), it was shown that components of the TGF- β signaling pathway are involved in *Brucella* infection. Figure 1 shows datasets from all siRNA screens that were performed thus far. Knockdown of TGFBR1 or TGFBR2, but not TGFBR3 caused a strong reduction in *Brucella* infection, having a similar effect as our positive control Rac1, known to be important for *Brucella* infection (40). Knockdown of other components of the pathway including the downstream signaling molecules Smad2, Smad3, Smad4 or the upstream ligands TGFB1 and TGFB2 also decreased infection but had a milder effect compared to the receptors. Smad7 that inhibits TGF- β signaling pathway showed an increase in infection upon knockdown while mock (transfection reagent only) that was used as a negative control showed no effect on *Brucella* infection (Figure 1).

Activation of TGF- β signaling by TGF- β 1 increases *Brucella abortus* infection

TGFBR2 is activated upon binding of its ligand, TGF- β 1, which further triggers its downstream signaling components of the pathway (11, 12). To validate the role of TGF- β signaling in the context of *Brucella* infection, cells were pre-incubated with TGF- β 1 containing medium prior to the infection assay. As shown in Figure 2, there was a mild but significant increase in the levels of *Brucella* infection upon pre-incubation with TGF- β 1 containing medium. A titration of the amount of TGF- β 1 added to the medium was performed, with only a slight increase in *Brucella* infection with increasing concentrations of TGF- β 1. Infection reached saturation at 0.5ng/ml of TGF- β 1, with further increase in concentration not resulting in a higher efficiency of *Brucella* infection in HeLa cells. Mv1Lu is a cell line that is sensitive to TGF- β stimulation and is widely used to study the TGF- β signaling pathway (41). Flow cytometry analysis was performed to investigate the levels of surface exposed TGFBR2 on HeLa cells compared to Mv1Lu. As shown, levels of surface TGFBR2 in HeLa cells (1.9%) were very low compared to Mv1Lu (79.9%) (Supplementary Figure 3A). Despite the difference in surface levels of TGFBR2, *Brucella* infection

rates were not significantly different between HeLa and Mv1Lu cells (Supplementary Figure 3B). Our results indicate that active TGF- β signaling is beneficial during *Brucella* infection in HeLa cells, with higher surface levels of TGFBR2 in Mv1Lu having similar infection rates as in HeLa cells.

Overexpression of wild type or dominant negative TGFBR increases and reduces *Brucella* infection, respectively

To further understand the role of the TGF- β signaling pathway receptors in *Brucella* infection, cells were transfected with wild type TGFBR1 or TGFBR2 cDNAs, (pCMV5-TGFBR1 or pCMV5B-TGFBR2) either individually or in combination of both receptors. Flow cytometry analysis indicated an increase in surface levels of TGFBR2 upon overexpression of the constructs in HeLa cells (Supplementary Figure 4). Overexpression of the individual receptors TGFBR1 or TGFBR2 for 1.5 days before infection significantly increased *Brucella* infection in HeLa cells compared to non-transfected cells. Unexpectedly, co-expression of both receptors did not further increase infection rates compared to single receptor expression (Figure 3A).

In a second step, overexpression studies were also performed with dominant negative mutants of TGFBR1 or TGFBR2. TGFBR2 K227R (pCMV5B-TGFBR2 K227R) and TGFBR1 K232R (pCMV5B-TGFBR1 K232R) have a single mutation that renders them to be kinase dead (42, 43). TGFBR2 Δ cyt is also incapable of phosphorylating and activating TGFBR1 due to a stop codon that was introduced downstream of the transmembrane domain (44). To further the understanding of TGF- β signaling in *Brucella* pathogenicity, these mutants were expressed in HeLa cells prior to *Brucella* infection. There was a significant but mild decrease in *Brucella* infection observed upon individual overexpression of dominant negative TGFBR2 and dominant negative TGFBR1. Similar to co-overexpression of wild type forms of both receptors, overexpressing a combination of dominant negative TGFBR2 and TGFBR1 did not further decrease *Brucella* infection. Infection rates remained similar between HeLa cells expressing the individual receptors or the combination of both receptors (Figure 3B). These results suggest the importance of having functional receptors in the TGF- β signaling pathway during *Brucella* infection, with an increase of kinase active receptor expression leading to an increased *Brucella* infection.

shRNA knockdown of TGFBR1, TGFBR2 or Rab7 reduces *Brucella* infection in HeLa cells

To confirm the effect of TGFBR1 and TGFBR2 knockdown on *Brucella* infection in a physiologically relevant cell line for *Brucella* infection, we chose THP-1 cells for our studies. THP-1 is a human monocytic cell line derived from an acute monocytic leukemia patient and could be differentiated into a macrophage-like cell line via induction with phorbol myristate acetate (PMA). *Brucella* infects THP-1 macrophages very efficiently, with high levels of infection at a much lower multiplicity of infection (MOI) compared to HeLa cells (Supplementary Figure 2). Due to the low efficiency of siRNA transfection in macrophages, a shRNA knockdown system was developed. pLKO.3G is a replication-incompetent lentiviral vector which allows expression of shRNA sequences as well as GFP, allowing the monitoring of cells that harbour the shRNA sequence. pLKO.3G can be introduced efficiently into HeLa cells and THP-1 macrophages via lentiviral particles (Supplementary Figure 5), allowing stable expression of shRNA sequences of interest.

shRNA sequences targeting *TGFBR1*, *TGFBR2* or *Rab7* were selected based on siRNA sequences that had a strong effect on *Brucella* infection in HeLa cells from our siRNA screens (Research Article II and Research Article III) or validated shRNA sequences that were available from Sigma's website (Supplementary Table 1). As THP-1 is a human cell line, siRNA sequences from the siRNA screen that target the human genome could be used for shRNA design. *Rab7* was included as it is a protein known to be important for *Brucella* trafficking to its replication niche in both HeLa cells and macrophages (45) and serve as a positive control for this assay.

The assay was first developed in HeLa cells before using the THP-1 system to confirm the efficiency of shRNAs designed. To this end, western blots and immunofluorescence studies were performed. As seen in Supplementary Figure 1A, shRNA sequence 1 and 3 targeting *Rab7* showed a significantly reduced *Rab7* staining compared to neighbouring cells which do not express the shRNA. shRNA sequence 2 expressing cells however did not show any reduction in *Rab7* levels. TGFBR2 antibodies available in the lab were not suitable for immunofluorescence staining (data not shown). Reduction in *Rab7*, TGFBR1 or TGFBR2 levels were also confirmed by western blot analysis for *Rab7* shRNA sequence 3 and a few of the

shRNA sequences of TGFBR1 (shRNA 1 and shRNA 1 sigma) and TGFBR2 (shRNA 1 sigma) (Supplementary Figure 1B).

Next, HeLa cells stably expressing shRNAs of interest were infected with DsRed-expressing *Brucella abortus*. Consistent with the results seen with the immunofluorescence studies, cells expressing Rab7 shRNA sequence 1 and 3 showed a significant decrease in *Brucella* infection while shRNA sequence 2 had no effect on infection (Figure 4 and Supplementary Figure 1A). A significant reduction in *Brucella* infection was also seen in cells expressing TGFBR1 shRNAs as well as TGFBR2 shRNAs. This confirms our genome-wide siRNA results that TGFBR1 and TGFBR2 are indeed important for *Brucella* infection in HeLa cells (Figure 4).

shRNA knockdown of Rab7 but not TGFBR1 reduces *Brucella* infection in THP-1 macrophage-like cell line

Using the shRNA sequences that showed an effect on *Brucella* infection in HeLa cells, shRNA knockdown was repeated in THP-1 macrophage-like cell line. Cells stably expressing Rab7 or TGFBR1 shRNAs were infected with DsRed-expressing *Brucella abortus*. TGFBR2 shRNA knockdown was not performed in this experiment. Rab7 shRNA 1 and 3 showed a mild decrease in *Brucella* infection in THP-1 cells as compared to the effect seen in HeLa cells, while there was no effect on *Brucella* infection upon knockdown of TGFBR1 in THP-1 cells (Figure 4). This confirms the role of Rab7 in *Brucella* infection of macrophages, with TGFBR1 knockdown having no effect on infection in this cell line.

***Brucella abortus* are not contained within Lamp1 or Rab7 compartments upon knockdown of TGFBR1 in HeLa cells**

Brucella interacts with the endocytic markers e.g. Rab7 and Lamp1 prior to its arrival at its replicative niche (45). Once arriving at its niche, *Brucella* is contained in an ER-derived compartment that is devoid of endosomal or lysosomal markers (46). As shown in Research article III, TGFBR1 and TGFBR2 are important for *Brucella* entry in HeLa cells. However, TGF- β signaling pathway could also be important in steps to reach their replicative niche post entry, e.g. the escape from the endo-lysosomal compartment. To understand this, cells depleted of TGFBR1 were infected with GFP-expressing *Brucella abortus* for 24 h, fixed and stained for Lamp1 that marks the

endo-lysosomal compartment. At 24 hours post infection (hpi), *Brucella* containing vacuoles (BCVs) should be negative for markers of this system and have already arrived at their replicative niche (7). As expected, cells treated with scrambled siRNA showed BCVs devoid of Lamp1 markers at 24 hpi. The same was seen with cells treated with TGFBR1 siRNAs, with most bacteria having no Lamp1 around their vacuole (Figure 5). TGFBR1 depleted HeLa cells were also stained for Rab7 at 24 hpi to check for its interaction with the late endosomal compartment. Again, at 24 hpi, most of the cells treated with scrambled siRNA and TGFBR1 siRNA did not acquire Rab7 marker around the BCV (Figure 6). This suggests that knockdown of TGFBR1 did not affect *Brucella* escape from the endo-lysosomal compartment.

Discussion and outlook

TGF- β signaling is involved in the regulation of various cellular processes including cell proliferation, development, immune system modulation and cell differentiation (3). Nevertheless, TGF- β signaling has also been shown to be important for pathogens in the course of infection, in terms of host immunosuppression for survival in macrophages as well as entry into mammalian cells (25, 26, 29, 30). Here, we report the importance of TGF- β signaling in *Brucella* entry and infection in non-phagocytic HeLa cells.

Knockdown of components of the TGF- β signaling pathway including receptors, R-Smads, Co-Smad as well as their upstream ligands resulted in a decrease in *Brucella* infection in HeLa cells. TGFBR2 and TGFBR1 forms a heteromeric complex on the cell surface upon pathway activation, leading to downstream signaling. The fact that there was a similar effect upon knockdown of each of these receptors individually validates the importance of this receptor during *Brucella* infection. The importance of TGF- β signaling in *Brucella* infection was further confirmed by a mild dose-dependent increase in infection upon pre-sensitization of HeLa cells with active TGF- β 1 ligand. This suggests that it is beneficial to have active TGF- β signaling pathway in the host cell prior to *Brucella* infection. However, the effect of TGF- β 1 towards *Brucella* infection is much milder in HeLa cells compared to that reported for *Trypanosoma cruzi* infection of Mv1Lu cells (29). There was a 6 fold increase in infection upon pre-incubating Mv1Lu cells for 24 h with TGF- β 1 (2ng/ml) before

infection with *T. cruzi*, while there was only a 1.5 fold increase with the same conditions in HeLa cells with *B. abortus*. Furthermore, there was no further increase in *Brucella* infection with higher doses of TGF- β 1 (up to 10ng/ml, data not shown). This could be due to the difference in sensitivity towards TGF- β 1 stimulation between Mv1Lu (TGF- β -sensitive cell line) (41) and HeLa cancer cell line. Since TGF- β 1 is a tumour suppressor, it is known that cancer cells reduce TGF- β signaling via reduced TGFBR expression, with a more abundant cytosolic compared to membrane pool of receptors (47, 48). This is consistent with our studies that HeLa cells express much less surface TGFBR2 compared to Mv1Lu (Supplementary Figure 3). Therefore, adding more TGF- β 1 to HeLa cells probably does not increase *Brucella* infection due to the limited number of surface TGFBR or limited downstream signaling that is available in this cancerous cell line. It could also be that TGF- β signaling has a more prominent role during *T. cruzi* infection than *Brucella abortus* infection of mammalian host cells, leading to a more significant increase of *T. cruzi* infection upon TGF- β 1 addition. Furthermore, we did not observe a higher infection rate in Mv1Lu as compared to HeLa cells (Supplementary Figure 4a) even though Mv1Lu has much higher levels of surface TGFBR2. It is possible that the system needs an additional activator, e.g. TGF- β . Additionally, these results might also suggest that *Brucella* does not attach to or enter host cells by TGFBR itself in Mv1Lu but rather requires the signaling events from this pathway. The role of TGFBR or TGF- β signaling in Mv1Lu during *Brucella* infection has also still to be confirmed.

Another line of evidence to support the role of TGF- β 1 signaling during *Brucella* infection, specifically the role of the TGFBRs, was observed when overexpression of TGFBR1 or TGFBR2 significantly increased *Brucella* infection and kinase dead dominant negative TGFBR mutants reduced *Brucella* infection. This emphasizes the importance of TGFBR kinase activity during *Brucella* infection. Again, only a mild increase in infection was observed (1.5 fold) upon overexpression of TGFBRs, suggesting other limiting factors of the system (e.g. limited amount of ligand) that prevent further increase in *Brucella* infection. Co-expression of both TGFBR1 and TGFBR2 did not seem to increase infection rates. Since there was no fluorescent labeling for TGFBR expression in mammalian cells, estimation of the efficiency of double transfection in the cells remains challenging. Possibly, double transfection was not as efficient as single transfection, thus masking expected additive effects.

Therefore, it would be interesting to assay infection rates on cells co-expressing both receptors in the presence of TGF- β signaling agonist, TGF- β 1. In order to gain deeper insights into the effect of TGF- β 1 agonist stimulation or overexpression of the receptors on the activation of TGF- β 1 signaling, we could perform experiments with HeLa cells or Mv1Lu cells stably transfected with a plasminogen activator inhibitor-1 (*PAI-1*) construct. This plasmid contains a truncated *PAI-1* promoter fused to the firefly luciferase reporter gene that leads to increased luciferase activity upon TGF- β stimulation (49) as a read-out for increase of TGF- β signaling.

shRNA knockdown of TGFBR2 or TGFBR1 in HeLa cells reduced *Brucella* infection while no significant effect was detected in THP-1 macrophage-like cell line. As the effect that was seen with the positive control Rab7 shRNA knockdown is milder in THP-1 than HeLa cells, it is possible that the overall knockdown efficiency is lower in THP-1 and has to be confirmed with western blot studies. The efficiency of the shRNA system to knockdown proteins of interest in HeLa cells also suggests that it is possible to use siRNA sequences of interest for shRNA design, with varying knockdown efficiencies between different sequences.

TGFBR2 and TGFBR1 were shown to be important for *Brucella abortus* entry into HeLa cells (Research Article III Figure 4). To test for other possible roles of these proteins downstream of entry, the interaction of BCVs with endo-lysosomal markers was investigated. *Brucella* did not acquire either Rab7 or Lamp1 markers at 24 hpi, suggesting it is still able to escape from the endocytic pathway to reach its replicative niche even under TGFBR1 depletion conditions. Therefore, TGF- β signaling benefits pathogenicity mostly on the level of *Brucella* entry. This is still to be confirmed using other markers such as early endosomal, ER or autophagosomal markers. The role of TGF- β signaling in the adhesion of *Brucella* on host cells also remains yet to be investigated.

In summary, together with the results from Research Article III, we show that active TGF- β signaling and kinase active receptors are important for *Brucella* entry into HeLa cells. It would now be interesting to investigate the role of this signaling pathway in other cell lines and to pinpoint the exact mechanism that is regulated by this pathway during *Brucella* infection.

Materials & Methods

Materials

HeLa (human cervical carcinoma epithelial cell line, ATCC, CCL-2); Mv1Lu (mink lung epithelial cell line, ATCC, CCL-64); THP-1 (human monocytic leukemia cell line, ATCC, TIB-202); human embryonic kidney 293T (HEK-293T)(from Hwain Cornelis's lab); Dulbecco Modified Eagle Medium (DMEM) (Sigma, D5796); Dulbecco Modified Eagle Medium Glutamax (DMEM Glutamax)(Gibco, 61965-026); minimum essential medium (MEM)(Sigma, M5650); RPMI-1640 medium (Sigma, R0883); Fetal Calf Serum (FCS)(Gibco, 10270): heat inactivated at 56°C for 30 min before use; Fetal Calf Serum (FCS)(Bioconcept, 2-01F30-I); tryptic soy broth (TSB)(Fluka, 22092); kanamycin sulfate (Sigma-Aldrich, 60615); ampicillin sodium salt (Applichem, A.8039.0025); gentamicin (Sigma, G1397); Triton-x-100, sigma-ultra (Sigma-Aldrich, T9284); DAPI (Roche, 10236276001); phalloidin-547 (Dyomics, 547PI-33); albumin from bovine serum (BSA)(Sigma, A9647); paraformaldehyde (Sigma, P6148); phosphate buffered saline (PBS)(Gibco, 20012); L-glutamine (Sigma-Aldrich, G7513); phorbol myristate acetate (PMA)(Sigma, P8139); PAC1 (New England Biolabs, R0547); EcoRI (New England Biolabs, R3101); BamHI (New England Biolabs, R3130); T4 DNA ligase (New England Biolabs, M0202); TGF β -1 (R&D Systems, 240-13); Fugene HD (Promega, E2312); 0.45 μ m membrane filter (Sarstedt, 83.1826); polybrene (Sigma, H9268); Hepes (Sigma, H3375); NaCl (Merck, 1.06404.1000); EDTA(Fluka Chemika, 03685); pepstatin (Applichem, A2205.0010); leupeptin (Applichem, A2183.0010); Pierce BCA protein assay kit (Thermoscientific 23225); nitrocellulose membrane (GE healthcare, RPN203D); ECL system (KPL, 54-69-00 and 54-70-00); mouse monoclonal anti-Lamp1 [H4A3] antibody (Abcam, ab25630); rabbit monoclonal anti-Rab7 (D95F2) antibody (Cell Signaling, 9367); rabbit polyclonal anti-TGFBR1 antibody (Abcam, ab31013); rabbit polyclonal anti-TGFBR2 antibody (milipore, 06-227); monoclonal mouse anti-actin antibody, clone C4 (milipore, MAB1501); ECL mouse IgG, HRP-linked whole antibody (GE healthcare, NA931V); ECL rabbit IgG, HRP-linked whole antibody (GE healthcare, NA934V); Alexa Fluor 546 Goat Anti-mouse IgG (Molecular probes, A-11030); Alexa Fluor 546 Goat Anti-rabbit IgG (Molecular probes, A-11010); propidium iodide (kind gift from Ton Rolink's lab);

Alexa Fluor 633 Rabbit Anti-mouse IgG (Molecular probes, A-21063), Qiagen all stars negative control (SI03650318); TGFBR1 siRNA (Qiagen Hs_TGFBR1_7).

Bacterial strains and cell lines

The bacterial strains used in this study include GFP expressing *Brucella abortus* 2308 that contains pJC43 with *gfp-mut3* gene under a constitutively active kanamycin resistance gene *aphA3* promoter (50) and DsRed expressing *Brucella abortus* 2308 that contains pJC44 with *DsRed_m* gene from pDsRed_m (Clontech) under a constitutively active kanamycin resistance gene *aphA3* promoter (45). DH5 α used for cloning experiments contains genotype ϕ 80dlacZ Δ M15, *recA1*, *endA1*, *gyrAB*, *thi-1*, *hsdR17* (rK-, mK+), *supE44*, *relA1*, *deoR*, Δ (lacZYA-argF) U169, *phoA* (N. Mantis, Institut Pasteur).

HeLa cells were grown in DMEM (Sigma) supplemented with 10% FCS (Gibco), Mv1Lu cells with MEM supplemented with 10% FCS (Bioconcept), THP-1 cells with RPMI-1640 medium supplemented with 10% FCS (Gibco) and 10mM L-glutamine and HEK293T cells with DMEM Glutamax (Gibco) supplemented with 10% FCS (Bioconcept). Cells were incubated at 37°C with 5% CO₂. THP-1 monocytes can be differentiated into a macrophage-like cell line with PMA at a final concentration of 10⁻⁷ M and 48 h incubation at 37°C with 5% CO₂ (51).

Plasmids

Plasmids pCMV5-TGFBR1 (addgene ID: 19161), pCMV5B-TGFBR2 (addgene ID: 11766 (42)), pCMV5B-TGFBR1 K232R (addgene ID: 11763 (43)), pCMV5-TGFBR2 Δ cyt (addgene ID: 14051 (44)) and pCMV5B-TGFBR2 K227R (addgene ID: 11762 (42)), pLKO.3G (addgene ID: 14748), psPAX2 (addgene ID: 12260) and pMD2.G (addgene ID: 12259) were obtained from Addgene (www.addgene.org). pWay19 was a gift from the Molecular Motion laboratory, Montana State University, Bozeman, MT.

A summary of the primers used in this study is listed in Table 1 in this results section. Primers were designed for the shRNA assay to be cloned into lentiviral vector pLKO.3G. For Rab7A shRNA 1 the primer pair prSL098 and prSL099 was used to produce pSL064, for Rab7 shRNA 2 prSL100 and prSL101 produced pSL065, for Rab7 shRNA 3 prSL102 and prSL103 produced pSL066, for TGFBR1 shRNA 1

prSL104 and prSL105 produced pSL076, for TGFBR1 shRNA 1 Sigma prSL129 and prSL130 produced pSL078 or pSL080, for TGFBR2 shRNA 1 prSL106 and prSL107 produced pSL067, for TGFBR2 shRNA 2 prSL108 and prSL109 produced pSL069, for TGFBR2 shRNA 3 prSL110 and prSL111 produced pSL068 and for TGFBR2 shRNA 1 sigma prSL133 and prSL134 produced pSL079. shRNA primers were annealed at 95⁰C for 4 min, incubated at 70⁰C for 10 min, after which the now double stranded fragments were slowly cooled to room temperature (RT) over several hours. pLKO.3G was digested with Pac1 for 3 h at 37⁰C, gel purified and digested again with EcoRI for 3 h at 37⁰C. 7 kb fragment from the digestion was gel purified and used for ligation with annealed oligos, for 13 h at 16⁰C using T4 DNA ligase. Ligated plasmids were transformed into DH5 α and plated on ampicillin agar plates. Clones were sequenced with primer prSL112.

Cell culture and Infection

Brucella abortus strains were grown in TSB medium containing 50 μ g/ml kanamycin for 24 h at 37⁰C and shaking (100 rpm) to an OD of 0.8- 1.1. Bacteria were added to cells with a final MOI of 10000 for HeLa cells or Mv1Lu and MOI 500 for THP-1 macrophage like cell line. Plates were then centrifuged at 400 g for 20 min at 4⁰C to synchronize bacterial entry. After 4 h incubation at 37⁰C and 5% CO₂, extracellular bacteria were killed by exchanging the infection medium with DMEM (Sigma)/10% FCS (Gibco) supplemented with 100 μ g/ml gentamicin. After a total infection time of 44 h, cells were fixed with 3.7% PFA for 20 min at RT.

Automated image analysis and infection scoring

Images were taken with Molecular Devices ImageXpress microscopes using the 10X S Fluor objective, after which automated image analysis and decision tree infection scoring was performed as described in Research Article III. Binary level infection detection (infected vs. non-infected) allows infection index (Infected cell / total cell number) to be defined.

TGF- β assay

HeLa cells (2800 cells / well) were seeded a day before infection in a 96 well format with DMEM (Sigma) 10% FCS (Gibco) containing TGF β -1. Cells were then incubated in 37°C, 5% CO₂ for another 24 h, after which infection was performed in absence of TGF β -1 in the culture medium.

Plasmid transfection

HeLa cells were seeded a day before transfection in a 6 well plate with 250,000 cells/well. Next morning, 0.9 μ g of plasmid or 0.45 μ g of each plasmid for double transfection (in 200 μ l of DMEM without FCS) were mixed with 8 μ l of Fugene HD (in 200 μ l of DMEM without FCS) and incubated for 15 min at RT. HeLa cells were exchanged with 1.5 ml of fresh medium and DNA-Fugene complex was added to the cells. The next morning, cells were exchanged with fresh medium and in the evening splitted into a 96 well format (2800 cells/well) for infections on the following day.

Lentiviral transduction

3x10⁶ HEK293T cells were grown in a 10 cm² dish and incubated at 37°C, 5% CO₂ for at least 6-8 hours. 1.7 μ g pLKO.3G, 1.25 μ g psPAX2 packing plasmid and 420 ng pMD2.G envelope plasmid in 600 μ l DMEM without FCS were then mixed with 25 μ l of Fugene HD in 600 μ l DMEM without FCS and incubated for 15 min at room temperature. The DNA-Fugene complex was then added to the cells that were replaced with 5 ml of fresh medium. Cells were exchanged with fresh medium the following day. 2 days later, the supernatant of HEK293T that now contain viruses was collected and filtered through a 0.45 μ m membrane filter. Viruses were then used directly for transduction of cells or stocked in tubes at -80°C. Aliquots of virus containing supernatant (1 ml or 2 ml volumes) were added to cells in presence of fresh medium to a total volume of 3 ml, and polybrene having a final concentration of 5 μ g/ml. Cells were exchanged with fresh medium the next day and used for further experiments.

shRNA assay

For HeLa cells, 70,000 cells were seeded in a 6 well plate one day before viral transduction. Around 100,000 cells of THP-1 monocytes were first transduced with viruses containing shRNA in a 6 well plate and exchanged with fresh medium the next day. In a 96 well format, 15,000-20,000 cells were differentiated 2 days before infection with PMA (final concentration of 10^{-7} M) in RPMI medium supplemented with 10% FCS (Gibco) and 10mM L-glutamine. Experiments were performed earliest after 3 days of incubation to allow sufficient time for shRNA expression. Cells were then infected with DsRed expressing *Brucella abortus* 2308 for around 44 h, fixed and stained with DAPI (final concentration 1 μ g/ml) for the nucleus. Images were acquired with Molecular Devices ImageXpress microscopes with the 10X S Fluor objective, after which automated image analysis and infection scoring was performed as described above. An additional feature during image analysis that was obtained in this assay is the GFP intensity of the cells that indicates shRNA expression. A threshold is set in Spotfire that allows separation of shRNA containing GFP-positive cells vs. GFP-negative cells. Infection indices were then compared only within the GFP-positive cell population, between the empty plasmid control that contains only GFP vs. shRNA expressing GFP-positive cells.

siRNA transfection

HeLa cells were transfected with siRNAs at a final concentration of 20 nM. 2 μ l of 20 μ M siRNA stock and 2 μ l of RNAimax were added to 200 μ l of DMEM (Sigma) without FCS, mixed and incubated for 30 min at RT. Transfection mixture was then added to 1.8 ml of cells (6×10^4 cells/ml) in a well of a 6 well plate and incubated for around 48 h. Next, cells were trypsinized and 50,000 cells/well were seeded in a 24 well plate with coverslip one day before infection.

Immunofluorescent labeling

HeLa cells on coverslips were permeabilized with 0.1% Triton-x-100 for 10 min at RT, washed with PBS before incubated with 0.5% BSA/PBS for 30 min at RT. Afterwards, cells were labeled for Lamp1 or Rab7 using mouse monoclonal anti-Lamp1 [H4A3] antibody (1:100) or rabbit monoclonal anti-Rab7 (D95F2) antibody

(1:50) respectively and secondary antibody Alexa Fluor 546 Goat Anti-mouse IgG (1:100) or Alexa Fluor 546 Goat Anti-rabbit IgG (1:300). For the TGF β -1 assay and overexpression studies, cells were permeabilized for 10 min with 0.1% Triton-x-100 and stained with DAPI (final concentration 1 μ g/ml) for nuclei and phalloidin-547 (1:250) for F-actin.

