RNA Interference Screening Identifies Host Factors Involved in Brucella Infection

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Alain Casanova

aus Calanca (Braggio), GR

Basel, 2016

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät

auf Antrag von

Prof. Dr. Christoph Dehio
Prof. Dr. Dirk Bumann

Basel, den 11.11.2014

Prof. Dr. Jörg Schibler
Dekan

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. Dirk Bumann
Prof. Dr. Michael N. Hall
Prof. Dr. Mihaela Zavolan

My thesis is written as a cumulative dissertation. It consists of a summary of the major findings of my thesis, a general introduction covering aspects relevant for this work followed by the result sections composed of a scientific publication, one submitted article, and three manuscripts in preparation. Supplementary information can be found on the CD attached to the thesis. Finally, the different aspects of my thesis are discussed and future directions of the project are suggested.

SUMMARY

Intracellular pathogens rely to varying extents on cellular functions of the host cell for their own propagation. A number of bacteria have evolved strategies to invade human cells and to establish an intracellular niche, which often consists of a cellular compartment that is modified by the pathogen to its own benefit. To understand the infection strategies of such organisms and eventually design new medical interventions, knowledge on the host factors exploited by the pathogens is critical. To this end, the InfectX consortium attempts to decipher the human infectome for a number of bacterial and viral pathogens.

In this framework, we study the zoonotic pathogen *Brucella,* which is able to invade phagocytic as well as non-phagocytic cells. The molecular mechanisms by which *Brucella* enters cells, evades lysosomal degradation, and finally replicates in an endoplasmic reticulum-like compartment, remain largely unknown.

To identify host factors involved in these processes, genome-wide microscopy-based RNA interference (RNAi) screens of *Brucella* entry and replication in HeLa cells were performed. To assign the function of the hits from the genome-wide screen to either early or late stages of *Brucella* infection, a follow-up assay suitable for high-throughput screening of *Brucella* entry was established. Both screening protocols are described in detail in *research article I*.

In-depth analysis of the genome-wide siRNA data generated within InfectX found that all screens including the one for *Brucella* infection show signs of miRNA-like off-target effects. *Research article II* focuses on the discovery and validation of this phenomenon in siRNAs screens and illustrates the potential of such an analysis to discover natural miRNAs and synthetic miRNA-like molecules that regulate the process of study. These findings motivated the screening of a library of human miRNA mimics for their involvement in *Brucella* infection. We identified miR-103 and miR-107 (miR-103/107), which strongly promote *Brucella* entry in non-phagocytic cells as presented in *research article III*. Interestingly, also the infection of other pathogens tested within InfectX was promoted by these miRNAs. Proteome and transcriptome analyses of cells with high levels of miR-103/107 indicated that this alters endocytic properties which manifested in reduced clathrin dependent uptake of transferrin in these cells. Furthermore, the abundance of several surface receptors required by different pathogens is increased. TGF-β receptor 2 showed elevated expression upon miR-103/107 transfection and independent experiments could confirm

that high levels of this transmembrane kinase promote *Brucella* infection.

Having analyzed the full scale of off-target effects, we set out to determine a strategy to validate candidate genes of the genome-wide screens. We thus assayed a set of human kinases with a total of eleven individual siRNAs and one siRNA pool. *Research article IV* shows that the true discovery rate is directly proportional to the number of siRNAs tested and that siRNA pools tend to give more reliable results than individual siRNAs. As a consequence of these findings, we used six independent siRNAs and one siRNA pool for the validation of genes discovered in the primary screen with one pooled and five single siRNAs. This allowed the identification of several host cell pathways relevant for *Brucella* infection. Besides previously known functions, which include actin cytoskeleton remodeling or maturation of endocytic vesicles, also novel ones such as FGF and TGF-β signaling were found. While most of these networks were connected to *Brucella* entry into HeLa cells, we were also able to identify retrograde trafficking between endosomes and the Golgi apparatus to regulate a post-entry process as presented in *research article IV*.

Altogether, the results of our studies presented here point out limitations as well as the potential of siRNA technology. If off-target effects are accounted for and experimental confirmation is applied carefully on identified factors, RNAi allows to successfully reveal genes and pathways hitherto unrelated to the mechanism of interest. Additionally, and if analyzed accordingly, off-target effects also constitute a rich source of information for the discovery of miRNAs and miRNA-like molecules that regulate a certain process. Applied to the presented screens for human factors taking part in *Brucella* infection, this led to the description of miRNAs and several host pathways, which support pathogenicity. By that our results contribute to the expansion of the currently described infectome for this intracellular pathogen.

CONTENTS

1 Introduction

1 INTRODUCTION

1.1 VESICULAR TRAFFICKING IN MAMMALIAN CELLS

A fundamental difference between eukaryotic and prokaryotic life is the compartmentalization of the cytosol. Eukaryotic cells contain several membrane-enclosed compartments such as the nucleus, endoplasmic reticulum (ER), or mitochondria. These compartments are characterized by specific molecular compositions and chemical properties. In order to exchange material between these compartments, cells have evolved different mechanisms. Channels and pores selectively transport small molecules, bigger gates, such as the nuclear pore, can accommodate macromolecules, and specialized transport vesicles can shuttle large volumes of cargo between compartments.

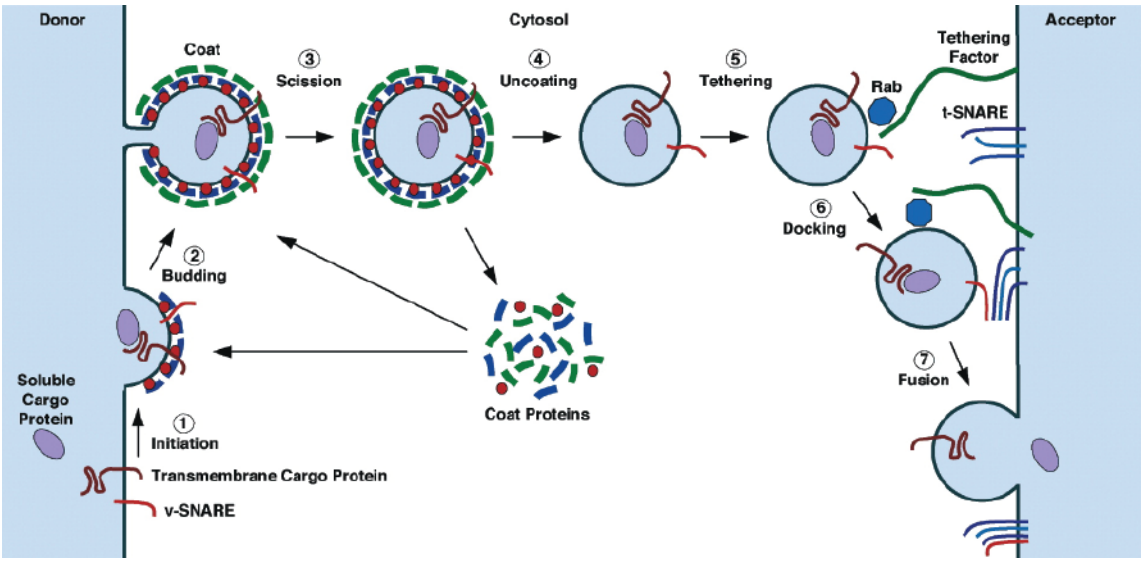


Figure 1. Vesicular budding and fusion. 1) Initiation of coat assembly. Inner layer of coat proteins (blue) binds to cargo and a membrane-associated small GTPase (e.g. Sar1) at a donor membrane. 2) Budding. Outer coat layer (green) assembles and forms the complete coat around the concentrated cargo. 3) Scission. Membrane between vesicle and donor membrane is severed to release the vesicle from the membrane. 4) Uncoating. Coat proteins are released from the vesicle by the action of several processes such as inactivation of the small GTPase. Coat proteins are reused on newly forming vesicles. 5) Tethering. Vesicle binds to the target membrane by interaction of a GTP bound Rab and a tethering factor. 6) Docking. The v- and t-SNARES assemble to a four-helix bundle pulling the vesicle closer to the membrane. 7) Fusion. The action of the SNARE complex promotes fusion of the vesicle and the target membrane. Soluble cargo is released into the acceptor compartment. Picture taken from (*1*).

Vesicular transport occurs by budding of vesicles from a donor compartment and subsequent fusion to an acceptor membrane (Figure 1). The concentration of selected cargo for transport and the bending of the membrane for budding are dependent on coat proteins. The correct targeting of the vesicle to the target membrane is determined by tethering and SNARE proteins. The latter also provide energy for fusion of the transport vesicle to the acceptor membrane. In order to maintain the integrity and identity of all compartments such as the ER, the Golgi apparatus, and the endo-lysosomal system, vesicular transport acts bidirectional to return membranes and resident proteins to their correct location (*1*).

1.1.1 Compartments of vesicular trafficking

The ER is a continuous membrane system which is connected to the nuclear membrane and forms tubular and cisternal structures. The rough ER is coated with ribosomes which translocate proteins dedicated to the secretory pathway into the lumen of the ER. These proteins encode a translocation signal in their N-terminus, which is recognized by the signal-recognition particle SRP as it emerges from the ribosome. The whole complex is then targeted to the translocon spanning the ER membrane and the protein is pushed into the lumen of the ER as it emerges from the ribosome (*2*). Membrane proteins are directly integrated into the ER membrane in a process which is believed to involve lateral escape of membrane spanning segments through the tranlocon (*3*). For the ER, luminal and membrane proteins are further distributed to lysosomes, Golgi apparatus, plasma membrane, or secreted to the extracellular space. The smooth ER is involved in the biosynthesis of lipids and steroids. Further functions of the ER include folding and modification of proteins, storage of calcium, and it serves as a source of membrane for the biogenesis of other organelles (*4*).

The Golgi apparatus serves as a central hub where proteins are sorted and get directed to the ER, the plasma membrane, and other organelles such as lysosomes. Furthermore, it is the compartment where the majority of chemical modifications on proteins and lipids are conducted. Long carbohydrate chains are linked to proteins of the extracellular matrix, sugar moieties are added to lipids, and sulfotransferases add sulfates to sugar chains or directly to proteins. The Golgi forms stacks and can be divided in *cis-*, *medial-*, and *trans*-Golgi depending on specific sets of resident proteins. The *cis*-Golgi faces the ER from where it receives newly synthesized proteins while the *trans*-Golgi sorts incoming and outgoing proteins for their final destination (*5, 6*).

Endosomes and lysosomes together form a highly dynamic network of vesicular structures that is scattered throughout the cell. Early endosomes are the first address for incoming vesicles from the plasma membrane and are involved in sorting their content; if the cargo is destined for degradation, early endosomes will mature to late endosomes and eventually lysosomes. This degradative compartment is equipped with hydrolytic enzymes and designed to break down virtually any biomolecule (*7*).

1.1.2 Trafficking routes

All vesicular compartments of the cell are highly interconnected by trafficking routes as illustrated in Figure 2. Among the best studied trafficking routes is the so called secretory pathway. It is used to transport proteins which are synthesized in the ER via the ERGIC, Golgi apparatus, and secretory granules to the plasma membrane or the extracellular space. The transport starts at so called ER exit sites (ERES) where cargo is recognized and locally concentrated by COPII coat components (*8*). The signal for export out of the ER is found on cytoplasmic domains of the cargo where they bind to a family of different COPII components which ensure the recognition of diverse signal sequences (*1*). The COPII coat mainly consists of the small GTPase Sar1 which together with Sec23 and Sec24 forms the inner layer, while Sec13 and Sec31 form the outer layer (*9*). Cargo is then delivered to the ER-Golgi intermediate compartment (ERGIC) and from there to the Golgi. GTP hydrolysis of Sar1 initiates the destabilization of the COPII complex, which dissociates from the vesicle (*10*). This allows the vesicle to interact with the acceptor membranes where the SNARE of the vesicle (v-SNARE) interacts with the target SNARE (t-SNARE) to form a four-helix bundle (*11*) which assures target specificity and provides the energy for vesicle fusion (Figure 1). From the *trans*-Golgi, the cargo continues either to the plasma membrane or other organelles such as the lysosomes. For these transport steps other coat proteins, e.g. clathrin, are used instead of COPII (*12*).

Retrograde transport from the Golgi apparatus back to the ER and intra-Golgi transport follow the same principle steps as COPII-mediated transport. Even though different proteins forming the coatomer (COPI complex) are used in the process, they share similar functions. Activation of the small GTPase Arf1 recruits the inner core components β-COP, γ-COP, δ-COP, and ζ-COP and the outer coat α-COP, β'-COP, and ε-COP (*13*). Similar to COPII, COPI also recognizes cytosolic signal sequences which bring back proteins to the ER (*14*).

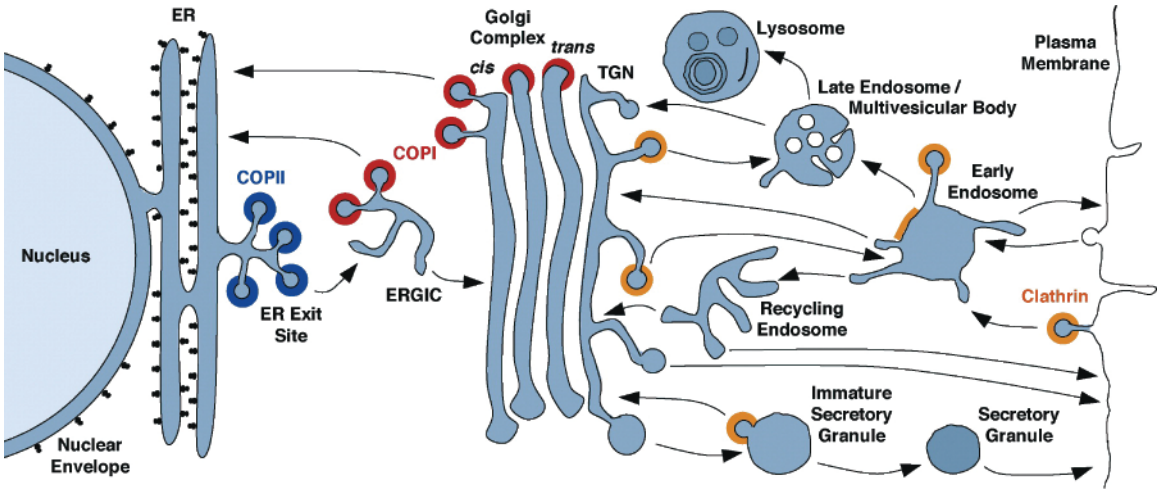


Figure 2. Intracellular trafficking routes. Scheme illustrates trafficking between the endoplasmic reticulum (ER), ER to Golgi intermediate compartment (ERGIC), Golgi apparatus including the *trans*-Golgi network (TGN), endo-lysosomal compartments, and the plasma membrane. The specific location of coat complexes is highlighted in color: COPI (red), COPII (blue), and clathrin (yellow). Transport pathways connecting the compartments are indicated by arrows. Picture taken from (*1*).

While COPII and COPI coated vesicles regulate anterograde and retrograde transport between ER and Golgi, many other pathways regulate trafficking between the Golgi and the plasma membrane. Early endosomes are the main sorting station for incoming cargo that is taken up by a diverse set of endocytic processes (described in detail in chapter 1.1.3.). From early endosomes, cargo can take a number of different routes either leading to degradation in lysosomes or targeting to other compartments. Following the degradative pathways, endosomes undergo a process of maturation. They are initially characterized by the presence of the small GTPase Rab5 which is exchanged for Rab7 during maturation to late endosomes (*15*). At the same time conversion of phosphatidylinositol-3-phosphate PtdIns(3)P to phosphatidylinositol-3,5-bisphosphate (PtdIns(3,5)P_2) occurs. Both events are crucial for maturation of the vesicle and ensure correct binding of effector proteins (*7*). The maturation process is also accompanied by gradual acidification of the vesicle by the action of the vesicular ATPase that supports ligand-receptor dissociation or the activity of lysosomal proteases (*16*).

In late endosomes, selected cargo is internalized in small vesicles leading to the formation of intraluminal vesicle (ILVs) by inward budding of membrane. This process is dependent on the function of the ESCRT complex (ESCRT-0, -I, -II, -III) which recognizes ubiquitin-tags on the

cytosolic domains of proteins (*17*). The selective internalization of the complete proteins and of lipids into the lumen of the late endosomes stops signaling of transmembrane receptors and promotes degradation of the content by hydrolytic enzymes (*7*).

Certain endocytosed receptors such as the transferrin receptor (TfR) (*18*) are generally reused at the plasma membrane rather than degraded after internalization. Trafficking back to the plasma membrane either occurs via recycling endosomes or directly from early endosomes (*19*). Another important alternative route which deviates from the degradative pathway involves the retromer complex which connects endosomes to the *trans*-Golgi network. 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 (*20*). Vps35 is the core component of the trimer and has a direct role in cargo binding (*21, 22*) while Vps26 and Vps29 independently associating at either end. Different types of Snx are known to interact with the CST and depending on their composition are involved in retrograde trafficking from early or late endosomes (*23*).

1.1.3 Endocytic processes

A wide variety of mechanisms are known to be involved in the transport of cargo from the extracellular space to the inside of a cell by the help of vesicular transport (Figure 3). Some of these pathways are constitutive, while others are selectively triggered by binding of certain ligands to surface receptors. Endocytic pathways fulfill essential functions in cellular homeostasis, however, they are also exploited by toxins, viruses, and bacteria to enter cells. The specific nature of these interactions will be discussed in detail in chapter 1.2.2.

Clathrin-mediated endocytosis

The best studied pathway among all endocytic processes is clathrin-mediated endocytosis (CME). It depends on the formation of a clathrin cage around the vesicle and was discovered almost 40 years ago (*24*). Owing to the use of different adapter proteins, this process is very versatile in terms of cargo which can be internalized. Many surface receptors such as epidermal growth factor receptor (EGFR) or transferrin receptor (TfR) are internalized by CME and thus the pathway is also referred to as receptor-mediated endocytosis (*25*).

The process is initiated by deformation of the plasma membrane to form a so called pit. It was traditionally thought that binding of the adaptor protein complex AP2 to cargo initiates this process, however, there is evidence that a class of membrane-bending proteins start the nucleation and membrane bending even before the adaptor is recruited (*26, 27*). These nucleation modules are thought to assemble at the plasma membrane solely due to binding to phosphatidylinositol-4,5-bisphosphates (PtdIns(4,5)P_2) independent of the cargo. Only then, the adaptor complex AP2 together with adaptor proteins are recruited to the initial nucleation modules. After clathrin, AP2 is the most abundant component of clathrin coated vesicles and serves as a central hub during the formation of the vesicle. It interacts directly with PtdIns(4,5)P_2 and cargo (*28, 29*). AP2 then recruits clathrin triskelia from the cytosol. This helps to stabilize the growing pit and promotes membrane bending by organizing curvature effectors (*30*).

Once the pit has fully matured, the neck of the vesicle must be pinched off to release it from the plasma membrane. This depends on the enzyme dynamin (*31*) which polymerizes around the neck of the vesicle. Dynamin undergoes a GTP hydrolysis driven conformational change which leads to membrane fission (*32, 33*). Once released from the membrane, the clathrin coat is released to the cytoplasm and can be recruited to other forming pits. The release of clathrin triskelia from the lattice is induced by the ATPase heat shock cognate 70 (HSC70) and its cofactor auxilin or GAK (*34, 35*).

Macropinocytosis

Macropinocytosis is a clathrin- and dynamin- independent pathway that mediates non-selective uptake of liquid phase. This process is generally induced by growth factor stimulation such as EGF (*36*) but some cell types such as antigen presenting cells are capable of constitutive macropinocytosis (*37, 38*). Hallmarks of macropinocytosis are the formation of large membrane ruffles on the plasma membrane (see Figure 3) and the folding back of lamellipodia on itself followed by fusion with the membrane. This leads to the uptake of large, irregular shaped vesicles between 0.2 μm and 5 μm in diameter (*39*). The process is dependent on actin remodeling and the activity of PI3 kinases (*40*), and small GTPases such as CDC42 (*41*).

Interestingly, macropinocytosis was found to be induced in cells when alternative endocytic processes were inactivated. In HeLa cells and *Drosophila* haemocytes, inhibition of dynamin-dependent processes induces macropinocytosis (*42, 43*). This indicates that cells compensate for the loss of certain endocytic routes by up-regulation of alternative pathways.

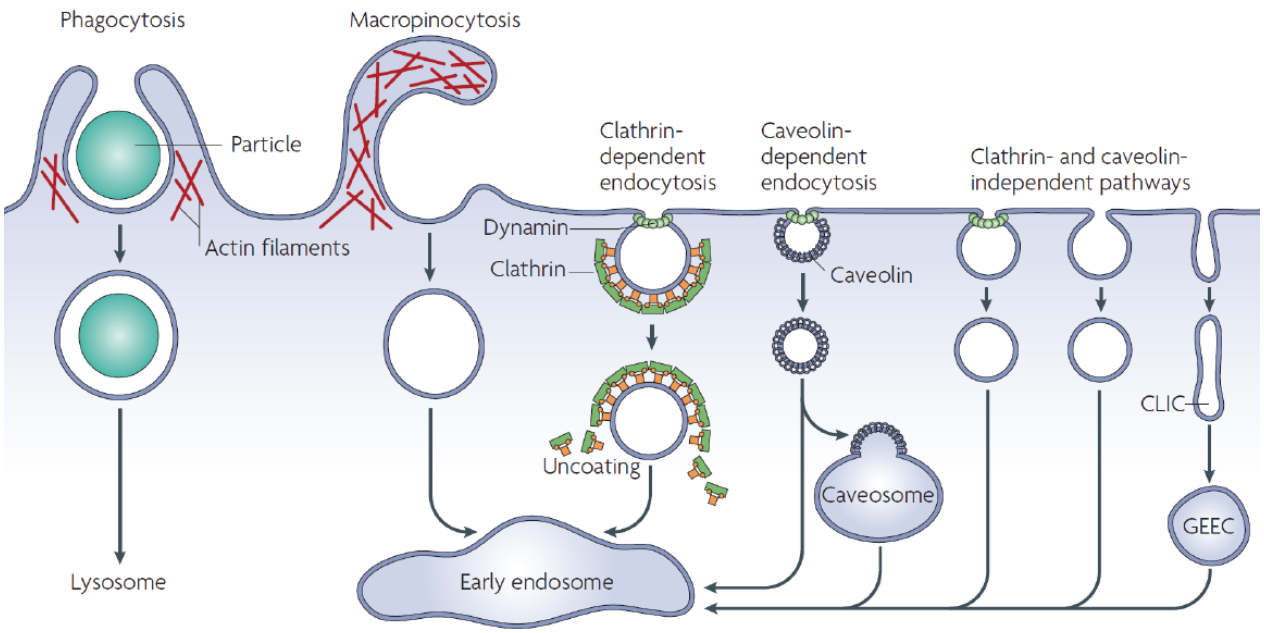


Figure 3. Endocytic pathways. Illustration of different endocytic pathways of eukaryotic cells. Large particles can be taken up by phagocytosis while fluid uptake is mediated by macropinocytosis. Both processes require actin remodeling and protrusions of the plasma membrane. The size of the vesicle that can be taken up is much larger compared to the other pathways. Clathrin mediated endocytosis requires the formation of a clathrin coat and dynamin for the scission. Caveolae are coated in caveolin instead of clathrin and also require dynamin for their release from the membrane. Alternative pathways include clathrin- and dynamin- independent carriers (CLICs) which traffic through glycosyl phosphatidylinositol-anchored protein enriched early endosomal compartments (GEEC) before reaching early endosomes. Picture taken from (*44*).

Phagocytosis

In contrast to macropinocytosis, phagocytosis is a receptor mediated process that allows the uptake of large particles. Only some specialized cell types such as neutrophils, macrophages, and dendritic cells, are capable of this particular type of endocytosis. Phagocytosis allows the uptake of a vast range of foreign bodies including viruses, bacteria, or apoptotic cells and cell debris. It involves a variety of different receptors to interact with specific ligands which then trigger very distinct downstream signaling events. Some receptors directly recognize pathogen-associated molecular patterns (PAMPs) such as peptidoglycan. Others act indirectly by recognizing opsonized particles. Opsonins are soluble factors such as complement fragments (iC3b) or immunoglobulins (IgG) that recognize foreign bodies and mark them for phagocytosis (*45*). Immunoreceptors such as the Fcγ receptor on the surface of professional phagocytic cells can recognize these opsonins and clustering of the receptors induces phagocytosis (*46*). Even though

the receptors and underlying signaling pathways are very distinct, all lead to the formation of actin polymerization which induces the growth of pseudopodia around the particle and culminate in its engulfment and uptake. Like macropinocytosis, the process of actin polymerization is dependent on the activation of small GTPases such as Rac1 and CDC42 (*47*). (*48*).

Clathrin-independent endocytosis

Several less well defined pathways that neither require actin polymerization nor clathrin are known. Among those, caveolae with a distinct size between 60-80 nm are the best studied (*49*). Caveolae are characterized by the enrichment of caveolins which are integral membrane proteins. These are essential for caveolae formation (*50*) and responsible for membrane curvature (*51*). In addition, cavins, a class of cytoplasmic proteins, are involved in the process (*52*). While there are many differences between CME and caveolae, also caveolae depend on dynamin for budding of the vesicle (*53*). The biological role of caveolae is diverse and the process includes not only endocytosis but also other functions like lipid regulation, integrin signaling, or mechanosensing and protection of the plasma membrane (*54*).

1.2 HOST-PATHOGEN INTERACTIONS

Host-pathogen interaction is the interplay between one organism, acting as a host, and a second organism exploiting the resources of that host. This relationship is associated with a negative impact on host fitness and stands in contrast to other interspecies interactions such as commensalism or symbiosis. Host-pathogen interactions result in selective pressure on the host to minimize the burden inflicted by the pathogen. This has led to the evolution of diverse immune strategies to combat pathogens which, in turn, are forced to evolve countermeasures to ensure their survival. As an evolutionary concept, this arms race is known as the Red Queen hypothesis first used by Leigh van Valen in 1973 as an analogy to the novel “Alice’s Adventures in Wonderland” by Charles Lutwidge Dodgeson. It is based on the statement “Now, *here*, you see, it takes all the running you can do, to keep in the same place” “keep in the same place” can be interpreted as species survival and “running” as the process of adaptation. This metaphor aims to explain the extinction of species if they fail to adapt fast enough to changing conditions (*55*).

Host-pathogen interactions are wide-spread throughout all branches of life. In the following I will focus on interactions involving mammalian hosts and viral or bacterial pathogens. From a

pathogen's perspective, several constraints have shaped infection strategies. One such constraint is that the pathogen must ensure its own survival, which comes at a fitness cost for the host cell, while also minimizing those fitness costs in order to maintain its replicative niche. This is crucial on an evolutionary time scale in order to avoid the extinction of the host. Some pathogens cause acute infections, which have a high impact on the host for a short period of time, but allow for rapid pathogen proliferation and spread. Others cause chronic infections which generally have a lower impact on the host over a long period of time, but also results in slow pathogen replication.
Examples of human pathogens that cause acute infections are *Shigella*, causing a food-borne disease associated with acute intestinal infections, or the influenza virus, which causes the flu (*56, 57*). These are generally accompanied by rapid onset of the disease and a high risk of spread to other individuals. Since the infection is generally self-limiting, the body is able to clear the disease within a few days. Chronic infections however, often will not cure without medical interventions and can remain unrecognized for years due to the absence of disease symptoms. Examples include *Tuberculosis* which persists in the lung of patients for decades or Eppstein-Barr virus which causes mononucleosis (*58, 59*).

1.2.1 Intracellular pathogens

In contrast to viruses, which are fully dependent on host cell functions, bacteria show varying degrees of host-dependency. Some pathogenic bacteria do not invade host cells but rather adhere to cells and survive in body fluids enriched in nutrients. A second class of bacteria are obligate intracellular and, like viruses, fully depend on a host cell for replication. Finally, the third class includes facultative intracellular bacteria which invade host cells only when it is beneficial. In the following, I will give an overview of the different strategies employed by bacterial pathogens with a focus on the host-cell interaction.
One of the advantages of an intracellular niche is the availability of metabolic resources. Obligate intracellular bacteria such as *Mycobacterium leprae* have lost many of their own catabolic pathways and take full advantage of metabolites of the host (*60*). While genome-reduction seems to be a general trend for intracellular pathogens, this can lead to a Muller's ratchet phenomenon, where gradual accumulation of deleterious mutations ultimately leads to extinction (*61*). Facultative intracellular bacteria on the other hand generally maintain most of their catabolic

processes but take advantage of the rich resources of a cell. They survive outside of host cells and can be transmitted to susceptible hosts from environmental niches.

Aside from favorable nutrient conditions, an intracellular niche also provides protection from immune functions such as the complement or antibodies. However, the intracellular niche is not devoid of threats to the intruder. Cells are equipped with a number of defense strategies that aim at the detection and destruction of foreign particles. These depend on the microenvironment of the pathogen with a major difference between a cytosolic and a vacuolar niche. A summary of the intracellular lifestyle of different bacterial pathogens is shown in Figure 4.

Bacteria living in the cytosol

A small number of bacteria are known to replicate within the cytosol, with *Listeria monocytogenes* (*Listeria)* and *Shigella flexneri* (*Shigella*) among the best studied. After invasion of the host cell these bacteria are enclosed within a vacuole that gradually acidifies along the endocytic pathway. Therefore, the first step for cytosolic bacteria is to escape from the vacuole. This process occurs rapidly after cell entry which allows bacteria to avoid contact with the microbicidal content of lysosomes (*63, 64*). Lysis of the vacuole is achieved by secretion of bacterial enzymes that attack the membrane. *Listeria* secretes listeriolysine O (LLO) (*65*) as well as type C phospholipases (*66*) which are highly regulated to protect the cell from damage that could be caused if wrong membranes were attacked (*67*). Many cytosolic bacteria use actin-based motility which allows cell-cell spreading (*68*). Pushing of the bacterium from one cell to the other subsequently requires lysis of a double-membrane to access the cytosol of a neighboring cell. Actin polymerization also protects bacteria from detection by the ubiquitin degradation system (*69*). Even though conditions in the cytosol are not as harsh as in lysosomes, the cytosolic environment also presents challenges. These include antimicrobial peptides (*70*) and, perhaps more importantly, immune receptors that recognize pathogen-associated molecular patterns (PAMPs). Examples are Nod-like receptors (NLRs) and Nod proteins that activate inflammatory responses to alarm the immune system (*71*). The host cell is able to recognize flagellin, components of bacterial secretion systems, dsDNA, and several other unique molecular patterns that reveal the presence of an intruder (*72*).

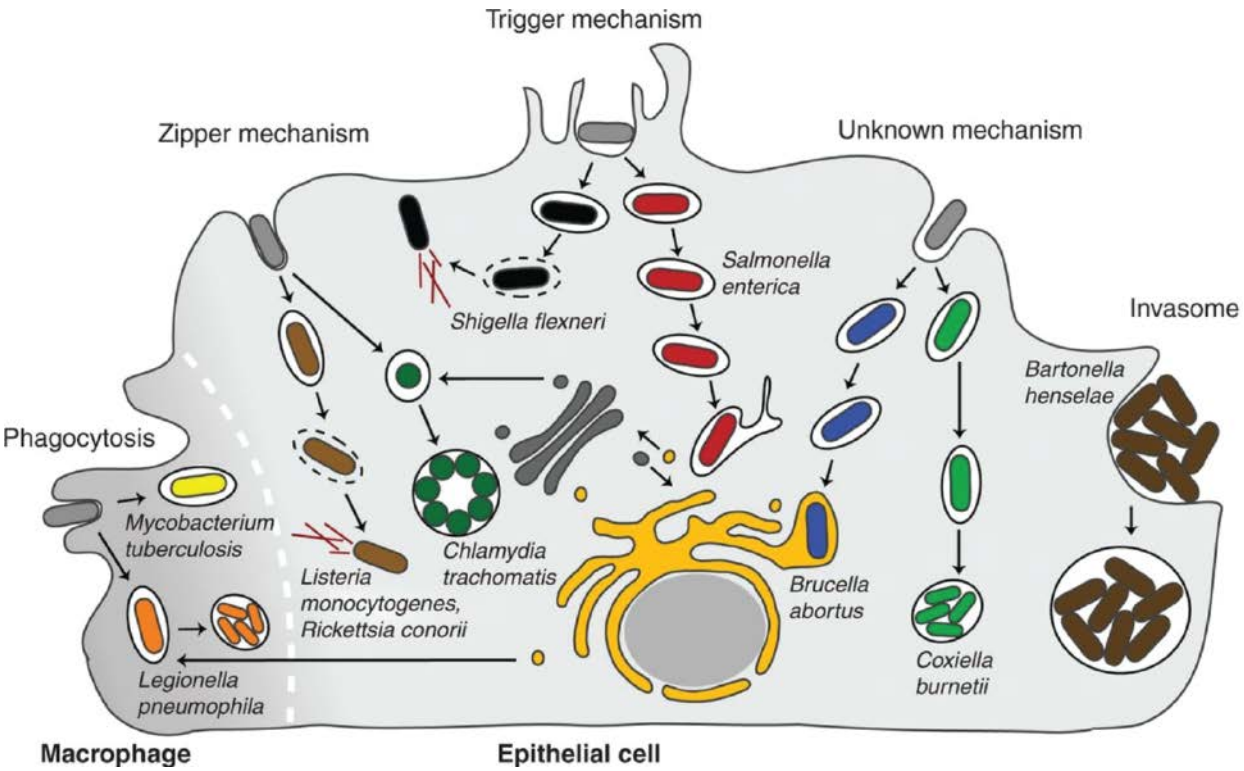


Figure 4. Intracellular pathogens. Illustration of entry mechanism and replicative niche of intracellular bacterial pathogens. *Listeria monocytogenes* (light brown) enters epithelial cells by a zipper mechanism induced by two internalins. It lyses the vacuolar membrane, replicates in the cytosol and performs actin-based motility. *Chlamydia trachomatis* (dark green) replicates in a compartment that intercepts membranes from the Golgi apparatus. *Shigella flexneri* (black) enters cells by a trigger mechanism induced by a type III secretion system (T3SS), lyses its membrane, replicates in the cytoplasm, and performs actin-based motility. *Salmonella enterica* (red) triggers entry by secretion of T3SS (SPI-1) secreted effectors and modulates its vacuole by a second T3SS (SPI-2) which induces the formation of membranous extensions. *Brucella abortus* (blue) replicates in an endoplasmic reticulum-derived compartment. *Coxiella burnetii* (light green) replicates in a lysosome-derived vacuole. *Bartonella henselae* (dark brown) enters endothelial cells as large aggregates in a process dependent on a type IV secretion system. *Mycobacterium tuberculosis* (yellow) block maturation of the phagosome in macrophages. *Legionella pneumophila* (orange) replicates in a vacuolar compartment containing endoplasmic reticulum markers. Picture taken from (*62*).

Until now, it is not clear why only a few bacterial pathogens chose the host cytoplasm as a niche. There are conflicting data regarding the ability of bacteria that generally do not live in the cytoplasm to survive and replicate in this compartment if introduced synthetically. It was found that non-pathogenic *Bacillus subtilis* expressing LLO or *Escherichia coli* that are surface coated with LLO and express a *Yersinia enterocolitica* invasion were both able to replicate in the cytosol of cells (*73, 74*). However, among several species that were microinjected into the cytoplasm of host cells only those bacteria that naturally live in this environment were able to replicate (*75*). It

remains to be investigated whether the way of entering the host cell cytoplasm primes the cell or the bacteria for cytoplasmic viability. It is also unclear whether only a few bacterial species possess the capacity to lyse their vacuole or whether most species are better adapted to intravacuolar replication and thus actively prevent vacuolar escape.

Bacteria in membranous compartments

While bacteria that escape their vacuole do not face potential destruction by the phagolysosome, those that remain in a vacuole have evolved various strategies to avoid such a fate. This generally involves interference with the normal endocytic pathway in favor of the pathogen by secretion of bacterial factors.

Organisms such as *Mycobacterium tuberculosis* or *Salmonella* block the maturation from early to late endosomes using phosphoinositide phosphatases (*76*, *77*). These prevent the phosphorylation of phosphatidylinositols (PtdIns) which accompanies endosomal maturation. The lack of $PtdInsP_2$ and $PtdInsP_3$, and a concomitant accumulation of PtdInsP, blocks the vacuoles in an early endosomal state. In the case of *Salmonella*, this is achieved by SopB in a type III secretion system (T3SS) dependent manner (*78*). In a later stage of infection, the *Salmonella* containing vacuole (SCV) is redirected to a perinuclear region in close proximity to the Golgi apparatus and induces the formation of membrane tubules. These structures are generated by the secreted effector SifA (*79*) and required for the maintenance of the SCV (*80*).

Other bacteria also use the strategy of deviation from the normal endocytic pathway with varying degrees of interaction with late endosomes. The *Brucella* containing vacuole (BCV) interacts with late endosomal markers such as Rab7 or Lamp1 (*81*). Acidification of the vacuole serves as a trigger for the expression of a type IV secretion system (T4SS) (*82*). Secretion of T4SS effectors are believe to be involved in redirecting the BCV to an ER-derived compartment, since T4SS-deficient mutants are degraded in lysosomes (*83*). Also *Legionella* replicates in an ER-associated compartment and redirects its vacuole with the help of a T4SS which secretes over 200 proteins with highly redundant functions into host cells (*84*).

A radically different strategy is followed by *Coxiella*. Instead of avoiding trafficking to a lysosomal compartment, *Coxiella* has learned to survive and replicate in this hostile environment (*85*). Even though the pathogen can withstand harsh conditions, it is very active in reshaping its own vacuole (*Coxiella* containing vacuole, CCV) which starts very early after entering the cell. While normal endocytic cargo can reach the lysosomes within minutes, fusion of lysosomes with

the CCV takes two hours (*86*) and is accompanied by acquisition of autophagosomal structures (*87*).

1.2.2 Cell invasion of pathogens

All intracellular pathogens must reach the inside of a host cell. Even though it seems intuitive to directly penetrate the plasma membrane, very few organisms use this route. Examples that follow such an entry strategy include enveloped animal viruses such as HIV (*88*). In this case the viral membrane fuses with the plasma membrane, directly releasing the viral content into the cytoplasm. However, this is the exception rather than the rule for intracellular pathogens. In most cases, an existing endocytic pathway is used to enter a cell. These include macropinocytosis, clathrin-mediated endocytosis (CME), caveolae, and clathrin- and caveolae-independent endocytosis. With the following examples I will illustrate how different bacterial and viral pathogens hijack these endocytic routes to gain access to the cell.

Clathrin-mediated endocytosis

The most common mechanism for viral entry is clathrin-mediated endocytosis. Here, the virus first binds to surface exposed receptors such as Integrins (Integrin αvβ3 and αvβ5) for Adenovirus 2 (*89*) or LDL-receptor for Rhinovirus (*90*) with the help of specific ligands exposed on the viral surface. Viruses can thus only bind to cells that express the corresponding ligands, which determine the tropism of a virus for certain cells. Binding leads to clustering of the receptors which induce intracellular signaling. A clathrin coated pit forms around the virus and leads to its uptake. Following CME, the coat falls off and the virus is enclosed in an endocytic vesicle. Depending on the size of the virus, this process requires actin remodeling to complete the internalization process (*91*).

Zipper mechanism

Bacteria also make use of clathrin for their uptake even though they are far larger than the normal size of a clathrin-coated vesicle of about 100 nm (*92*). Clathrin was shown to be involved in early entry steps of bacteria that enter by a so called zipper mechanism (*93*). To activate this process, bacterial surface structures interact with plasma membrane receptors on the surface of a host cell and induce signaling events that result in the endocytosis of the bacterium. The best studied organism using this type of invasion is *Listeria monocytogenes.* This gram-positive bacterium

expresses eight internalin proteins, of which Internalin A (InlA) and Internalin B (InlB) are well characterized and involved in bacterial entry. InlA binds the cell-cell adhesion molecule E-cadherin (*94*) and InlB is a ligand for c-Met, the hepatocyte growth factor receptor (*95*). Binding of either receptor leads to activation of downstream signaling which, in turn, activates Rho GTPases and Arp2/3-dependent actin remodeling (*96*). A similar mechanism is used by *Yersinia pseudotuberculosis* which expresses the outer-membrane protein Invasin to interact with β1-integrins (*97*). Integrins usually bind to extracellular matrix components like fibronectin and are implicated in cell adherence (*98*). Invasin induces integrin clustering and downstream signaling which activates the Arp2/3 complex leading to actin rearrangement and bacterial uptake (*99*).

Trigger mechanism

While, in the zipper mechanism, the plasma membrane wraps around the bacteria until they are internalized, the so called trigger mechanism induces large, actin-rich membrane ruffles on the cell surface. These resemble macropinosomes and can be induced by different viral and bacterial pathogens. One example is *Salmonella* which is able to utilize this mode of entry via the secretion of T3SS effectors. *Salmonella* first adhere to the cell surface with the fimbrial adhesin FimH and with the SPI-1 T3SS (*100*). SPI-1 is then used to translocate a cocktail of effectors that aim at remodeling the actin cytoskeleton. SopE and SopE2 are guanine nucleotide exchange factors that activate Cdc42 and Rac1 (*101*). SipC is part of the T3SS translocon required for effector translocation and directly involved in nucleation and bundling of actin (*102*). Formed actin fibers are then prevented from disassembly by the effector SipA that inhibits the depolymerizing factors cofilin and the severing factor gelsolin (*103*). The synergistic action of these effectors efficiently induces the uptake of bacteria into host cells.

Invasome

A very intriguing mode of entry is employed by the zoonotic pathogen *Bartonella henselae* which causes cat scratch disease or angiomatosis-peliosis in humans (*104*). In addition to classical invasion of single bacteria into cells, *Bartonella henselae* can also enter endothelial cells as a large aggregate of bacteria known as the invasome (*105*). This process requires the action of T4SS secreted effectors and is accompanied by massive actin rearrangement surrounding the bacterial aggregate (*106, 107*). Both effector secretion and invasome formation are dependent on Integrin β1 signaling (*108*).

1.3 RNA INTERFERENCE

1.3.1 Mechanisms of RNA interference

RNA interference (RNAi) is an RNA-mediated gene silencing mechanism. Different types of RNAi have been discovered and the exact mode of action depends on the organism and auxiliary proteins involved in the process. Despite these differences, each mechanism retains the same critical feature in which a long double stranded RNA (dsRNA) , either synthesized by the cell or introduced from the outside, is processed into shorter dsRNA by endoribonuclease III enzymes (RNase III). Next, the RNA duplexes are unwound and one of the two strands (known as the guide strand) is bound to the RNA-induced silencing complex (RISC). This complex scans the transcriptome to find potential targets. The guide strand determines the target specificity of the RNAi response by complete or partial base paring to the target transcript which is then silenced by RISC.

In the following I will focus on the biogenesis and action of two classes of RNAi in mammalian cells (Figure 5), small interfering RNAs (siRNAs) and microRNAs (miRNA). While each has a distinct origin, and they vary to some extent in their function, it is increasingly evident that the boundaries between these pathways are blurring.

MicroRNAs are transcribed by RNA polymerase II and thus part of a cell's repertoire of translational regulators. Most miRNAs are encoded by introns of non-coding or coding transcripts but miRNAs that originate from exonic regions have also been described (*109*). Often, several miRNAs are encoded in close proximity and processed from a long polycistronic transcript (*110*). The structure of the primary miRNA (pri-miRNA) contains one or several stem-loops, where the RNA folds back on itself to form specific hairpins. Pri-miRNAs are processed in the nucleus by Drosha, an RNase III-type endonuclease similar to Dicer, by cropping the 5' and 3' ends below the stem loop (*111*). Drosha contains two RNase III domains that cleave one arm at the stem leaving a two nucleotide-long 3' overhang (*112, 113*). The roughly 70 nucleotide-long precursor RNA (pre-miRNA) is then exported from the nucleus to the cytoplasm with the help of exportin 5 (*114*). In the cytoplasm, Dicer removes the loop from the pre-miRNA thus generating a mature double-stranded miRNA approximately 20 to 25 nucleotides in length (*115, 116*). miRNAs often contain unpaired regions unlike siRNAs that are fully complementary aside from the 5' overhangs.

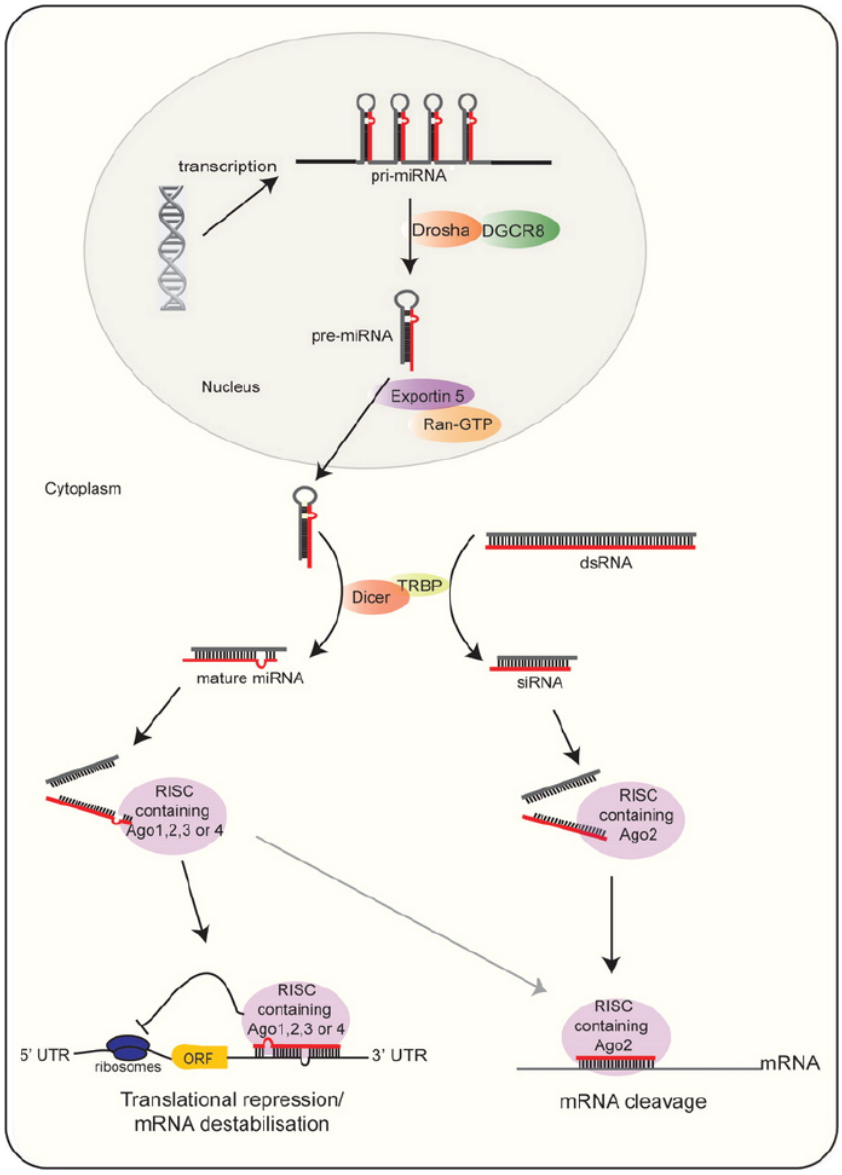


Figure 5. RNA interference pathways. Endogenous primary microRNAs (pri-miRNAs) are transcribed in the nucleus and form stem-loop structures by folding back on themselves. Processing by the endonuclease Drosha generates ~70 nt long precursor miRNAs (pre-miRNAs) which are exported to the cytoplasm by Exportin 5 and Ran-GTP and further trimmed by Dicer to the mature miRNA. These contain 3' overhangs and often unpaired nucleotides. Dicer also produces small interfering RNAs (siRNAs) from long dsRNAs. One strand (guide) of the miRNA or siRNA is assembled into the RNA-induced silencing complex (RISC) and directs the complex to target mRNAs that show full or partial complementarity to the guide strand. RISC binding causes translational inhibition or mRNA destabilization if bound by partial complementarity. Full complementarity leads to cleavage of the target mRNA. The major component of RISC is a member of the Argonaut (Ago) family of proteins of which human cells encode four. All Ago members can cause translational repression and mRNA destabilization, while only Ago2 has endonuclease activity which is required for direct cleavage of the target mRNA. Picture taken from (*117*).

To become active, the guide strand of the mature miRNA must be loaded into an Argonaut (AGO) protein part of the RISC complex, while the other strand is degraded. In humans, four AGO proteins can incorporate siRNAs or miRNAs with no strict sorting system (*118*). While only AGO2 is able to cleave the target transcript (*119*), all AGO proteins are able to induce translational inhibition and messenger RNA (mRNA) decay (*120*). The selection of the guide strand is primarily based on the differential thermodynamic stability at the ends of the miRNA. The strand with the weaker stability at its 5' end is generally incorporated into RISC (*121, 122*). The other strand, known as passenger strand, is then quickly removed and degraded. AGO2 can cleave the passenger strand if the miRNA is matched in the center. However, this is not the general case for miRNAs since most contain central mismatches and human AGO1, AGO3, and AGO4 do not possess slicer activity (*119*). Duplexes generally unwind without cleavage and mismatches between both strands in positions 2-8 and 12-15 promote this process (*123*).

The mature RISC complex, loaded with the guide miRNA, will then scan the cell's transcriptome for targets sharing some degree of complementarity to the guide RNA. The established model for target-miRNA interaction states that miRNAs mainly act by partial binding of the miRNA to the 3' untranslated region (UTR) of target mRNAs by Watson-Crick base paring. The binding specificities of an miRNA is largely determined by complementarity of the so called seed region to the target mRNA (*124*). The seed region comprises a stretch of 6-8 nucleotides located in the 5' end of the miRNA. This model is based on the initial discovery that the *lin-4* miRNA binds to multiple conserved sites in the 3' UTR of its target *lin-14* (*125, 126*). Later on, crystal structures of Argonaut proteins bound to the guide RNA showed that the complex is in a conformation which is primed for base pairing in the seed region (*127*). Furthermore several structures indicate that the first nucleotide of the miRNA does not engage in target interaction (*128, 129*). Even though seed paring seems to be a common mechanism of target binding, many exceptions to this rule have been found. Some miRNAs compensate partial mismatches in seed pairing by compensatory sites in the 3' end of the miRNA. This has been discovered for miR-196 and its target *HOXB8* (*130*). In other cases central sequences, rather than the seed, have been found to dictate binding specificity (*131*). Aside from differences in target interactions, the idea that binding sites on the mRNA occur solely in the 3' UTR is over simplified. Certain miRNAs were found to bind to the 5' UTR or coding regions of mRNAs (*132, 133*).

Several other important features that affect target interactions have also been identified. For example, binding sites in unstructured regions on the mRNA, located in sequences rich in

adenosine und uridine bases (AU-rich) and outside the center of large 3' UTRs, are more readily accessed (*134*). In addition, the presence of RNA-binding proteins can largely affect the efficiency of miRNA interaction. Blocking of miRNA binding sites presents a mechanism for the cell to protect mRNAs from miRNA regulation during specific conditions (*135*).

The general effect of miRNA-guided RISC binding to a target mRNA is a net reduction of protein synthesis. Originally, it was reported that this is exclusively caused by translational repression while target mRNA levels were not affected (*136*). Later on, microarray analysis showed that miRNAs have a prominent effect on target mRNA abundance (*137*). Since both translational repression and mRNA degradation affect protein synthesis, the question of the relative contributions of these effects remains. While in the majority of studies mRNA degradation was found to account for most of the effect (*138*), there are some indications that translational repression without mRNA destabilization occurs early on when miRNA levels are rising (*139*). This indicates that miRNA binding initially reduces translation of the transcript, which is followed by destabilization and degradation of the mRNA (*140, 141*).

While a steadily increasing number of miRNAs are discovered thanks to high throughput sequencing techniques, the identification of the direct targets is more challenging. Methods such as quantitative proteomics or transcriptomics can be used to obtain a global picture of the changes which a cell undergoes upon perturbation of a certain miRNA. Overexpression of an miRNA which is normally not expressed in a specific cell type was found to alter the transcriptome of this cell in the direction of cells which express this miRNA (*137*). This showed that the cellular changes induced by a miRNA can be identified by high throughput methods. However, it is not directly obvious which of the altered transcripts are directly targeted. In this case the use of bioinformatics prediction tools can help to narrow down the list of potential candidates always considering that non-canonical or non-conserved targets won't be recognized easily. Currently used prediction methods consider the seed sequence, free energy of the miRNA-mRNA interaction, local mRNA content and secondary structure, and often evolutionary conservation (*134, 142, 143*).

In contrast to miRNAs, siRNAs are generated from long dsRNAs only by the action of Dicer. Under natural conditions, these dsRNAs are introduced into the cell from the outside, e.g. by a virus. The siRNA is then introduced into the RISC complex and binding to a perfectly complementary target sequence induces RNA cleavage of the target by Argonaut 2 (Ago2). This

mode of antiviral defense is very prominent in many species, while it was only observed in a subset of human cells (*144*).

1.3.2 Use of RNA interference for system biology studies

Since the discovery of RNA interference for which Craig Mellow and Andrew Fire received the Nobel Prize in 2006, this technology has revolutionized genetic manipulation of many organisms. In comparisons to simpler organisms, loss-of-function studies in cultured mammalian cells were extremely difficult. This has changed with the discovery that small double stranded RNA molecules introduced into cells will use the cell's silencing machinery to deplete target transcripts. Since many cell lines in the lab are easily transfected, this technology provides the tool to systematically target a large number of genes. siRNAs can be synthesized to match any target sequence which guides the siRNA loaded RISC complex to the corresponding RNA to induce degradation.

Early on it was realized that siRNAs also induce unwanted off-target effects. Some siRNAs were found to trigger an immune response in a cell type specific manner. It was found that Toll-like receptors 7 (TRL7) and 8 (TLR8) were mainly responsible for this phenomenon. Certain sequence patterns including uridine- and guanosine-rich sequences were preferentially recognized (*145, 146*). Avoiding these sequence patterns in siRNAs enabled the design of molecules with fewer immunogenic properties (*147*) even though a certain level of sequence independent TLR-response seems to be inherent to non-modified dsRNA molecules (*148*). However, these considerations are mainly of importance when working with immune cells or for therapeutic applications.

Of more concern to the field was the discovery of sequence specific off-target effects. Microarray analysis of different siRNAs targeting the same gene showed very distinct effects on the transcriptome. Sequence analysis of off-target transcripts revealed that several 3' UTRs showed complementarity to the 5' end of the guide siRNAs. This indicated that a mechanism of action reminiscent of miRNAs causes the off-target effects (*149*). These studies were expanded later and it was found that mutations in the seed sequence of an siRNA could reverse the effect on the original off-target transcripts and, instead, target a new set of mRNAs (*150, 151*). While such off-target effects can be addresses in small scale studies where individual genes are tested in detail, they are a more serious concern for large scale studies such as genome-wide siRNA screens.

Screens are often found to show a high false positive discovery rate and the overlap between comparable siRNAs screens can be extremely low (*152*). A meta-analysis of three genome-wide HIV screens showed that the overlap between the identified host factors is minimal (*153*). Among the combined list of 842 genes that reduce HIV replication in all three screens, only 34 genes were found in at least two siRNA screens and only three were found in all screens. Apart from differences in the cell lines or the details of the screens, it is likely that miRNA-like off-target effects account for a large part of this observation. A comparison of the enriched pathways that were found to contribute to HIV infection of the respective screens showed much higher agreement. This clearly shows that there is on-target information in these screens which enables identification of relevant host functions involved in HIV infection. However, on the level of individual genes, the false positive and false negative discovery rates are high.

To tackle the problem of miRNA-like off-target effects, strategies aiming at optimized RNAi reagents have been developed. siRNA reagents have been modified on several positions to improve the on-target to off-target specificity or to avoid the incorporation of the passenger strand into the RISC complex (*154, 155*). Other strategies include the pooling of several different siRNAs against the same target. While the different siRNA in the pool have an individual off-target spectrum, all are able to bind to the intended gene. This should reduce off-targets by competition between the different sequences without affecting the knock down of the intended target (*156*). Such competition seems to be important, since simple dilution of the siRNA is not sufficient to selectively reduce off-target effects (*149, 150*).

While advanced RNAi reagents can improve future experiments, several methods have been developed that help improve the analysis of large scale siRNA screens. RNAi screens that were performed with several siRNAs per gene were classically analyzed by relatively simple rules. A gene was considered a hit if a certain fraction of all siRNAs for this particular target exceeded a certain threshold in a given readout (e.g. 2 out of 3, 3 out of 4, etc. rules) (*157*). The drawback of such methods is that it requires a cutoff and genes with effects consistently below this threshold will not be considered. Alternative approaches use statistical methods to rank siRNAs and genes and provide p-values that reflect the statistical probability. An example is the commonly used Redundant siRNA Activity (RSA) method (*158*). Other methods aim at identifying seed-driven off-target effects in siRNAs screens (*159*). This information can then be used to remove affected siRNAs from the analysis or to correct for the off-target effect (*160*). Excluding untrustworthy siRNAs reduces the number of false positives, while correction of the data potentially helps

recover some of the false negatives. There have also been attempts to even go one step further and use the information of the off-target effects to infer which genes are involved in the biology of the studied process (*161, 162*). This relies on the prediction of the targets for the active seed sequences and a mathematical model that determines which of those are most likely to contribute to the phenotype.

Aside from its use in improving the rate of true positive hit genes, seed analysis of siRNA screens directly allows the identification of natural miRNAs or synthetic miRNA-like molecules involved in a certain process (*152, 163*). This is based on the observation that siRNAs that share the seed sequence with a natural miRNA elicit similar biological functions in a cell as their endogenous counterparts.

1.4 THE HUMAN PATHOGEN BRUCELLA

1.4.1 *Brucella* species and brucellosis

Brucella species are gram-negative, facultative intracellular pathogens that belong to the class of α-Proteobacteria. They were first isolated in Malta in 1887 by Sir David Bruce from the spleens of soldiers that died from brucellosis (*164*). The natural hosts of *Brucella* include a wide range of wild and domestic animals. *Brucella* causes abortions and infertility in their natural hosts such as cattle, goats, or sheep, resulting in severe economic losses in endemic areas. Human brucellosis is one of the most important zoonotic diseases worldwide with over half a million new cases annually (*165*). Transmission to human is most commonly caused by direct contact with infected animals or by ingestion of contaminated food such as unpasteurized milk, while human to human transmission is very rare. The most important species associated with human disease are *Brucella melitensi* and *Brucella abortus* that infect cattle, goats, or sheep, *Brucella suis* that infect pigs, and to a minor extent *Brucella canis* infecting dogs. Other *Brucella* species have been reported to infect humans, but these events are uncommon, most likely due to the limited contact between humans and the natural host species carrying these strains (*166*).

Due to the direct connection of animal and human brucellosis, the most effective intervention to prevent the disease is the control of brucellosis in domestic animals. Even though an effective live vaccine is available for animals, animal brucellosis is still abundant in many regions of the

world being endemic in many parts of the Middle East, Africa, Latin America, central Asian, and several countries around the Mediterranean basin (*167*).

Human brucellosis begins as a flu-like disease with symptoms such as fever and body pain. Due to the relatively unspecific symptoms, brucellosis is often not correctly diagnosed and treated. This allows bacteria do disseminate throughout the body where they can infect diverse tissues. At this point, the disease can become chronic and is difficult to treat. A common complication of brucellosis is arthritis, which is often seriously debilitating. More life threatening complications include liver abscess formation, endocarditis, or neuropathy. Antibiotic treatment is normally effective if started early during infection. However, a combination therapy of at least two antibiotics over a prolonged period is required to prevent complications or relapses. Due the high risk of serious disease and the ability of *Brucella* to be spread by aerosols *Brucella* species are classified as biosafety level 3 organisms (*168*).

1.4.2 Intracellular lifestyle of *Brucella*

Adhesion and entry into host cells

On the cellular level, *Brucella* is able to invade phagocytic and non-phagocytic cells (Figure 6). Depending on the cell type, different host-pathogen interactions are involved in adhesion of bacteria to the cell surface or entry into the host cell. Lipid rafts are required for the uptake of non-opsonized bacteria in macrophages (*169, 170*). Two receptors have been proposed to be involved in this process; the class A scavenger receptor (SR-A) which interacts with Brucella *lipopolysaccharide* (LPS) (*171*) and the cellular prion receptor PrPc which was proposed to bind to Hsp60 (*172*). However, the latter interaction is debated since the results could not be confirmed by an independent study (*173*).

Receptors containing sialic acid residues are important for adhesion of *Brucella* to phagocytic and non-phagocytic cells by interactions with the surface protein 41 (SP41) (*174, 175*). Furthermore, other *Brucella* genes have been found to be involved in binding to non-phagocytic cells. These include a gene cluster that harbors a bacterial immunoglobulin-like protein (*176*) and different autotransporters (*177, 178*). The autotransporters showed localization to the new pole of *Brucella* formed after cell division (*178*), which is in agreement with the finding that most bacteria that infect cells are G1-arrested newborn bacteria (*179*).

In phagocytic and non-phagocytic cells bacterial internalization involves actin rearrangements, which requires phosphatidylinositide 3-kinase PI3K, the small GTPases Rac, Rho, and direct activation of Cdc42 (*180*). In murine trophoblast giant cells, surface located Hsc70 and ezrin, a factor which tethers actin filaments to the plasma membrane, are required for *Brucella* uptake (*181, 182*). Overall, there is no indication that *Brucella* actively induces its own uptake via secretion of effectors. In line with this, the T4SS is only induced after the internalization of bacteria (*82*).