Western blotting

Cells were trypsinized from a 25cm² culture flask, washed once with cold PBS and cell pellet used directly or kept at -80°C. 50 μ l of ice cold lysis buffer (50 mM HEPES pH 7.5, 250 mM NaCl, 1 mM EDTA, 1 μ M pepstatin, 1 μ M leupeptin and 1% Triton-x-100) was used to resuspend the cell pellet and suspension was incubated 30 min on ice. Cell lysate was pelleted at 13,000 rpm for 10 min at 4°C and supernatant was collected in a new eppendorf tube. Protein concentration was measured with the BCA kit and around 50 μ g – 80 μ g of protein was loaded in each well, separated by 10% SDS-PAGE and transferred onto a Hybond-C Extra nitrocellulose membrane. The membranes were blotted for Rab7 using rabbit monoclonal anti-Rab7 (D95F2) antibody (1:1000), TGFBR1 using rabbit polyclonal anti-TGFBR1 antibody (1:500), TGFBR2 using rabbit polyclonal anti-TGFBR2 antibody (1:500) and actin as loading control using monoclonal mouse anti-actin antibody, clone C4 (1:10000). Proteins were visualized using the ECL system (GE Healthcare) with ECL mouse IgG, HRP-linked whole antibody (1:10000) or ECL rabbit IgG, HRP-linked whole antibody (1:2000).

Cell surface staining and flow cytometry

For cell surface staining of TGFBR2, cells were trypsinized and 1-2x10⁶ cells were collected for each sample. Cells were washed once with FACS buffer (1% FCS (Bioconcept) in PBS) and incubated with rabbit polyclonal anti-TGFBR2 antibody (final concentration 20 μ g / ml) for 1 h at 4°C, washed with FACS buffer and further incubated with Alexa Fluor 633 Rabbit Anti-mouse IgG (final concentration 20 μ g / ml) for 1 h at 4°C. After washing cells with FACS buffer, samples were incubated with propidium iodide (PI) for few minutes to label dead cells before flow cytometry was performed. Flow cytometry was performed using the FACS Calibur (BD

Results: TGF- β signaling and *Brucella* infection

Biosciences) and data were analyzed using the FlowJo software. Only cells that were negative for PI staining were used for analysis.

Table 1: Primers used in this study

Name	Sequence
prSL098	5'-AATTCTGCTGCGTTCTGGTATTTGACTCGAGTCAAATACC AGAACGCAGCAGTTTTTTTAT-3'
prSL099	5'-AAAAAACTGCTGCGTTCTGGTATTTGACTCGAGTCAAAT ACCAGAACGCAGCAG-3'
prSL100	5'-AATTGAGGTGGAGCTGTACAACGAACTCGAGTTCGTTGTA CAGCTCCACCTCTTTTTTTAT-3'
prSL101	5'-AAAAAAGAGGTGGAGCTGTACAACGAACTCGAGTTCGT TGTACAGCTCCACCTC-3'
prSL102	5'-AATTCACGTAGGCCTTCAACACAATCTCGAGATTGTGTTG AAGGCCTACGTGTTTTTTTAT-3'
prSL103	5'-AAAAAACACGTAGGCCTTCAACACAATCTCGAGATTGT GTTGAAGGCCTACGTG-3'
prSL104	5'- AATTTGGGATTGTACTATAACCAGTACTCGAGTACTGGTAT AGTACAATCCCATTTTTTTTAT-3'
prSL105	5'- AAAAAAATGGGATTGTACTATAACCAGTACTCGAGTACTG GTATAGTACAATCCCA-3'
prSL129	5'- AATTGATCATGATTACTGTGATAACTCGAGTTATCGACA GTAATCATGATCTTTTTTTTAT-3'
prSL130	5'- AAAAAAAGATCATGATTACTGTGATAACTCGAGTTATC GACAGTAATCATGATC-3'
prSL106	5'- AATTTACCATGACTTTATTCTGGA ACTCGAGTTCAGAAT AAAGTCATGGTATTTTTTTTAT-3'
prSL107	5'- AAAAAAATACCATGACTTTATTCTGGA ACTCGAGTTCGA GAATAAAGTCATGGTA-3'
prSL108	5'- AATTATGGAGAAAGAATGACGAGAACTCGAGTTCCTCGTC ATTCTTTCTCCATTTTTTTTAT-3'
prSL109	5'- AAAAAAATGGAGAAAGAATGACGAGAACTCGAGTTCCT CGTCATTCTTTCTCCAT-3'
prSL110	5'- AATTCTCCAATATCCTCGTGAAGAACTCGAGTTCCTCACG AGGATATTGGAGTTTTTTTAT-3'

Results: TGF- β signaling and *Brucella* infection

prSL111 5'-AAAAAACTCCAATATCCTCGTGAAGAACTCGAGTTCTT
CACGAGGATATTGGAG-3'

prSL112 5'- GACTATCATATGCTTACCGT-3'

prSL133 5'- AATTCTCTAGGCTTTATCGTGTTTACTCGAGTAAACACGA
TAAAGCCTAGAGTTTTTTTAT-3'

prSL134 5'- AAAAAAACTCTAGGCTTTATCGTGTTTACTCGAGTAAACA
CGATAAAGCCTAGAG-3'

Figure legends

Figure 1. Knockdown of TGF- β signaling components reduces *Brucella* infection in HeLa cells. Box plot represents Z-score normalized infection index (described in material and methods of Research Article III) of data points from kinome screen (Ambion library), primary genome-wide screens (Dharmacon and Qiagen library) and secondary screens (Ambion and Sigma library). Details of these libraries can be found in Research Article III. The median is represented at the middle of the box. Positive and negative controls Rac1 or scrambled siRNA are included for reference. Whiskers and outliers of boxplot are calculated with the Tukey method.

Figure 2. Activation of TGF- β signaling by TGF- β 1 increases *Brucella* infection. HeLa cells were pre-incubated for 24 h with DMEM containing TGF- β 1 before infected with GFP-expressing *Brucella abortus* for 44 h. Plates were imaged with DAPI channel for nuclei, GFP channel for bacteria and RFP channel for cell body with F-actin. Automated image analysis with CellProfiler and decision tree infection scoring was performed with nucleus, perinucleus and cell body as objects. Data represents normalized infection in reference to cells that were not pre-incubated with TGF- β 1, the mean \pm STDEV of at least three independent experiments.

Figure 3. Overexpression of wild type or dominant negative TGFBR increases and reduces *Brucella abortus* infection respectively. HeLa cells were transfected with A) wild type TGFBR (pCMV5-TGFBR1 or pCMV5B-TGFBR2) or B) dominant-negative TGFBR (pCMV5B-TGFBR1 K232R, pCMV5 TGFBR2 Δ cyt, pCMV5B-TGFBR2 K227R) either individually or in combination of both TGFBR1 and TGFBR2. After 1.5 days of overexpression, cells were infected with GFP-expressing *Brucella abortus* for around 44 h. Plates were imaged with DAPI channel for nuclei, GFP channel for bacteria and RFP channel for cell body with F-actin. Automated image analysis with CellProfiler and decision tree infection scoring was performed with nucleus, perinucleus and cell body as objects. Data represents normalized infection in reference to non- transfected cells, the mean \pm STDEV at least four independent experiments. For B), dataset with overexpression of both receptors shows the mean of only two independent experiments.

Figure 4. shRNA knockdown in HeLa and THP-1 cells. HeLa cells were lentivirally transduced with pLKO.3G empty vector or pLKO.3G containing indicated shRNAs respectively. After 6 – 12 days, cells were infected with DsRed-expressing *Brucella abortus* and fixed after around 44 hours post infection (hpi). THP-1 monocytes were lentivirally transduced with pLKO.3G empty vector or pLKO.3G containing shRNAs respectively. On day two, monocytes were differentiated by addition of PMA to the medium. On day four, cells were infected with DsRed-expressing *Brucella abortus* and were fixed after around 44 hpi. Plates were imaged with DAPI channel for nuclei, GFP channel for cells expressing shRNA, and RFP channel for bacteria. Automated image analysis with CellProfiler and decision tree infection scoring was performed with nucleus, perinucleus and voronoi cell as objects. GFP-positive cells were selected in SpotFire software for infection analysis. Data represents normalized infection in reference to the empty vector, the mean \pm STDEV of at least three independent experiments. TGFBR2 shRNA 2 dataset has only two independent experiments.

Figure 5. *Brucella abortus* are not contained within Lamp1 compartments upon knockdown of TGFBR1 in HeLa cells. Representative images of HeLa cells infected with GFP-expressing *Brucella abortus* for 24 h after 72 h of transfection with scrambled or TGFBR1 siRNA. Cells were immunostained for Lamp1 (red) and images were taken with the 60x objective and FEI MORE with TIRF microscope. Images in stacks were deconvolved with HUVGENs remote manager and one represented slice around the middle of a stack is shown. Scale bars represent 10 μ m.

Figure 6. *Brucella abortus* are not contained within Rab7 compartments upon knockdown of TGFBR1 in HeLa cells. Representative images of HeLa cells infected with GFP-expressing *Brucella abortus* for 24 h after 72 h of transfection with scrambled or TGFBR1 siRNA. Cells were immunostained for Rab7 (red) and images were taken with the 60x objective and FEI MORE with TIRF microscope. Images in stacks were deconvolved with HUVGENs remote manager and one represented slice around the middle of a stack is shown. Scale bars represent 10 μ m.

Supplementary Figure 1. Validation of shRNA knockdown by western blot or immunofluorescence studies. A) Comparison of endogenous Rab7 staining of Rab7 shRNA expressing cells (GFP positive) versus neighbouring non transduced cells. Rab7 shRNA or pLKO.3G empty vector transduced HeLa cells were stained with Rab7 rabbit antibody (1:50, Cell signaling) and imaged with the Andor confocal microscope. A representative slice around the middle of a stack is presented. At least three independent experiments were performed. Scale bar represents 10 μ m. B) Validation of shRNA knockdown by western blot with actin as loading control.

Supplementary Figure 2. *Brucella* infection of THP-1 cell macrophage-like cell line. A) Image represents THP-1 macrophage infected by GFP-expressing *Brucella abortus* at around 44 hpi. Scale bar represents 10 μ m. B) Bar graph represents MOI titration curve of GFP-expressing *Brucella abortus* infection of THP-1 macrophages at around 44 hpi. THP-1 monocytes were differentiated two days before infection with PMA.

Supplementary Figure 3. HeLa cells expresses low levels of surface TGFBR2 compared to Mv1Lu. A) HeLa cells and Mv1Lu were stained for surface TGFBR2, without cell permeabilization using 1^o TGFBR2 antibody and 2^o antibody Alexa 633. Cells were then subjected to flow cytometry analysis. B) Representative image of HeLa cells or Mv1Lu cells infected with GFP-expressing *Brucella abortus*. Image was taken with 10x objective using Molecular Devices ImageXpress microscope. Scale bar represents 50 μ m.

Supplementary Figure 4. TGFBR2 overexpression increases the levels of surface TGFBR2 in HeLa cells. HeLa cells were transfected with cDNA encoding TGFBR2 (pCMV5B-TGFBR2) and stained for surface TGFBR2, without cell permeabilization using 1^o TGFBR2 antibody and 2^o antibody Alexa 633. Cells were then subjected to flow cytometry analysis. pWay19 is a GFP alone-expressing plasmid that was co-transfected with pCMV5B-TGFBR2 to monitor transfection efficiency. There was generally an increase in TGFBR surface expression with both 1 day and 2 days of cDNA overexpression, seen in GFP expressing and non- expressing cells.

Supplementary Figure 5. Efficiency of shRNA transduction in HeLa cells and THP-1 monocytes. HeLa cells and THP-1 monocytes transduced with shRNA were analyzed with flow cytometry to estimate the efficiency of transduction in these cells. shRNA expressing cells also express GFP, since GFP is encoded in the empty vector.

Supplementary Table 1. Sequences that were used for shRNA design. Table shows sequences that were used for shRNA design, selected from different sources.

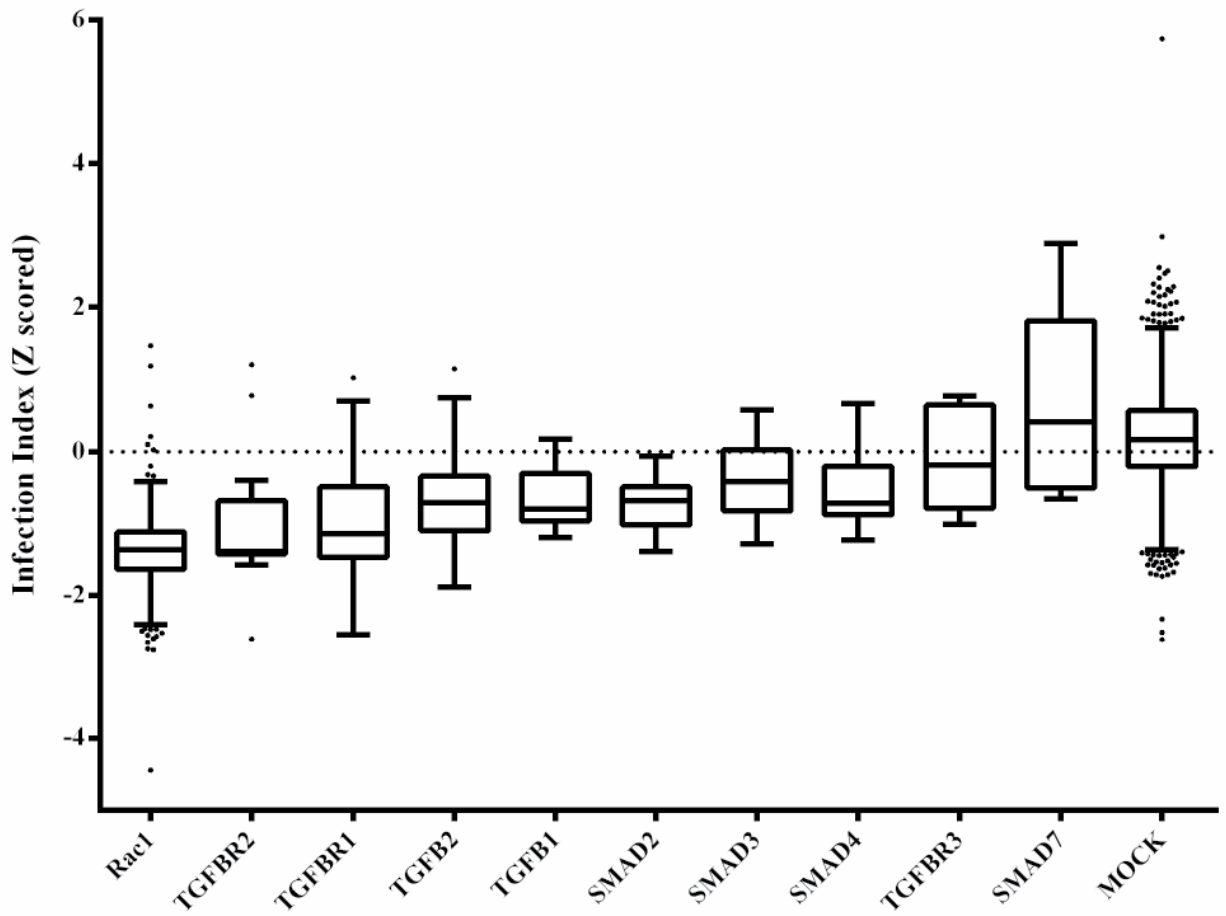


Figure 1

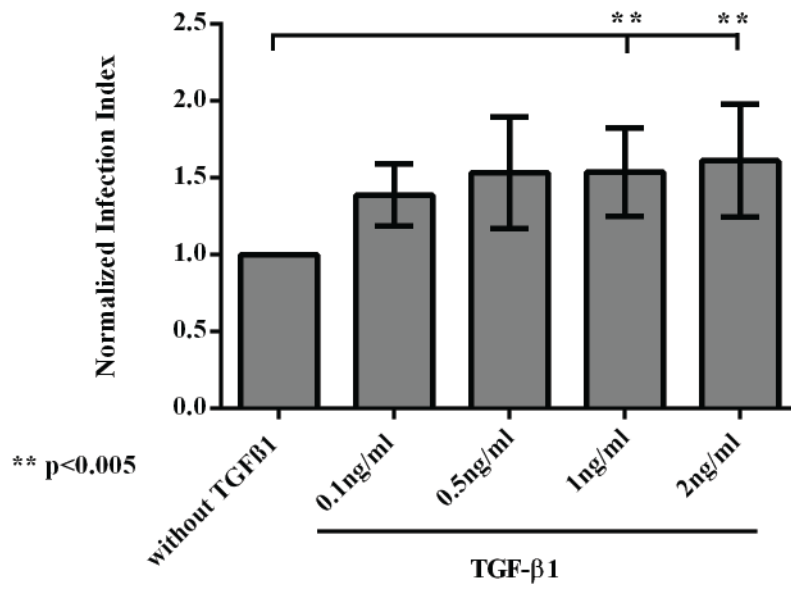


Figure 2

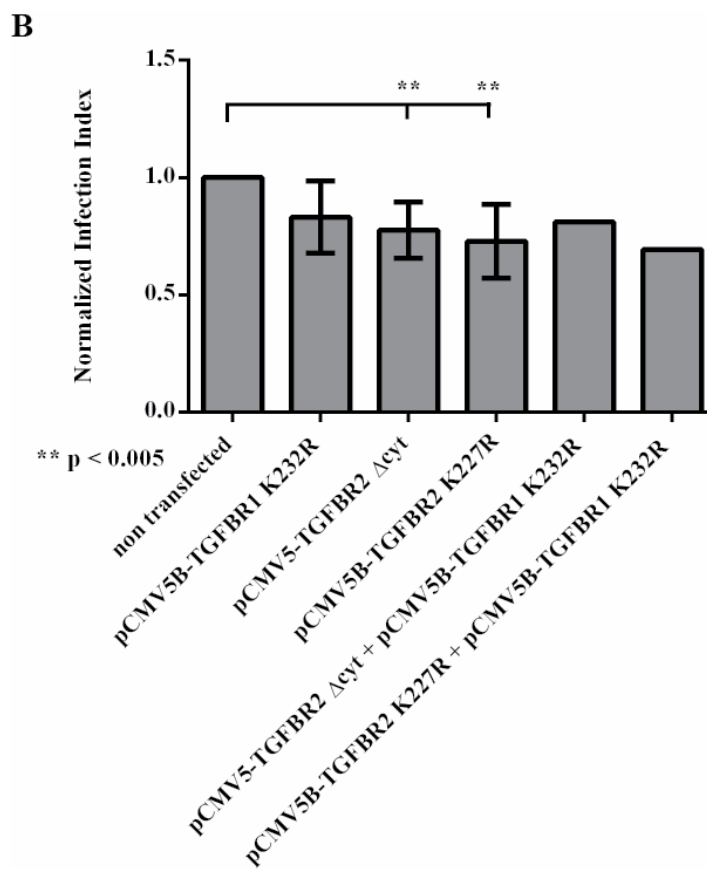
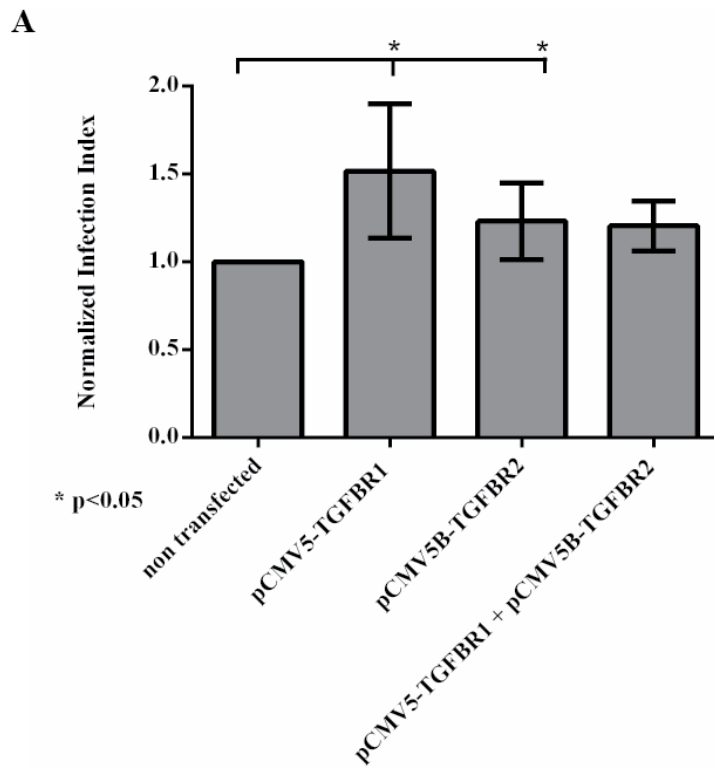


Figure 3

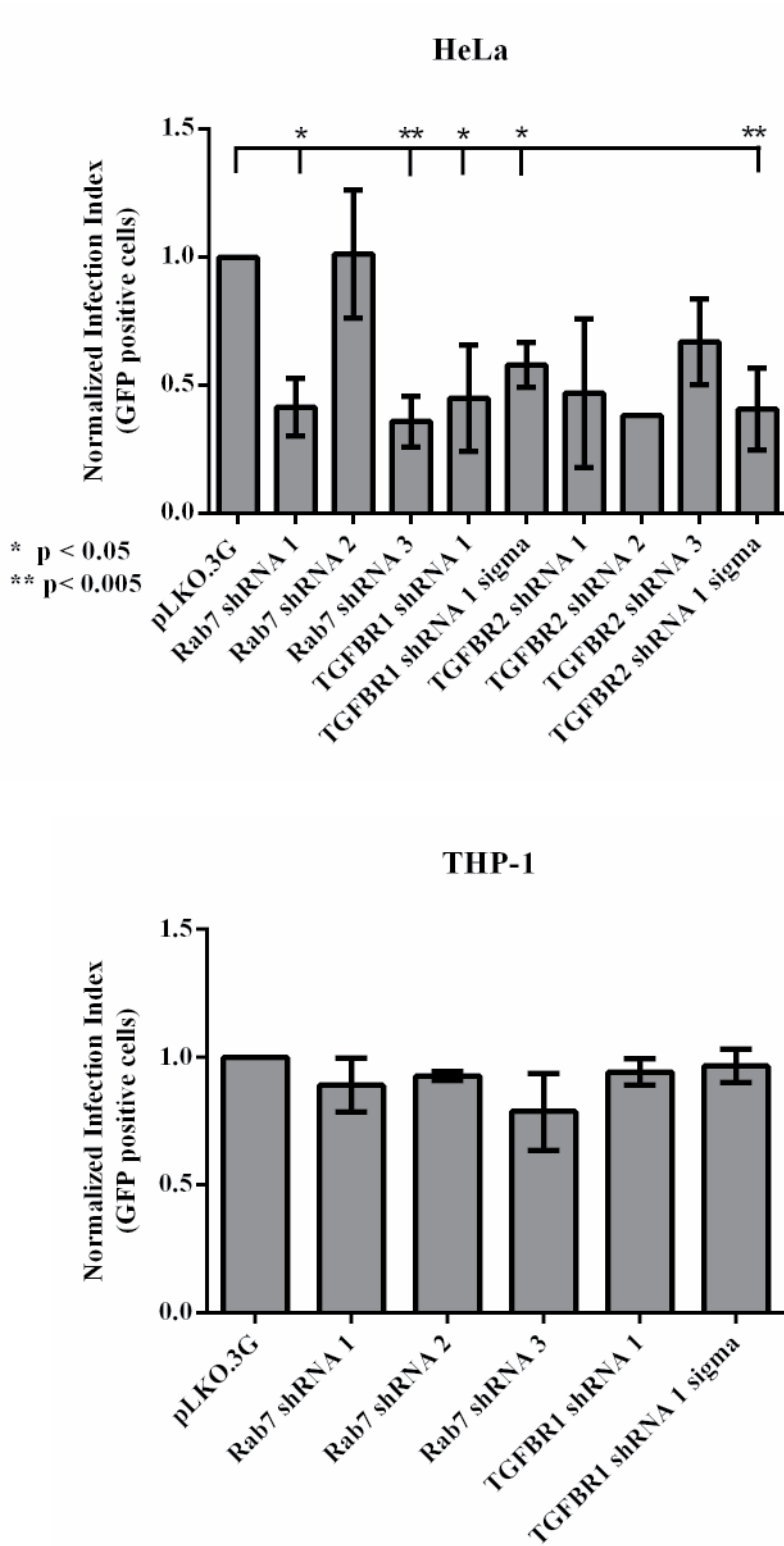


Figure 4

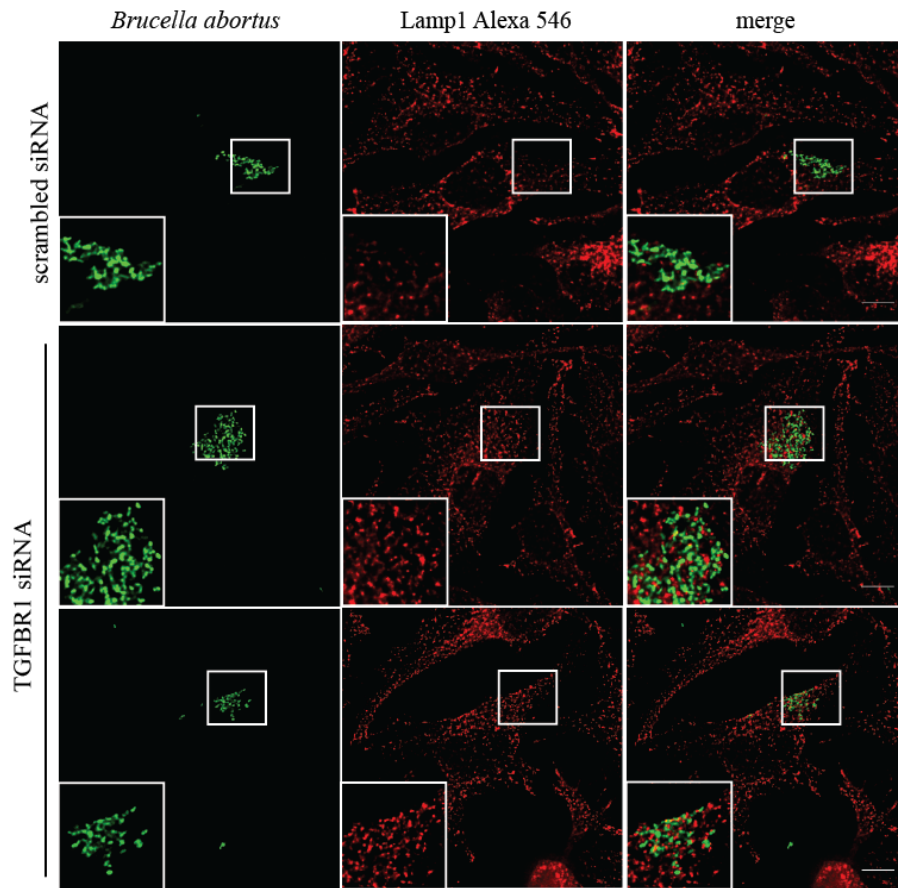


Figure 5

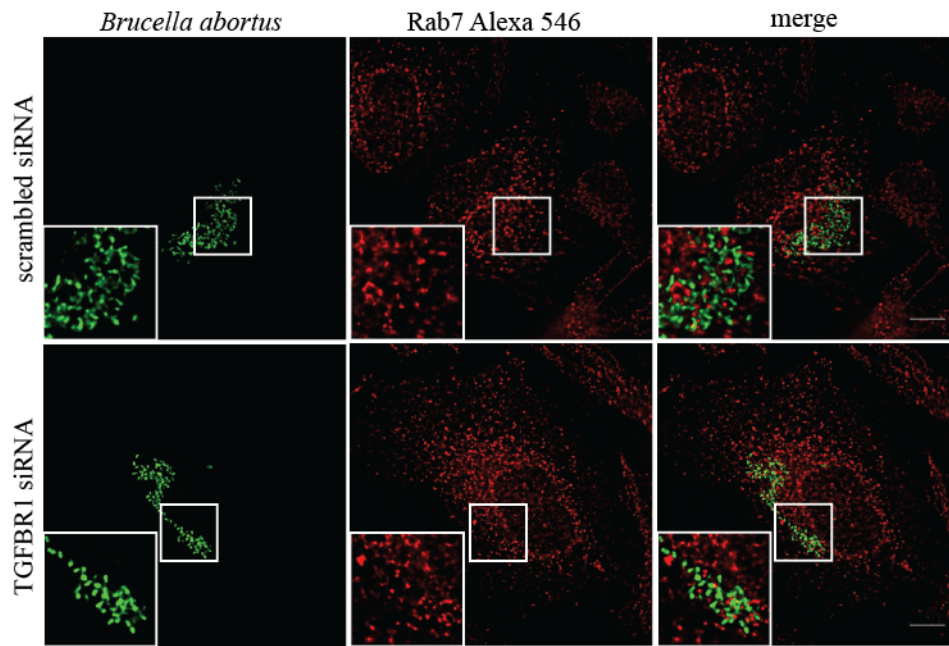
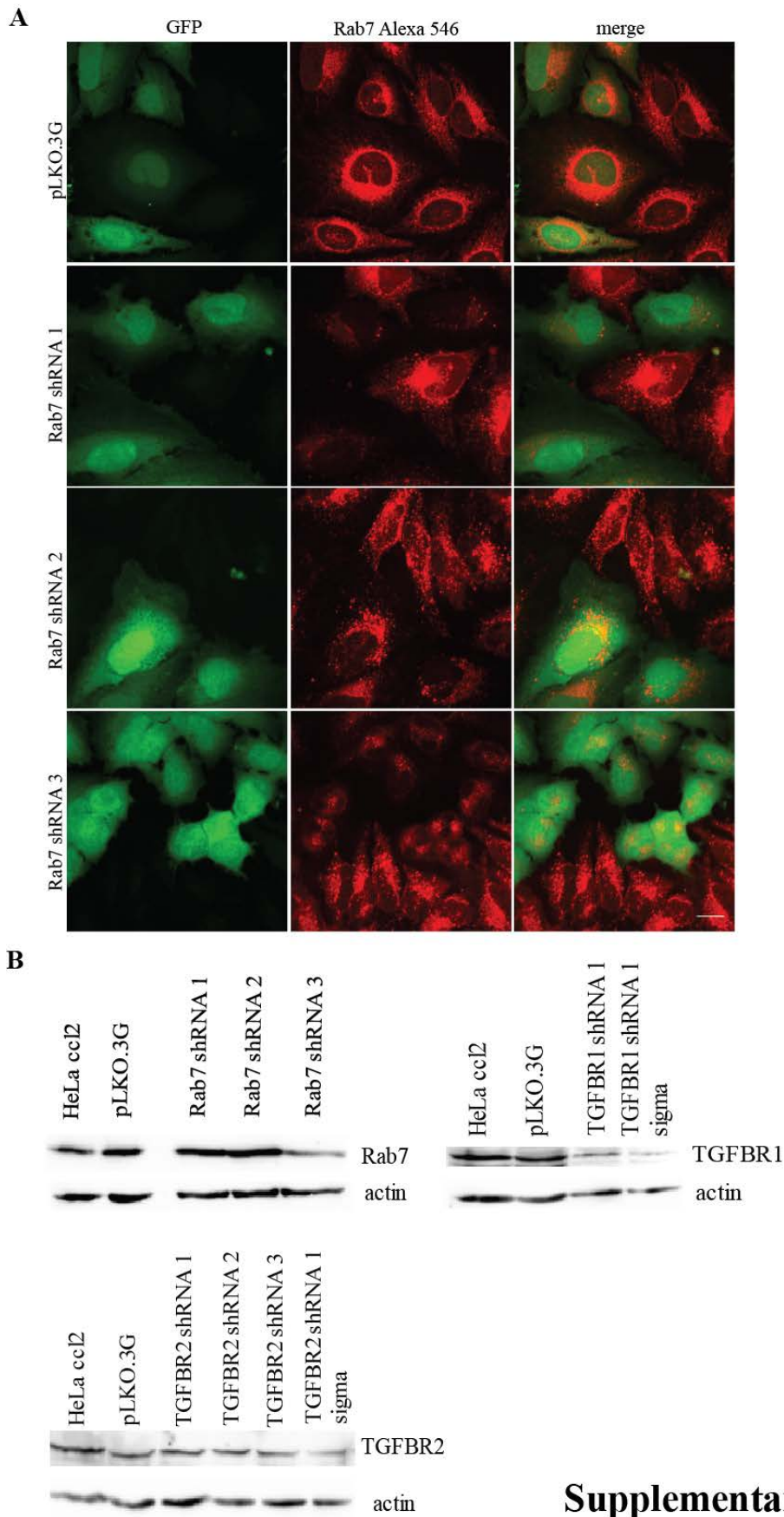
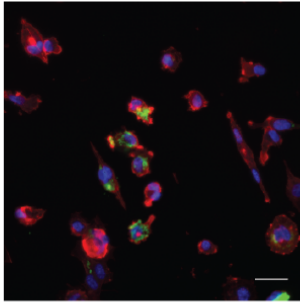