Intracellular trafficking

Inside the eukaryotic host cell, *Brucella* remains enclosed within a membrane, known as the *Brucella* containing vacuole (BCV). Directly after entry the BCV is enriched in cholesterol and the lipid raft-associated signaling molecule flotilin-1 (*183*). Bacteria interact with markers of early endosomes such as Rab5 or EEA1 (*83, 184, 185*). The *Brucella* secreted cyclic β-1,2-glucan is involved in inhibition of phagosome-lysosome fusion and proposed to act by extracting cholesterol from the BCV (*183, 186*). The BCV then traffics along the endocytic pathway, acquiring Rab-7, the Rab7 effector RILP, and Lamp-1 (*81*). Despite a certain degree of interaction with lysosomes, bacteria manage to avoid degradation (*81*). Since no enrichment of luminal lysosomal enzymes such as cathepsin D can be found in BCV, Brucella seems to be able to avoid full lysosomal fusion (*83, 187, 188*). Nevertheless, the BCV undergoes acidification during trafficking along the endocytic pathway. This is dependent on the action of the vesicular ATPase and is required to induce the expression of the bacterial T4SS (*82*). It is believed that secretion of effectors is essential for *Brucella* to establish its replicative niche, since deletion of the T4SS (*83*) or inhibition of the vesicular ATPase lead to defects in the establishment of the intracellular niche (*189*).

During later stages of trafficking, several factors involved in ER to Golgi bidirectional trafficking are required for *Brucella* to successfully reach its replicative niche. Bacteria are found to interact with Sar1 and COPII complex components and inhibition of Sar1 activity blocked intracellular replication (*190*). Sar1 initiates budding of vesicles from the ER that are destined for the Golgi apparatus. It is thus conceivable that *Brucella* intercepts vesicles at ER exit site.

Furthermore, components involved in retrograde trafficking have been found to be required for intracellular replication. The small GTPase Rab2, which is required for the maturation of the ER-Golgi intermediate compartment (ERGIC) and trafficking from the Golgi to the ER, was found to

localize to the BCV (*191*). This study could further identify several factors (GAPDH, COPI, and Protein kinase C) that interact with Rab2 and regulate retrograde trafficking to be required for infection. In agreement with the notion that *Brucella* hijacks ER-Golgi trafficking, the secreted protein RicA was found to interact with the GDP-bound form of Rab2. It remains unclear how this interaction affects intracellular trafficking of the BCV. However, it is intriguing that a mutant of *ricA* showed enhanced intracellular proliferation giving good indications that this protein is indeed involved in the regulation of trafficking (*192*). It was proposed that interaction of RicA with Rab2 could promote the silent or stealthy infection strategy of *Brucella* by preventing excessive proliferation.

Replication and intercellular spread

During the phase of trafficking and until *Brucella* establishes an ER-like replicative niche, bacteria do not divide (*179*). Once proliferation is initiated, the BCV is found in close association with ER markers such as calnexin and glucose-6-phosphatase (*83*). Even though *Brucella* replicates to very high numbers inside cells, basic cellular processes are not affected and cells can still undergo cell division (*185*). Similar to other stealthy pathogens, *Brucella* is able to inhibit apoptosis of infected cells. The exact mechanisms are unclear, but it was shown that the anti-apoptotic factor BCL2A1 is up-regulated (*193*) and mitochondrial genes involved in apoptosis are repressed in infected cells (*194*).

Cell-to-cell spread has only recently been addressed where autophagy factors were found to be involved. Interestingly, only autophagy initiation factors such as ULK1, Beclin1, ATG14L, and PI3K but not elongation factors are required for this process. Furthermore, it was observed that some BCVs acquire Rab7 during late stages of infection which coincides with the potential of *Brucella* to spread to neighboring cells. It was found that cell-to-cell spreading can be blocked by the cell-impermeable antibiotic gentamycin indicating an extracellular stage during the spread. However, the exact mechanism of egress remains elusive (*195*).

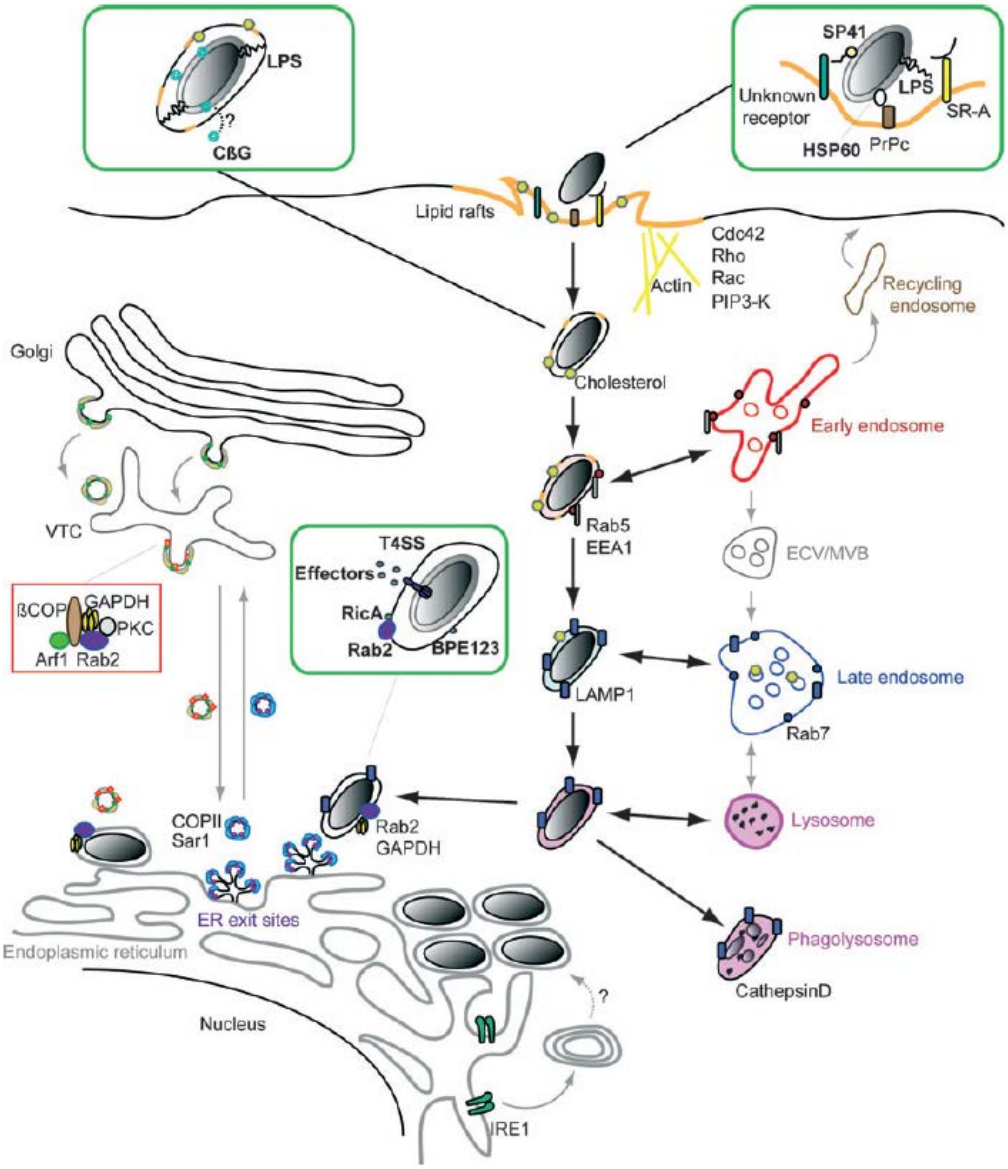


Figure 6. *Brucella* intracellular life. Scheme depicting the entry, trafficking, and intracellular replication of *Brucella*. Green zooms indicated bacterial factors involved in the corresponding step. Bacteria enter host cells via lipid rafts (orange) enriched in cholesterol (yellow dots). Cyclic β-1,2-glucans (CβB) are thought to deplete the *Brucella* containing vacuole (BCV) from cholesterol. The BCV traffics along the endocytic pathway interacting with early and late endosomes. Acidification of the BCV leads to the expression of the type IV secretion system (T4SS) which secretes effectors such as BPE123 or RicA which interacts with Rab2. Avirulent mutants (e.g. carrying deletions in the *virB* T4SS) are degraded in phagolysosomes. The BCV interacts with components of ER exit sites (Sar1, COPII) as well as retrograde trafficking factors (Arf1, Rab2, PKC, GAPDH, βCOP) which constitute COPI coated vesicles that mediate vesicular transport from the vesicular tubular clusters (VTC) to the ER. Bacteria also interact with IRE1 involved in autophagy and finally replicate in an ER-derived compartment. Picture taken from (*167*).

1.5 REFERENCES

1. J. S. Bonifacino, B. S. Glick, The mechanisms of vesicle budding and fusion. *Cell* **116**, 153 (Jan 23, 2004).
2. Y. Nyathi, B. M. Wilkinson, M. R. Pool, Co-translational targeting and translocation of proteins to the endoplasmic reticulum. *Biochimica et biophysica acta* **1833**, 2392 (Nov, 2013).
3. F. Cymer, G. von Heijne, S. H. White, Mechanisms of Integral Membrane Protein Insertion and Folding. *J Mol Biol*, (Sep 30, 2014).
4. O. Baumann, B. Walz, Endoplasmic reticulum of animal cells and its organization into structural and functional domains. *International review of cytology* **205**, 149 (2001).
5. J. Klumperman, Architecture of the mammalian Golgi. *Cold Spring Harb Perspect Biol* **3**, (Jul, 2011).
6. P. Stanley, Golgi glycosylation. *Cold Spring Harb Perspect Biol* **3**, (Apr, 2011).
7. J. Huotari, A. Helenius, Endosome maturation. *Embo J* **30**, 3481 (Aug 31, 2011).
8. P. Malkus, F. Jiang, R. Schekman, Concentrative sorting of secretory cargo proteins into COPII-coated vesicles. *J Cell Biol* **159**, 915 (Dec 23, 2002).
9. C. Barlowe *et al.*, COPII: a membrane coat formed by Sec proteins that drive vesicle budding from the endoplasmic reticulum. *Cell* **77**, 895 (Jun 17, 1994).
10. T. Oka, A. Nakano, Inhibition of GTP hydrolysis by Sar1p causes accumulation of vesicles that are a functional intermediate of the ER-to-Golgi transport in yeast. *J Cell Biol* **124**, 425 (Feb, 1994).
11. R. B. Sutton, D. Fasshauer, R. Jahn, A. T. Brunger, Crystal structure of a SNARE complex involved in synaptic exocytosis at 2.4 A resolution. *Nature* **395**, 347 (Sep 24, 1998).
12. S. Deborde *et al.*, Clathrin is a key regulator of basolateral polarity. *Nature* **452**, 719 (Apr 10, 2008).
13. S. Hara-Kuge *et al.*, En bloc incorporation of coatomer subunits during the assembly of COP-coated vesicles. *J Cell Biol* **124**, 883 (Mar, 1994).
14. P. Cosson, F. Letourneur, Coatomer interaction with di-lysine endoplasmic reticulum retention motifs. *Science* **263**, 1629 (Mar 18, 1994).
15. J. Rink, E. Ghigo, Y. Kalaidzidis, M. Zerial, Rab conversion as a mechanism of progression from early to late endosomes. *Cell* **122**, 735 (Sep 9, 2005).
16. V. Marshansky, M. Futai, The V-type H+-ATPase in vesicular trafficking: targeting, regulation and function. *Current opinion in cell biology* **20**, 415 (Aug, 2008).
17. I. Roxrud, H. Stenmark, L. Malerod, ESCRT & Co. *Biol Cell* **102**, 293 (May, 2010).
18. J. D. Bleil, M. S. Bretscher, Transferrin receptor and its recycling in HeLa cells. *Embo J* **1**, 351 (1982).
19. T. Galvez, J. Gilleron, M. Zerial, G. A. O'Sullivan, SnapShot: Mammalian Rab proteins in endocytic trafficking. *Cell* **151**, 234 (Sep 28, 2012).
20. T. Wassmer *et al.*, The retromer coat complex coordinates endosomal sorting and dynein-mediated transport, with carrier recognition by the trans-Golgi network. *Developmental cell* **17**, 110 (Jul, 2009).
21. S. F. Nothwehr, S. A. Ha, P. Bruinsma, Sorting of yeast membrane proteins into an endosome-to-Golgi pathway involves direct interaction of their cytosolic domains with Vps35p. *J Cell Biol* **151**, 297 (Oct 16, 2000).

22. C. N. Arighi, L. M. Hartnell, R. C. Aguilar, C. R. Haft, J. S. Bonifacino, Role of the mammalian retromer in sorting of the cation-independent mannose 6-phosphate receptor. *J Cell Biol* **165**, 123 (Apr, 2004).
23. C. Burd, P. J. Cullen, Retromer: a master conductor of endosome sorting. *Cold Spring Harb Perspect Biol* **6**, (Feb, 2014).
24. B. M. Pearse, Clathrin: a unique protein associated with intracellular transfer of membrane by coated vesicles. *Proc Natl Acad Sci U S A* **73**, 1255 (Apr, 1976).
25. C. R. Hopkins, K. Miller, J. M. Beardmore, Receptor-mediated endocytosis of transferrin and epidermal growth factor receptors: a comparison of constitutive and ligand-induced uptake. *J Cell Sci Suppl* **3**, 173 (1985).
26. W. M. Henne *et al.*, FCHo proteins are nucleators of clathrin-mediated endocytosis. *Science* **328**, 1281 (Jun 4, 2010).
27. H. E. Stimpson, C. P. Toret, A. T. Cheng, B. S. Pauly, D. G. Drubin, Early-arriving Syp1p and Ede1p function in endocytic site placement and formation in budding yeast. *Molecular biology of the cell* **20**, 4640 (Nov, 2009).
28. B. T. Kelly *et al.*, A structural explanation for the binding of endocytic dileucine motifs by the AP2 complex. *Nature* **456**, 976 (Dec 18, 2008).
29. B. M. Collins, A. J. McCoy, H. M. Kent, P. R. Evans, D. J. Owen, Molecular architecture and functional model of the endocytic AP2 complex. *Cell* **109**, 523 (May 17, 2002).
30. F. Tebar, T. Sorkina, A. Sorkin, M. Ericsson, T. Kirchhausen, Eps15 is a component of clathrin-coated pits and vesicles and is located at the rim of coated pits. *J Biol Chem* **271**, 28727 (Nov 15, 1996).
31. T. Kosaka, K. Ikeda, Reversible blockage of membrane retrieval and endocytosis in the garland cell of the temperature-sensitive mutant of Drosophila melanogaster, shibirets1. *J Cell Biol* **97**, 499 (Aug, 1983).
32. S. M. Sweitzer, J. E. Hinshaw, Dynamin undergoes a GTP-dependent conformational change causing vesiculation. *Cell* **93**, 1021 (Jun 12, 1998).
33. A. Roux, K. Uyhazi, A. Frost, P. De Camilli, GTP-dependent twisting of dynamin implicates constriction and tension in membrane fission. *Nature* **441**, 528 (May 25, 2006).
34. D. M. Schlossman, S. L. Schmid, W. A. Braell, J. E. Rothman, An enzyme that removes clathrin coats: purification of an uncoating ATPase. *J Cell Biol* **99**, 723 (Aug, 1984).
35. E. Ungewickell *et al.*, Role of auxilin in uncoating clathrin-coated vesicles. *Nature* **378**, 632 (Dec 7, 1995).
36. H. T. Haigler, J. A. McKanna, S. Cohen, Rapid stimulation of pinocytosis in human carcinoma cells A-431 by epidermal growth factor. *J Cell Biol* **83**, 82 (Oct, 1979).
37. C. C. Norbury, L. J. Hewlett, A. R. Prescott, N. Shastri, C. Watts, Class I MHC presentation of exogenous soluble antigen via macropinocytosis in bone marrow macrophages. *Immunity* **3**, 783 (Dec, 1995).
38. F. Sallusto, M. Cella, C. Danieli, A. Lanzavecchia, Dendritic cells use macropinocytosis and the mannose receptor to concentrate macromolecules in the major histocompatibility complex class II compartment: downregulation by cytokines and bacterial products. *J Exp Med* **182**, 389 (Aug 1, 1995).
39. L. J. Hewlett, A. R. Prescott, C. Watts, The coated pit and macropinocytic pathways serve distinct endosome populations. *J Cell Biol* **124**, 689 (Mar, 1994).
40. M. J. Clague, C. Thorpe, A. T. Jones, Phosphatidylinositol 3-kinase regulation of fluid phase endocytosis. *FEBS Lett* **367**, 272 (Jul 3, 1995).

41. S. Sabharanjak, P. Sharma, R. G. Parton, S. Mayor, GPI-anchored proteins are delivered to recycling endosomes via a distinct cdc42-regulated, clathrin-independent pinocytic pathway. *Developmental cell* **2**, 411 (Apr, 2002).
42. H. Damke, T. Baba, A. M. van der Bliek, S. L. Schmid, Clathrin-independent pinocytosis is induced in cells overexpressing a temperature-sensitive mutant of dynamin. *J Cell Biol* **131**, 69 (Oct, 1995).
43. A. Guha, V. Sriram, K. S. Krishnan, S. Mayor, Shibire mutations reveal distinct dynamin-independent and -dependent endocytic pathways in primary cultures of Drosophila hemocytes. *J Cell Sci* **116**, 3373 (Aug 15, 2003).
44. S. Mayor, R. E. Pagano, Pathways of clathrin-independent endocytosis. *Nature reviews. Molecular cell biology* **8**, 603 (Aug, 2007).
45. M. van Lookeren Campagne, C. Wiesmann, E. J. Brown, Macrophage complement receptors and pathogen clearance. *Cell Microbiol* **9**, 2095 (Sep, 2007).
46. D. H. Jones, J. Nusbacher, C. L. Anderson, Fc receptor-mediated binding and endocytosis by human mononuclear phagocytes: monomeric IgG is not endocytosed by U937 cells and monocytes. *J Cell Biol* **100**, 558 (Feb, 1985).
47. A. D. Hoppe, J. A. Swanson, Cdc42, Rac1, and Rac2 display distinct patterns of activation during phagocytosis. *Molecular biology of the cell* **15**, 3509 (Aug, 2004).
48. R. Levin, S. Grinstein, D. Schlam, Phosphoinositides in phagocytosis and macropinocytosis. *Biochimica et biophysica acta*, (Sep 16, 2014).
49. R. G. Parton, K. Simons, The multiple faces of caveolae. *Nature reviews. Molecular cell biology* **8**, 185 (Mar, 2007).
50. M. Drab *et al.*, Loss of caveolae, vascular dysfunction, and pulmonary defects in caveolin-1 gene-disrupted mice. *Science* **293**, 2449 (Sep 28, 2001).
51. P. J. Walser *et al.*, Constitutive formation of caveolae in a bacterium. *Cell* **150**, 752 (Aug 17, 2012).
52. M. M. Hill *et al.*, PTRF-Cavin, a conserved cytoplasmic protein required for caveola formation and function. *Cell* **132**, 113 (Jan 11, 2008).
53. J. R. Henley, E. W. Krueger, B. J. Oswald, M. A. McNiven, Dynamin-mediated internalization of caveolae. *J Cell Biol* **141**, 85 (Apr 6, 1998).
54. R. G. Parton, M. A. del Pozo, Caveolae as plasma membrane sensors, protectors and organizers. *Nature reviews. Molecular cell biology* **14**, 98 (Feb, 2013).
55. L. V. Valen, A new evolutionary law. *Evolutionary Theory* **1**, 1 (1973).
56. T. L. Hale, G. T. Keusch, in *Medical Microbiology*, S. Baron, Ed. (Galveston (TX), 1996).
57. M. Boyd, K. Clezy, R. Lindley, R. Pearce, Pandemic influenza: clinical issues. *Med J Aust* **185**, S44 (Nov 20, 2006).
58. D. N. McMurray, in *Medical Microbiology*, S. Baron, Ed. (Galveston (TX), 1996).
59. R. J. Whitley, in *Medical Microbiology*, S. Baron, Ed. (Galveston (TX), 1996).
60. V. D. Vissa, P. J. Brennan, The genome of Mycobacterium leprae: a minimal mycobacterial gene set. *Genome Biol* **2**, REVIEWS1023 (2001).
61. S. G. Andersson, C. G. Kurland, Reductive evolution of resident genomes. *Trends Microbiol* **6**, 263 (Jul, 1998).
62. P. Cossart, A. Helenius, Endocytosis of viruses and bacteria. *Cold Spring Harb Perspect Biol* **6**, (Aug, 2014).
63. P. J. Sansonetti, A. Ryter, P. Clerc, A. T. Maurelli, J. Mounier, Multiplication of Shigella flexneri within HeLa cells: lysis of the phagocytic vacuole and plasmid-mediated contact hemolysis. *Infect Immun* **51**, 461 (Feb, 1986).

64. J. T. Myers, A. W. Tsang, J. A. Swanson, Localized reactive oxygen and nitrogen intermediates inhibit escape of Listeria monocytogenes from vacuoles in activated macrophages. *J Immunol* **171**, 5447 (Nov 15, 2003).
65. D. A. Portnoy, P. S. Jacks, D. J. Hinrichs, Role of hemolysin for the intracellular growth of Listeria monocytogenes. *J Exp Med* **167**, 1459 (Apr 1, 1988).
66. G. A. Smith *et al.*, The two distinct phospholipases C of Listeria monocytogenes have overlapping roles in escape from a vacuole and cell-to-cell spread. *Infect Immun* **63**, 4231 (Nov, 1995).
67. I. J. Glomski, M. M. Gedde, A. W. Tsang, J. A. Swanson, D. A. Portnoy, The Listeria monocytogenes hemolysin has an acidic pH optimum to compartmentalize activity and prevent damage to infected host cells. *J Cell Biol* **156**, 1029 (Mar 18, 2002).
68. J. M. Stevens, E. E. Galyov, M. P. Stevens, Actin-dependent movement of bacterial pathogens. *Nat Rev Microbiol* **4**, 91 (Feb, 2006).
69. A. J. Perrin, X. Jiang, C. L. Birmingham, N. S. So, J. H. Brumell, Recognition of bacteria in the cytosol of Mammalian cells by the ubiquitin system. *Curr Biol* **14**, 806 (May 4, 2004).
70. P. S. Hiemstra, M. T. van den Barselaar, M. Roest, P. H. Nibbering, R. van Furth, Ubiquicidin, a novel murine microbicidal protein present in the cytosolic fraction of macrophages. *J Leukoc Biol* **66**, 423 (Sep, 1999).
71. L. M. Delbridge, M. X. O'Riordan, Innate recognition of intracellular bacteria. *Curr Opin Immunol* **19**, 10 (Feb, 2007).
72. Y. Koizumi *et al.*, Inflammasome activation via intracellular NLRs triggered by bacterial infection. *Cell Microbiol* **14**, 149 (Feb, 2012).
73. J. Bielecki, P. Youngman, P. Connelly, D. A. Portnoy, Bacillus subtilis expressing a haemolysin gene from Listeria monocytogenes can grow in mammalian cells. *Nature* **345**, 175 (May 10, 1990).
74. D. M. Monack, J. A. Theriot, Actin-based motility is sufficient for bacterial membrane protrusion formation and host cell uptake. *Cell Microbiol* **3**, 633 (Sep, 2001).
75. M. Goetz *et al.*, Microinjection and growth of bacteria in the cytosol of mammalian host cells. *Proc Natl Acad Sci U S A* **98**, 12221 (Oct 9, 2001).
76. I. Vergne *et al.*, Mechanism of phagolysosome biogenesis block by viable Mycobacterium tuberculosis. *Proc Natl Acad Sci U S A* **102**, 4033 (Mar 15, 2005).
77. M. A. Bakowski *et al.*, The phosphoinositide phosphatase SopB manipulates membrane surface charge and trafficking of the Salmonella-containing vacuole. *Cell Host Microbe* **7**, 453 (Jun 25, 2010).
78. L. D. Hernandez, K. Hueffer, M. R. Wenk, J. E. Galan, Salmonella modulates vesicular traffic by altering phosphoinositide metabolism. *Science* **304**, 1805 (Jun 18, 2004).
79. M. A. Stein, K. Y. Leung, M. Zwick, F. Garcia-del Portillo, B. B. Finlay, Identification of a Salmonella virulence gene required for formation of filamentous structures containing lysosomal membrane glycoproteins within epithelial cells. *Mol Microbiol* **20**, 151 (Apr, 1996).
80. C. R. Beuzon *et al.*, Salmonella maintains the integrity of its intracellular vacuole through the action of SifA. *Embo J* **19**, 3235 (Jul 3, 2000).
81. T. Starr, T. W. Ng, T. D. Wehrly, L. A. Knodler, J. Celli, Brucella intracellular replication requires trafficking through the late endosomal/lysosomal compartment. *Traffic* **9**, 678 (May, 2008).
82. M. L. Boschiroli *et al.*, The Brucella suis virB operon is induced intracellularly in macrophages. *Proc Natl Acad Sci U S A* **99**, 1544 (Feb 5, 2002).

83. J. Celli *et al.*, Brucella evades macrophage killing via VirB-dependent sustained interactions with the endoplasmic reticulum. *J Exp Med* **198**, 545 (Aug 18, 2003).
84. A. Hubber, C. R. Roy, Modulation of host cell function by Legionella pneumophila type IV effectors. *Annu Rev Cell Dev Biol* **26**, 261 (2010).
85. D. E. Voth, R. A. Heinzen, Lounging in a lysosome: the intracellular lifestyle of Coxiella burnetii. *Cell Microbiol* **9**, 829 (Apr, 2007).
86. D. Howe, L. P. Mallavia, Coxiella burnetii exhibits morphological change and delays phagolysosomal fusion after internalization by J774A.1 cells. *Infect Immun* **68**, 3815 (Jul, 2000).
87. W. Beron, M. G. Gutierrez, M. Rabinovitch, M. I. Colombo, Coxiella burnetii localizes in a Rab7-labeled compartment with autophagic characteristics. *Infect Immun* **70**, 5816 (Oct, 2002).
88. L. A. Cooley, S. R. Lewin, HIV-1 cell entry and advances in viral entry inhibitor therapy. *J Clin Virol* **26**, 121 (Feb, 2003).
89. T. J. Wickham, P. Mathias, D. A. Cheresh, G. R. Nemerow, Integrins alpha v beta 3 and alpha v beta 5 promote adenovirus internalization but not virus attachment. *Cell* **73**, 309 (Apr 23, 1993).
90. R. Fuchs, D. Blaas, Productive entry pathways of human rhinoviruses. *Adv Virol* **2012**, 826301 (2012).
91. D. K. Cureton, R. H. Massol, S. P. Whelan, T. Kirchhausen, The length of vesicular stomatitis virus particles dictates a need for actin assembly during clathrin-dependent endocytosis. *PLoS Pathog* **6**, e1001127 (2010).
92. H. T. McMahon, E. Boucrot, Molecular mechanism and physiological functions of clathrin-mediated endocytosis. *Nature reviews. Molecular cell biology* **12**, 517 (Aug, 2011).
93. E. Veiga, P. Cossart, Listeria hijacks the clathrin-dependent endocytic machinery to invade mammalian cells. *Nat Cell Biol* **7**, 894 (Sep, 2005).
94. J. Mengaud, H. Ohayon, P. Gounon, R. M. Mege, P. Cossart, E-cadherin is the receptor for internalin, a surface protein required for entry of L. monocytogenes into epithelial cells. *Cell* **84**, 923 (Mar 22, 1996).
95. Y. Shen, M. Naujokas, M. Park, K. Ireton, InIB-dependent internalization of Listeria is mediated by the Met receptor tyrosine kinase. *Cell* **103**, 501 (Oct 27, 2000).
96. P. Cossart, J. Pizarro-Cerda, M. Lecuit, Invasion of mammalian cells by Listeria monocytogenes: functional mimicry to subvert cellular functions. *Trends Cell Biol* **13**, 23 (Jan, 2003).
97. G. T. Van Nhieu, R. R. Isberg, The Yersinia pseudotuberculosis invasin protein and human fibronectin bind to mutually exclusive sites on the alpha 5 beta 1 integrin receptor. *J Biol Chem* **266**, 24367 (Dec 25, 1991).
98. D. S. Harburger, D. A. Calderwood, Integrin signalling at a glance. *J Cell Sci* **122**, 159 (Jan 15, 2009).
99. R. R. Isberg, Z. Hamburger, P. Dersch, Signaling and invasin-promoted uptake via integrin receptors. *Microbes Infect* **2**, 793 (Jun, 2000).
100. B. Misselwitz *et al.*, Salmonella enterica serovar Typhimurium binds to HeLa cells via Fim-mediated reversible adhesion and irreversible type three secretion system 1-mediated docking. *Infect Immun* **79**, 330 (Jan, 2011).
101. W. D. Hardt, L. M. Chen, K. E. Schuebel, X. R. Bustelo, J. E. Galan, S. typhimurium encodes an activator of Rho GTPases that induces membrane ruffling and nuclear responses in host cells. *Cell* **93**, 815 (May 29, 1998).

102. R. D. Hayward, V. Koronakis, Direct nucleation and bundling of actin by the SipC protein of invasive Salmonella. *Embo J* **18**, 4926 (Sep 15, 1999).
103. E. J. McGhie, R. D. Hayward, V. Koronakis, Control of actin turnover by a salmonella invasion protein. *Mol Cell* **13**, 497 (Feb 27, 2004).
104. C. Dehio, Molecular and cellular basis of bartonella pathogenesis. *Annu Rev Microbiol* **58**, 365 (2004).
105. C. Dehio, M. Meyer, J. Berger, H. Schwarz, C. 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* **110 (Pt 18)**, 2141 (Sep, 1997).
106. T. A. Rhomberg, M. C. Truttmann, P. Guye, Y. Ellner, C. Dehio, A translocated protein of Bartonella henselae interferes with endocytic uptake of individual bacteria and triggers uptake of large bacterial aggregates via the invasome. *Cell Microbiol* **11**, 927 (Jun, 2009).
107. M. C. Truttmann, T. A. Rhomberg, C. Dehio, Combined action of the type IV secretion effector proteins BepC and BepF promotes invasome formation of Bartonella henselae on endothelial and epithelial cells. *Cell Microbiol* **13**, 284 (Feb, 2011).
108. M. C. Truttmann *et al.*, Bartonella henselae engages inside-out and outside-in signaling by integrin beta1 and talin1 during invasome-mediated bacterial uptake. *J Cell Sci* **124**, 3591 (Nov 1, 2011).
109. J. Han *et al.*, Posttranscriptional crossregulation between Drosha and DGCR8. *Cell* **136**, 75 (Jan 9, 2009).
110. Y. Lee, K. Jeon, J. T. Lee, S. Kim, V. N. Kim, MicroRNA maturation: stepwise processing and subcellular localization. *Embo J* **21**, 4663 (Sep 2, 2002).
111. Y. Lee *et al.*, The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425**, 415 (Sep 25, 2003).
112. J. Han *et al.*, The Drosha-DGCR8 complex in primary microRNA processing. *Genes & development* **18**, 3016 (Dec 15, 2004).
113. H. Zhang, F. A. Kolb, L. Jaskiewicz, E. Westhof, W. Filipowicz, Single processing center models for human Dicer and bacterial RNase III. *Cell* **118**, 57 (Jul 9, 2004).
114. E. Lund, S. Guttinger, A. Calado, J. E. Dahlberg, U. Kutay, Nuclear export of microRNA precursors. *Science* **303**, 95 (Jan 2, 2004).
115. E. Bernstein, A. A. Caudy, S. M. Hammond, G. J. Hannon, Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**, 363 (Jan 18, 2001).
116. G. Hutvagner *et al.*, A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science* **293**, 834 (Aug 3, 2001).
117. A. Kelly, A. F. Hurlstone, The use of RNAi technologies for gene knockdown in zebrafish. *Brief Funct Genomics* **10**, 189 (Jul, 2011).
118. A. Dueck, C. Ziegler, A. Eichner, E. Berezikov, G. Meister, microRNAs associated with the different human Argonaute proteins. *Nucleic Acids Res* **40**, 9850 (Oct, 2012).
119. J. Liu *et al.*, Argonaute2 is the catalytic engine of mammalian RNAi. *Science* **305**, 1437 (Sep 3, 2004).
120. H. Su, M. I. Trombly, J. Chen, X. Wang, Essential and overlapping functions for mammalian Argonautes in microRNA silencing. *Genes & development* **23**, 304 (Feb 1, 2009).
121. A. Khvorova, A. Reynolds, S. D. Jayasena, Functional siRNAs and miRNAs exhibit strand bias. *Cell* **115**, 209 (Oct 17, 2003).
122. D. S. Schwarz *et al.*, Asymmetry in the assembly of the RNAi enzyme complex. *Cell* **115**, 199 (Oct 17, 2003).

123. T. Kawamata, H. Seitz, Y. Tomari, Structural determinants of miRNAs for RISC loading and slicer-independent unwinding. *Nat Struct Mol Biol* **16**, 953 (Sep, 2009).
124. B. P. Lewis, I. H. Shih, M. W. Jones-Rhoades, D. P. Bartel, C. B. Burge, Prediction of mammalian microRNA targets. *Cell* **115**, 787 (Dec 26, 2003).
125. R. C. Lee, R. L. Feinbaum, V. Ambros, The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75**, 843 (Dec 3, 1993).
126. B. Wightman, I. Ha, G. Ruvkun, Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. *Cell* **75**, 855 (Dec 3, 1993).
127. Y. Wang, G. Sheng, S. Juranek, T. Tuschl, D. J. Patel, Structure of the guide-strand-containing argonaute silencing complex. *Nature* **456**, 209 (Nov 13, 2008).
128. J. B. Ma *et al.*, Structural basis for 5'-end-specific recognition of guide RNA by the A. fulgidus Piwi protein. *Nature* **434**, 666 (Mar 31, 2005).
129. J. S. Parker, S. M. Roe, D. Barford, Structural insights into mRNA recognition from a PIWI domain-siRNA guide complex. *Nature* **434**, 663 (Mar 31, 2005).
130. S. Yekta, I. H. Shih, D. P. Bartel, MicroRNA-directed cleavage of HOXB8 mRNA. *Science* **304**, 594 (Apr 23, 2004).
131. C. Shin *et al.*, Expanding the microRNA targeting code: functional sites with centered pairing. *Mol Cell* **38**, 789 (Jun 25, 2010).
132. I. Lee *et al.*, New class of microRNA targets containing simultaneous 5'-UTR and 3'-UTR interaction sites. *Genome Res* **19**, 1175 (Jul, 2009).
133. J. Hausser, A. P. Syed, B. Bilen, M. Zavolan, Analysis of CDS-located miRNA target sites suggests that they can effectively inhibit translation. *Genome Res* **23**, 604 (Apr, 2013).
134. A. Grimson *et al.*, MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell* **27**, 91 (Jul 6, 2007).
135. S. N. Bhattacharyya, R. Habermacher, U. Martine, E. I. Closs, W. Filipowicz, Relief of microRNA-mediated translational repression in human cells subjected to stress. *Cell* **125**, 1111 (Jun 16, 2006).
136. P. H. Olsen, V. Ambros, The lin-4 regulatory RNA controls developmental timing in Caenorhabditis elegans by blocking LIN-14 protein synthesis after the initiation of translation. *Dev Biol* **216**, 671 (Dec 15, 1999).
137. L. P. Lim *et al.*, Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* **433**, 769 (Feb 17, 2005).
138. D. Baek *et al.*, The impact of microRNAs on protein output. *Nature* **455**, 64 (Sep 4, 2008).
139. M. Selbach *et al.*, Widespread changes in protein synthesis induced by microRNAs. *Nature* **455**, 58 (Sep 4, 2008).
140. A. A. Bazzini, M. T. Lee, A. J. Giraldez, Ribosome profiling shows that miR-430 reduces translation before causing mRNA decay in zebrafish. *Science* **336**, 233 (Apr 13, 2012).
141. S. Djuranovic, A. Nahvi, R. Green, miRNA-mediated gene silencing by translational repression followed by mRNA deadenylation and decay. *Science* **336**, 237 (Apr 13, 2012).
142. A. Krek *et al.*, Combinatorial microRNA target predictions. *Nature genetics* **37**, 495 (May, 2005).
143. B. John *et al.*, Human MicroRNA targets. *PLoS biology* **2**, e363 (Nov, 2004).

144. B. R. Cullen, S. Cherry, B. R. tenOever, Is RNA interference a physiologically relevant innate antiviral immune response in mammals? *Cell Host Microbe* **14**, 374 (Oct 16, 2013).
145. S. S. Diebold, T. Kaisho, H. Hemmi, S. Akira, C. Reis e Sousa, Innate antiviral responses by means of TLR7-mediated recognition of single-stranded RNA. *Science* **303**, 1529 (Mar 5, 2004).
146. F. Heil *et al.*, Species-specific recognition of single-stranded RNA via toll-like receptor 7 and 8. *Science* **303**, 1526 (Mar 5, 2004).
147. A. D. Judge *et al.*, Sequence-dependent stimulation of the mammalian innate immune response by synthetic siRNA. *Nat Biotechnol* **23**, 457 (Apr, 2005).
148. S. S. Diebold, Activation of dendritic cells by toll-like receptors and C-type lectins. *Handb Exp Pharmacol*, 3 (2009).
149. A. L. Jackson *et al.*, Expression profiling reveals off-target gene regulation by RNAi. *Nat Biotechnol* **21**, 635 (Jun, 2003).
150. A. L. Jackson *et al.*, Widespread siRNA "off-target" transcript silencing mediated by seed region sequence complementarity. *Rna* **12**, 1179 (Jul, 2006).
151. A. Birmingham *et al.*, 3' UTR seed matches, but not overall identity, are associated with RNAi off-targets. *Nature methods* **3**, 199 (Mar, 2006).
152. N. Schultz *et al.*, Off-target effects dominate a large-scale RNAi screen for modulators of the TGF-beta pathway and reveal microRNA regulation of TGFBR2. *Silence* **2**, 3 (2011).
153. F. D. Bushman *et al.*, Host cell factors in HIV replication: meta-analysis of genome-wide studies. *PLoS Pathog* **5**, e1000437 (May, 2009).
154. A. L. Jackson *et al.*, Position-specific chemical modification of siRNAs reduces "off-target" transcript silencing. *Rna* **12**, 1197 (Jul, 2006).
155. J. W. Engels, Gene silencing by chemically modified siRNAs. *New biotechnology* **30**, 302 (Mar 25, 2013).
156. R. Kittler *et al.*, Genome-wide resources of endoribonuclease-prepared short interfering RNAs for specific loss-of-function studies. *Nature methods* **4**, 337 (Apr, 2007).
157. J. Mercer *et al.*, RNAi screening reveals proteasome- and Cullin3-dependent stages in vaccinia virus infection. *Cell reports* **2**, 1036 (Oct 25, 2012).
158. R. Konig *et al.*, A probability-based approach for the analysis of large-scale RNAi screens. *Nature methods* **4**, 847 (Oct, 2007).
159. S. Marine, A. Bahl, M. Ferrer, E. Buehler, Common seed analysis to identify off-target effects in siRNA screens. *J Biomol Screen* **17**, 370 (Mar, 2012).
160. R. Zhong *et al.*, Computational detection and suppression of sequence-specific off-target phenotypes from whole genome RNAi screens. *Nucleic Acids Res* **42**, 8214 (2014).
161. E. Buehler *et al.*, siRNA off-target effects in genome-wide screens identify signaling pathway members. *Sci Rep* **2**, 428 (2012).
162. F. D. Sigoillot *et al.*, A bioinformatics method identifies prominent off-targeted transcripts in RNAi screens. *Nature methods* **9**, 363 (Apr, 2012).
163. A. Franceschini *et al.*, Specific inhibition of diverse pathogens in human cells by synthetic microRNA-like oligonucleotides inferred from RNAi screens. *Proc Natl Acad Sci U S A* **111**, 4548 (Mar 25, 2014).
164. J. Theodorides, [Historical brief on the fever of Malta]. *Hist Sci Med* **30**, 87 (1996).
165. G. Pappas, P. Papadimitriou, N. Akritidis, L. Christou, E. V. Tsianos, The new global map of human brucellosis. *Lancet Infect Dis* **6**, 91 (Feb, 2006).
166. E. Moreno, Retrospective and prospective perspectives on zoonotic brucellosis. *Front Microbiol* **5**, 213 (2014).

167. K. von Bargen, J. P. Gorvel, S. P. Salcedo, Internal affairs: investigating the Brucella intracellular lifestyle. *FEMS microbiology reviews* **36**, 533 (May, 2012).
168. V. L. Atluri, M. N. Xavier, M. F. de Jong, A. B. den Hartigh, R. E. Tsolis, Interactions of the human pathogenic Brucella species with their hosts. *Annu Rev Microbiol* **65**, 523 (2011).
169. A. Naroeni, F. Porte, Role of cholesterol and the ganglioside GM(1) in entry and short-term survival of Brucella suis in murine macrophages. *Infect Immun* **70**, 1640 (Mar, 2002).
170. M. Watarai *et al.*, Macrophage plasma membrane cholesterol contributes to Brucella abortus infection of mice. *Infect Immun* **70**, 4818 (Sep, 2002).
171. S. Kim *et al.*, Lipid raft microdomains mediate class A scavenger receptor-dependent infection of Brucella abortus. *Microbial pathogenesis* **37**, 11 (Jul, 2004).
172. M. Watarai, Interaction between Brucella abortus and cellular prion protein in lipid raft microdomains. *Microbes Infect* **6**, 93 (Jan, 2004).
173. P. Fontes *et al.*, Absence of evidence for the participation of the macrophage cellular prion protein in infection with Brucella suis. *Infect Immun* **73**, 6229 (Oct, 2005).
174. E. I. Castaneda-Roldan *et al.*, Adherence of Brucella to human epithelial cells and macrophages is mediated by sialic acid residues. *Cell Microbiol* **6**, 435 (May, 2004).
175. E. I. Castaneda-Roldan *et al.*, Characterization of SP41, a surface protein of Brucella associated with adherence and invasion of host epithelial cells. *Cell Microbiol* **8**, 1877 (Dec, 2006).
176. C. Czibener, J. E. Ugalde, Identification of a unique gene cluster of Brucella spp. that mediates adhesion to host cells. *Microbes Infect* **14**, 79 (Jan, 2012).
177. D. M. Posadas, V. Ruiz-Ranwez, H. R. Bonomi, F. A. Martin, A. Zorreguieta, BmaC, a novel autotransporter of Brucella suis, is involved in bacterial adhesion to host cells. *Cell Microbiol* **14**, 965 (Jun, 2012).
178. V. Ruiz-Ranwez *et al.*, BtaE, an adhesin that belongs to the trimeric autotransporter family, is required for full virulence and defines a specific adhesive pole of Brucella suis. *Infect Immun* **81**, 996 (Mar, 2013).
179. M. Deghelt *et al.*, G1-arrested newborn cells are the predominant infectious form of the pathogen Brucella abortus. *Nat Commun* **5**, 4366 (2014).
180. C. Guzman-Verri *et al.*, GTPases of the Rho subfamily are required for Brucella abortus internalization in nonprofessional phagocytes: direct activation of Cdc42. *J Biol Chem* **276**, 44435 (Nov 30, 2001).
181. K. Watanabe, M. Tachibana, S. Kim, M. Watarai, EEVD motif of heat shock cognate protein 70 contributes to bacterial uptake by trophoblast giant cells. *Journal of biomedical science* **16**, 113 (2009).
182. K. Watanabe, M. Tachibana, S. Kim, M. Watarai, Participation of ezrin in bacterial uptake by trophoblast giant cells. *Reprod Biol Endocrinol* **7**, 95 (2009).
183. B. Arellano-Reynoso *et al.*, Cyclic beta-1,2-glucan is a Brucella virulence factor required for intracellular survival. *Nat Immunol* **6**, 618 (Jun, 2005).
184. J. Pizarro-Cerda *et al.*, Brucella abortus transits through the autophagic pathway and replicates in the endoplasmic reticulum of nonprofessional phagocytes. *Infect Immun* **66**, 5711 (Dec, 1998).
185. E. Chaves-Olarte *et al.*, Activation of Rho and Rab GTPases dissociates Brucella abortus internalization from intracellular trafficking. *Cell Microbiol* **4**, 663 (Oct, 2002).

186. G. Briones *et al.*, Brucella abortus cyclic beta-1,2-glucan mutants have reduced virulence in mice and are defective in intracellular replication in HeLa cells. *Infect Immun* **69**, 4528 (Jul, 2001).
187. J. Pizarro-Cerda, E. Moreno, V. Sanguedolce, J. L. Mege, J. P. Gorvel, Virulent Brucella abortus prevents lysosome fusion and is distributed within autophagosome-like compartments. *Infect Immun* **66**, 2387 (May, 1998).
188. D. J. Comerci, M. J. Martinez-Lorenzo, R. Sieira, J. P. Gorvel, R. A. Ugalde, Essential role of the VirB machinery in the maturation of the Brucella abortus-containing vacuole. *Cell Microbiol* **3**, 159 (Mar, 2001).
189. F. Porte, J. P. Liautard, S. Kohler, Early acidification of phagosomes containing Brucella suis is essential for intracellular survival in murine macrophages. *Infect Immun* **67**, 4041 (Aug, 1999).
190. J. Celli, S. P. Salcedo, J. P. Gorvel, Brucella coopts the small GTPase Sar1 for intracellular replication. *Proc Natl Acad Sci U S A* **102**, 1673 (Feb 1, 2005).
191. E. Fugier *et al.*, The glyceraldehyde-3-phosphate dehydrogenase and the small GTPase Rab 2 are crucial for Brucella replication. *PLoS Pathog* **5**, e1000487 (Jun, 2009).
192. M. de Barsy *et al.*, Identification of a Brucella spp. secreted effector specifically interacting with human small GTPase Rab2. *Cell Microbiol* **13**, 1044 (Jul, 2011).
193. A. Gross, A. Terraza, S. Ouahrani-Bettache, J. P. Liautard, J. Dornand, In vitro Brucella suis infection prevents the programmed cell death of human monocytic cells. *Infect Immun* **68**, 342 (Jan, 2000).
194. Y. He *et al.*, Brucella melitensis triggers time-dependent modulation of apoptosis and down-regulation of mitochondrion-associated gene expression in mouse macrophages. *Infect Immun* **74**, 5035 (Sep, 2006).
195. T. Starr *et al.*, Selective subversion of autophagy complexes facilitates completion of the Brucella intracellular cycle. *Cell Host Microbe* **11**, 33 (Jan 19, 2012).

2 Aim of the thesis

2 AIM OF THE THESIS

The overall aim of my thesis, started in November 2010, was to identify host factors that are involved in *Brucella* infection of human cells. This project was embedded in the InfectX and later TargetInfectX consortium that aims at deciphering the human infectome of diverse bacterial and viral pathogens. To this end, we performed large-scale, microscopy-based RNA interference screening for infection in HeLa cells, where I was part of the team performing the *Brucella* screens. The work performed during my thesis can be separated in two parts which are tightly interconnected.

The first part intended at understanding miRNA-like off-target effects in siRNA screening and the role of natural and synthetic miRNAs in *Brucella* infection. To validate miRNA-like off-target effects, I tested custom designed oligonucleotides and a human miRNA mimic library for their effect on *Brucella* infection. Furthermore, I used proteome and transcriptome analyses to investigate the cellular changes induced by selected miRNAs.

The second part aims at the identification and characterization of genes involved in *Brucella* infection. Identification of host factors relied on the knowledge gained on off-target effects and a validation strategy for candidate genes which was developed from smaller screen covering all human kinases. The *Brucella* genome-wide screen was designed to cover multiple aspects of the intracellular lifecycle, covering entry, trafficking, and intracellular replication. To gain a deeper understanding of the precise role of the identified host factors, I developed and employed a microscopy-based entry assay to separate early from late effects.

3 Results

3 RESULTS

3.1 RESEARCH ARTICLE I (IN PREPARATION)

High-throughput microscopy-based RNA interference assays to study host factors involved in *Brucella* infection of HeLa cells

Alain Casanova [1*], Shyan Huey Low[1*], Mario Emmenlauer[1,4*], Raquel Conde-Alvarez[3*], Suzana Salcedo[2], Jean Pierre Gorvel[2], Christoph Dehio[1]

* These authors contributed equally to this work

Manuscript in preparation

3.1.1 Statement of own contribution

The entry assay was conducted and developed by me. The endpoint assay was developed by Dr. Shyan H. Low, Dr. Raquel Conde-Alvarez, and Dr. Suzana Salcedo. Image analysis pipelines were developed by Mario Emmenlauer. The manuscript was written by me, Dr. Shyan H. Low, and Mario Emmenlauer.

3.1.2 Manuscript

High-throughput microscopy-based RNA interference assays to study host factors involved in *Brucella* infection of HeLa cells

Authors

Alain Casanova [1*], Shyan Huey Low[1*], Mario Emmenlauer[1,4*], Raquel Conde-Alvarez[3*], Suzana Salcedo[2], Jean Pierre Gorvel[2], Christoph Dehio[1]

[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

[4] Mario Emmenlauer BioDataAnalysis, Munich, Germany

* 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

Summary

Here we describe two phenotypic assays applicable for high-throughput, high-content RNA interference (RNAi) screening of host factors involved in *Brucella* infection of HeLa cells. The assays are based on detection of fluorescently labeled bacteria and host cells using automated wide-field microscopy. The entry assay allows detection of host factors involved in *Brucella* entry and the endpoint assay covers all steps required for *Brucella* intracellular replication.

Abstract

Brucella species are facultative intracellular pathogens that infect domestic animals as their natural hosts. Transmission to humans is most commonly caused by direct contact with infected animals or by ingestion of contaminated food and can lead to severe chronic infections.

Brucella is able to invade professional and non-professional phagocytic cells and replicates within endoplasmic reticulum- (ER) derived intracellular vacuoles. The molecular mechanisms by which *Brucella* enters cells, avoids lysosomal degradation, and finally replicates in an ER-like compartment remain largely unknown. Here we describe two microscopy-based assays for RNAi screens of *Brucella* entry and replication in HeLa cells. The assays are based on the detection of fluorescently labeled bacteria in fluorescently labeled host cells using automated wide-field microscopy. The fluorescent images are analyzed using a standardized image analysis pipeline in CellProfiler which allows single cell-based infection scoring.

In the endpoint assay, intracellular replication is measured two days after infection. This allows bacteria to traffick to their replicative niche where intracellular proliferation is initiated around 12 h after bacterial entry. *Brucella* which have successfully established an intracellular niche will thus have strongly proliferated inside the host cell by the end of this assay and occupy a large part of the cell. Since intracellular bacteria will greatly outnumber individual extracellular or intracellular non-replicative bacteria, a strain constitutively expressing GFP is used in this assay. The strong GFP signal in infected cells is then used to automatically identify infected cells.

In contrast to the endpoint assay, it is essential to differentiate between intracellular and extracellular bacteria for the entry assay. Here, we use a *Brucella* strain encoding for a tetracycline-inducible GFP. Induction of GFP with simultaneous inactivation of extracellular bacteria by gentamicin enables the differentiation between intracellular and extracellular bacteria based on the GFP signal, with only intracellular bacteria being able to express GFP. This allows the robust detection of single intracellular bacteria before intracellular proliferation is initiated.

Introduction

Brucella species are gram-negative, facultative intracellular pathogens belonging to the class of α-Proteobacteria. They cause abortions and infertility in their natural hosts such as cattle, goats, or sheep resulting in severe economic losses in endemic areas. Brucellosis is one of the most important zoonotic diseases worldwide with over half a million new cases annually[1]. Transmission to human is most commonly caused by direct contact with infected animals or by ingestion of contaminated food such as unpasteurized milk. Symptoms of the febrile disease are unspecific, which cause difficulties in the diagnosis of brucellosis. If untreated, patients can develop a chronic infection with more severe symptoms such as arthritis, endocarditis, and neuropathy[2].

On the cellular level *Brucella* is able to invade phagocytic and non-phagocytic cells and replicates within an intracellular compartment known as the *Brucella*-containing vacuole (BCV). Internalization of bacteria requires actin cytoskeleton rearrangement by Rac, Rho, and direct activation of Cdc42[3]. Inside the eukaryotic host cell, the BCV traffics along the endocytic pathway and despite the interaction with lysosomes, bacteria manage to avoid degradation[4]. Acidification of the BCV by the vesicular ATPase is required to induce the expression of the bacterial type IV secretion system (T4SS)[5]. It is believed that bacterial effectors secreted by the T4SS are essential for *Brucella* to establish its replicative niche, since deletion of the T4SS[6] or inhibition of the vesicular ATPase lead to defects in the establishment of the intracellular niche[7]. Bacteria do not replicate during the phase of trafficking or until *Brucella* establishes an ER-like replicative niche[8]. Once intracellular proliferation occurs, the BCV is found in close association with ER markers such as calnexin and glucose-6-phosphatase[6].

The molecular mechanisms by which *Brucella* enters cells, avoids lysosomal degradation, and finally replicates in an ER-like compartment remain largely unknown. Host factors involved in different steps of infection have mainly been identified by targeted approaches or a small scale siRNA screen performed in Drosophila cells[9]. These have shed light on the contribution of individual host factors during *Brucella* infection but we are still far from a comprehensive understanding of the entire process.

Here we present a protocol that allows the identification of human host factors using large-scale RNAi screening in combination with automated wide-field fluorescence microscopy and automated image analysis. The reverse siRNA transfection protocol was optimized in the

Research and Technology Development (RTD) project InfectX in the frame of SystemsX.ch, the Swiss Initiative for Systems Biology. Reverse siRNA transfection of HeLa cells is performed with minor modifications as described by Kühbacher *et al.*[10]. We have developed an endpoint assay that covers a large part of the *Brucella* intracellular lifecycle except egress and infection of neighboring cells. To further characterize the hits identified in the endpoint assay, we use a modified protocol to identify factors involved in early steps of the infection.

A *Brucella abortus* strain that constitutively expresses GFP is used for the endpoint assay where bacteria are allowed to infect cells for two days. During this time, bacteria can enter cells, traffic to the ER-derived replicative niche, and replicate in the peri-nuclear space. High levels of bacterial GFP signal inside infected cells can then be used to reliably detect infected cells on a single cell level.

To study *Brucella* entry in a high throughput assay, it is important to be able to distinguish between intracellular and extracellular bacteria. Here we present a method that circumvents differential antibody staining of intracellular and extracellular bacteria. It is based on a *Brucella* strain expressing a tetracycline-inducible GFP in combination with constitutively expressed dsRed. The presence of a constitutive dsRed marker allows identification of all bacteria that are present in the sample. GFP expression is induced by the addition of the non-toxic tetracycline analog anhydrotetracycline (aTc) simultaneously with inactivation of extracellular bacteria by gentamicin (gent). While the cell-impermeable antibiotic gent kills extracellular bacteria, aTc can enter the host cell and induce GFP expression selectively in intracellular bacteria. This allows the robust detection of single intracellular bacteria using wide-field microscopy. Similar induction schemes to selectively express GFP in intracellular bacteria has been used previously to study intracellular *Shigella*[11].

Protocol

1. Preparation of screening plates and culturing of bacteria and cells

1.1 Preparation of *Brucella abortus* starter cultures

Work with live *Brucella abortus* strains must be performed inside a biosafety level 3 (BSL3) laboratory considering all required regulations and safety precautions.

1. Streak *Brucella abortus 2308* (*B. abortus*) from -80°C milk stock on a Tryptic Soy Agar (TSA) plate containing 50 µg/ml kanamycin (TSA/Km). Incubate plate for 3-4 days at 37°C.
    a. Endpoint assay: *Brucella abortus* 2308 pJC43 (*B. abortus* pGFP)[12]
    b. Entry assay: *Brucella abortus* pAC042.08 (*B. abortus* pTetR-GFP)
2. Restreak bacteria on four TSA/Km plates covering the full plates and incubate for 2-3 days at 37°C.
3. Using a disposable plastic loop, transfer and resuspend bacteria from the TSA/Km plates in 10 ml 10% autoclaved skim milk. Make sure that the bacteria are well resuspended to ensure consistent bacterial concentrations in all starter cultures.
4. Aliquot 250 µl of the bacterial suspension into 2 ml screw cap tubes and freeze at -80°C.
5. Thaw an aliquot of the starter culture and transfer different amounts of the bacterial starter culture to 50 ml Tryptic Soy Broth (TSB) containing 50 µg/ml kanamycin (TSB/Km) in 250 ml screw cap bottles.
6. Incubate the bottles overnight in an orbital shaker at 100 rpm and 37°C.
7. Measure the optical density at 600 nm to determine the volume of starter culture required to reach an OD_{600} = 0.8-1.1.

1.2 Preparation of HeLa cells

1. Grow HeLa cells in Dulbecco Modified Eagle Medium (DMEM) supplemented with 10% fetal calf serum (FCS).

1.3 Preparation of screening plates with siRNAs

1. Dilute siRNAs in RNase-free water to a final concentration of 0.32 µM and transfer 5 µl to black clear-well flat-bottom 384 well-plates. Use columns 1, 2, 23, and 24 for standard controls. Negative controls include non-targeting siRNA (scrambled) and mock wells without siRNAs. Transfection controls include an siRNA that induces cell death (Kif11), as well as positive controls of known host factors involved in *Brucella* infection (e.g. ArpC3, a component of the Arp2/3 complex involved in actin polymerization). Commercially available siRNA libraries or custom siRNAs are transferred to the remaining wells.
2. Seal plates with peelable aluminum foils.

3. Plates can be stored at -20°C for at least 6 months and up to years depending on the manufacturer's recommendations.

2. Reverse siRNA transfection

1. Thaw a black 384-well plate by centrifugation at 300 x *g* for 20 min at room temperature. This brings down all the liquid which might adhere to the lid or the side of the wells.
2. Prepare the transfection medium by diluting Lipofectamine RNAiMAX 1:200 in DMEM without FCS at room temperature. Prepare the transfection medium no longer than 20 min prior to use. Carefully mix the solution before use.
3. Add 25 µl transfection medium to each well and mix the solutions by moving the plate back and forth.
4. Incubate the plate for 1 h at room temperature to allow siRNA/Lipofectamine RNAiMAX complex formation.
5. In the meantime, prepare HeLa cells by washing the sub-confluent cells of a 75 cm^2 flask once with 2.5 ml trypsin.
6. Add 1.5 ml trypsin and transfer the flask to 37°C for 2-3 min until the cells round up.
7. Resuspend the cells in 10 ml prewarmed DMEM /16% FCS.
8. Count cells and prepare a cell suspension of 10'000 cells per ml in DMEM/16% FCS.
9. Add 50 µl of the cell suspension to each well resulting in 500 cells/well.
10. Move the plate back and forth to achieve even distribution of the cells. Leave the plate at room temperature for 5-10 min to allow for the cells to settle down.
11. Seal the plate with parafilm and incubate it for 72 h on a pre-warmed aluminum plate in a 37°C humid incubator with 5% CO_2. Pre-warmed aluminum plate allows equal temperature distribution throughout the whole plate.

3. Infection and fixation

1. One day prior to infection inoculate the appropriate amount of *B. abortus* starter culture in 50 ml TSB/Km medium in a 250 ml screw cap bottle. Grow bacteria overnight at 37°C on an orbital shaker to OD_{600} = 0.8-1.1 at 100 rpm.
2. Measure OD_{600} of the bacterial culture and prepare the infection medium by diluting the bacterial culture in DMEM /10% FCS to achieve the desired multiplicity of infection

(MOI) (for 2000 HeLa cells a MOI of 10'000 is estimated. MOI titration curve should be performed for each cell type to obtain optimal MOI for infection in respective cell types).

3. Exchange the transfection medium of the 384-well plate with 50 µl of the infection medium using an automated plate washer (e.g. ELx50-16, BioTek).
4. Centrifuge the 384-well plate at 400 x *g* and 4°C for 20 min to synchronize the infection process.
5. Seal the plate with parafilm and incubate it on a pre-warmed aluminum plate in a 37°C humid incubator with 5% CO_2 for 4 h.
6. Wash the cells with DMEM/10% FCS containing 100 µg/ml gent to inactivate extracellular bacteria (use 200 µl/well overflow wash with ELx50-16).
7. For the entry assay, wash cells a second time using DMEM/10% FCS containing 100 µg/ml gent and 100 ng/ml aTc (use 200 µl/well overflow wash with ELx50-16).
8. Seal the plate with parafilm and return it to a pre-warmed aluminum plate in the 37°C humid incubator with 5% CO_2 for another 40 h or 4 h, for the endpoint assay and the entry assay, respectively.
9. For fixation, wash the wells with PBS (use 200 µl/well overflow wash with ELx50-16).
10. Exchange PBS with 50 µl 3.7% PFA (in 0.2 M HEPES, pH 7.4) using an automated plate washer. Each batch of PFA must be tested prior to use to ensure efficient inactivation of *B. abortus.*
11. Incubate for 20 min at room temperature.
12. Exchange the fixation medium with 50 µl PBS. The samples are now ready to be taken out from the BSL3 if needed.

4. Staining

1. Wash cells twice with 50 µl PBS.
2. Permeabilize cells in 50 µl 0.1% Triton-X-100 in PBS for 10 min.
3. Add 20 µl of staining solution and incubate for 30 min at room temperature.
    a. Endpoint assay: PBS containing 1 µg/ml DAPI, 1.5 U/ml DY-547-phalloidin (optional)
    b. Entry assay: PBS containing 1 µg/ml DAPI, 2 U/ml DY-647-phalloidin (optional)
4. Wash cells three times in PBS and protect the staining from light with an aluminum foil cover.

5. Imaging

Set up automated wide-field microscope for image acquisition. The settings for a Molecular Device ImageXpress microscope using MetaXpress software are described below.

1. Select 10 X objective, camera binning = 1, gain = 1
2. Select plate format corresponding to the 384-well plate
3. Acquire 9 sites per well
4. Use “Enable laser based focus” and “focus on plate and well bottom”
5. Set “First well” as initial well for finding sample and “All sites” for site autofocus
6. Use the DAPI channel to manually set Z-Offset for focus.
7. Manually select the Z-offset from DAPI for all other channels.
8. Manually correct the exposure time of all channels to ensure a wide dynamic range with low overexposure.
9. Select the following channels:
    a. Endpoint assay: DAPI, GFP, RFP (optional)
    b. Endpoint assay: DAPI, GFP, RFP, CY5 (optional)

6. Automated image analysis

We employ CellProfiler 2 [13] to segment cellular and bacterial objects and perform automated measurements within the identified objects. To follow our protocol, install CellProfiler 2.1.1 or a newer version of CellProfiler. Then, load our provided pipeline into CellProfiler and follow the instructions below to adjust the required parameters.

6.1 Endpoint assay

In this image analysis we will apply shading correction to reduce the effects of an inhomogeneous light path from the microscope. Computing the shading model at the beginning of the pipeline is a lengthy process, but analysis results will be of higher accuracy.

1. Go to the Menu “File”, select “Import” -> “Pipeline from File…”, and select our provided pipeline “BrucellaEndpoint-Ver005.cppipe”. When asked if you want to convert the legacy pipeline, answer “Don’t convert”.
2. Go to “Output” -> “View output settings” and select an input folder with the images and an output folder for the resulting Excel sheet.

3. The analysis modules work as follows:
    a. Module (1) “LoadImages” load images with DAPI and GFP channels. Adjust the name of the “Text that these images have in common” to match the image names.
    b. Modules (2) – (3) “CorrectIlluminationCalculate” create a shading model by reading all images and creating a median-averaged image.
    c. Module (4) “CorretIlluminationApply” applies the shading model to the images.
    d. Modules (5) – (6) “ImageMath” correct 12bit images to use the full dynamic range [0, 1]. Adjust the parameter “Multiply the first image by” matching the bit depth of your microscope camera. For 8 and 16bit images set the parameter to “1.0”, for 12bit images set the parameter to “16.0”.
    e. Module (7) “MeasureImageIntensity” measures the lower quartile intensity of the DAPI and GFP images. The lower quartile intensity is a robust measure for the image background in this assay.
    f. Modules (8) – (9) “ImageMath” subtract the lower quartile intensity from the image. This is done to remove the background from the images.
    g. Module (10) “ImageMath” subtracts a fraction of the GFP image from the DAPI image. This is done to suppress the *Brucella* signal in the DAPI staining. Adjust the parameter “Multiply the second image by” to a value that will suppress the *Brucella* signal in the DAPI staining without suppressing the Nuclei. To identify a good value, step over the module, display the output image, right-click the image and set “Image contrast” to “Log normalized”. Set the parameter sufficiently high that *Brucella* signal is suppressed, but low enough to avoid black areas that indicate over-subtraction.
    h. Module (11) “IdentifyPrimaryObjects” identifies the nuclei in the DAPI stained images from which the *Brucella* signal was suppressed. Set the parameter “Lower bound on threshold” high enough that only nuclei are segmented, and background is ignored. To identify a good value, step over the module, display the input image, right-click the image and display the histogram. The peak of the histogram typically indicates background. Set the parameter high enough to ignore background, which is important for empty sites.
    i. Module (12) “ExpandOrShrinkObjects” creates expanded nuclei with an expansion of 8 pixels.

j. Module (13) “IdentifyTertiaryObjects” creates peri-nuclear rings around the nuclei from the expanded nuclei, by removing the nuclei from the expanded nuclei.
k. Module (14) “ExpandOrShrinkObjects” creates a Voronoi cell body around the nuclei. A Voronoi cell body is an expansion of the nucleus by 25 pixels.
l. Module (15) “MeasureObjectIntensity” measures the intensity of the GFP image at the nuclei.
m. Module (16) “FilterObjects” filters for nuclei of infected cells based on the minimum mean GFP intensity in the nuclei. Adjust the parameter “Minimum value” so that cells with clearly visible *Brucella* at the nucleus are kept, and all others are filtered away. In this step, ignore *Brucella* outside the nucleus. Start with a low value that identifies too many cells, and set it stepwise higher until only cells with clearly visible *Brucella* at the nucleus are kept.
n. Module (17) “FilterObjects” filters for peri-nuclei of infected cells based on the minimum mean GFP intensity in the peri-nuclei. Adjust the parameter “Minimum value” so that cells with clearly visible *Brucella* at the peri-nucleus are kept, and all other cells are filtered away. Approach similar as in the previous module.
o. Module (18) “FilterObjects” filters for Voronoi cell body of infected cells based on the minimum upper quartile GFP intensity in the Voronoi cell body. Adjust the parameter “Minimum value” so that cells with clearly visible *Brucella* at the Voronoi cell body are kept, and all other cells are filtered away. Approach similar as in the previous module.
p. Modules (19) – (24) combine the previously found “infected nuclei”, “infected peri-nuclei” and “infected Voronoi cells” into a single object “InfectedCells”
q. Module (25) “ExportToSpreadsheet” writes an Excel CSV sheet with the summarized readouts for each site.
r. Modules (26) – (30) are optional, and create color overlay png images with infected and uninfected cells outlined on the microscope images. These images can be helpful for later inspection, and help in optimization of the analysis.
s. Optional: Perform fluorescent intensity, texture, and distribution measurements in the remaining objects depending on availability. These can provide further information on bacterial distribution, F-actin morphology, etc.

4. In the resulting Excel CSV table, the readout “Count_Nuclei” is used to quantify the number of cells, and the readout “Count_InfectedNuclei” is used to quantify infected cells.

6.2 Entry assay

In this image analysis we do not apply shading correction because the adversary effect of shading is less pronounced in the Entry assay. Users interested in highest accuracy may add shading correction at their own discretion. Note that with the low density of objects in the GFP channel, much care has to be taken to arrive at a smooth shading model.

1. Go to the Menu “File”, select “Import” -> “Pipeline from File…”, and select our provided pipeline “BrucellaEntry-Ver003.cppipe”. When asked if you want to convert the legacy pipeline, answer “Don’t convert”.
2. Go to “Output” -> “View output settings” and select an input folder with the images and an output folder for the resulting Excel sheet.
3. The Analysis modules work as follows:
    a. Module (1) “LoadImages” loads images with DAPI and GFP channels. Adjust the name of the “Text that these images have in common” to match the image names.
    b. Module (2) – (3) “ImageMath” correct 12bit images to use the full dynamic range [0, 1]. Adjust the parameter “Multiply the first image by” matching the bit depth of your microscope camera. For 8 and 16bit images set the parameter to “1.0”, for 12bit images set the parameter to “16.0”.
    a. Module (4) “IdentifyPrimaryObjects” identifies the nuclei in the DAPI stained images from which the *Brucella* signal was suppressed. Set the parameter “Lower bound on threshold” high enough that only nuclei are segmented, and background is ignored. To identify a good value, step over the module, display the input image, right-click the image and display the histogram. The peak of the histogram typically indicates background. Set the parameter high enough to ignore background, which is important for empty sites.
    c. Module (5) “ExpandOrShrinkObjects” creates a Voronoi cell body around the nuclei. A Voronoi cell body is an expansion of the nucleus by 25 pixels.

d. Module (6) “IdentifyPrimaryObjects” identifies the intracellular *Brucella* colonies in the GFP image. Set the parameter “Lower bound on threshold” high enough that only true *Brucella* are segmented and background is ignored. To identify a good value, step over the module, display the input image, right-click the image and display the histogram. The peak of the histogram typically indicates background. Set the parameter high enough to ignore background in empty sites.
e. Module (7) “MeasureObjectIntensity” measures the intensity of segmented *Brucella* colonies in the GFP image.
f. Module (8) “MeasureObjectSizeShape” measures the area of segmented *Brucella* objects.
g. Module (9) “FilterObjects” filters for segmented *Brucella* colonies that have a minimum area of 2 pixels. Smaller *Brucella* objects are discarded to reduce the effects of pixel noise. The Module subsequently filters for segmented *Brucella* colonies that have a minimum upper quartile intensity. Adjust the parameter “Minimum value” so that background and artefacts are filtered away, and only true *Brucella* colonies are kept.
h. Module (10) “MeasureObjectIntensity” measures the intensity of the filtered *Brucella* colonies in the GFP image.
i. Module (11) “RelateObjects” relates the filtered *Brucella* colonies to the Voronoi cell body.
j. Module (12) “FilterObjects” filters for infected cells, which are defined to be all cells that have a *Brucella* colony as child of the Voronoi cell body.
k. Module (13) “FilterObjects” filters for *Brucella* colonies that are children of a Voronoi cell body.
l. Module (14) “MeasureImageAreaOccupied” measures the integrated area of *Brucella* colonies that are children of a Voronoi cell body.
m. Module (15) “ExportToSpreadsheet” writes an Excel CSV sheet with the summarized readouts for each site.
n. Optional: Perform fluorescent intensity, texture, and distribution measurements in the remaining objects depending on availability. These can give further information on bacterial distribution, F-actin morphology, etc.

4. In the resulting Excel CSV table, the readout “Count_Nuclei” is used to quantify the number of cells, the readout “Count_InfectedCells” is used to quantify infected cells, and the readout “AreaOccupied_AreaOccupied_TrueIntPathogenInCells” is used to quantify the area covered by *Brucella* in infected cells.

7. Infection scoring

7.1 Endpoint assay

1. In CellProfiler, adjust the GFP thresholds for the nucleus, peri-nucleus, and Voronoi cell body. See section 6.1 for details. If the GFP intensity in any of the three objects exceeds the threshold, the cell is considered infected.
2. In Excel, calculate a per well infection rate by dividing the number of infected cells by the total number of cells. If multiple sites per well have been imaged, it is required to first summarize all sites for each well, to build a per-well number of infected cells and a per-well total number of cells.
3. Perform Z scoring normalization to account for plate-to-plate variations if the plates contain sufficient non-hit siRNAs. A sufficient number of non-hit siRNAs per plate can be assumed if full libraries are screened.

$$Zscore = \frac{Infection\ rate\ (well) - mean\ infection\ rate\ (plate)}{standard\ deviation\ (plate)}$$

7.2 Entry assay

1. In CellProfiler, adjust the GFP threshold for intensity and the minimum area of 2 pixels of intracellular bacteria. This allows the exclusion of bacteria that are too dim and potentially extracellular or single pixel artifacts. The remaining bacterial objects are considered “true intracellular bacteria”. See section 6.2 for details. A cell is considered infected if at least one true intracellular bacterial object is identified within the Voronoi cell body.
2. In Excel, calculate a per well infection rate by dividing the number of infected cells by the total number of cells. If multiple sites per well have been imaged, it is required to first

summarize all sites for each well, to build a per-well number of infected cells and a per-well total number of cells.

3. To quantify the bacterial load of infected cells, the average area occupied by true intracellular bacteria per cell is estimated. To this end, the integrated area of all true intracellular bacteria overlapping with infected cells is summed up per site. To account for artefacts, the median of all sites is used as the well readout. This readout is divided by the total number of infected cells per well, to gather a readout of average bacterial load per infected cell.
4. If this assay is used as a follow up assay containing a large number of genes involved in *Brucella* infection Z score normalization should not be applied. Instead, normalization using mock wells as a reference can be used to account for plate to plate variation.

Representative Results

Figure 1A shows an example of image analysis used to automatically identify infected cells in the endpoint assay. Nuclei of HeLa cells nuclei stained with DAPI were identified and a peri-nucleus of 8 pixels width surrounding the nucleus and a Voronoi cell body by extension of the nucleus by 25 pixels was calculated. Since bacteria mainly proliferate in the peri-nuclear space, the GFP intensity in this area of the cell is the most robust measure to discriminate between infected and non-infected cells. In some cases, bacteria are found to proliferate outside of the peri-nucleus or overlay to a large extent with the nucleus. Therefore, these two additional objects were also considered for the identification of infected cells. Segmentation as well as GFP intensity measurements were performed with the image analysis software CellProfiler 2.

Brucella requires actin rearrangement for successful invasion of host cells. Thus, depletion of actin remodeling components is suitable to serve as positive controls for siRNA screening. Here we used ArpC3, a component of the Arp2/3 complex involved in actin polymerization as our positive control. As seen in Figure 1B, depletion of ArpC3 reduced the number of cells that show proliferating bacteria two days after infection. Applying the automated image analysis pipeline to these images allowed the quantification of the observed effect (Figure 1C). The data which are shown here originate from the control layout of a genome-wide siRNA screen where they served as controls. Z scoring was applied to account for plate-to-plate variations.

Figure 2A illustrates the image analysis used to identify infected cells and measure the intracellular bacteria load in the entry assay. In contrast to the endpoint assay, the entry assay uses bacterial segmentation and a Voronoi cell body as cellular object. Only the GFP signal of intracellular *Brucella* was used to segment the pathogen in this image analysis pipeline. True intracellular bacteria were defined by a minimal size of 2 pixels and a GFP intensity which exceeds the GFP intensity of extracellular bacteria, which show a dim background GFP signal. A cell was considered infected, if at least one intracellular bacterial object overlaps with its Voronoi cell body. Furthermore, the bacterial load is estimated by calculating the average area of infected cells that is covered by intracellular bacteria.

As for the endpoint assay, depletion of ArpC3 showed the expected reduction in *Brucella* infection in the entry assay compared to cells treated with a control siRNA (Figure 2B). Quantification of the infection rate confirmed that the number of infected cells (Figure 2C) as well as the number of intracellular bacteria in infected cells (Figure 2D) was reduced by depletion of ArpC3.

Discussion

Bacterial pathogens have evolved numerous strategies to manipulate eukaryotic host cells to their benefit. Pathogens causing acute infections often show rapid proliferation which is accompanied by significant alarming of the immune system and loss of viability of infected cells. In contrast, *Brucella* and other pathogens that cause chronic infections manage to establish long-lasting interactions within host cells. Therefore, bacteria need to fine tune host cell functions to their benefit without disrupting the cells homeostasis. A common strategy to identify the host cell factors involved in an infection process on a genome scale is the use of systematic gene depletion by siRNAs coupled to a functional readout of infection. Here we present protocols suitable for high-content high-throughput screening based on detection of fluorescent bacteria.

An important aspect of successful large-scale screening involves a robust automated image analysis and infection scoring. The methods described here rely on the detection of a cell nucleus by DAPI staining which is generally very robust, and the definition of an area of interest where the pathogen signal is detected. The endpoint assay makes use of the detection of proliferating bacteria in the peri-nucleus, nucleus, and Voronoi cell body. This allows robust detection of

infected cells. Furthermore, it allows defining different thresholds for the individual compartments. While the thresholds can be set relatively low in the nucleus and peri-nucleus, higher thresholds should be employed for the Voronoi cell body. This helps to avoid the detection of GFP signal from neighboring cells which is important since the endpoint assay lasts for five days during which HeLa cells have proliferated and grown to confluency.

The entry assay in contrast only lasts for three days and HeLa cells are thus not as densely grown. Here, only a Voronoi cell body of 25 pixels (16.125 μm) is chosen for three main reasons. First, early during infection bacteria might not yet have reached the nuclear or peri-nuclear area and may not be detected by measurements within these objects. Second, in contrast to the endpoint assay, bacteria have not started intracellular replication. Thus, using the Voronoi cell body object, we avoid missing individual bacteria which are located further away from the nucleus. Third, the endpoint assay makes use of quantification of the bacterial load which makes it desirable to cover as many bacteria as possible with the Voronoi cell body.

A cell body stain such as the phalloidin staining suggested in this protocol is not essential for the detection of infected cells in either assay. Irrespective of the need for a cell body stain for infection scoring, we would suggest to make use of this staining if possible. Even though it comes with some additional costs of material and imaging time, it provides a number of valuable advantages. In both assays, a cell body stain will give additional information of morphological changes induced by the siRNA treatment. Altered cell size, cell shape, cell-cell-contacts, or actin distributions can be very important indicators of the underlying mechanism of altered pathogen infection[14]. Furthermore, it allows adjusting the size of the Voronoi cell body to fit the actual cell size especially when another cell type is used for this protocol. For the entry assay, a cell body staining with phalloidin would benefit the accuracy of infection scoring, especially for estimating the bacterial load of infected cells on a single cell level. This will be relevant if either a cell line with a very irregular shape is chosen or if the siRNA treatment induces morphological changes of the cell.

In the entry assay, the dsRed fluorescent marker which is constitutively expressed in all bacteria is not directly used for infection scoring. However, it serves as important quality control. Visual inspection of images gives a good indication whether bacteria that are outside cells show no or only a very weak signal compared to bacteria in close proximity of cells, indicative of an intracellular location. This is an important control to make sure that induction of GFP by aTc as

well as inactivation of extracellular bacteria by gentamicin was efficient. Here, it has to be considered that in the absence of aTc bacteria express very low levels of GFP.
In addition, extracellular bacteria (red signal and no/weak GFP signal) that overlap with a HeLa cell body can provide further information. In this case, bacteria could either be bound to the cell surface or they were internalized but failed to induce the GFP signal, indicative of very rapid killing of bacteria inside a host cell. These hypotheses can then be further tested with alternative assays such as counting of colony forming units (CFU) on TSA plates.
A further advantage of using a tetracycline inducible system over differential antibody staining to visualize intracellular bacteria is its applicability for live cell imaging. Depletion of host factors potentially affects various stages of the infection process which can be resolved by quantifying the dynamics of intracellular stages. To this end, the tetracycline inducible system can be used with live cell imaging to measure the onset of replication or to provide the information of the percentage of intracellular bacteria that initiate replication.

Disclosures

The authors declare no conflicts of interest.

Acknowledgements

This work was supported by grant 51RT 0_126008 for the Research and Technology Development (RTD) project InfectX in the frame of SystemsX.ch, the Swiss Initiative for Systems Biology. We acknowledge grant 31003A-132979 from the Swiss National Science Foundation (SNSF). Work of S.H.L and A.C. was supported by the International PhD Program "Fellowships for Excellence" of the Biozentrum. Simone Muntwiler is acknowledged for technical assistance. We would like to thank Dirk Bumann from the Biozentrum for providing pNF106. We acknowledge Damian Murezzan for optimization of image analysis parameters.

Materials

Name	Company	Catalog number	Comment
Tryptic Soy Agar (TSA)	BD	236950	
Tryptic Soy Broth (TSB)	Fluka	22092	
Kanamycin sulfate	Sigma-Aldrich	60615	
Skim milk			
250 ml screw cap bottle	Corning	8396	
DMEM	Sigma-Aldrich	D5796	
Fetal Calf Serum (FCS)	Gibco	10270	Heat-inactivated
Greiner CELLSTAR 384-well plate	Sigma-Aldrich	M2062	
Peelable aluminum foil	Costar	6570	
Lipofectamine RNAiMAX	Invitrogen	13778-150	
Gentamicin	Sigma-Aldrich	G1397	
Anhydrotetracycline hydrochloride	Sigma-Aldrich	37919	100 ug/ml solution in 100% ethanol is kept at -20°C protected from light in aluminum-foil
PBS	Gibco	20012	
Paraformaldehyde	Sigma-Aldrich	P6148	Dissolve in 0.2 M HEPES buffer, pH 7.4. Store at -20°C and thaw freshly the day before use. Each batch must be validated to ensure efficient inactivation of *B. abortus*

HEPES	Sigma-Aldrich	H3375	
Triton X-100	Sigma-Aldrich	T9284	
DAPI	Roche	10236276001	
DY-547-Phalloidin	Dyomics	547-33	Optional
DY-647-Phalloidin	Dyomics	647-33	Optional
Scrambled siRNA	Dharmacon	D-001810-10	
Kif11 siRNA	Dharmacon	L-003317-00	
ARPC3 siRNA	Dharmacon	L-005284-00	
Brucella abortus 2308			
pJC43 (*aphT::GFP*)			Celli *et al.*[12]
pAC042.08 (*aphT::dsRed, tetO::tetR-GFP*)			Construction: pJC44 [4] was digested with EcoRI followed by generation of blunt ends with Klenov enzyme and subsequent digestion with SalI. TetR-GFP was amplified from pNF106 using primer prAC090 and prAC092. Following digestion with SalI, the TetR-GFP product was ligated to the digested pJC44 vector.
Primer prAC090	Sigma-Aldrich		TTTTTGAATTCTGGCAATTCCGAC GTCTAAGAAACC
Primer prAC092	Sigma-Aldrich		TTTTTGTCGACTTTGTCCTACTCA GGAGAGCGTTC
HeLa CCL-2	ATCC	CCL-2	

References

1 Pappas, G., Papadimitriou, P., Akritidis, N., Christou, L. & Tsianos, E. V. The new global map of human brucellosis. *Lancet Infect Dis* **6**, 91-99, doi:10.1016/S1473-3099(06)70382-6 (2006).

2 Atluri, V. L., Xavier, M. N., de Jong, M. F., den Hartigh, A. B. & Tsolis, R. E. Interactions of the human pathogenic Brucella species with their hosts. *Annu Rev Microbiol* **65**, 523-541, doi:10.1146/annurev-micro-090110-102905 (2011).

3 Guzman-Verri, C. *et al.* GTPases of the Rho subfamily are required for Brucella abortus internalization in nonprofessional phagocytes: direct activation of Cdc42. *J Biol Chem* **276**, 44435-44443, doi:10.1074/jbc.M105606200 (2001).

4 Starr, T., Ng, T. W., Wehrly, T. D., Knodler, L. A. & Celli, J. Brucella intracellular replication requires trafficking through the late endosomal/lysosomal compartment. *Traffic* **9**, 678-694, doi:10.1111/j.1600-0854.2008.00718.x (2008).

5 Boschiroli, M. L. *et al.* The Brucella suis virB operon is induced intracellularly in macrophages. *Proc Natl Acad Sci U S A* **99**, 1544-1549, doi:10.1073/pnas.032514299 (2002).

6 Celli, J. *et al.* Brucella evades macrophage killing via VirB-dependent sustained interactions with the endoplasmic reticulum. *J Exp Med* **198**, 545-556, doi:10.1084/jem.20030088 (2003).

7 Porte, F., Liautard, J. P. & Kohler, S. Early acidification of phagosomes containing Brucella suis is essential for intracellular survival in murine macrophages. *Infect Immun* **67**, 4041-4047 (1999).

8 Anderson, T. D. & Cheville, N. F. Ultrastructural morphometric analysis of Brucella abortus-infected trophoblasts in experimental placentitis. Bacterial replication occurs in rough endoplasmic reticulum. *Am J Pathol* **124**, 226-237 (1986).

9 Qin, Q. M. *et al.* RNAi screen of endoplasmic reticulum-associated host factors reveals a role for IRE1alpha in supporting Brucella replication. *PLoS Pathog* **4**, e1000110, doi:10.1371/journal.ppat.1000110 (2008).

10 Kuhbacher, A., Gouin, E., Cossart, P. & Pizarro-Cerda, J. Imaging InlC secretion to investigate cellular infection by the bacterial pathogen Listeria monocytogenes. *J Vis Exp*, e51043, doi:10.3791/51043 (2013).

11 Kentner, D. *et al.* Shigella reroutes host cell central metabolism to obtain high-flux nutrient supply for vigorous intracellular growth. *Proc Natl Acad Sci U S A* **111**, 9929-9934, doi:10.1073/pnas.1406694111 (2014).

12 Celli, J., Salcedo, S. P. & Gorvel, J. P. Brucella coopts the small GTPase Sar1 for intracellular replication. *Proc Natl Acad Sci U S A* **102**, 1673-1678, doi:10.1073/pnas.0406873102 (2005).

13 Kamentsky, L. *et al.* Improved structure, function and compatibility for CellProfiler: modular high-throughput image analysis software. *Bioinformatics* **27**, 1179-1180, doi:10.1093/bioinformatics/btr095 (2011).

14 Snijder, B. *et al.* Population context determines cell-to-cell variability in endocytosis and virus infection. *Nature* **461**, 520-523, doi:10.1038/nature08282 (2009).

Figure legends

Figure 1. Analysis of HeLa cells infected by *B. abortus* pGFP in the endpoint assay

A) Endpoint assay: Fluorescence image showing example of endpoint assay (green = *B. abortus* expressing GFP, blue = nuclei stained with DAPI, scale bar = 50 μm). GFP expressing *B. abortus* was allowed to enter HeLa cells for 4 h followed by killing of extracellular bacteria with gentamicin. Further incubation for 40 h allows bacteria to replicate inside HeLa cells. Automated image analysis: CellProfiler is used to segment DAPI stained nuclei followed by calculation of the peri-nucleus (ring of 8 pixels surrounding the nucleus) and Voronoi cell body (extension of the nucleus by 25 pixels). A cell is considered infected if the integrated GFP intensity exceeds a least one of the thresholds set for the peri-nucleus, nucleus, or Voronoi cell body.

B) Representative images of a full microscopic site of HeLa cells infected with GFP expressing *B. abortus* 44 h after infection. Cells were either transfected with a scrambled (non-targeting) control or siRNA targeting ArpC3. Scale bar = 50 μm

C) Quantification of infection of HeLa cells depleted for ArpC3. Data are represented as box plots of Z scored infection rate (bar = mean; whiskers and outliers of boxplot calculated with Tukey method; ***p<0.001; Mann-Whitney test; n = 7)

Figure 2. Quantification of HeLa cells infection by *B. abortus* pTetR-GFP in the entry assay

A) Entry assay: Fluorescence image showing example of entry assay (green = intracellular *B. abortus*, blue = nuclei stained with DAPI, scale bar = 50 μm). *B. abortus* pTetR-GFP was allowed to enter HeLa cells for 4 h followed by inactivation of extracellular bacteria and simultaneous induction of GFP in intracellular bacteria with aTc for 4 h.

B) Automated image analysis: CellProfiler was used to detect DAPI stained nuclei followed by calculation of a Voronoi cell body by radial extension of the nucleus by 25 pixels. Bacteria were segmented based on the GFP signal. A cell is considered infected if its Voronoi cell body overlaps with at least one segmented bacterial object. The bacterial load in infected cells is estimated by the area of segmented intracellular bacteria in infected cells (Feature used: AreaOccupied_AreaOccupied_TrueIntPathogen. No number is calculated in non-infected cells).

C) Example images showing decreased infection of *B. abortus* pTetR-GFP in cells transfected with siRNA targeting ArpC3 compared to control treated cells (scrambled siRNA). The number

of infected cells as well as the average number of intracellular bacteria in infected cells is reduced upon ARPC3 knock down.

D) Quantification of infection rate of HeLa cells with ArpC3 depletion. The data are represented as bar graphs (bar = mean; normalization to mock; error bars = standard error of the mean; *$p<0.05$; Mann-Whitney test; n = 4)

E) Estimation of bacterial load of infected HeLa cells with ArpC3 depletion. The data are represented as bar graphs (bar = mean; normalization to mock; error bars = standard error of the mean; *$p<0.05$; Mann-Whitney test; n = 4)

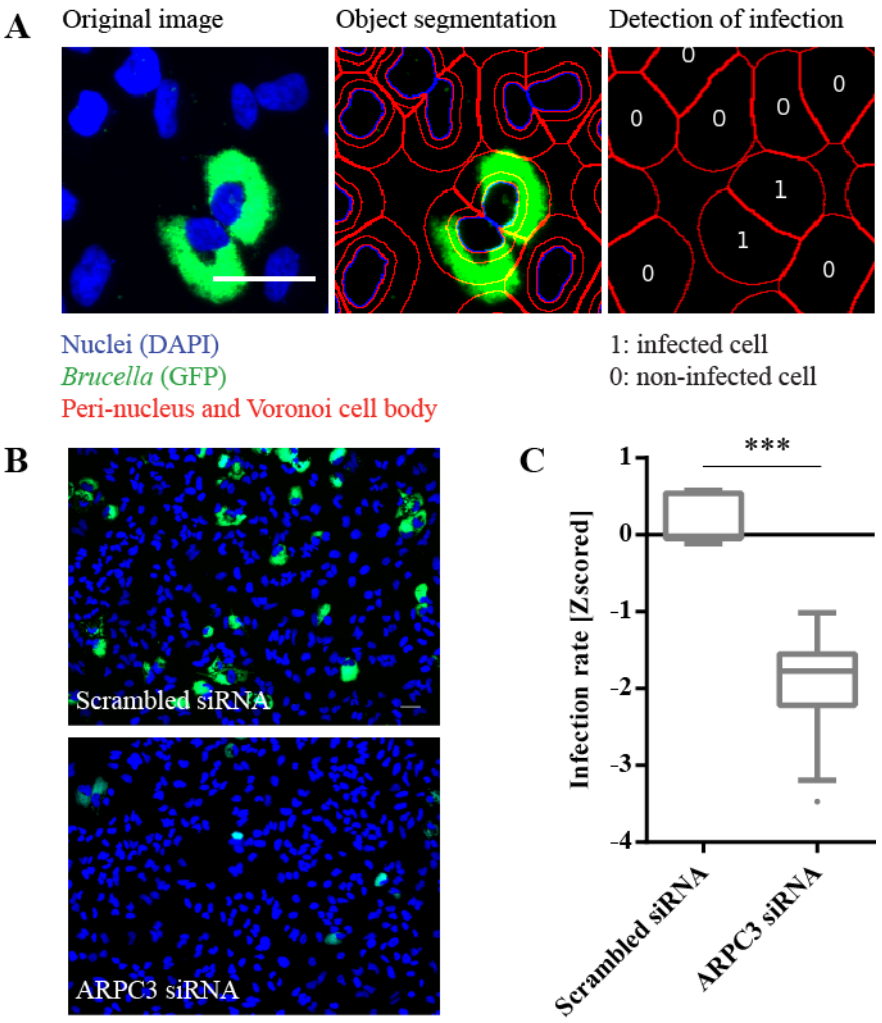

A
Original image
Object segmentation
Detection of infection
0
0
0
0
0
1
1
0
0
Nuclei (DAPI)
Brucella (GFP)
Peri-nucleus and Voronoi cell body
1: infected cell
0: non-infected cell
B
Scrambled siRNA
ARPC3 siRNA
C
***
Infection rate [Zscored]
1
0
-1
-2
-3
-4
Scrambled siRNA
ARPC3 siRNA

Figure 1

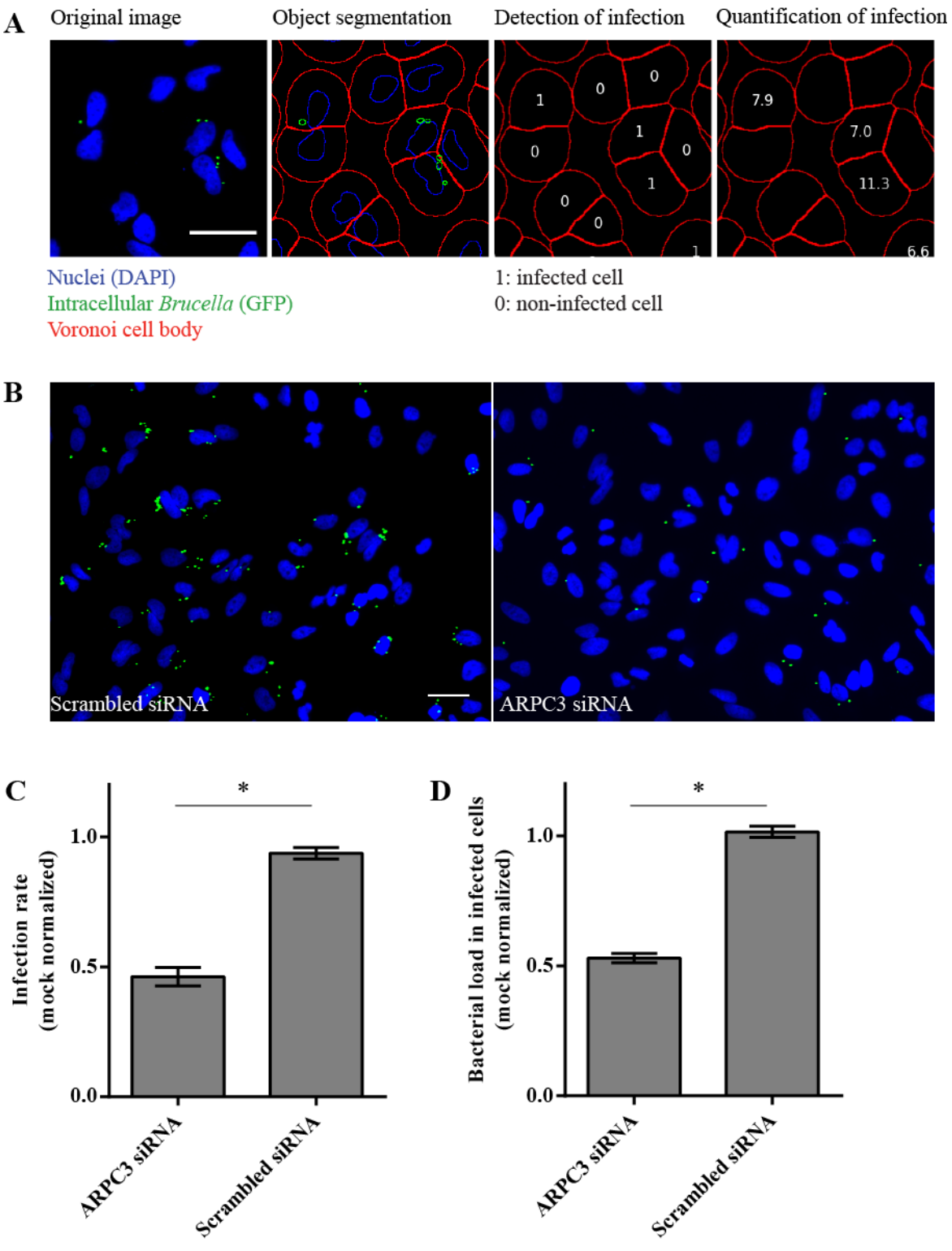

A
Original image
Object segmentation
Detection of infection
Quantification of infection
1
0
0
1
0
0
1
0
0
1
7.9
7.0
11.3
6.6
Nuclei (DAPI)
Intracellular Brucella (GFP)
Voronoi cell body
1: infected cell
0: non-infected cell
B
Scrambled siRNA
ARPC3 siRNA
C
Infection rate (mock normalized)
1.0
0.5
0.0
*
ARPC3 siRNA
Scrambled siRNA
D
Bacterial load in infected cells (mock normalized)
1.0
0.5
0.0
*
ARPC3 siRNA
Scrambled siRNA

Figure 2

3.2 RESEARCH ARTICLE II (PUBLISHED)

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 Emmenlauer, 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

3.2.1 Statement of own contribution

Genome-wide siRNA screens for *Brucella* infection used for seed analysis were performed together with Dr. Shyan H. Low. I performed the *Brucella* infection experiment for the validation experiment and was involved in the design and preparation of the layout for all pathogen screens of the latter.

3.2.2 Manuscript

PNAS

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[c], Ari Helenius[b,4], Wolf-Dietrich Hardt[d], Christoph Dehio[c], and Christian von Mering[a,4]

[a]Institute of Molecular Life Sciences and Swiss Institute of Bioinformatics, University of Zurich, CH-8057 Zurich, Switzerland; [b]Institute of Biochemistry, Eidgenössische Technische Hochschule Zurich, CH-8093 Zurich, Switzerland; [c]Biozentrum, University of Basel, CH-4056 Basel, Switzerland; [d]Institute of Microbiology, Eidgenössische Technische Hochschule Zurich, CH-8093 Zurich, Switzerland; and [e]Institute 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.

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

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

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

[4]To whom correspondence may be addressed. E-mail: ari.helenius@bc.biol.ethz.ch or mering@imls.uzh.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 pathogens' 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

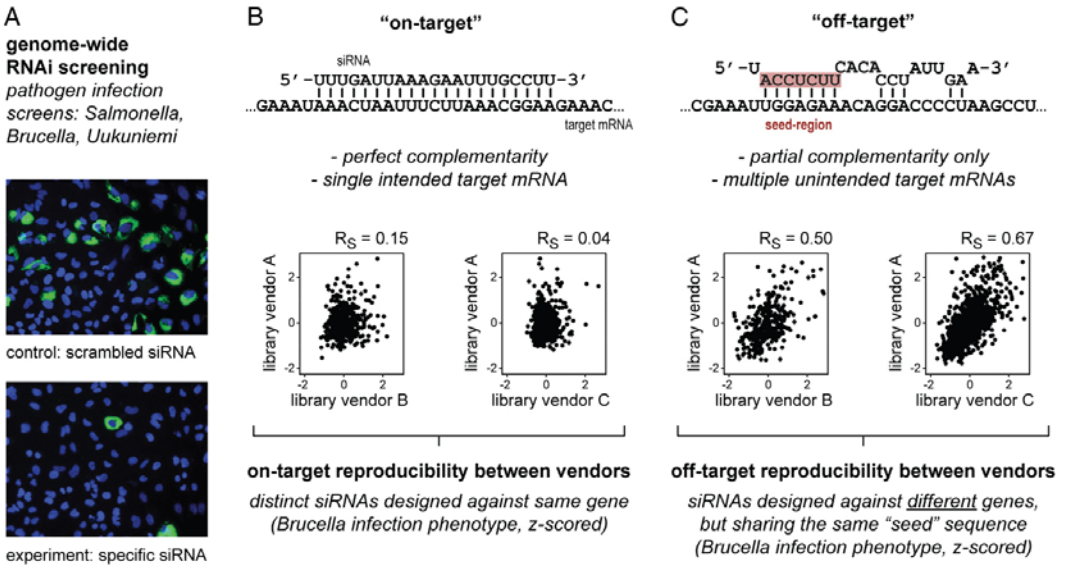


Fig. 1. Off-target effects in RNAi pathogen infection screens. (*A*) Experimental setup. HeLa cells were screened for host factors required for pathogen entry. Microscopy images from two separate wells of a typical perturbation experiment are shown (DAPI-stained HeLa cell nuclei in blue; successful pathogen 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 *C* are highly significant ($P \leq 10^{-50}$).

MICROBIOLOGY

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 four 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 compared 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).

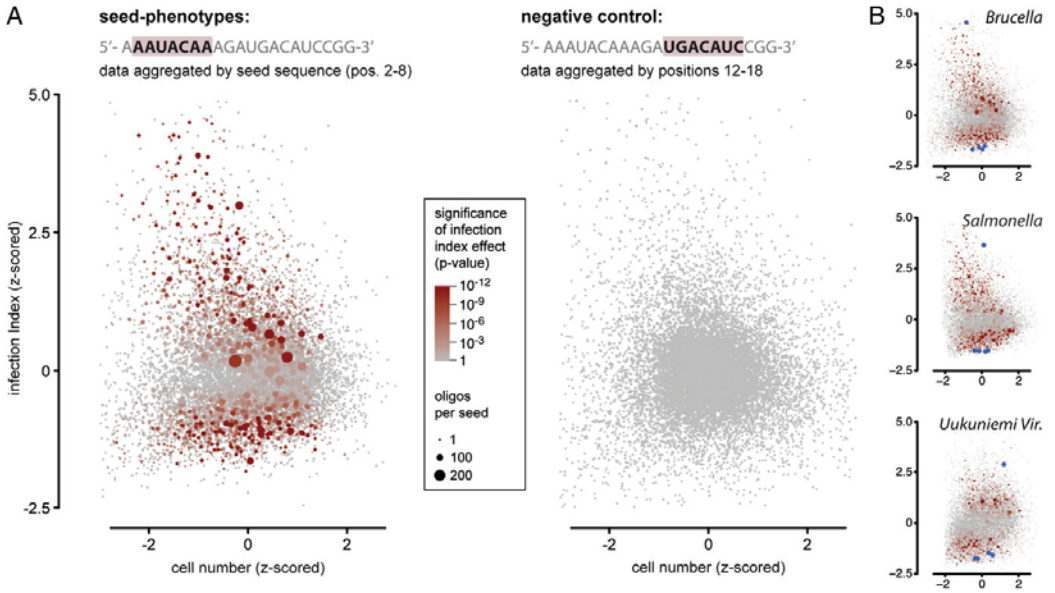


Fig. 2. Genome-wide screening data aggregated by shared seed sequences. (*A*) Visualization of the entire genome-wide data of the infection screen for *B. abortus*, aggregated by the seed sequences found in the various siRNA oligos. Each data point represents one heptameric seed sequence, showing the 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.

To experimentally confirm our findings and to formally separate the off-target and on-target contributions to each phenotype, we selected a number of seed sequences for detailed follow-up. For each of the three pathogens, four seeds were selected that were predicted to reduce infection, plus one seed that would enhance it (all marked in blue in Fig. 2*B*). Although the seeds were selected to have strong phenotypes in the infection readout, they were also chosen such that they had little effect on the cell number (seed effects on infection and on host-cell viability were often orthogonal). For each of the selected seeds, we first reordered four standard inventory oligos from the genome-wide library; in such inventory oligos, both the off-target and the intended on-target component should still be present. Importantly, we also designed novel oligos for each seed; for these oligos, the nucleotide sequences outside the seed were arbitrarily set to a random string of nucleotides (drawn from the background distribution of all oligos in the genome-wide library). The design of these latter oligos formally excludes any intended "on-target" component. For controls, we reordered a population of arbitrary inventory oligos chosen at random, as well as a set of inventory oligos with seeds predicted to have no phenotype per se. Additionally, for some seeds, we custom-designed oligos that were similar to the corresponding inventory oligos, except at one position within the seed region where they differed by a single point mutation (presumably, this should abolish any specific seed-mediated off-target effects).

Upon rescreening all three pathogen-specific assays using the new set of oligos, we indeed observed that the predicted phenotypes were clearly reproducible, both in the presence and in the absence of any specific on-target component (Fig. 3 and Fig. S3). The custom-designed oligos that featured arbitrary sequences outside the seed were blocking infection just as effectively as the corresponding inventory oligos that still had a designed on-target (Fig. 3; dark blue vs. light blue). By comparison, the overall effects of the oligos on the cell-number phenotypes were mild (Fig. S4) and often insignificant. We were able to design oligos not only to block infection, but also to enhance it if appropriate seeds were selected (orange colors in Fig. 3).

In all three screens, we observed that some of the seed sequences that showed significant phenotypic effects coincided with seed sequences known to be present in endogenous human miRNAs. This raised the possibility of predicting the overexpression phenotypes of such miRNAs—under the assumption that the target-gene specificity of endogenous miRNAs is similarly dictated to a large extent by the seed region. For the *B. abortus* screen, we set out to test this prediction by selecting eight distinct seed sequences shown to strongly block infection, which were represented in the siRNA libraries at least 10 times, and corresponded to exactly a single known human miRNA (we did not consider matches to miRNA families having multiple members that shared the same seed). Likewise, we chose eight seeds that strongly enhanced infection and eight seeds that were predicted to be neutral. For all 24 corresponding human miRNAs, we ordered commercially available, double-stranded RNA molecules intended to mimic the native miRNA. Indeed, in all cases, the predicted overexpression phenotype was confirmed experimentally (Fig. 4 and Fig. S5).

Finally, we analyzed the specificity of the observed seed effects. To test the sequence specificity, we introduced single-point mutations into the seed regions; these mutations indeed completely abolished the intended activity of the corresponding siRNAs (Fig. 5*A*). To test the pathogen specificity, we searched for seeds that would influence one pathogen, but not the other two. This was based on the rationale that the distinct pathogens should have different sequence- and pathway-specific requirements, and this should be reflected in the seed phenotypes. Indeed, at a significance level of $P \leq 10^{-6}$, the majority of active seeds (78%) affected only one pathogen. Nineteen percent of active seeds affected two pathogens, and only 3% affected all three pathogens significantly. Effectively, in our genome-wide analysis, the observations for each seed sequence describe a

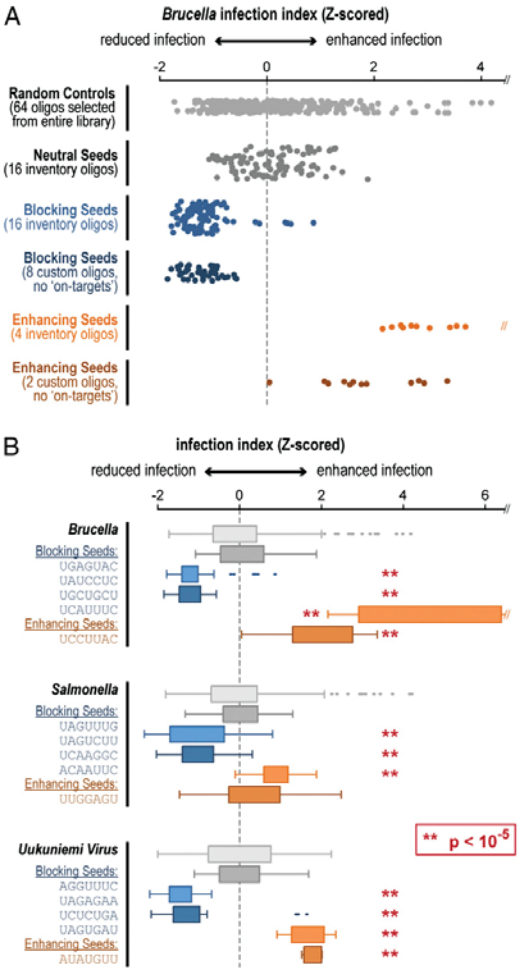
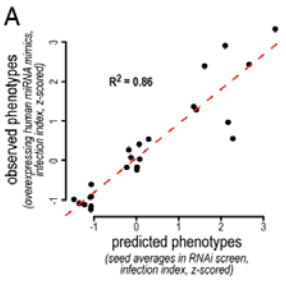


Fig. 3. Experimental confirmation of predicted seed phenotypes. (*A*) Detailed phenotypes measured for *B. abortus* (six replicates per oligo). (*B*) Summary of phenotypes measured for each of the three pathogens. The siRNA oligos predicted to block infection are shown in blue (dark blue for those that were designed not to have any on-targets), and oligos predicted to enhance infection are shown in orange (again, dark orange if lacking on-targets by design). The full sequences of all oligos in this experiment are given in Fig. S3.

vector of six phenotypes: three distinct infection phenotypes ("infection index") and three independent replicates of the cell viability/proliferation phenotype ("cell number"). Principal-component analysis of this space reveals that the three cell-number dimensions neatly fold into one component, capturing about half of the variance (Fig. 5*B*). The remainder of the phenotypes mostly discriminate between the pathogens—with the virus being on one side and the two bacteria on the other (often somewhat closer to each other than to the virus).

Overall, these results show that genome-wide datasets enable the design of novel RNAs (which we term "seed drugs") that reproducibly block infection by one or more pathogens, without



B

selected seed sequence	seed occurrence in library	seed phenotype (average)	corresponding human miRNA	miRNA sequence	observed miRNA overexpr. phenotype
UAGGGAG	10	-1.065	hsa-miR-3162	UUAGGGAGUAGAAGGGUGGGGAG	-0.611
UGAUAAG	12	-1.070	hsa-miR-3161	CUGAUAAGAACAGAGGCCCAGAU	-1.163
UAAAGCU	13	-1.474	hsa-miR-9*	AUAAAGCUAGAUAACCGAAAGU	-0.985
AGAUGUG	13	-1.076	hsa-miR-576-3p	AAGAUGUGGAAAAAUUGGAAUC	-1.241
AGUAGAC	14	-1.089	hsa-miR-411	UAGUAGACCGUAUAGCGUACG	-0.936
UGCACUU	24	-1.224	hsa-miR-4325	UUGCACUUGUCUCAGUGA	-1.122
UCUCCUC	30	-1.349	hsa-miR-4279	CUCUCCUCCCGGCUUC	-1.086
UCAGGGC	35	-1.066	hsa-miR-3127	AUCAGGGCUUGUGGAAUGGGAAG	-0.916

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 cell's 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

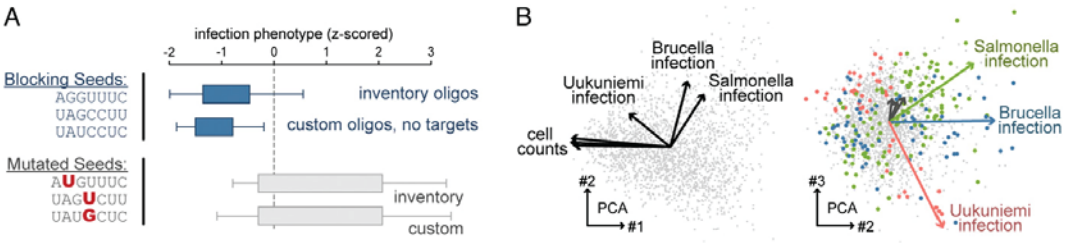


Fig. 5. Specificity of seed effects. (*A*) Effects of single-point mutations located in the seed regions. For each of the three pathogens, one seed was chosen that was predicted to block infection (data shown in blue). Shown in gray are data for the corresponding seeds that have been mutated at one position. 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 51RT 0_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.

1. Nagy A, Perrimon N, Sandmeyer S, Plasterk R (2003) Tailoring the genome: The power of genetic approaches. *Nat Genet* 33(Suppl):276–284.
2. Cong L, et al. (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science* 339(6121):819–823.
3. Mali P, et al. (2013) RNA-guided human genome engineering via Cas9. *Science* 339(6121):823–826.
4. Miller JC, et al. (2011) A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol* 29(2):143–148.
5. Zhang F, et al. (2011) Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nat Biotechnol* 29(2):149–153.
6. Mohr SE, Perrimon N (2012) RNAi screening: New approaches, understandings, and organisms. *Wiley Interdiscip Rev RNA* 3(2):145–158.
7. Brass AL, et al. (2008) Identification of host proteins required for HIV infection through a functional genomic screen. *Science* 319(5865):921–926.
8. 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.
9. 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.
10. Karlas A, et al. (2010) Genome-wide RNAi screen identifies human host factors crucial for influenza virus replication. *Nature* 463(7282):818–822.
11. Krishnan MN, et al. (2008) RNA interference screen for human genes associated with West Nile virus infection. *Nature* 455(7210):242–245.
12. Mercer J, et al. (2012) RNAi screening reveals proteasome- and Cullin3-dependent stages in vaccinia virus infection. *Cell Rep* 2(4):1036–1047.
13. Misselwitz B, et al. (2011) RNAi screen of Salmonella invasion shows role of COPI in membrane targeting of cholesterol and Cdc42. *Mol Syst Biol* 7:474.
14. Tai AW, et al. (2009) A functional genomic screen identifies cellular cofactors of hepatitis C virus replication. *Cell Host Microbe* 5(3):298–307.
15. Zhou H, et al. (2008) Genome-scale RNAi screen for host factors required for HIV replication. *Cell Host Microbe* 4(5):495–504.
16. 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.
17. Bushman FD, et al. (2009) Host cell factors in HIV replication: Meta-analysis of genome-wide studies. *PLoS Pathog* 5(5):e1000437.
18. 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.
19. Jackson AL, et al. (2006) Widespread siRNA "off-target" transcript silencing mediated by seed region sequence complementarity. *RNA* 12(7):1179–1187.
20. 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.
21. Buehler E, et al. (2012) siRNA off-target effects in genome-wide screens identify signaling pathway members. *Sci Rep* 2:428.
22. Jackson AL, et al. (2003) Expression profiling reveals off-target gene regulation by RNAi. *Nat Biotechnol* 21(6):635–637.
23. 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.
24. 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 TGFBR2. *Silence* 2:3.
25. Sigoillot FD, et al. (2012) A bioinformatics method identifies prominent off-targeted transcripts in RNAi screens. *Nat Methods* 9(4):363–366.
26. Schmaljohn C, Nichol S (2007) Bunyaviridae. *Virology*, ed Fields KD (Lippincott, Williams & Wilkins, Philadelphia), pp 1741–1788.
27. Snijder B, et al. (2009) Population context determines cell-to-cell variability in endocytosis and virus infection. *Nature* 461(7263):520–523.
28. Snijder B, et al. (2012) Single-cell analysis of population context advances RNAi screening at multiple levels. *Mol Syst Biol* 8(1):579.
29. Garcia DM, et al. (2011) Weak seed-pairing stability and high target-site abundance decrease the proficiency of lsy-6 and other microRNAs. *Nat Struct Mol Biol* 18(10):1139–1146.
30. Vejnar CE, Zdobnov EM (2012) MiRmap: Comprehensive prediction of microRNA target repression strength. *Nucleic Acids Res* 40(22):11673–11683.
31. Hsu JB, et al. (2011) miRTar: An integrated system for identifying miRNA-target interactions in human. *BMC Bioinformatics* 12:300.
32. Shamu CE, Wiemann S, Boutros M (2012) On target: A public repository for large-scale RNAi experiments. *Nat Cell Biol* 14(2):115.
33. 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

a correlations between siRNA libraries from different vendors

screen / phenotype	*Vendor A vs. Vendor B* on-target correlation	*Vendor A vs. Vendor B* off-target correlation	*Vendor A vs. Vendor C* on-target correlation	*Vendor A vs. Vendor C* off-target correlation	*Vendor B vs. Vendor C* on-target correlation	*Vendor B vs. Vendor C* off-target correlation
Brucella: infection index phenotype	0.148*	0.496***	0.040	0.668***	-0.027	0.564***
Brucella: cell number phenotype	0.146*	0.377**	0.112	0.509***	0.167*	0.544***
Salmonella: infection index phenotype	0.026	0.311**	0.044	0.252***	0.041	0.119**
Salmonella: 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)
Brucella: infection index phenotype	0.092	0.592***	0.585***
Brucella: cell number phenotype	0.074	0.444**	0.448**
Salmonella: infection index phenotype	0.095	0.627***	0.629***
Salmonella: cell number phenotype	0.099	0.478**	0.479**

Vendor B

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

Vendor C

	on-target correlation	off-target correlation	off-target (downsampled)
Brucella: infection index phenotype	0.071***	0.750***	0.649***
Brucella: cell number phenotype	0.394***	0.646***	0.538***
Salmonella: infection index phenotype	0.021	0.766***	0.690***
Salmonella: cell number phenotype	0.108***	0.679***	0.578***
Uukun. Virus: infection index phenotype	0.235***	0.727***	0.623***
Uukun. Virus: 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.

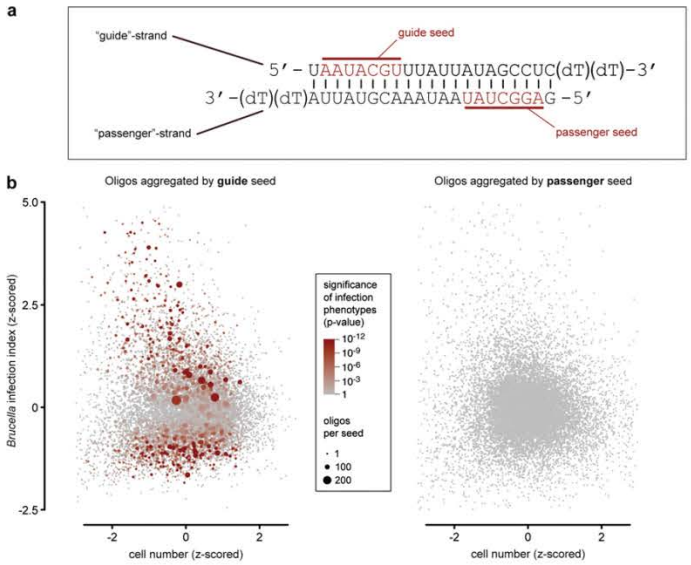
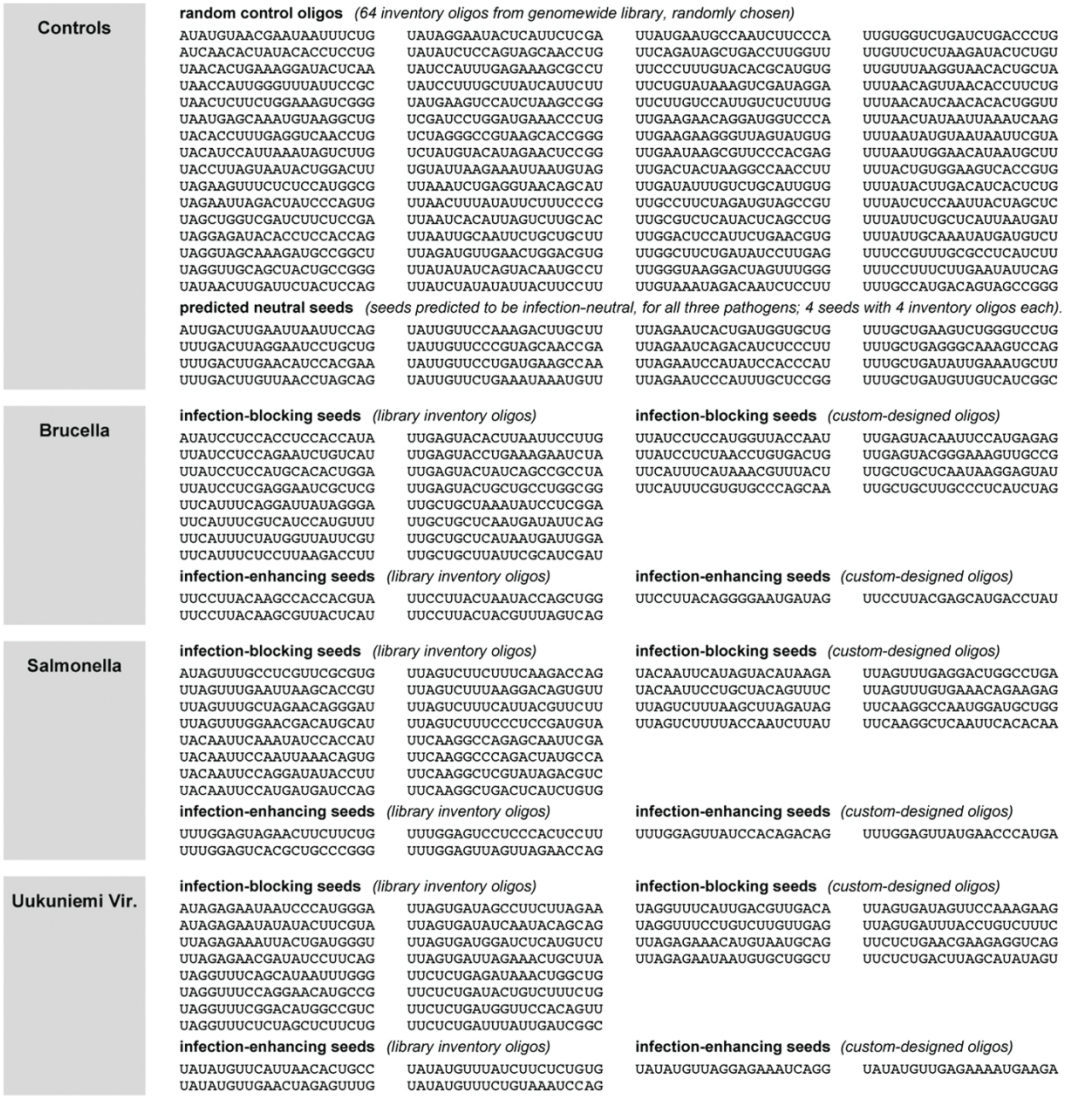


Fig. S2. Guide strands versus passenger strands. (*A*) Assuming a double-stranded design of the siRNA reagents, there are two potential active seed sequences (i.e., after the strands have separated). (*B*) siRNA oligo phenotypes aggregated per seed, as in Fig. 2. Only the guide seed shows strongly significant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



Controls

random control oligos *(64 inventory oligos from genomewide library, randomly chosen)*

AUAUGUAACGAAUAAUUUCUG	UAUAGGAAUACUCAUUCUCGA	UUAUGAAUGCCAAUCUUCCCA	UUGUGGUCUGAUCUGACCCUG
AUCAACACUAUACACCUCCUG	UAUAUCUCCAGUAGCAACCUG	UUCAGAUAGCUGACCUUGGUU	UUGUUCUCUAAGAUACUCUGU
UAACACUGAAAGGAUACUCAA	UAUCCAUUUGAGAAAGCGCCU	UUCCCUUUGUACACGCAUGUG	UUGUUUAAGGUAACACUGCUA
UAACCAUUGGGUUUAUUCCGC	UAUCCUUUGCUUAUCAUUCUU	UUCUGUAUAAAGUCGAUAGGA	UUUAACAGUUAACACCUUCUG
UAACUCUUCUGGAAAGUCGGG	UAUGAAGUCCAUCUAAGCCGG	UUCUUGUCCAUUGUCUCUUUG	UUUAACAUCAACACACUGGUU
UAAUGAGCAAAUGUAAGGCUG	UCGAUCCUGGAUGAAACCCUG	UUGAAGAACAGGAUGGUCCCA	UUUAACUAUAAUUAAAUCAAG
UACACCUUUGAGGUCAACCUG	UCUAGGGCCGUAAGCACCGGG	UUGAAGAAGGGUUAGUAUGUG	UUUAAUAUGUAAUAAUUCGUA
UACAUCCAUUAAAUAGUCUUG	UCUAUGUACAUAGAACUCCGG	UUGAAUAAGCGUUCCCACGAG	UUUAAUUGGAACAUAAUGCUU
UACCUUAGUAAUACUGGACUU	UGUAUUAAGAAAUUAAUGUAG	UUGACUACUAAGGCCAACCUU	UUUACUGUGGAAGUCACCGUG
UAGAAGUUUCUCUCCAUGGCG	UUAAAUCUGAGGUAACAGCAU	UUGAUAUUUGUCUGCAUUGUG	UUUAUACUUGACAUCACACUCG
UAGAAUUAGACUAUCCCAGUG	UUAACUUUAUAUUCUUUCCCG	UUGCCUUCUAGAUGUAGCCGU	UUUAUCUCCAAUUACUAGCUC
UAGCUGGUCGAUCUUCUCCGA	UUAAUCACAUUAGUCUUGCAC	UUGCGUCUCAUACUCAGCCUG	UUUAUUCUGCUCAUUAAUGAU
UAGGAGAUACACCUCCACCAG	UUAAUUGCAAUUCUGCUGCUU	UUGGACUCCAUUCUGAACGUG	UUUAUUGCAAAUAUGAUGUCU
UAGGUAGCAAAGAUGCCGGCU	UUAGAUGUUGAACUGGACGUG	UUGGCUUCUGAUAUCCUUGAG	UUUCCGUUUGCGCCUCAUCUU
UAGGUUGCAGCUACUGCCGGG	UUAUAUAUCAGUACAAUGCCU	UUGGGUAAGGACUAGUUUGGG	UUUCCUUUCUUGAAUAUUCAG
UAUAACUUGAUUCUACUCCAG	UUAUCUAUAUAUUACUUCCUU	UUGUAAAUAGACAAUCUCCUU	UUUGCCAUGACAGUAGCCGGG

predicted neutral seeds *(seeds predicted to be infection-neutral, for all three pathogens; 4 seeds with 4 inventory oligos each).*

AUUGACUUGAAUUAAUUCCAG	UAUUGUUCCAAAGACUUGCUU	UUAGAAUCACUGAUGGUGCUG	UUUGCUGAAGUCUGGGUCCUG
UUUGACUUAGGAAUCCUGCUG	UAUUGUUCCCGUAGCAACCGA	UUAGAAUCAGACAUCUCCCUU	UUUGCUGAGGGCAAAGUCCAG
UUUGACUUGAACAUCCACGAA	UAUUGUUCCUGAUGAAGCCAA	UUAGAAUCCAUAUCCACCCAU	UUUGCUGAUAUUGAAAUGCUU
UUUGACUUGUUAACCUAGCAG	UAUUGUUCUGAAAUAAAUGUU	UUAGAAUCCCAUUUGCUCCGG	UUUGCUGAUGUUGUCAUCGGC

Brucella

infection-blocking seeds *(library inventory oligos)*

AUAUCCUCCACCUCCACCAUA	UUGAGUACACUUAAUUCCUUG	UUAUCCUCCAUGGUUACCAAU	UUGAGUACAAUUCCAUGAGAG
UUAUCCUCCAGAAUCUGUCAU	UUGAGUACCUGAAAGAAUCUA	UUAUCCUCUAACCUGUGACUG	UUGAGUACGGGAAAGUUGCCG
UUAUCCUCCAUGCACACUGGA	UUGAGUACUAUCAGCCGCCUA	UUCAUUUCAUAAACGUUUACU	UUGCUGCUCAAUAAGGAGUAU
UUAUCCUCGAGGAAUCGCUCG	UUGAGUACUGCUGCCUGGCGG	UUCAUUUCGUGUGCCCAGCAA	UUGCUGCUUGCCCUCAUCUAG
UUCAUUUCAGGAUUAUAGGGA	UUGCUGCUAAAUAUCCUCGGA		
UUCAUUUCGUCAUCCAUGUUU	UUGCUGCUCAAUGAUAUUCAG		
UUCAUUUCUAUGGUUAUUCGU	UUGCUGCUCAUAAUGAUUGGA		
UUCAUUUCUCCUUAAGACCUU	UUGCUGCUUAUUCGCAUCGAU		

infection-enhancing seeds *(library inventory oligos)*

UUCCUUACAAGCCACCACGUA	UUCCUUACUAAUACCAGCUGG	UUCCUUACAGGGGAAUGAUAG	UUCCUUACGAGCAUGACCUAU
UUCCUUACAAGCGUUACUCAU	UUCCUUACUACGUUUAGUCAG		

Salmonella

infection-blocking seeds *(library inventory oligos)*

AUAGUUUGCCUCGUUCGCGUG	UUAGUCUUCUUUCAAGACCAG	UACAAUUCAUAGUACAUAAGA	UUAGUUUGAGGACUGGCCUGA
UUAGUUUGAAUUAAGCACCGU	UUAGUCUUUAAGGACAGUGUU	UACAAUUCCUGCUACAGUUUC	UUAGUUUGUGAAACAGAAGAG
UUAGUUUGCUAGAACAGGGAU	UUAGUCUUUCAUUACGUUCUU	UUAGUCUUUAAGCUUAGAUAG	UUCAAGGCCAAUGGAUGCUGG
UUAGUUUGGAACGACAUGCAU	UUAGUCUUUCCCUCCGAUGUA	UUAGUCUUUUACCAAUCUUAU	UUCAAGGCUCAAUUCACACAA
UACAAUUCAAAUAUCCACCAU	UUCAAGGCCAGAGCAAUUCGA		
UACAAUUCCAAUUAAACAGUG	UUCAAGGCCCAGACUAUGCCA		
UACAAUUCCAGGAUAUACCUU	UUCAAGGCUCGUAUAGACGUC		
UACAAUUCCAUGAUGAUCCAG	UUCAAGGCUGACUCAUCUGUG		

infection-enhancing seeds *(library inventory oligos)*

UUUGGAGUAGAACUUCUUCUG	UUUGGAGUCCUCCCACUCCUU	UUUGGAGUUAUCCACAGACAG	UUUGGAGUUAUGAACCCAUGA
UUUGGAGUCACGCUGCCCGGG	UUUGGAGUUAGUUAGAACCAG		

Uukuniemi Vir.

infection-blocking seeds *(library inventory oligos)*

AUAGAGAAUAAUCCCAUGGGA	UUAGUGAUAGCCUUCUUAGAA	UAGGUUUCAUUGACGUUGACA	UUAGUGAUAGUUCCAAAGAAG
AUAGAGAAUAUAUACUUCGUA	UUAGUGAUAUCAAUACAGCAG	UAGGUUUCCUGUCUUGUUGAG	UUAGUGAUUUACCUGUCUUUC
UUAGAGAAAUUACUGAUGGGU	UUAGUGAUGGAUCUCAUGUCU	UUAGAGAAACAUGUAAUGCAG	UUCUCUGAACGAAGAGGUCAG
UUAGAGAACGAUAUCCUUCAG	UUAGUGAUUAGAAACUGCUUA	UUAGAGAAUAAUGUGCUGGCU	UUCUCUGACUUAGCAUAUAGU
UAGGUUUCAGCAUAAUUUGGG	UUCUCUGAGAUAAACUGGCUG		
UAGGUUUCCAGGAACAUGCCG	UUCUCUGAUACUGUCUUUCUG		
UAGGUUUCGGACAUGGCCGUC	UUCUCUGAUGGUUCCACAGUU		
UAGGUUUCUCUAGCUCUUCUG	UUCUCUGAUUUAUUGAUCGGC		

infection-enhancing seeds *(library inventory oligos)*

UAUAUGUUCAUUAACACUGCC	UAUAUGUUUAUCUUCUCUGUG	UAUAUGUUAGGAGAAAUCAGG	UAUAUGUUGAGAAAAUGAAGA
UAUAUGUUGAACUAGAGUUUG	UAUAUGUUUCUGUAAAUCCAG		

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.