Figure 6



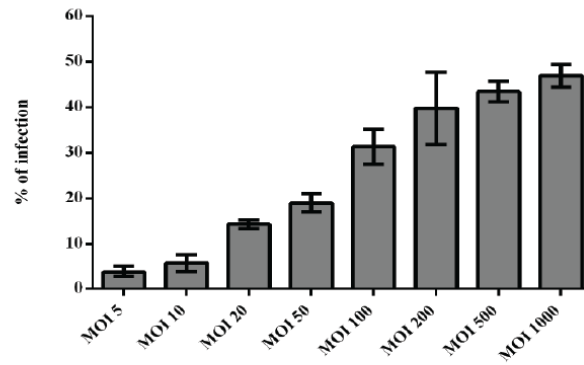
Supplementary Figure 1

A



Nuclei (DAPI)
Brucella abortus 2308 (GFP)
F-actin (phalloidin-547)

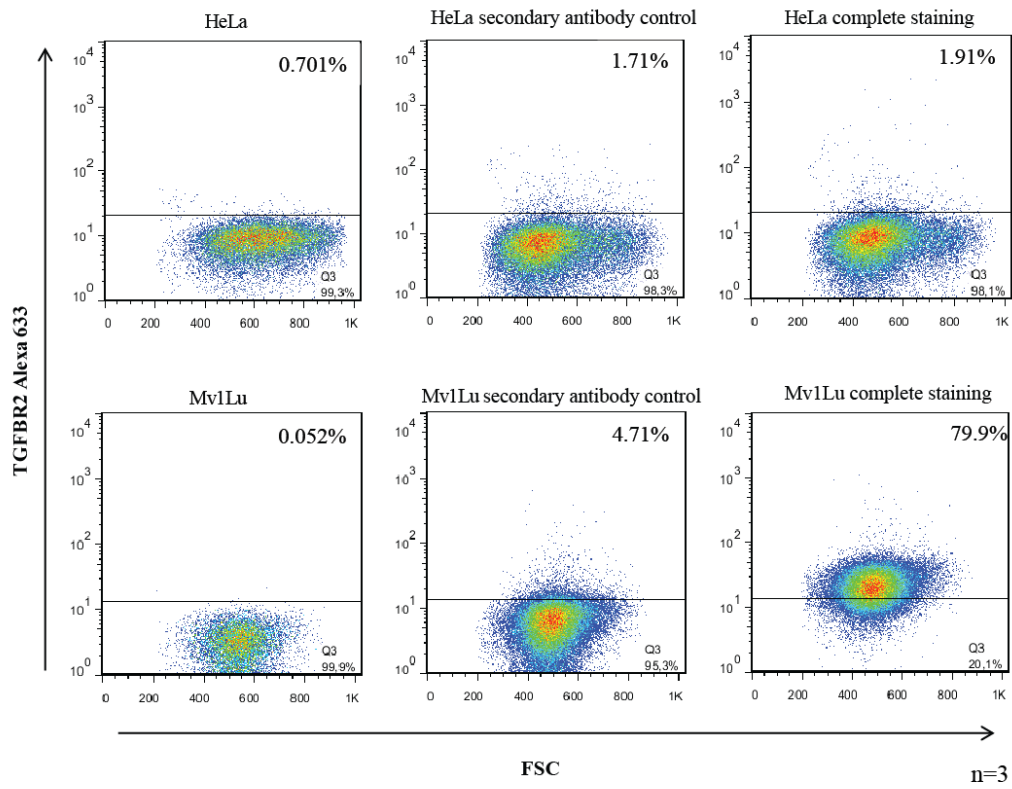
B



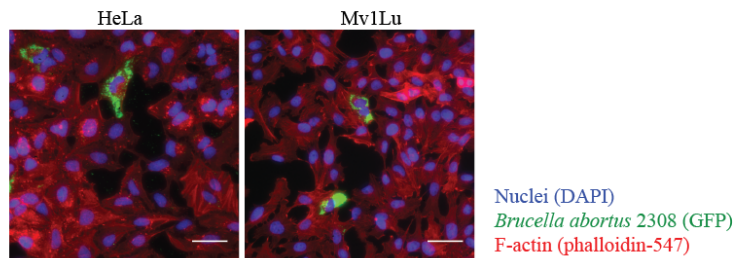
Supplementary Figure 2

Results: TGF- β signaling and *Brucella* infection

A

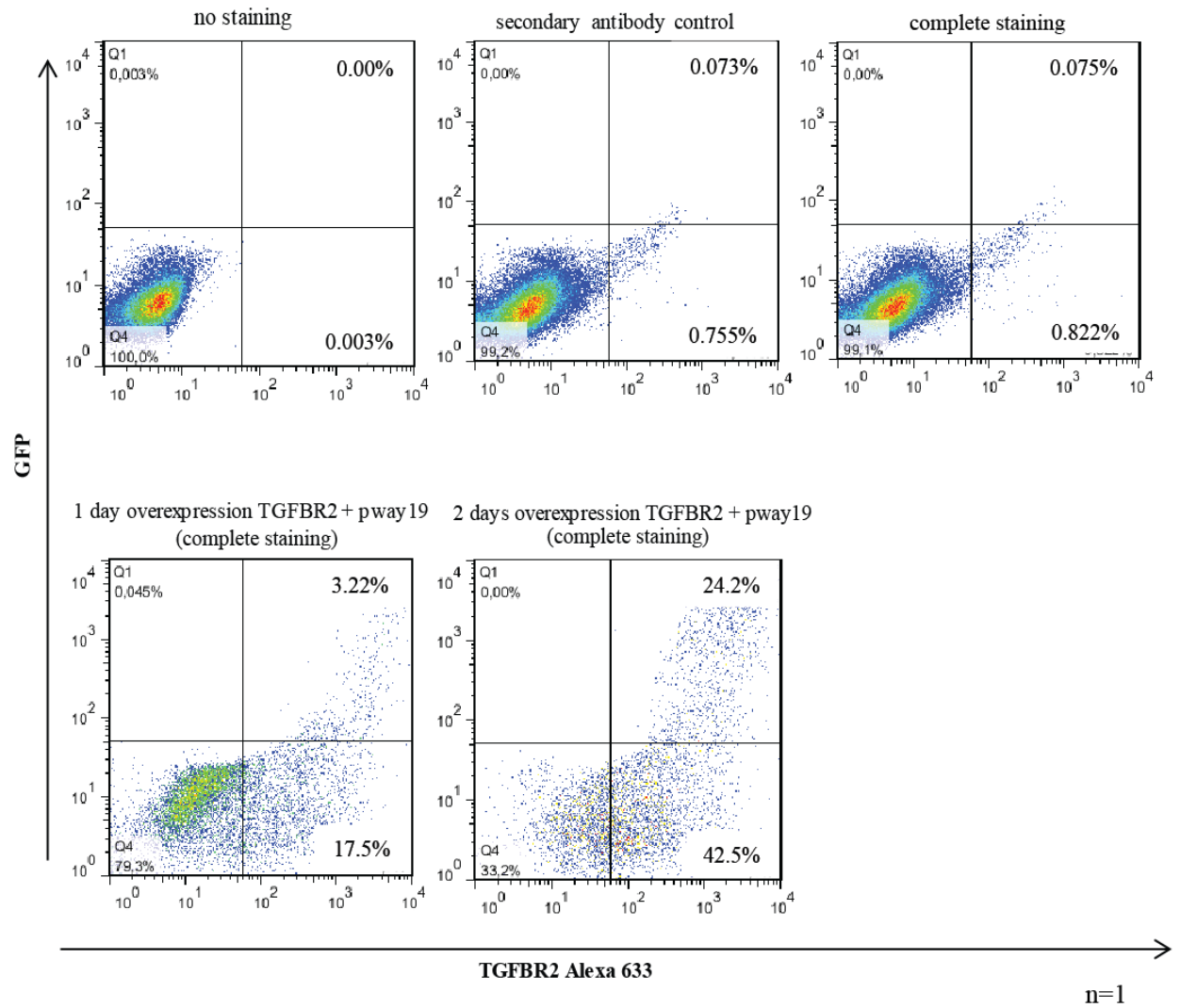


B



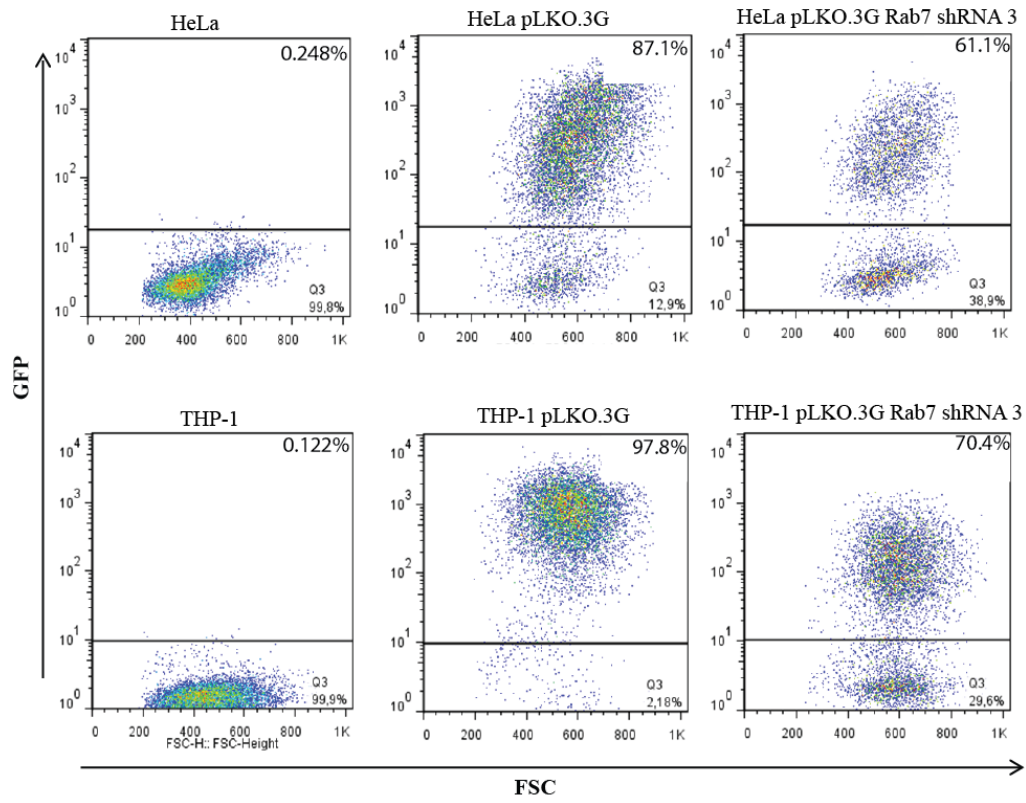
Supplementary Figure 3

Results: TGF- β signaling and *Brucella* infection



Supplementary Figure 4

Results: TGF- β signaling and *Brucella* infection



Supplementary Figure 5

Results: TGF- β signaling and *Brucella* infection

	shRNA sequence	Target region	Infection Index (Z scored)	Sequence source
Rab7 shRNA 1	CTGCTGCGTTCTGGTATTGA	ORF	-1.33	Qiagen unpooled library
Rab7 shRNA 2	GAGGTGGA GCTGTA CAACGAA	ORF	-1.32	Qiagen unpooled library
Rab7 shRNA 3	CACGTAGGCCTTCAACACAAT	3' UTR	-1.14	Qiagen unpooled library
TGFBR1 shRNA 1	TGGGATTGTA CTATACCA GTA	3' UTR	-1.11	Qiagen unpooled library
TGFBR1 shRNA 1 sigma	GATCATGATTACTGTCGATAA	3' UTR	NA	Sigma website*
TGFBR2 shRNA 1	TACCATGACTTTATTCTG GAA	ORF	-1.39	Qiagen unpooled library
TGFBR2 shRNA 2	ATGGGA GAAA GAAUGA CGA GAA	ORF	-1.42	Ambion unpooled library
TGFBR2 shRNA 3	CTCCAATA TCCTCGTGAAGAA	ORF	-1.42	Qiagen unpooled library
TGFBR2 shRNA 1 sigma	CTCTAGGCTTTATCGTGTTA	3' UTR	NA	Sigma website*

NA : not available, ORF : open reading frame, 3' UTR : 3' untranslated region
 * sequence information no longer available on website

Supplementary Table 1

3.5 Unpublished results: The role of retromer complex in *Brucella* infection

Introduction

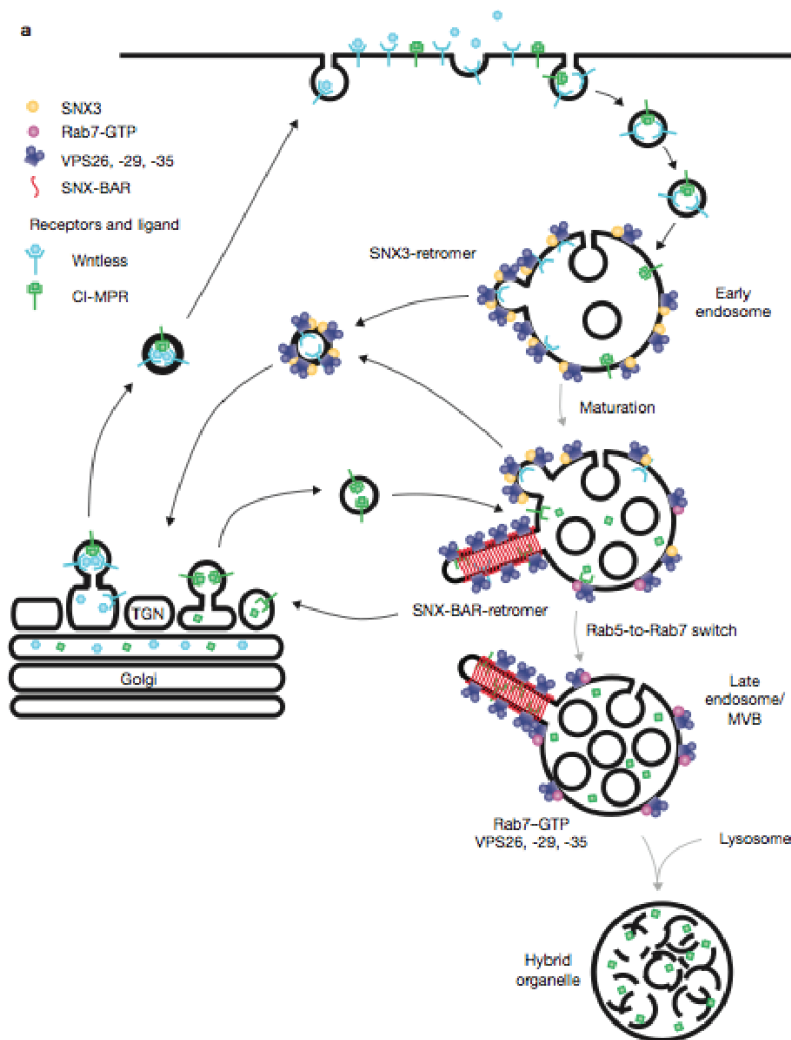
The retromer complex was first identified in yeast *Saccharomyces cerevisiae* (52, 53) and later in higher eukaryotes e.g. mammalian cells, being highly conserved from yeast to mammals. It is an essential component of the endosomal protein sorting machinery. Besides having a major role in transporting cargoes via the endosome-Golgi route, it has also been reported to play a role in recycling of cargoes through Rab4-dependent endosome-plasma membrane pathway (52-54). Well-studied cargoes of the retromer complex include cation independent mannose 6-phosphate receptor (CIMPR) (55), iron transporter (DMT1) (56), sortilin-related receptor (SorL1/ SorLA) that binds to amyloid precursor protein (57), Shiga toxin (58), etc. Retromer cargoes have been shown to traffic to the Golgi via different routes. Recycling endosomes are important intermediates of Shiga toxin retrograde transport while CIMPR travels from late endosomes to Golgi since late endosomal Rab9 GTPase is required for its recycling (59-61). Syntaxin 10 is required for CIMPR recycling while STX16 and Rab6 GTPase are required for toxins transport to the Golgi (62, 63).

In mammalian cells, the retromer complex is composed of two functional subcomplexes: a trimer Vps26-Vps29-Vps35 that is involved in cargo selection and is generally known as the cargo selective trimer (CST), and proteins from the sorting nexin (Snx) family (64). Vps35 is the core component of the trimer that has a direct role in cargo binding (65, 66), with Vps26 and Vps29 independently associating at either end. Vps26 has also been shown to directly bind retromer cargo SorL1 / SorLA (67). There are two paralogues of Vps26 in mammals: Vps26a and Vps26B (68). Recent evidence showed that endosome to Golgi retrieval of (CIMPR) requires Vps26a and not Vps26b, suggesting preferences of the paralogues for different cargo (69).

There are two different types of SNX that associate with CST (Figure 1). SNX1 or SNX2 that dimerizes with SNX5 or SNX6 belong to the SNX-BAR subfamily due to their membrane curvature sensing BAR (Bin/amphiphysin/Rv) domain and

Results: Retromer complex in *Brucella* infection

phosphatidylinositol 3-phosphate [PtdIns(3)P] or phosphatidylinositol 3,5-bisphosphate [PtdIns(3,5)P₂] binding PX (phox homology) domain (70, 71). The SNX heterodimer is therefore able to induce and stabilize the formation of membrane tubules (72, 73). This complex of SNX-BAR and the CST is termed the SNX-BAR-retromer (74). SNX3 has also been reported to associate with CST (75). However, SNX3 does not contain a BAR domain but has a PX domain that binds with high affinity to PtdIns(3)P (76). The complex of SNX3 and CST is termed the SNX3-retromer (74) (Figure 1). Not all cargoes have been assigned to specific retromer complex and it could be that only one or both of these retromer complexes are needed for the trafficking of each cargo (75).



(64)(64)⁶⁴(64)(64)

Figure 1. A model that shows differential cargo sorting between SNX3-retromer and SNX-BAR retromer pathways. Wntless and CIMPR are sorted by SNX3-retromer and SNX-BAR-retromer complexes respectively. Authors from this review (74) suggest that transport of these cargoes are not

Results: Retromer complex in *Brucella* infection

exclusively only through these respective pathways but reflects the steady state that seems to be dependent on specific retromer complexes. Picture taken from (74) and adapted.

Retromer complex can be recruited to the endosomal membrane via different ways. Rab7-GTP has been shown to be required for the recruitment of the CST (77, 78), with interference of Rab7 (late endosomal marker) function causing dissociation of CST but not SNX from the membranes. Perturbation of Rab5 (early endosomal marker) also causes dissociation of both SNX dimer and CST from membranes through inhibition of phosphoinositide 3- kinase (PI3K) pathway. Therefore, Rab5 and Rab7 act together to regulate retromer recruitment to the endosome (78). SNX3 also recruits CST to the endosome as silencing of SNX3 displaces CST from the membrane (75). In the SNX3-retromer complex, SNX3 associates with Vps35 and both SNX3 and Rab7 are required for recruitment of the CST(79). Therefore, this places the action of the retromer at a region where both SNX3 and Rab7 are present (at early to late endosome transition)(80). For the SNX-BAR retromer complex, SNX1 and SNX2 are important to recruit CST and knockdown of both displaces CST from the membrane(81). SNX-BAR-retromer also functions at endosomes undergoing early to late endosome transition by binding to Rab7-GTP (74). Rab7 is a GTPase that cycles between the GTP active form and GDP inactive form, with Rab GTPase activating proteins (RabGAPs) proteins that stimulate GTP hydrolysis inhibiting its activity. TBC1D5, a member of the Tre2-Bub2-Cdc16 (TBC) family of Rab GTPase activating proteins (GAPs) (77) has been shown to directly interact with the CST by binding directly to Vps29 (82), regulating assembly and turnover of the nucleation complex by fine-tuning the Rab7 GTPase cycle.

The retromer plays an important role in different physiological processes e.g. cell polarity, iron transporter recycling, regulation of G protein-coupled receptor (GPCR) signaling etc (80). There are also many examples of the retromer or in more general the retrograde trafficking from endosomes to Golgi being hijacked or manipulated by pathogens during their infectious processes. *Salmonella* inhibits retrograde trafficking of CIMPR and lysosome function (83), Shiga toxin / cholera toxin uses the retrograde trafficking pathway to reach the Golgi and later the ER (84), *Legionella* effector RidL promotes intracellular replication by binding directly to Vps29 and PtdIns(3)P to inhibit retromer's function (85) and human papillomavirus (HPV) has been shown to

form a stable complex with the retromer during cell entry, later requiring the retromer to reach a Golgi-like compartment (86).

From our genome-wide siRNA screen (Research Article III), we showed that components of the retromer complex Vps35, Vps26a and to a lesser extent Vps29 had decreased *Brucella* infection upon knockdown. Furthermore, Vps35 did not show an effect in *Brucella* entry, suggesting that it is involved in a post-entry step. It is still unknown how *Brucella* traffics from an endosomal compartment to its ER-derived replicative niche. Components of the retrograde trafficking pathway have also not been implicated in *Brucella* trafficking to its replicative niche. Therefore, it would be interesting to investigate the role of the retromer complex in this context.

Results

Retromer complex and its interacting partner Rab7 are required for *Brucella* infection

Figure 1 depicts datasets from genome-wide screens (primary and secondary screens) for components of the retromer complex and some other components in context of the retrograde trafficking from endosomes to Golgi. Rac1 and mock (transfection reagent only) controls from all screens performed thus far were also shown. As seen, knockdown of individual components of the cargo selective trimer (CST) of the retromer complex that are the Vps35, Vps26a, and to a lesser extent Vps29 led to decreased *Brucella* infection, while Vps26b showed a much milder effect compared to Vps26a. Mock had no strong effect on *Brucella* infection while Rac1 that is known to be important for *Brucella* entry in non-phagocytic cells showed reduced infection upon knockdown (40). Sorting nexin proteins (SNX) associate with the CST to form the retromer complex (64). As seen, there was a mild effect of SNX3 and SNX2 in reducing *Brucella* infection upon knockdown, SNX5 and SNX6 depletion did not show a strong effect on infection, while SNX1 seemed to increase *Brucella* infection upon knockdown. Rab7, one of the factors involved in recruiting the retromer complex to the endosomal membrane via interaction with Vps35 (77) also had a strong effect on *Brucella* infection upon knockdown. Depletion of TBC1D5, which is a Rab GAP protein for Rab7 did not show an effect on infection while USP6NL, a Rab GAP protein for Rab43 shown to be involved in Shiga toxin transport from

Results: Retromer complex in *Brucella* infection

endosomes to Golgi (87) led to increased infection upon knockdown. These results suggest the importance of the retromer complex and its interacting partner Rab7 in *Brucella* infection.

Vps35, component of the retromer complex has a role in *Brucella* infection

To confirm the results from our siRNA screens, shRNA experiment was repeated and specificity of the knockdown was validated by complementation with cDNA encoding wild type Vps35. Figure 2A shows the constructs that were used in this experiment (88). This system allows co-expression of shRNA and the rescue cDNA, increasing the probability that efficient rescue occurs in the cell expressing the shRNA of interest. As shown in Figure 2B, shRNA against Vps35 reduced *Brucella* infection, validating the results from our RNAi screens. Expression of cDNA of wild type Vps35 is able to rescue this phenotype, with cells having similar infection levels as the empty vector. Rescue was also performed with cDNA of Vps35 mutant that lacks amino acids 237-252 (hVps35 Δ 6), the region where Vps35 associates with Rab7. It was shown that hVps35 Δ 6 localizes to the cytosol as compared to wild type Vps35 that goes to punctate endosomes. hVps35 Δ 6 is also not able to transport CIMPR from the endosomes to the perinuclear region(88). Therefore, hVps35 Δ 6 mutant upon losing its interaction with Rab7 loses its localization to endosomes and its function in retrograde sorting (88). Due to the large error bars with the dataset from the rescue experiment with hVps35 Δ 6, the experiment will be repeated before any conclusive statement is made of the role of Vps35's interaction with Rab7 on *Brucella* infection. Therefore, the ability of wild type Vps35 to rescue the phenotype of Vps35 knockdown confirms the specificity of our knockdown experiments and the role of this component in *Brucella* infection.