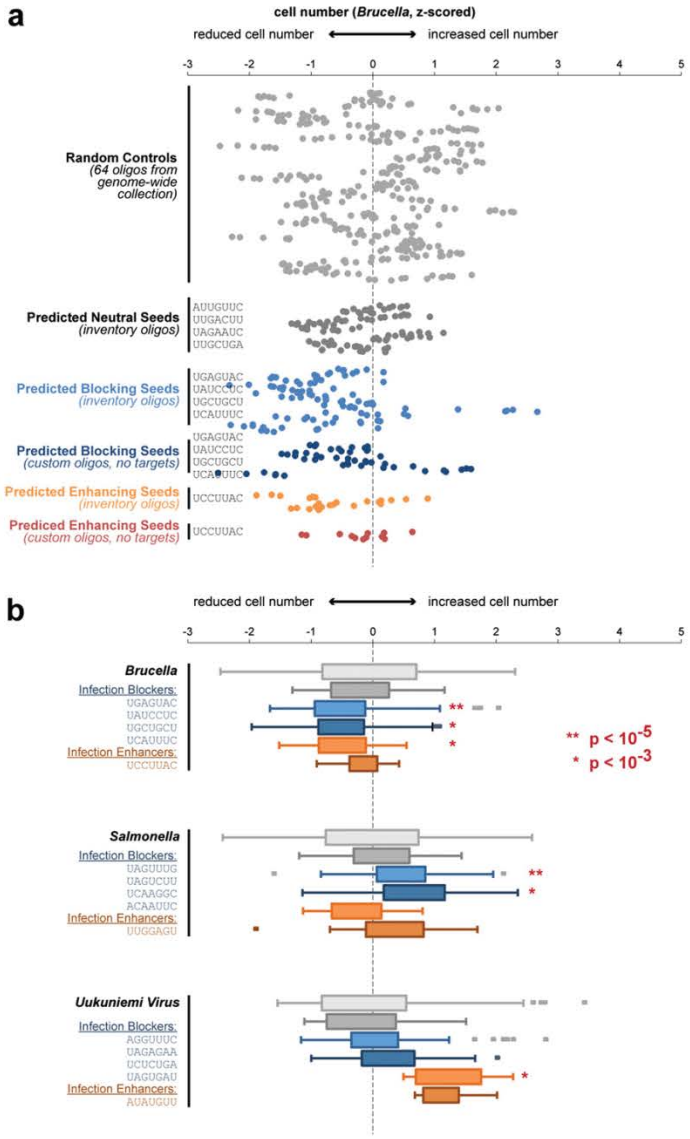


Fig. S4. Cell-number phenotypes of selected seeds. This figure corresponds to Fig. 3, except that the phenotype is now the cell number (not the infection). The observed cell-number phenotypes are much less pronounced than the infection phenotypes, and often insignificant.

	selected seed sequence	seed occurrence in library	seed phenotype (average)	corresponding human miRNA	mature human miRNA sequence	observed miRNA overexpr. phenotype
predicted to block infection	UAGGGAG	10	-1.065	hsa-miR-3162	UUAGGGAGUAGAAGGGUGGGGAG	-0.611
	UGAUAAG	12	-1.070	hsa-miR-3161	CUGAUAAGAACAGAGGCCCAGAU	-1.163
	UAAAGCU	13	-1.474	hsa-miR-9*	AUAAAGCUAGAUAACCGAAAGU	-0.985
	AGAUGUG	13	-1.076	hsa-miR-576-3p	AAGAUGUGGAAAAAUUGGAAUC	-1.241
	AGUAGAC	14	-1.089	hsa-miR-411	UAGUAGACCGUAUAGCGUACG	-0.936
	UGCACUU	24	-1.224	hsa-miR-4325	UUGCACUUGUCUCAGUGA	-1.122
	UCUCCUC	30	-1.349	hsa-miR-4279	CUCUCCUCCCGGCUUC	-1.086
	UCAGGGC	35	-1.066	hsa-miR-3127	AUCAGGGCUUGUGGAAUGGGAAG	-0.916
predicted to enhance infection	UCUGGAG	17	1.355	hsa-miR-516b	AUCUGGAGGUAAGAAGCACUUU	1.361
	UGACUGA	20	1.611	hsa-miR-3136	CUGACUGAAUAGGUAGGGUCAUU	2.387
	UGAAAUG	23	2.098	hsa-miR-203	GUGAAAUGUUUAGGACCACUAG	2.903
	UGUUGAA	23	3.281	hsa-miR-653	GUGUUGAAACAAUCUCUACUG	3.323
	UGUACAU	25	2.279	hsa-miR-493*	UUGUACAUGGUAGGCUUUCAUU	0.553
	UGUGCUU	28	2.166	hsa-miR-218	UUGUGCUUGAUCUAACCAUGU	0.965
	UGACUGU	54	2.665	hsa-miR-943	CUGACUGUUGCCGUCCUCCAG	2.433
	UUCUUGU	86	1.413	hsa-miR-578	CUUCUUGUGCUCUAGGAUUGU	1.282
predicted to have no effect	AAAGUUU	36	-0.227	hsa-miR-561	CAAAGUUUAAGAUCCUUGAAGU	-0.175
	UUCAAAU	36	0.017	hsa-miR-607	GUUCAAAUCCAGAUCUAUAAC	-0.170
	UUGGUUC	41	0.079	hsa-miR-659	CUUGGUUCAGGGAGGGUCCCCA	0.034
	AAAUGAA	46	0.290	hsa-miR-3646	AAAAUGAAAUGAGCCCAGCCCA	0.537
	UUAACAU	51	0.011	hsa-miR-302c*	UUUAACAUGGGGGUACCUGCUG	-0.233
	UUAGGAU	53	-0.128	hsa-miR-651	UUUAGGAUAAGCUUGACUUUUG	0.077
	AUAUUAU	55	-0.178	hsa-miR-656	AAUAUUAUACAGUCAACCUCU	0.272
	UUAAGAA	55	0.068	hsa-miR-3658	UUUAAGAAAACACCAUGGAGAU	0.407

Fig. S5. Overexpression phenotypes of endogenous miRNAs. This figure provides further details on the data shown in Fig. 4. Listed are all 24 endogenous human miRNAs that have been selected to be tested for their ability to affect *Brucella* cell entry. To be selected, human miRNAs had to contain a seed sequence that had already been characterized in the genome-wide RNAi screen, occurring in that screen independently in at least 10 siRNA oligos directed at protein-coding genes. Eight miRNAs each were selected to correspond to the top seeds shown to block infections, enhance infections, or to have no influence on infections (neutral). miRNAs that were a part of larger families sharing the same seed sequences were not considered. All miRNAs were transfected into HeLa cells as "*miRIDIAN microRNA Mimics*" (Thermo Scientific). Their effect on *Brucella* infections is shown in the last column, as the average of four replicate experiments. The observed phenotypes corresponded very well to the predicted outcome (see graph in Fig. 4).

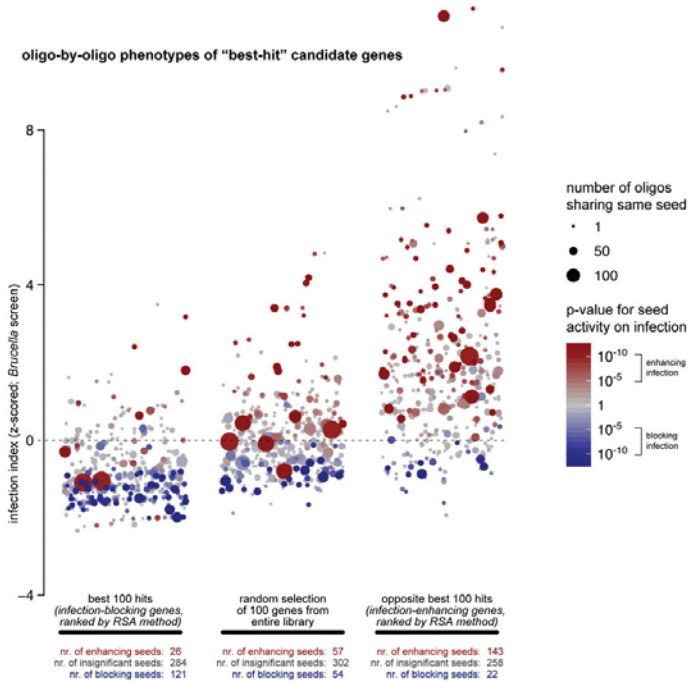


Fig. S6. Influence of the seed effect on "best hit" lists. For the genome-wide *Brucella* screen, all genes were ranked according to phenotypic strength and consistency by the widely used "Redundant siRNA Activity" ("RSA") method (1), and the best hits are shown here (together with a random set for comparison). 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.

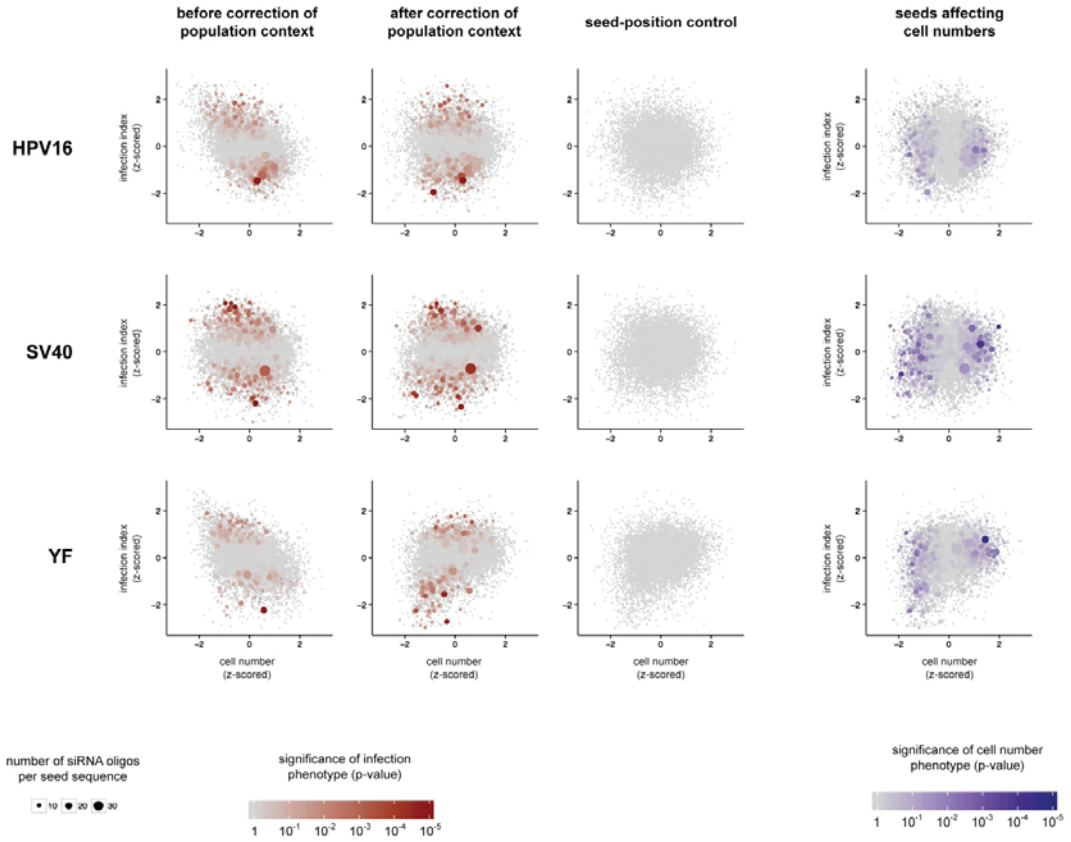


Fig. S7. Off-target effects and cellular population context. Of the seven "druggable genome" screens that Snijder et al. had reported on earlier (1), the three with the strongest seed effects are shown here. In each plot, the data points represent one heptameric seed sequence each, showing the averaged phenotypes over all siRNA oligos that happen to share that seed. The color code indicates the statistical significance of the observed phenotypes (red, infection phenotype; blue, cell number phenotype). Column 3 shows the negative control for seed effects, by incorrectly assuming the seed position at residues 12–18 in each oligo (similar to Fig. 2). In all columns except the first column, the infection phenotypes have been corrected for their "population context": i.e., for subtle dependencies between the observed infection index and other cellular phenotypes (such as absolute number of cells, local clustering of cells, and aspects of the cellular morphology). For all three screens, population context correction fails to diminish the off-target seed effects on infection. Note that these data are not from genome-wide screens, but from druggable genome screens only; therefore, there is a somewhat reduced statistical power in detecting seed effects. HPV16, human papilloma virus 16 (*Papillomaviridae*); SV40, simian virus 40 (*Polyomaviridae*); YF, yellow fever virus (*Flaviviridae*).

1. Snijder B, et al. (2012) Single-cell analysis of population context advances RNAi screening at multiple levels. *Mol Syst Biol* 8:579.

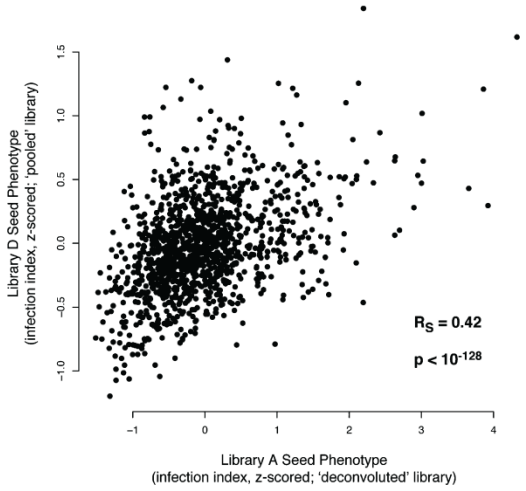


Fig. S8. "Pooled" versus "deconvoluted" libraries. Each data point represents one seed, restricted here to seeds that are represented in both libraries at least five times. The position of each data point represents the average infection phenotype of the oligos that contain a given seed, in both libraries (screening data are from *Brucella abortus*). For the pooled library, the phenotype of any given oligo corresponds strictly to the phenotype of the entire pool that it is contained in.

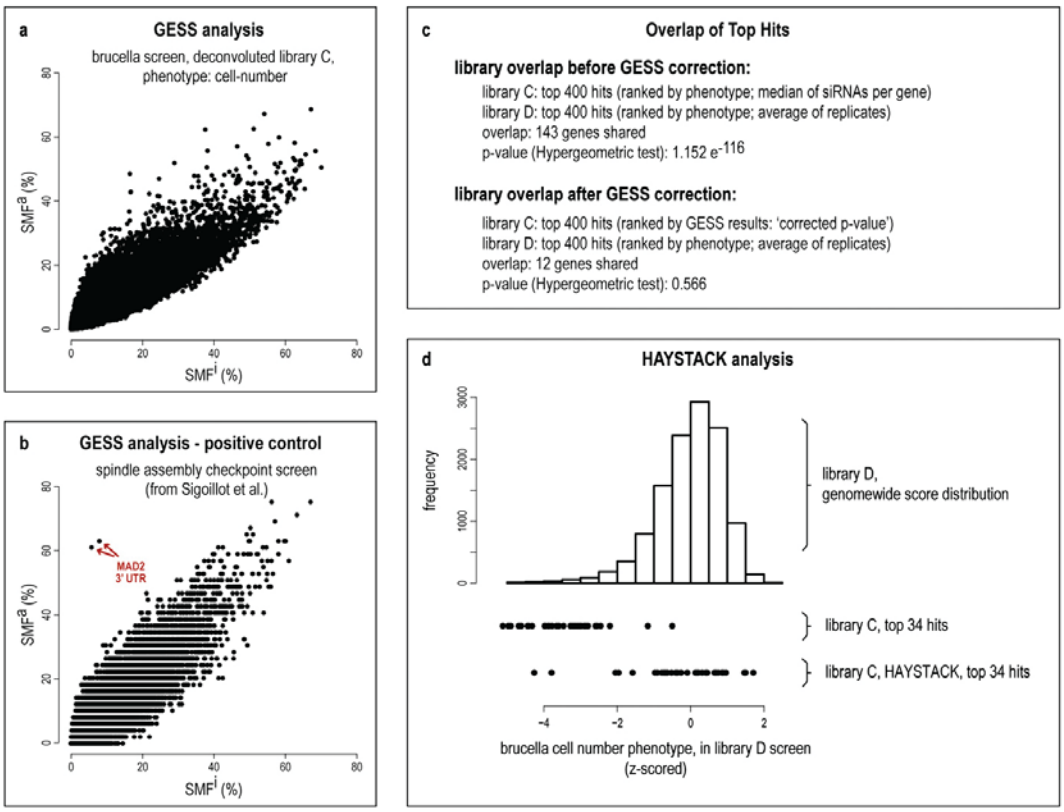


Fig. S9. *GESS* and *Haystack* analysis of off-target effects. The software packages *GESS* ("Genome-wide Enrichment of Seed Sequence Matches") (1) and *Haystack* (2) have been demonstrated to be capable of selecting true positive hits from genome-wide RNAi screens, at least in some instances. They were 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. SMF^i] 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. Sigoillot 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 gene(s) sharing same seed sequence	TargetScan: number of predicted target genes	TargetScan: overlap with primary screens	TargetScan: expected overlap by chance	miRmap: number of predicted target genes	miRmap: overlap with primary screens	miRmap: expected overlap by chance	miRTar: number of predicted target genes	miRTar: overlap with primary screens	miRTar: expected overlap by chance
Brucella Infection Reduction	UAUACAG	-1.33	hsa-let-7f-2-3p	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-4460	209	4	8.1	158	7	6.1	-	-	-
	UUCUGCA	-1.31	hsa-miR-544a	224	6	8.7	280	9	10.9	-	-	-
	UCUGCUG	-1.18	hsa-miR-6071	292	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	UUGGUCC	-1.30	hsa-miR-133a-3p	364	18	14.1	286	8	11.1	51	1	2.0
	AAGGCAU	-1.32	hsa-miR-3910	156	3	6.1	373	12	14.5	-	-	-
	UCCUUGC	-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-5587-5p	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-1273h-3p	167	5	6.5	-	-	-	-	-	-
	UAAUGCU	2.28	hsa-miR-155-5p	91	2	3.5	51	1	2.0	829	34	32.1
	UGAAAUG	2.10	hsa-miR-203a	1	0	0.0	76	1	2.9	23	3	0.9
	UGUGCUU	2.16	hsa-miR-218-5p	202	8	7.8	316	14	12.2	88	4	3.4
	UGACUGA	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	-	-	-
	UGUGGAU	3.63	hsa-miR-4752	152	9	5.9	156	4	6.0	-	-	-
	UUGCCAU	3.31	hsa-miR-4776-3p	239	15	9.2	228	8	8.8	-	-	-
	UGUACAU	2.28	hsa-miR-493-5p	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	-	-	-
	ACAAGGU	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-6833-5p	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	-	-	-	-	-	-
	UCUGUCU	2.57	hsa-miR-6888-3p	149	2	5.8	-	-	-	-	-	-
	UUCCUGU	1.71	hsa-miR-7152-5p	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	AAAUGAG	1.94	hsa-miR-1252-3p	9	0	0.3	-	-	-	-	-	-
	UGAAAUG	2.38	hsa-miR-203a	1	0	0.0	76	1	2.9	23	3	0.9
	UGUGCUU	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	-	-	-
	AGAACAG	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	-	-	-
	UGUAUGC	2.46	hsa-miR-675-3p	141	8	5.5	48	3	1.9	-	-	-
	UGUGGAA	1.71	hsa-miR-6833-5p	175	11	6.8	-	-	-	-	-	-
	UAAAGCU	2.15	hsa-miR-9-3p	46	2	1.8	35	1	1.4	2	0	0.1
	UUGGCAC	2.19	hsa-miR-96-5p	311	9	12.0	472	14	18.3	100	5	3.9

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 "top hits" 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 given 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 mirTar, we 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.

3.3 RESEARCH ARTICLE III (IN PREPARATION)

MicroRNAs miR-103 and miR-107 promote infection of *Brucella* and other intracellular pathogens in non-phagocytic cells

Alain Casanova [1], Shyan Huey Low[1], Mario Emmenlauer[1,4] , Pauli Rämö [1], Alexander Schmidt[3], Nitish Mittal[2], Andreas Gruber[2], Mihaela Zavolan[2], Christoph Dehio[1]

Manuscript in preparation

3.3.1 Statement of own contribution

miRNA mimics and genome-wide siRNA screens for *Brucella* infection were conducted together with Dr. Shyan H. Low. Entry assays, MEF infection, and the transferrin uptake assay were performed by myself. Transciptomics data were obtained in collaboration with Dr. Nitish Mittal and Andreas Gruber. Proteomics analysis was performed together with Dr. Alexander Schmidt. Mario Emmenlauer devised image analysis pipelines and Dr. Pauli Rämö calculated the data normalization. TGF-β stimulation and TGF-β receptor overexpression assays were performed by Dr. Shyan H. Low. Infection assays with pathogens other than *Brucella* were conducted by the following people: Salmonella (Dr. Saskia Kreibich, Prof. Dr. Wolf-Dietrich Hardt; ETH Zurich), Bartonella (Dr. Simone Eicher, Prof. Dr. Christoph Dehio; Biozentrum Basel), Listeria (Dr. Andreas Kühbacher, Dr. Javier Pizarro-Cerda, Prof. Dr. Pascale Cossart; Institut Pasteur Paris), adenovirus and rhinovirus (Dr. Daria Seiler, Bettina Cardel, Dr. Artur Yakimovich, Prof. Dr. Urs Greber; University of Zurich). Figures were generated by myself except for Figure 8 and Supplementary Figures S1-S3, which were created by Dr. Shyan H. Low and Andreas Gruber, respectively. The manuscript was written by me.

3.3.2 Manuscript

MicroRNAs miR-103 and miR-107 promote infection of *Brucella* and other intracellular pathogens in non-phagocytic cells

Alain Casanova [1], Shyan Huey Low[1], Mario Emmenlauer[1,4] , Pauli Rämö [1], Alexander Schmidt[3], Nitish Mittal[2], Andreas Gruber[2], Mihaela Zavolan[2], Christoph Dehio[1]

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

[2] Focal Area Computational & Systems Biology, Biozentrum, University of Basel, Basel, Switzerland

[3] Proteomics Core Facility, Biozentrum, University of Basel, Basel, Switzerland

[4] Mario Emmenlauer BioDataAnalysis, Munich, Germany

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

MicroRNAs are small non-coding RNAs that regulate gene expression in eukaryotic cells. It has become evident in recent years that miRNAs regulate central processes involved in viral and bacterial pathogenesis. In order to identify miRNAs that are involved in *Brucella* infection, we performed a functional miRNA mimic screen for *Brucella* entry, trafficking, and replication in HeLa cells. This revealed miRNAs miR-103 and miR-107 (miR-103/107) as positive regulators of *Brucella* invasion of non-phagocytic cells. The infection of other intracellular pathogens, such as *Salmonella* Typhimurium and adenovirus, were also enhanced by miR-103/107 emphasizing their importance in pathogen infection. Proteome and transcriptome studies of cells transfected with miR-103/107 mimics revealed an increase of cell surface components as well as an altered expression of factors involved in endocytic processes. In agreement with a change in endocytic properties, cells transfected with miR-103/107 mimics showed a decrease in clathrin-mediated uptake of transferrin. Depletion of several components of clathrin-mediated endocytosis (CME) was found previously to enhance *Brucella* infection. This supports the notion that CME negatively affects *Brucella* infection. Furthermore, it was found that TGF-β receptor 2, a host factor required for several intracellular pathogens, is upregulated by miR-103/107. Overexpression of TGF-β receptors or stimulation of the pathway by TGF-β was enhancing *Brucella* infection, further confirming the importance of this pathway. Altogether, this suggests that miR-103/107 controls several components of the host cell in favor of *Brucella* invasion and infection with other intracellular pathogens.

Introduction

MicroRNAs (miRNA) are important post-transcriptional regulators of eukaryotic gene expression. They negatively affect gene expression by translational inhibition and destabilization of target mRNAs. Most mammalian miRNAs are generated from short transcripts that fold back on themselves to form specific hairpin structures. These are consecutively cleaved by the endonucleases drosha and dicer and are ultimately incorporated in the RNA induced silencing complex (RISC) (*1*). miRNAs mainly act by partial binding of the miRNA to the 3' untranslated region (UTR) of target mRNAs via Watson-Crick base paring. The binding specificity of a miRNA is largely determined by complementarity of the so-called seed region to the target mRNA (*2*). The seed region comprises a stretch of 6-8 nucleotides located at the 5' end of the miRNA. However, apart from canonical binding sites, some miRNAs were found to bind to the 5' UTR or coding regions of mRNAs (*3, 4*) and in some cases miRNA-mRNA interactions are determined by base-paring outside the seed region (*5*).

miRNAs play a central role as regulators of almost any cellular process ranging from reprogramming of cells during differentiation and development to the maintenance of cell homeostasis and cell cycle control. It is thus not surprising that also functions related to pathogen infection are under the control of miRNAs. Immune cells have been described to respond to pathogen infections by differentially regulating several miRNAs in order to fine-tune the immune response (*6*). Further, RNA interference (RNAi) is directly involved in viral degradation in certain cell types (*7*). In return, pathogens make use of host miRNA pathways to their own benefit. Viral pathogens have evolved multiple strategies to interfere with the host RNAi machinery or they even encode their own miRNAs (*8*). Bacteria were found to manipulate cellular miRNA expression by secretion of effector proteins to manipulate host cell functions such as the cell cycle or cell differentiation (*9, 10*).

Pathogenic bacteria and viruses have evolved to survive inside mammalian cells with a varying degree of dependence on host factors. While all viruses are obligate intracellular organisms, many intracellular bacteria are able to survive outside of a host cell. However, an intracellular lifestyle offers protection from host immune components such as the complement system or antibodies and provides resources favorable for growth and replication. Host cells in return have adapted to the threat of pathogen invasion by numerous cellular functions that allow the detection of pathogen-specific molecules or damage associated with the attack (*11*). Depending on the

infection strategy of the pathogen the impact on cell viability can differ considerably. Pathogens that cause acute infections often lead to the death of infected host cells, which is associated with rapid proliferation and spread of the pathogen. In contrast, pathogens that cause chronic infections often protect their host cell from cell death to establish a long lasting niche for their survival and replication (*12*).

Independent of the infection strategy, all intracellular pathogens need to gain access to the interior of the cell. Bacteria and viruses have evolved to make use of the full spectrum of endocytic pathways that a cell offers. These include macropinocytosis as well as clathrin- and caveolin-dependent and -independent mechanisms (*13*). Some pathogens rely on signaling induced on the cell surface by direct or indirect interaction of pathogen molecules and cell surface receptor. This applies to all viruses as well as bacterial pathogens that enter cells by a zipper-like mechanism (e.g.: *Listeria monocytogenes*). Other bacterial pathogens make use of specialized secretion systems to deliver effector proteins into the host cell that trigger their uptake. Examples are *Salmonella* and *Shigella* that induce large, actin-rich membrane ruffles on the cell surface which resemble macropinosomes (*14*).

Brucella species are facultative intracellular pathogens that infect domestic animals as their natural hosts. Transmission to humans is most commonly caused by direct contact with infected animals or by ingestion of contaminated food and can lead to severe chronic infections (*15*). *Brucella* is able to invade phagocytic and non-phagocytic cells and replicates within an intracellular compartment known as the *Brucella*-containing vacuole (BCV). Following the internalization of bacteria into a eukaryotic host cell, the BCV is transported along the endocytic pathway and, despite the interaction with lysosomes, bacteria manage to avoid degradation (*16*). During the phase of trafficking and until *Brucella* establishes an endoplasmic reticulum (ER)-like replicative niche (*17*), bacteria do not divide. During intracellular proliferation, BCVs are found in close association with ER markers such as calnexin and glucose-6-phosphatase (*18*).

Despite the importance of miRNAs in pathogen infection, there is very little knowledge around as to what impact miRNAs might have on *Brucella* infection. We thus screened a library of human miRNA mimics for their effect on *Brucella* infection of HeLa cells. By this, several miRNAs including miR-103 and miR-107 (miR-103/107) were identified that enhance *Brucella* infection of non-phagocytic cells. Furthermore, these miRNAs also enhanced infection of HeLa cells by different bacterial and viral pathogens. To investigate which cellular processes are regulated by miR-103/107, we measured changes of the proteome and transcriptome induced by high levels of

miR-103/107. Several cell surface factors, partially already described to be important during the infection process of certain intracellular pathogens, were identified. Among those was TGF-β receptor 2 (TGFBR2) which showed increased protein levels upon transfection of miR-103/107. This kinase had been independently identified as crucial for *Brucella* infection in a siRNA-mediated mRNA-depletion study (*research article IV*), where cells were investigated under the same experimental conditions as used for the miRNA mimic screen. We could show here, that overexpression of the receptor or activation of the pathway by addition of the ligand indeed promotes *Brucella* infection. We speculated that altered endocytosis or recycling of cell surface receptors could underlie the observed changes of several cell surface proteins. In agreement with this hypothesis CME of transferrin was reduced in miR103/107-treated cells. It will be interesting to further dissect the underlying principle of how the increased miR-103/107 levels enhance the susceptibility of non-phagocytic cells to *Brucella* infection and to link these findings to cell surface receptors and the endocytic trafficking routes.

Results

Screening for miRNAs identifies miR-103 and miR-107 to promote *Brucella* infection

In order to study the effect of human miRNAs on *Brucella* infection we performed a high-content high-throughput miRNA mimic screen in HeLa cells covering over 1200 human miRNAs. HeLa cells were transfected for three days with miRNA mimics followed by infection with GFP-expressing *Brucella abortus*. After bacterial internalization for 4 h the remaining extracellular bacteria were inactivated by gentamycin and the infectious process was continued for a total of 44 h. This endpoint assay allows the identification of miRNAs involved in *Brucella* entry, trafficking, and replication. It was observed that many miRNAs showed an effect on *Brucella* infection (Fig. 1A). The strongest up and down hits are summarized in Table 1 (a complete list can be found in Suppl. Table 1).

miR-103/107 was found to cause a striking increase in *Brucella* infection (Fig. 1A, 1B). These miRNAs belong to a conserved miRNA family and share an identical sequence except for a single nucleotide difference at position 22. Even though miRNAs preferentially interact with their mRNA targets by base pairing in the seed region, different non-canonical miRNA-mRNA interactions involving base-pairing outside the seed have been found (5). These options cannot be readily distinguished for miR-103 and miR-107, since they share most of their sequence. We

previously performed siRNA screens using the same experimental setup as for the present miRNA mimic screen and identified prominent seed-driven off-target effects (*19*). We found that siRNAs and miRNA mimics that share the same seed sequence also affect *Brucella* infection in a comparable manner. We thus searched for siRNAs with a seed sequence identical to the one of miR-103/107 (GCAGCAU) and found two oligos (Fig. 1C, highlighted in bold). Both siRNAs showed a strong positive effect on *Brucella* infection similar to the one observed for miR-103/107. These oligos belonged to a genome-wide siRNA library from Qiagen with four siRNAs per target gene. The other three siRNAs of the Qiagen library designed against the same targets (MMP13 and CTAG1B) did not show an effect on infection, indicating that depletion of these genes does not affect *Brucella* infection *per se* (data not shown). This suggests that the two siRNAs act like synthetic miR-103/107 miRNAs. Since these oligos only share the seed sequence but not the rest of the sequence with miR-103/107 we concluded that the seed sequence determines binding specificity for those targets that promote *Brucella* infection.

Interestingly, the third strongest up hit, miR-1184, shares with miR-103/107 a GCAGC motif in its seed sequence which is shifted within the seed by two positions (Fig. 1C). Two miRNAs with an identical sequence motif within the seed which is shifted by one or several positions could in principle bind to the same mRNA if the nucleotides on the mRNA which flank the sequence motif are complementary to both miRNA seeds. This potentially allows miR-1184 and miR-103/107 to regulate a partially overlapping set of target genes. Even though the effect of miR-1184 was not as pronounced as for miR-103/107, it is conceivable that some of the cellular targets underlying the observed increase in infection might be shared. Looking at all available siRNAs and miRNA mimics of our screens that contain a GCAGC motif in their seed we found a number of similar sequences that increase *Brucella* infection (Fig. 1C). An alignment of all seed sequences with a Z-scored infection index ≥ 4 identified a UGCAGC motif (Fig. 1D). Even though the complete sequence space of possible seeds containing a GCAGC motif is not fully covered in our siRNA screens, it is evident that certain seed families such the well represented miR-15 family (miR-15a, miR-15b, miR-16, miR-195, miR-424, miR-497) do not regulate *Brucella* infection.

miR-103 and miR-107 promote *Brucella* entry into non-phagocytic cells

To better characterize the effect of miR-103/107 during *Brucella* infection, we performed an image-based follow-up assay to study the effect of these miRNAs on bacterial entry (Fig. 2A).

This assay makes use of a tetracycline-inducible GFP to selectively mark intracellular bacteria after infection. Following incubation of bacteria with host cells, extracellular bacteria are inactivated by the cell-impermeable antibiotic gentamycin. At the same time the non-toxic tetracycline analog, anhydrotetracycline (aTc), is added. Since aTc is cell-permeable, it will induce GFP expression in intracellular bacteria which are not inactivated by gentamycin. This results in GFP expression solely in viable intracellular bacteria. We found that cells treated with either a miR-103 mimic or siRNAs sharing the same or a similar seed sequence were more strongly infected by *Brucella* in the entry assay (Fig. 2B-2E). The number of cells containing intracellular *Brucella* as well as the bacterial load measured by the integrated GFP intensity of intracellular bacteria in infected cells was increased. This shows that bacterial entry is enhanced by high levels of miR-103/107.

Since the miR-103/107 family is evolutionary conserved in vertebrates (*20*), we were interested to see whether similar effects on infection would be observed in species other than human. To this end, we repeated the experiment in mouse embryonic fibroblasts (MEFs). The transfection efficiency as measured by cell death induced by transfection of an siRNA targeting kif11 was not as high as for HeLa cells (Fig. 3A). Despite the reduced transfection efficiency, the effect of the positive control, an ARPC3 depletion to interfere with actin filament networks, led to a detectably lower infection with *Brucella*. We thus went on to test the effect of miR-103 mimics in MEF cells and preliminary results (n=2) indicate an increase in the intracellular bacterial load for all infection intervals tested. Enhanced infection relative to the scrambled control siRNA (non-targeting control) was observed as early as 8 h after infection and remained constant up to 48 h indicating that also in MEF cells miR-103 promotes *Brucella* uptake.

Infection of different intracellular bacterial and viral pathogens is promoted by miR-103/107

Since all intracellular pathogens have to enter host cells, we wondered whether the miR-103/107 family positively affects the infection of other intracellular pathogens. Similar miRNA mimic screens using largely standardized protocols were performed on a number of different intracellular viral and bacterial pathogens (full screens will be published elsewhere). These pathogens use a wide variety of infection strategies throughout their life cycle. Entry into host cells is achieved by a trigger mechanism (*Salmonella* Typhimurium), zipper mechanism (*Listeria monocytogenes*), invasome formation (*Bartonella henselae*), or receptor-mediated endocytosis

(adenovirus, rhinovirus) (*13*). Additionally, each screen covers further aspects of the intracellular lifecycle of the corresponding organism. With the exception of *Bartonella henselae* invasome formation (*21*) we consistently observed an increase in infection for *S.* Typhimurium, *L. monocytogenes*, adenovirus, and rhinovirus (Fig. 4). Even though most effects were not as pronounced as for *Brucella*, it shows that this family of miRNAs alters host cell function in favor of infection with the tested pathogens.

Identification of proteome and transcriptome changes induced by miR-103/107

To gain a comprehensive understanding of the cellular changes induced by miR-103/107 we performed proteome as well as transcriptome analyses of HeLa cells transfected with miRNA mimics. In a first approach we measured the proteome changes induced by miR-103 or miR-107 compared to scrambled controls. We were able to identify over 7'000 proteins across samples. No differentially expressed proteins could be identified when comparing the samples from miR-103 and miR-107 treated cells (Suppl. Table 3) and shows that the effect of both miRNAs on the detectable proteome is largely identical. In contrast, about 500 proteins were differentially expressed in miR-103/107 treated cells compared to the scrambled control when a false discovery rate below 1% was used as a threshold (Fig. 5A).

For the transcriptomics approach, miR-103 and scrambled treated HeLa cells were compared and over 14'000 mRNA transcripts could be identified by Illumina sequencing. Since miRNAs act by direct binding to target mRNAs leading to inhibition of translation and a destabilization of the mRNA we expected to observe similar changes for direct targets using both approaches. Indeed the overall correlation of mRNA and protein changes induced by miR-103 compared to scrambled-treated cells was good and an even higher agreement was found for predicted direct targets of miR-103 (Suppl. Fig. S1). Both approaches showed that miR-103 targets predicted by TargetScan (*22*) or Elmmo (*23*) software are predominantly down-regulated (Fig. 5A, Suppl. Fig. S2, Suppl. Fig. S3), which is agreement with the action of miRNAs. Consistent with previous reports we could identify DICER and CDK6 to be strongly down-regulated (*24, 25*).

miR-103/107 alters protein abundance of cell surface components and affects cell cycle

To obtain a global picture of the changes induced by miR-103/107 we performed pathway analysis with those factors of the proteomics study which were significantly changed. This revealed a significant enrichment of upregulated proteins with terms associated to the plasma

membrane, enzyme-linked receptor signaling, and cell adhesion (Fig. 6A). A similar picture was observed when string analysis was used to identify interacting proteins among the altered factors. The strongest clusters of upregulated proteins were EGF signaling, TGF-β signaling, as well as adhesion and extracellular matrix (Fig. 7) supporting the notion that the cell surface undergoes changes when miR-103/107 levels are increased. For proteins down-regulated upon miR-103/107 treatment in turn, several components associated with microtubules and vesicular transport came up and particularly a very strong effect on cell cycle components was found. Gene ontology terms (GO terms) such as DNA replication, cell cycle, and DNA metabolic processes were strongly enriched (Fig. 6B). String analysis revealed interactions among down-regulated proteins involved in cell cycle regulation such as the cyclin-dependent kinases CDK4 and CDK6 which both promote G1/S phase transition. Furthermore, all six components of the DNA helicase (MCM2 - MCM7) and all four subunits of DNA polymerase delta (POLD1 - POLD4) showed reduced protein levels. Likely related to an influence on the cell cycle, a treatment with miR-103/107 led to a mild negative effect on cell number after three days of transfection and the effect became more prominent with a transfection-duration of five days (Suppl. Fig. S4).

TGF- β receptor 2 is upregulated by miR-103/107 and promotes *Brucella* infection

In order to narrow down candidate genes that could account for the increased susceptibility to *Brucella* infection caused by miR-103/107, we used data from a genome-wide siRNA screen for *Brucella* infection that was previously performed in the lab (*research article V*). Overlay of the top 500 up- and down -hits from the screen with proteomics data identified eight candidate genes that increase infection when depleted by siRNAs and that were also down-regulated by miR-103/107 (with decreasing strength: NFIB, KIAA0664, SNX1, ASNA1, PDE8B, SDF2, MAP4, PEX14). Among the proteins upregulated by miR-103/107 we found eight genes that had a negative effect on infection, suggesting that increasing these host factors might benefit *Brucella* infection (with decreasing strength: TGFBR2, SLC3A2, STK4, STMN3, ADO, LAMB1, UBE2D1, CDC34). Interestingly, TGFBR2 was also found among the top hits in a comparative kinome-wide siRNA screen covering several intracellular pathogens (*research article IV*). There we found that depletion of TGFBR2 as well as TGFBR1, which form a functional heterotetramer during TGF-β signaling, reduced infection of several pathogens whose infection is promoted by miR-103/107. This suggests that increased TGFBR2 levels could promote pathogen infection observed here. Indeed, overexpression of wild-type TGFBR2 or TGFBR1 in HeLa cells led to a

mild yet significant increase in *Brucella* infection, while dominant negative alleles caused a reduction in infection. Furthermore, we found that activation of the pathway by addition of the ligand TGF-β1 also promoted infection (Fig. 8). Altogether, these data indicate that elevated levels of TGFBR2 could account partially for the increased susceptibility of cells to infection when levels of miR-103/107 are increased.

Clathrin-mediated endocytosis is reduced by miR-103/107

Since several cell surface components that undergo CME, including EGFR and TGFBR2, were upregulated by miR-103/107, we wanted to assess whether the rate of CME is altered (*26*, *27*). To this end, HeLa cells were transfected with a miR-103 mimic for three days and incubated with fluorescently labeled transferrin (Tf). Comparison to cells transfected with a non-targeting control oligo revealed that miR-103/107 reduced Tf-uptake (Fig. 9). As the levels of transferrin receptor were not significantly altered by miR-103/107 (miR-103/107 vs. control: log2=0.24) we conclude that the CME-dependent Tf-uptake is indeed reduced in presence of elevated miR-103/107 levels. Interestingly, we previously found that siRNA-induced depletion of several components involved in either the formation or the uncoating of clathrin coated pits (AP2S1, AP2A1, CLTC, EPN1 (*research article V*) and GAK (*research article IV*), respectively) enhance *Brucella* infection. There is thus independent evidence that reduced CME benefits *Brucella* infection.

Discussion

miRNAs are well studied regulators of gene expression and have gained increasing importance in contributing to cellular traits relevant for host-pathogen interactions. Here, we tested a library of over 1200 human miRNAs for their role in *Brucella* infection of HeLa cells. This identified candidate miRNAs involved in entry, trafficking, or replication of this intracellular pathogen and presents a rich resource for further investigations.

We identified two closely related microRNAs, miR-103 and miR-107, that strongly promote *Brucella* infection and we could show that already the stage of bacterial entry is enhanced by increased levels of these miRNAs. The family of miR-103/107 miRNAs is broadly conserved across species and accordingly, a similar reduction in *Brucella* entry upon a raise of miR-103/107 could be observed in mouse embryonic fibroblasts. This suggests that not only the miRNAs are

conserved, but also the targets which underlie the increased susceptibility of non-phagocytic cells to *Brucella* infection when miR-103/107 levels are high.

Interestingly, infection of several other intracellular viral and bacterial pathogens was also positively affected by elevated levels of miR-103/107. These pathogens use very distinct infection strategies and each of the assays covered aspects of the infection cycle which go beyond invasion of the host cell. Therefore, several scenarios are possible that would explain how one family of miRNAs could affect such diverse pathogens. Either, the miRNAs affect multiple cellular functions that promote a different relevant step for each pathogen or a single cellular mechanism, which benefits all pathogens, is altered. If the latter is the case, it is likely that the entry process is affected. Indicative for this is first our finding that this is the stage *Brucella* infection is promoted and second, because all intracellular pathogens share the need of cell invasion.

To shed more light on the cellular changes induced by high levels if miR-103/107, we analyzed transfected cells for the changes induced on the proteome and transcriptome levels. Both methods showed good correlation, especially for predicted direct targets of miR-103/107. This indicates that the miRNAs mainly act via the classical pathways involving translational inhibition as well as destabilization of target mRNAs. Proteins deregulated by miR-103/107 populated distinct cellular pathways when analyzed by Gene enrichment and protein-protein interaction algorithms. Several cell surface-located factors were found to be upregulated. Interestingly, these include the transmembrane kinase TGFBR2, which was previously found to be involved in the infection of several intracellular pathogens (*research article IV*). In that study, siRNA depletion of TGFBR2 was found to decrease pathogen infection. Here, we confirmed that increasing TGF-β receptor levels or activation of the pathway by addition of the ligand TGF-β1 enhances *Brucella* infection. Even though the effects were not as pronounced as seen for transfection of miR-103/107, this shows that elevated TGFBR2 levels could partially account for the observed effect. It remains to be shown if the TGF-β pathway also accounts for the facilitated entry of other pathogens which showed enhanced infection by miR-103/107 transfection.

Arguing that a change in endocytic properties of a cell could benefit a broader spectrum of pathogen entry processes, we investigated the effects of miR-103/107 on CME. We found that transferrin uptake was negatively affected in cells transfected with miR-103/107. At this point it remains open whether transferrin uptake is reduced or recycling to the cell surface is increased. Since TGFBR2 as well as EGFR are endocytosed in a clathrin-dependent manner, any of the

above-mentioned scenarios could result in a net increase of the protein due to altered degradation. Along these lines, we found that additional cell surface proteins relevant for other pathogens were also upregulated. MET is mildly, but significantly, upregulated and is the only receptor for *Listeria monocytogenes* entry in HeLa cells that lack E-cadherin (*28*). Furthermore, we found receptors involved in adenovirus entry to be upregulated by miR-103/107. These include Integrinβ3 (ITB3) (*29*), which was among the proteins with the strongest increase, as well as coxsackie and adenovirus receptor (CXDAR) which was mildly upregulated (*30*). Altogether, we provide preliminary evidence for the hypothesis that altered miR-103/107 might have a broad impact on endocytic processes which, in consequence, might result in changed cell surface display of different pathogen receptors. It will be interesting to verify this further in context of the respective organisms.

While high levels of miR-103/107 increased protein levels of all above-mentioned surface receptors, they are unlikely to be directly targeted by the miRNAs, since this generally results in a negative regulation. Among the 8 candidate genes that promoted *Brucella* infection in a genome-wide siRNA screen and also showed reduced protein levels in the presence of miR-103/107, three are predicted to be direct targets. These are nuclear factor I B (NFIB), phosphodiesterase 8B (PDE8B), and microtubule-associated protein 4 (MAP4). NFIB is a site-specific transcription factor involved in positive and negative regulation of various cellular processes from stem cell biology to brain development (*31, 32*). PDE8B is involved in regulation of cAMP signaling by catalyzing the hydrolysis of cAMP (*33*). MAP4 promotes microtubule assembly, counteracts microtubule catastrophe, and has thus major impact on cell cycle regulation (*34*). Since siRNA mediated depletion of any of these genes in isolation did only mildly increase *Brucella* infection, it remains to be investigated if their combinatorial knockdown might phenocopy the effect of miR-103/107.

Cell cycle related proteins were found to be most strongly enriched among the down-regulated proteins in miR-103/107 treated cells. Two cell cycle dependent kinases, CDK4 and CDK6, as well as one of the interacting cyclins (CCND3) were found down-regulated. These factors are responsible for initiation of G1/S phase transition by phosphorylation of the retinoblastoma protein RB1, which in its unphosphorylated form inhibits the action of E2F transcription factors. Activity of E2F is required for the synthesis of genes involved in cell cycle progression (*35*). Among those, many were also found to be down-regulated by miR-103/107, indicative of a secondary effect of CDK4/6 depletion and a G1 arrest. Comparable effects on the cell cycle were

also reported for miRNAs with seed sequences closely related to the seed of miR-103/107, like the miR-15 family (miR-15a, miR-15b, miR-16, miR-195, miR-424, miR-497) (*36*). Nevertheless, there is currently no support of the notion that a G1 arrest could be the underlying cause for increased susceptibility of cells to pathogen infection as identified here. The miR-15 family miRNAs did not affect *Brucella* infection and a report by Maudet and colleagues recently reported inhibitory effects of miR-15 on *Salmonella* infection, although only at later stages during infection, which were not tested in the *Salmonella* screen presented here (*9*).

microRNAs miR-103 and miR-107 have previously been implicated in multiple biological processes including cancer (*24*), cell cycle regulation (*37*), or insulin sensitivity (*38*). Here, we found that *Brucella* invasion of non-phagocytic cells is strongly promoted by this family of miRNAs. Infection with other intracellular pathogens was also enhanced and CME of the host cell was reduced. Altogether, this indicates that the endocytic properties of the host cell are altered in favor of diverse pathogenic infections which are related to different endocytic routes. It will thus be interesting to study the effects of miR-103/107 on further endocytic pathways such as macropinocytosis or caveolae in non-phagocytic cells and correlate these to entry of diverse pathogens.

Materials and Methods

Culture of bacteria and cells

Brucella abortus 2308 strains used in this study contain plasmids pJC43 (*39*) and pAC042.08 (*research article I*) and were grown in tryptic soy broth (TSB) medium (Fluka, 22092) containing 50 ug/ml kanamycin (Sigma-Aldrich, 60615).

Salmonella Typhimurium strain *S.*Tm^{SopE_pM975} was grown in 0.3 M LB medium containing 50 ug/ml streptomycin and 50 ug/ml ampicillin.

Bartonella henselae $ATCC49882^{T}$ containing plasmid pCD353 (*40*) was grown on Columbia base agar (CBA; Thermo Scientific, CM0331B) plates supplemented with 5% defibrinated sheep blood (Thermo Scientific, SR0051) and 50 ug/ml kanamycin.

Listeria monocytogenes EGDe.PrfA*GFP (*41*) was grown in liquid Bacto Brain Heart Infusion (BHI) (BD, 237500).

HeLa cells (human cervical carcinoma epithelial cell line; ATCC, CCL-2) and mouse embryonic fibroblasts (MEFs) were grown in Dulbecco´s Modified Eagle Medium (DMEM) (Sigma, D5796) supplemented with 10 % heat-inactivated Fetal Calf Serum (FCS; Gibco, 10270).

Reverse transfection

Reverse siRNA and miRNA mimic transfection was performed as described in *research article I*. Unless otherwise stated, the following conditions were used: 1.6 pmol siRNA or miRNA mimic were used per well of a 384 well plate. RNAiMAX (Invitrogen, 13778-150) was used 1:250 in DMEM without FCS and RNA:RNAiMAX complex formation was allowed for 1 h at room temperature. 500 HeLa cells or the indicated number of MEF cells were reverse transfected in DMEM/10% FCS for three and two days prior to infection, respectively.

For screening of human miRNAs the Human miRIDIAN miRNA Mimic Library version 16.0 (Dharmacon, CS-00100-E2-01) was used.

Brucella infection

Brucella infection for the miRNA mimic screen and the entry assay in HeLa cells was performed as described in *research article I* and is summarized here.

HeLa cells were infected with *B. abortus* 2308 pJC43 (*aphT::GFP*) at an MOI of 10^4. The infection process was synchronized by centrifugation at 400 *g* and 4°C for 20 min. After 4 h of

incubation at 37°C and 5% CO_2, extracellular bacteria were killed by addition of 100 ug/ml gentamicin (Sigma, G1397) in DMEM/10% FCS. After a total of 44 h, cells were washed in PBS (Gibco, 20012) and fixed in 3.7% paraformaldehyde (PFA) (Sigma, P6148) for 20 min at RT. After permeabilization in 0.1% Triton- X-100 (Sigma-Aldrich, T9284), cell were stained with DAPI (1 ug/ml; Roche, 10236276001) and DY-547-phalloidin (1.5 U/ml; Dyomics, 547PI-33).
For the entry assay *Brucella abortus* 2308 pAC042.08 (*aphT::dsRed, tetO::tetR-GFP*) was used as described above. After the first gentamicin wash, cell were washed again with DMEM/10% FCS containing 100 ug/ml gentamicin and 100 ng/ml anhydrotetracycline (Sigma, 37919) for 4 h to induce GFP expression in intracellular bacteria. After fixation, cells were stained with DAPI (1 ug/ml).
Mouse embryonic fibroblasts (MEFs) were infected with *B. abortus* 2308 pAC042.08 as described for the entry assay. Induction of GFP expression in intracellular bacteria started at 4 h after inactivation of extracellular bacteria by gentamycin treatment and continued throughout the experiment. Fixed cells were stained with DAPI (1 ug/ml).

Imaging, automated image analysis

Imaging, image analysis, and infection scoring was performed as described in *research article IV* with minor modifications as described here.
Images were acquired using a Molecular Device ImageXpress microscope with a 10X S Fluor objective with 0.45 NA. Laser-based autofocusing on each site was used for acquisition of a 3x3 image array per site without gain and with a 12 bit dynamic range. Z-offsets for each channel and exposure times were manually defined ensuring a high dynamic range and low overexposure. DNA was imaged in the DAPI channel, pathogens in GFP, and, depending on the assay, F-actin in RFP or Cy5. The *Listeria* screen uses an additional Cy3 staining as described in detail by Kühbacher et al. (*41*).
Images were scaled such that pixel intensities of a full plate lie between 0 and 1. Images were then corrected for shading (flat field correction, vignetting correction) by applying a shading model to the image pixels. 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. On the corrected images, object detection was performed using CellProfiler (*42*). Firstly, nucleus objects labeled “Nuclei“ were segmented in the DAPI channel using OTSU’s method (CellProfiler module

IdentifyPrimAutomatic). Secondly, a 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 assays using *B. abortus* 2308 pAC042.08 that allows the detection of intracellular bacteria, the cell body was not stained with a fluorescent marker and only a Voronoi cell body is used. Intracellular bacteria were segmented in CellProfiler using the GFP signal from intracellular bacteria based on the OTSU method or wavelets.

On the segmented objects, measurements were performed using CellProfiler. From all segmented objects (Nuclei, PeriNuclei, Cells, VoronoiCells, Bacteria), shape measurements were extracted. Intensity and texture measurements were extracted from all available channels (DAPI, Actin, and Pathogen).

Infection scoring for *Brucella* infections with *B. abortus 2308* pJC43

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 ration of infected cells to total cell number in a well. We selected a number of image analysis single cell features that were most sensitive to the infection phenotype. To quantify *B. abortus* infection for the endpoint assay, GFP intensity was measured in the objects Nuclei, PeriNuclei, and Cells using CellProfiler module MeasureObjectIntensity. Thresholds for GFP intensity for each of the objects were set manually and a cell was considered infected if any of the thresholds was reached.

Segmentation-based infection scoring for assays using *B. abortus* 2308 pAC042.08

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. The bacterial load is quantified using the median integrated GFP intensity.

Infection assay and analysis of other pathogen infections

Infection screens in HeLa cells with adenovirus (Ad2_ΔE3B-eGFP (*43*)), Bartonella (*Bartonella henselae* ATCC49882^{T} containing plasmid pCD353 (*40*)), Listeria (*Listeria monocytogenes* EGDe.PrfA*GFP (*41*)), rhinovirus (human rhinovirus serotype 1a (HRV1a (*44, 45*)), Salmonella SopE (*Salmonella* Typhimurium strain *S.*Tm^{SopE_pM975}) were performed and analyzed as described in *research article IV*.

Transferrin uptake assay

Transferrin (Tf) uptake assay was performed as described by Hirschmann et al. (personal communication: Hirschmann, D.T., Kasper, C.A., and Spiess, M. (in press). Quantitative analysis of Tf cycling by automated fluorescence microscopy. Methods Mol. Biol.) with some modifications in image analysis. Scrambled siRNA or miR-103 miRNA mimic were mixed with RNAiMAX (1:250) in DMEM without FCS and 90 ul were added per well of a 96 well plate. After 30 min of incubation at room temperature, 150 ul of a HeLa CCl-2 cell suspension of 16'000 cells in DMEM/16% FCS was added, resulting in 2400 cell per well and a final FCS concentration of 10%. Cells were incubated at 37°C and 5% CO_2 for 3 days. The cells were washed once with assay medium (DMEM without FCS, buffered with 20 mM HEPES (Gibco, 15630-080)) and incubated in assay medium for 2 h at 37°C and 5% CO_2. The plates were transferred to a 37°C water bath and the medium was replaced by 50 ul of assay medium containing 50 ug/ml of fluorescently labeled transferrin (568-Tf, Invitrogen, T-23365) for the time indicated. Endocytosis was stopped by rapid transfer of the assay plate to a metal plate precooled on ice and the addition of 150 ul/well ice-cold PBS. Cells were twice quickly washed with ice-cold PBS followed by a wash with ice-cold stripping buffer (150 mM NaCl, 50 mM NaAcetate, pH 3.5) to remove surface bound 568-Tf. Cells were fixed with 170 ul/well 3% PFA solution for 5 min on ice followed by 10 min at room temperature. Cells were washed once with PBS, once with 50 mM ammonium chloride in PBS, followed by 3 washes with PBS. Cells were permeabilized for 1 min in 100 ul/well 0.1% Triton-X-100 in PBS and washed four times in PBS. Nuclei were stained by DAPI (1 ug/ml) and the actin cytoskeleton by DY-495-Phalloidin (1 U/ml; Dyomics, 495-33) followed by four PBS washes. Unless otherwise stated, 200 ul/well are used for all washing steps. Imaging and image analysis were performed as described above, acquiring 25 sites/well. Nuclei and cell body were segmented using the signals from the DAPI and the DY-488-Phalloidin signals, respectively, applying the image analysis pipeline as

described above. Internalized Tf was quantified by measuring the mean Tf-568 fluorescence intensity per cell averaged over all cells per condition.

Plasmid transfection

HeLa cells were seeded a day before transfection in a 6 well plate with $2.5*10^5$ cells/well. For double transfection, 0.9 µg of plasmid or 0.45 µg of each plasmid (in 200 µl of DMEM without FCS) were mixed with 8 µl of Fugene HD (Promega, E2312) (in 200 µl of DMEM without FCS) and incubated for 15 min at RT. The medium on HeLa cells was exchanged with 1.5 ml of fresh medium and the DNA-Fugene complex was added to the cells. The next morning, cells were exchanged with fresh medium and in the evening seeded into a 96 well format (2800 cells/well) for infections on the following day.

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 TGFβ-1 (R&D Systems, 240-13). Cells were then incubated at 37°C and 5% CO_2 for another 24 h, after which infection was performed in absence of TGFβ-1 in the culture medium.

Quantitative proteomics

Scrambled siRNA or miRNA mimic were mixed with RNAiMAX (1:250) in DMEM without FCS and 600 ul were added per well of a 24 well plate. After 30 min of incubation at RT, 10^5 HeLa CCl-2 cells were seeded per well. Cells were incubated at 37°C and 5% CO_2 for 3 days. Cells were washed with 500 ul/well trypsin and detached from the well by incubation in 500 ul of trypsin at 37°C and 5% CO_2 for 2-3 min. Cells were washed twice in PBS, the supernatant was removed and the cell pellet was frozen in liquid nitrogen and kept at -80°C until processing. Cells from 2 wells per sample were pooled for proteomics analysis.

Sample preparation: For each sample, cells were resuspended in 100 µl of lysis buffer (2% sodium deoxycholate (DOC), 100 mM ammonium bicarbonate), sonicated for 2x10 s using a vial tweeter and spinned down. A small aliquot of the supernatant was taken to determine the protein concentration using a BCA assay (Thermo Fisher Scientific). Aliquots containing 50 ug of proteins were taken from each sample, were reduced with 5 mM TCEP for 15 min at 95°C and alkylated with 10 mM iodoacetamide for 30 min in the dark at 25°C. After quenching the

reaction with 12 mM N-acetyl-cysteine the samples were diluted with 100 mM ammonium bicarbonate buffer to a final DOC concentration of 1%. Proteins were digested by incubation with sequencing-grade modified trypsin (1/50, w/w; Promega, Madison, Wisconsin) overnight at 37°C. Then, the samples were acidified with 2 M HCl to a final concentration of 50 mM, incubated for 15 min at 37°C and the precipitated detergent was removed by centrifugation at 10,000 *g* for 15 min. Subsequently, peptides were desalted on C18 reversed-phase spin columns according to the manufacturer's instructions (Microspin, Harvard Apparatus) and dried under vacuum. Dried peptide samples were labeled with an isobaric tag (TMT 10-plex, Thermo Fisher Scientific) according to the manufacturer's instructions. The pooled sample was again desalted on C18 reversed-phase spin columns according to the manufacturer's instructions (Macrospin, Harvard Apparatus) and dried under vacuum.

Off-Gel electrophoresis: TMT-labeled samples were resolubilized to a final concentration of 1 mg/ml in Off-Gel electrophoresis buffer containing 6.25% glycerol and 1.25% IPG buffer (GE Healthcare). Peptides were separated on a 12 cm pH 3-10 IPG strip (GE Healthcare) with a 3100 OFFGEL fractionator (Agilent) as previously described (*46*) using a protocol of 1 h rehydration at max. 500 V, 50 µA, and 200 mW. Peptides were separated at max. 8000 V, 100 µA, and 300 mW until 20 kVh were reached. Subsequently, each of the 12 peptide fractions was desalted using C18 reversed-phase columns according to the manufacturer's instructions (Microspin, Harvard Apparatus), dried under vacuum and subjected to LC-MS/MS analysis.

Mass spectrometric analysis: The setup of the µRPLC-MS system was as described previously (*47*). Chromatographic separation of peptides was carried out using an EASY nano-LC 1000 system (Thermo Fisher Scientific), equipped with a heated RP-HPLC column (75 µm x 50 cm) packed in-house with 1.9 µm C18 resin (Reprosil-AQ Pur, Dr. Maisch). Aliquots of 1 µg of total peptides were analyzed per LC-MS/MS run using a linear gradient ranging from 95% solvent A (0.15% formic acid, 2% acetonitrile) and 5% solvent B (98% acetonitrile, 2% water, 0.15% formic acid) to 30% solvent B over 180 min at a flow rate of 200 nl/min. Mass spectrometry analysis was performed on a dual pressure LTQ-Elite Orbitrap mass spectrometer equipped with a nanoelectrospray ion source (both Thermo Fisher Scientific). Each MS1 scan was followed by high-collision-dissociation (HCD, both acquired in the Orbitrap) of the 10 most abundant precursor ions with dynamic exclusion for 60 sec. Total cycle time was approximately 2 s. For MS1, 10E6 ions were accumulated in the Orbitrap cell over a maximum time of 300 ms and scanned at a resolution of 30,000 FWHM (at 400 m/z). MS2 scans were acquired at a target

setting of 50,000 ions, accumulation time of 100 ms and a resolution of 15,00 FWHM (at 400 m/z). Singly charged ions and ions with unassigned charge state were excluded from triggering MS2 events. The normalized collision energy was set to 35%, and one microscan was acquired for each spectrum.

Database searching and protein quantification: Acquired raw-files were converted to the mascot generic file (mgf) format using the msconvert tool (part of ProteoWizard, version 3.0.4624 (2013-6-3)). Using the MASCOT algorithm (Matrix Science, Version 2.4.0), the mgf files were searched against a decoy database containing normal and reverse sequences of the predicted SwissProt entries of *Homo sapiens* (www.ebi.ac.uk, release date 16/05/2012) and commonly observed contaminants (in total 41,250 sequences) generated using the SequenceReverser tool from the MaxQuant software (Version 1.0.13.13). The precursor ion tolerance was set to 10 ppm and fragment ion tolerance was set to 0.02 Da. The search criteria were set as follows: full tryptic specificity was required (cleavage after lysine or arginine residues unless followed by proline), 2 missed cleavages were allowed, carbamidomethylation (C), TMT6plex (K and peptide N-terminus) were set as fixed modification and oxidation (M) as a variable modification. Next, the database search results were imported to the Scaffold Q+ software (version 4.3.3, Proteome Software Inc., Portland, OR) and the protein false identification rate was set to 1% based on the number of decoy hits. Specifically, peptide identifications were accepted if they could be established at greater than 93.0% probability to achieve an FDR less than 1.0% by the scaffold local FDR algorithm. Protein identifications were accepted if they could be established at greater than 5.0% probability to achieve an FDR less than 1.0% and contained at least 1 identified peptide. Protein probabilities were assigned by the Protein Prophet program (*48*). Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony. Proteins sharing significant peptide evidence were grouped into clusters. Acquired reporter ion intensities in the experiments were employed for automated quantification and statically analyzed using a modified version of our in-house developed SafeQuant R script (*47*). In brief, reporter ion intensities were corrected for isotopic impurities according to the manufacturer's instructions. Intensities for each peptide and protein identification were summed, globally normalized across all acquisition runs, and employed for ratio calculation and statistical analysis.

Mapping and counting of sequencing reads per gene (represented by the longest transcript of each gene) were done with CLIPZ (*49*). Subsequently, the relative expression level of each gene

(RPKM) was calculated by normalizing the raw read count by the transcript length as well as by the size of the sequencing library (sum of all considered reads). Finally, log2 fold-changes were calculated by subtracting the log2 mean expression of control cells from the log2 mean expression in of miR-103 transfected cells.

Directional mRNA sequencing

HeLa cells were transfected as described above for proteomics analysis. Cells from one well were used per condition and three biological replicates for scrambled and miR-103 miRNA mimic treated cells were processed in parallel during mRNA preparation and sequencing.

Libraries for mRNA sequencing were prepared using the "Directional mRNA-seq sample preparation" protocol from Illumina with minor modifications. In brief, poly(A)+ RNA was isolated directly from cells using Dynabeads® mRNA DIRECT™ Kit (61011, Life Technologies) according to the manufacturer protocol. After isolation, 50 ng of mRNA was chemically fragmented by incubating the mRNA solution with twice the volume of alkaline hydrolysis buffer (50 mM sodium carbonate [NaHCO3/Na2Co3] pH 9.2, 1 mM EDTA) at 95°C for 5 min to get the fragments of approximately 200-300 bases. Fragmented mRNA was immediately purified with the RNeasy MinElute Cleanup Kit (74204, Qiagen) to stop the reaction and to remove small RNA fragments (<100 bases). Purified fragmented mRNA was treated with thermo-sensitive alkaline phosphatase FastAP (EF0651, Fermentas) at 37°C for 30 min and then at 75°C for 5 min to inactivate FastAP. Fragmented mRNA was further incubated with ATP and T4 polynucleotide kinase (EK0032, Fermentas) at 37°C for 1 h and subsequently purified. Ligation of RNA 3'-adapter (RA3, part # 15013207, Illumina) was done using T4 RNA Ligase 2, truncated K227Q (M0351L, New England Biolabs Inc) according to the Illumina protocol. The ligation step is followed by RNA purification as mentioned above to remove free 3'-adapters. The RNA 5'-adapter (RA5, part # 15013205, Illumina) was ligated using T4 RNA ligase (EL0021, Fermentas) according to the Illumina protocol followed by an RNA purification step to remove free 5'-adapters. cDNA was synthesized using RNA RT Primer (RTP, part # 15013981, Illumina) using SuperScript III (18080044, Invitrogen) as per Illumina protocol. Libraries were amplified for 14 cycles of PCR using a forward PCR primer (RNA PCR Primer (RP1), part # 15005505 Illumina) and a reverse PCR primer (Illumina PCR Primer, Index). Different indexed reverse PCR primers were used for library preparation from different samples to facilitate multiplexing. Libraries were sequenced for 51 cycles on an Illumina HiSeq 2000 machine.

Acknowledgements

This work was supported by grant 51RT 0_126008 for the Research and Technology Development (RTD) project InfectX in the frame of SystemsX.ch, the Swiss Initiative for Systems Biology. We acknowledge grant 31003A-132979 from the Swiss National Science Foundation (SNSF). Work of S.H.L and A.C. was supported by the International PhD Program "Fellowships for Excellence" of the Biozentrum. Simone Muntwiler is acknowledged for technical assistance. We would like to thank Prof. Dr. Dirk Bumann from the Biozentrum for providing pNF106. Thank goes to the following collaborators for sharing infection data with miR-103/107 miRNA mimics: *Salmonella* (Dr. Saskia Kreibich, Prof. Dr. Wolf-Dietrich Hardt; ETH Zurich), *Bartonella* (Dr. Simone Eicher; Biozentrum Basel), *Listeria* (Dr. Andreas Kühbacher, Dr. Javier Pizarro-Cerda, Prof. Dr. Pascale Cossart; Institut Pasteur Paris), adenovirus and rhinovirus (Dr. Daria Seiler, Bettina Cardel, Dr. Artur Yakimovich, Prof. Dr. Urs Greber; University of Zurich).

References

1. D. P. Bartel, MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281 (Jan 23, 2004).
2. B. P. Lewis, I. H. Shih, M. W. Jones-Rhoades, D. P. Bartel, C. B. Burge, Prediction of mammalian microRNA targets. *Cell* **115**, 787 (Dec 26, 2003).
3. I. Lee *et al.*, New class of microRNA targets containing simultaneous 5'-UTR and 3'-UTR interaction sites. *Genome Res* **19**, 1175 (Jul, 2009).
4. J. Hausser, A. P. Syed, B. Bilen, M. Zavolan, Analysis of CDS-located miRNA target sites suggests that they can effectively inhibit translation. *Genome Res* **23**, 604 (Apr, 2013).
5. A. Lal *et al.*, miR-24 Inhibits cell proliferation by targeting E2F2, MYC, and other cell-cycle genes via binding to "seedless" 3'UTR microRNA recognition elements. *Mol Cell* **35**, 610 (Sep 11, 2009).
6. C. Staedel, F. Darfeuille, MicroRNAs and bacterial infection. *Cell Microbiol* **15**, 1496 (Sep, 2013).
7. B. R. Cullen, S. Cherry, B. R. tenOever, Is RNA interference a physiologically relevant innate antiviral immune response in mammals? *Cell Host Microbe* **14**, 374 (Oct 16, 2013).
8. B. R. Cullen, MicroRNAs as mediators of viral evasion of the immune system. *Nat Immunol* **14**, 205 (Mar, 2013).
9. C. Maudet *et al.*, Functional high-throughput screening identifies the miR-15 microRNA family as cellular restriction factors for Salmonella infection. *Nat Commun* **5**, 4718 (2014).
10. C. Belair *et al.*, Helicobacter pylori interferes with an embryonic stem cell micro RNA cluster to block cell cycle progression. *Silence* **2**, 7 (2011).
11. J. C. Tam, D. A. Jacques, Intracellular immunity: finding the enemy within--how cells recognize and respond to intracellular pathogens. *J Leukoc Biol* **96**, 233 (Aug, 2014).
12. D. S. Merrell, S. Falkow, Frontal and stealth attack strategies in microbial pathogenesis. *Nature* **430**, 250 (Jul 8, 2004).
13. P. Cossart, A. Helenius, Endocytosis of viruses and bacteria. *Cold Spring Harb Perspect Biol* **6**, (Aug, 2014).
14. J. Pizarro-Cerda, P. Cossart, Bacterial adhesion and entry into host cells. *Cell* **124**, 715 (Feb 24, 2006).
15. V. L. Atluri, M. N. Xavier, M. F. de Jong, A. B. den Hartigh, R. E. Tsolis, Interactions of the human pathogenic Brucella species with their hosts. *Annu Rev Microbiol* **65**, 523 (2011).
16. T. Starr, T. W. Ng, T. D. Wehrly, L. A. Knodler, J. Celli, Brucella intracellular replication requires trafficking through the late endosomal/lysosomal compartment. *Traffic* **9**, 678 (May, 2008).
17. T. D. Anderson, N. F. Cheville, Ultrastructural morphometric analysis of Brucella abortus-infected trophoblasts in experimental placentitis. Bacterial replication occurs in rough endoplasmic reticulum. *Am J Pathol* **124**, 226 (Aug, 1986).
18. J. Celli *et al.*, Brucella evades macrophage killing via VirB-dependent sustained interactions with the endoplasmic reticulum. *J Exp Med* **198**, 545 (Aug 18, 2003).
19. M. R. Franceschini A, Casanova A, Kreibich S, Daga N, Andritschke D, Dilling S, Rämö P, Emmenlauer M, Kaufmann A, Conde-Alvarez R, Low SH, Pelkmans L, Helenius A, Hardt

WD, Dehio C, von Mering C, Specific inhibition of diverse pathogens in human cells by synthetic microRNA-like oligonucleotides inferred from genome-wide RNAi screens. *manuscript in preparation*, (2014).
20. J. R. Finnerty *et al.*, The miR-15/107 group of microRNA genes: evolutionary biology, cellular functions, and roles in human diseases. *J Mol Biol* **402**, 491 (Sep 24, 2010).
21. C. Dehio, M. Meyer, J. Berger, H. Schwarz, C. 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* **110 (Pt 18)**, 2141 (Sep, 1997).
22. A. Grimson *et al.*, MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell* **27**, 91 (Jul 6, 2007).
23. D. Gaidatzis, E. van Nimwegen, J. Hausser, M. Zavolan, Inference of miRNA targets using evolutionary conservation and pathway analysis. *BMC Bioinformatics* **8**, 69 (2007).
24. G. Martello *et al.*, A MicroRNA targeting dicer for metastasis control. *Cell* **141**, 1195 (Jun 25, 2010).
25. K. H. Lee *et al.*, Epigenetic silencing of MicroRNA miR-107 regulates cyclin-dependent kinase 6 expression in pancreatic cancer. *Pancreatology* **9**, 293 (2009).
26. L. K. Goh, A. Sorkin, Endocytosis of receptor tyrosine kinases. *Cold Spring Harb Perspect Biol* **5**, a017459 (May, 2013).
27. F. Huang, Y. G. Chen, Regulation of TGF-beta receptor activity. *Cell Biosci* **2**, 9 (2012).
28. Y. Shen, M. Naujokas, M. Park, K. Ireton, InIB-dependent internalization of Listeria is mediated by the Met receptor tyrosine kinase. *Cell* **103**, 501 (Oct 27, 2000).
29. T. J. Wickham, P. Mathias, D. A. Cheresh, G. R. Nemerow, Integrins alpha v beta 3 and alpha v beta 5 promote adenovirus internalization but not virus attachment. *Cell* **73**, 309 (Apr 23, 1993).
30. J. M. Bergelson *et al.*, Isolation of a common receptor for Coxsackie B viruses and adenoviruses 2 and 5. *Science* **275**, 1320 (Feb 28, 1997).
31. C. Y. Chang *et al.*, NFIB is a governor of epithelial-melanocyte stem cell behaviour in a shared niche. *Nature* **495**, 98 (Mar 7, 2013).
32. S. Mason, M. Piper, R. M. Gronostajski, L. J. Richards, Nuclear factor one transcription factors in CNS development. *Mol Neurobiol* **39**, 10 (Feb, 2009).
33. M. Hayashi *et al.*, Molecular cloning and characterization of human PDE8B, a novel thyroid-specific isozyme of 3',5'-cyclic nucleotide phosphodiesterase. *Biochem Biophys Res Commun* **250**, 751 (Sep 29, 1998).
34. P. Holmfeldt, G. Brattsand, M. Gullberg, MAP4 counteracts microtubule catastrophe promotion but not tubulin-sequestering activity in intact cells. *Curr Biol* **12**, 1034 (Jun 25, 2002).
35. S. Lapenna, A. Giordano, Cell cycle kinases as therapeutic targets for cancer. *Nat Rev Drug Discov* **8**, 547 (Jul, 2009).
36. P. S. Linsley *et al.*, Transcripts targeted by the microRNA-16 family cooperatively regulate cell cycle progression. *Mol Cell Biol* **27**, 2240 (Mar, 2007).
37. Y. Takahashi *et al.*, MiR-107 and MiR-185 can induce cell cycle arrest in human non small cell lung cancer cell lines. *PLoS One* **4**, e6677 (2009).
38. M. Trajkovski *et al.*, MicroRNAs 103 and 107 regulate insulin sensitivity. *Nature* **474**, 649 (Jun 30, 2011).

39. J. Celli, S. P. Salcedo, J. P. Gorvel, Brucella coopts the small GTPase Sar1 for intracellular replication. *Proc Natl Acad Sci U S A* **102**, 1673 (Feb 1, 2005).
40. M. Dehio, A. Knorre, C. Lanz, C. Dehio, Construction of versatile high-level expression vectors for Bartonella henselae and the use of green fluorescent protein as a new expression marker. *Gene* **215**, 223 (Jul 30, 1998).
41. A. Kuhbacher, E. Gouin, P. Cossart, J. Pizarro-Cerda, Imaging InlC secretion to investigate cellular infection by the bacterial pathogen Listeria monocytogenes. *J Vis Exp*, e51043 (2013).
42. A. E. Carpenter *et al.*, CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol* **7**, R100 (2006).
43. A. Yakimovich *et al.*, Cell-free transmission of human adenovirus by passive mass transfer in cell culture simulated in a computer model. *J Virol* **86**, 10123 (Sep, 2012).
44. A. Jurgeit *et al.*, An RNA replication-center assay for high content image-based quantifications of human rhinovirus and coxsackievirus infections. *Virol J* **7**, 264 (2010).
45. A. Jurgeit *et al.*, Niclosamide is a proton carrier and targets acidic endosomes with broad antiviral effects. *PLoS Pathog* **8**, e1002976 (2012).
46. M. Beck *et al.*, The quantitative proteome of a human cell line. *Mol Syst Biol* **7**, 549 (2011).
47. T. Glatter *et al.*, Large-scale quantitative assessment of different in-solution protein digestion protocols reveals superior cleavage efficiency of tandem Lys-C/trypsin proteolysis over trypsin digestion. *J Proteome Res* **11**, 5145 (Nov 2, 2012).
48. A. I. Nesvizhskii, A. Keller, E. Kolker, R. Aebersold, A statistical model for identifying proteins by tandem mass spectrometry. *Anal Chem* **75**, 4646 (Sep 1, 2003).
49. M. Khorshid, C. Rodak, M. Zavolan, CLIPZ: a database and analysis environment for experimentally determined binding sites of RNA-binding proteins. *Nucleic Acids Res* **39**, D245 (Jan, 2011).
50. G. E. Crooks, G. Hon, J. M. Chandonia, S. E. Brenner, WebLogo: a sequence logo generator. *Genome Res* **14**, 1188 (Jun, 2004).
51. W. Huang da, B. T. Sherman, R. A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44 (2009).
52. W. Huang da, B. T. Sherman, R. A. Lempicki, Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* **37**, 1 (Jan, 2009).
53. A. Franceschini *et al.*, STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res* **41**, D808 (Jan, 2013).

Figure and table legends

Figure 1. miR-103 and miR-107 promote *Brucella* infection of HeLa cells

HeLa cells were transfected for 3 days, followed by infection with *B. abortus* 2308 pJC43 for 44 h. Cells were stained for DNA with DAPI and F-actin with DY-547-phalloidin. Automated image analysis and infection scoring was applied to quantify cell counts and infection rate for each condition.

A) Dotplot of miRNA mimic screen using Dharmacon miRIDIAN miRNA mimics. Dots represent the average of two independent replicates for Z score normalized cell number and infection index. miR-103 and miR-107 are highlighted in green and the sequences of the mature miRNAs are shown. The seed sequences are underlined and the GCAGC motif shared with miR-1184 is highlighted in green.

B) Fluorescence image showing *B. abortus* infection of HeLa cells transfected with miR-103 and miR-107 miRNA mimics compared to scrambled (non-targeting) control (green = *B. abortus*, blue = nuclei stained with DAPI, scale bar = 50 um).

C) Effect of siRNAs (black circle) and miRNA mimics (red circle) containing a GCAGC motif within their seed sequence on *Brucella* infection. Independent replicates are averaged and the vertical line indicates the mean effect of the different RNAi molecules of the respective seed sequence. The seed sequence of siRNAs identical to the seed of miR-103 and miR-107 is in bold. A complete list of readouts and RNAi reagents can be found in Suppl. Table 2.

D) Sequence logo (http://weblogo.berkeley.edu, (*50*)) of siRNA or miRNA mimic seeds that increased *Brucella* infection by a Z score of at least 4.

Figure 2. High levels of mir-103 or siRNAs containing similar seed sequences increase entry of *Brucella* in HeLa cells

A) Entry assay: Fluorescence image showing example of entry assay (green = intracellular *B. abortus*, blue = nuclei stained with DAPI, scale bar = 50 μm). TetR-GFP expressing *B. abortus* was allowed to enter HeLa cells for 4 h followed by killing of extracellular bacteria and simultaneous induction of GFP in intracellular bacteria with anhydrotetracycline for 4 h.

Automated image analysis: CellProfiler pipeline is used to detect DAPI-stained nuclei followed by calculation of a Voronoi cell body by the radial extension of the nucleus by 25 pixels. Bacteria are segmented based on the GFP signal. A cell is considered infected if at least one segmented

bacterial object of sufficient size and intensity overlaps with the Voronoi cell body. The bacterial load in infected cells is estimated by the integration of the GFP intensity of all segmented bacterial objects.

B) Example images showing increased infection rates of TetR-GFP expressing *B. abortus* in cells transfected with miR-103 miRNA mimics or molecules with similar seed sequences.

C) *Brucella* infection rate of HeLa cells. Data are represented as bar graphs (bar = mean; normalization to mock; error bars = SEM; ***p<0.001, **p<0.01; Mann-Whitney test; n = 7 (SCRAMBLED, miR-103 mimic), n=5 (ALLSTARS, siRNA 1, siRNA 2))

D) Estimate of bacterial load of infected cells. The data are represented as bar graphs (bar = mean; normalization to mock; error bars = SEM; ***p<0.001, **p<0.01; Mann-Whitney test; n = 7 (SCRAMBLED, miR-103 mimic), n=5 (ALLSTARS, siRNA 1, siRNA 2)).

E) Sequences used for transfection with the corresponding underlined seed region. The following materials were used: miR-103 mimic (Dharmacon, C-300523-03); siRNA1 (Qiagen, SI03075947); siRNA2 (Qiagen, SI03070529).

Figure 3. Preliminary results indicate that high levels of miR-103 increase *Brucella* entry in mouse embryonic fibroblasts (MEF)

A) Cell counts of MEF cells transfected with the cell killer control kif11 for the time indicated. Data are represented as bar graphs (bar = mean; error bars = range; n = 2).

B) Estimate of bacterial load in MEF cells infected with *Brucella* for 8 h, 24 h, and 48 h. Cells were transfected for 48 h with the RNAi reagents as indicated. *B. abortus* pAC42.08 expressing a tetracycline-inducible GFP was added for 4 h, followed by inactivation of extracellular bacteria by gentamicin and induction of GFP by anhydrotetracycline for 4 h, 20 h, or 44 h. Data are represented as bar graphs (bar = mean; bacterial load = infection rate * mean area of intracellular bacteria in infected cells; error bars = range; n = 2).

C) Fluorescence images illustrating *Brucella* infection of MEF cells treated with either a scrambled control siRNA, ARPC3 siRNA, or a miR-103 miRNA mimic. Total infection time = 8 h; scale bar = 50 μM; green = intracellular *Brucella*; blue = DAPI staining of DNA.

Figure 4. miR-103/107 enhance infection of different intracellular viral and bacterial pathogens

Effect of miR-103 and miR-107 miRNA mimics on infection rate of intracellular pathogens in HeLa cells. Z scored infection index of two independent replicates are averaged and the vertical line indicates the mean effect of miR-103 and miR-107 miRNA mimics. Pathogens used: adenovirus (Ad2_ΔE3B-eGFP), Bartonella (*Bartonella henselae* ATCC49882^{T} containing plasmid pCD353), Brucella (*Brucella abortus* 2308 pJC43 (*aphT::GFP*), Listeria (*Listeria monocytogenes* EGDe.PrfA*GFP), rhinovirus (human rhinovirus serotype 1a (HRV1a), Salmonella (*Salmonella* Typhimurium strain *S.*Tm^{SopE_pM975})

Figure 5. Proteome changes induced by elevated levels of miR-103/107 levels

A) Volcano plot showing differential protein levels in cells transfected with either miR-103 or miR-107 miRNA mimics compared to control cells transfected with scrambled control. False discovery rate (FDR) representing Benjamin-Hochberg correction is calculated comparing miR-103 (n=3) together with miR-107 (n=3) to scrambled (n=3). Direct targets of miR-103/107 predicted by TargetScan are highlighted (TargetScan Release 6.2, Total context+ score, predicted targets with only conserved sites). Color code represents confidence of target prediction (red = high confidence, white = low confidence).

B) Effect of siRNA depletion of corresponding targets on *Brucella* infection. Among the identified proteins only those are shown, which appear in the top 500 up or down hit list of a genome-wide siRNA screen for *Brucella* infection (*research article V*). The green color gradient represents targets that we found to increase *Brucella* infection upon target depletion, the red gradient those targets that reduce *Brucella* infection upon depletion.

Figure 6. Gene ontology (GO) enrichment terms of proteins changed by elevated miR-103/107

GO analysis was performed using DAVID (http://david.abcc.ncifcrf.gov/ (*51, 52*). As input, genes which showed significantly altered protein levels ($q < 0.01$) were used. Selection of non-similar, significantly enriched GO terms ($q < 0.05$) is shown as bar graphs of $-\mathrm{Log}_{10}(q)$. In brackets the number of genes found in the sample is shown. BP = Biological function; CC = cellular compartment, KEGG = KEGG pathways.

A) GO enrichment for upregulated targets enriched for terms involving cell surface related functions.

B) GO enrichment for down-regulated targets enriched for cell cycle related terms.

Figure 7. String network of protein changes induced by miR-103/107

Proteins significantly changed ($q < 0.01$) in cells with high miR-103/107 levels were used as input for STRING analysis (http://string-db.org/ (*53*)). Proteins with at least one high confidence (> 0.95) edge are displayed. Red and green color gradients in the boxes indicate the strength of down-regulation and upregulation, respectively. Colors outlining the boxes are based on functional pathways.

Figure 8. Effect of TGF-β signaling components on Brucella infection

A) Activation of TGF-β signaling by TGF-β1 increases *Brucella* infection. HeLa cells were pre-incubated for 24 h with DMEM containing TGF-β1 before infection with GFP-expressing *B. abortus* for 44 h. Data represent normalized infection rates in reference to cells that were not pre-incubated with TGF-β1 with the mean ± STDEV of at least three independent experiments.

B and C) Overexpression of wild type or dominant negative TGFBR increase or reduce *Brucella abortus* infection, respectively. HeLa cells were transfected with B) wild type TGFBR (pCMV5-TGFBR1 or pCMV5B-TGFBR2) or C) dominant-negative TGFBR (pCMV5B-TGFBR1 K232R, pCMV5 TGFBR2 Δcyt, pCMV5B-TGFBR2 K227R). After 1.5 days of overexpression, cells were infected with GFP-expressing *B. abortus* for 2 days. Data represent normalized infection rates in reference to non-transfected cells with the mean ± STDEV of at least four independent experiments.

Figure 9. miR-103 reduces transferrin uptake by HeLa cells

A) Time course of transferrin (Tf) uptake by HeLa cells. HeLa cells were transfected with a miR-103 mimic (black) or scrambled control siRNA (grey) for 3 days followed by incubation with fluorescent transferrin (Tf-568) for the time indicated. Nuclei and actin cytoskeleton were stained with DAPI and 488-Phalloidin, respectively. An automated image analysis pipeline was used to segment nuclei and cell bodies in the acquired fluorescence images. Internalized Tf was quantified by measuring the mean Tf-568 intensity per cell averaged over all cells per condition. Error bars = standard error of the mean; n= 5).

B) Fluorescent images showing internalized transferrin 20 min after addition to HeLa cells. Scale bar = 50 μM.

Supplementary Figure S1. Correlation of protein and mRNA levels after miR-103/107 treatment

mRNA log_2 fold-changes versus log2 protein ratios after filtering out genes with absolute protein ratios higher than 5.0. Scatter plot of all available genes (grey) versus the top 100 miR-103 targets from TargetScan (black). 'n' specifies the number of genes plotted and 'r' the Pearson correlation coefficient.

Supplementary Figure S2. Cumulative distributions of protein log_2 ratios

The cumulative distribution of the log_2 fold-changes of all genes, the predicted top 150 Elmmo hsa-miR-103 targets and the top 150 TargetScan hsa-miR-103 targets are shown. Targets predicted by Elmmo as well as by TargetScan are down-regulated significantly upon miR-103 transfection.

Supplementary Figure S3. Cumulative distributions of mRNA log_2-foldchanges

The cumulative distribution of the log_2 fold-changes of all genes, the predicted top 150 Elmmo miR-103 targets and the top 150 TargetScan miR-103 targets are shown. Both, Elmmo as well as TargetScan targets are down-regulated significantly upon miR-103 transfection.

Supplementary Figure S4. Effect on cell number and *Brucella* infection of different concentrations of miR-103 miRNA mimic, scrambled, and kif11

RNAi reagents used: kif11 (Dharmacon, L-003317-00), miR103 mimic (Dharmacon, C-300523-03), SCRAMBLED (Dharmacon, D-001810-10). HeLa cells were transfected for 3 days prior to infection.

A) Entry assay showing the relative bacterial load in HeLa cells infected with *B. abortus* pAC42.08 and number of HeLa cells. Bacterial load is calculated by multiplication of the normalized infection rate multiplied by the normalized median of the integrated GFP intensity of infected cells. The data are represented as bar graphs (bar = mean; normalization to mock; error bars = SEM; n = 4)

B) Endpoint assay showing the relative infection rate in HeLa cells infected with GFP expressing *B. abortus* and number of HeLa cells. The data are represented as bar graphs (bar = mean; normalization to mock; error bars = SEM; n = 4)

Supplementary Table 1. Top ten up and down hits in miRIDIAN miRNA mimic screen regulating *Brucella* infection

Z scored effect of miRNA mimic on *Brucella* infection averaged over two independent replicates. Negative and positive values indicate decreased or increased *Brucella* infection, respectively. Names of mature miRNAs with their corresponding seed region covering nucleotides at position 2-8 from the 5' end are shown.

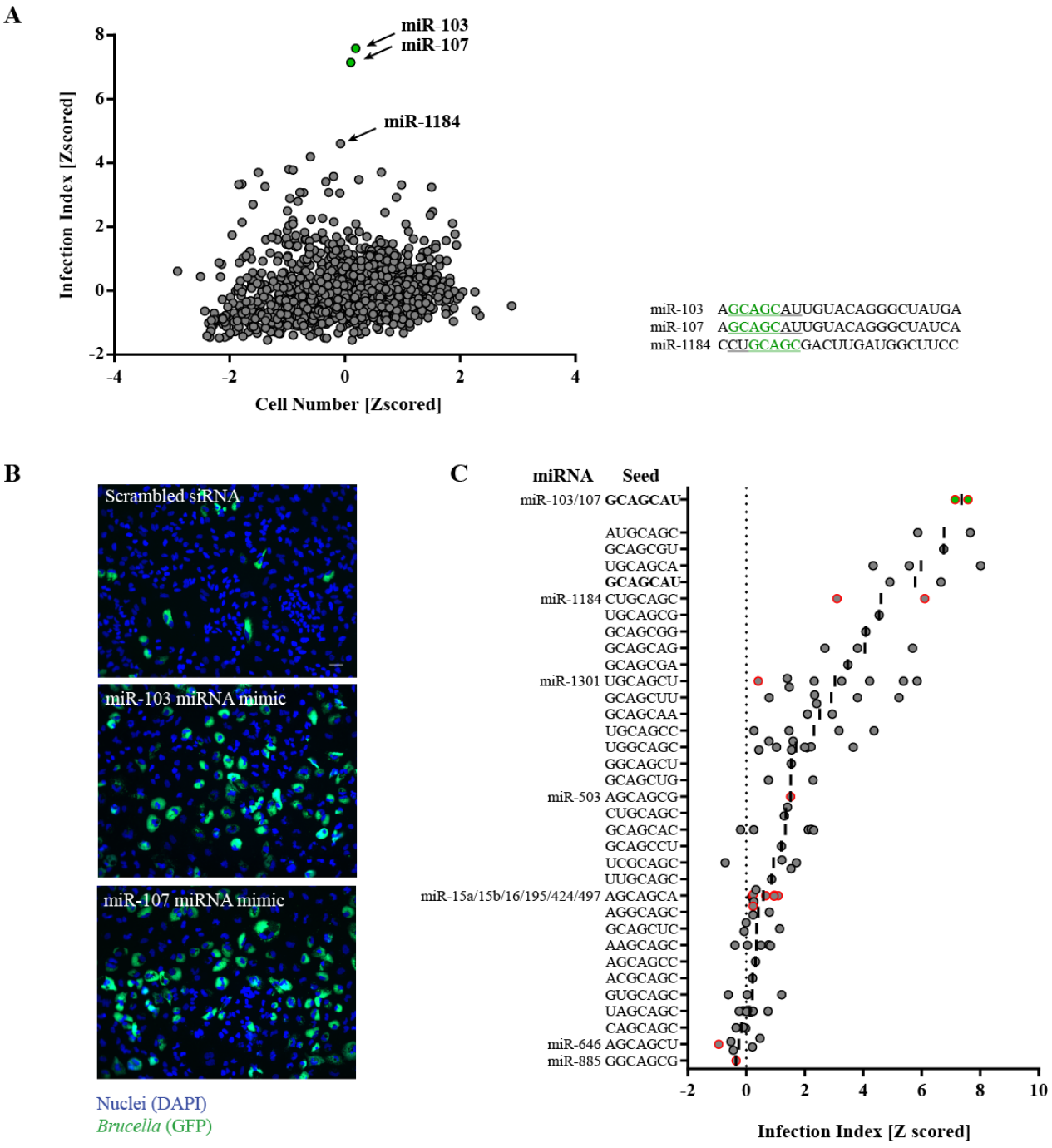

A
miR-103
miR-107
miR-1184
Infection Index [Zscored]
Cell Number [Zscored]
miR-103 AGCAGCAUUGUACAGGGCUAUGA
miR-107 AGCAGCAUUGUACAGGGCUAUCA
miR-1184 CCUGCAGCGACUUGAUGGCUUCC
B
Scrambled siRNA
miR-103 miRNA mimic
miR-107 miRNA mimic
Nuclei (DAPI)
Brucella (GFP)
C
miRNA Seed
miR-103/107 GCAGCAU
AUGCAGC
GCAGCGU
UGCAGCA
GCAGCAU
miR-1184 CUGCAGC
UGCAGCG
GCAGCGG
GCAGCAG
GCAGCGA
miR-1301 UGCAGCU
GCAGCUU
GCAGCAA
UGCAGCC
UGGCAGC
GGCAGCU
GCAGCUG
miR-503 AGCAGCG
CUGCAGC
GCAGCAC
GCAGCCU
UCGCAGC
UUGCAGC
miR-15a/15b/16/195/424/497 AGCAGCA
AGGCAGC
GCAGCUC
AAGCAGC
AGCAGCC
ACGCAGC
GUGCAGC
UAGCAGC
CAGCAGC
miR-646 AGCAGCU
miR-885 GGCAGCG
Infection Index [Z scored]
D
bits
UGCAGC

Figure 1

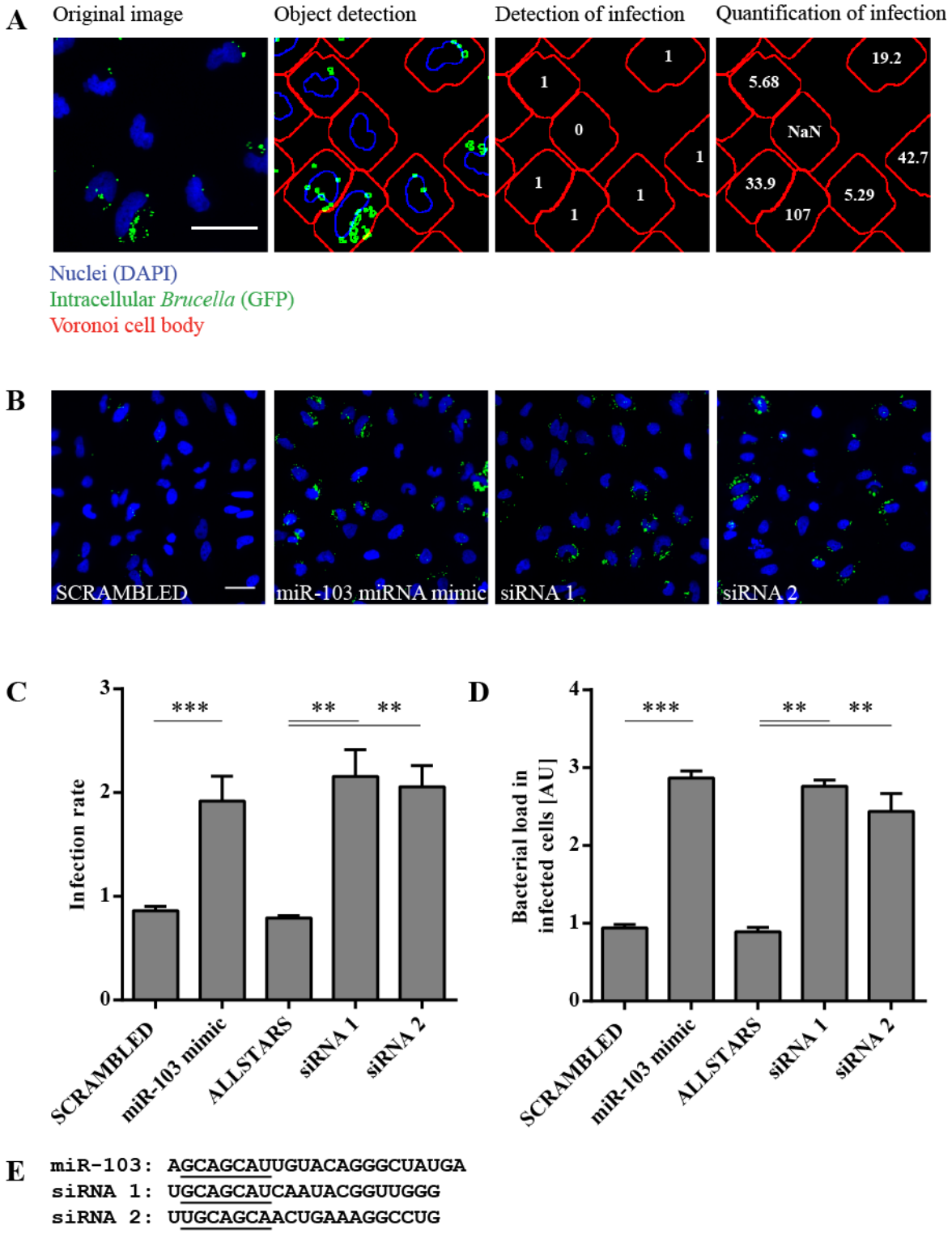

A
Original image
Object detection
Detection of infection
Quantification of infection
1
1
0
1
1
1
1
19.2
5.68
NaN
42.7
33.9
5.29
107
Nuclei (DAPI)
Intracellular Brucella (GFP)
Voronoi cell body
B
SCRAMBLED
miR-103 miRNA mimic
siRNA 1
siRNA 2
C
Infection rate
***
**
**
SCRAMBLED
miR-103 mimic
ALLSTARS
siRNA 1
siRNA 2
D
Bacterial load in infected cells [AU]
***
**
**
SCRAMBLED
miR-103 mimic
ALLSTARS
siRNA 1
siRNA 2
E
miR-103: AGCAGCAUUGUACAGGGCUAUGA
siRNA 1: UGCAGCAUCAAUACGGUUGGG
siRNA 2: UUGCAGCAACUGAAAGGCCUG

Figure 2

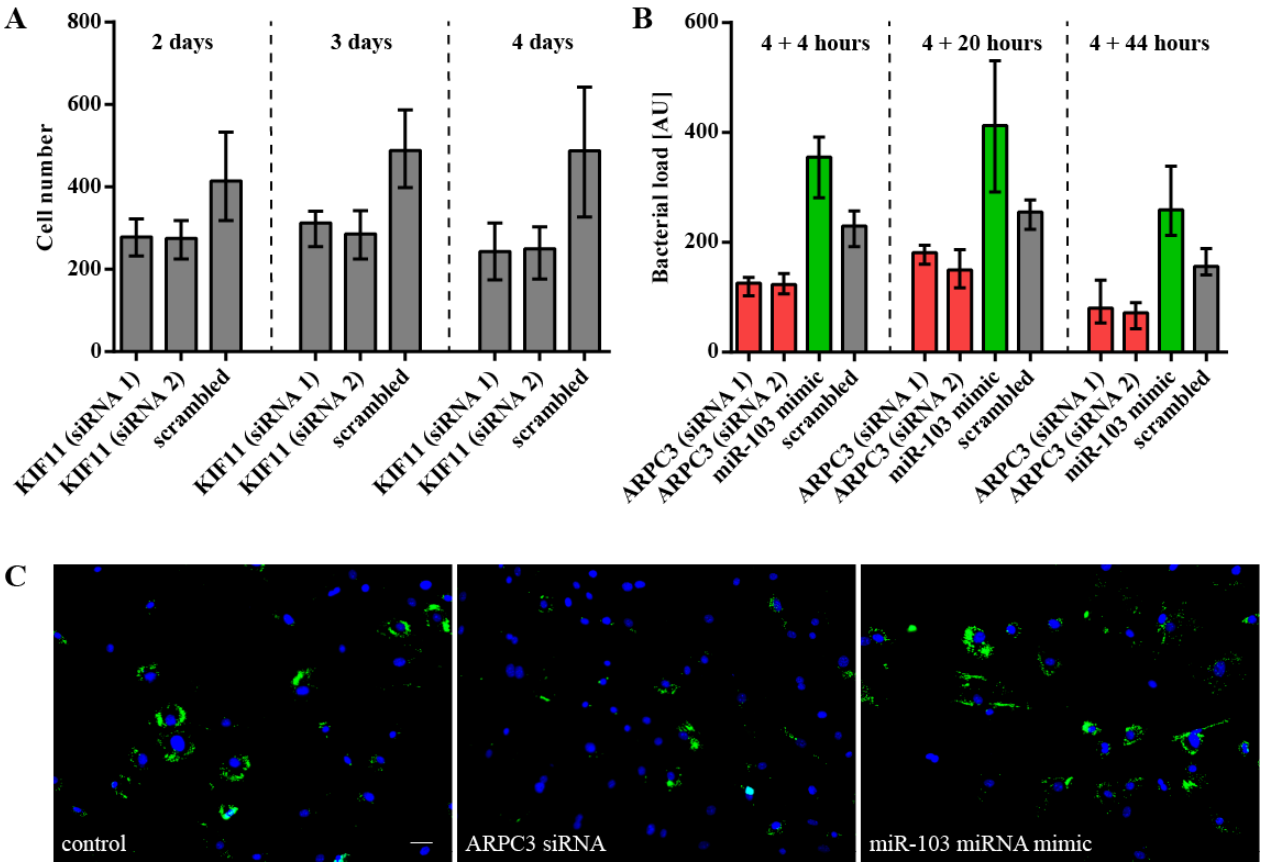

A
800
600
400
200
0
Cell number
2 days
3 days
4 days
KIF11 (siRNA 1)
KIF11 (siRNA 2)
scrambled
B
600
400
200
0
Bacterial load [AU]
4 + 4 hours
4 + 20 hours
4 + 44 hours
ARPC3 (siRNA 1)
ARPC3 (siRNA 2)
miR-103 mimic
scrambled
C
control
ARPC3 siRNA
miR-103 miRNA mimic

Figure 3

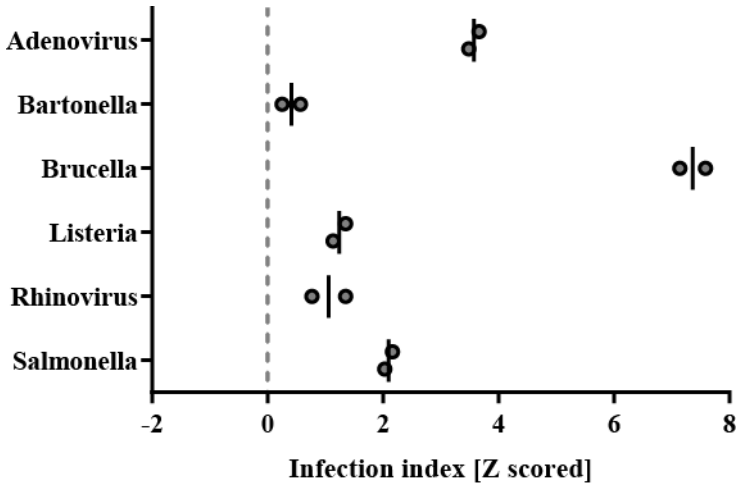

Adenovirus
Bartonella
Brucella
Listeria
Rhinovirus
Salmonella
-2
0
2
4
6
8
Infection index [Z scored]

Figure 4

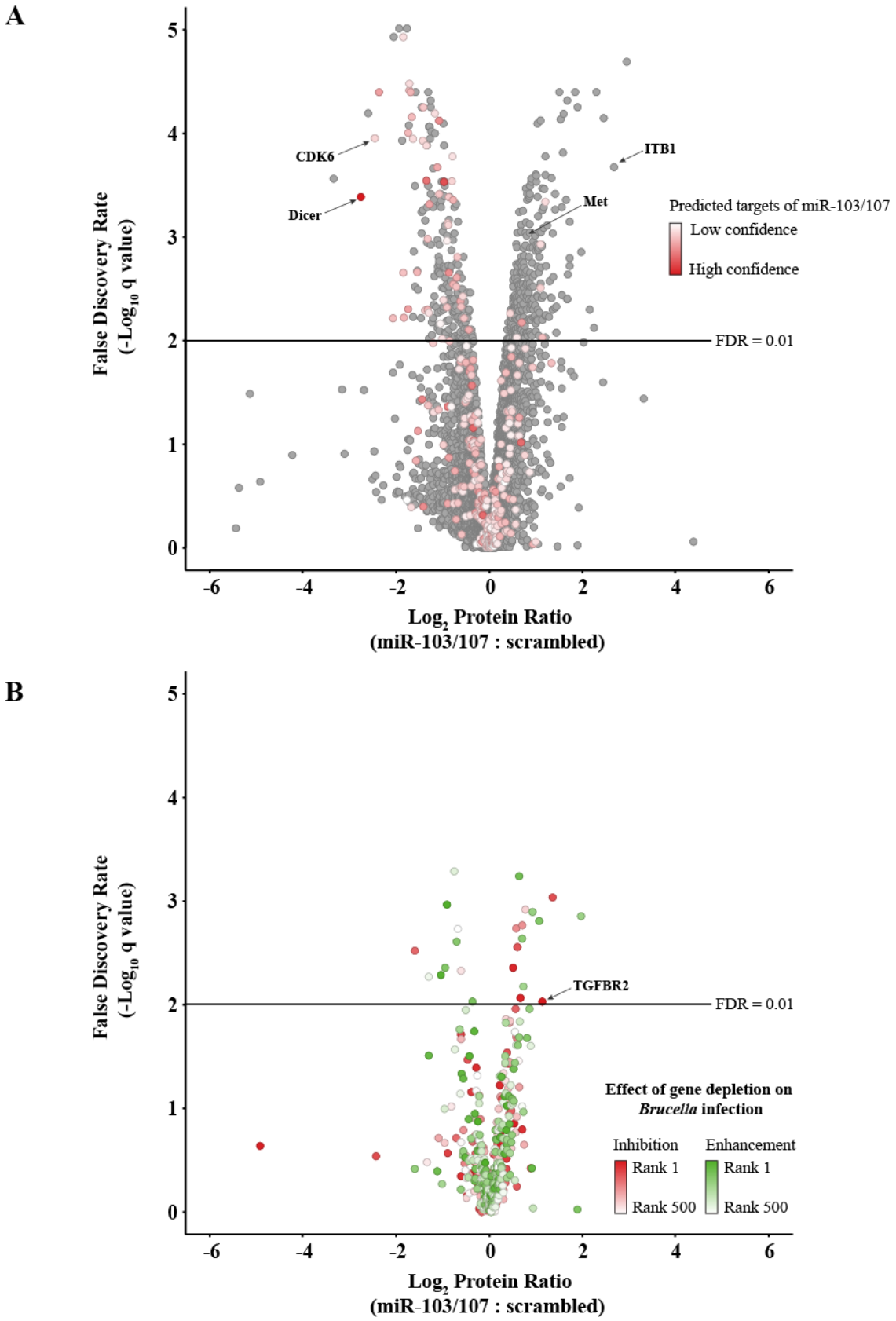

A
False Discovery Rate
(-Log10 q value)
Log2 Protein Ratio
(miR-103/107 : scrambled)
CDK6
Dicer
ITB1
Met
Predicted targets of miR-103/107
Low confidence
High confidence
FDR = 0.01
B
False Discovery Rate
(-Log10 q value)
Log2 Protein Ratio
(miR-103/107 : scrambled)
TGFBR2
FDR = 0.01
Effect of gene depletion on
Brucella infection
Inhibition
Rank 1
Rank 500
Enhancement
Rank 1
Rank 500

Figure 5

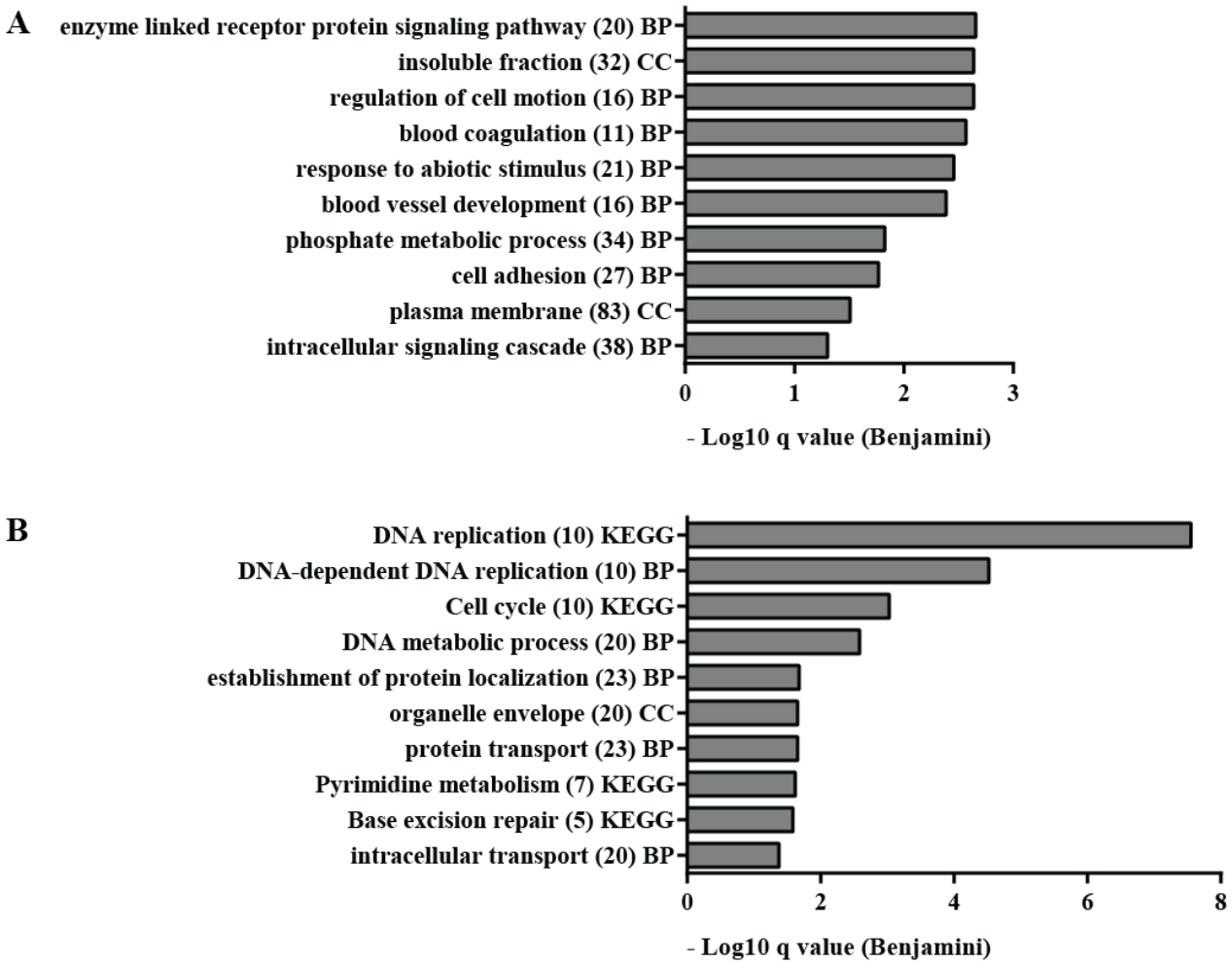

A
enzyme linked receptor protein signaling pathway (20) BP
insoluble fraction (32) CC
regulation of cell motion (16) BP
blood coagulation (11) BP
response to abiotic stimulus (21) BP
blood vessel development (16) BP
phosphate metabolic process (34) BP
cell adhesion (27) BP
plasma membrane (83) CC
intracellular signaling cascade (38) BP
0
1
2
3
- Log10 q value (Benjamini)
B
DNA replication (10) KEGG
DNA-dependent DNA replication (10) BP
Cell cycle (10) KEGG
DNA metabolic process (20) BP
establishment of protein localization (23) BP
organelle envelope (20) CC
protein transport (23) BP
Pyrimidine metabolism (7) KEGG
Base excision repair (5) KEGG
intracellular transport (20) BP
0
2
4
6
8
- Log10 q value (Benjamini)

Figure 6

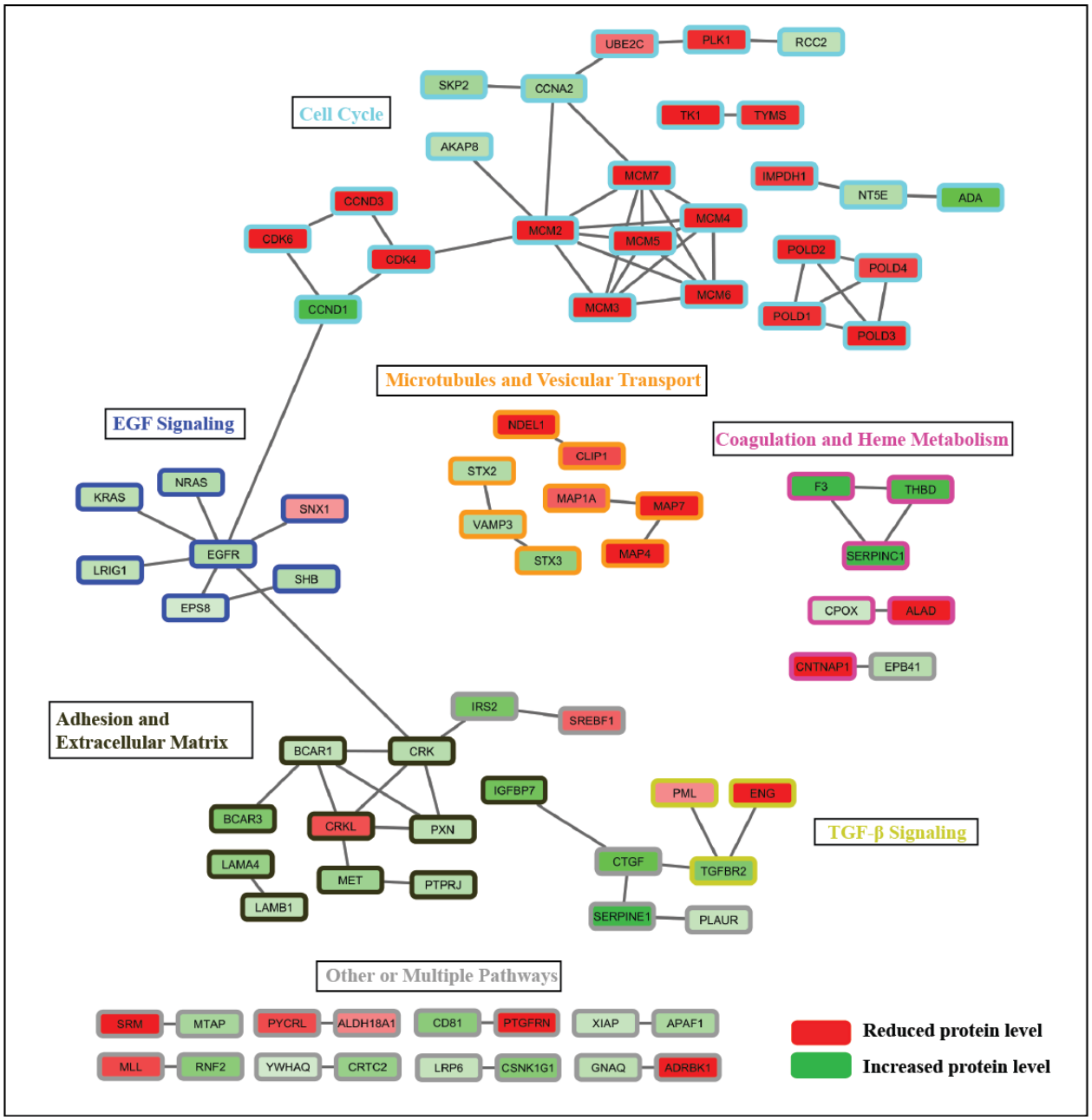

Cell Cycle
UBE2C
PLK1
RCC2
SKP2
CCNA2
TK1
TYMS
AKAP8
MCM7
IMPDH1
NT5E
ADA
CCND3
CDK6
CDK4
MCM2
MCM4
MCM5
MCM6
MCM3
POLD2
POLD4
POLD1
POLD3
CCND1
Microtubules and Vesicular Transport
NDEL1
CLIP1
STX2
MAP1A
MAP7
VAMP3
STX3
MAP4
EGF Signaling
KRAS
NRAS
SNX1
EGFR
LRIG1
SHB
EPS8
Coagulation and Heme Metabolism
F3
THBD
SERPINC1
CPOX
ALAD
CNTNAP1
EPB41
Adhesion and Extracellular Matrix
IRS2
SREBF1
BCAR1
CRK
BCAR3
CRKL
PXN
LAMA4
MET
PTPRJ
LAMB1
IGFBP7
PML
ENG
TGF-β Signaling
CTGF
TGFBR2
SERPINE1
PLAUR
Other or Multiple Pathways
SRM
MTAP
PYCRL
ALDH18A1
CD81
PTGFRN
XIAP
APAF1
MLL
RNF2
YWHAQ
CRTC2
LRP6
CSNK1G1
GNAQ
ADRBK1
Reduced protein level
Increased protein level