Retrograde transport is required for *Brucella* infection

To understand the general role of retrograde transport from endosomes to Golgi in *Brucella* infection, experiments were performed with Retro-2 inhibitor. Retro-2 has been shown to selectively block retrograde trafficking of toxin ricin, Shiga-like toxin, cholera toxin B subunit (CTxB) and human papillomaviruses from endosomes to trans-Golgi-network (TGN) (86, 89), without affecting retrograde cargoes, morphology of compartments or other trafficking steps. This compound has also been

shown to protect mice from lethal nasal exposure to ricin (89). Retro-2 was added to HeLa cells together with the addition of GFP-expressing *Brucella abortus* at 0 hour post infection (hpi) (Figure 3A) or at 4 hpi when cells were washed with gentamicin containing medium to kill extracellular bacteria (Figure 3B). The inhibitor was then kept throughout the experiment. As seen in both Figure 3A and 3B, preliminary data suggests that there was a dose-dependent decrease in *Brucella* infection upon increased doses of the inhibitor. This suggests a role of retrograde trafficking from endosomes to TGN in *Brucella* infection and since inhibition of infection is still seen when drug was added 4 hpi, this might suggest that this pathway is needed at a post-entry step during *Brucella* infection.

***Brucella* are contained within Lamp1 compartments upon Vps35 knockdown in HeLa cells**

As shown in Research Article III, Vps35 knockdown did not affect entry of *Brucella* into HeLa cells, suggesting a post-entry role of this component during infection. To understand this, cells depleted of Vps35 were fixed after 24 hpi with GFP-expressing *Brucella abortus* and stained for Lamp1 that marks the endo-lysosomal compartment. At 24 hpi, most of the BCVs are negative for markers of the endosomal-lysosomal system and have reached their replicative niche that is an ER-derived compartment (46). As expected, cells treated with scrambled siRNA showed BCVs devoid of Lamp1 markers at 24 hpi. For cells depleted of Vps35, GFP-expressing *Brucella abortus* were still found in Lamp1 positive compartments at 24 hpi, with some of the infected cells showing multiple bacteria in a vacuole. This preliminary result suggests that *Brucella* is unable to divert from the endosomal-lysosomal system upon knockdown of Vps35.

Discussion and outlook

The retromer complex plays an important role in different physiological processes, e.g. cell polarity, recycling of cargoes from endosomes to plasma membrane, having a major role in cargo transport from endosomes to Golgi (80). There are many examples of pathogens hijacking the retromer complex or in general the retrograde trafficking from endosomes to Golgi in benefit of their infectious process (83-86, 90).

Results: Retromer complex in *Brucella* infection

Here, our data suggests a role of the retromer complex component, Vps35 during *Brucella* infection, with preliminary results showing *Brucella* ending up in an endo-lysosomal compartment upon depletion of Vps35. We also showed that retrograde trafficking from endosome to Golgi is needed during *Brucella* infection.

Our results from siRNA screens showed the role of retromer complex and its associated components in *Brucella* infection. The consistent effect of reducing *Brucella* infection upon individual knockdown of Vps35, Vps26a and Vps29, which are components of the CST of retromer complex, confirms the role of the retromer complex during infection. It has been reported that Vps26a and Vps26b have preferences towards different cargoes (69). As we did not observe a similar reduction in *Brucella* infection upon depletion of Vps26b compared to Vps26a, it is possible that only Vps26a is needed during *Brucella* infection. SNX1, SNX2, SNX5 and SNX6 are components of the SNX-BAR-retromer complex (introduced in the introduction of this results part) while SNX3 is a component of the SNX3-retromer complex. SNX3 and SNX2 showed a mild decrease in infection upon knockdown while there is no strong effect with depletion of SNX5 or SNX6. SNX1 however increased infection upon depletion. The varying effects on *Brucella* infection of these proteins suggest there might be different roles among the SNX proteins. To confirm this, efficiency of siRNAs targeting these proteins should also be investigated with western blot studies. Combinatorial knockdown could be also done to check for redundancy between these SNX proteins. It would also be interesting to compare the roles of the SNX-BAR-retromer complex or SNX3-retromer complex during *Brucella* infection. USP6NL, Rab GAP protein for Rab43 that has been shown to be involved in Shiga toxin transport from endosomes to the TGN (87) increased infection upon knockdown. It could be that *Brucella* needs USP6NL in a similar manner as Shiga toxin transport and could be investigated in future.

To confirm the specificity of Vps35 knockdown and its effect on *Brucella* infection, we performed shRNA knockdown against Vps35 and rescue with complementary cDNA in HeLa cells. Vps35 is the core component of the CST that has a direct role in cargo binding (65, 66). It is also the first component that associates with SNX3 on endosomal membranes, responsible for the recruitment of the CST to an area where both SNX3 and Rab7 are present, forming the SNX3-retromer complex (79, 80). Therefore, depletion of Vps35 would perturb the formation and function of the

Results: Retromer complex in *Brucella* infection

retromer complex. We showed that shRNA knockdown of Vps35 is rescued upon complementation with wild type Vps35 containing a silent mutation in its sequence that renders it insensitive to shRNA knockdown. These results confirm the specificity of our knockdown and that the Vps35 subunit that is important during *Brucella* infection. To investigate the significance of Rab7-retromer complex interaction during *Brucella* infection, rescue was also performed with a mutant hVps35 Δ 6 that lost its interaction with Rab7. As Rab7 has been shown to be important to interact with Vps35 and its recruitment to the endosomes (77, 78), we would expect that this mutant is not able to rescue the effect of Vps35 shRNA knockdown. Other Vps35 mutants that are unable to interact with Vps29 subunit of the retromer complex (91) could also be used in a similar experiment to further investigate the importance of retromer assembly in the context of *Brucella* infection.

In addition, we have shown that Retro-2 inhibitor is able to reduce *Brucella* infection in a dose dependent manner. As this inhibitor is known to inhibit transport of ricin, Shiga-like toxin, CTxB and human papillomaviruses from early endosomes to the TGN (86, 89), its effects on *Brucella* infection indicates that the pathogen possibly utilizes similar components of the retrograde trafficking pathway or that it travels via a similar pathway during its course of infection. Addition of the inhibitor at 0 hpi or 4 hpi showed a similar effect in reducing *Brucella* infection, suggesting that the retrograde trafficking pathway is most probably needed post internalization during *Brucella* infection. This is consistent with our results that Vps35 was not required for *Brucella* entry in HeLa cells (Research Article III).

Furthermore, we showed that Vps35 knockdown caused *Brucella* to localize within a Lamp1 positive compartment at 24 hpi. This is unexpected since *Brucella* normally reaches the ER-derived replicative niche that is devoid of lysosomal markers at this time point (46). Therefore, this suggests that the retromer complex is important for *Brucella* trafficking to its replicative niche, with a dysfunction of this complex leading to *Brucella* ending up in an endo-lysosomal compartment. To analyze the properties this compartment, cells could be stained for cathepsin D that labels lysosomal enzymes. Interaction of *Brucella* with other known compartmental markers, e.g. for early endosomal markers, ER markers or autophagosomal markers should also be investigated upon depletion of Vps35.

In summary, we identified and confirmed a novel role of a component of the retromer complex, Vps35 during *Brucella* infection. Nevertheless, preliminary results showed that a functional retromer complex is crucial for *Brucella* escape of the endocytic pathway to reach its replicative niche since *Brucella* ends up in an endo-lysosomal compartment upon Vps35 depletion. Since it is still unclear how *Brucella* traffics from an endocytic compartment to an ER-derived replicative niche, future studies would focus on the role of the retromer complex or retrograde trafficking from endosome to Golgi in this context.

Materials & Methods

Materials

HeLa (human cervical carcinoma epithelial cell line, ATCC, CCL-2); Dulbecco Modified Eagle Medium (DMEM) (Sigma, D5796); Fetal Calf Serum (FCS)(Gibco, 10270): heat inactivated at 56°C for 30 min before use; tryptic soy broth (TSB)(Fluka, 22092); kanamycin sulfate (Sigma-Aldrich, 60615); gentamicin (Sigma, G1397); Triton-x-100, sigma-ultra (Sigma-Aldrich, T9284); DAPI (Roche, 10236276001); phalloidin-547 (Dyomics, 547PI-33); albumin from bovine serum (BSA)(Sigma, A9647); paraformaldehyde (Sigma, P6148); phosphate buffered saline (PBS)(Gibco, 20012); L-glutamine (Sigma-Aldrich, G7513); Fugene HD (Promega, E2312); mouse monoclonal anti-Lamp1 [H4A3] antibody (Abcam, ab25630); rabbit monoclonal anti-Rab7 (D95F2) antibody (Cell Signaling, 9367); Qiagen all stars negative control (SI03650318); Vps35 siRNA (Qiagen Hs_Vps35_2); Retro-2 (Calbiochem, 554715).

Plasmids

pCMS3.H1p/HA.YFP,pCMS3.H1p.shVPS35,pCMS3.H1p.shVPS35/HA.YFP.VPS35, pCMS3.H1p.shVPS35/HA.YFP.VPS35-Rab7mut re-exp are kind gifts from Daniel D. Billadeau (88).

Bacterial strains and cell lines

The bacterial strains used in this study include GFP expressing *Brucella abortus* 2308 that contains pJC43 with *gfp-mut3* gene under a constitutively active kanamycin resistance gene *aphA3* promoter (50) and DsRed expressing *Brucella abortus* 2308

Results: Retromer complex in *Brucella* infection

that contains pJC44 with *DsRed_m* gene from pDsRed_m (Clontech) under a constitutively active kanamycin resistance gene *aphA3* promoter (45).

HeLa cells were grown in DMEM (Sigma) supplemented with 10% FCS (Gibco), Cells were incubated at 37°C with 5% CO₂.

Cell culture and Infection

Brucella abortus were grown in TSB medium containing 50 µg/ml kanamycin for 24 h at 37°C and shaking (100 rpm) to an OD of 0.8- 1.1. Bacteria were added to cells with a final multiplicity of infection (MOI) of 10000 for HeLa cells. Plates were then centrifuged at 400xg for 20 min at 4°C to synchronize bacterial entry. After 4 h incubation at 37°C and 5% CO₂, extracellular bacteria were killed by exchanging the infection medium with DMEM (Sigma)/10% FCS (Gibco) supplemented with 100 µg/ml gentamicin. After a total infection time of 44 h cells were fixed with 3.7% PFA for 20 min at room temperature (RT).

Automated image analysis and Infection scoring

Images were taken with Molecular Devices ImageXpress microscopes using the 10X S Fluor objective, after which automated image analysis and decision tree infection scoring was performed as described in Research Article III. Binary level infection detection (infected vs. non infected) allows infection index (Infected cell / total cell number) to be defined.

Rescue experiment

Plasmids from the suppression / rescue system include empty vector, shVPS35, shVPS35/WT rescue and shVPS35 /Δ6 rescue are kind gifts from Daniel Billadeau (88). HeLa cells were seeded a day before transfection in a 6 well plate with 125,000 cells / well. The next day morning, 0.9 µg of plasmid (in 200 µl of DMEM without FCS) were mixed with 8 µl of Fugene HD (in 200 µl of DMEM without FCS) and incubated 15 min at RT. HeLa cells were exchanged with 1.5ml of fresh medium and DNA-fugene complex was added to the cells. The next day morning, cells were exchanged with fresh medium and 1 day later in the evening splitted into a 96 well format (2800 cells / well) for infection the next day.

Inhibitor experiment

HeLa cells were seeded in 96 well plates (2800 cells/well) one day before infection. Retro-2 was added to cells together with GFP expressing *Brucella abortus* 2308 or during gentamicin wash at 4 hpi and cells were maintained at 37°C with 5% CO₂. Retro-2 was kept throughout the experiment.

Immunofluorescent labeling

HeLa cells on coverslips were permeabilized with 0.1% Triton-x-100 for 10 min at RT, washed with PBS before incubated with 0.5% BSA/PBS for 30 min at RT. Afterwards, cells were labeled for Lamp1 using mouse monoclonal anti-Lamp1 [H4A3] antibody (1:100) and secondary antibody Alexa Fluor 546 Goat Anti-mouse IgG (1:100). For some experiments, cells were permeabilized for 10 min with 0.1% Triton-x-100 and stained with DAPI (final concentration 1µg/ml) for nucleus and/or phalloidin-547 (1:250) for F-actin.

Figure legends

Figure 1. Retromer complex and associated components in *Brucella* infection.

Box plot represents Z-score normalized infection scores of all available data points from all screens performed in Research Article II and Research Article III. Positive and negative controls Rac1 or scrambled siRNA are included for reference. The median is being represented at the middle of the box. Whiskers and outliers of boxplot are calculated with the Tukey method.

Figure 2. Vps35, component of the retromer complex has a specific role in *Brucella* infection.

A) Diagram of shRNA suppression / rescue constructs taken from (88). B) HeLa cells were transfected with the shRNA/rescue constructs and expressed for 2.5 days before infection with GFP-expressing *Brucella abortus* for 44 h. Automated image analysis with CellProfiler and decision tree infection scoring was performed with nucleus, perinucleus and voronoi cell as objects (details is described in materials and methods of Research Article III). A cell is detected as infected if

Results: Retromer complex in *Brucella* infection

pathogen intensity in any of these objects exceeds a certain threshold that is set manually. Data represents normalized infection in reference to empty vector, the mean \pm STDEV of three independent experiments.

Figure 3. Retrograde transport is required for *Brucella* infection. HeLa cells were incubated with Retro-2 A) during addition of bacteria to the cells (0 hpi) or B) at 4 hpi when cells were washed with gentamicin containing medium. The inhibitor was kept in the medium throughout the rest of the experiment. Cells were fixed around 44 hpi. Automated image analysis with CellProfiler and decision tree infection scoring was performed with nucleus, perinucleus and cell body as objects. A cell is detected as infection if pathogen intensity in any of these objects exceeds a certain threshold that is set manually. Data represents infection rates of two independent replicas for A and only one replica for B. Error bars of each bar graph represents mean \pm STDEV of at least six technical replicates within the experiment.

Figure 4. *Brucella* is contained within Lamp1 compartments at 24 hpi upon silencing of Vps35 in HeLa cells. Representative images of HeLa cells infected with GFP-expressing *Brucella abortus* for 24 h after 72 h of transfection with scrambled or Vps35 siRNA. Cells were then immunostained for Lamp1 (red) and images were taken with the 60x objective and FEI MORE with TIRF microscope. Image in stacks were deconvolved with HUVGENs remote manager and one represented slice around the middle of a stack is shown. Scale bars represent 10 μ m.

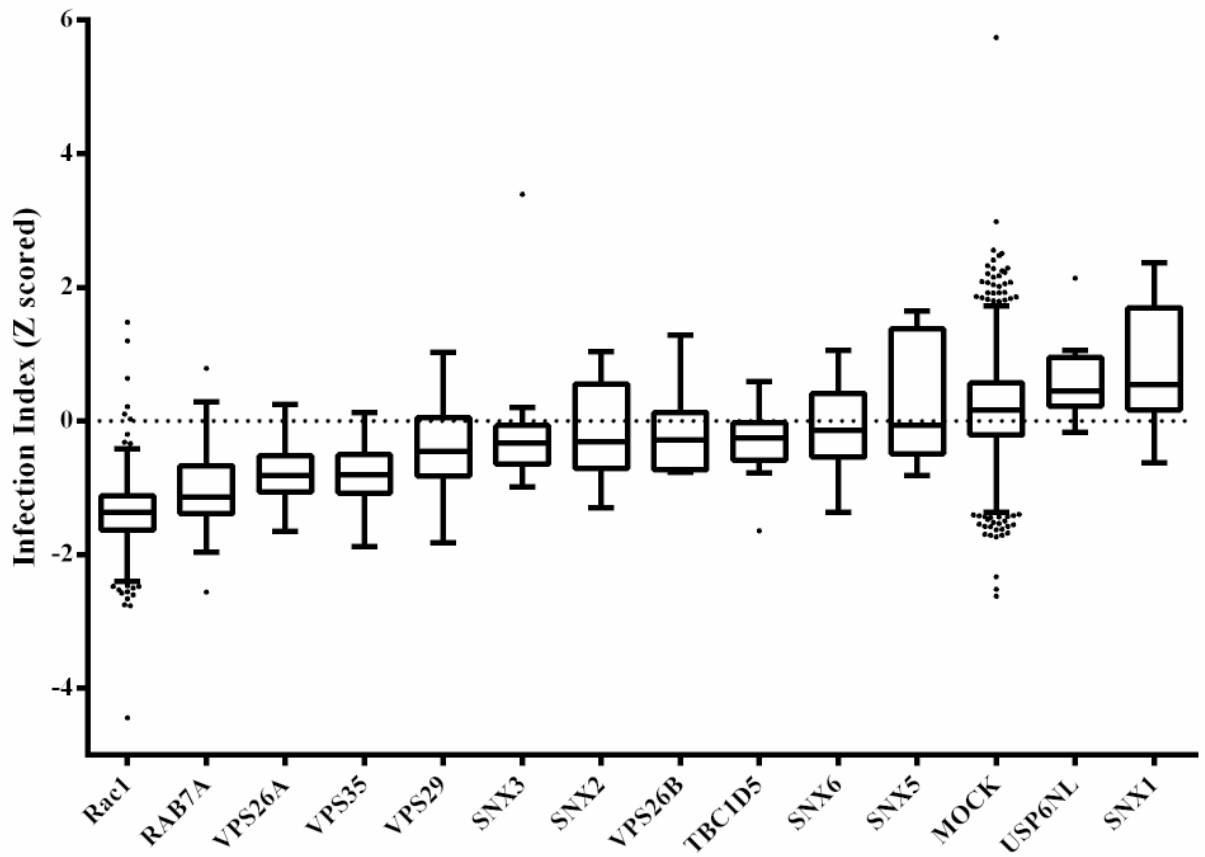
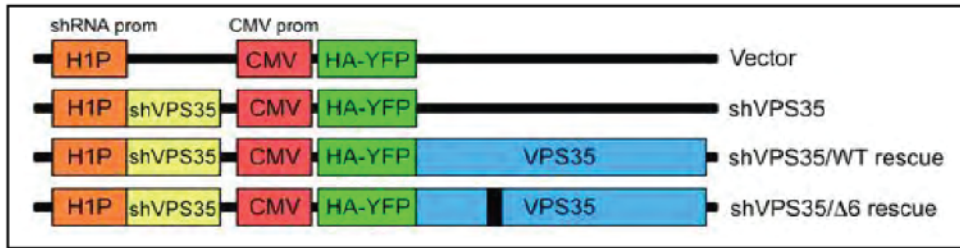


Figure 1

Results: Retromer complex in *Brucella* infection

A



adapted from Liu *et al* Mol Biol Cell 2012

B

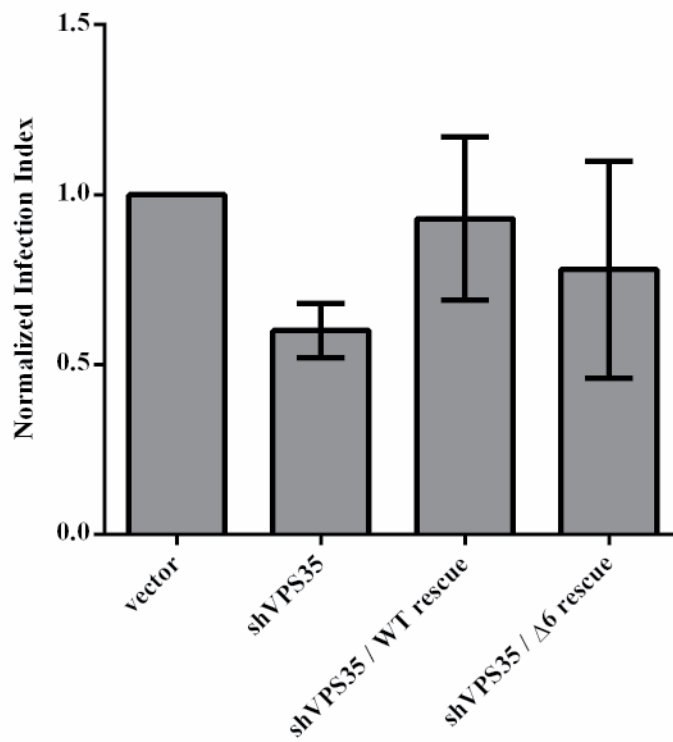
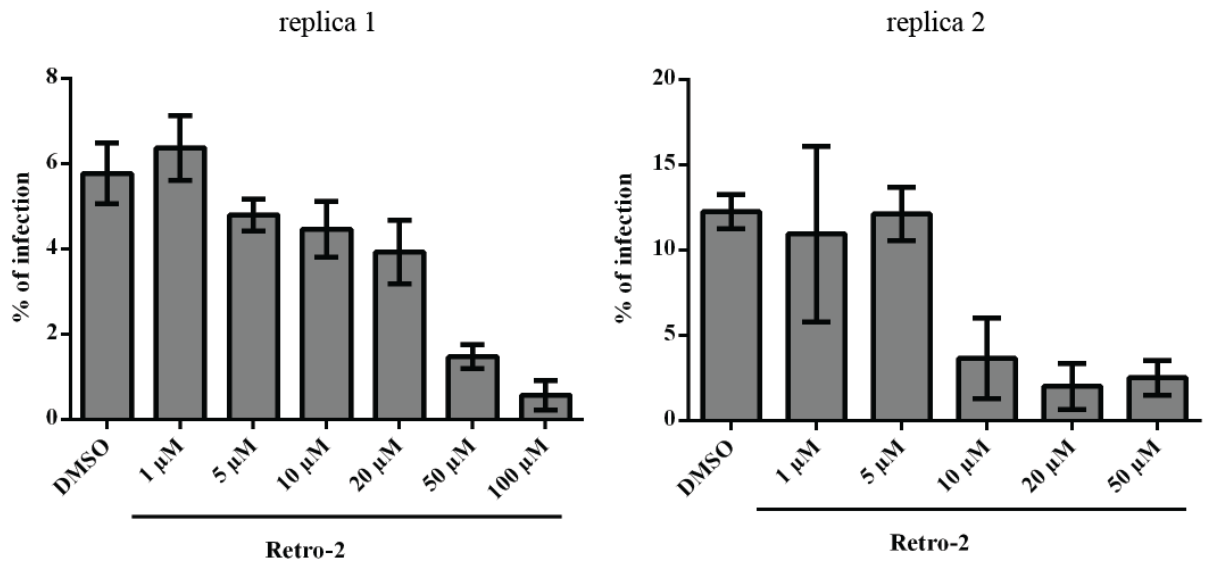


Figure 2

Results: Retromer complex in *Brucella* infection

A



B

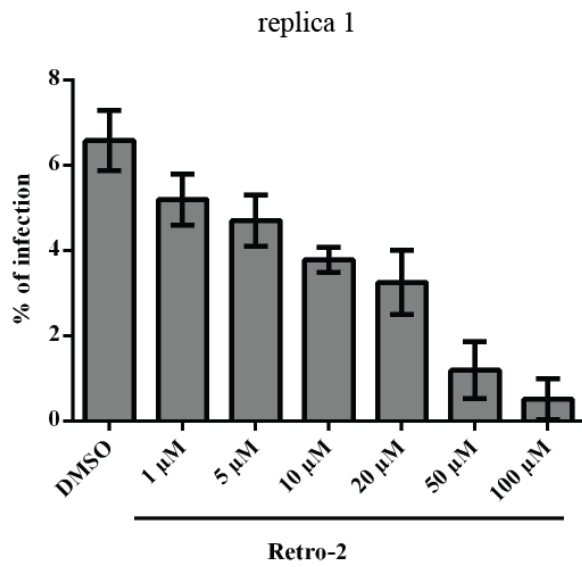


Figure 3

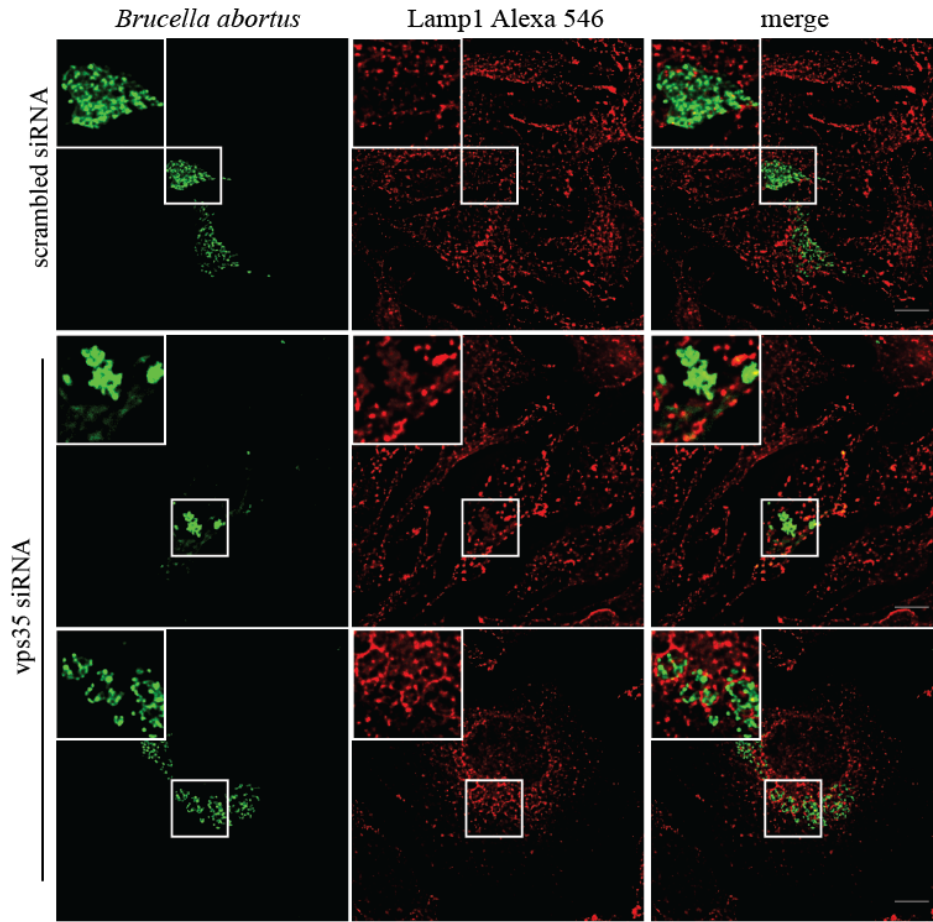


Figure 4

3.6 Additional tools developed – understanding *Brucella* intracellular lifestyle by fluorescent microscopy and / or electron microscopy

Results

HeLa cells stably expressing fluorescently labeled cellular compartments

To further understand *Brucella* interaction with different compartments of the host cell, HeLa cells stably expressing fluorescent markers labeling the entire endocytic pathway, endoplasmic reticulum (ER) and ER-Golgi intermediate compartment (ERGIC) were generated. A general introduction can be found in the Section 1.1 of the components of these trafficking pathways. These include the markers for lipid rafts (caveolin-GFP), early endosomes (pmRFP1-Rab5, TfR-GFP), late endosomes and / or lysosomes (Rab7a-GFP, pmRFP1-Rab7, Igp120-mCherry, Lamp1 GFP, Lamp1-YFP), ER (RFP-RTN2, calnexin-GFP) and ERGIC (Rab2A-GFP, Rab2B-GFP, GFP-ERGIC53). Multiple marker cell lines expressing e.g. pmRFP1-Rab7 and GFP-ERGIC53 were also successfully generated (Figure 1). Stable expression of the gene of interest was achieved via lentiviral transduction. This method has been introduced in Section 1.3.1.2 under the section of shRNA that is also delivered via the same method into the host cell. Stable expression of the gene of interest labeled with fluorescent markers is advantageous for live cell imaging and correlative light and electron microscopy (CLEM) studies.

HeLa pmRFP1-Rab7: *Brucella* containing vacuole (BCV) interacts with Rab7 at early and late time points of the infection

Rab7 is a late endosomal marker that has been shown to be required for *Brucella* infection (45, 92). BCVs interact with Rab7 at early time points of the infection (with a peak at around 6 hpi) and the marker is lost before arriving at its replicative niche (45). Additionally, it has also been reported that at late time points of infection (72 hpi) BCVs re-acquire Rab7 (92). The role of this interaction at late time points is unclear. Since autophagosomal like membranes were also found at later time point to surround 20% of the BCVs (92), we would like to investigate whether Rab7 positive BCVs are

Results: Additional tools developed

the ones that also possess autophagosomal features. BCVs with autophagosomal features at late time points are suggested to be responsible for egression of *Brucella* (92).