Figure 7

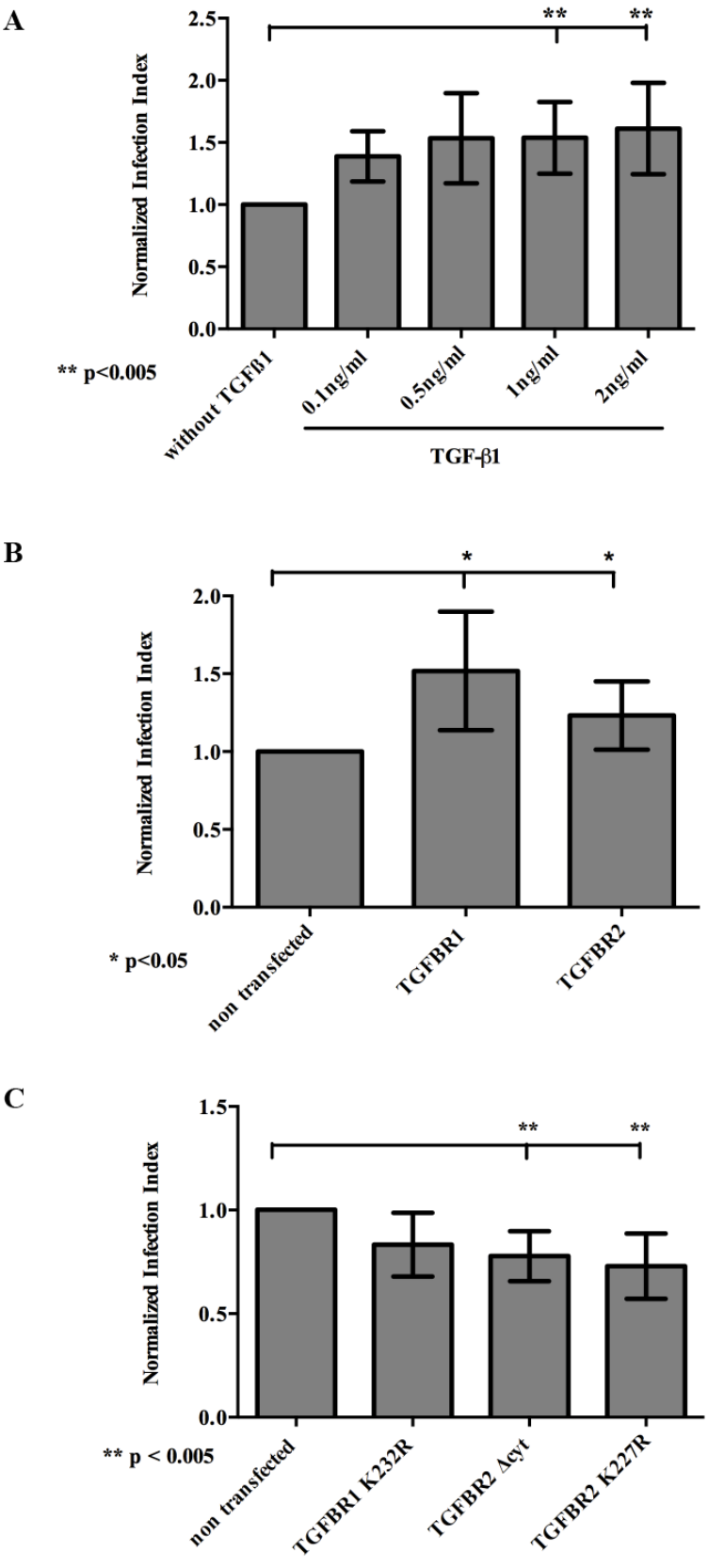

A
Normalized Infection Index
2.5
2.0
1.5
1.0
0.5
0.0
**
**
** p<0.005
without TGFß1
0.1ng/ml
0.5ng/ml
1ng/ml
2ng/ml
TGF-β1
B
Normalized Infection Index
2.0
1.5
1.0
0.5
0.0
*
*
* p<0.05
non transfected
TGFBR1
TGFBR2
C
Normalized Infection Index
1.5
1.0
0.5
0.0
**
**
** p < 0.005
non transfected
TGFBR1 K232R
TGFBR2 Δcyt
TGFBR2 K227R

Figure 8

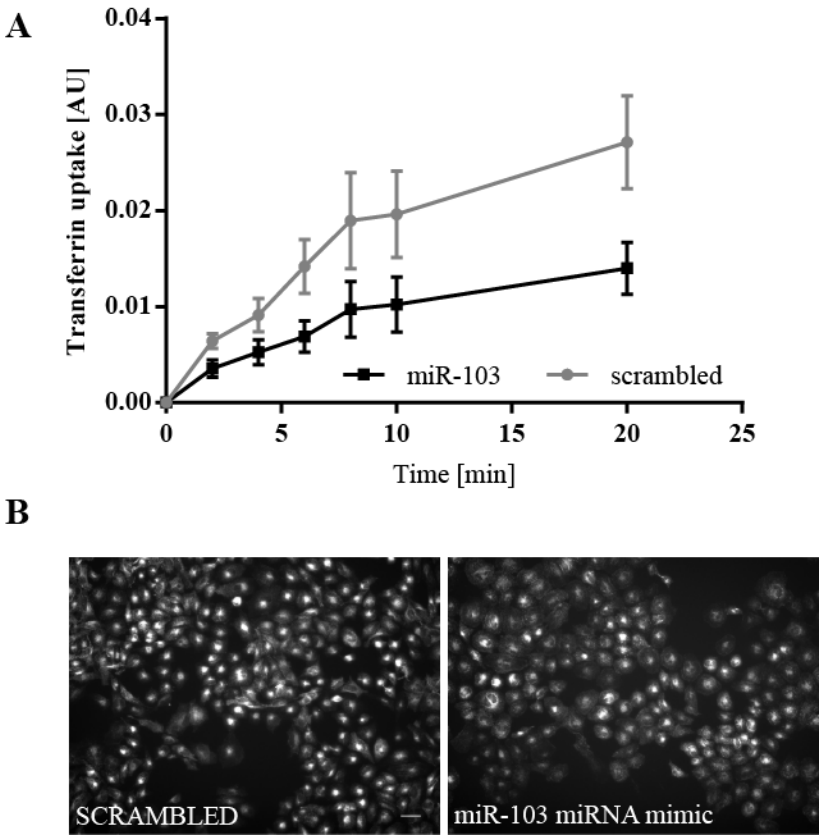
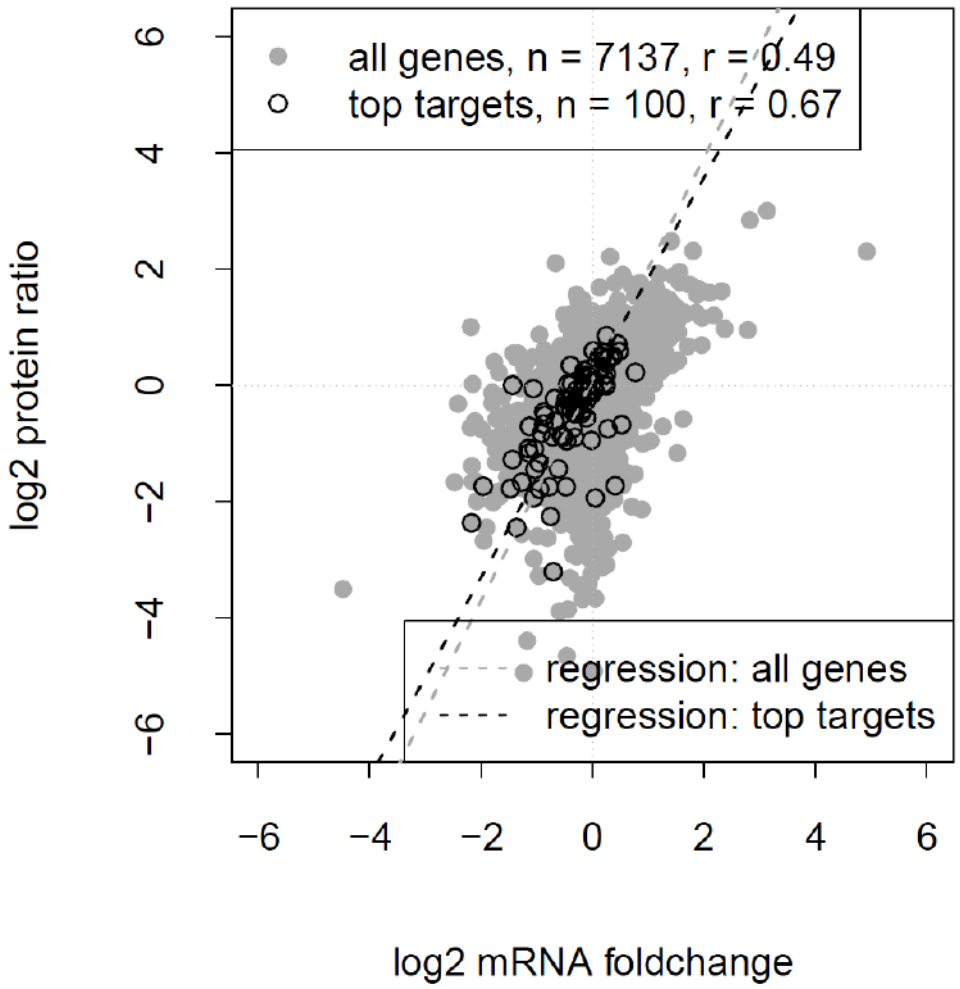
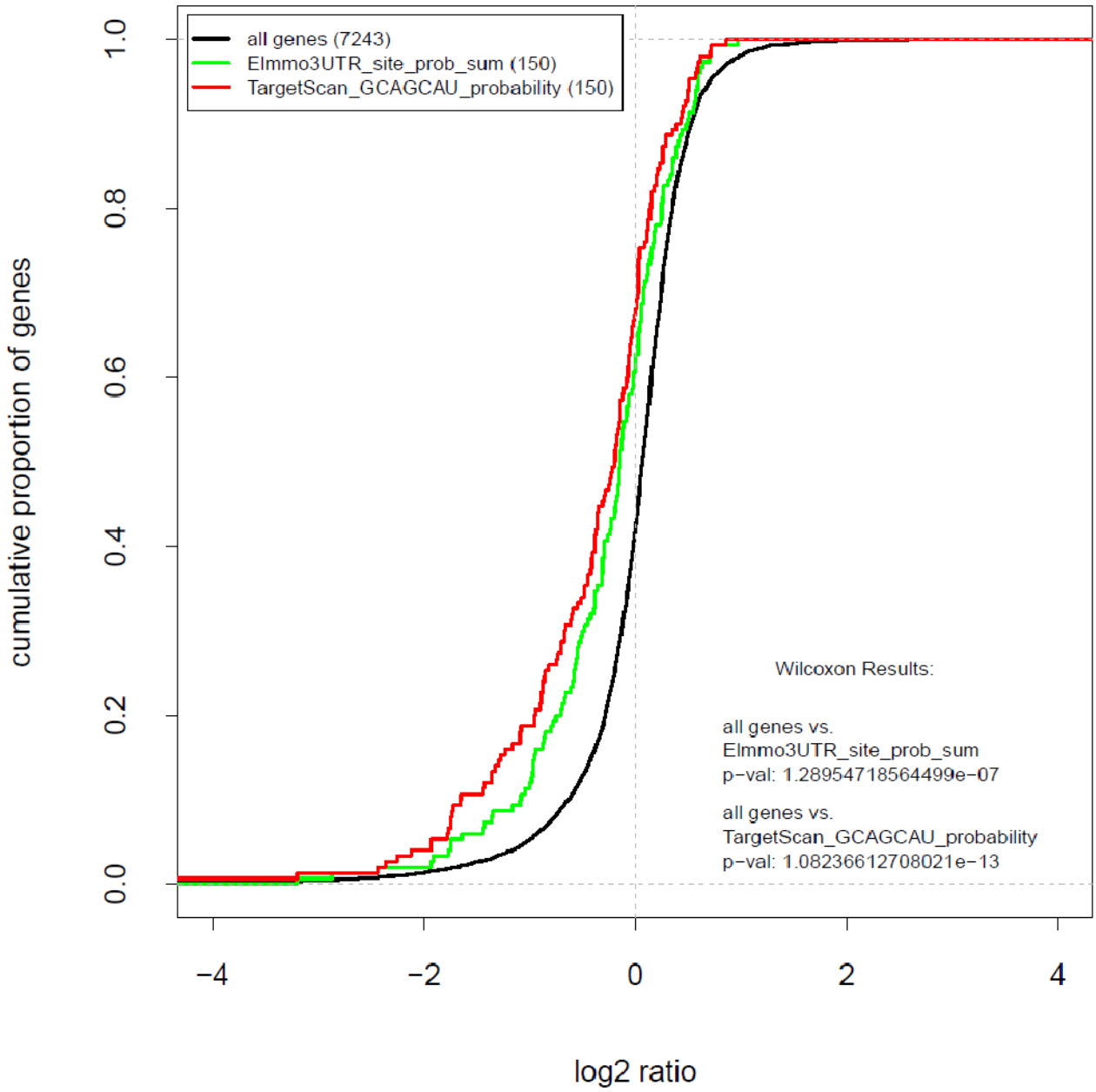

A
0.04
0.03
0.02
0.01
0.00
Transferrin uptake [AU]
0
5
10
15
20
25
Time [min]
miR-103
scrambled
B
SCRAMBLED
miR-103 miRNA mimic

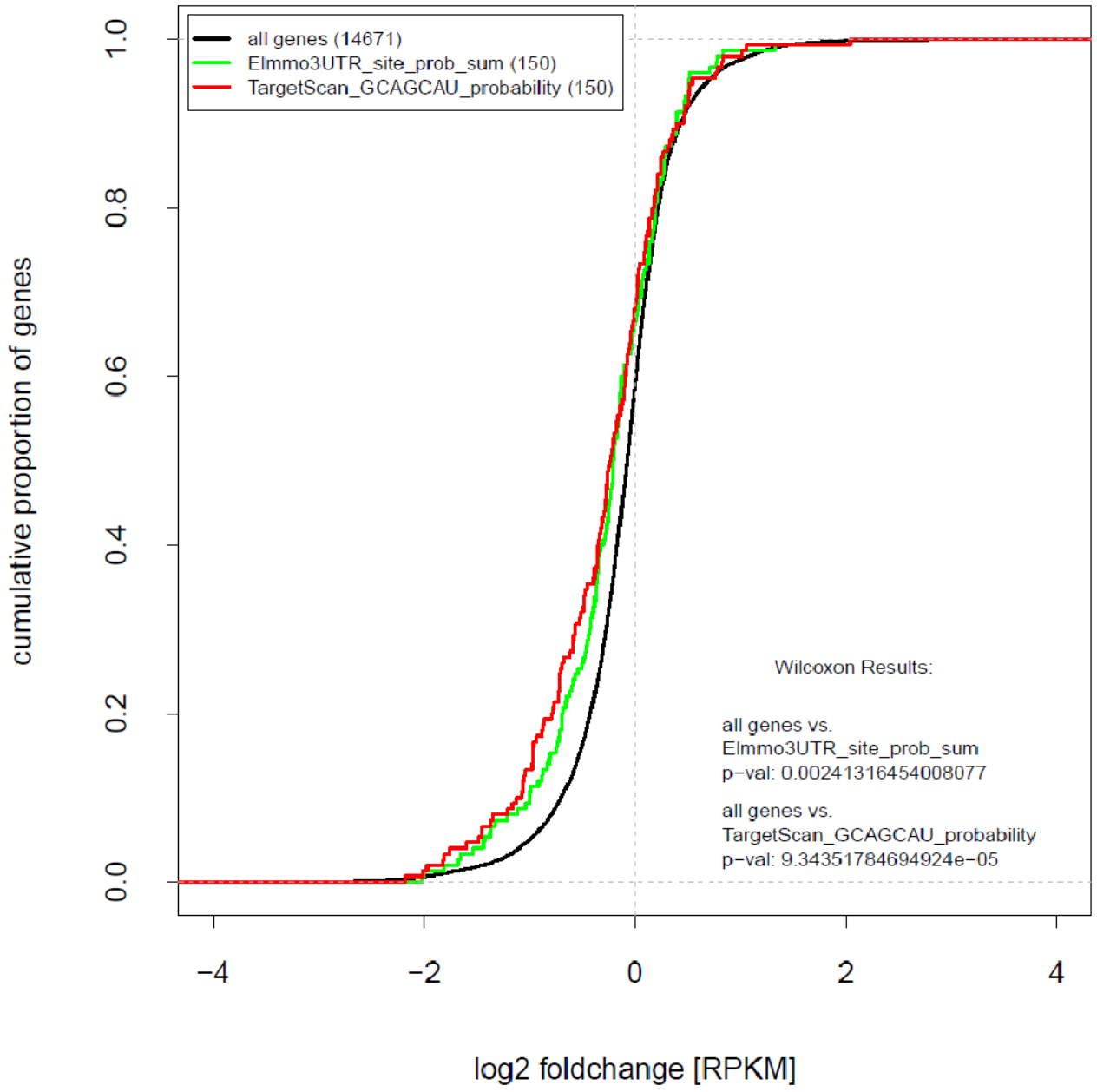
Figure 9



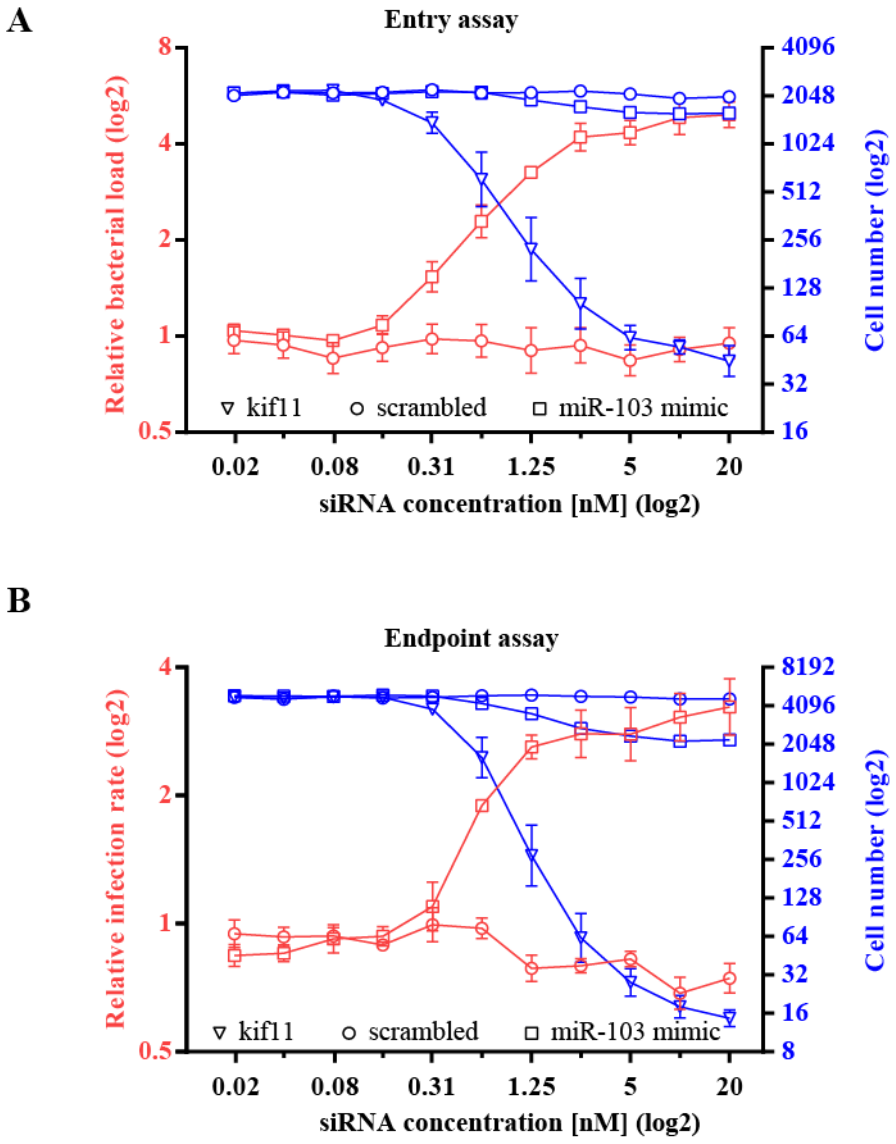
Supplementary Figure 1



Supplementary Figure 2



Supplementary Figure 3



Supplementary Figure 4

miRNA	Seed sequence	Infection rate (Z scored)
hsa-miR-103a-3p	GCAGCAU	7.59
hsa-miR-107	GCAGCAU	7.14
hsa-miR-1184	CUGCAGC	4.60
hsa-miR-653	UGUUGAA	4.19
hsa-miR-3177-3p	GCACGGC	3.80
hsa-miR-520b	AAGUGCU	3.78
hsa-miR-3117-3p	UAGGACU	3.71
hsa-miR-210	UGUGCGU	3.70
hsa-miR-30a-5p	GUAAACA	3.57
hsa-miR-17-3p	CUGCAGU	3.48
hsa-miR-664a-3p	AUUCAUU	-1.49
hsa-miR-92b-3p	AUUGCAC	-1.49
hsa-miR-511	UGUCUUU	-1.51
hsa-miR-301a-3p	AGUGCAA	-1.51
hsa-miR-1293	GGGUGGU	-1.52
hsa-miR-148a-5p	AAGUUCU	-1.55
hsa-miR-629-3p	UUCUCCC	-1.55
hsa-miR-548v	GCUACAG	-1.61
hsa-miR-1908	GGCGGGG	-1.66
hsa-miR-3678-5p	CCGUACA	-2.15

Table 1

3.4 RESEARCH ARTICLE IV (PUBLISHED)

Simultaneous analysis of large-scale RNAi screens for pathogen entry

Pauli Rämö[†,1], Anna Drewek[†,2], Cécile Arrieumerlou[1], Niko Beerenwinkel[4,5], Houchaima Ben-Tekaya[1], Bettina Cardel[3], Alain Casanova[1], Raquel Conde-Alvarez[7], Pascale Cossart[6], Gábor Csúcs[11], Simone Eicher[1], Mario Emmenlauer[1], Urs Greber[3], Wolf-Dietrich Hardt[9], Ari Helenius[8], Christoph Kasper[1], Andreas Kaufmann[11], Saskia Kreibich[9], Andreas Kühbacher[6], Peter Kunszt[13], Shyan Huey Low[1], Jason Mercer[8], Daria Mudrak[3], Simone Muntwiler[1], Lucas Pelkmans[3], Javier Pizarro-Cerdá[6], Michael Podvinec[12], Eva Pujadas[12], Bernd Rinn[4,5], Vincent Rouilly[12], Fabian Schmich[4], Juliane Siebourg-Polster[4], Berend Snijder[3], Michael Stebler[11], Gabriel Studer[1], Ewa Szczurek[4,5], Matthias Truttmann[1], Christian von Mering[3], Andreas Vonderheit[10], Artur Yakimovich[3], Peter Bühlmann[2] & Christoph Dehio*[,1]

[†]Shared first authors

3.4.1 Statement of own contribution

Brucella infection screens were conducted together with Dr. Shyan H. Low and Dr. Raquel Conde-Alvarez.

3.4.2 Manuscript

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

BMC Genomics

RESEARCH ARTICLE Open Access

Simultaneous analysis of large-scale RNAi screens for pathogen entry

Pauli Rämö[1†], Anna Drewek[2†], Cécile Arrieumerlou[13], Niko Beerenwinkel[4,5], Houchaima Ben-Tekaya[1], Bettina Cardel[3], Alain Casanova[1], Raquel Conde-Alvarez[7], Pascale Cossart[6], Gábor Csúcs[11], Simone Eicher[1], Mario Emmenlauer[1], Urs Greber[3], Wolf-Dietrich Hardt[9], Ari Helenius[8], Christoph Kasper[1], Andreas Kaufmann[11], Saskia Kreibich[9], Andreas Kühbacher[6], Peter Kunszt[14], Shyan Huey Low[1], Jason Mercer[8], Daria Mudrak[3], Simone Muntwiler[1], Lucas Pelkmans[3], Javier Pizarro-Cerdá[6], Michael Podvinec[12], Eva Pujadas[12], Bernd Rinn[4,5], Vincent Rouilly[12], Fabian Schmich[4], Juliane Siebourg-Polster[4], Berend Snijder[3], Michael Stebler[11], Gabriel Studer[1], Ewa Szczurek[4,5], Matthias Truttmann[1], Christian von Mering[3], Andreas Vonderheit[10], Artur Yakimovich[3], Peter Bühlmann[2] and Christoph Dehio[1*]

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-targets 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 the 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 resource 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.dehio@unibas.ch
†Equal contributors
[1]Focal Area Infection Biology, Biozentrum, University of Basel, Klingelberstrasse 70, CH-4056 Basel, Switzerland
Full list of author information is available at the end of the article

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-TARGETpIus) with 4 siRNAs per gene. Additionally, we performed screens where we targeted each kinase with a pool of the four Dharmacon siRNAs (Human ON-TARGETpIus SMARTpooI). 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 for *L. 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

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

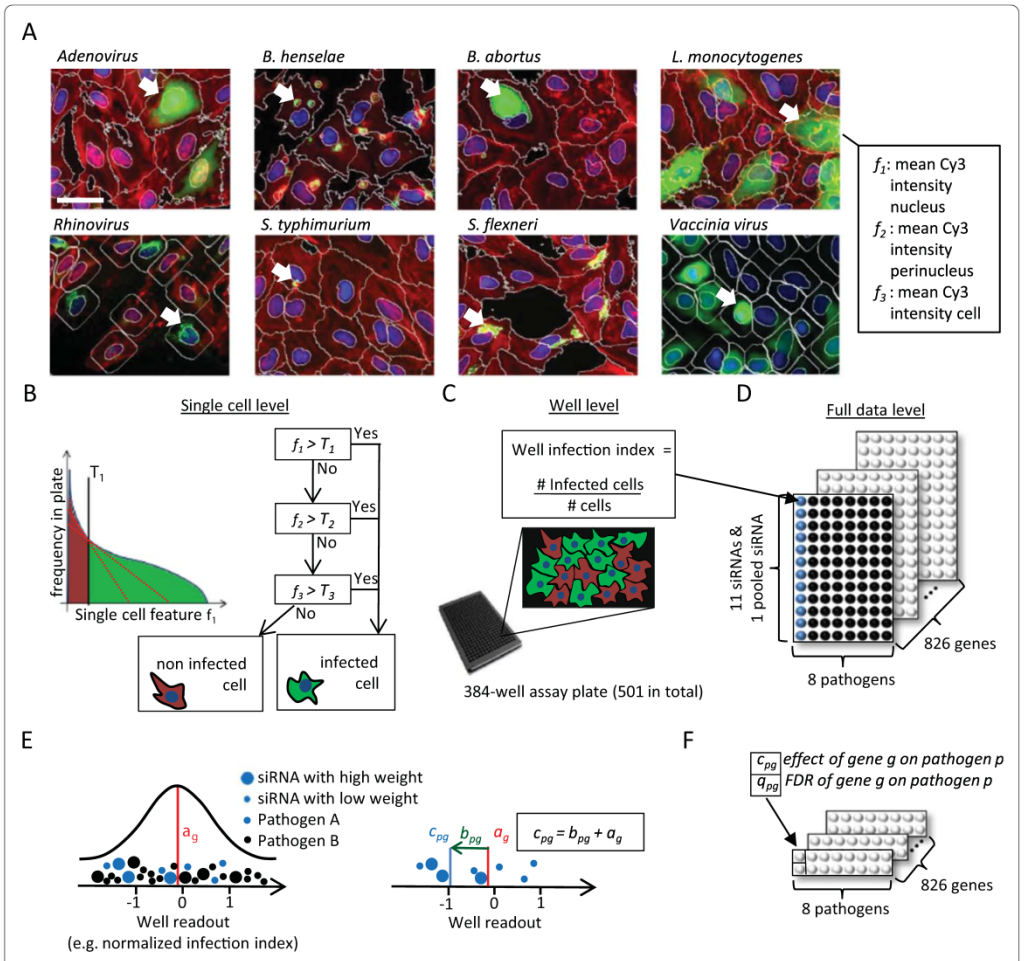


Figure 1 Overview of InfectX high-content datasets, image analysis, and Parallel Mixed Model (PMM). (A) The figure shows example images of the different pathogens after siRNA transfection and the infection phase. The arrows indicate typical infectious phenotypes for each pathogen. The list shows an example of three single cell features that we extracted to identify infected cells for *L. monocytogenes*. The scale bar has a length of 50 μm. **(B)** For each selected feature, we defined the optimal threshold that separated best between uninfected and infected cells via histograms. We used the thresholds in the Decision Tree Infection Scoring (DTIS) algorithm to classify between infected (green) and non-infected cells (red). We optimized this procedure for each pathogen separately. **(C)** For each well in a 384-well assay plate, we calculated the infection index by dividing the number of infected cells (green) by the total number of cells (green and red). **(D)** The figure shows a schematic representation of the input data for the statistical analysis. Each point represents the average infection index over all its replicate wells (wells with the same siRNA set targeting the same gene and pathogen). **(E)** The Parallel Mixed Model (PMM) algorithm fits via a normal distribution for an overall effect a_g to all data of gene g. The second plot shows the correction of the overall effect a_g within every pathogen by an estimate b_{pg} in order to obtain to an pathogen and gene specific effect c_{pg}. The different sizes of the data points refer to weights w_s which can be incorporated in the PMM to depict the quality of the siRNA. **(F)** The figure shows a schematic representation of the final output of PMM. The model estimates gene effects c_{pg} for each gene and pathogen and provides corresponding local False Discovery Rates q_{pg}.

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

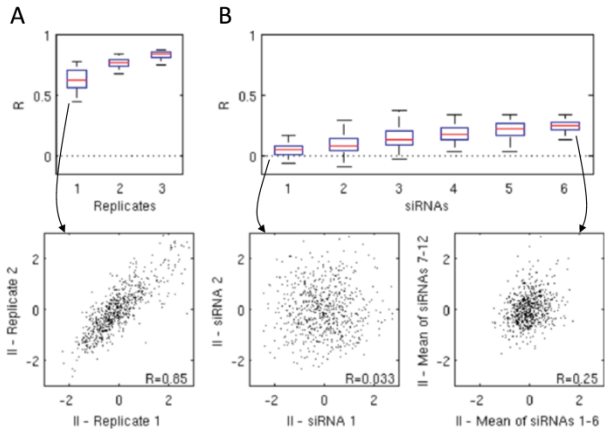


Figure 2 Using more siRNAs adds power and yields reproducible results. (A) The three boxplots show Pearson correlation coefficients *R* between screens performed using the same siRNA set. The numbers 1 to 3 correspond to the total number of replicate screens that we averaged and compared to another distinct set of replicate screens, averaged over the same number. We resampled the replicate screens up to 500 times. The scatter plot shows an example for the correlation of infection indices from the duplicate of *Adenovirus* Dharmacon pooled screen. **(B)** The set of six boxplots shows the Pearson correlation coefficients of the averaged readouts from 1 to 6 siRNA sets. The scatter plots depict the correlation of infection indices for *Adenovirus*, the first between two different single siRNAs and the second between each an average over six siRNAs.

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 8×12 = 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:

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

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

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

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

$$c_{pg} = a_g + b_{pg}.$$

A positive estimated c_{pg} effect means that the infection level was enhanced if the corresponding gene g is knocked down. A negative effect means that the infection level was reduced. To distinguish hit genes, PMM provides as second step an estimate q_{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.

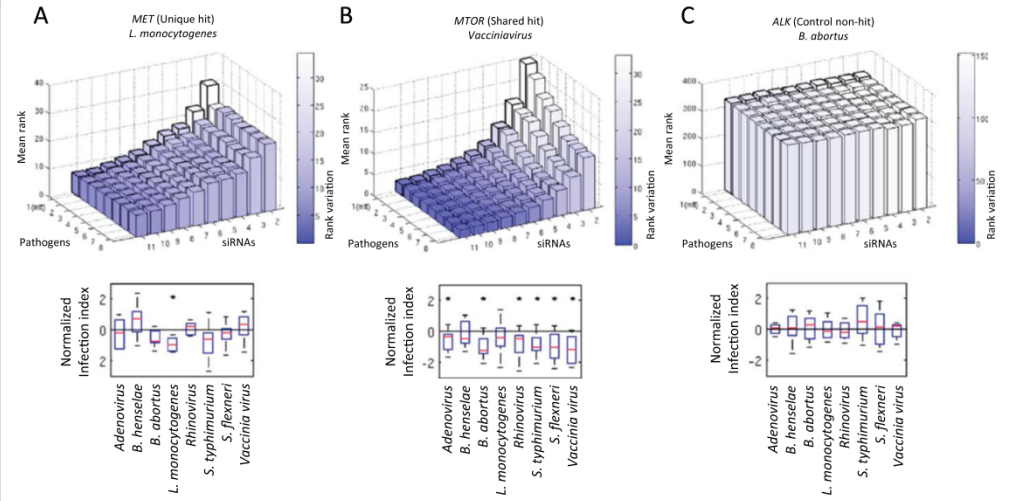


Figure 3 Parallel screens add power to find more shared hits. (A) We varied the number of used siRNAs and pathogens and calculated the rank of *MET* for *L. monocytogenes* in the ordered list of hit genes. We used PMM (and MTT for the case of one pathogen) over 1000 random resampling rounds with replacement. 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 in **(A)**, but now with *MTOR* for *Vaccianiavirus*. The boxplot shows that *MTOR* is a shared significant hit for several pathogens. **(C)** The figure shows the same experiment as in **(A)** but with non-hit *ALK* for *B. abortus* for control.

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

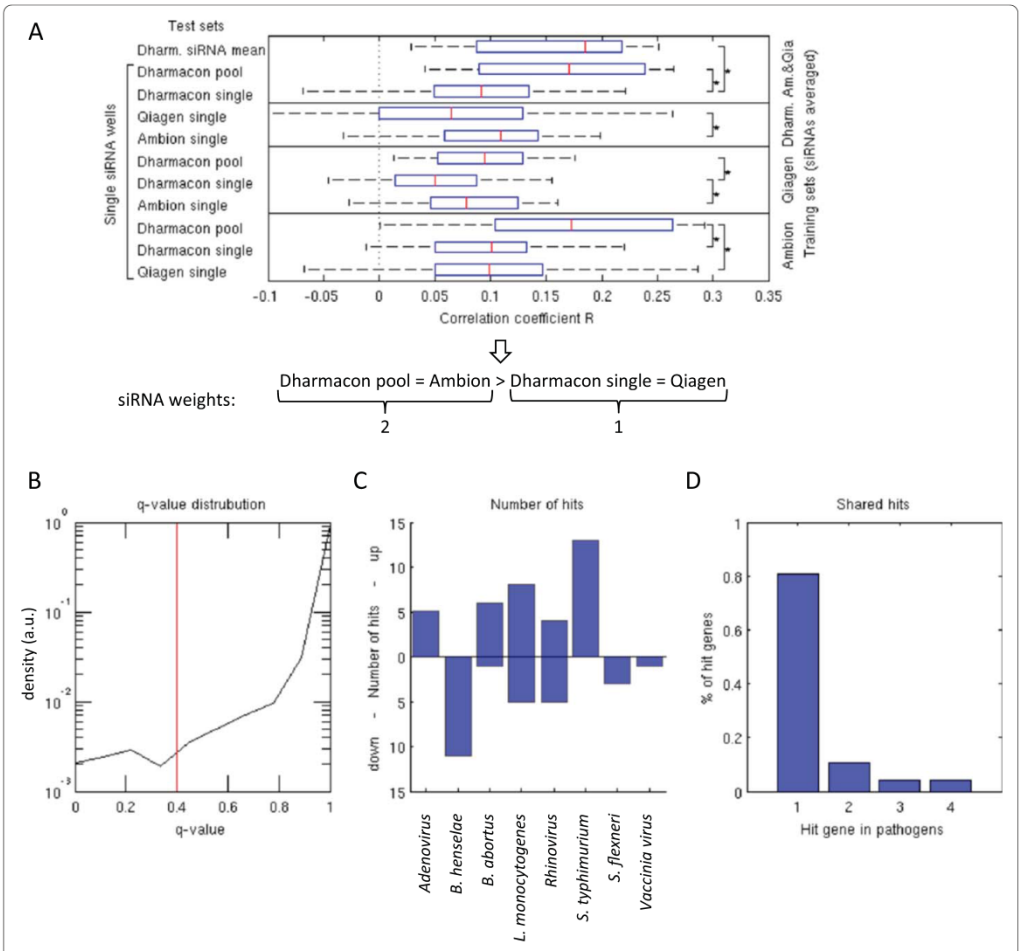


Figure 4 Statistics on used siRNA libraries and hits. (A) We weighted siRNAs based on their library quality. Each vertical compartment in the plot corresponds to a training set of siRNAs. We averaged data in the training set from the siRNAs of the specific manufacturers. Each boxplot corresponds to a test set of single siRNAs from different manufacturers (except "Dharm. siRNA mean" which is the average of 4 Dharmacon unpooled siRNAs). Y-axis refers to Pearson correlation coefficients *R* between the training and test sets. A star corresponds to significant differences in the correlation coefficients (Mann–Whitney-U-test $p < 0.05$) between pairs of manufacturers. We used all screens, infection index, and cell number well readouts in the analysis. We used the results of this analysis to assign siRNA weights to siRNAs from different library manufacturers as shown below the plot. **(B)** The histogram shows obtained FDR q-values from all screens using the infection index readouts. The red line shows the FDR-threshold of 0.4. **(C)** The bar shows number of up and down hits for different pathogens. **(D)** The bar plot shows the number of hit genes that were shared between pathogens.

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 manufactures (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

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

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

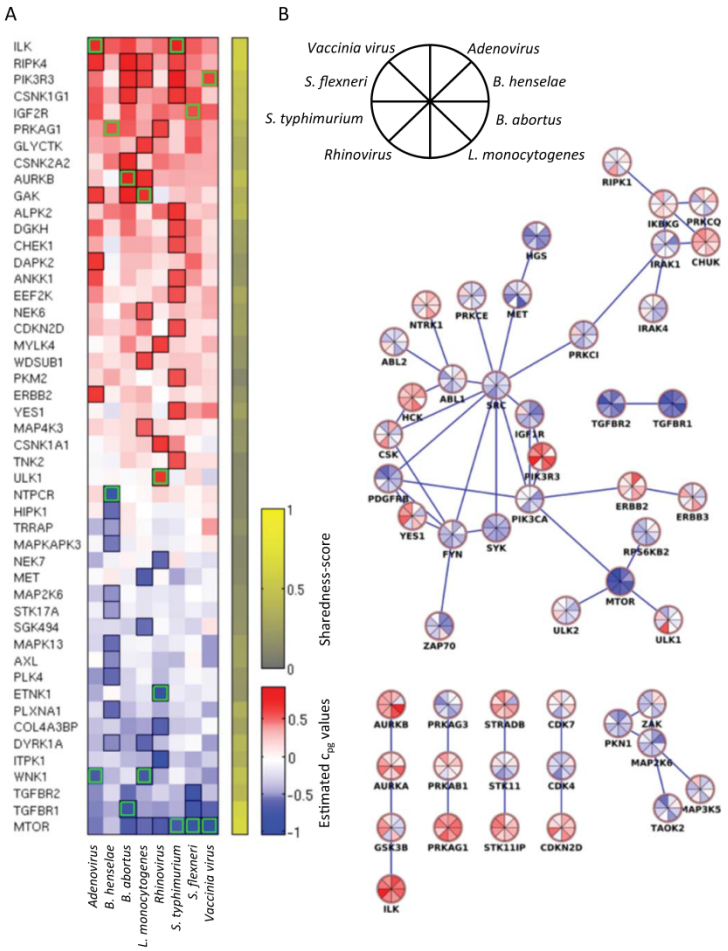


Figure 5 Summary of screening hits for all pathogens. (A) The heat map shows all genes which were significant (FDR < 0.4) at least for one pathogen. We ordered the genes by their averaged c-values over all pathogens. The colors correspond to the estimated c-values. The black outlines indicate significant hits (FDR < 0.4) and the green outlines high-light 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:

$$s_g = \frac{1}{2}\left(\left(1 - Mean_p\left(q_{pg}\right)\right) + \frac{\sum_p\left(q_{pg} < 1\right)}{P}\right).$$

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 c_{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 (prob > 0.9 and prob > 0.7) in the top of the ordered lists of genes based on their estimated c_{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.

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

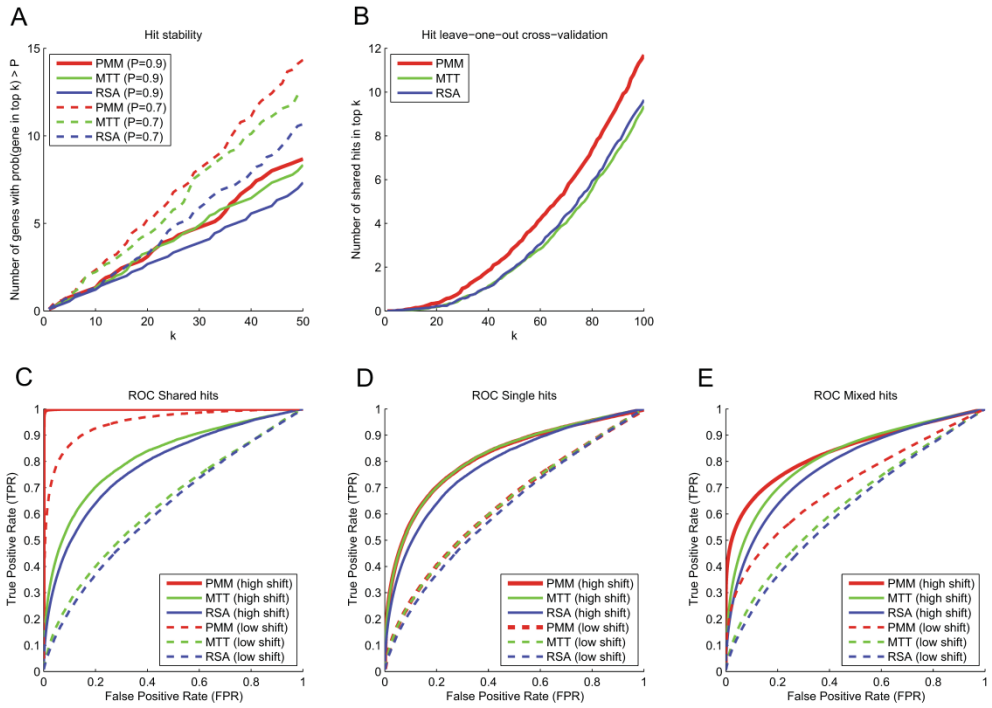


Figure 6 Performance statistics of hit ranking methods. (A) The figure shows stability curves using the three different methods (PMM, MTT and RSA). The y-axis denotes the number of genes that were found with probability higher than 0.7 (dashed lines) and 0.9 (solid lines) in the top k (x-axis) of the list of ranked genes. The curves show the average over all eight pathogens. **(B)** The figure shows hit overlaps of cross-validated siRNA sets between the set of 10 unpooled siRNA libraries and the remaining siRNA library using the three tested gene ranking methods as a function of hit threshold *k*. The curves show the average over all eight pathogens. **(C)** The figure shows ROC-curves for PMM, MTT and RSA applied on simulated data containing only hits that were shared between all pathogens. The dashed and solid lines indicate whether the shifts were generated by a low or high shift away from zero. The PMM method outperformed the reference hit detection method. **(D)** The figure shows ROC-curves for PMM, MTT and RSA applied on simulated data containing only unique hits for all pathogens. PMM and Moderated T-Test performed equally well. **(E)** The figure shows ROC-curves for simulated data with a mixed hit structure of both unique and shared hits. The PMM method outperformed the reference hit detection method.

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

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

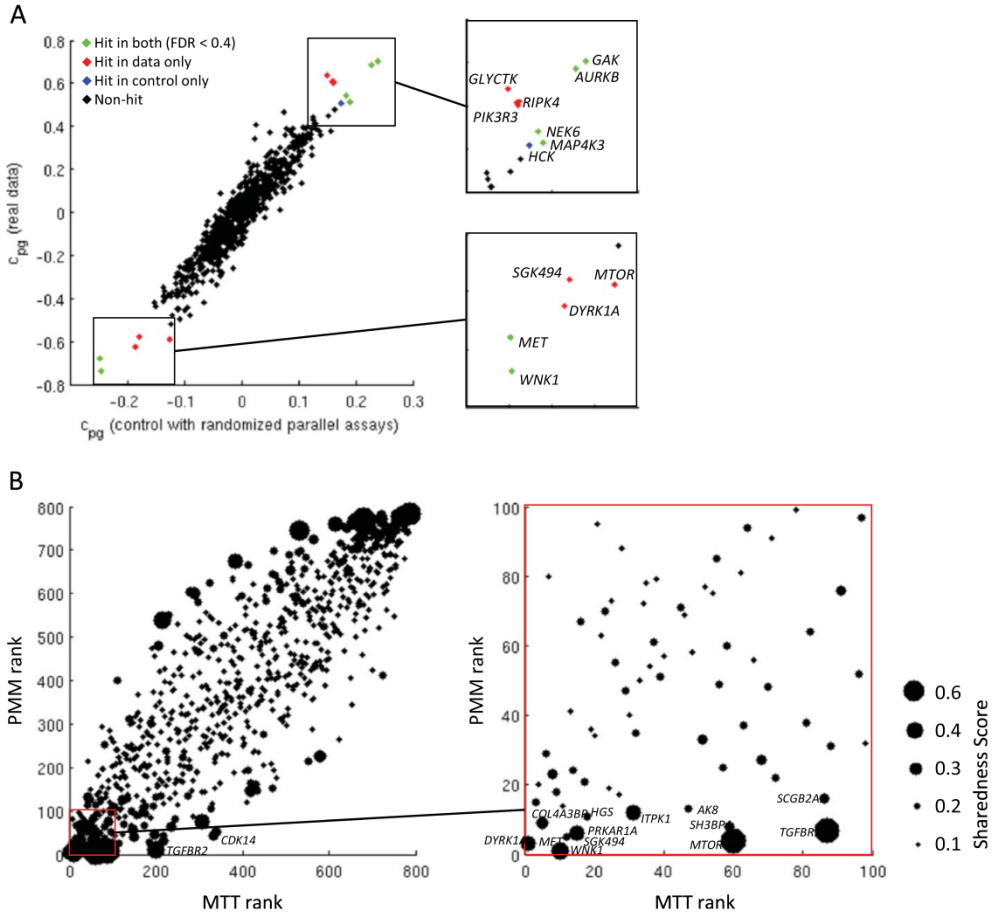


Figure 7 Summary of differences of PMM top hits compared to other hit scoring methods. (A) Y-axis shows the PMM gene ranking for *L. monocytogenes*. X-axis is the same, but we randomized the other 7 parallel assays. The colors correspond to hit genes (FDR < 0.4) in different cases. Parallelism yielded only a slight effect on the ranking, but added genes to the list of significant hit genes. **(B)** The scatter plot shows PMM hit ranking (y-axis) compared to the MTT hit ranking (x-axis) for *L. monocytogenes*. 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 ($q_{pg} < 0.4$) 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*, *Enterovirus71*, *Coxackievirus, 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 *TGFB* 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 *ETNK1* 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* (*InlA*) or *InlB* 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 microRNA 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 gene's 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% CO_2 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 ddH_2O. 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% CO_2 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 Ad2_ΔE3B-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 NH_4Cl, permeabilized with 25 μl 0.1% Triton X-100 (Pharmaciebiothek). 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/NaN_3.

Bartonella henselae -specific protocol

Bacterial strain SEB0109: *Bartonella henselae* ATCC49882^T Δ*bepG* 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% CO_2 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 (Applichem) and were incubated at 35°C in 5% CO_2 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 ug/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% CO_2, 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% CO_2 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% CO_2 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 *InlC*. 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 μl PBS per well cells were stained with 10 μl/well of the secondary staining solution (0.2% saponin, PBS) containing Alexa Fluor-546 coupled anti-rabbit antibody (1:250, Invitrogen), DAPI (0.7 μg/ml, Roche), and DY-647-Phalloidin (2 U/ml, Dyomics). After four washes with 40 μl PBS per well, the cells were kept in 40 μl 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 μl of an infection media/BSA (DMEM supplemented with GlutaMAX, 30 mM $MgCl_2$ and 0.2% BSA, Invitrogen). Screening plates were incubated for 7 h at 37°C, and cells were fixed by adding 33 μl 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 NH_4Cl, permeabilized with 50 μl 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 μg/ml) in PBS/1% BSA. Cells were washed 3 times with PBS and incubated with 25 μl secondary staining solution (PBS/1% BSA) containing Alexa Fluor 488 secondary antibody (1 μg/ml, Invitrogen), DAPI (1 μg/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 μl PBS/NaN_3.

Salmonella typhimurium -specific protocol

All liquid handing 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 μg/ml streptomycin and 50 μg/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 μg/ml streptomycin, which reached an $OD_{600nm} \approx 1.0$ after the respective 4 h of incubation time. To perform the infection, 16 μl of diluted *S. typhimurium* (moi = 80) were added to the HeLa cells. After 20 min of incubation at 37°C and 5% CO_2, the *S. typhimurium*-containing media was replaced by 60 μl DMEM/10% FCS containing 50 μg/μl streptomycin and 400 μg/μl gentamicin to kill all remaining extracellular bacteria. After additional 3 h 40 min incubation at 37°C and 5% CO_2, cells were fixed by adding 35 μl 4% PFA, 4% sucrose in PBS for 20 min at RT. The fixation solution was removed by adding 60 μl PBS containing 400 μg/ml gentamicin. Cells were permeabilized for 5 min with 40 μl 0.1% Triton X-100 (Sigma-Aldrich). Afterwards 24 μl 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 μl PBS containing 400 μg/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 μl 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% CO_2. After 30 min of infection, 75 μl were aspirated from each well and monensin (Sigma) and gentamicin (Gibco) were added to a final concentration of 66.7 μM and 66.7 μg/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 dispension 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 μg/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 handing 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

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

RNAi-transfected cell plates and replaced with 40 μl 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 μl DMEM/10% FCS. 8 h after infection 40 μl of DMEM/10%FCS containing 20 μM 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 μl 18% PFA for 30 min followed by two PBS washes of 80 μl. For immunofluorescence staining of EGFP, cells were incubated for 2 h in 30 μl 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 μl PBS, followed by the addition of 30 μl 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 μl PBS after 1 h incubation in secondary staining solution followed by the addition of 80 μl H_2O.

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 3×3 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 g as y_{pgs}. The linear mixed model of PMM is defined as the following linear model

$$y_{pgs} = \mu_p + a_g + b_{pg} + \varepsilon_{pgs},$$
$$a_g \sim N\left(0, \sigma_a^2\right),\ b_{pg} \sim N\left(0, \sigma_b^2\right),\ \varepsilon_{pgs} \sim N\left(0, \sigma_\varepsilon^2\right),$$

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 pathogens, 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 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

$$fdr(c) = P(\text{No Hit} \mid c) = \frac{\pi_0 f_0(c)}{f(c)}$$
$$f_0 \sim N\left(0, \sigma_a^2 + \sigma_b^2\right), f_1 \sim N\left(\theta, \sigma_a^2 + \sigma_b^2\right),$$
$$f(c) = \pi_0 f_0(c) + \pi_1 f_1(c)$$

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 ($n_s = 2,\ldots,11$) and a fixed number of pathogens ($n_p = 2,\ldots,8$) 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_s and n_p. As a last step we calculated for each combination the mean and variance of the rank for gene g 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 ($prob > 0.9$ *and* $prob > 0.7$) in the top k ($k = 1,\ldots,50$) 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 ($k = 1,\ldots,100$) 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 ($k = 1,\ldots,826$) and computed the true positive rate (TPR = FP/(FP + TN)) and the false positive rate (FPR = FP/(FP + 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 und 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 51RT 0_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-132979 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 MIE-2009-SignRupVac, Fondation Le Roch and Fondation Jeantet. A.K. is a recipient of a scholarship from the Pasteur-Paris University International Doctoral Program/Institut Carnot 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 12–1) (to E.S. and N.B.). F.S. has been supported by SystemsX.ch, the Swiss initiative in systems biology, under IPhd grant No. 2009/025 (to N. 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 *Adenovirus* 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 Greber lab was funded by a grant from the SNSF (31003A_141222/1) and an iPhD fellowship to partly support A.Y. (SXPHIO_142001/1). We acknowledge support from Dr. A. Jurgeit in the development of the *Rhinovirus* infection assay. We thank Dr. Jean Pierre Gorvel (Centre d'immunolgie 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
[1]Focal Area Infection Biology, Biozentrum, University of Basel, Klingelberstrasse 70, CH-4056 Basel, Switzerland. [2]Seminar for Statistics, ETH Zurich, Zurich, Switzerland. [3]Institute of Molecular Life Sciences, University of Zurich, Zurich, Switzerland. [4]Department of Biosystems Science and Engineering, ETH Zurich, Zurich, Switzerland. [5]Swiss Institute of Bioinformatics, Basel, Switzerland. [6]Institut Pasteur, Unité des Interactions Bactéries Cellules; INSERM, U604; INRA, USC2020, Paris, France. [7]Institute for Tropical Health and Departamento de Microbiología y Parasitología, Universidad de Navarra, Pamplona, Spain. [8]Institute of Biochemistry, ETH Zurich, Zurich, Switzerland. [9]Department of Biology, Institute of Microbiology, ETH Zurich, Zurich, Switzerland. [10]Institute of Molecular Biology, Mainz, Germany. [11]Light Microscopy and Screening Center, ETH Zurich, Zurich, Switzerland. [12]Research IT, Biozentrum, University of Basel, Basel, Switzerland. [13]Institut Cochin, INSERM U1016, CNRS 8104, Université Paris Descartes, Paris, France. [14]SyBIT, SystemsX.ch, Zurich, Switzerland.

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

References

1. Conrad C, Gerlich DW: **Automated microscopy for high-content RNAi screening.** *J Cell Biol* 2010, **188**(4):453–461.
2. Mohr S, Bakal C, Perrimon N: **Genomic screening with RNAi: results and challenges.** *Annu Rev Biochem* 2010, **79**:37–64.
3. Mohr SE, Perrimon N: **RNAi screening: new approaches, understandings, and organisms.** *Wiley Interdiscip Rev RNA* 2012, **3**(2):145–158.
4. Simpson KJ, Davis GM, Boag PR: **Comparative high-throughput RNAi screening methodologies in C. elegans and mammalian cells.** *N Biotechnol* 2012, **29**(4):459–470.
5. Elbashir SM, Harborth J, Lendeckel W, Yalcin A, Weber K, Tuschl T: **Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells.** *Nature* 2001, **411**(6836):494–498.
6. Mercer J, Snijder B, Sacher R, Burkard C, Bleck CK, Stahlberg H, Pelkmans L, Helenius A: **RNAi screening reveals proteasome- and cullin3-dependent stages in vaccinia virus infection.** *Cell Rep* 2012, **2**(4):1036–1047.
7. Stertz S, Shaw ML: **Uncovering the global host cell requirements for influenza virus replication via RNAi screening.** *Microbes Infect* 2011, **13**(5):516–525.
8. Misselwitz B, Dilling S, Vonaesch P, Sacher R, Snijder B, Schlumberger M, Rout S, Stark M, von Mering C, Pelkmans L, Hardt WD: **RNAi screen of Salmonella invasion shows role of COPI in membrane targeting of cholesterol and Cdc42.** *Mol Syst Biol* 2011, **7**:474.
9. Snijder B, Sacher R, Ramo P, Liberali P, Mench K, Wolfrum N, Burleigh L, Scott CC, Verheije MH, Mercer J, Moese S, Heger T, Theusner K, Jurgeit A, Lamparter D, Balistreri G, Schelhaas M, De Haan CAM, Marjomäki V, Hyypiä T, Rottier PJM, Sodeik B, Marsh M, Gruenberg J, Amara A, Greber U, Helenius A, Pelkmans L: **Single-cell analysis of population context advances RNAi screening at multiple levels.** *Mol Syst Biol* 2012, **8**:579.
10. Schmidt EE, Pelz O, Buhlmann S, Kerr G, Horn T, Boutros M: **GenomeRNAi: a database for cell-based and in vivo RNAi phenotypes, 2013 update.** *Nucleic Acids Res* 2013, **41**(Database issue):D1021–D1026.
11. Pache L, Konig R, Chanda SK: **Identifying HIV-1 host cell factors by genome-scale RNAi screening.** *Methods* 2011, **53**(1):3–12.
12. Sigoillot FD, King RW: **Vigilance and validation: keys to success in RNAi screening.** *ACS Chem Biol* 2011, **6**(1):47–60.
13. Bushman FD, Malani N, Fernandes J, D'Orso I, Cagney G, Diamond TL, Zhou H, Hazuda DJ, Espeseth AS, König R, Bandyopadhyay S, Ideker T, Goff SP, Krogan NJ, Frankel AD, Young JA, Chanda SK: **Host cell factors in HIV replication: meta-analysis of genome-wide studies.** *PLoS Pathog* 2009, **5**(5):e1000437.
14. Heynen-Genel S, Pache L, Chanda SK, Rosen J: **Functional genomic and high-content screening for target discovery and deconvolution.** *Expert Opin Drug Discov* 2012, **7**(10):955–968.
15. 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.
16. Buehler E, Chen YC, Martin S: **C911: A bench-level control for sequence specific siRNA off-target effects.** *PLoS One* 2012, **7**(12):e51942.
17. Bassik MC, Kampmann M, Lebbink RJ, Wang S, Hein MY, Poser I, Weibezahn J, Horlbeck MA, Chen S, Mann M, Hyman AA, Leproust EM, McManus MT, Weissman JS: **A systematic mammalian genetic interaction map reveals pathways underlying ricin susceptibility.** *Cell* 2013, **152**(4):909–922.
18. Kittler R, Surendranath V, Heninger AK, Slabicki M, Theis M, Putz G, Franke K, Caldarelli A, Grabner H, Kozak K, Wagner J, Rees E, Korn B, Frenzel C, Sachse C, Sönnichsen B, Guo J, Schelter 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):337–344.
19. Collinet C, Stoter M, Bradshaw CR, Samusik N, Rink JC, Kenski D, Habermann B, Buchholz F, Henschel R, Mueller MS, Nagel WE, Fava E, Kalaidzidis Y, Zerial M: **Systems survey of endocytosis by multiparametric image analysis.** *Nature* 2010, **464**(7286):243–249.
20. Marine S, Bahl A, Ferrer M, Buehler E: **Common seed analysis to identify off-target effects in siRNA screens.** *J Biomol Screen* 2012, **17**(3):370–378.
21. Buehler E, Khan AA, Marine S, Rajaram M, Bahl A, Burchard J, Ferrer M: **siRNA off-target effects in genome-wide screens identify signaling pathway members.** *Sci Rep* 2012, **2**:428.
22. Sigoillot FD, Lyman S, Huckins JF, Adamson B, Chung E, Quattrochi B, King RW: **A bioinformatics method identifies prominent off-targeted transcripts in RNAi screens.** *Nat Methods* 2012, **9**(4):363–366.
23. Dehio C, Meyer M, Berger J, Schwarz H, Lanz C: **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* 1997, **110**(Pt 18):2141–2154.
24. Cossart P, Sansonetti PJ: **Bacterial invasion: the paradigms of enteroinvasive pathogens.** *Science* 2004, **304**(5668):242–248.
25. Meier O, Boucke K, Hammer SV, Keller S, Stidwill RP, Hemmi S, Greber UF: **Adenovirus triggers macropinocytosis and endosomal leakage together with its clathrin-mediated uptake.** *J Cell Biol* 2002, **158**(6):1119–1131.
26. Mercer J, Helenius A: **Vaccinia virus uses macropinocytosis and apoptotic mimicry to enter host cells.** *Science* 2008, **320**(5875):531–535.
27. Malo N, Hanley JA, Cerquozzi S, Pelletier J, Nadon R: **Statistical practice in high-throughput screening data analysis.** *Nat Biotechnol* 2006, **24**(2):167–175.
28. Carpenter AE, Jones TR, Lamprecht MR, Clarke C, Kang IH, Friman O, Guertin DA, Chang JH, Lindquist RA, Moffat J, Golland P, Sabatini DM: **Cell profiler: image analysis software for identifying and quantifying cell phenotypes.** *Genome Biol* 2006, **7**(10):R100.
29. Birmingham A, Selfors LM, Forster T, Wrobel D, Kennedy CJ, Shanks E, Santoyo-Lopez J, Dunican DJ, Long A, Kelleher D, Smith Q, Beijersbergen 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, Merdes G, Misselwitz B, Hardt WD, Beerenwinkel N: **Stability of gene rankings from RNAi screens.** *Bioinformatics* 2012, **28**(12):1612–1618.
31. Snijder B, Pelkmans L: **Origins of regulated cell-to-cell variability.** *Nat Rev Mol Cell Biol* 2011, **12**(2):119–125.
32. 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.
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 S4 Classes.* 2013. http://cran.r-project.org/package=lme4.
35. Efron B, Turnbull BB, Narasimhan B: *locfdr: Computes Local False Discovery Rates.* 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-Cerda J, Kuhbacher A, Cossart P: **Entry of listeria monocytogenes in Mammalian Epithelial Cells: an updated view.** *Cold Spring Harbor Perspect Med* 2012, **2:**11.
38. Tattoli I, Philpott 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. Konig R, Chiang CY, Tu BP, Yan SF, DeJesus PD, Romero A, Bergauer T, Orth A, Krueger U, Zhou Y, Chanda SK: **A probability-based approach for the analysis of large-scale RNAi screens.** *Nat Methods* 2007, **4**(10):847–849.
40. Zhang XD, Ferrer M, Espeseth AS, Marine SD, Stec EM, Crackower MA, Holder DJ, Heyse JF, Strulovici 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, Muller J, Bork P, Jensen LJ, von Mering C: **The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored.** *Nucleic Acids Res* 2011, **39**(Database issue):D561–D568.
44. Elfaki MG, Al-Hokail AA: **Transforming growth factor beta production correlates with depressed lymphocytes function in humans with chronic brucellosis.** *Microbes Infect* 2009, **11**(14–15):1089–1096.
45. Shen Y, Naujokas K, Park M, Ireton K: **InlB-dependent internalization of Listeria is mediated by the Met receptor tyrosine kinase.** *Cell* 2000, **103**(3):501–510.
46. Zheng 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 P: **RNAi screen of endoplasmic reticulum-associated host factors reveals a role for IRE1alpha in supporting Brucella replication.** *PLoS Pathog* 2008, **4**(7):e1000110.
48. Cain RJ, Vanhaesebroeck B, Ridley AJ: **The PI3K p110alpha isoform regulates endothelial adherens junctions via Pyk2 and Rac1.** *J Cell Biol* 2010, **188**(6):863–876.
49. Kakumoto 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. Grassme H, Riehle A, Wilker B, Gulbins E: **Rhinoviruses 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, Hershey 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, Boucke K, Bianchi S, Engel DA, Greber UF: **A 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, 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.
54. Dehio M, Knorre 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. Celli J, Salcedo SP, Gorvel JP: **Brucella coopts the small GTPase Sar1 for intracellular replication.** *Proc Natl Acad Sci U S A* 2005, **102**(5):1673–1678.
56. 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.** *Virol J* 2010, **7:**264.
57. Jurgeit A, McDowell R, Moese S, Meldrum E, Schwendener R, Greber UF: **Niclosamide is a proton carrier and targets acidic endosomes with broad antiviral effects.** *PLoS Pathog* 2012, **10:**8.
58. Mosser AG, Brockman-Schneider R, Amineva S, Burchell L, Sedgwick JB, Busse WW, Gern 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: Rämö *et al.*: Simultaneous analysis of large-scale RNAi screens for pathogen entry. *BMC Genomics* 2014 **15**:1162.

3.5 RESEARCH ARTICLE V (IN PREPARATION)

Genome-wide siRNA screen in HeLa cells reveals role of endosome to Golgi transport during *Brucella* infection

A. Casanova[1*], S. H. Low[1*], R. Conde-Alvarez[3], M. Emmenlauer[1], P. Rämo[1], H.B. Tekaya[1], S. Muntwiler[1], S. Salcedo[2], J.P. Gorvel[2] and C. Dehio[1]

* These authors contributed equally to this work

Manuscript in preparation

3.5.1 Statement of own contribution

siRNA screens and MOI titrations were performed together with Dr. Shyan H. Low, except for the Qiagen druggable library, which was conducted by Dr. Shyan H. Low and Dr. Raquel Conde-Alvarez. The entry assay was developed and conducted by me. Image analysis was performed by Mario Emmenlauer and data normalization by Dr. Pauli Rämö. Vps35 rescue and Retro-2 inhibitor studies were carried out by Dr. Shyan H. Low. Figures are produced by me and Dr. Shyan H. Low, except the representation of the decision tree infection scoring in Figure 1 which is provided by Mario Emmenlauer. The manuscript is written by me and Dr. Shyan H. Low.

3.5.2 Manuscript

Genome-wide siRNA screen in HeLa cells reveals role of endosome to Golgi transport during *Brucella* infection

A. Casanova[1*], S. H. Low[1*], R. Conde-Alvarez[3], M. Emmenlauer[1], P. Rämo[1], H.B. Tekaya[1], S. Muntwiler[1], S. Salcedo[2], J.P. Gorvel[2] and C. Dehio[1]

[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 a 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 infected as incidental hosts, suffer from 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, some bacteria manage to escape degradation. 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 employed to study early stages of the infection. We showed that TGF-β and FGF signaling pathways are involved in bacterial entry into non-phagocytic cells. Furthermore, we identified a novel host factor, Vps35, to be involved in a post-entry process during *Brucella* infection. Vps35 is a component of the retromer complex that mediates transport between endosomes and the trans-Golgi network (TGN). Furthermore, we showed that Retro-2, an inhibitor which blocks retrograde transport of toxins for endosomes to the Golgi apparatus, significantly reduced *Brucella* infection in a dose-dependent manner. This further confirms the importance of endosome to Golgi transport 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 only rarely 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 vironectin (*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 the lipid rafts component flotillin-1 (*7-10*). Next, the 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 endoplasmic reticulum (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. 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 and remain as one of the important unresolved questions of *Brucella* intracellular trafficking. Furthermore, host factors that are needed for the maintenance of the BCV in its replicative niche are still largely unexplored.

The retromer complex is an essential component of the endosomal protein sorting machinery. It is involved in vesicular transport from endosomes to the trans-Golgi network (TGN) and recycling of cargo to the plasma membrane (*20-22*). In mammalian cells, the retromer complex is composed of two functional subcomplexes: 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 (*23*) that interact with the CST. Vps35 is the core component of the CST with a direct role in cargo binding (*24, 25*) while Vps26 and Vps29 independently associates at either end of the complex. There are two different types of retromer complexes depending on the components of sorting nexins associated to it. The SNX-BAR-retromer, which is comprised of Snx1, Snx2, Snx5, and Snx6 associated with the CST, is able to induce and stabilize the formation of membrane tubules (*26-28*). This is due to their membrane curvature sensing BAR (Bin/amphiphysin/Rv) domain and phosphatidylinositol 3-phosphate [PtdIns(3)P] or phosphatidylinositol 3,5-bisphosphate [PtdIns(3,5)P2] binding PX (phox homology) domain (*29, 30*). The SNX3-retromer however only consists of the complex of SNX3 and CST (*28, 31*). SNX3 does not contain a BAR domain but has a PX domain that binds with high affinity to PtdIns(3)P (*32*).

While the retromer serves critical functions in cellular homeostasis, different toxins and pathogens have also been identified to hijack or manipulate retrograde endosome-Golgi transport to their own benefit. Shiga and cholera toxins require retrograde trafficking to reach the Golgi and further the ER (*33*) while human papillomavirus (HPV) stably interacts with the retromer during cell entry, which is later required for reaching a Golgi-like compartment (*34*). Also, *Coxiella burnetii* which replicates in a lysosome-like vacuole, requires the retromer complex for intracellular proliferation (*35*). In other cases, bacteria intercept with the endosome-Golgi pathway. For instance, *Salmonella* inhibits retrograde trafficking of cation-independent mannose 6-phosphate receptor (CIMPR) and lysosome function (*36*) while *Legionella* effector RidL promotes intracellular replication by binding directly to Vps29 and PtdIns(3)P to inhibit the function of the retromer (*37*).

To understand on 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 including components of the endosome-Golgi transport pathway. Knockdown of components of the retromer complex, especially the CST, led to a decrease in *Brucella* infection. Furthermore, our entry assay suggests that Vps35 is involved in a post-entry step of the infection cycle. Inhibitor studies using Retro-2 which blocks retrograde transport of toxins between endosomes and Golgi (*38*) showed a significant decrease in *Brucella* infection. Therefore, our studies indicate that *Brucella* requires functional retrograde transport between the endocytic pathway and the Golgi apparatus during its infection.

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 performed a high-throughput microscopy-based RNAi assay in HeLa cells as described previously (*research article I*). As seen in Supplementary Figure 1, the conditions that we used in our screens did not reach saturation and intracellular trafficking of *Brucella* under these conditions is consistent with previous studies (*39*). Wild-type *B. abortus* as well as a *virB9* mutant acquire Lamp1 markers at 6 hours post

infection (hpi). At 24 hpi, the *virB9* mutant remains in a Lamp1 containing endo-lysosomal compartment while wild-type *B. abortus* is excluded from this compartment (Supplementary Figure 2i and 2ii).

The experimental workflow used in our siRNA screens is summarized in Figure 1A. Reverse siRNA transfection was performed for 72 h in HeLa cells after which cells were infected with GFP-expressing *B. abortus* 2308. After 4 h of infection, extracellular bacteria were inactivated by gentamicin-containing medium. 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 the formation of a micro-colony (Figure 1B). Cells were fixed at 44 hpi and stained with DAPI and phalloidin-547 for nuclei and F-actin, respectively. Fluorescence images were acquired with automated microscopes and image analysis was performed with CellProfiler (*40*). This allows for an automated detection of objects such as the nucleus, perinucleus (8 pixels or 5.16 μm wide zone surrounding the nucleus), or cell body (Figure 1A and B), and the extraction of features for example pathogen intensity (GFP intensity) from these objects. Decision tree based infection scoring was performed with the extracted features (Figure 1B) to calculate single cell infection scores which were used to determine a well-based infection rate. In short, if the mean GFP intensity in at least one 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 which contain a high number of bacteria, indicative of intracellular proliferation, will be detected as infected. Finally, Z scoring normalization was performed to account for plate-to-plate variations.

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 siRNA screens were performed in HeLa cells. Three and one replicates of a Dharmacon and a Qiagen library, respectively, were screened. The Dharmacon library contains a single pool of four siRNAs targeting each gene while the Qiagen library comprises four individual siRNAs for each target.

Each plate contained control siRNAs to monitor the quality of the screens. 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 3B). Positive control siRNAs targeting known host factors of *Brucella* infection such as Rac1, Cdc42, and ATP6V1A (*6, 12*), showed an expected reduction of infection compared to mock (transfection reagent only) and scrambled siRNA without specific host targets (Supplementary Figure 3A). Furthermore, we compared the results from the independent replicates of the Dharmacon library. 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 the Dharmacon pooled library. Altogether, this shows that with our experimental workflow we were 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 technology, we performed statistical analysis of the primary screening data with the Redundant siRNA Analysis (RSA) algorithm (*41*). This reduces the number of false positives caused by off-target effects of single siRNAs and favors genes with a reproducible phenotype from different siRNAs, indicating that the effect observed on *Brucella* infection is due to the depletion of the intended gene.

Using datasets obtained from screening with Dharmacon and Qiagen libraries as input for RSA analysis, we were able to rank genes according to statistical significance (P value). Figure 1C summarizes the general workflow that was used for analyzing the screening data and selecting genes for further validation. Some of the genes that were in the top ranks of our RSA analysis were directly selected for validation. To further prioritize, we performed gene ontology (GO) enrichment studies of the top hits using the DAVID functional annotation database (*42*). As seen in Figure 2A, genes that appear in the top 200 RSA ranks for reducing *Brucella* infection upon knockdown show 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 they did not show the strongest phenotype 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* infection

Secondary screens for all selected targets were performed with up to three siRNAs each from the Ambion Silencer and Ambion Silencer Select unpooled libraries as well as one esiRNA from the pooled Sigma MISSION library (replicate correlations are shown in Supplementary Figure 4ii). This strategy to confirm selected genes with many additional siRNAs was proven successful in previous kinome-wide screens (*research article IV*), to reduce the false discovery rate caused by off-target effects. Finally, RSA analysis was performed with genes reducing or increasing infection, using the combination of all data from both primary and secondary screens (Supplementary Table 1). Figure 3 represents the high confidence STRING database protein-protein interaction network of top ranking genes from the RSA analysis that reduce and increase *Brucella* infection upon knockdown (*43*). 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 pathway, ER-Golgi bidirectional transport, ubiquitin conjugation, nucleotide metabolism, cell cycle, and clathrin coated pits were found in our top ranking gene lists suggesting a role of these signaling pathways during *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 which is regulated by the identified genes during *Brucella* infection, we performed an assay to study early steps of infection in a high-throughput format (described in *research article I*). The entry assay is based on the infection of HeLa cells with *B. abortus* that express GFP under a tetracycline inducible system and dsRed from a constitutive promoter. Since induction of GFP expression is performed simultaneous to gentamicin addition to the medium, only intracellular bacteria are able to express GFP while extracellular bacteria are inactivated by gentamicin in the medium. As shown in Figure 4A, all *Brucella* express dsRed while only

intracellular bacteria induce GFP expression. For the entry assay, bacteria were allowed to enter cells for 4 h, after which GFP expression was induced for another 4 h. 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 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 between the results of the entry assay and the 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 a role of these pathways 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*. Interestingly, Vps35 showed no effect on *Brucella* entry upon knockdown even though there was a significant reduction in the number of cells that showed the formation of a micro-colony two days after infection (endpoint assay). This suggests that Vps35, a key component of the retromer complex is involved in a post-entry step during *Brucella* infection.

Endosome to Golgi transport is required for *Brucella* infection

Besides Vps35, Vps26a knockdown also strongly reduced *Brucella* infection in the genome-wide screen, comparable to the effect of the positive controls ArpC3 and Rab7a (Figure 5A). siRNA depletion of Vps29 and Vps26b, the paralogue of Vps26a, only led to a mild reduction in *Brucella* infection. In contrast, depletion of Snx1 and Snx5, two of the four sorting nexins associated with the Snx-BAR-retromer enhanced infection while the other two components, SNX2 and SNX6 did not show an effect on *Brucella* infection upon knockdown. SNX3, component of the SNX3-retromer showed a mild reduction in infection while USP6NL, a Rab GTPase activating protein (GAP) for Rab43 shown to be involved in Shiga toxin endosome to TGN trafficking (*44*) increased *Brucella* infection upon knockdown.

To confirm the specificity of Vps35 knockdown and its effect on *Brucella* infection, a complementation experiment with a Vps35 cDNA insensitive to the shRNA was performed. While shRNA knockdown of Vps35 inhibited *Brucella* infection, co-expression of the shRNA-

insensitive cDNA of Vps35 could rescue the phenotype (Figure 5B), confirming that depletion of Vps35 indeed negatively affects *Brucella* infection.

To further investigate the role of retrograde transport from endosomes to Golgi in *Brucella* infection, experiments were performed with Retro-2. This is a specific inhibitor of retrograde trafficking from endosomes to the TGN of ricin toxin, Shiga-like toxin, cholera toxin B subunit, and human papillomaviruses (*34, 38*). The effect of Retro-2 on this transport pathway is selective, with no effect on most retrograde cargoes except syntaxin 6, syntaxin 5, and syntaxin 16 and no perturbation of the morphology of compartments or other trafficking steps (*38*). Furthermore, this compound has been shown to protect mice from lethal nasal exposure to ricin, illustrating the potential *in vivo* applications of this drug (*38*). For our studies, Retro-2 was added to HeLa cells either together with GFP-expressing *B. abortus* (Figure 6A) or 4 h after infection when cells were washed with gentamicin-containing medium to inactivate extracellular bacteria (Figure 6B). The inhibitor was then kept throughout the experiment. *Brucella* infection decreased in a dose-dependent manner upon increased doses of the inhibitor. This supports a role of retrograde trafficking from endosomes to TGN in *Brucella* infection. Inhibition of infection was also seen when the drug was added after bacteria have entered cells, suggesting that the endosome-Golgi pathway is important for a post-entry process during *Brucella* infection.

Discussion

Studies that have been performed thus far to understand *Brucella* interaction with the host were mainly hypothesis driven, include a small-scale RNAi screen in Drosophila S2 cell, or proteomics studies to identify host components of the BCV (*7, 11, 15, 16, 18, 19*). To identify on a systems level host factors involved in *Brucella* infection, we performed a genome-wide RNAi screen in HeLa cells. We could identify novel host factors covering different signaling pathways to have a role in *Brucella* infection and using a high-throughput entry assay, host factors could be further separated according to their impact on *Brucella* entry. We identified Vps35, a component of the retromer complex to be involved in a post-entry process during *Brucella* infection. Importantly, the role of endosome-Golgi transport during a post-entry step in *Brucella* infection could be independently validated using a drug that has been shown to inhibit transport of toxins via this pathway.

Vps35 is a component of the retromer complex that regulates endosome to Golgi transport (*23*). Furthermore, other components of the retromer complex including Vps26a and to a lesser extent Vps29 and Vps26b also reduced *Brucella* infection upon knockdown in our genome-wide siRNA screen. It has been reported that Vps26a and Vps26b have preferences towards different cargoes (*45*). As we generally did not observe a strong effect upon depletion of Vps26b compared to Vps26a, it is possible that mainly Vps26a is needed during *Brucella* infection. The varying effects on *Brucella* infection of the different Snxs suggest that they might have different roles during infection. Alternatively, the efficiency of different siRNA knockdowns might vary and should be validated. Combinatorial knockdown of Snxs will give a better understanding to whether there are redundancies in the function between these different Snx proteins.

Furthermore, inhibition of retrograde trafficking with the drug Retro-2 caused a defect in *Brucella* infection, even when the drug was added to cells when bacteria were already internalized which suggests that the endosome-Golgi retrograde trafficking pathway is required after bacteria have invaded the host cell. This is consistent with our entry assay results that showed that Vps35 is dispensable for *Brucella* entry in HeLa cells. 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. It is conceivable that *Brucella* uses a similar pathway as ricin or Shiga-like toxin. In support of this notion, USP6NL, a Rab GAP that is involved in Shiga toxin transport from endosomes to the trans-Golgi network by regulating Rab43 (*44*), led to an increase in *Brucella* infection upon knockdown. It is thus tempting to speculate that USP6NL regulation of Shiga toxin transport is needed by *Brucella* in a similar manner. Alternatively, the retromer complex could be involved in the establishment or maintenance of the replicative niche potentially providing host factors which follow retrograde trafficking.

Besides the retromer complex, we could identify several other pathways which are involved in *Brucella* infection. Among the most prominent clusters are 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 type IV secretion system (*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*).

The entry assay showed that depletion of components of other pathways than endosome-Golgi transport affect *Brucella* entry into host cells. Genes associated with actin-remodeling (Rac1, Cdc42, CYFIP1, NCKAP1, ACTR3), TGF-β signaling (TGFBR1, TGFBR2, SMAD4), endocytic pathway (RAB7A), and Golgi to ER transport (COPG) showed reduced *Brucella* entry and a subsequent decrease in the formation of an intracellular micro-colony upon knockdown.

Studies with drug inhibitors have shown the importance of Rac1 and Cdc42 in *Brucella* uptake into eukaryotic cells (*6*) and the role of these factors was confirmed in our RNAi screens. In addition, we identified numerous 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 decreases infection upon knockdown. Upstream or downstream components of Rac1 such as the WAVE complex (NCKAP1, CYFIP1, ABI1), 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 were found in the genome-wide screen and all tested factors were 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 extension of host factors in this signaling process.

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 negatively affect *Brucella* replication (*16*). In this same study, the authors reported that *Brucella* replication was affected by prolonged treatment with Brefeldin A. This drug causes a redistribution of the Golgi to the ER, suggesting that Golgi to ER trafficking is important for *Brucella* replication. 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 of 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 COPG1 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 (*46*). It is conceivable that *Brucella* entry follows a similar route as described for *Salmonella*, requiring COP components for membrane ruffling (*46*) and should be investigated with further studies. We also identified a component of the COPII complex (Sec13) to reduce *Brucella* infection upon depletion. This is consistent with previous studies showing the importance of the COPII complex and ER exit sites for the interaction of *Brucella* with the ER and subsequent replication (*15, 16*). Taken together, these findings show the importance of 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 the composition of the plasma membrane.

Members of the TGF-β and FGF signaling pathways were also found to 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 heterotetrameric receptor complex, TGFBR1 and TGFBR2, showed an effect on *Brucella* infection upon individual knockdown strengthens the notion that TGF-β signaling is important for *Brucella* infection. It has previously been reported that patients with brucellosis show higher levels of TGF-β1 in their sera that is correlated with depressed function of T cell responses (*47*). B cells were also shown to produce TGF-β at early stages of infection with *Brucella* in mice (*48*). Therefore, this suggests a role of this pathway during *Brucella* infection in terms of immunosuppression of the host. However, since our RNAi screen was performed in HeLa cells, it is conceivable 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. Components of the FGF signaling pathway (FGFR1, FGF2 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 (HSP-G) dependent manner. The pathogen additionally stimulates production of FGF2 that enhances subsequent rounds of infection (*49*). 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 trophoblast cell lines 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, and EPN1) caused an increase in *Brucella* infection in our screen. This 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, chloropromazine (*50*). 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 end point 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 were able to identify novel signaling pathways as well as novel host factors of pathways that have been previously associated with *Brucella* infection. Altogether, many additional genes were identified and will provide a rich resource for future discoveries. Most of the identified hits that were tested in the entry assay were found to be involved in *Brucella* entry, including the TGF-β signaling pathway or factors involved in actin remodeling. Importantly, we identified retrograde transport between the endocytic pathway and the Golgi to be involved in a post-entry process during *Brucella* infection. This trafficking route has not been described for *Brucella* infection before and provides the hypothesis that *Brucella* requires this pathway to arrive at or maintain its replicative niche.

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); Retro-2 (Calbiochem, 554715), Fugene HD (Promega, E2312), mouse monoclonal anti-Lamp1 [H4A3] antibody (Abcam, ab25630), Anhydrotetracycline hydrochloride (Sigma, 37919)

Cloning of pAC042.08 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 *B. abortus* 2308 by conjugation.

siRNA reverse transfection

Reverse siRNA transfection was performed as describe in *research article I* with minor modifications. 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 µl (QU: 15 ul) 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 µl (QU: 30 ul) DMEM/16% FCS, resulting in a final concentration of 10% FCS. Plates were incubated at 37°C and 5% CO_2 for 72 h prior to infection.

Infection

Infections were performed as describe in *research article I.* In short, *B. abortus* 2308 pJC43 (*aphT::GFP*) (*15*) was grown in 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% 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% CO_2, extracellular bacteria were killed by exchanging the infection medium by 50 µl DMEM/10% FCS 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 RT.

For the entry assay, *B. 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 and kept throughout the experiment.

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 *B. abortus* 2308 or during gentamicin wash at 4 hpi and cells were maintained at 37°C with 5% CO_2. Retro-2 was kept throughout the experiment.

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 (*51*). HeLa cells were seeded in the morning of the day of transfection in a 6 well plate with 125,000 cells / well. In the evening, , 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 1.5 ml 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.

Staining

Cells were washed twice with PBS and permeabilized with 0.1% Triton-x-100 for 10 min. After washing twice with PBS, 20 µl of staining solution that contains DAPI (1 µg/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 µl 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 (*40*)**.** 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 *B. 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 (*52*). 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, we chose non-control based data normalization methods for primary and secondary screens. 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. 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 (*41*). 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. RSA was run using the R-package ↓"RSA↓" release 1.3 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.

Acknowledgements

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

References

1. G. Pappas, P. Papadimitriou, N. Akritidis, L. Christou, E. V. Tsianos, The new global map of human brucellosis. *The Lancet infectious diseases* **6**, 91 (Feb, 2006).
2. V. L. Atluri, M. N. Xavier, M. F. de Jong, A. B. den Hartigh, R. M. Tsolis, Interactions of the human pathogenic Brucella species with their hosts. *Annu Rev Microbiol* **65**, 523 (2011).
3. J. Ariza *et al.*, Perspectives for the treatment of brucellosis in the 21st century: the Ioannina recommendations. *PLoS medicine* **4**, e317 (Dec, 2007).
4. E. I. Castaneda-Roldan *et al.*, Adherence of Brucella to human epithelial cells and macrophages is mediated by sialic acid residues. *Cell Microbiol* **6**, 435 (May, 2004).
5. E. I. Castaneda-Roldan *et al.*, Characterization of SP41, a surface protein of Brucella associated with adherence and invasion of host epithelial cells. *Cell Microbiol* **8**, 1877 (Dec, 2006).
6. C. Guzman-Verri *et al.*, GTPases of the Rho subfamily are required for Brucella abortus internalization in nonprofessional phagocytes: direct activation of Cdc42. *The Journal of biological chemistry* **276**, 44435 (Nov 30, 2001).
7. J. Celli *et al.*, Brucella evades macrophage killing via VirB-dependent sustained interactions with the endoplasmic reticulum. *The Journal of experimental medicine* **198**, 545 (Aug 18, 2003).
8. B. H. Bellaire, R. M. Roop, 2nd, J. A. Cardelli, Opsonized virulent Brucella abortus replicates within nonacidic, endoplasmic reticulum-negative, LAMP-1-positive phagosomes in human monocytes. *Infect Immun* **73**, 3702 (Jun, 2005).
9. J. Pizarro-Cerda *et al.*, Brucella abortus transits through the autophagic pathway and replicates in the endoplasmic reticulum of nonprofessional phagocytes. *Infection and immunity* **66**, 5711 (Dec, 1998).
10. B. Arellano-Reynoso *et al.*, Cyclic beta-1,2-glucan is a Brucella virulence factor required for intracellular survival. *Nat Immunol* **6**, 618 (Jun, 2005).
11. T. Starr, T. W. Ng, T. D. Wehrly, L. A. Knodler, J. Celli, Brucella intracellular replication requires trafficking through the late endosomal/lysosomal compartment. *Traffic* **9**, 678 (May, 2008).
12. F. Porte, J. P. Liautard, S. Kohler, Early acidification of phagosomes containing Brucella suis is essential for intracellular survival in murine macrophages. *Infection and immunity* **67**, 4041 (Aug, 1999).
13. M. L. Boschiroli *et al.*, The Brucella suis virB operon is induced intracellularly in macrophages. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 1544 (Feb 5, 2002).
14. D. J. Comerci, M. J. Martinez-Lorenzo, R. Sieira, J. P. Gorvel, R. A. Ugalde, Essential role of the VirB machinery in the maturation of the Brucella abortus-containing vacuole. *Cell Microbiol* **3**, 159 (Mar, 2001).
15. J. Celli, S. P. Salcedo, J. P. Gorvel, Brucella coopts the small GTPase Sar1 for intracellular replication. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 1673 (Feb 1, 2005).
16. E. Fugier *et al.*, The glyceraldehyde-3-phosphate dehydrogenase and the small GTPase Rab 2 are crucial for Brucella replication. *PLoS Pathog* **5**, e1000487 (Jun, 2009).

17. M. de Barsy *et al.*, Identification of a Brucella spp. secreted effector specifically interacting with human small GTPase Rab2. *Cellular microbiology* **13**, 1044 (Jul, 2011).
18. Q. M. Qin *et al.*, RNAi screen of endoplasmic reticulum-associated host factors reveals a role for IRE1alpha in supporting Brucella replication. *PLoS pathogens* **4**, e1000110 (Jul, 2008).
19. T. Starr *et al.*, Selective subversion of autophagy complexes facilitates completion of the Brucella intracellular cycle. *Cell Host Microbe* **11**, 33 (Jan 19, 2012).
20. M. N. Seaman, E. G. Marcusson, J. L. Cereghino, S. D. Emr, Endosome to Golgi retrieval of the vacuolar protein sorting receptor, Vps10p, requires the function of the VPS29, VPS30, and VPS35 gene products. *J Cell Biol* **137**, 79 (Apr 7, 1997).
21. M. N. Seaman, J. M. McCaffery, S. D. Emr, A membrane coat complex essential for endosome-to-Golgi retrograde transport in yeast. *J Cell Biol* **142**, 665 (Aug 10, 1998).
22. P. Temkin *et al.*, SNX27 mediates retromer tubule entry and endosome-to-plasma membrane trafficking of signalling receptors. *Nat Cell Biol* **13**, 715 (Jun, 2011).
23. T. Wassmer *et al.*, The retromer coat complex coordinates endosomal sorting and dynein-mediated transport, with carrier recognition by the trans-Golgi network. *Developmental cell* **17**, 110 (Jul, 2009).
24. S. F. Nothwehr, S. A. Ha, P. Bruinsma, Sorting of yeast membrane proteins into an endosome-to-Golgi pathway involves direct interaction of their cytosolic domains with Vps35p. *J Cell Biol* **151**, 297 (Oct 16, 2000).
25. C. N. Arighi, L. M. Hartnell, R. C. Aguilar, C. R. Haft, J. S. Bonifacino, Role of the mammalian retromer in sorting of the cation-independent mannose 6-phosphate receptor. *J Cell Biol* **165**, 123 (Apr, 2004).
26. J. Carlton *et al.*, Sorting nexin-1 mediates tubular endosome-to-TGN transport through coincidence sensing of high- curvature membranes and 3-phosphoinositides. *Curr Biol* **14**, 1791 (Oct 26, 2004).
27. J. G. Carlton *et al.*, Sorting nexin-2 is associated with tubular elements of the early endosome, but is not essential for retromer-mediated endosome-to-TGN transport. *J Cell Sci* **118**, 4527 (Oct 1, 2005).
28. P. J. Cullen, H. C. Korswagen, Sorting nexins provide diversity for retromer-dependent trafficking events. *Nat Cell Biol* **14**, 29 (Jan, 2012).
29. B. J. Peter *et al.*, BAR domains as sensors of membrane curvature: the amphiphysin BAR structure. *Science* **303**, 495 (Jan 23, 2004).
30. G. E. Cozier *et al.*, 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. *J Biol Chem* **277**, 48730 (Dec 13, 2002).
31. M. Harterink *et al.*, A SNX3-dependent retromer pathway mediates retrograde transport of the Wnt sorting receptor Wntless and is required for Wnt secretion. *Nat Cell Biol* **13**, 914 (Aug, 2011).
32. J. W. Yu, M. A. Lemmon, All phox homology (PX) domains from Saccharomyces cerevisiae specifically recognize phosphatidylinositol 3-phosphate. *J Biol Chem* **276**, 44179 (Nov 23, 2001).
33. K. Sandvig, T. Skotland, B. van Deurs, T. I. Klokk, Retrograde transport of protein toxins through the Golgi apparatus. *Histochemistry and cell biology* **140**, 317 (Sep, 2013).
34. A. Lipovsky *et al.*, Genome-wide siRNA screen identifies the retromer as a cellular entry factor for human papillomavirus. *Proc Natl Acad Sci U S A* **110**, 7452 (Apr 30, 2013).

35. J. A. McDonough *et al.*, Host pathways important for Coxiella burnetii infection revealed by genome-wide RNA interference screening. *mBio* **4**, e00606 (2013).
36. K. McGourty *et al.*, Salmonella inhibits retrograde trafficking of mannose-6-phosphate receptors and lysosome function. *Science* **338**, 963 (Nov 16, 2012).
37. I. Finsel *et al.*, The Legionella effector RidL inhibits retrograde trafficking to promote intracellular replication. *Cell Host Microbe* **14**, 38 (Jul 17, 2013).
38. B. Stechmann *et al.*, Inhibition of retrograde transport protects mice from lethal ricin challenge. *Cell* **141**, 231 (Apr 16, 2010).
39. J. Celli *et al.*, Brucella evades macrophage killing via VirB-dependent sustained interactions with the endoplasmic reticulum. *J Exp Med* **198**, 545 (Aug 18, 2003).
40. A. E. Carpenter *et al.*, CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome biology* **7**, R100 (2006).
41. R. Konig *et al.*, A probability-based approach for the analysis of large-scale RNAi screens. *Nature methods* **4**, 847 (Oct, 2007).
42. W. Huang da *et al.*, DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. *Nucleic Acids Res* **35**, W169 (Jul, 2007).
43. A. Franceschini *et al.*, STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res* **41**, D808 (Jan, 2013).
44. E. Fuchs *et al.*, Specific Rab GTPase-activating proteins define the Shiga toxin and epidermal growth factor uptake pathways. *J Cell Biol* **177**, 1133 (Jun 18, 2007).
45. A. Bugarcic *et al.*, Vps26A and Vps26B subunits define distinct retromer complexes. *Traffic* **12**, 1759 (Dec, 2011).
46. B. Misselwitz *et al.*, RNAi screen of Salmonella invasion shows role of COPI in membrane targeting of cholesterol and Cdc42. *Mol Syst Biol* **7**, 474 (Mar 15, 2011).
47. M. G. Elfaki, A. A. Al-Hokail, Transforming growth factor beta production correlates with depressed lymphocytes function in humans with chronic brucellosis. *Microbes Infect* **11**, 1089 (Dec, 2009).
48. R. Goenka, M. A. Parent, P. H. Elzer, C. L. Baldwin, B cell-deficient mice display markedly enhanced resistance to the intracellular bacterium Brucella abortus. *The Journal of infectious diseases* **203**, 1136 (Apr 15, 2011).
49. J. H. Kim, S. Jiang, C. A. Elwell, J. N. Engel, Chlamydia trachomatis co-opts the FGF2 signaling pathway to enhance infection. *PLoS Pathog* **7**, e1002285 (Oct, 2011).
50. J. J. Lee *et al.*, Interplay between clathrin and Rab5 controls the early phagocytic trafficking and intracellular survival of Brucella abortus within HeLa cells. *J Biol Chem* **288**, 28049 (Sep 27, 2013).
51. T. T. Liu, T. S. Gomez, B. K. Sackey, D. D. Billadeau, C. G. Burd, Rab GTPase regulation of retromer-mediated cargo export during endosome maturation. *Molecular biology of the cell* **23**, 2505 (Jul, 2012).
52. A. Birmingham *et al.*, Statistical methods for analysis of high-throughput RNA interference screens. *Nature methods* **6**, 569 (Aug, 2009).

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 image analysis, infection scoring, and data normalization. B) Image on the right represents HeLa cells infected with GFP-expressing *B. 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 –LogP values of enrichment terms for top 200 RSA lists of A) genes that reduce *Brucella* infection upon siRNA knockdown or B) genes that increase *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 5 components of our list and with a P value lower than 0.05 are shown. A higher –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 a Z scored cell number < -1 were removed before RSA analysis. To illustrate the interaction network within our hit lists, RSA top 350 genes that reduce *Brucella* infection and RSA top 350 genes that increase *Brucella* infection were added to the STRING

database. The edges between genes indicate high-confidence (>0.95) STRING database interactions and only genes that contain at least one interacting partner are shown. Genes that reduce *Brucella* infection upon knockdown are surrounded by a blue outline, while a red outline indicates genes that increase 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 *B. 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 3 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).

Figure 5. Retromer components are involved in *Brucella* infection

A) Effect of siRNA knockdown of retromer components on *Brucella* infection. Each dot represents Z scored infection index averaged over all independent replicates with a certain siRNA reagent. All siRNAs tested during primary and validation screens are shown. Horizontal line indicates median over all data points per gene. B) Rescue experiment of Vps35 depletion. HeLa cells were transfected with the shRNA/rescue constructs and expressed for 2.5 days before infection with GFP-expressing *B. abortus* for 44 h. Data represents normalized infection in reference to empty vector, the mean ± STDEV of three independent experiments.

Figure 6. Retrograde endosome-Golgi 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. Data represents normalized infection in reference to non-treated HeLa cells, the mean ± STDEV of three independent experiments for Figure 6A (except for 20 µM dataset which has only two independent experiments) and at least four independent experiments for Figure 6B.

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 ± STDEV of three independent experiments.

Supplementary Figure 2. *B. abortus* Δ*virB9* mutant interacts with the endo-lysosomal compartment at 24 hpi, while most *B. abortus* avoids this compartment, with low or high MOI of bacteria

Images represent infection of HeLa cells with i) GFP expressing *B. 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 representative slice around the middle of a stack is shown. Scale bar represents 10 µm.

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 results 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 *B. abortus* and stained with DAPI (blue). Scale bar represents 50 µm.

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. RSA list for genes that i) reduce or ii) increase *Brucella* infection

Table shows list of genes that are ranked according to their P value. Data represents output from RSA analysis using dataset from the primary screen (Dharmacon and Qiagen libraries), secondary screen (Ambion and Sigma libraries) and kinome screens (Ambion library). Independent replicates were averaged before RSA analysis was performed.

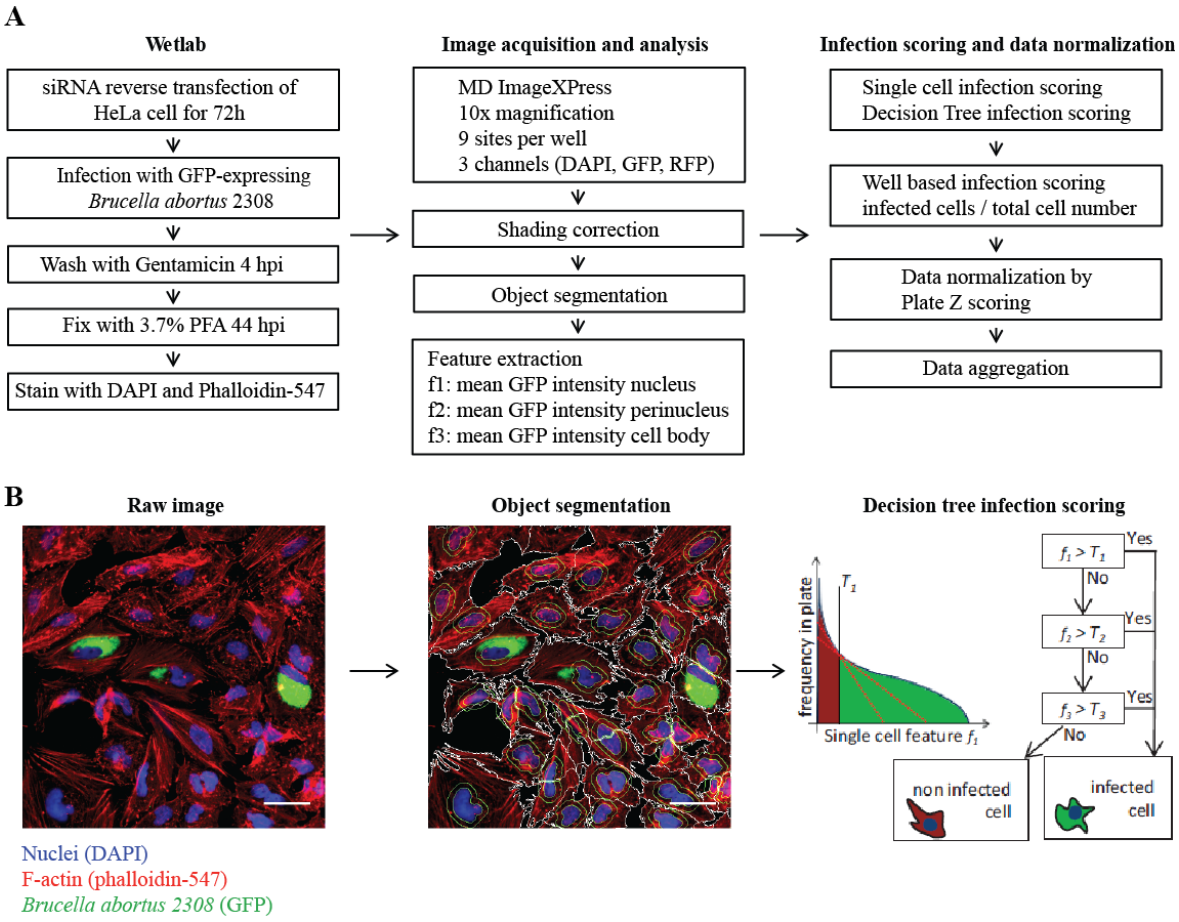
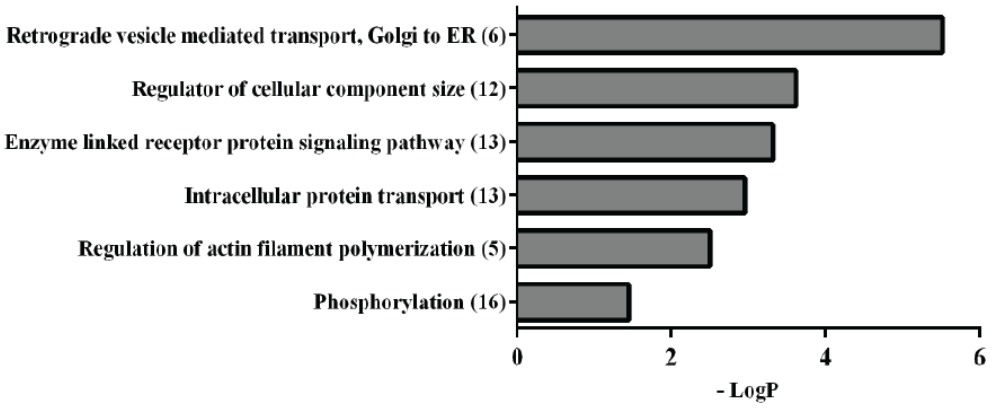

A
Wetlab
siRNA reverse transfection of HeLa cell for 72h
Infection with GFP-expressing *Brucella abortus* 2308
Wash with Gentamicin 4 hpi
Fix with 3.7% PFA 44 hpi
Stain with DAPI and Phalloidin-547
Image acquisition and analysis
MD ImageXPress
10x magnification
9 sites per well
3 channels (DAPI, GFP, RFP)
Shading correction
Object segmentation
Feature extraction
f1: mean GFP intensity nucleus
f2: mean GFP intensity perinucleus
f3: mean GFP intensity cell body
Infection scoring and data normalization
Single cell infection scoring
Decision Tree infection scoring
Well based infection scoring
infected cells / total cell number
Data normalization by
Plate Z scoring
Data aggregation
B
Raw image
Object segmentation
Decision tree infection scoring
frequency in plate
T_1
Single cell feature f_1
$f_1 > T_1$
Yes
No
$f_2 > T_2$
Yes
No
$f_3 > T_3$
Yes
No
non infected cell
infected cell
Nuclei (DAPI)
F-actin (phalloidin-547)
Brucella abortus 2308 (GFP)
C
Primary screen (~18'690 genes)
Dharmacon pooled library (1 pool of 4 siRNAs/gene)
Qiagen unpooled library (~4 siRNAs/gene)
Redundant siRNA analysis (RSA) and pathway analysis for up and down hits
Genes selected by other research groups in the consortium
Secondary screen
Ambion unpooled library (~6 siRNAs/gene)
Sigma esiRNA pooled library (1 pool/gene)
RSA and pathway analysis of primary and secondary screens for up and down hits

Figure 1

A


Retrograde vesicle mediated transport, Golgi to ER (6)
Regulator of cellular component size (12)
Enzyme linked receptor protein signaling pathway (13)
Intracellular protein transport (13)
Regulation of actin filament polymerization (5)
Phosphorylation (16)
0
2
4
6
- LogP

B

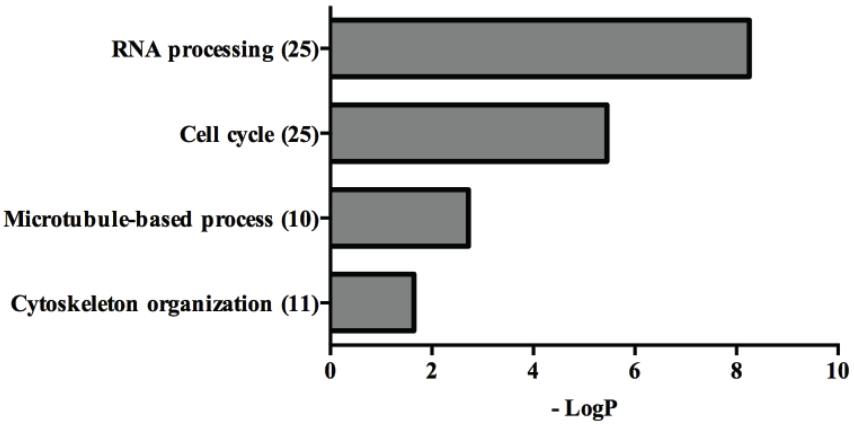

RNA processing (25)
Cell cycle (25)
Microtubule-based process (10)
Cytoskeleton organization (11)
0
2
4
6
8
10
- LogP

Figure 2

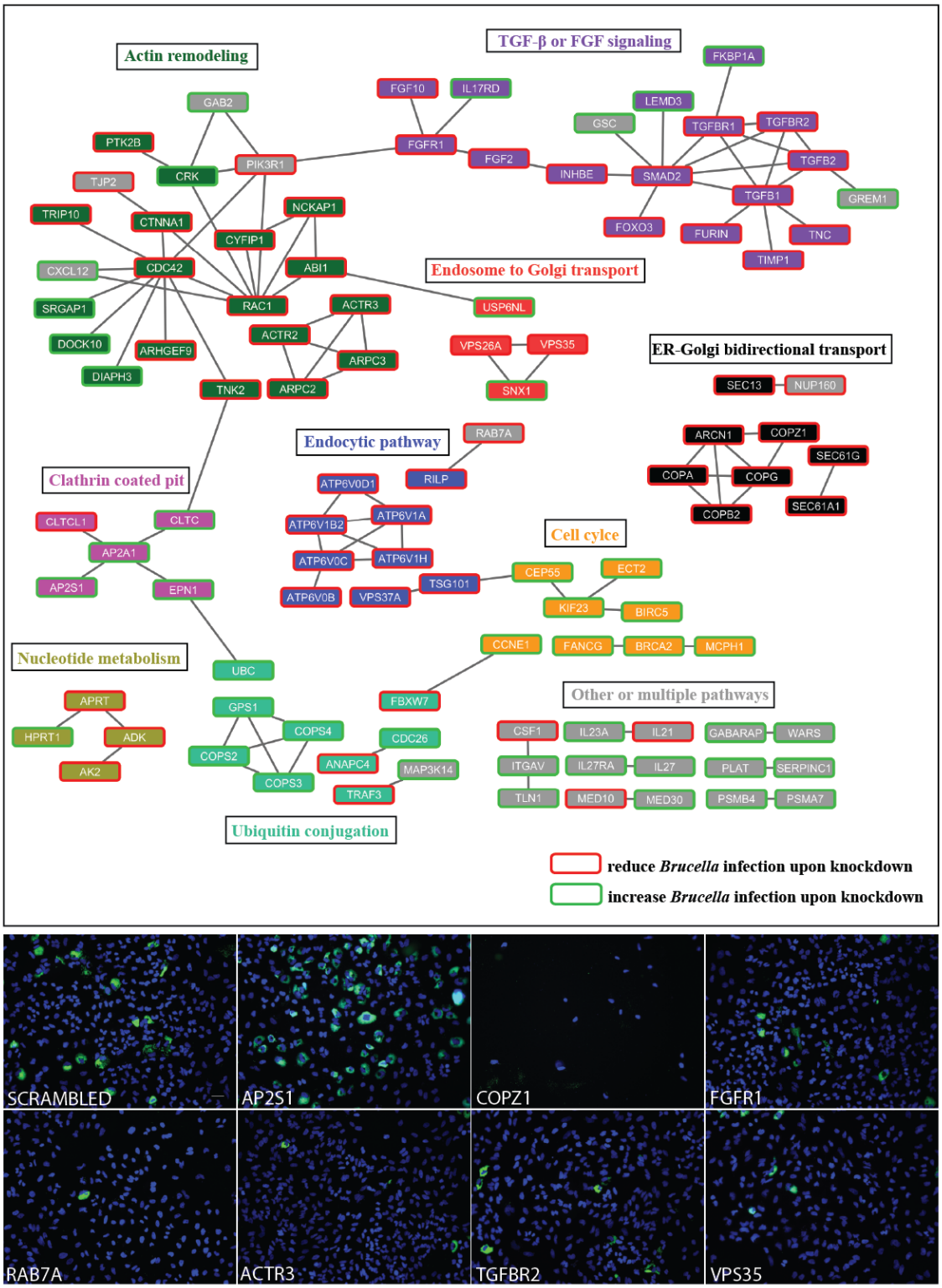
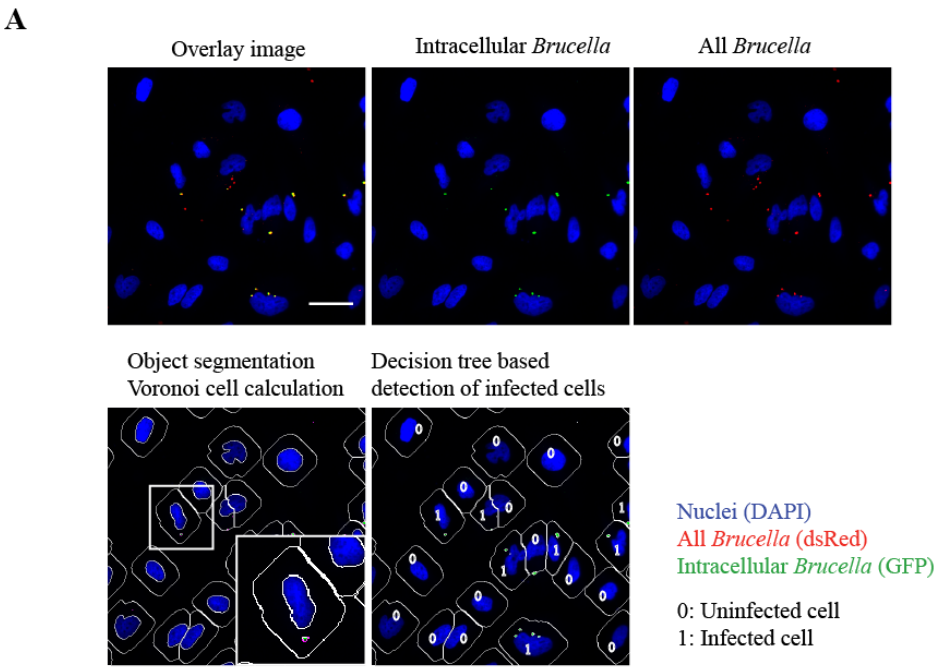

Actin remodeling
GAB2
PTK2B
CRK
PIK3R1
TJP2
TRIP10
CTNNA1
NCKAP1
CYFIP1
CXCL12
CDC42
ABI1
SRGAP1
RAC1
ACTR3
ACTR2
DOCK10
ARHGEF9
ARPC3
DIAPH3
TNK2
ARPC2
TGF-β or FGF signaling
FGF10
IL17RD
FGFR1
FGF2
INHBE
FKBP1A
LEMD3
GSC
TGFBR1
TGFBR2
SMAD2
TGFB2
TGFB1
GREM1
FOXO3
FURIN
TNC
TIMP1
Endosome to Golgi transport
USP6NL
VPS26A
VPS35
SNX1
ER-Golgi bidirectional transport
SEC13
NUP160
ARCN1
COPZ1
SEC61G
COPA
COPG
SEC61A1
COPB2
Endocytic pathway
RAB7A
RILP
ATP6V0D1
ATP6V1B2
ATP6V1A
ATP6V0C
ATP6V1H
ATP6V0B
VPS37A
TSG101
Clathrin coated pit
CLTCL1
CLTC
AP2A1
AP2S1
EPN1
Cell cylce
CEP55
ECT2
KIF23
BIRC5
CCNE1
FANCG
BRCA2
MCPH1
Nucleotide metabolism
UBC
APRT
HPRT1
ADK
AK2
GPS1
COPS4
COPS2
COPS3
FBXW7
CDC26
ANAPC4
MAP3K14
TRAF3
Other or multiple pathways
CSF1
IL23A
IL21
GABARAP
WARS
ITGAV
IL27RA
IL27
PLAT
SERPINC1
TLN1
MED10
MED30
PSMB4
PSMA7
Ubiquitin conjugation
reduce Brucella infection upon knockdown
increase Brucella infection upon knockdown
SCRAMBLED
AP2S1
COPZ1
FGFR1
RAB7A
ACTR3
TGFBR2
VPS35

Figure 3


A
Overlay image
Intracellular *Brucella*
All *Brucella*
Object segmentation
Voronoi cell calculation
Decision tree based
detection of infected cells
Nuclei (DAPI)
All *Brucella* (dsRed)
Intracellular *Brucella* (GFP)
0: Uninfected cell
1: Infected cell

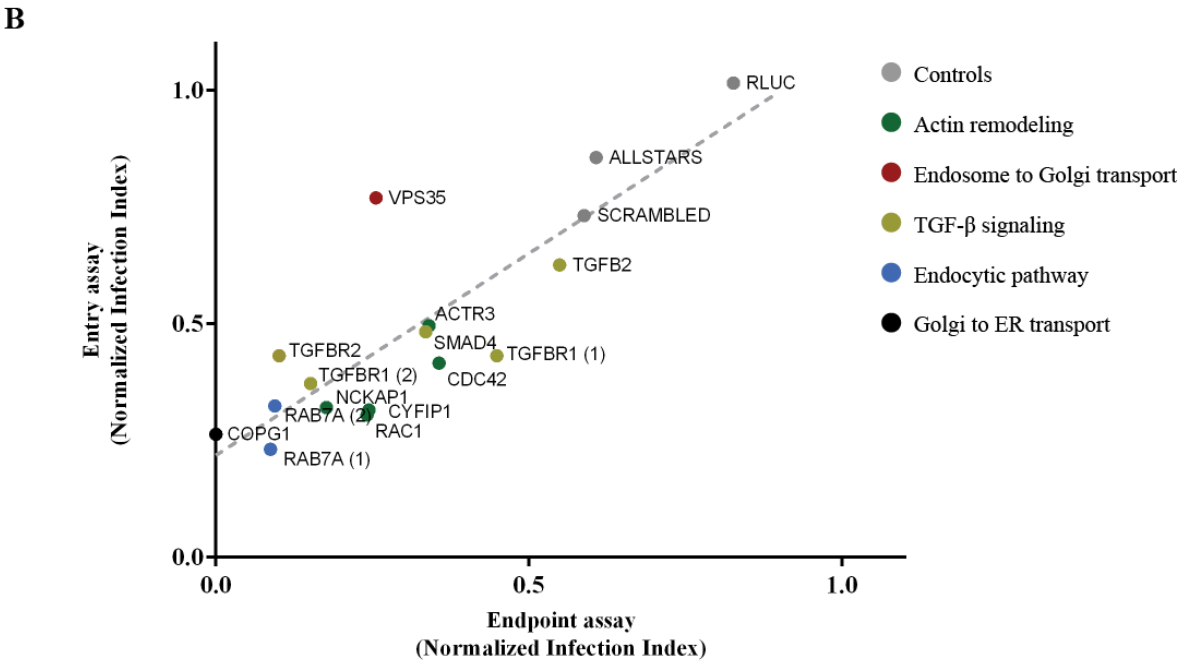

B
Entry assay
(Normalized Infection Index)
Endpoint assay
(Normalized Infection Index)
1.0
0.5
0.0
RLUC
ALLSTARS
VPS35
SCRAMBLED
TGFB2
ACTR3
SMAD4
TGFBR1 (1)
TGFBR2
TGFBR1 (2)
CDC42
NCKAP1
RAB7A (2)
CYFIP1
RAC1
COPG1
RAB7A (1)
Controls
Actin remodeling
Endosome to Golgi transport
TGF-β signaling
Endocytic pathway
Golgi to ER transport

Figure 4

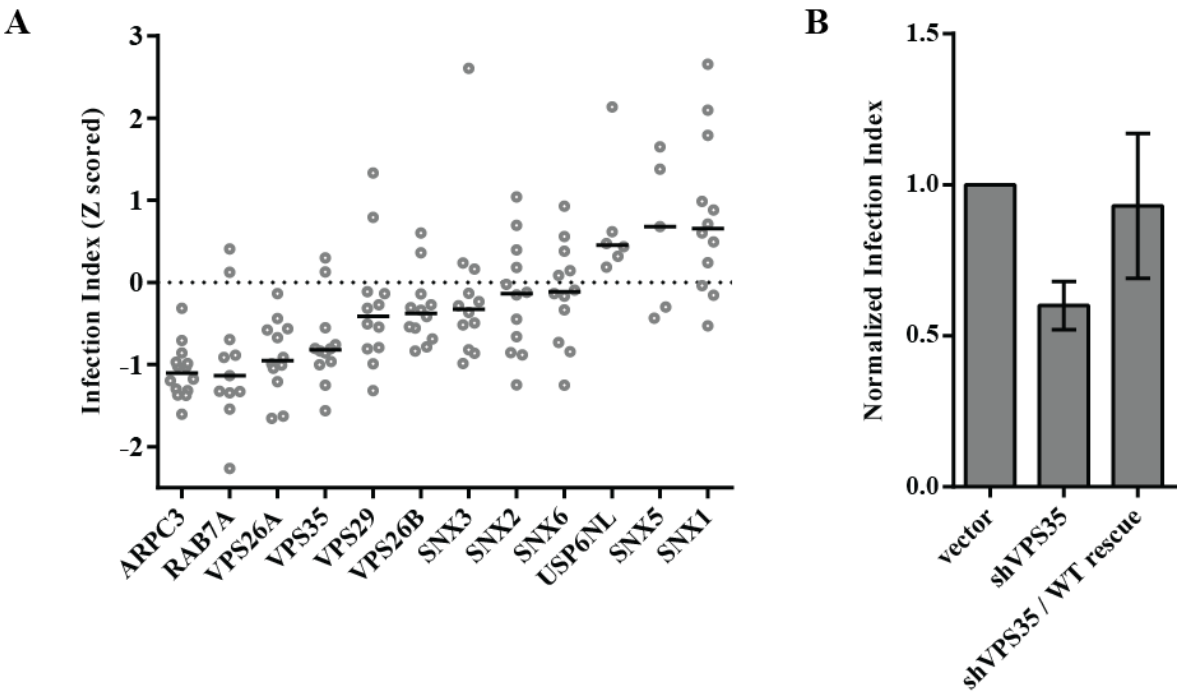

A
Infection Index (Z scored)
3
2
1
0
-1
-2
ARPC3
RAB7A
VPS26A
VPS35
VPS29
VPS26B
SNX3
SNX2
SNX6
USP6NL
SNX5
SNX1
B
Normalized Infection Index
1.5
1.0
0.5
0.0
vector
shVPS35
shVPS35 / WT rescue