To address this question, HeLa pmRFP1-Rab7 cells were used. HeLa pmRFP1-Rab7 stable cell line was validated for its localization at late endosomes by staining for Lamp1 that also labels late endosomes. A perfect co-localization of Rab7-RFP and endogenous Lamp1 was seen suggesting a proper localization of this marker via this system (Figure 2iA). At 6 hours post infection (hpi), Rab7 was acquired by some of the BCVs, with most of the BCVs not having this marker at 24 hpi. At 48 hpi and 72 hpi, most of the BCVs do not contain Rab7. However, there are some BCVs that are surrounded by Rab7 (Figure 2iB).

Using the example of a sample from 6 hpi, focused-ion beam-scanning electron microscope (FIB-SEM) was used to scan the area of interest in an infected cell (electron microscopy studies are all performed by Jaraslow Sedzicki, PhD student in Henning Stahlberg's group). With the focused ion beam, sections of the cell could be obtained, after which images were scanned, finally giving the 3D volume of the cell (93). Electron micrographs of the area that is highlighted with a box in the image taken by fluorescent microscopy (Figure 2iiCi) were taken in a stack with Helios Nanolab 650 dual beam microscope. As seen in Figure 2iiCi, there are six bacteria labeled 1-6, with bacteria 3 and 5 being negative for Rab7, while the other bacteria 1, 2, 4, 6 are all surrounded by Rab7 positive membrane. 3D model of the bacteria at the region of interest was obtained from the EM images (Figure 2iiCii), which is similar to the 3D model obtained with the fluorescent image (data not shown). These models allow correlation of the electron micrographs with the fluorescent images. As seen in Figure 2iiCiii, Rab7 positive BCVs (bacteria 3 and 5) showed only a single membrane surrounding the entire volume of this BCV, similar to a neighboring BCV that is not surrounded by Rab7 (bacteria 1, 2, 4, 6) (Figure 2Ciii). Experiments with 24 hpi, 48 hpi and 72 hpi are still ongoing and are not shown here. These preliminary results suggest that CLEM could be successfully performed and is definitely a good tool for future investigations of Rab7 and its role in different stages of *Brucella* infection. The CLEM method could also be extended to understand the intracellular lifestyle of *Brucella* under different experimental conditions.

HeLa Igp120 mCherry: Validation with wild type *Brucella abortus* or *Brucella abortus* $\Delta virB9$ mutant infection

In literature, it is known that BCV is surrounded by Lamp1 markers at early time points (6 hpi) in both non phagocytic and phagocytic cells and gradually loses it with time, with most of the BCVs excluding the Lamp1 marker at 24 hpi (46). To ensure that HeLa Igp120 (another name for Lamp1) mCherry stable cell line behaves like endogenous Lamp1, cells were infected with GFP-expressing *Brucella abortus* or GFP expressing *Brucella abortus* $\Delta virB9$ and fixed at 6 hpi or 24 hpi (Figure 3). *Brucella abortus* $\Delta virB9$ mutant contains a deletion in the *virB9* gene of the virB type IV secretion system (T4SS), leading to a dysfunctional secretion system. $\Delta virB9$ mutant is unable to replicate in mammalian cells and has been shown to retain Lamp1 at 24 hpi (46, 94). As expected, at 6 hpi GFP-expressing *Brucella abortus* and GFP expressing *Brucella abortus* $\Delta virB9$ were both surrounded by Lamp1. At 24 hpi, most of the GFP-expressing *Brucella abortus* were negative for the Lamp1 while all GFP-expressing *Brucella abortus* $\Delta virB9$ were in Lamp1 positive compartment (Figure 3). This is consistent with what has been shown for *Brucella* interaction with endogenous Lamp1, suggesting that HeLa Igp120-mCherry behaves similar to endogenous Lamp1 during *Brucella* infection and could be used for further experiments.

HeLa / THP-1 calnexin-GFP: *Brucella* is contained within calnexin positive vesicles at late time points of the infection

As mentioned in the general introduction section 1.2, *Brucella* interacts and replicates within an ER-like compartment (46). Aerolysin toxin treatment experiments that vacuolated the ER caused individual BCVs to fuse into giant vacuoles covered with ribosomes, suggesting that BCVs display ER membrane properties (46, 95). However, till date, it is still unclear whether *Brucella* fuses and replicates within the ER or only within a sub-compartment of the ER since not all ER proteins were detected on the BCVs. For example, ER proteins that are commonly seen surrounding BCVs include calnexin (integral protein and ER chaperone), sec61 β (ER translocon), calreticulin (ER chaperone that might also serve functions outside of the ER), protein disulfide-isomerase (PDI) (enzyme ER resident protein) and KDEL (tag for ER proteins) (46, 95, 96), while BiP (chaperone in the ER lumen) or ribophorin (glycoproteins only

present on the membrane of rough but not smooth ER) were reported to not be present on the BCVs (95).

To understand the interaction of *Brucella* with the ER, first experiments were performed to investigate localization of BCV with ER marker calnexin at 44 hpi. This was performed with stable cell lines HeLa and THP-1 macrophage-like cell line stably expressing calnexin-GFP. THP-1 cell line is a human monocytic cell line that could be differentiated with phorbol myristate acetate (PMA) into macrophage like cells. To understand *Brucella* infection of macrophages that are physiologically relevant cell line, we also performed experiments with THP-1 macrophages. As seen in Figure 4A, each BCV was surrounded by ER marker calnexin at 44 hpi in HeLa cells. Consistent with previous studies (97), BCVs were also surrounded by calnexin-GFP in THP-1 suggesting that *Brucella* also replicates in an ER positive compartment in THP-1 (Figure 4B). THP-1 macrophage-like cell line stably expressing calnexin-GFP and infected with *Brucella* was labeled with gold particles against GFP for EM studies (performed by former PhD student of Henning Stahlberg's group - Christopher Bleck). As shown in Figure 4C, gold particles were surrounding the BCV membrane suggesting that BCV is surrounded by calnexin-GFP. Therefore, consistent with previous studies that show *Brucella* surrounded by calnexin positive structures (46, 96), calnexin-GFP stable cell lines shows colocalization with *Brucella* at 44 hpi, suggesting it as a promising tool to study *Brucella* interaction with the ER.

HeLa RFP-reticulon 2 (RTN2): *Brucella* is surrounded by RTN2 positive membrane at late time points of the infection

To further understand the interaction of *Brucella* with the ER, HeLa cells stably expressing another ER marker RFP-RTN2 was generated. As introduced in Section 1.1, the ER is comprised of different subcompartments. RTN2 is a tubular ER-shaping protein that is largely restricted to tubular ER and excluded from the continuous sheets of the nuclear envelope and peripheral ER (98). HeLa RFP-RTN2 was infected with GFP-expressing *Brucella abortus* for 24 h or 48 h as seen in Figure 5A. At 24 h, most of the *Brucella* was not seen to interact with RTN2, while at 48 h it seems that RTN2 surrounds multiple bacteria in a bigger vacuole. This suggests an interaction of RTN2 with *Brucella*, but in a different manner as was seen with calnexin (Figure 4).

HeLa RFP-RTN2 was stained with anti-BAP31 or anti-CLIMP63 antibodies to compare its localization with endogenous ER proteins. CLIMP63 is an integral membrane protein that localizes to the reticular part of the ER and links the ER to microtubules (99, 100) while BAP31 is a component of the ER quality control compartment that moves between the peripheral ER and juxtannuclear ER or ER-related compartment along the microtubule tracks (101). As seen in Figure 5B, RTN2 showed a similar localization to BAP31 but not to CLIMP63, suggesting different localization patterns of these ER proteins. To compare interaction of GFP-expressing *Brucella-abortus* between these different ER markers, cells were stained with anti-BAP31 or anti-CLIMP63 antibodies. At 24 hpi and 48 hpi, most of the GFP-expressing *Brucella abortus* did not interact with BAP31 or CLIMP63, with only a few seen having these ER markers surrounding the BCV (Figure 6 and Figure 7). This is different from what we have seen with calnexin and RFP-RTN2, where at 44 hpi, each bacterium was surrounded by calnexin (Figure 4) and for RTN2 at 48 hpi, multiple *Brucellae* were already seen in a bigger vacuole positive for RTN2 (Figure 5). At 72 hpi, multiple *Brucellae* were seen to localize to BAP31 or CLIMP63 positive vacuoles similar to that seen with RTN2 at 48 hpi (Figure 6 and 7).

Discussion and Outlook

To understand the interaction of *Brucella* with different intracellular compartments of the host cell, HeLa stable cell lines expressing markers for the compartments that are known to interact with *Brucella* were generated. These cell lines have been successfully used for live cell imaging studies by Houchaima Ben Tekaya (postdoc in our group) to understand the dynamics of *Brucella* interaction with the host and also for CLEM studies (in collaboration with Jarek Sedzicki, PhD student in Henning Stahlberg's group). Experiments that were performed by myself with these stable cell lines include confirming the interaction of *Brucella* with Rab7 at early and late time points with CLEM studies, validation of stable cell line Igp120-mcherry with *Brucella* infection and understanding the interaction of *Brucella* with different ER markers calnexin-GFP, RFP-reticulon2 (RTN2), and endogenous CLIMP63 or BAP31.

Results: Additional tools developed

HeLa cells stably expressing pmRFP1-Rab7 were used to understand the different roles of Rab7 interaction with the BCV at different time points. BCV's association with Rab7 markers at early and late time points (Figure 2iB) was consistent with previous studies (45, 92). With CLEM studies using FIB-SEM in the electron microscopy part, we were able to show that at 6 hpi Rab7 positive BCVs have a similar membrane structure as Rab7 negative BCVs throughout the volume of a BCV, with a single membrane surrounding the bacteria (Figure 2iiCiii). This is expected since it has not been reported that Rab7 positive and Rab7 negative BCVs should be surrounded by a different membrane structure at this early time point of infection. Next, we would like to focus on the difference in the membrane surrounding Rab7 positive and negative BCVs at later time points of infection: 48 hpi and 72 hpi. Since autophagosomal structures have been shown to be enriched at late time points of infection, at a similar time point where egression and Rab7 acquisition was seen (92), it could be that Rab7 positive BCVs are the ones that also possess autophagosomal structures, represented normally as multiple membrane structures. This will be investigated in the near future. Nevertheless, we could also repeat this experiment with HeLa Igp120-mcherry stable cell line since Lamp1 was also acquired again similar to Rab7 at late time points of *Brucella* infection (92). The ability to perform CLEM studies allows us to specifically correlate fluorescent images of interest with electron microscopy studies. Together with the stable cell lines available, we have powerful tools that would allow us to understand at an ultrastructural level how *Brucella* interacts with different compartments of the host. This could be extended to different experimental conditions for example siRNA knockdown or drug treatment. Experiments were also performed with GFP-expressing *Brucella abortus* and GFP-expressing *Brucella abortus* $\Delta virB9$ to validate HeLa cells stably expressing Igp120-mcherry. Localization of these Lamp1 markers at BCVs were consistent with published reports, with GFP-expressing *Brucella abortus* devoid of Lamp1 markers while GFP-expressing *Brucella abortus* $\Delta virB9$ (46). This suggests that the ectopically expressed Igp120-mcherry has a similar behavior as the endogenous Lamp1 and the cell line could be used for example to validate our screen, to identify genes that upon knockdown cause *Brucella* to stay in an endo-lysosomal compartment and unable to reach its replicative niche.

Results: Additional tools developed

The nature of the replicative niche of *Brucella* is still not well understood. It is still unclear whether the BCV fuses with the ER or it acquires certain markers of the ER. If the latter is true, it is not clear why certain ER proteins are preferably acquired by the BCV than others. Calnexin is a protein that has been used commonly as a marker that localizes to the BCV (45, 46, 95, 96) while the interaction with other ER markers RTN2, CLIMP63 and BAP31 has not been investigated so far. The other ER markers, RTN2 and CLIMP63 are proteins that shape the ER tubules and ER sheets respectively (102). Therefore, we would expect that if *Brucella* is in the ER, we should be able to see these proteins surrounding the BCV. BCV showed a different colocalization pattern with RFP-RTN2 as compared to calnexin-GFP. At 48 hpi, *Brucella* was surrounded by RFP-RTN2. However, compared to calnexin that showed each bacterium surrounded by a calnexin-positive vacuole, multiple bacteria seemed to be in a big vacuole positive for RFP-RTN2. Comparison of localization of RTN2 with CLIMP63 and BAP31 showed RTN2 localizing at a similar ER subcompartment as BAP31 but not CLIMP63. Since RTN2 is normally in ER tubules while CLIMP63 is in ER sheets (102), it is expected that they show a different localization in the ER. BAP31 is a protein that regulates protein export from the ER, having shown similar localization as RTN2 suggest that it is also at the tubular network of the ER. BAP31 and CLIMP63 showed a similar localization as RTN2 at 72 hpi, with *Brucella* seen in a big vacuole surrounded by these markers. It would be interesting to see with CLEM studies whether there are indeed multiple bacteria in the same vacuole positive for RTN2, CLIMP63 or BAP31 at late time points of infection, possibly suggesting that fusion with the ER occurred at these time points. In summary, *Brucella* showed different interaction patterns with the different ER markers that were investigated here.

Proteins that have been reported to interact with the BCV include ER translocon sec61 β (95), ER chaperones calnexin, calreticulin, protein disulfide isomerase (PDI) (45, 46, 95, 96) and KDEL which is a tag that recycles between the Golgi and the ER to return ER resident proteins to the ER (96, 103). It is unclear why certain ER chaperones are interacting with the BCV while others, for example binding immunoglobulin protein (BiP), do not interact with *Brucella* (95). It could be that calnexin or sec61 β being an integral protein of the ER are more easily accessible to the BCV than the luminal BiP. This also might be the case for proteins such as

calreticulin (104) or KDEL that are able to exit the ER and might have higher possibilities to localize around the BCV. Since all these proteins that have been shown to localize around the BCV are absent from the ER exit sites (ERES) but on the other hand it has been shown that BCV interacts also with the ERES (50), this suggests again that *Brucella* interacts with different subcompartments of the ER. Since there was reorganization of the ER seen at later time points with BAP31 or CLIMP63 (Figure 6 and 7), it could be that *Brucella* first acquires specific ER markers due to their proximity to the BCV at earlier time points and later fuses with the ER at later time points of infection. Further studies would be required to understand whether ER markers are partially or completely surrounding the BCV. It could be that at earlier time points, there is only a partial interaction with the ER while at later time points, the massive replication of bacteria leads to complete acquisition of ER membranes around the BCV. This could be investigated with CLEM studies with specific markers of interest. Also, it would be interesting to see if *Brucella* starts to fuse with the ER at later time points of infection, in that case, we would expect to see multiple bacteria in a big vacuolar compartment at late time points of infection.

In summary, HeLa cells stably expressing fluorescent markers for various cellular compartments are useful tools that could be used for live cell imaging, CLEM as well high-throughput assays. Some of the stable cell lines are validated to be functional as endogenous proteins and localize to BCVs at expected time points. It would now be valuable to utilize these cell lines to understand and follow the effect of different experimental conditions on the intracellular trafficking pathway of *Brucella*.

Materials and Methods

Materials

In-fusion HD cloning kit (Clontech, 639649), HeLa (human cervical carcinoma epithelial cell line, ATCC, CCL-2), THP-1 (human monocytic leukemia cell line, ATCC, TIB-202); human embryonic kidney 293T (HEK-293T)(from Hwain Cornelis's lab); Dulbecco Modified Eagle Medium (DMEM) (Sigma, D5796); Dulbecco Modified Eagle Medium Glutamax (DMEM Glutamax)(Gibco, 61965-026); RPMI-1640 medium (Sigma, R0883); Fetal Calf Serum (FCS)(Gibco, 10270): heat

Results: Additional tools developed

inactivated at 56°C for 30 min before use; Fetal Calf Serum (FCS)(Bioconcept, 2-01F30-I); tryptic soy broth (TSB)(Fluka, 22092); kanamycin sulfate (Sigma-Aldrich, 60615); ampicillin sodium salt (Applichem, A.8039.0025); gentamicin (Sigma, G1397); DAPI (Roche, 10236276001); paraformaldehyde (Sigma, P6148); phosphate buffered saline (PBS)(Gibco, 20012); L-glutamine (Sigma-Aldrich, G7513); phorbol myristate acetate (PMA)(Sigma, P8139); EcoRI (New England Biolabs, R3101); BamHI (New England Biolabs, R3130); polybrene (Sigma, H9268); Triton-x-100, sigma-ultra (Sigma-Aldrich, T9284); albumin from bovine serum (BSA)(Sigma, A9647); mouse monoclonal anti-Lamp1 [H4A3] antibody (Abcam, ab25630), anti-CLIMP63 mouse monoclonal antibody (kind gift from Hauri lab) (105); anti-BAP31 mouse monoclonal antibody (kind gift from Hauri lab)(106); Alexa Fluor 546 Goat Anti-mouse IgG (Molecular probes, A-11030); Alexa Fluor 488 Anti-mouse IgG (Molecular probes); Formaldehyde (Electron Microscopy Sciences (EMS), 15710); Glutaraldehyde (EMS, 16000); PIPES (Sigma, P8203); HEPES (AppliChem, A3724); EGTA (Fluka , 03779); MgCl₂ hexahydrate (M9272); sodium cacodylate (SERVA, 1554002); calcium chloride anhydrous (499609); potassium ferrocyanate (Sigma, P3289); osmium tetroxide (EMS, 19170); thiocarbohydrazide (Sigma, 88535); uranyl acetate (Fluka, 73943), coverslips for CLEM studies (LUCERNA-CHEM AG, 72265-25)

Walton's lead aspartate solution preparation:

10ml 0.03M L-Aspartic acid solution (Sigma, A9256) at 60°C

add 0.066g Lead nitrate (EMS, 17900), leave in 60°C oven for 30 min (mix from time to time), filter through a 0.22um filter

Durcupan resin preparation (purchased as 4 different components with commercial names):

10g Part A (Fluka, 44611)

10g Part B (Fluka, 44612)

0.3g Part D (Fluka, 44614)

mix everything

add 16 drops of activator DMP-30 (EMS, 13600)

mix again

embed samples and polymerize for 48h at 60°C

Bacterial strains and cell lines

The bacterial strains used in this study include GFP expressing *Brucella abortus* 2308 that contains pJC43 with *gfp-mut3* gene under a constitutively active kanamycin resistance gene *aphA3* promoter (50), DsRed expressing *Brucella abortus* 2308 that contains pJC44 with *DsRed_m* gene from pDsRed_m (Clontech) under a constitutively active kanamycin resistance gene *aphA3* promoter (45) and GFP expressing *Brucella abortus* 2308 $\Delta virB9$ mutant that contains pJC43 (50). DH5 α used for cloning experiments contains genotype $\phi 80\text{dlacZ}\Delta M15$, *recA1*, *endA1*, *gyrAB*, *thi-1*, *hsdR17* (rK⁻, mK⁺), *supE44*, *relA1*, *deoR*, $\Delta(\text{lacZYA-argF})$ U169, *phoA* (N. Mantis, Institut Pasteur).

HeLa cells were grown in DMEM (Sigma) supplemented with 10% FCS (Gibco), THP-1 cells with RPMI-1640 medium supplemented with 10% FCS (Gibco) and 10mM L-glutamine and HEK293T cells with DMEM Glutamax (Gibco) supplemented with 10% FCS (Bioconcept). Cells were incubated at 37°C with 5% CO₂. THP-1 monocytes could be differentiated into a macrophage-like cell line with PMA at a final concentration of 10⁻⁷ M and 48 h incubation at 37°C with 5% CO₂ (51).

Plasmids

Different genes of interest fused to fluorescent markers were inserted into lentiviral vector pMDK124 (kind gift of Professor Oliver Pertz, unpublished). pMDK124 was digested with EcoRI and BamHI overnight at 37°C and gel purified. Polymerase chain reaction (PCR) was performed with gene-specific primers that contain 15bp extensions complementary to the vector digested ends and gel purified. Using the gel purified PCR products and digested pMDK124, In-Fusion recombination was performed with 200ng of vector and PCR product (amount depends on the size of the PCR product, <0.5kb: 10-50ng, 0.5kb to 10kb: 50-100ng) using In-Fusion enzyme mixture from In-Fusion HD cloning kit, incubated 15 min at 50°C and then transferred to ice. 5 μ l of the reaction mixture was transformed into 100 μ l of DH5 α and plated on ampicillin containing LA plates.

Results: Additional tools developed

PCR products were produced from different templates with primers that are listed in Table 1 of this results section, the PCR products were then recombined with pMDK124, giving the final lentiviral vectors containing genes fused to fluorescent markers. For pMDK124 expressing caveolin-1 GFP (pSL012), prSL061 and prSL062 were used for PCR amplification with template caveolin-GFP in pEGFP-N1 (a kind gift from Professor Ari Helenius (107)) before recombination with pMDK124. For pMDK124 expressing pmRFP1-Rab5a (pSL055), prSL032 and prSL036 were used for PCR amplification with template pmRFP1-Rab5a in pEGFP-C3 (a kind gift from Professor Ari Helenius (108)). For pMDK124 expressing TfR-eGFP (pSL051), prSL001 and prSL002 were used for PCR amplification with template TfR-eGFP in pNF314 (a kind gift from Professor Gary Banker (109)). For pMDK124 expressing Rab7A-GFP (pSL054), prSL032 and prSL053 was used to amplify template Rab7A-GFP (a kind gift from Dr. P. Boquet). For pMDK124 expressing pmRFP1-Rab7 (pSL057), prSL032 and prSL033 were used for PCR amplification from template pmRFP1-Rab7 in pEGFP-C3 (a kind gift from Professor Ari Helenius (108)). For pMDK124 expressing Igp120 mcherry (pSL047), prSL003 and prSL004 were used to PCR amplify template Igp120 mcherry (a kind gift from Dr. G. Patterson). For pMDK124 expressing Lamp1-GFP (pSL056), prSL039 and prSL040 were used to PCR amplify template Lamp1-GFP (a kind gift from Dr. P. Boquet). For pMDK124 expressing Lamp1-YFP (pSL058), prSL078 and prSL079 were used to PCR amplify Lamp1-YFP in pEYFP-N1 (addgene ID: 1816 (110)). For pMDK124 expressing RFP-RTN2 (pSL037), prSL072 and prSL073 were used to PCR amplify from template RFP-RTN2 (source is unknown). For pMDK124 expressing calnexin-GFP (pSL060), prSL086 and prSL087 were used to PCR amplify template calnexin-GFP in pEGFP-N1 (kind gift from Dr. G. Van der Goot). For pMDK124 expressing Rab2A-GFP (pSL052) or Rab2B-GFP (pSL053), prSL045 and prSL046 were used to PCR amplify template Rab2A-GFP (a kind gift from Professor Francis A. Barr (111)), while prSL045 and prSL050 were used to PCR amplify template Rab2B-GFP (a kind gift from Professor Francis A. Barr (111)). For pMDK124 expressing GFP-ERGIC53 (pSL049), prSL021 and prSL022 were used to PCR amplify from template PRL-EGFP-ERGIC53 (from Houchaima Ben Tekaya (112)). Helper plasmids pVSV, pMDL and pRev are kind gifts from Oliver Pertz's group (unpublished).

Infection

Brucella abortus were grown in TSB medium containing 50 µg/ml kanamycin for 24 h at 37°C and shaking (100 rpm) to an OD of 0.8- 1.1. Bacteria were added to cells with a final multiplicity of infection (MOI) of 10000 for HeLa cells and MOI 1000 for THP-1 macrophage like cell line. Plates were then centrifuged at 400xg for 20 min at 4°C to synchronize bacterial entry. After 4 h incubation at 37°C and 5% CO₂, extracellular bacteria were killed by exchanging the infection medium with DMEM (Sigma)/10% FCS (Gibco) supplemented with 100µg/ml gentamicin. After a total infection time of 44 h cells were fixed with 3.7% PFA for 20 min at RT.

Lentiviral transduction

3x10⁶ HEK293T cells were grown in a 10cm dish with DMEM supplemented with 10% FCS and incubated at 37°C, 5% CO₂ for at least 6-8 hours. 2.2 µg of lentiviral vector pMDK124 containing respective gene of interest, 0.75 µg of pVSV, 1.5 µg of pMDL, 0.5 µg of pREV in 600 µl DMEM without FCS were then mixed with 25µl of Fugene HD (Promega) in 600 µl DMEM without FCS and incubated for 15 min at room temperature. The DNA-fugene complex was then added to the cells that were replaced with 5ml of fresh medium. Cells were exchanged with fresh medium the following day. 2 days later, supernatant of HEK293T that now contain viruses was collected and filtered through a 0.45µm membrane filter. Viruses were then used directly for transduction of cells or stocked in tubes at -80⁰C. Viral containing supernatant (1ml or 2ml volumes) were added to cells in presence of fresh medium to a total volume of 3ml, and polybrene with final concentration of 5µg/ml. Cells were exchanged with fresh medium the next day and could be used for further experiments.

Immunofluorescent labeling

HeLa cells on coverslips were permeabilized with 0.1% TritonX for 10 min at room temperature, washed with PBS before incubated with 0.5% bovine serum albumin (cat no) for 30 min at room temperature. Afterwards, cells were labeled for Lamp1, CLIMP63, BAP31 or Rab7 antibody. For electron microscopy experiments, cells were stained with DAPI (Roche, final concentration 1µg/ml) for DNA of bacteria or cell without permeabilization of the cell.

Correlative Light and Electron microscopy (CLEM)

HeLa cells were grown and infected on gridded coverslips in a 6 well plate, fixed with 4% formaldehyde and 0.1% glutaraldehyde in 1xPHEM buffer for a total of 110 mins. Next, samples were passed for electron microscopy studies. Specimens were fixed in PHEM buffer (4% formaldehyde, 0.2% glutaraldehyde, 80mM PIPES 25mM HEPES adjusted to pH 6.9, 10mM EGTA, 2mM MgCl₂) for 30 min at room temperature. After several rinses in cold buffer (150mM sodium cacodylate pH 7.4, 2mM calcium chloride), the specimens were immersed in freshly prepared reduced osmium tetroxide buffer (3% potassium ferrocyanate, 150mM cacodylate, 4mM calcium chloride, 2% osmium tetroxide) for 1 h on ice. After several water rinses at room temperature, the samples were immersed in freshly prepared 0.01% thiocarbohydrazide solution for 20 minutes at room temperature. The samples were then washed several times with water and postfixed with 2% aqueous osmium tetroxide for 30 min at room temperature. After subsequent water rinses, the samples were placed in 1% aqueous uranyl acetate and stored overnight at 4°C. On the following day, the samples were immersed in freshly prepared Walton's lead aspartate solution (30mM lead aspartate, adjusted to pH 5.5) for 30 min at 60°C. The coverslips was then washed with water, and dehydrated with ethanol, followed by embedding in Durcupan.

The coverslips were removed from polymerized resin block using liquid nitrogen. Cells of interest found previously using fluorescent microscopy were traced back under a light microscope. The sample blocks were trimmed and attached to SEM stubs. FIB-SEM data was obtained using a Helios Nanolab 650 dual beam microscope (FEI, Eindhoven, the Netherlands).

The images were aligned using TrackEM2 (Fiji package). Further, IMOD was used to mark outlines of the bacteria in the stack images. The 3D model obtained in this way was used to map the volume with the same location depicted in the fluorescent image.

Immunogold labeling and Electron microscopy

Cells were fixed in 4% formaldehyde and 0.1% glutaraldehyde in 1x PHEM buffer for 90 min (113). Cryo-sectioning and immunolabelling were performed as described

Results: Additional tools developed

elsewhere (114). In brief, ultrathin sections (50–70nm) from gelatin-embedded and frozen cell pellets were obtained using an FC7/UC7-ultramicrotome (Leica, Vienna, Austria). Immunogold labelling was carried out on thawed sections with anti-GFP (Rockland, 600-101-215) antibody and 10nm protein A-gold (UMC Utrecht University, Utrecht, Netherlands) (1:50). Sections were examined with a CM10 Philips transmission electron microscope with an Olympus ‘Veleta’ 2kx2k side-mounted TEM CCD camera.

Table 1: List of primers used in this study

Name	Sequence
prSL001	5' - CGACTCTAGAGGATCCGCCACCATGGATCAAGCTA GATCAG -3'
prSL002	5' - GATTGTCGACGAATTCTTACTTGTACAGCTCGTCCA TGC-3'
prSL003	5' - CGACTCTAGAGGATCCCGCCACCATGGCGGCCCC-3'
prSL004	5' - GATTGTCGACGAATTCTTACTTGTACAGCTCGTCCAT GCCG-3'
prSL021	5' - CGACTCTAGAGGATCCCCACCATGGACAGCAAAGG TTCG-3'
prSL022	5' - GATTGTCGACGAATTCTCAAAGAATTTTTTGGCAGCTGC TTCT-3'
prSL032	5' - CGACTCTAGAGGATCCGCCACCATGGCCTCCTCC-3'
prSL033	5' - GATTGTCGACGAATTCTTAACAAGTGCAGCTTTCTGC GGAGG-3'
prSL036	5' -GATTGTCGACGAATTCTTAGTTACTACAACACTGATTC CTGGTTGG-3'
prSL039	5' - CGACTCTAGAGGATCCGCCACCATGGCGGCCCCCGG CAGC-3'
prSL040	5' -GATTGTCGACGAATTCTTACTTGTACAGCTCGTCCATG CCGAGAGT-3'
prSL045	5' - CGACTCTAGAGGATCCGCCACCATGGTGAGCAAGGGCG AGGA-3'
prSL046	5' - GATTGTCGACGAATTCTCAACAGCAGCCGCCCCCAG-3'
prSL050	5' - GATTGTCGACGAATTCTCAGCAGCAGCCAGAGTTGG-3'
prSL053	5' - GATTGTCGACGAATTCTTAACAAGTGCAGCTTTCTGC GGAG-3'
prSL061	5' -CGACTCTAGAGGATCCGCCACCATGTCTGGGGGCAAATA CGTAGA-3'
prSL062	5' -GATTGTCGACGAATTCTTACTTGTACAGCTCGTCCATGC-3'

Results: Additional tools developed

prSL072 5'- CGACTCTAGAGGATCCGCCACCATGGACAACACCGAGG
ACGT-3'

prSL073 5'- GATTGTCGACGAATTCTCATTTCGGCTTGGCTTTGGAT-3'

prSL078 5'- CGACTCTAGAGGATCCGCCACCATGGCGGCCCCCGGGCG
CC-3'

prSL079 5'-GATTGTCGACGAATTCTTACTTGTACAGCTCGTCCATGCC
GAGAGT-3'

prSL086 5'-CGACTCTAGAGGATCCGCCACCATGGAAGGGAAGTGGTT
GCTGTGTATGTTA-3'

prLS087 5'- GATTGTCGACGAATTCTTACTTGTACAGCTCGTCCATGC
CGAG-3'

Figure legends

Figure 1. Summary of HeLa cell lines stably – expressing fluorescently labeled cellular compartments. Representative images of HeLa stable cell lines that have been generated to express fluorescently labeled cellular compartments of interest. Stable cell lines were generated via transduction of lentiviruses containing DNA of interest fused to fluorescent markers. Images were taken with the Andor confocal microscope with 60x objective, a slice around the middle of a stack is represented. Scale bar represents 10 μm .

Figure 2. BCV interacts with Rab7 at early and late time points of infection in HeLa cells. A) Validation of the localization of pmRFP1-Rab7 as compared to endogenous Lamp1 that also labels late endosomes. HeLa pmRFP1-Rab7 cells were fixed and stained with anti-Lamp1 antibody and secondary antibody conjugated with Alexa 488. Image was taken with Andor confocal microscope and figure represents a slice in the middle of a stack with scale bar 10 μm . B) HeLa cells stably expressing pmRFP1-Rab7 were infected with *Brucella abortus* (DAPI) and fixed at 6 hpi, 24 hpi, 48 hpi or 72 hpi. Fluorescent images were obtained with API DeltaVision Core Microscope and deconvolved with the DeltaVision software. Figure shows representative slice around the middle of a stack. Scale bar represents 10 μm . C) Electron micrograph studies of a Rab7-RFP positive BCV at 6 hpi. i) Fluorescent image was first taken with API DeltaVision Core Microscope and deconvolved with the DeltaVision software. ii) samples were processed for electron microscopy studies with Helios Nanolab 650 dual beam microscope microscope after which a volume of the section was obtained. IMOD was used to mark outlines of the bacteria in the stack images. The 3D model shown in the figure was used to map the volume with the same location depicted in the fluorescent image. Scale bar represents 800nm. Images are provided by Jaroslaw Sedzicki. iii) Electron micrographs representing slices from a stack, a figure is shown for every 30 slices of the entire volume. Bacteria are labeled in numbers 1-6 according to the numbering in the fluorescent image. Scale bar represents 800nm. Images are provided by Jaroslaw Sedzicki.

Figure 3. Validation of HeLa Igp120-mCherry stable cell line with wild type *Brucella abortus* or *Brucella abortus* ΔvirB9 mutant. HeLa cells stably expressing

Igp120-mCherry were infected with GFP-expressing *Brucella abortus* or GFP-expressing *Brucella abortus* Δ virB9 and fixed at 6 hpi or 24 hpi. Confocal images were obtained with Andor confocal microscope. A representative slice from a stack is presented. Scale bar represents 10 μ m.

Figure 4. *Brucella abortus* is contained within calnexin positive vesicles in HeLa and THP-1 macrophage-like cell line at late time points of infection. A) HeLa or B) THP-1 macrophage-like cell line stably expressing calnexin-GFP were infected with DsRed-expressing *Brucella abortus* and fixed around 44 hpi. Fluorescent images were obtained with API DeltaVision Core Microscope and deconvolved with the DeltaVision software. A representative slice from around the middle of a stack is presented. Scale bar represents 10 μ m. C) Electron micrograph of *Brucella abortus* infected THP-1 stably expressing calnexin-GFP at around 44 hpi. Immunogold labeling was performed with antibody against GFP. Images are provided by Dr. Christopher Bleck.

Figure 5. *Brucella abortus* is surrounded by RTN2 positive membrane at late time points of infection. A) HeLa cells stably expressing RFP-RTN2 were infected with GFP-expressing *Brucella abortus* and fixed at 6 h or 24 hpi. B) RFP-RTN2 was stained with ER specific antibodies, anti-BAP31 or anti-CLIMP63. Confocal images were obtained with Andor confocal microscope. A representative slice from the middle of a stack is presented. Scale bar represents 10 μ m.

Figure 6. *Brucella abortus* is surrounded by BAP31 positive membrane at late time points of infection. HeLa cells were infected with GFP-expressing *Brucella abortus* and fixed at 24 hpi, 48 hpi or 72 hpi. Cells were stained with anti-BAP31 antibody and secondary antibody Alexa Fluor 546 Goat Anti-mouse IgG. Confocal images were obtained with Andor confocal microscope. A representative slice the middle of a stack is presented. Scale bar represents 10 μ m.

Figure 7. *Brucella abortus* is surrounded by CLIMP63 positive membrane at late time points of infection. HeLa cells were infected with GFP-expressing *Brucella abortus* and fixed at 24 hpi, 48 hpi or 72 hpi. Cells were stained with anti-CLIMP63

Results: Additional tools developed

antibody and secondary antibody Alexa Fluor 546 Goat Anti-mouse IgG. Confocal images were obtained with Andor confocal microscope. A representative slice from the middle of a stack is presented. Scale bar represents 10 μm .

Results: Additional tools developed

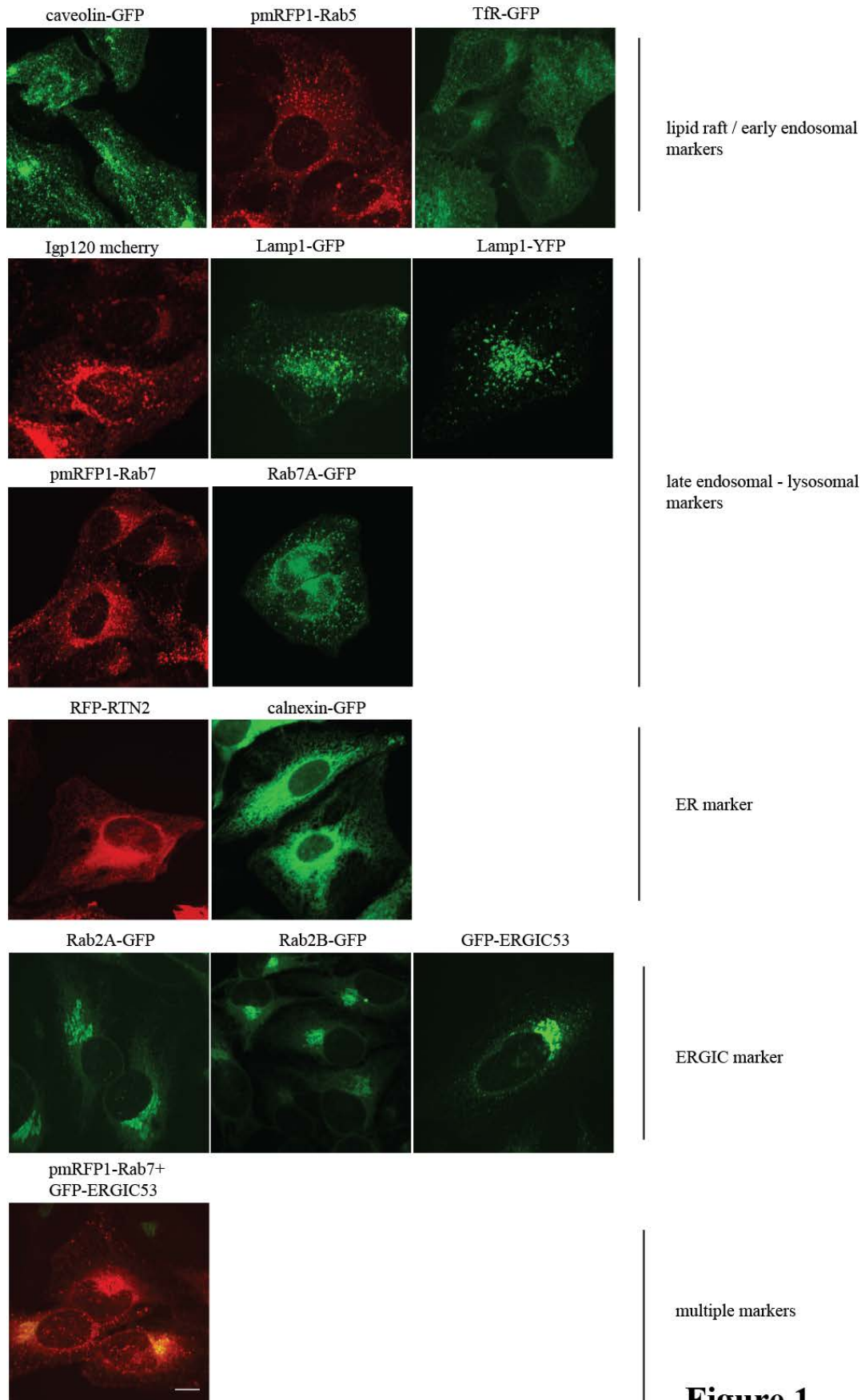
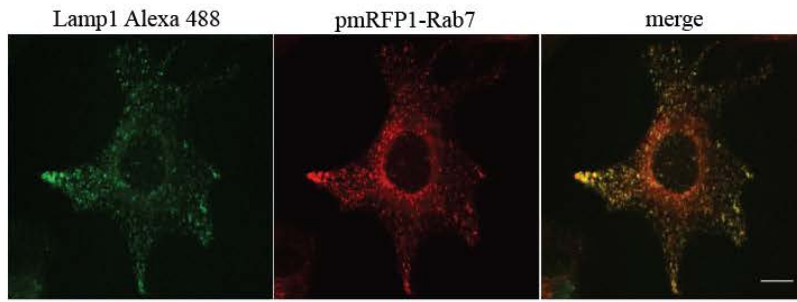


Figure 1

A



B

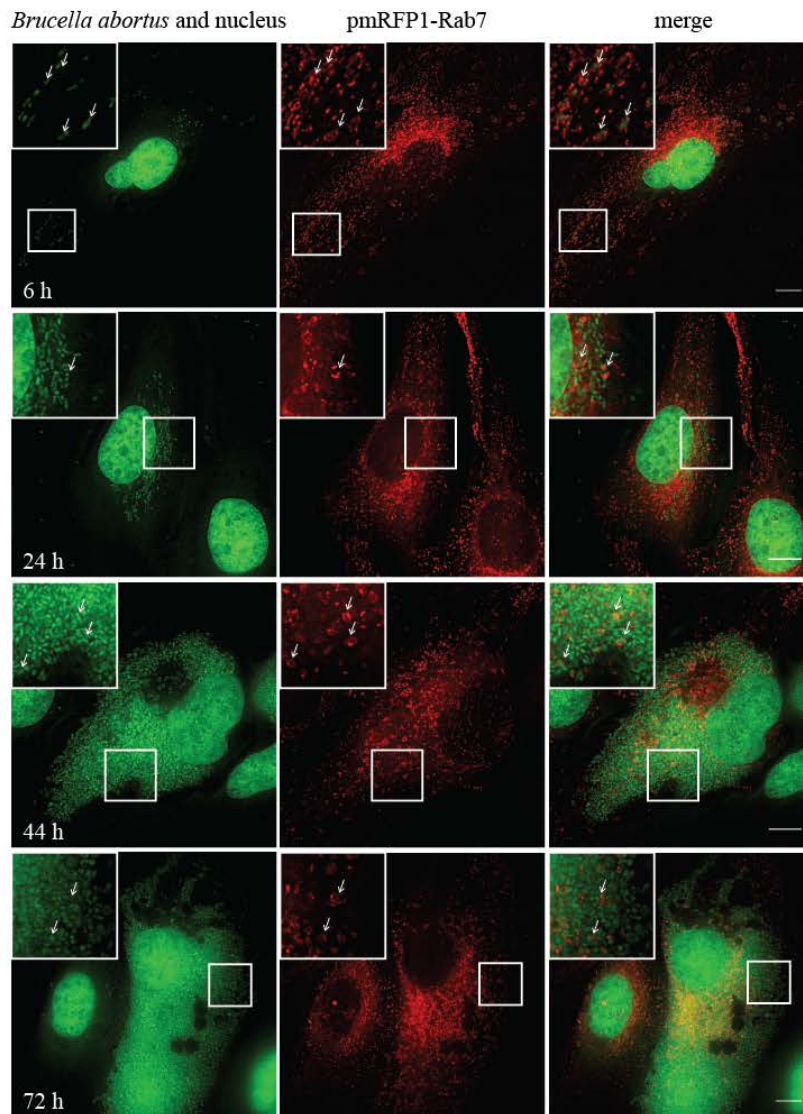
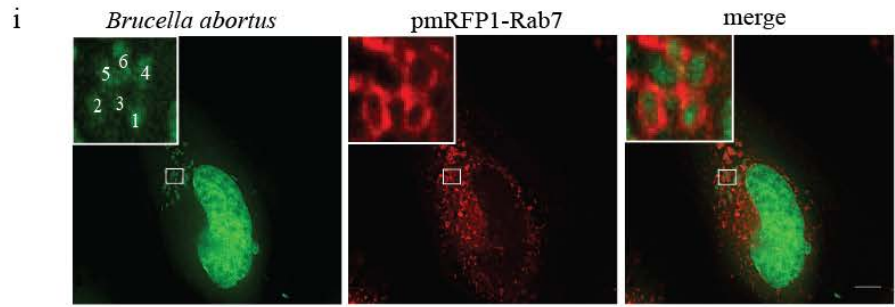
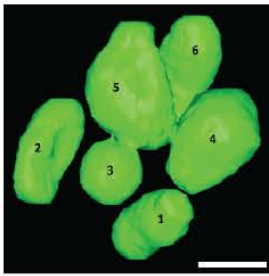


Figure 2i

C



ii



iii

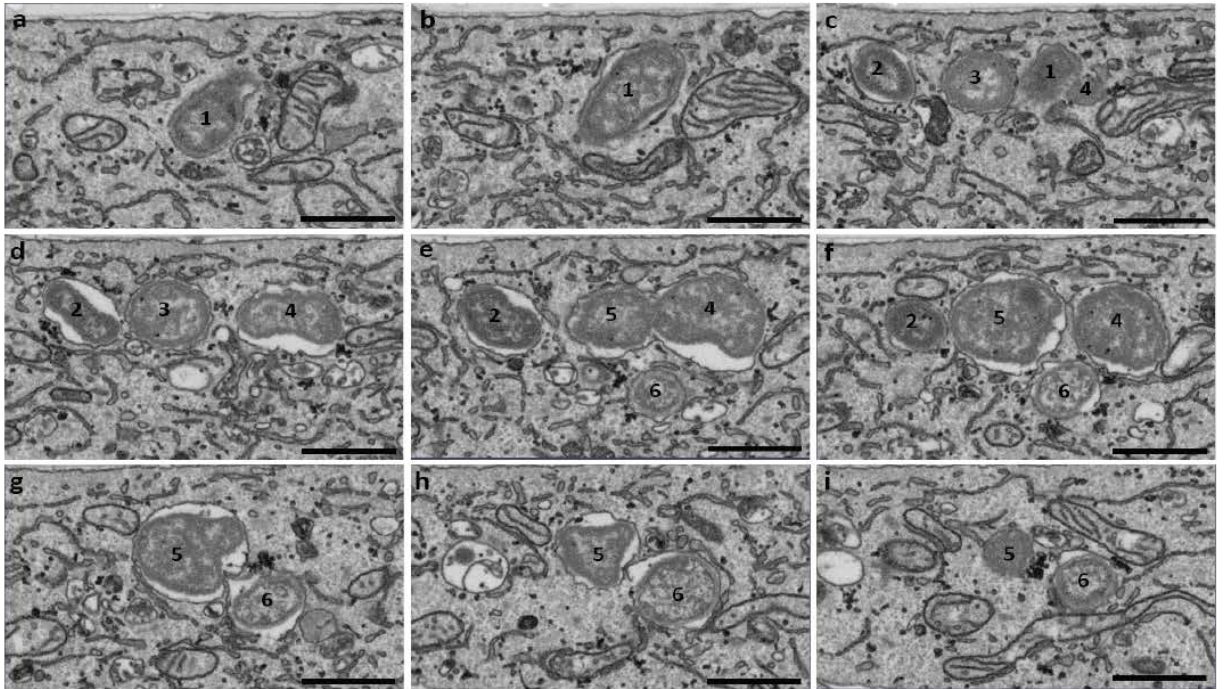


Figure 2ii

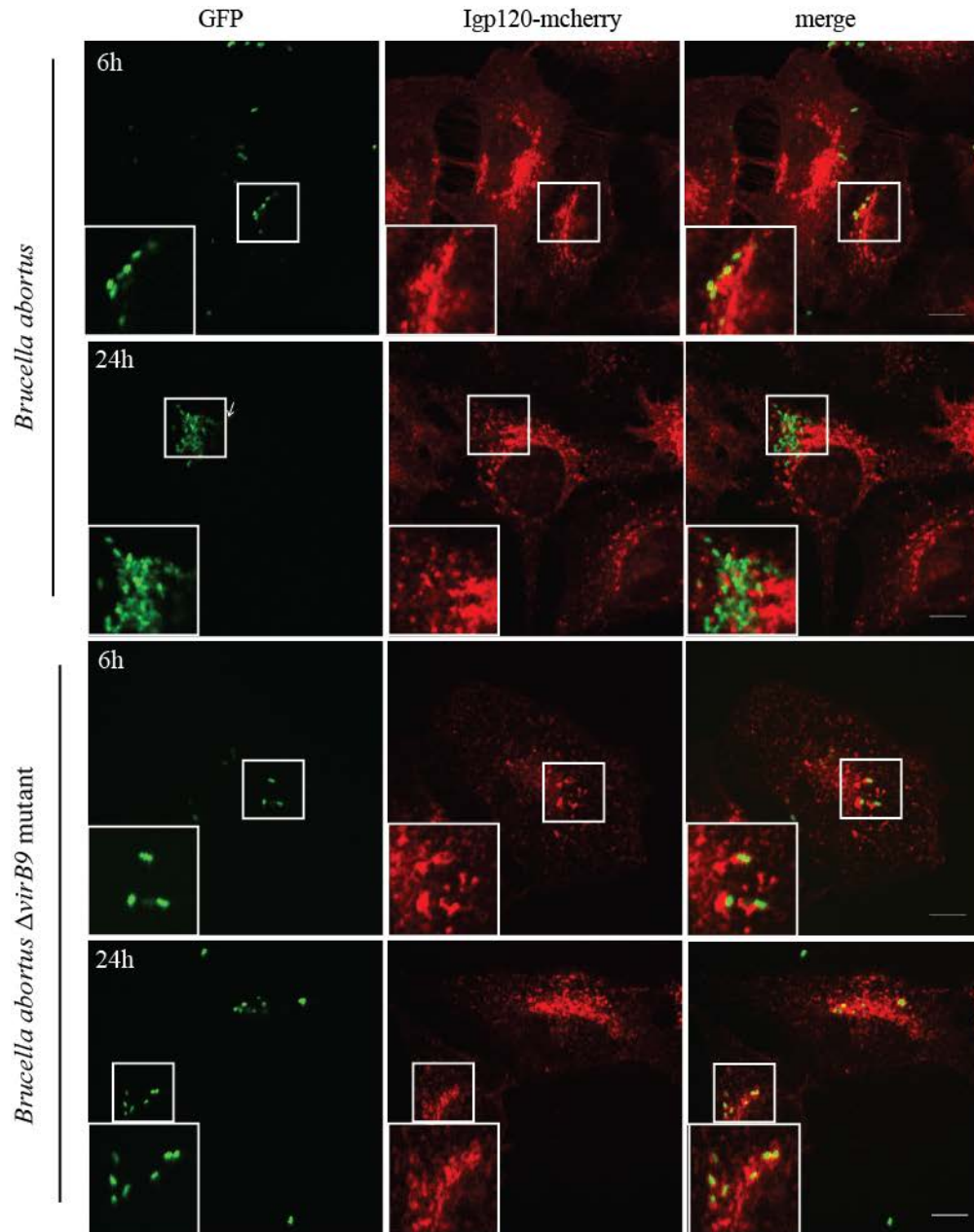


Figure 3

Results: Additional tools developed

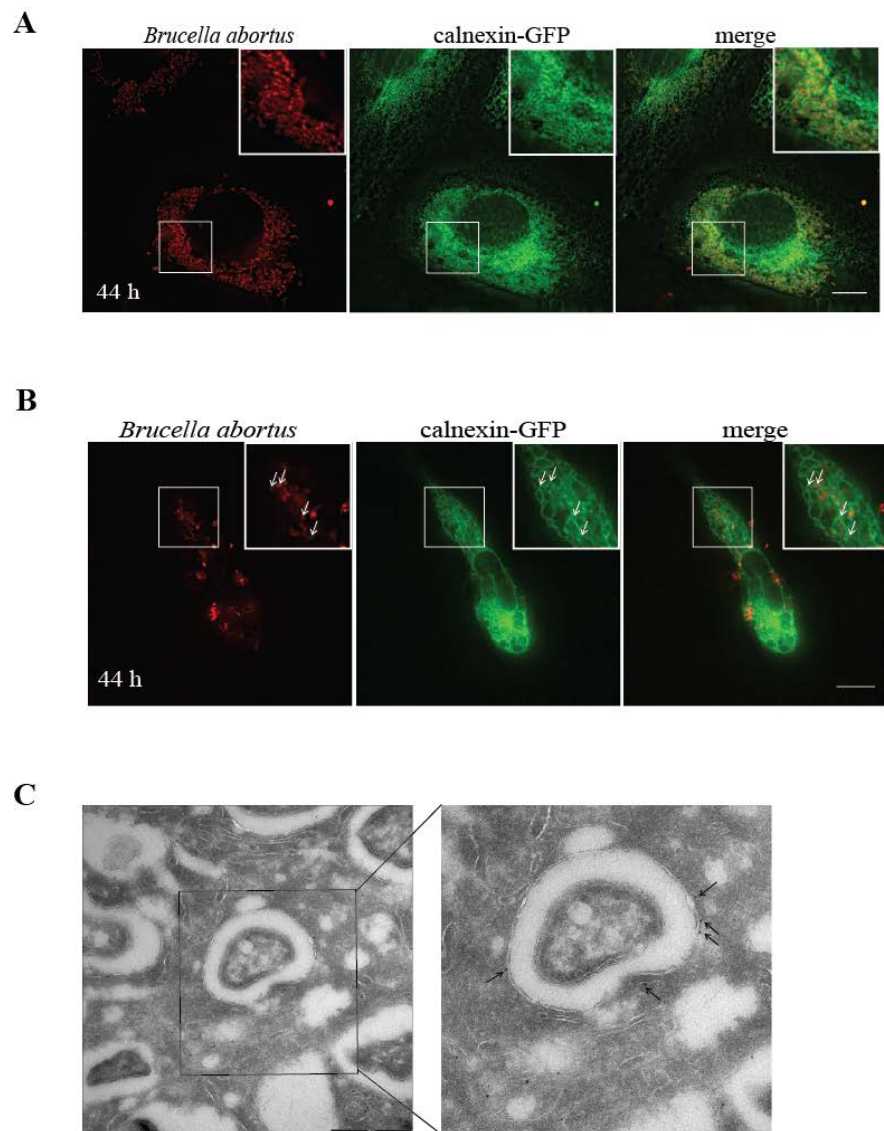


Figure 4

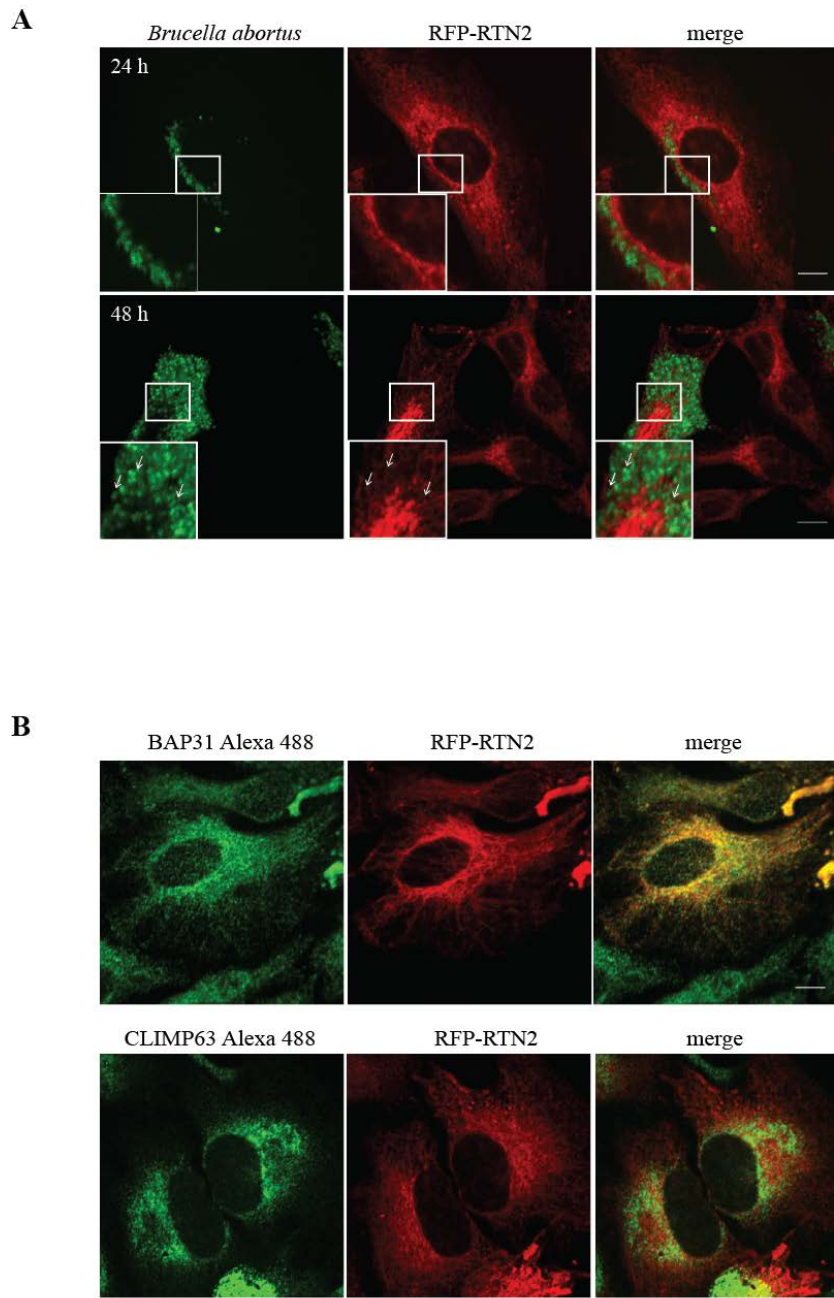


Figure 5

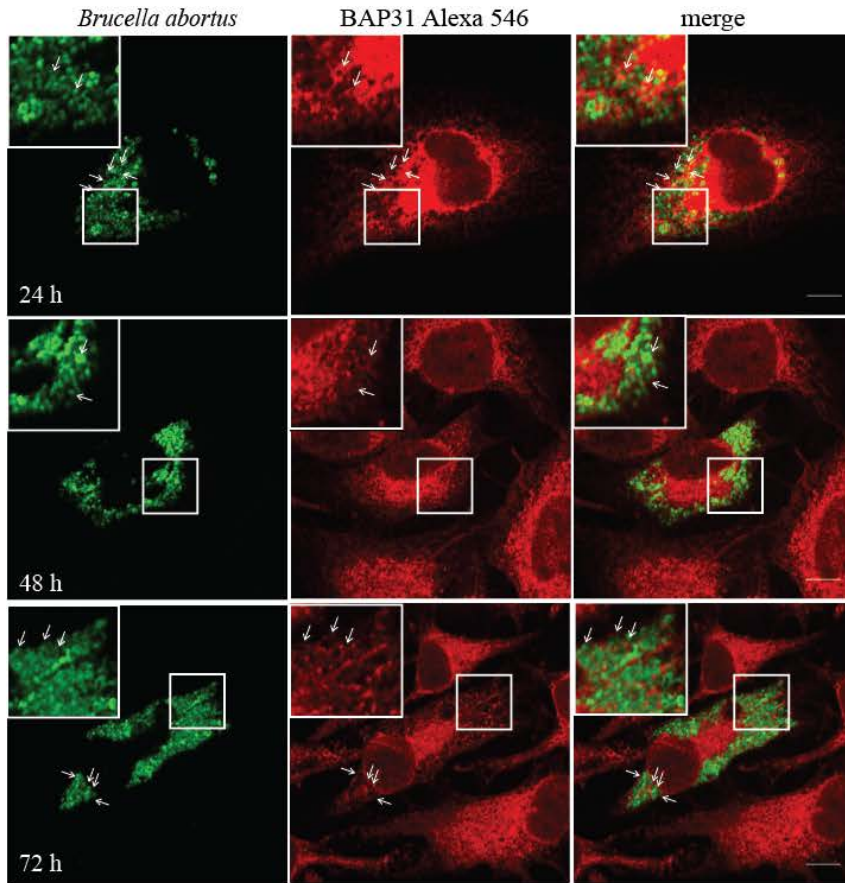


Figure 6

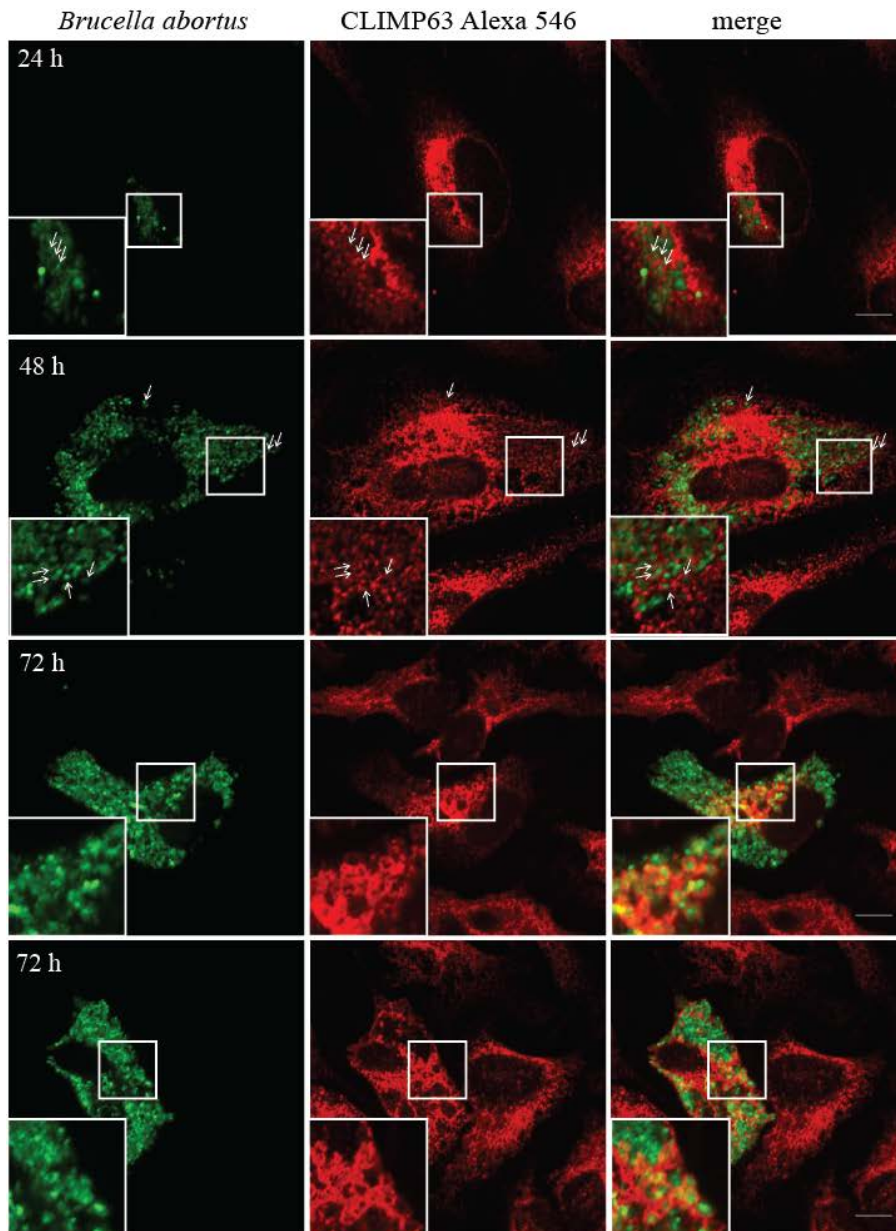


Figure 7

GENERAL CONCLUSIONS AND OUTLOOK

4. GENERAL CONCLUSIONS AND OUTLOOK

4.1 Genome-wide siRNA screen reveals novel host signaling pathways involved in *Brucella* infection

To identify host factors involved in *Brucella* infection, we established a high-throughput, high-content microscopy based RNA interference (RNAi) assay for *Brucella* infection in HeLa cells. Pilot screens were performed with kinome-based libraries, with independent replicates in the kinome screens having reproducible results confirming the robustness of our assay. Also, with multiple individual siRNAs (up to 11) targeting each gene available in the kinome screen, we were able to show that screening with more independent siRNAs increased the quality of our data in terms of correlation coefficient between different siRNA libraries. A statistical model, Parallel Mixed Model (PMM), was also developed using the kinome-based screens which allows analysis of multiple large scale RNAi screens that are performed under similar experimental conditions (Research Article II) (115). This model allows comparison between kinome-based screens that were performed in parallel between different groups within the InfectX consortium using the same siRNA libraries and standardized experimental conditions but different pathogens, discovering unique as well as shared hits between these pathogens.

With the established experimental workflow, genome-wide siRNA screens were performed with Dharmacon pooled and Qiagen unpooled siRNA libraries. Using datasets from the Qiagen unpooled library as well as kinome-based screens with Ambion and Dharmacon unpooled libraries, we identified off-target driven effects in siRNA screens due to the seed region of the siRNA oligo (nucleotides 2-8). This region could bind with partial complementarity to multiple mRNAs, in a way that is similar to endogenous microRNAs (116). Furthermore, we were able to design novel RNAs that block or increase infection of one or multiple pathogens, by having seed sequences that are known to affect pathogen infection but with no specific gene target. This allows potential drug design that does not cause toxicity to the host but has an effect on bacterial pathogenicity (116) (Research Article I).

Due to the off-target effect prone siRNA screens, we performed statistical analysis of our primary screening dataset using the Redundant siRNA analysis (RSA) method

General conclusions and Outlook

(117). This allowed us to rank genes according to the consistency of effects from multiple individual siRNAs. Higher P values were given to genes that have a consistent effect on infection upon knockdown with multiple individual siRNA, suggesting an on-target effect. With the top ranking up and down hits from the primary screen, we were able to identify enrichment of expected pathways known to be involved in *Brucella* infection, suggesting that our workflow allows the finding of relevant pathways in *Brucella* infection. Next, for selected genes of interest, we performed secondary screens with more individual siRNAs to gain confidence on our datasets as recommended from the kinome-based screen results. Finally, we discovered novel host signaling pathways involved in *Brucella* infection (Research Article III).

It would now be interesting to study the molecular details of these pathways in the context of *Brucella* infection. Genes of interest could be validated with siRNA-independent approaches e.g. inhibitors or knockout experiments with the CRISPR-Cas9 system (118, 119). Rescue experiments by expressing complementary cDNA upon siRNA knockdown could also be done to validate the specificity of siRNA knockdown. To understand the *Brucella* trafficking steps that are affected upon knockdown of certain host factors, study of interaction with Lamp1 endo-lysosomal compartmental marker would allow us to identify genes that are involved in deviating *Brucella* from the endocytic pathway and for subsequent interaction and replication within its replicative niche. In addition to the siRNA screen that was performed together with *Brucella* infection, we also performed mock screens in which cells were only treated with siRNA without addition of bacteria. Cells from the mock screen were stained with TGN46 and ERGIC53 antibodies, to understand the effect of the siRNA treatment towards trafficking pathways from the plasma membrane to the Golgi as well as from the Golgi to the ER. By combining the information from the mock screen, we would be able to understand the role of the endocytic or exocytic pathways in the context of *Brucella* infection. Finally, certain groups in the consortium are currently developing methods to correct for off-target effects within the siRNA screens. With that, we would like to compare our hit list before and after off-target correction, possibly identifying new pathways involved in *Brucella* infection.

4.2 Active TGF- β signaling increases *Brucella* entry in HeLa cells

One of the signaling pathways that were chosen for follow up studies was the TGF- β signaling pathway since most of the components of this signaling pathway including ligands, surface receptors, and downstream signaling components showed reduction in *Brucella* infection upon depletion. Surface receptors TGFBR1 and TGFBR2 were strong hits that reduced infection of almost every pathogen in the InfectX consortium, with a stronger effect on *Adenovirus*, *Brucella abortus*, *Shigella flexneri* and *Vacciniavirus* and a milder effect on *Bartonella henselae*, *Listeria monocytogenes*, *Rhinovirus* and *Salmonella typhimurium* infection (Research Article II Figure 7). This suggests that there is probably a more general role of TGF- β signaling that is exploited by different pathogens for their interaction with the host.

It has been reported that in mice infected with *Brucella abortus*, TGF- β were produced by B cells during early stages of the infection (28). Also, in patients with chronic brucellosis, there was a correlation between increased TGF- β production and depressed function of T cell responses (27). This suggests a role of TGF- β during *Brucella* infection in terms of immunosuppression of the host. For *Trypanosoma cruzi*, it has been reported that active TGF- β signaling is important for entry in mammalian cells (29, 30). Since we observe a decrease in *Brucella* infection in HeLa cells that is an epithelial cell line, it is likely that TGF- β signaling has also a non-immunological role during *Brucella* infection. Our experiments confirmed that activation of TGF- β signaling pathway as well as overexpression of wild type TGFBRs increased *Brucella* infection in HeLa cells. Not only the expression of the receptors but also their kinase activity is important for *Brucella* infection. More specifically, we showed that TGF- β signaling pathway is probably mainly involved in the entry step during *Brucella* infection in HeLa cells (Results section 3.4).

Next, we would like to identify the pathways that are regulated by TGF- β signaling for the benefit of infection by *Brucella* and possibly by other pathogens. It has been reported that TGF- β signaling is involved in the regulation of endocytosis (9) and actin remodeling via activation of Rho GTPases (120, 121). Possibly, *Brucella* and other pathogens utilize TGF- β signaling to modulate endocytosis and actin polymerization for their entry into the host cell. To study this, endocytosis could be

General conclusions and Outlook

monitored with fluorescent latex beads or specific cargoes using the receptor-mediated endocytic pathway in combination with TGFBR depletion to monitor the effect on endocytosis. Receptor-mediated endocytosis and lipid rafts are reported to be important for *Brucella* entry in non-phagocytic (122, 123) and phagocytic cells, respectively (124-126). It would be interesting to study the role of TGF- β signaling towards these entry routes. Changes in actin remodeling upon TGFBR depletion and its effect on *Brucella* entry could also be monitored at early time points of infection with cell lines stably expressing mcherry-Lifeact that allows time lapsed imaging of the actin cytoskeleton.

TGF- β 1 exists as a latent form and upon activation by pathogens (37-39) or proteolytic digestion by host factors binds to TGF- β receptors on the cell surface, triggering associated downstream signaling and transcription of regulated genes. The ability of *Brucella* in activating TGF- β has not been studied so far. Therefore, it would also be interesting to investigate whether *Brucella* is also able to activate latent TGF- β directly as seen with other pathogens that exploit the TGF- β signaling pathway (38, 39). This could be done by incubating *Brucella* with latent TGF- β in a cell free system, detecting active TGF- β with ELISA (39) or comparing by western blot the levels of active versus latent TGF- β using specific antibodies (38).

Furthermore, it would be interesting to investigate the role of TGF- β signaling in various cell types. Possibly, the extent of TGF- β signaling activity of a cell line determines the infectivity levels of *Brucella*. With that, different non-phagocytic cell lines could be tested for TGF- β signaling activity and *Brucella* infection levels in correlation to TGF- β signaling. Also, it would be interesting to confirm the role of TGF- β signaling in more relevant cell lines for *Brucella* infection e.g. the macrophages. We initiated our studies with THP-1 human monocytic cell line to take advantage of genome-wide siRNA sequences that are targeting the human genome and the results from the genome-wide siRNA screens. However, since THP-1 is a monocytic cell line that has to be differentiated to become macrophage-like, the properties may not be the same as they would be expected for primary cells or bona fide macrophage cell lines. Therefore, it would be useful to analyze the role of TGF- β signaling in relevant cell lines, e.g. bone marrow derived macrophages (BMDM) or mouse macrophage cell line RAW264.7.

In summary, we identified TGF- β signaling pathway to be involved in the entry step of *Brucella* infection of HeLa cells, with active TGF- β signaling and kinase active receptors being involved in *Brucella* infection in this cell line.

4.3 Vps35, a retromer complex component is required for *Brucella* trafficking to its replicative niche

Importantly, with our genome-wide siRNA screens, we were also able to discover a novel host factor, Vps35 that is a component of the retromer complex involved in endosome to Golgi transport. The specificity of Vps35 knockdown was confirmed with complementary cDNA being able to rescue the knockdown phenotype. Furthermore, treatment of HeLa cells with Retro-2 inhibitor that specifically inhibits transport from endosome to Golgi also showed a significant decrease in *Brucella* infection, suggesting a general role of this transport pathway in infection. Preliminary results showed that a knock down of Vps35 caused *Brucella* to stay in endo-lysosomal compartments at time points where it is expected to reach its replicative niche (Results section 3.5). Nevertheless, we also identified a Rab GTPase-activating protein (GAP), USP6NL, in our genome-wide siRNA screen to increase *Brucella* infection upon depletion. USP6NL has been reported to be involved in Shiga toxin transport from endosomes to Golgi via regulation of Rab43 (87), suggesting that possibly a similar path is exploited by *Brucella* during infection. It would now be important to understand the molecular mechanisms underlying this process and components of this transport pathway that are needed during *Brucella* infection.

Retro-2 inhibitor has been shown to be very specific in inhibiting toxin transport from endosomes to the Golgi without affecting the endocytic or recycling pathway, secretory pathway or trafficking of major cargoes. However, it was found that Retro-2 inhibitor alters the localization of SNARE proteins, syntaxin 5 and syntaxin 6 (89). Syntaxin 5 is normally localized at the Golgi and receives traffic from the ER, playing a role between the Golgi and ER. Treatment of cells with Retro-2 caused relocalization of syntaxin 5 and, to a lesser extent, syntaxin 6 from the Golgi to the cytoplasm (89). Syntaxin 5 and syntaxin 16 are also known to be important for retrograde transport of Shiga toxin, CIMPR, cholera toxin and ricin (127). Therefore,

General conclusions and Outlook

it would be interesting to understand the role of these SNARE proteins in the context of *Brucella* infection, specifically in trafficking to the replicative niche.

Despite recent advances, it remains unknown how *Brucella* traffics from the endocytic pathway to an ER-derived replicative niche. There are studies that show controversial roles of the Golgi or the retrograde pathway from the Golgi to the ER during *Brucella* infection (46, 50, 128). Experiments that were performed to study this pathway were done with Brefeldin A (BFA) treatment that inhibits protein transport from ER to Golgi, ultimately causes a collapse of the Golgi. In these studies, Fugier *et al* reported an effect on *Brucella* infection upon prolonged BFA treatment while Celli *et al* showed no effect on *Brucella* replication with BFA treatment not longer than 3 h at 30min before infection as well as 30min, 2 h, 5 h or 8 h after infection (46, 128). It would be important to confirm the role of the Golgi during *Brucella* infection by repeating this experiment with BFA treatment in our system, and to include time points later than 8 h in our studies.

Furthermore, the role of endosomes to Golgi traffic has not been explored. It would be important to study in detail the role of the Golgi prior to *Brucella* arrival at its replicative niche. Since interaction with the Golgi could be transient, experiments with live cell imaging using stable cell lines expressing Golgi markers could be utilized to identify specific time points where this interaction could be observed. With that, correlative light and electron microscopy studies (Results Section 3.6) could be performed to confirm this interaction in an ultrastructural level.

Another possibility to be considered is that *Brucella* does not traffic by itself to the Golgi, but only requires a functional retrograde trafficking pathway due to its need for distinct components from this system. *Legionella* effector RidL for example has been shown to inhibit retrograde trafficking by binding to Vps29 subunit of the retromer complex for the benefit of its infection. The molecular mechanism of this inhibition is still unknown (85). It also remains unclear whether the retrograde trafficking pathway or its associated components are only required to be present or are actively manipulated during the course of infection. Therefore, it would be interesting to test the effect of *Brucella* infection on the retrograde trafficking pathway using different cargoes e.g. mannose 6-phosphate receptor (CIMPR) or toxins (Shiga toxin B-subunit, cholera toxin B-subunit) that were known to utilize this pathway to arrive at the Golgi

General conclusions and Outlook

(84, 129). Experiments could be repeated with *Brucella abortus* $\Delta virB9$ mutant to understand the role of the VirB type IV secretion system in this infection context.

In summary, we discovered a novel host factor, Vps35, which is a component of the retromer complex to be important for *Brucella* trafficking to its replicative niche, with *Brucella* ending up in an endo-lysosomal compartment upon Vps35 depletion. Furthermore, interfering retrograde trafficking from endosome to Golgi also caused a significant decrease in *Brucella* infection suggesting a general role of this pathway in the course of infection.

All in all, genome wide-siRNA screen with an endpoint infection assay allows identification of relevant and novel signalling pathways covering the entire intracellular life cycle of *Brucella* infection. It is now crucial to understand in a deeper level how these pathways contribute or inhibit *Brucella* infection. With available tools of HeLa cells stably expressing fluorescently labeled cellular compartmental markers of interest, we will be able to gain a deeper understanding with fluorescence microscopy, live cell imaging and CLEM studies the intracellular trafficking pathway of *Brucella* and its interaction with the host.

REFERENCES

5. REFERENCES

1. Jackson AL & Linsley PS (2010) Recognizing and avoiding siRNA off-target effects for target identification and therapeutic application. (Translated from eng) *Nature reviews. Drug discovery* 9(1):57-67 (in eng).
2. Birmingham A, *et al.* (2006) 3' UTR seed matches, but not overall identity, are associated with RNAi off-targets. (Translated from eng) *Nature methods* 3(3):199-204 (in eng).
3. Massague J (2012) TGFbeta signalling in context. (Translated from eng) *Nat Rev Mol Cell Biol* 13(10):616-630 (in eng).
4. Moustakas A & Heldin CH (2008) Dynamic control of TGF-beta signaling and its links to the cytoskeleton. (Translated from eng) *FEBS Lett* 582(14):2051-2065 (in eng).
5. Munger JS & Sheppard D (2011) Cross talk among TGF-beta signaling pathways, integrins, and the extracellular matrix. (Translated from eng) *Cold Spring Harb Perspect Biol* 3(11):a005017 (in eng).
6. Roberts AB, McCune BK, & Sporn MB (1992) TGF-beta: regulation of extracellular matrix. (Translated from eng) *Kidney Int* 41(3):557-559 (in eng).
7. Ignatz RA, Heino J, & Massague J (1989) Regulation of cell adhesion receptors by transforming growth factor-beta. Regulation of vitronectin receptor and LFA-1. (Translated from eng) *J Biol Chem* 264(1):389-392 (in eng).
8. Ignatz RA & Massague J (1987) Cell adhesion protein receptors as targets for transforming growth factor-beta action. (Translated from eng) *Cell* 51(2):189-197 (in eng).
9. Collinet C, *et al.* (2010) Systems survey of endocytosis by multiparametric image analysis. (Translated from eng) *Nature* 464(7286):243-249 (in eng).
10. Huang F & Chen YG (2012) Regulation of TGF-beta receptor activity. (Translated from eng) *Cell Biosci* 2:9 (in eng).
11. Massague J & Chen YG (2000) Controlling TGF-beta signaling. (Translated from eng) *Genes Dev* 14(6):627-644 (in eng).
12. Wrighton KH, Lin X, & Feng XH (2009) Phospho-control of TGF-beta superfamily signaling. (Translated from eng) *Cell Res* 19(1):8-20 (in eng).

References

13. Chen X & Xu L (2011) Mechanism and regulation of nucleocytoplasmic trafficking of smad. (Translated from eng) *Cell Biosci* 1(1):40 (in eng).
14. L.Friedman AYHaS (2003) The transforming growth factor b (TGF-b) signaling pathway. *Expert Reviews in Molecular Medicine* 5.
15. Di Guglielmo GM, Le Roy C, Goodfellow AF, & Wrana JL (2003) Distinct endocytic pathways regulate TGF-beta receptor signalling and turnover. (Translated from eng) *Nat Cell Biol* 5(5):410-421 (in eng).
16. Razani B, *et al.* (2001) Caveolin-1 regulates transforming growth factor (TGF)-beta/SMAD signaling through an interaction with the TGF-beta type I receptor. (Translated from eng) *J Biol Chem* 276(9):6727-6738 (in eng).
17. Ebisawa T, *et al.* (2001) Smurf1 interacts with transforming growth factor-beta type I receptor through Smad7 and induces receptor degradation. (Translated from eng) *J Biol Chem* 276(16):12477-12480 (in eng).
18. Hayashi H, *et al.* (1997) The MAD-related protein Smad7 associates with the TGFbeta receptor and functions as an antagonist of TGFbeta signaling. (Translated from eng) *Cell* 89(7):1165-1173 (in eng).
19. Kavsak P, *et al.* (2000) Smad7 binds to Smurf2 to form an E3 ubiquitin ligase that targets the TGF beta receptor for degradation. (Translated from eng) *Mol Cell* 6(6):1365-1375 (in eng).
20. Gao S, *et al.* (2009) Ubiquitin ligase Nedd4L targets activated Smad2/3 to limit TGF-beta signaling. (Translated from eng) *Mol Cell* 36(3):457-468 (in eng).
21. Su Y, *et al.* (2007) The evolutionally conserved activity of Dapper2 in antagonizing TGF-beta signaling. (Translated from eng) *FASEB J* 21(3):682-690 (in eng).
22. Hao X, *et al.* (2011) SNX25 regulates TGF-beta signaling by enhancing the receptor degradation. (Translated from eng) *Cell Signal* 23(5):935-946 (in eng).
23. Shi W, *et al.* (2004) GADD34-PP1c recruited by Smad7 dephosphorylates TGFbeta type I receptor. (Translated from eng) *J Cell Biol* 164(2):291-300 (in eng).
24. Batut J, *et al.* (2008) Two highly related regulatory subunits of PP2A exert opposite effects on TGF-beta/Activin/Nodal signalling. (Translated from eng) *Development* 135(17):2927-2937 (in eng).

References

25. Bogdan C, Paik J, Vodovotz Y, & Nathan C (1992) Contrasting mechanisms for suppression of macrophage cytokine release by transforming growth factor-beta and interleukin-10. (Translated from eng) *J Biol Chem* 267(32):23301-23308 (in eng).
26. Gazzinelli RT, Oswald IP, Hieny S, James SL, & Sher A (1992) The microbicidal activity of interferon-gamma-treated macrophages against *Trypanosoma cruzi* involves an L-arginine-dependent, nitrogen oxide-mediated mechanism inhibitable by interleukin-10 and transforming growth factor-beta. (Translated from eng) *Eur J Immunol* 22(10):2501-2506 (in eng).
27. Elfaki MG & Al-Hokail AA (2009) Transforming growth factor beta production correlates with depressed lymphocytes function in humans with chronic brucellosis. (Translated from eng) *Microbes Infect* 11(14-15):1089-1096 (in eng).
28. Goenka R, Parent MA, Elzer PH, & Baldwin CL (2011) B cell-deficient mice display markedly enhanced resistance to the intracellular bacterium *Brucella abortus*. (Translated from eng) *J Infect Dis* 203(8):1136-1146 (in eng).
29. Ming M, Ewen ME, & Pereira ME (1995) Trypanosome invasion of mammalian cells requires activation of the TGF beta signaling pathway. (Translated from eng) *Cell* 82(2):287-296 (in eng).
30. Hall BS & Pereira MA (2000) Dual role for transforming growth factor beta-dependent signaling in *Trypanosoma cruzi* infection of mammalian cells. (Translated from eng) *Infect Immun* 68(4):2077-2081 (in eng).
31. Toossi Z, Gogate P, Shiratsuchi H, Young T, & Ellner JJ (1995) Enhanced production of TGF-beta by blood monocytes from patients with active tuberculosis and presence of TGF-beta in tuberculous granulomatous lung lesions. (Translated from eng) *J Immunol* 154(1):465-473 (in eng).
32. Rodel J, Straube E, Lungershausen W, Hartmann M, & Groh A (1998) Secretion of cytokines by human synoviocytes during in vitro infection with *Chlamydia trachomatis*. (Translated from eng) *J Rheumatol* 25(11):2161-2168 (in eng).
33. Michelson S, *et al.* (1994) Human cytomegalovirus infection induces transcription and secretion of transforming growth factor beta 1. (Translated from eng) *J Virol* 68(9):5730-5737 (in eng).

References

34. Yoo YD, *et al.* (1996) Regulation of transforming growth factor-beta 1 expression by the hepatitis B virus (HBV) X transactivator. Role in HBV pathogenesis. (Translated from eng) *J Clin Invest* 97(2):388-395 (in eng).
35. Araujo-Jorge TC, *et al.* (2002) Implication of transforming growth factor-beta1 in Chagas disease myocardiopathy. (Translated from eng) *J Infect Dis* 186(12):1823-1828 (in eng).
36. Barral-Netto M, *et al.* (1992) Transforming growth factor-beta in leishmanial infection: a parasite escape mechanism. (Translated from eng) *Science* 257(5069):545-548 (in eng).
37. Schultz-Cherry S & Hinshaw VS (1996) Influenza virus neuraminidase activates latent transforming growth factor beta. (Translated from eng) *J Virol* 70(12):8624-8629 (in eng).
38. Somanna A, Mundodi V, & Gedamu L (2002) Functional analysis of cathepsin B-like cysteine proteases from *Leishmania donovani* complex. Evidence for the activation of latent transforming growth factor beta. (Translated from eng) *J Biol Chem* 277(28):25305-25312 (in eng).
39. Waghbi MC, Keramidas M, Feige JJ, Araujo-Jorge TC, & Bailly S (2005) Activation of transforming growth factor beta by *Trypanosoma cruzi*. (Translated from eng) *Cell Microbiol* 7(4):511-517 (in eng).
40. Guzman-Verri C, *et al.* (2001) GTPases of the Rho subfamily are required for *Brucella abortus* internalization in nonprofessional phagocytes: direct activation of Cdc42. (Translated from eng) *J Biol Chem* 276(48):44435-44443 (in eng).
41. Satterwhite DJ, Aakre ME, Gorska AE, & Moses HL (1994) Inhibition of cell growth by TGF beta 1 is associated with inhibition of B-myb and cyclin A in both BALB/MK and Mv1Lu cells. (Translated from eng) *Cell Growth Differ* 5(8):789-799 (in eng).
42. Wrana JL, *et al.* (1992) TGF beta signals through a heteromeric protein kinase receptor complex. (Translated from eng) *Cell* 71(6):1003-1014 (in eng).
43. Wieser R, Wrana JL, & Massague J (1995) GS domain mutations that constitutively activate T beta R-I, the downstream signaling component in the TGF-beta receptor complex. (Translated from eng) *EMBO J* 14(10):2199-2208 (in eng).

References

44. Siegel PM, Shu W, Cardiff RD, Muller WJ, & Massague J (2003) Transforming growth factor beta signaling impairs Neu-induced mammary tumorigenesis while promoting pulmonary metastasis. (Translated from eng) *Proc Natl Acad Sci U S A* 100(14):8430-8435 (in eng).
45. Starr T, Ng TW, Wehrly TD, Knodler LA, & Celli J (2008) Brucella intracellular replication requires trafficking through the late endosomal/lysosomal compartment. (Translated from eng) *Traffic* 9(5):678-694 (in eng).
46. Celli J, *et al.* (2003) Brucella evades macrophage killing via VirB-dependent sustained interactions with the endoplasmic reticulum. (Translated from eng) *J Exp Med* 198(4):545-556 (in eng).
47. Colasante A, Aiello FB, Brunetti M, & di Giovine FS (2003) Gene expression of transforming growth factor beta receptors I and II in non-small-cell lung tumors. (Translated from eng) *Cytokine* 24(5):182-189 (in eng).
48. Koli KM & Arteaga CL (1997) Predominant cytosolic localization of type II transforming growth factor beta receptors in human breast carcinoma cells. (Translated from eng) *Cancer Res* 57(5):970-977 (in eng).
49. Abe M, *et al.* (1994) An assay for transforming growth factor-beta using cells transfected with a plasminogen activator inhibitor-1 promoter-luciferase construct. (Translated from eng) *Anal Biochem* 216(2):276-284 (in eng).
50. Celli J, Salcedo SP, & Gorvel JP (2005) Brucella coopts the small GTPase Sar1 for intracellular replication. (Translated from eng) *Proc Natl Acad Sci U S A* 102(5):1673-1678 (in eng).
51. Ittig S, *et al.* (2012) The lipopolysaccharide from *Campylobacter jejuni* reveals an unexpected role of the core-oligosaccharide in MD-2 binding. (Translated from eng) *PLoS Pathog* 8(5):e1002667 (in eng).
52. Seaman MN, Marcusson EG, Cereghino JL, & Emr SD (1997) Endosome to Golgi retrieval of the vacuolar protein sorting receptor, Vps10p, requires the function of the VPS29, VPS30, and VPS35 gene products. (Translated from eng) *J Cell Biol* 137(1):79-92 (in eng).
53. Seaman MN, McCaffery JM, & Emr SD (1998) A membrane coat complex essential for endosome-to-Golgi retrograde transport in yeast. (Translated from eng) *J Cell Biol* 142(3):665-681 (in eng).

References

54. Temkin P, *et al.* (2011) SNX27 mediates retromer tubule entry and endosome-to-plasma membrane trafficking of signalling receptors. (Translated from eng) *Nature cell biology* 13(6):715-721 (in eng).
55. Seaman MN (2004) Cargo-selective endosomal sorting for retrieval to the Golgi requires retromer. (Translated from eng) *The Journal of cell biology* 165(1):111-122 (in eng).
56. Tabuchi M, Yanatori I, Kawai Y, & Kishi F (2010) Retromer-mediated direct sorting is required for proper endosomal recycling of the mammalian iron transporter DMT1. (Translated from eng) *J Cell Sci* 123(Pt 5):756-766 (in eng).
57. Nielsen MS, *et al.* (2007) Sorting by the cytoplasmic domain of the amyloid precursor protein binding receptor SorLA. (Translated from eng) *Mol Cell Biol* 27(19):6842-6851 (in eng).
58. Lieu ZZ & Gleeson PA (2010) Identification of different itineraries and retromer components for endosome-to-Golgi transport of TGN38 and Shiga toxin. (Translated from eng) *Eur J Cell Biol* 89(5):379-393 (in eng).
59. Wilcke M, *et al.* (2000) Rab11 regulates the compartmentalization of early endosomes required for efficient transport from early endosomes to the trans-golgi network. (Translated from eng) *J Cell Biol* 151(6):1207-1220 (in eng).
60. Riederer MA, Soldati T, Shapiro AD, Lin J, & Pfeffer SR (1994) Lysosome biogenesis requires Rab9 function and receptor recycling from endosomes to the trans-Golgi network. (Translated from eng) *J Cell Biol* 125(3):573-582 (in eng).
61. Lombardi D, *et al.* (1993) Rab9 functions in transport between late endosomes and the trans Golgi network. (Translated from eng) *EMBO J* 12(2):677-682 (in eng).
62. Ganley IG, Espinosa E, & Pfeffer SR (2008) A syntaxin 10-SNARE complex distinguishes two distinct transport routes from endosomes to the trans-Golgi in human cells. (Translated from eng) *J Cell Biol* 180(1):159-172 (in eng).
63. Mallard F, *et al.* (2002) Early/recycling endosomes-to-TGN transport involves two SNARE complexes and a Rab6 isoform. (Translated from eng) *J Cell Biol* 156(4):653-664 (in eng).

References

64. Wassmer T, *et al.* (2009) The retromer coat complex coordinates endosomal sorting and dynein-mediated transport, with carrier recognition by the trans-Golgi network. (Translated from eng) *Dev Cell* 17(1):110-122 (in eng).
65. Nothwehr SF, Ha SA, & Bruinsma P (2000) Sorting of yeast membrane proteins into an endosome-to-Golgi pathway involves direct interaction of their cytosolic domains with Vps35p. (Translated from eng) *J Cell Biol* 151(2):297-310 (in eng).
66. Arighi CN, Hartnell LM, Aguilar RC, Haft CR, & Bonifacino JS (2004) Role of the mammalian retromer in sorting of the cation-independent mannose 6-phosphate receptor. (Translated from eng) *J Cell Biol* 165(1):123-133 (in eng).
67. Fjorback AW & Andersen OM (2012) SorLA is a molecular link for retromer-dependent sorting of the Amyloid precursor protein. (Translated from eng) *Commun Integr Biol* 5(6):616-619 (in eng).
68. Kerr MC, *et al.* (2005) A novel mammalian retromer component, Vps26B. (Translated from eng) *Traffic* 6(11):991-1001 (in eng).
69. Bugarcic A, *et al.* (2011) Vps26A and Vps26B subunits define distinct retromer complexes. (Translated from eng) *Traffic* 12(12):1759-1773 (in eng).
70. Peter BJ, *et al.* (2004) BAR domains as sensors of membrane curvature: the amphiphysin BAR structure. (Translated from eng) *Science* 303(5657):495-499 (in eng).
71. Cozier GE, *et al.* (2002) The phox homology (PX) domain-dependent, 3-phosphoinositide-mediated association of sorting nexin-1 with an early sorting endosomal compartment is required for its ability to regulate epidermal growth factor receptor degradation. (Translated from eng) *J Biol Chem* 277(50):48730-48736 (in eng).
72. Carlton J, *et al.* (2004) Sorting nexin-1 mediates tubular endosome-to-TGN transport through coincidence sensing of high- curvature membranes and 3-phosphoinositides. (Translated from eng) *Curr Biol* 14(20):1791-1800 (in eng).
73. Carlton JG, *et al.* (2005) Sorting nexin-2 is associated with tubular elements of the early endosome, but is not essential for retromer-mediated endosome-to-TGN transport. (Translated from eng) *J Cell Sci* 118(Pt 19):4527-4539 (in eng).

References

74. Cullen PJ & Korswagen HC (2012) Sorting nexins provide diversity for retromer-dependent trafficking events. (Translated from eng) *Nat Cell Biol* 14(1):29-37 (in eng).
75. Harterink M, *et al.* (2011) A SNX3-dependent retromer pathway mediates retrograde transport of the Wnt sorting receptor Wntless and is required for Wnt secretion. (Translated from eng) *Nat Cell Biol* 13(8):914-923 (in eng).
76. Yu JW & Lemmon MA (2001) All phox homology (PX) domains from *Saccharomyces cerevisiae* specifically recognize phosphatidylinositol 3-phosphate. (Translated from eng) *J Biol Chem* 276(47):44179-44184 (in eng).
77. Seaman MN, Harbour ME, Tattersall D, Read E, & Bright N (2009) Membrane recruitment of the cargo-selective retromer subcomplex is catalysed by the small GTPase Rab7 and inhibited by the Rab-GAP TBC1D5. (Translated from eng) *J Cell Sci* 122(Pt 14):2371-2382 (in eng).
78. Rojas R, *et al.* (2008) Regulation of retromer recruitment to endosomes by sequential action of Rab5 and Rab7. (Translated from eng) *J Cell Biol* 183(3):513-526 (in eng).
79. Vardarajan BN, *et al.* (2012) Identification of Alzheimer disease-associated variants in genes that regulate retromer function. (Translated from eng) *Neurobiol Aging* 33(9):2231 e2215-2231 e2230 (in eng).
80. Seaman MN (2012) The retromer complex - endosomal protein recycling and beyond. (Translated from eng) *J Cell Sci* 125(Pt 20):4693-4702 (in eng).
81. Rojas R, Kametaka S, Haft CR, & Bonifacino JS (2007) Interchangeable but essential functions of SNX1 and SNX2 in the association of retromer with endosomes and the trafficking of mannose 6-phosphate receptors. (Translated from eng) *Mol Cell Biol* 27(3):1112-1124 (in eng).
82. Harbour ME, Breusegem SY, & Seaman MN (2012) Recruitment of the endosomal WASH complex is mediated by the extended 'tail' of Fam21 binding to the retromer protein Vps35. (Translated from eng) *Biochem J* 442(1):209-220 (in eng).
83. McGourty K, *et al.* (2012) Salmonella inhibits retrograde trafficking of mannose-6-phosphate receptors and lysosome function. (Translated from eng) *Science* 338(6109):963-967 (in eng).

References

84. Sandvig K, Skotland T, van Deurs B, & Klokke TI (2013) Retrograde transport of protein toxins through the Golgi apparatus. (Translated from eng) *Histochem Cell Biol* 140(3):317-326 (in eng).
85. Finsel I, *et al.* (2013) The Legionella effector RidL inhibits retrograde trafficking to promote intracellular replication. (Translated from eng) *Cell Host Microbe* 14(1):38-50 (in eng).
86. Lipovsky A, *et al.* (2013) Genome-wide siRNA screen identifies the retromer as a cellular entry factor for human papillomavirus. (Translated from eng) *Proc Natl Acad Sci U S A* 110(18):7452-7457 (in eng).
87. Fuchs E, *et al.* (2007) Specific Rab GTPase-activating proteins define the Shiga toxin and epidermal growth factor uptake pathways. (Translated from eng) *J Cell Biol* 177(6):1133-1143 (in eng).
88. Liu TT, Gomez TS, Sackey BK, Billadeau DD, & Burd CG (2012) Rab GTPase regulation of retromer-mediated cargo export during endosome maturation. (Translated from eng) *Mol Biol Cell* 23(13):2505-2515 (in eng).
89. Stechmann B, *et al.* (2010) Inhibition of retrograde transport protects mice from lethal ricin challenge. (Translated from eng) *Cell* 141(2):231-242 (in eng).
90. McDonough JA, *et al.* (2013) Host pathways important for *Coxiella burnetii* infection revealed by genome-wide RNA interference screening. (Translated from eng) *MBio* 4(1):e00606-00612 (in eng).
91. Hierro A, *et al.* (2007) Functional architecture of the retromer cargo-recognition complex. (Translated from eng) *Nature* 449(7165):1063-1067 (in eng).
92. Starr T, *et al.* (2012) Selective subversion of autophagy complexes facilitates completion of the *Brucella* intracellular cycle. (Translated from eng) *Cell Host Microbe* 11(1):33-45 (in eng).
93. Drobne D (2013) 3D imaging of cells and tissues by focused ion beam/scanning electron microscopy (FIB/SEM). (Translated from eng) *Methods Mol Biol* 950:275-292 (in eng).
94. de Barsey M, *et al.* (2011) Identification of a *Brucella* spp. secreted effector specifically interacting with human small GTPase Rab2. (Translated from eng) *Cell Microbiol* 13(7):1044-1058 (in eng).

References

95. Pizarro-Cerda J, *et al.* (1998) *Brucella abortus* transits through the autophagic pathway and replicates in the endoplasmic reticulum of nonprofessional phagocytes. (Translated from eng) *Infect Immun* 66(12):5711-5724 (in eng).
96. Salcedo SP, *et al.* (2008) *Brucella* control of dendritic cell maturation is dependent on the TIR-containing protein Btp1. (Translated from eng) *PLoS Pathog* 4(2):e21 (in eng).
97. Bellaire BH, Roop RM, 2nd, & Cardelli JA (2005) Opsonized virulent *Brucella abortus* replicates within nonacidic, endoplasmic reticulum-negative, LAMP-1-positive phagosomes in human monocytes. (Translated from eng) *Infect Immun* 73(6):3702-3713 (in eng).
98. Voeltz GK, Prinz WA, Shibata Y, Rist JM, & Rapoport TA (2006) A class of membrane proteins shaping the tubular endoplasmic reticulum. (Translated from eng) *Cell* 124(3):573-586 (in eng).
99. Klopfenstein DR, Kappeler F, & Hauri HP (1998) A novel direct interaction of endoplasmic reticulum with microtubules. (Translated from eng) *EMBO J* 17(21):6168-6177 (in eng).
100. Schweizer A, Rohrer J, Hauri HP, & Kornfeld S (1994) Retention of p63 in an ER-Golgi intermediate compartment depends on the presence of all three of its domains and on its ability to form oligomers. (Translated from eng) *J Cell Biol* 126(1):25-39 (in eng).
101. Wakana Y, *et al.* (2008) Bap31 is an itinerant protein that moves between the peripheral endoplasmic reticulum (ER) and a juxtannuclear compartment related to ER-associated Degradation. (Translated from eng) *Mol Biol Cell* 19(5):1825-1836 (in eng).
102. Lin S, Sun S, & Hu J (2012) Molecular basis for sculpting the endoplasmic reticulum membrane. (Translated from eng) *Int J Biochem Cell Biol* 44(9):1436-1443 (in eng).
103. Munro S & Pelham HR (1987) A C-terminal signal prevents secretion of luminal ER proteins. (Translated from eng) *Cell* 48(5):899-907 (in eng).
104. Afshar N, Black BE, & Paschal BM (2005) Retrotranslocation of the chaperone calreticulin from the endoplasmic reticulum lumen to the cytosol. (Translated from eng) *Mol Cell Biol* 25(20):8844-8853 (in eng).
105. Schweizer A, Ericsson M, Bachi T, Griffiths G, & Hauri HP (1993) Characterization of a novel 63 kDa membrane protein. Implications for the

References

- organization of the ER-to-Golgi pathway. (Translated from eng) *Journal of cell science* 104 (Pt 3):671-683 (in eng).
106. Klumperman J, *et al.* (1998) The recycling pathway of protein ERGIC-53 and dynamics of the ER-Golgi intermediate compartment. (Translated from eng) *Journal of cell science* 111 (Pt 22):3411-3425 (in eng).
107. Pelkmans L, Kartenbeck J, & Helenius A (2001) Caveolar endocytosis of simian virus 40 reveals a new two-step vesicular-transport pathway to the ER. (Translated from eng) *Nat Cell Biol* 3(5):473-483 (in eng).
108. Vonderheit A & Helenius A (2005) Rab7 associates with early endosomes to mediate sorting and transport of Semliki forest virus to late endosomes. (Translated from eng) *PLoS Biol* 3(7):e233 (in eng).
109. Burack MA, Silverman MA, & Banker G (2000) The role of selective transport in neuronal protein sorting. (Translated from eng) *Neuron* 26(2):465-472 (in eng).
110. Sherer NM, *et al.* (2003) Visualization of retroviral replication in living cells reveals budding into multivesicular bodies. (Translated from eng) *Traffic* 4(11):785-801 (in eng).
111. Short B, *et al.* (2001) A GRASP55-rab2 effector complex linking Golgi structure to membrane traffic. (Translated from eng) *J Cell Biol* 155(6):877-883 (in eng).
112. Ben-Tekaya H, Miura K, Pepperkok R, & Hauri HP (2005) Live imaging of bidirectional traffic from the ERGIC. (Translated from eng) *J Cell Sci* 118(Pt 2):357-367 (in eng).
113. Schliwa M, Euteneuer U, Bulinski JC, & Izant JG (1981) Calcium lability of cytoplasmic microtubules and its modulation by microtubule-associated proteins. (Translated from eng) *Proceedings of the National Academy of Sciences of the United States of America* 78(2):1037-1041 (in eng).
114. Tokuyasu KT (1973) A technique for ultracryotomy of cell suspensions and tissues. (Translated from eng) *The Journal of cell biology* 57(2):551-565 (in eng).
115. Pauli Rämö AD, Cécile Arrieumerlou, Niko Beerenwinkel, Houchaima Ben-Tekaya, Bettina Cardel, Alain Casanova, Raquel Conde-Alvarez, Pascale Cossart, Gábor Csúcs, Simone Eicher, Mario Emmenlauer, Urs Greber, Wolf-Dietrich Hardt, Ari Helenius, Christoph Kasper, Andreas Kaufmann, Saskia

References

- Kreibich, Andreas Kühbacher, Peter Kunszt, Shyan Huey Low, Jason Mercer, Daria Mudrak, Simone Muntwiler, Lucas Pelkmans, Javier Pizarro-Cerdá, Michael Podvinec, Eva Pujadas, Bernd Rinn, Vincent Rouilly, Fabian Schmich, Juliane Siebourg-Polster, Berend Snijder, Michael Stebler, Gabriel Studer, Ewa Szczurek, Matthias Truttmann, Christian von Mering, Andreas Vonderheit, Artur Yakimovich, Peter Bühlmann & Christoph Dehio (submitted manuscript) Parallel Mixed Model for Simultaneous Analysis of Multiple Large-Scale RNAi Screens.
116. Franceschini A, *et al.* (2014) Specific inhibition of diverse pathogens in human cells by synthetic microRNA-like oligonucleotides inferred from RNAi screens. (Translated from eng) *Proc Natl Acad Sci U S A* 111(12):4548-4553 (in eng).
117. Konig R, *et al.* (2007) A probability-based approach for the analysis of large-scale RNAi screens. (Translated from eng) *Nat Methods* 4(10):847-849 (in eng).
118. Mali P, *et al.* (2013) RNA-guided human genome engineering via Cas9. (Translated from eng) *Science* 339(6121):823-826 (in eng).
119. Cong L, *et al.* (2013) Multiplex genome engineering using CRISPR/Cas systems. (Translated from eng) *Science* 339(6121):819-823 (in eng).
120. Edlund S, Landstrom M, Heldin CH, & Aspenstrom P (2002) Transforming growth factor-beta-induced mobilization of actin cytoskeleton requires signaling by small GTPases Cdc42 and RhoA. (Translated from eng) *Mol Biol Cell* 13(3):902-914 (in eng).
121. Vardouli L, Moustakas A, & Stournaras C (2005) LIM-kinase 2 and cofilin phosphorylation mediate actin cytoskeleton reorganization induced by transforming growth factor-beta. (Translated from eng) *J Biol Chem* 280(12):11448-11457 (in eng).
122. Dettleux PG, Deyoe BL, & Cheville NF (1990) Entry and intracellular localization of *Brucella* spp. in Vero cells: fluorescence and electron microscopy. (Translated from eng) *Vet Pathol* 27(5):317-328 (in eng).
123. Dettleux PG, Deyoe BL, & Cheville NF (1991) Effect of endocytic and metabolic inhibitors on the internalization and intracellular growth of *Brucella abortus* in Vero cells. (Translated from eng) *Am J Vet Res* 52(10):1658-1664 (in eng).

References

124. Watarai M, *et al.* (2002) Macrophage plasma membrane cholesterol contributes to *Brucella abortus* infection of mice. (Translated from eng) *Infect Immun* 70(9):4818-4825 (in eng).
125. Naroeni A & Porte F (2002) Role of cholesterol and the ganglioside GM(1) in entry and short-term survival of *Brucella suis* in murine macrophages. (Translated from eng) *Infect Immun* 70(3):1640-1644 (in eng).
126. Kim S, *et al.* (2004) Lipid raft microdomains mediate class A scavenger receptor-dependent infection of *Brucella abortus*. (Translated from eng) *Microb Pathog* 37(1):11-19 (in eng).
127. Amessou M, *et al.* (2007) Syntaxin 16 and syntaxin 5 are required for efficient retrograde transport of several exogenous and endogenous cargo proteins. (Translated from eng) *J Cell Sci* 120(Pt 8):1457-1468 (in eng).
128. Fugier E, *et al.* (2009) The glyceraldehyde-3-phosphate dehydrogenase and the small GTPase Rab 2 are crucial for *Brucella* replication. (Translated from eng) *PLoS Pathog* 5(6):e1000487 (in eng).
129. van Rahden VA, *et al.* (2012) The 5-phosphatase OCRL mediates retrograde transport of the mannose 6-phosphate receptor by regulating a Rac1-cofilin signalling module. (Translated from eng) *Hum Mol Genet* 21(23):5019-5038 (in eng).

ACKNOWLEDGEMENTS

6. ACKNOWLEDGEMENTS

This work was carried out in the group of Prof. Christoph Dehio in the Focal Area Infection Biology at the Biozentrum of University of Basel, Switzerland.

Firstly, I would like to thank my supervisor Prof. Christoph Dehio for giving me the opportunity to work in his group and with this project. I am grateful for the guidance, positive criticism and support that were given throughout this project.

I would like to thank my thesis committee members, Prof. Jean Pieters and Prof. Martin Spiess for their constructive inputs during the PhD committee meetings and for discussions whenever needed.

I would like to thank Dr. Raquel Conde-Alvarez who has introduced me to the *Brucella* field. I learnt a lot under her guidance that laid a foundation for my work.

I would like to thank Alain Casanova for being a great teammate in the screening projects. We are the best team and it was a lot of fun working with you!

I would like to Dr. Houchaima Ben Tekaya and Dr. Simone Eicher for great discussions, especially in the field of trafficking and cell biology.

I would like to thank Mario Emmenlauer and Dr. Pauli Ramo for the huge support provided throughout this project in terms of high-throughput data analysis.

I would like to thank Alain Casanova, Dr. Simone Eicher, Dr. Kathrin Pieleles and Dr. Rusudan Okujava for careful reading of my thesis.

I would like to thank Jaroslaw Sedzicki and Dr. Christopher Bleck for the great support with the electron microscopy studies.

I would like to thank Dr. Sonia Borrell for guidance in the BSL3 laboratory.

Acknowledgements

Furthermore, I would like to thank former and current members of the Dehio group : Dr. Matthias Truttmann, Dr. Kathrin Pieles, Alain Casanova, Yun-Yueh Lu, Dr. Rusudan Okujava, Claudia Mistl, Dr. Houchaima-Ben-Tekaya, Alexander Harms, Dr. Simone Eicher, Frederic Stanger, Dr. Isabel Sorg, Dr. Anne-Cecile Hiebel, Dr. Tjaart De Beer, Dr. Maxime Quebatte, Mario Emmenlauer, Dr. Pauli Ramo, Dr. Sabrina Siamer, Simon Marlaire, Therese Tschon, Dr. Christoph Schmutz, Jonas Korner, Sarah Stiegeler, Damian Murezzan for the wonderful and fun working atmosphere in the lab.

Special thanks to Claudia Mistl, Simone Muntwiler and Julia Feldmann for technical support in the general lab and BSL3 respectively, also to secretaries Claudia Erbel, Michaela Hanisch for always helping with everything. I would also like to thank Andreas Hefti for helping always with IT related problems, Roger Sauder and Marina Kuhn for technical assistance.

I would like to thank Kathrin Pieles, Simone Eicher, Sarah Stiegeler, Alain Casanova, Rusudan Okujava, Raquel Conde, Claudia Mistl, Ying Ying Lee, Li Juan Pang, Yun-Yueh Lu, Madhan Selvaraj and Ivana Dinulovic for the great times in Basel.

I would also like to thank everyone from the 4th floor who has helped me during this entire project. I would like to thank the Imaging facility for their great support with my experiments.

Most importantly, I am very grateful to my parents Low Poh Fee and Yap Kwee Chee, who have given me immense support throughout my life, and for always believing in me.

CURRICULUM VITAE

7. CURRICULUM VITAE

Shyan Huey Low

PERSONAL DETAILS

Nationality : Malaysian
Date of Birth : 13th September 1984
Address : Klybeckstrasse 18, Basel,
Switzerland
Email Address : shyan.low@unibas.ch

EDUCATION

7/2010 - present PhD student at Biozentrum Basel, Department of Molecular Microbiology, Focal area Infection Biology
2/2010 - 6/2010 Masters student at Biozentrum Basel
2003-2006 BSc (Hons) Biochemistry with Molecular Biology and Biotechnology at University of Bristol, UK
2002 – 2003 A levels at Taylor's College, Malaysia
1997- 2001 Catholic High School, Petaling Jaya, Malaysia

PUBLICATIONS

2014 Andreas Franceschini*, Roger Meier*, Alain Casanova*, Saskia Kreibich*, Neha Daga, Daniel Andritschke, Sabrina Dilling, Pauli Ramo, Mario Emmelauer, Andreas Kaufmann, Raquel Conde-Alvarez, Shyan Huey Low, Lucas Pelkmans, Ari Helenius, Wolf-Dietrich Hardt, Christoph Dehio, and Christian von Mering
Specific inhibition of diverse pathogens in human cells by synthetic microRNA-like oligonucleotides inferred from RNAi screens
PNAS, March 25 2014, vol. 111 no 12

2012 Jayadev R, Kuk CY, Low SH, Murata-Hori M.
Calcium sensitivity of alpha-actinin is required for equatorial actin assembly during cytokinesis
Cell Cycle. 2012 May 15;11(10):1929-37

2010 Zhong XY, Zhang B, Asadollahi R, Low SH, Holzgreve W.
Umbilical cord blood stem cells: what to expect
Ann N Y Acad Sci. 2010 Sep;1205:17-22

Curriculum vitae

- 2010 Low SH, Mukhina S, Srinivas V, Ng CZ, Murata-Hori M.
Domain analysis of alpha actinin reveals new aspects of its association with
F-actin during cytokinesis.
Exp Cell Res. 2010 Jul 15;316(12):1925-34

HONOURS AND AWARDS

- Fellow of International PhD Program 'Fellowship for Excellence'
- Awarded Nuffield Foundation Undergraduate Research Bursary (UK) 2005

TEACHING EXPERIENCE

- 2011-2014 Assistant and tutor for undergraduate students in practical courses
'Microbiology'
Focal area infection biology, Biozentrum, University of Basel

POSTER PRESENTATIONS

- 2013 Poster presentation at 'Microbial Pathogenesis & Host Response' Cold
Spring Harbor Laboratory meeting
- 2013 Poster presentation at Systems Biology of Infection Symposium
- 2012 Poster presentation at 'Cellular organization and functions and their
subversion by pathogens' EMBO conference
- 2011 Poster presentation at FEMS conference, Geneva

ADDITIONAL EXPERIENCE

Assistant Research Officer, Mammalian Cell Biology Group, Temasek Life Sciences
Laboratory, Singapore Nov 2006 – Jan 2010

Project: Regulation of the interaction of equatorial F actin and alpha actinin along the
cleavage furrow during cytokinesis in mammalian cells

Honours research project, University of Bristol, UK Oct 2005- Jan 2006

Project: Investigation of the role of ER to Golgi transport in cellular reorganization
during wound healing

Summer Internship, University of Bristol, UK June - July 2004

Project: Characterization of *C. elegans* mutants generated through *mos-1* transposon
mobilization and to investigate efficiency of *mos-1* transposon jumping in *C. elegans*
genome through biolistic transformation