Figure 5

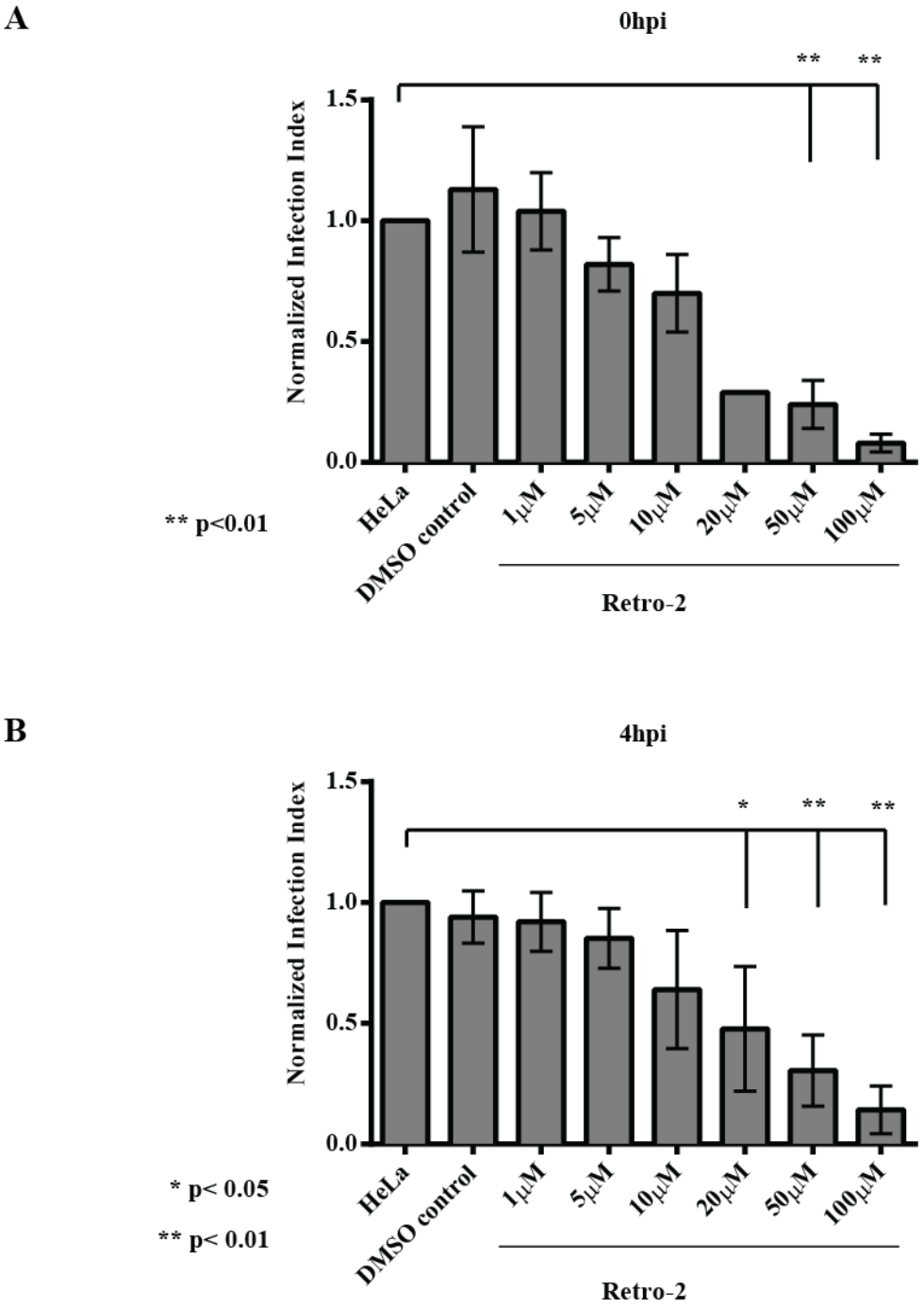
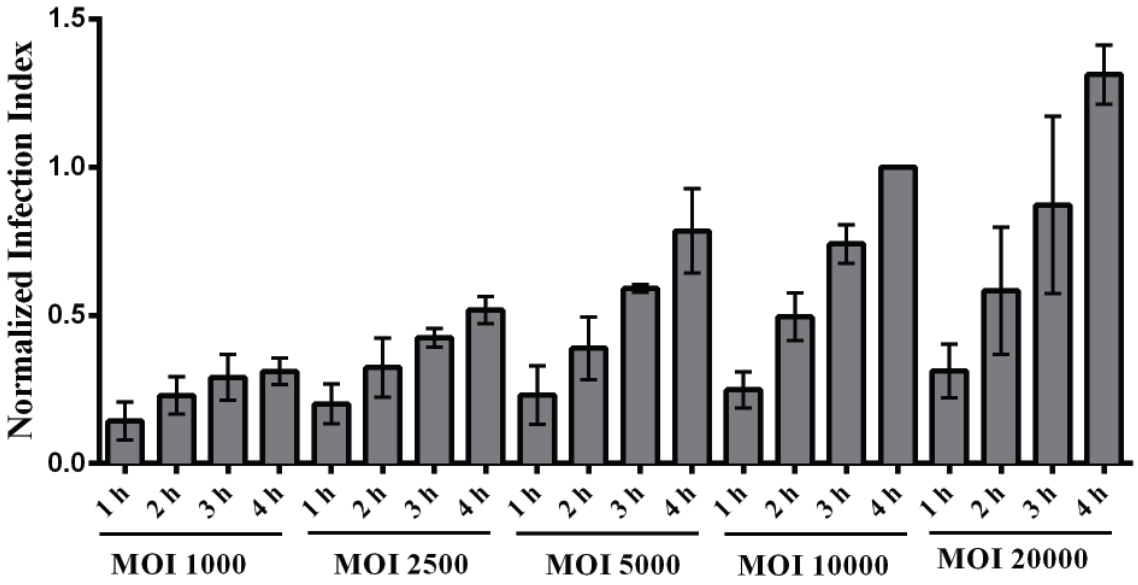
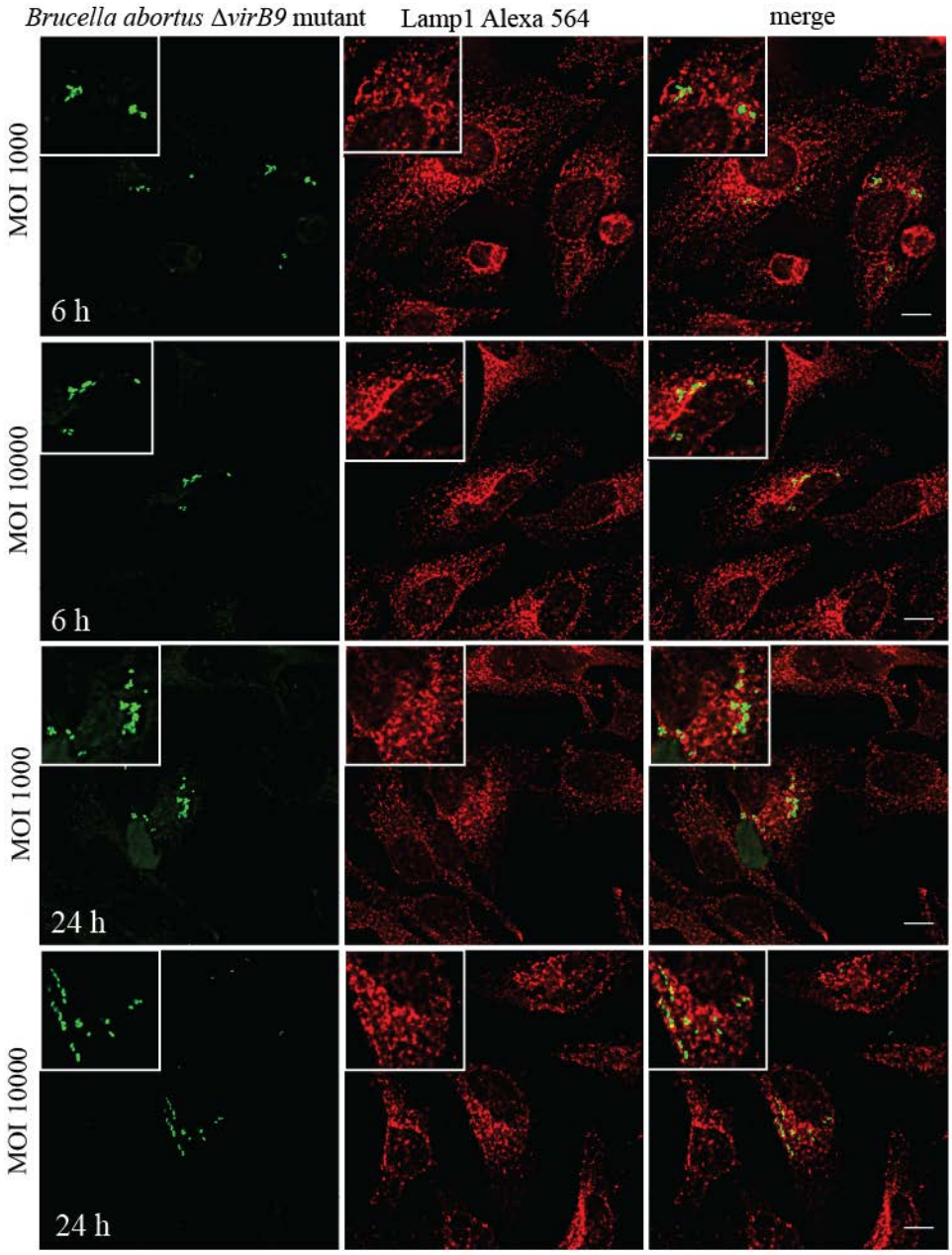

A
0hpi
** **
Normalized Infection Index
1.5
1.0
0.5
0.0
HeLa
DMSO control
1μM
5μM
10μM
20μM
50μM
100μM
Retro-2
** p<0.01
B
4hpi
* ** **
Normalized Infection Index
1.5
1.0
0.5
0.0
HeLa
DMSO control
1μM
5μM
10μM
20μM
50μM
100μM
Retro-2
* p< 0.05
** p< 0.01

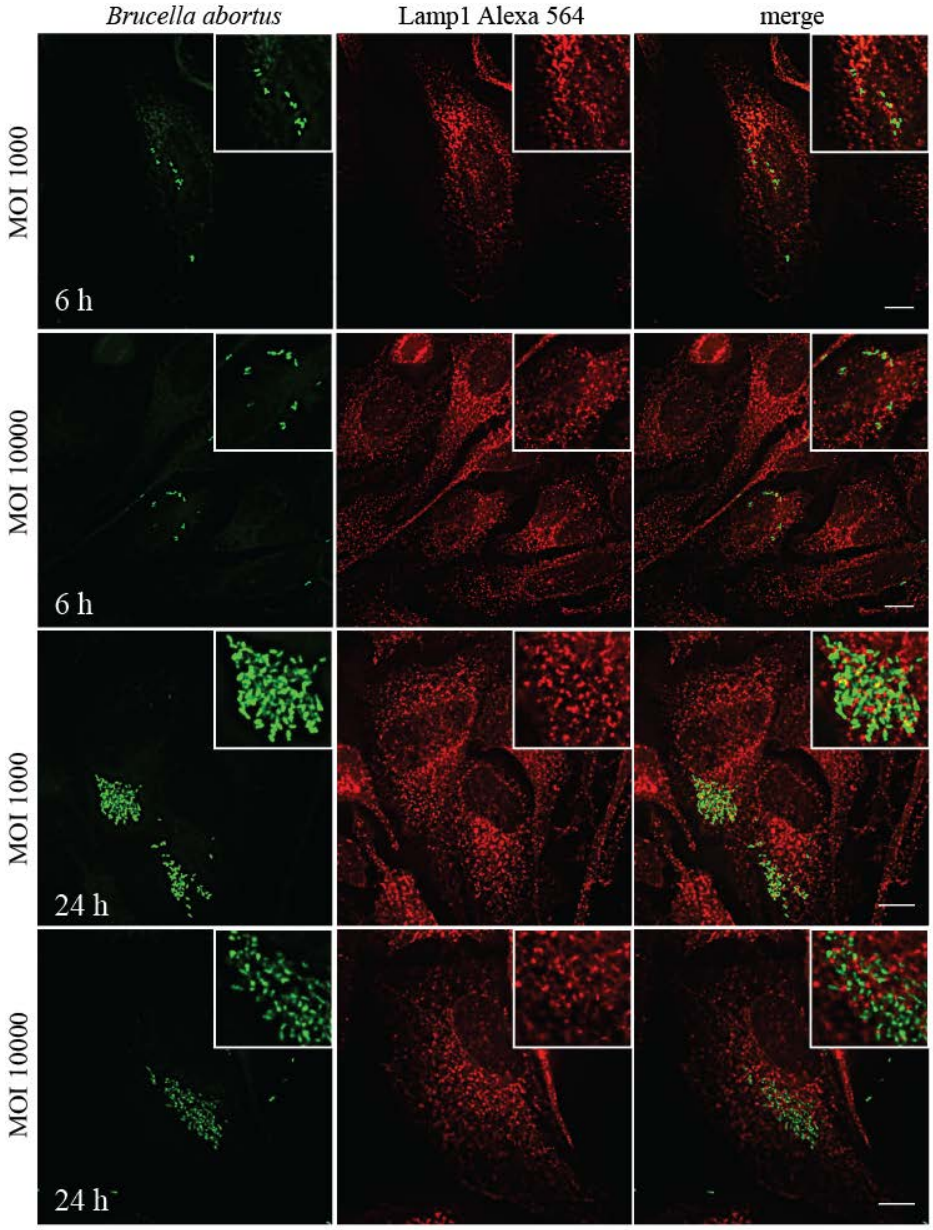
Figure 6



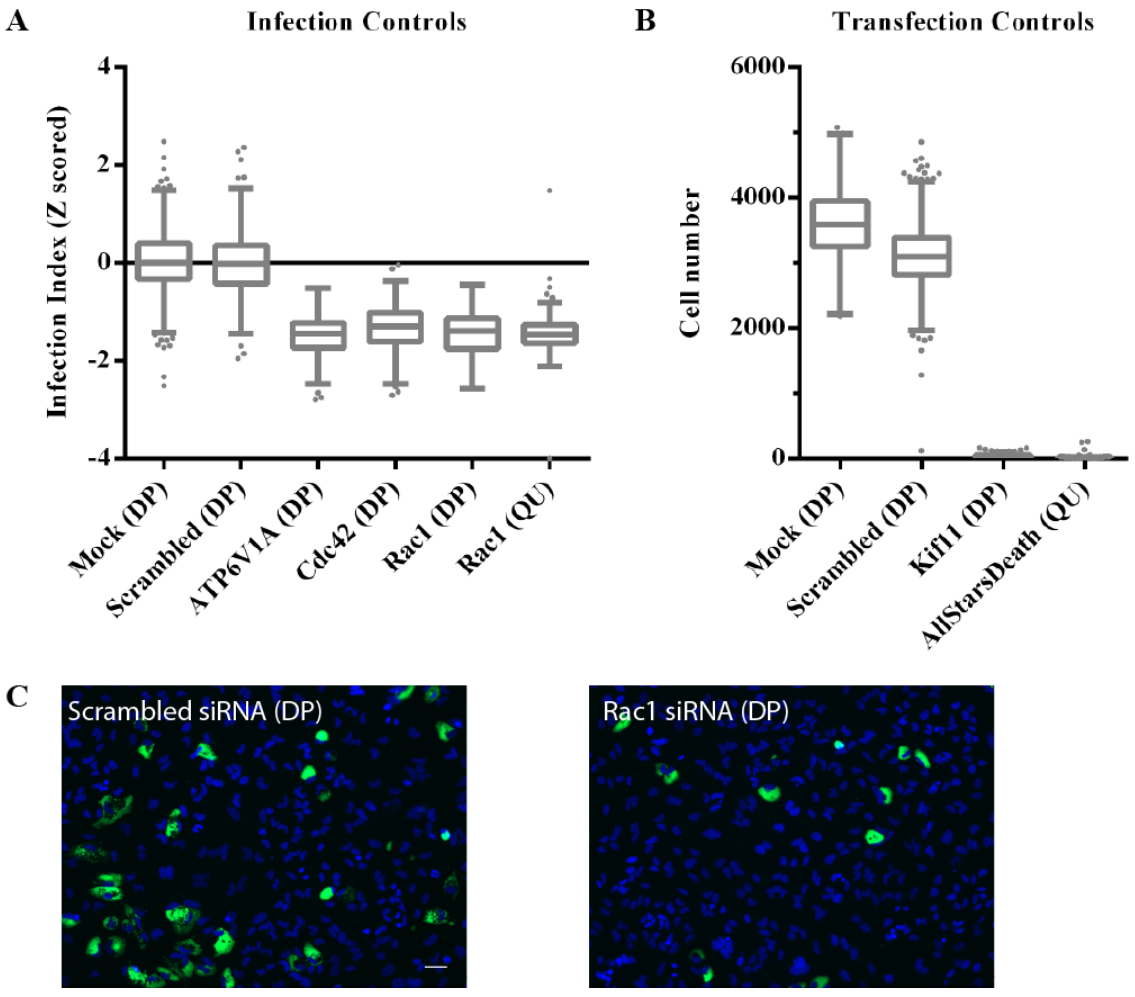
Supplementary Figure 1



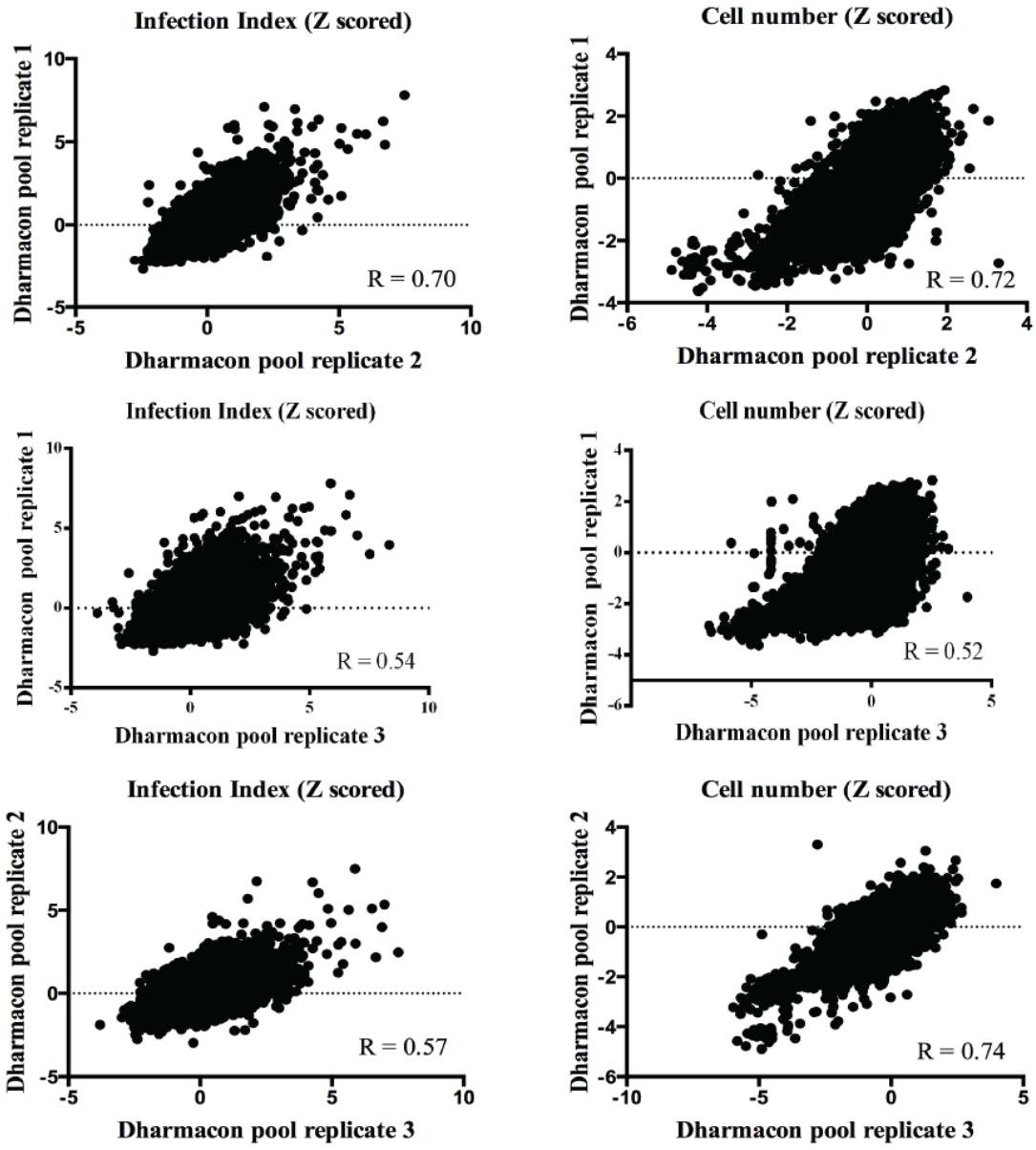
Supplementary Figure 2i



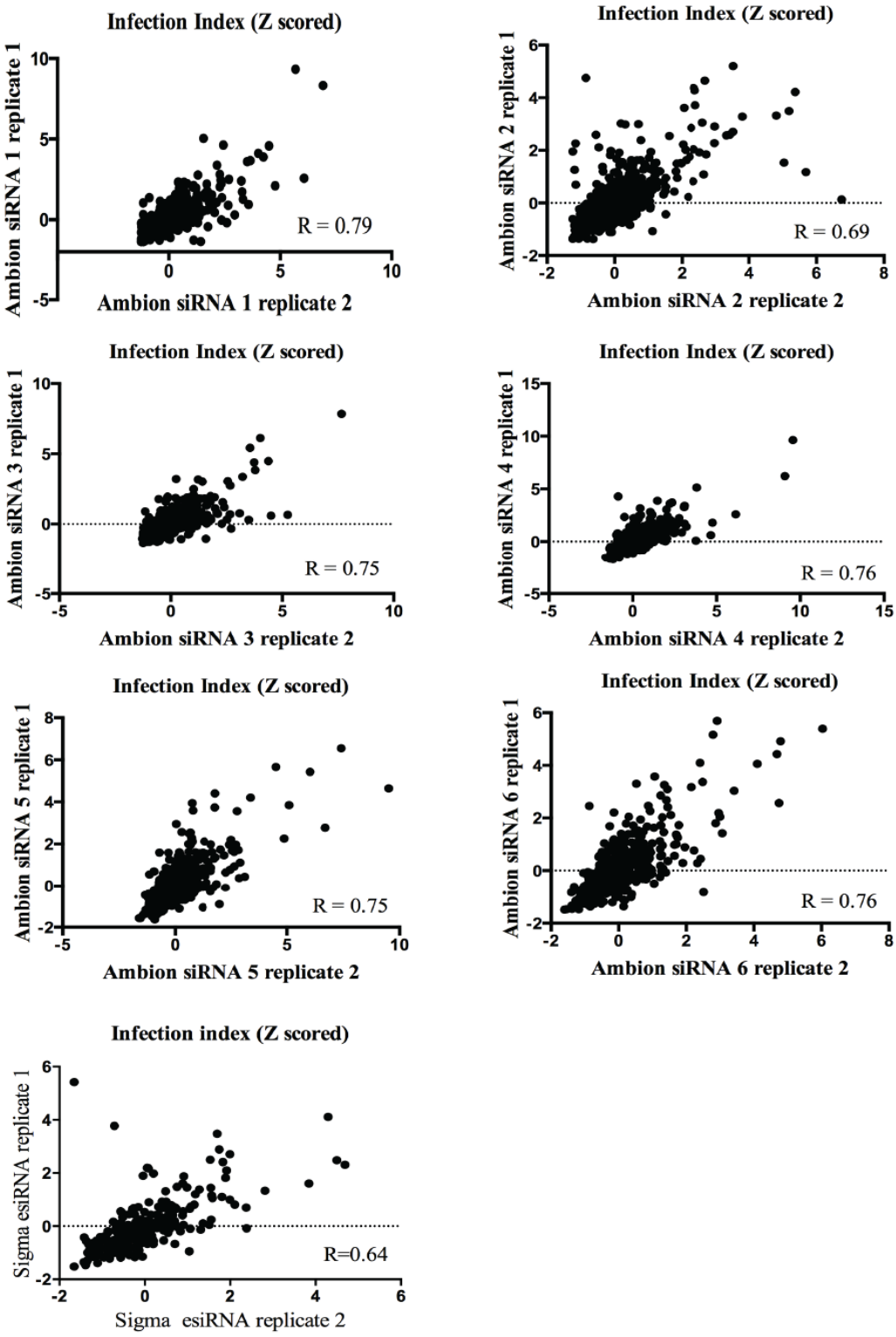
Supplementary Figure 2ii



Supplementary Figure 3



Supplementary Figure 4i



Supplementary Figure 4ii

4 Concluding remarks

4 CONCLUDING REMARKS

4.1 GENOME-WIDE SIRNA SCREENING

4.1.1 Sequence specific off-target effects – challenges and opportunities

High-throughput, genome-wide perturbation studies using siRNAs are valuable tools to identify novel genes and pathways of biological functions in organisms where knockouts are technically challenging. One of the mayor difficulties of siRNA screening are so called off-target effects. These include all unwanted effects which do not relate to the depletion of the intended target of the siRNA (on-target). Screens often show a high false positive discovery rate and the overlap between comparable siRNA screens can be extremely low (*1*). A meta-analysis of three genome-wide HIV screens showed that the overlap between identified host factors was only mildly better than expected by chance (*2*). Soon after the introduction of siRNA technology, transcriptome analysis of siRNA treated cells identified miRNA-like off-target effects (*3*). In this case, siRNAs guide the RISC complex to mRNAs with partial complementarity and reduce expression of those genes, which is reminiscent of the action of miRNAs.

When analyzing genome-wide siRNA screens for host factors that are involved in the infection of intracellular pathogens, only low correlations between different siRNA libraries were found (*research article II*). Overall, testing of two different siRNAs which are designed to target the same gene often resulted in very distinct effects on cell number and pathogen infections. However, when we compared siRNAs designed against different targets that contain the same heptameric sequence at position 2-8 at the 5' end of the siRNA, we found a very high correlation. Position 2-8 of a natural miRNA is called the seed region, which largely determines the target specificity of the miRNA (*4*). This indicates that the siRNAs in our screens often act like miRNAs. In *research article II* we could validate this finding by designing custom siRNAs which only contain active seed sequences but cannot bind to any mRNA by perfect complementarity. This allowed us to identify seed sequences with a variety of biological functions that were covered by the analysis of siRNA induced phenotypes. We found seeds which inhibited or promoted infection of one or several pathogens, reduced or increased cell number, or affected several phenotypes simultaneously. Some of the identified seed sequences are found in natural

human miRNAs and testing of miRNA mimics could confirm the effect that was predicted from off-target analysis of the genome-wide siRNA screens. Altogether, seed analysis allowed us to identify natural miRNAs and synthetic miRNA-like molecules which regulate the growth of cells and infection of bacterial and viral pathogens. Importantly, this analysis can be applied to published siRNA screens to identify miRNAs which are involved in the corresponding processes if the sequences of the siRNAs are published.

While the knowledge gained from off-target effects can be valuable, it may not answer the question of which genes underlie the process studied by the siRNA screen. Here, a number of strategies can be considered to counteract off-target effects and optimize the true positive discovery rate. For screens where several siRNAs per gene have been tested, statistical methods that rank siRNAs and genes according to statistical probability can be used. These do not depend on sequence information of the siRNAs and can also be applied to siRNA pools. An example is the commonly used Redundant siRNA Activity (RSA) method (*5*). Similarly, we could show in *research article IV* that statistical analysis of different siRNA screens which are performed using identical sets of siRNAs can improve the discovery of hits for each individual screens. With this approach, genes which are involved in the biology of different screens gain statistical power while factors uniquely involved in a single screen remain unaffected.

Other methods depend on the identification of seed-driven off-target effects (*6*). Seed analysis only yields valuable information for those seeds which are present in many different siRNAs. In this case, the effect of the seed sequence can be well estimated. Optimally, siRNA libraries would thus be designed in a way that only a limited number of seeds are use, but all of them in sufficiently high numbers that the seed effect can be quantified during each screen. The seed information can then be used to remove seed-driven siRNAs from the analysis or to correct for the off-target effect (*7*). While excluding siRNAs mainly reduces the number of false positives, correction of the data potentially helps to then recover some of the false negatives. However, this approach requires an estimate of how strongly the original data should be corrected. It assumes that on- and off-target phenotypes are to some degree additive, which might not always be the true. Depending on the underlying biology, on- and off-target effects can be synergistic or epistatic which complicates their disentanglement. To avoid over-correction, only a subset of the full data should be corrected and compared to the remaining uncorrected part. Improved correlation between the two sets indicates that the correction recovered some of the on-target

information. Yet it is important to remember that corrections of individual siRNAs might not always be beneficial due to the reasons mentioned above.

Alternative methods even go one step further and use the information of the off-target effects to directly infer the genes that are involved in the process of study (*8, 9*). These rely on the prediction of all potential genes which might be targeted by an siRNA. The measured phenotype of a certain siRNA is assumed to be the sum of the effects for all those targets which are involved in the phenotype. Thus, the challenge is to identify among the long list of genes which can be targeted by the siRNA those, which affect the phenotype. If such an analysis is performed for a large number of siRNAs, a mathematical model can be used to fit phenotypic readouts to each gene until the model fits the actual measurements of the siRNAs from the screen. The result of such an analysis is a list of genes and their corresponding effect on the phenotype.

In summary, multiple published methods are available that can improve genome-wide siRNA screening data. These methods are based on very different approaches and rely to varying extents on the quantity and type of siRNA products used. It is thus likely that every method reveals a subset of all genes that are involved in a specific biological process.

For smaller screens, off-target effects cannot be readily calculated and alternative approaches are required. In such cases, experimental setups can be used to directly control for miRNA-like off-target effects. siRNA-specific controls which carry certain mutations in the siRNA but keep the seed region intact can be used. If the phenotype is not affected by mutations outside the seed sequence, a miRNA-like off-target effect has to be assumed and no conclusion on the depletion of the intended target can be drawn (*10, 11*).

An alternative approach to experimentally reduce sequence specific off-target effects is pooling of siRNAs. While every siRNA in the pool has an individual off-target spectrum, all can bind to the intended target. This should then reduce off-target effects by competition between the different sequences without affecting the knockdown efficiency on the intended target (*12*). Competition seems to be important, since simple dilution of the siRNA is not sufficient to selectively reduce off-target effects (*3, 13*). Also the number of siRNAs which are combined is important and increasing the number of oligos per pool correlates with reduced off-targets (*12*). In agreement with this, we could still observe off-target signal in pools consisting of four siRNAs although with reduced strength compared to unpooled siRNAs.

A logical consequence of the above statement is that testing many different siRNAs (with different seed sequences) for the same gene individually should also average out off-target effects

and show the effect of depletion of the gene of interest. This approach is generally applied to validate primary hits from genome-wide screens and can be performed in high-throughput. In *research article IV* we have tested up to 12 different RNAi reagents for all human kinases. We also found that the confidence to identify true positive hits directly correlates with the number of siRNAs tested.

Once interesting candidate genes have been found, it is advisable to perform alternative validations. Complementation of the knockdown with an siRNA-resistant construct or inactivation of the target with small molecules are suitable approaches. Lately, the development of efficient genome-editing tools presents an attractive alternative to RNAi even though these are not yet readily applicable for genome-wide screening (*14-16*). Nevertheless, full loss-of-function studies of non-essential genes have become possible in a wide variety of mammalian cell types.

4.1.2 Identification of host factors and pathways involved in *Brucella* infection

Having identified the off-target effects in our siRNA screens, we set out to identify genes and pathways that are implicated in *Brucella* infection of human cells. Since we performed genome-wide screens with both, pooled and unpooled siRNAs, we decided to analyze the data using the statistical method RSA which is independent of seed analysis and can also be applied to siRNA pools (*5*). It was previously reported that even though the overlap between hit lists of comparable screens is very low due to off-targets, identified pathways show a much better overlap (*2*). This is likely due to different off-target effects which cause distinct false positives and negatives depending on the siRNA library which is used. However, if relevant pathways are sufficiently well described in literature, they can be consistently identified even though some components might be missed in one or the other screen. Therefore, we performed gene enrichment analysis with the top ranking genes of the hit list generated by RSA. This identified several significantly enriched pathways, some of which were known to be important for *Brucella* infection.

To validate individual genes we decided to re-screen interesting candidates with six additional siRNAs and one pool of siRNAs. This is a relatively high number of additional siRNAs considering that the primary screen already consisted of five different RNAi reagents. The aim was to reduce the false positive discovery rate and is based on our finding that hit detection directly correlates with the number of siRNAs tested per gene (*research article IV*). Some of the genes that were in the top ranks of the RSA analysis were directly selected for validation. A

second set of genes was chosen from enriched pathways. In this case, we also included pathway members that did not show a strong effect on *Brucella* infection in the primary screen reasoning that they might have been missed in the primary screen due to technical or off-target related reasons. Since these screens were performed in the framework of the InfectX consortium, we also re-screened all the genes that were selected by the other research groups which study different intracellular pathogens. This is beneficial in several ways; it provides sufficient negative controls which are important for plate normalization, it optimizes the use of siRNA reagents, and potentially allows for the identification of host factors shared between different intracellular pathogens.

Applying this validation strategy, we could identify genes involved in *Brucella* infection in several cellular pathways. Some of these were previously reported, such as subunits of the v-ATPase complex, Rab7A, Rac1, and Cdc42 (*17-19*). Furthermore, multiple components of pathways which have not been connected to *Brucella* infection previously could be identified. involved in TGF-β or FGF signaling, actin remodeling, endosome to Golgi transport, endocytic pathway, ER-Golgi bidirectional transport, ubiquitin conjugation, nucleotide metabolism, cell cycle, and clathrin coated pits were found in our top ranking gene lists suggesting a role of these pathways in *Brucella* infection.

The genome-wide screen covered multiple aspects of the *Brucella* infection from entry into the host cell up to proliferation inside the replicative niche. To further understand the molecular role of the identified host factors during the infection process we performed an entry assay on some of the hits covering the major pathways. Most of the host factors that were selected showed a direct correlation of infection between the results of both assays indicating that depletion of these factors affects bacterial entry. This might seem surprising, because some of these genes have well known functions in intracellular trafficking, such as ER-Golgi bidirectional trafficking or endosome maturation. However, it is important to consider that depletion of vesicular trafficking components over the course of three days can cause secondary effects on several other trafficking routes including endocytic pathways. This was illustrated by the finding that the COPII complex, involved in ER-Golgi transport, is required for *Salmonella* invasion since it maintains cholesterol, sphingolipids, Rac1, and CDC42 at the plasma membrane (*20*). Similar phenomena could also apply to prolonged depletion of other components of vesicular trafficking that were tested here. As an example, depletion of Rab7, which has previously been implicated in *Brucella* trafficking (*19*), was found here to affect *Brucella* entry into HeLa cells.

To fully understand the role(s) of the identified host factors, additional experiments will be required. Especially for factors related to trafficking it will be important to study their effects on *Brucella* infection using methods that induce rapid perturbation to circumvent secondary effects. If available, the use of drugs or temperature sensitive mutants could be suitable approaches to unravel the exact role if individual components. In addition it will be interesting to follow *Brucella* infection in cells depleted for some of these host factors over time using live microcopy. This will help to understand whether depletion of a gene that was found to reduce invasion also perturbs later steps during infection. Tracking of individual bacteria will show whether the onset of cell division, intracellular survival rates, or the speed of intracellular replication are affected by depletion of certain host factors.

Among the tested genes, only depletion of Vps35 did not affect *Brucella* entry of HeLa cells but reduced infection in the endpoint assay. Vps35 is an essential component of the retromer which is involved in vesicular transport from endosomes to the trans-Golgi network (TGN) and recycling of cargo to the plasma membrane (*21*-*23*). The importance of this pathway for *Brucella* infection was further validated by the use of the specific inhibitor Retro-2. This drug has been shown to selectively block retrograde trafficking of different toxins and human papillomaviruses from endosomes to the trans-Golgi-network (*24*, *25*), without affecting the trafficking of other retrograde cargoes, the morphology of compartments, or other trafficking routes. We found that addition of the drug after *Brucella* entry into host cells could block intracellular replication. This emphasizes the importance of this pathway during a post-entry process of *Brucella.*

It remains to be investigated whether the BCV traffics through this compartment before reaching its replicative niche, or whether this transport route delivers membranes or host factors which are required for intracellular proliferation. Until now, it has not been described that *Brucella* traffics through Golgi compartments, yet, the interaction might be very short-lived. Live imaging using fluorescently labeled Golgi and retromer markers might reveal such interactions and give indications whether bacteria that ultimately manage to replicate within a cell showed previous interactions with the retromer. Alternatively, *Brucella* might depend on retrograde transport to establish or maintain the replicative niche. In this case, chemical inhibition at later stages during the infection will give further insight into the exact mechanism of retrograde trafficking during *Brucella* infection.

4.2 ROLE OF MIRNAS IN PATHOGEN INFECTION

4.2.1 miR-103/107 promote *Brucella* entry in non-phagocytic cells

miRNAs are small RNA molecules that belong to the repertoire of regulatory elements for gene regulation in eukaryotic cells. They negatively affect target gene transcription by translational inhibition and destabilization of the mRNA (*26*). miRNAs act by partial base pairing of the miRNA to target mRNAs which allows each miRNA to regulate a large number of different transcripts. The binding specificity of a miRNA is largely determined by complementarity of the seed region to the target mRNA (*4*). The seed comprises a stretch of 6-8 nucleotides located at the 5' end of the miRNA. While miRNAs are well known for their roles in developmental regulation or cell differentiation, there is increasing evidence for their involvement during pathogen infection (*27*).

In order to investigate the role of human miRNAs in *Brucella* infection, we screened a library of over 1200 human miRNA mimics. We found that a large number of miRNAs affect *Brucella* infection when transfected into HeLa cells. Among those miR-103 and miR-107 (miR103/107) caused a striking increase in infection (*research article III*). This effect was visible early during infection, indicating that already entry into the host cell was enhanced. miR103/107 belong to a broadly conserved family of miRNAs (*28*) and transfection of a miR-103 mimic into mouse embryonic fibroblasts resulted in a similar increase of *Brucella* uptake. This suggests that in addition to the conservation of the miRNA itself, also the targets which protect cells from infection and are repressed by miR-103/107 are conserved. It will be interesting to test whether this effect is restricted to non-phagocytic cells or whether miR-103/107 also promotes *Brucella* infection in professional phagocytes such as macrophages.

Even though miRNAs preferentially act by target recognition of the seed sequence, alternative miRNA-mRNA interactions have been described (*29, 30*). Because miR-103 and miR-107 share almost the same sequence it is not possible to distinguish whether the targets relevant for increased infection are recognized by the seed or other parts of the miRNA. Off-target analysis of the genome-wide siRNA screens indicated that miR-103/107 act via the seed sequence, since siRNAs sharing the seed with these miRNAs also strongly increased *Brucella* infection. To formally proof this, the seed sequence of miR-103 or miR-107 could be mutated and these variants could be tested for their effect on *Brucella* infection.

The information that the miRNA dependent phenotype is seed mediated can be used to predict potential targets since most prediction software rely on seed based miRNA-mRNA target interactions (*31*). However, the list of predicted targets is generally in the range of several hundred genes complicating direct target-phenotype associations. To experimentally determine the effect of miR-103/107 on cells we analyzed the proteome and transcriptome changes induced by transfection of miRNA mimics. We found an overall good correlation between mRNA and protein changes, which was even higher for the genes which are predicted to be direct targets. It is expected to find a higher correlation for direct targets, because secondary effects can be very diverse with distinct influences on protein and transcript abundance. Furthermore, both approaches showed that miR-103 targets predicted by different algorithms were down-regulated (*31, 32*). This is in agreement with the general notion that miRNAs destabilize the target mRNA und thereby reduce protein biosynthesis (*33, 34*).

Proteome analysis revealed that several factors associated with the cell surface, such as TGF-β or FGF receptors, showed increased protein levels. This is also reflected in the gene ontology analysis of upregulated proteins, where terms such as insoluble fraction, cell adhesion, or plasma membrane were found to be significantly enriched. This indicated that the endocytic properties of cells could be altered when miR-103/107 levels are high, which is in agreement with the observation that clathrin-mediated endocytosis (CME) of transferrin was reduced in HeLa cells upon miR-103 transfection. It remains to be investigated whether this is caused by a reduced endocytosis rate or increased recycling of the receptor to the cell surface. Both scenarios could explain why certain proteins such as the TGF-β receptor 2 (TGFBR2), which undergo CME, were enriched (*35*). In this respect, it will be important to test the uptake of additional ligands which are endocytosed in a clathrin-dependent manner. Furthermore, protein quantification of the surface composition of cells transfected with miR-103/107 will show whether receptors are indeed enriched on the cell surface or stabilized in intracellular compartments.

Interestingly, depletion of several components involved in CME was found to enhance *Brucella* infection in the genome-wide siRNA screen (*research article V*). This would be in line with the notion that reduced CME benefits *Brucella* infection but contradicts recent work by Lee *et al*. who found opposing effects (*36*). It remains to be investigated whether the different HeLa cell line or the specific biovar of *B. abortus* used in this study account for these inconsistencies. However, it is conceivable that *Brucella* favors clathrin-independent endocytic mechanisms to enter non-phagocytic cells. These could be induced to compensate for a reduction of CME.

Macropinocytosis, which is well-known for its role in bacterial uptake (*37*), was found to be upregulated in HeLa cells when CME was inhibited (*38*). It will thus be interesting to study the contribution of macropinocytosis and other endocytic mechanisms in the presence and absence of high levels of miR-103/107 to *Brucella* uptake in HeLa cells.

Among the proteins upregulated by miR-103/107, TGFBR2 was found previously (*research article IV*) to be required for *Brucella* infection, and depletion of several components of the TGF-β signaling pathway negatively affected *Brucella* entry into HeLa cells (*research article V*). We thus hypothesized that increased activity of this pathway could benefit *Brucella* infection. Overexpression of TGF-β receptors 1 and 2 or pretreatment of cells with the ligand TGF-β1 indeed increased *Brucella* infection. The effects were not as pronounced as seen with transfection of miR-103/107 mimics, but it is conceivable that the increased TGFBR2 levels could account for parts of the phenomenon. Epistasis experiments with simultaneous transfection of miR-103/107 mimics and siRNA depletion of TGFBR2 could give indications whether this hypothesis holds true.

The expression of direct targets of a miRNA is generally repressed in the presence of the miRNA, while the surface receptors described above represent proteins which were upregulated upon increased cellular levels of miR-103/107. Thus, they are unlikely to qualify as direct targets. Among the genes which were repressed by miR-103/107 the strongest enrichment was found for proteins which are involved in cell cycle progression from G1 to S phase. In agreement with a function on cell cycle, it was previously reported that miR-103/107 can repress the growth of tumor cells (*39, 40*). Consistently, we found a mild reduction in HeLa cells number three days after transfection with miR-103/107 mimics, which was aggravated during the following two days. Since the reduction in cell number at the time of infection is only very mild, we can exclude that the observed increase in infection is caused by an increased ratio of bacteria to cells. It remains unclear whether alterations of the cell cycle are involved in the increased susceptibility of HeLa cells to *Brucella* infection. However, other miRNAs, such as members of the miR-15/16 family, which also affect G1/S phase transition in a similar manner as miR-103/107, did not promote *Brucella* infection (*41-43*).

Beside the cell cycle genes we could also identify factors which were down-regulated by miR-103/107 on the protein level and identified in the genome-wide siRNA screen to promote infection upon depletion. Among those, three are predicted to be direct targets of miR-103/107 and thus candidates that could underlie the phenotype. The most promising is NFIB, since

siRNA-mediated depletion caused a stronger increase in *Brucella* infection compared to knockdowns of MAP4 or PDE8B. NFIB is a transcriptional regulator of the nuclear factor I family which regulate a large number of genes involved in processes ranging from stem cell biology to brain development (*44, 45*). Again, siRNA mediated depletion of NFIB did only show a mild increase in *Brucella* infection suggesting that it could only account for parts of the phenotype.

Taken together, it seems that not a single target underlies the strong induction of *Brucella* infection by transfection of miR-103/107 to non-phagocytic cells. This notion is supported by the observation that such as strong increase in infection could not be observed for any gene-specific knockdown in a consistent manner in the full genome-wide siRNA screen. Either, we were not able to identify this gene due to technical limitation such as off-target effects or this factor does not exist. The latter would suggest that simultaneous depletion of several host factors by miR-103/107 acts synergistically. miRNAs are known for their pleiotropic effects and their capacity to strongly alter the transcriptome of a cell in the direction of cells which naturally express this miRNA (*33*). While miR-103/107 are broadly expressed throughout different tissues, it is difficult to directly correlate its role to cell type specific functions (*28*). It is conceivable that miR-103/107 alter the endocytic capacities or the surface composition of non-phagocytic cells in favor of *Brucella* uptake.

To unravel the molecular network of miR-103/107 targets that underlie the full phenotype, combinatorial siRNA depletion studies could be employed. Promising candidates would be factors which were found depleted in miR-103/107 treated cells and are predicted to be direct targets. These include several transcription factors, cell cycle regulators, and individual genes with functions that could relate to membrane dynamics. In addition to siRNA depletion, also the use of chemical interference of certain functions such as the cell cycle, or specific endocytic pathways could give further indications of their importance during *Brucella* infection.

4.2.2 miR-103/107 enhance infection of diverse intracellular pathogens

Within InfectX, several research groups screened the miRNA mimics library for their effect on different bacterial and viral pathogens. We found that in addition to *Brucella*, also several other pathogens showed enhanced infection upon miR-103/107 transfection (*research article III*). These include *Salmonella* Typhimurium, *Listeria monocytogenes*, adenovirus, and rhinovirus,

which use a wide variety of different entry mechanisms and intracellular replication strategies (*37*). The screens that were performed with these organisms all covered cell invasion and further aspects of the intracellular life for the corresponding pathogens. It thus remains to be investigated whether pathogen entry into cells or a subsequent step of the infection is promoted by miR-103/107. However, since all these pathogens need to gain access to the interior of the host cell and because this is the stage which we found affected during *Brucella* infection, it is conceivable that also the uptake of other organisms is promoted by miR-103/107. In support of this, several cell surface proteins which are required by distinct pathogens showed elevated protein levels in HeLa cells upon miR-103/107 transfection (*research article III*). We thus speculate that similar to *Brucella*, elevated entry levels could underlie the observed increase in infection for other pathogens. This would support the notion, that high levels of miR-103/107 affect cells in a way that generally promotes pathogen infection which could be explained by changed endocytic properties.

Apart from *Brucella*, genome-wide siRNA screens were also performed with all the above mentioned bacterial and viral pathogens. This information has so far remained untapped for the efforts to identify the direct target(s) which underlie the effects of miR-103/107 on pathogen infection. The *Brucella* siRNA screen provided only limited candidates which at the same time showed increased infection upon siRNA knockdown, decreased protein levels upon miR-103/107 transfection, and qualify as direct miR-103/107 targets. However, a similar analysis for the other pathogen screens could reveal additional factors which might have been missed in the *Brucella* screen for different reasons, one of which could be off-target effects. While we found that certain miRNA-like off-target effects have an impact on the infection of several pathogens, others can be very specific (*research article II*). Therefore, they could mask the identification of a host factor in one screen while others might be unaffected. A similar analysis of all those screens as presented here on the example of *Brucella* could thus help to identify the molecular components which underlie the increased susceptibility of cells to pathogen infection induced by miR-103/107.

Because miR-103/107 are involved in the infection process of diverse intracellular pathogens they could be attractive targets for the development of novel intervention strategies to combat infectious diseases. Most antimicrobials directly target the pathogen, which is often associated with a rapid appearance of resistances (*46*). Strategies which interfere with host factors that are essential for the infection could potentially circumvent this problem and might also be used against different pathogens that depend on the same host function. To further investigate whether

miR-103/107 could be used as a target against infectious diseases, it will be important so study whether depletion of miR-103/107 renders cells less susceptible to pathogen infection. Especially cells which are infected by the pathogen in a natural setting should be studied. In the case of human brucellosis, bacteria reside in macrophages or dendritic cells before the spread to various tissues in later stages of the infection (*47*, *48*).

Because miR-103/107 is broadly conserved across vertebrates and we found that they promote *Brucella* infection in mouse cells it would be interesting to test whether inhibition of miR-103/107 could protect mice from *Brucella* infection. To this end, miRNA inhibitors could be delivered to specific regions of the body, for example the liver, which contains high titer of *Brucella* in the mouse model (*49-51*).

4.2.3 Perspectives on miRNA-like off-target analysis

For the identification of direct targets which underlie the phenotype regulated by a miRNA different approaches can be followed. Bioinformatics predictions are readily available but contain many false positives, might miss relevant non-canonical targets, and from the long list of candidates it can be difficult to identify the one(s) involved in the process. Experimental procedures, such as mRNA sequencing of cells perturbed by the miRNA of interest, will yield a more comprehensive understanding of the cellular changes, but these also include indirect effects. Since such experiments are labor and resource intense, they can currently only be applied for selected miRNAs.

For large RNAi datasets, like the ones that we have generated with the *Brucella* screens, it might be possible to narrow down the number of potential direct targets underlying the effect of certain miRNAs from the long list of bioinformatics predictions. We observed that miR-103/107 (seed: **GCAGC**AU) and miR-1184 (seed: CU**GCAGC**) strongly promote *Brucella* infection and all contain a GCAGC motif within their seed sequence (*research article III*). Two miRNAs with an identical sequence motif within the seed which is shifted by one or several positions could in principle bind to the same site on an mRNA if the nucleotides on the mRNA which flank the sequence motif are complementary to both miRNA seeds (in this case complementary to CU**GCAGC**AU). This potentially allows miR-1184 and miR-103/107 to regulate a partially overlapping set of target genes. Since only a subset of all possible miR-103/107 targets would also meet binding criteria for miR-1184, this information could be used to narrow down

candidates. Such an approach could be applied to all those miRNAs where shifted seed motifs in miRNA mimics and siRNAs are available in sufficiently high numbers to perform off-target analysis and gain confidence that the observed effects are dependent on the seeds of the molecules.

Overall, we found that the effects of miRNA mimics strongly correlate with the mean effects of all siRNAs that share the same seed sequence, provided that the number of siRNA is sufficiently large to average out the on-target effects (*research article II*). This indicates that miRNAs mainly act via their seed sequence in our screens. However, for some miRNA mimics we found that this is not the case (data not shown). Technical problems of individual miRNA mimics or the unlikely case that many siRNAs that share the same seed sequence also target genes that are involved in the phenotype could account for such observations. Alternatively, the miRNA might not regulate the underlying targets of the phenotype via the seed region, but by alternative miRNA-mRNA interactions (*29*, *30*). Therefore, it might be possible to identify miRNAs which affect a specific phenotype (e.g. cell viability or pathogen infection) by non-canonical miRNA-mRNA interactions. Depending on the availability of siRNAs which match different parts of this particular miRNA, it might even be possible to narrow down which part of the miRNA is required to bind the target(s). This could be achieved by searching for siRNAs which share certain conserved motifs with the miRNA and cause the same phenotypic effect upon transfection. In analogy to the seed-analysis performed in *research article II* this could identify non-canonical miRNA-like off-target effects for very specific cases and shed more light on the molecular mechanisms of target recognition of miRNAs which are not seed mediated.

4.3 REFERENCES

1. N. Schultz *et al.*, Off-target effects dominate a large-scale RNAi screen for modulators of the TGF-beta pathway and reveal microRNA regulation of TGFBR2. *Silence* **2**, 3 (2011).
2. F. D. Bushman *et al.*, Host cell factors in HIV replication: meta-analysis of genome-wide studies. *PLoS Pathog* **5**, e1000437 (May, 2009).
3. A. L. Jackson *et al.*, Expression profiling reveals off-target gene regulation by RNAi. *Nat Biotechnol* **21**, 635 (Jun, 2003).
4. B. P. Lewis, I. H. Shih, M. W. Jones-Rhoades, D. P. Bartel, C. B. Burge, Prediction of mammalian microRNA targets. *Cell* **115**, 787 (Dec 26, 2003).
5. R. Konig *et al.*, A probability-based approach for the analysis of large-scale RNAi screens. *Nature methods* **4**, 847 (Oct, 2007).
6. S. Marine, A. Bahl, M. Ferrer, E. Buehler, Common seed analysis to identify off-target effects in siRNA screens. *J Biomol Screen* **17**, 370 (Mar, 2012).
7. R. Zhong *et al.*, Computational detection and suppression of sequence-specific off-target phenotypes from whole genome RNAi screens. *Nucleic Acids Res* **42**, 8214 (2014).
8. E. Buehler *et al.*, siRNA off-target effects in genome-wide screens identify signaling pathway members. *Sci Rep* **2**, 428 (2012).
9. F. D. Sigoillot *et al.*, A bioinformatics method identifies prominent off-targeted transcripts in RNAi screens. *Nature methods* **9**, 363 (Apr, 2012).
10. A. Franceschini *et al.*, Specific inhibition of diverse pathogens in human cells by synthetic microRNA-like oligonucleotides inferred from RNAi screens. *Proc Natl Acad Sci U S A* **111**, 4548 (Mar 25, 2014).
11. E. Buehler, Y. C. Chen, S. Martin, C911: A bench-level control for sequence specific siRNA off-target effects. *PLoS One* **7**, e51942 (2012).
12. R. Kittler *et al.*, Genome-wide resources of endoribonuclease-prepared short interfering RNAs for specific loss-of-function studies. *Nature methods* **4**, 337 (Apr, 2007).
13. A. L. Jackson *et al.*, Widespread siRNA "off-target" transcript silencing mediated by seed region sequence complementarity. *Rna* **12**, 1179 (Jul, 2006).
14. L. Cong *et al.*, Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819 (Feb 15, 2013).
15. P. Mali *et al.*, RNA-guided human genome engineering via Cas9. *Science* **339**, 823 (Feb 15, 2013).
16. T. Li *et al.*, TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. *Nucleic Acids Res* **39**, 359 (Jan, 2011).
17. C. Guzman-Verri *et al.*, GTPases of the Rho subfamily are required for Brucella abortus internalization in nonprofessional phagocytes: direct activation of Cdc42. *The Journal of biological chemistry* **276**, 44435 (Nov 30, 2001).
18. F. Porte, J. P. Liautard, S. Kohler, Early acidification of phagosomes containing Brucella suis is essential for intracellular survival in murine macrophages. *Infection and immunity* **67**, 4041 (Aug, 1999).
19. T. Starr, T. W. Ng, T. D. Wehrly, L. A. Knodler, J. Celli, Brucella intracellular replication requires trafficking through the late endosomal/lysosomal compartment. *Traffic* **9**, 678 (May, 2008).

20. B. Misselwitz *et al.*, RNAi screen of Salmonella invasion shows role of COPI in membrane targeting of cholesterol and Cdc42. *Mol Syst Biol* **7**, 474 (Mar 15, 2011).
21. M. N. Seaman, E. G. Marcusson, J. L. Cereghino, S. D. Emr, Endosome to Golgi retrieval of the vacuolar protein sorting receptor, Vps10p, requires the function of the VPS29, VPS30, and VPS35 gene products. *J Cell Biol* **137**, 79 (Apr 7, 1997).
22. M. N. Seaman, J. M. McCaffery, S. D. Emr, A membrane coat complex essential for endosome-to-Golgi retrograde transport in yeast. *J Cell Biol* **142**, 665 (Aug 10, 1998).
23. P. Temkin *et al.*, SNX27 mediates retromer tubule entry and endosome-to-plasma membrane trafficking of signalling receptors. *Nat Cell Biol* **13**, 715 (Jun, 2011).
24. B. Stechmann *et al.*, Inhibition of retrograde transport protects mice from lethal ricin challenge. *Cell* **141**, 231 (Apr 16, 2010).
25. A. Lipovsky *et al.*, Genome-wide siRNA screen identifies the retromer as a cellular entry factor for human papillomavirus. *Proc Natl Acad Sci U S A* **110**, 7452 (Apr 30, 2013).
26. G. Meister, T. Tuschl, Mechanisms of gene silencing by double-stranded RNA. *Nature* **431**, 343 (Sep 16, 2004).
27. C. Staedel, F. Darfeuille, MicroRNAs and bacterial infection. *Cell Microbiol* **15**, 1496 (Sep, 2013).
28. J. R. Finnerty *et al.*, The miR-15/107 group of microRNA genes: evolutionary biology, cellular functions, and roles in human diseases. *J Mol Biol* **402**, 491 (Sep 24, 2010).
29. S. Yekta, I. H. Shih, D. P. Bartel, MicroRNA-directed cleavage of HOXB8 mRNA. *Science* **304**, 594 (Apr 23, 2004).
30. C. Shin *et al.*, Expanding the microRNA targeting code: functional sites with centered pairing. *Mol Cell* **38**, 789 (Jun 25, 2010).
31. A. Grimson *et al.*, MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell* **27**, 91 (Jul 6, 2007).
32. D. Gaidatzis, E. van Nimwegen, J. Hausser, M. Zavolan, Inference of miRNA targets using evolutionary conservation and pathway analysis. *BMC Bioinformatics* **8**, 69 (2007).
33. L. P. Lim *et al.*, Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* **433**, 769 (Feb 17, 2005).
34. D. Baek *et al.*, The impact of microRNAs on protein output. *Nature* **455**, 64 (Sep 4, 2008).
35. F. Huang, Y. G. Chen, Regulation of TGF-beta receptor activity. *Cell Biosci* **2**, 9 (2012).
36. J. J. Lee *et al.*, Interplay between clathrin and Rab5 controls the early phagocytic trafficking and intracellular survival of Brucella abortus within HeLa cells. *J Biol Chem* **288**, 28049 (Sep 27, 2013).
37. P. Cossart, A. Helenius, Endocytosis of viruses and bacteria. *Cold Spring Harb Perspect Biol* **6**, (Aug, 2014).
38. H. Damke, T. Baba, A. M. van der Bliek, S. L. Schmid, Clathrin-independent pinocytosis is induced in cells overexpressing a temperature-sensitive mutant of dynamin. *J Cell Biol* **131**, 69 (Oct, 1995).
39. Y. Takahashi *et al.*, MiR-107 and MiR-185 can induce cell cycle arrest in human non small cell lung cancer cell lines. *PLoS One* **4**, e6677 (2009).
40. P. S. Linsley *et al.*, Transcripts targeted by the microRNA-16 family cooperatively regulate cell cycle progression. *Mol Cell Biol* **27**, 2240 (Mar, 2007).
41. Q. Liu *et al.*, miR-16 family induces cell cycle arrest by regulating multiple cell cycle genes. *Nucleic Acids Res* **36**, 5391 (Sep, 2008).

42. N. Bandi *et al.*, miR-15a and miR-16 are implicated in cell cycle regulation in a Rb-dependent manner and are frequently deleted or down-regulated in non-small cell lung cancer. *Cancer research* **69**, 5553 (Jul 1, 2009).
43. M. Lerner *et al.*, DLEU2, frequently deleted in malignancy, functions as a critical host gene of the cell cycle inhibitory microRNAs miR-15a and miR-16-1. *Experimental cell research* **315**, 2941 (Oct 15, 2009).
44. C. Y. Chang *et al.*, NFIB is a governor of epithelial-melanocyte stem cell behaviour in a shared niche. *Nature* **495**, 98 (Mar 7, 2013).
45. S. Mason, M. Piper, R. M. Gronostajski, L. J. Richards, Nuclear factor one transcription factors in CNS development. *Mol Neurobiol* **39**, 10 (Feb, 2009).
46. G. M. Rossolini, F. Arena, P. Pecile, S. Pollini, Update on the antibiotic resistance crisis. *Curr Opin Pharmacol* **18C**, 56 (Sep 22, 2014).
47. K. von Bargen, J. P. Gorvel, S. P. Salcedo, Internal affairs: investigating the Brucella intracellular lifestyle. *FEMS microbiology reviews* **36**, 533 (May, 2012).
48. E. Billard, C. Cazevieille, J. Dornand, A. Gross, High susceptibility of human dendritic cells to invasion by the intracellular pathogens Brucella suis, B. abortus, and B. melitensis. *Infect Immun* **73**, 8418 (Dec, 2005).
49. M. Jayaraman *et al.*, Maximizing the potency of siRNA lipid nanoparticles for hepatic gene silencing in vivo. *Angew Chem Int Ed Engl* **51**, 8529 (Aug 20, 2012).
50. K. J. Rayner *et al.*, Antagonism of miR-33 in mice promotes reverse cholesterol transport and regression of atherosclerosis. *The Journal of clinical investigation* **121**, 2921 (Jul, 2011).
51. K. J. Rayner *et al.*, Inhibition of miR-33a/b in non-human primates raises plasma HDL and lowers VLDL triglycerides. *Nature* **478**, 404 (Oct 20, 2011).

5 Acknowledgements

5 ACKNOWLEDGEMENTS

First and foremost I would like to thank my supervisor Prof. Christoph Dehio. I am very grateful for his constant support throughput the project and very much appreciated that he allowed me to explore own ideas. Furthermore, I would also express my gratitude for the many opportunities he gave me to broaden my horizon, inside and outside science.

I would also like to thanks my committee members Prof. Dirk Bumann, Prof. Mike Hall, and Prof. Mihalea Zavolan. The input on many aspects of my work was extremely valuable and helped significantly to advance the projects.

A big thank you goes to Shyan H. Low. She was such a great teammate when we spent all those hours screening. We had so many fun moments together and even though I'm not very skilled in learning Chinese, she managed to teach me some of the essentials, thanks!

I would like to thank Mario Emmenlauer who greatly contributed to the success of the project and has become a great friend.

I very much enjoyed the work together with Pauli Rämö. He was involved in many aspects of this work. I'm very grateful for everything that I could learn from him. Besides science, also his yoga lessons had a great positive impact on my body-mind balance which goes on hanks to Simone and Lisa.

I'm very happy to have been able to spend many fun moments with Simone, Kathrin, Shyan, Lu, and Claudia joking in the office, climbing a mountain or enjoying great food.

Furthermore, I would like to thank the whole group for a truly great working atmosphere, help, support, fun, countless great moments and plenty of sugar: Matthias Truttmann, Kathrin Pieles, Shyan Huey Low, Yun-Yueh Lu, Rusudan Okujava, Claudia Mistl, Houchaima-Ben-Tekaya, Alexander Harms, Simone Eicher, Frederic Stanger, Arnaud Goepfert, Isabel Sorg, Anne-Cecile Hiebel, Maxime Quebatte, Mario Emmenlauer, Pauli Rämö, Sabrina Siamer, Simon Marlaire, Therese Tschon, Christoph Schmutz, Jonas Körner, Sarah Stiegeler, Damian Murezzan, Tjaart De Beer, Kevin Charles Smith, Nikolaus Schmitz, Clement Barbier, Caroline Grey Hackett.

I thank Simone, Kathrin, Caroline, and Houchaima for careful reading of the thesis.

I thank Nitish Mittal and Andreas Gruber for their help with transcriptomics analyses, Alex Schmidt for proteomics support and, Christoph Kasper who often came in rescue when we needed help with lab automation, Claudia Erbel-Sieler, Mihaela Hanisch, Angelika Klarer, Rachel Benzies, and Anne-Cécile Hiebel for organizational support, Andi Hefti for keeping my computers up to date, Marina Kuhn Rüfenacht, Roger Sauder, Patric Haenni and Leonardo Faletti to keep machines and orderings running.

I would like to thank everyone working in the BSL3 at the Swiss THP. I very much appreciate the friendly and helpful atmosphere. Special thank goes to Sonia Borrell for her great organization of the lab and keeping everyone safe and to Julia Feldmann whose support is of great help for the daily working routine.

I am very happy to have meet Lisa, Nura, Max, Ricardo, Roland and Julia. I enjoyed our time together during lunch breaks, hanging out on the river, or climbing a wall. Lisa, many thanks for the great Rhein-lunch-swimming session, refreshing coffee breaks, and of course the pipetting monkeys.

I am grateful for funding of the first three years of my PhD by the “Fellowship for Excellence”.

I thank my flat mates, especially Fabia, for everything. During busy times, I might have starved without her...

I would also like thank my family for their love and support.

Many thanks to nature for keeping me fascinated every day.

6 Curriculum vitae

6 CURRICULUM VITAE

Personal data

Alain Casanova, M. Sc.

Education

11/2010 – 11/2014	**PhD thesis** Focal Area of Infection Biology, Biozentrum, University of Basel Laboratory: Prof. Dr. Christoph Dehio Fellowship: "Fellowship for Excellence"
07/2009 – 06/2010	**Scientific associate** (Civil Service) National TB Reference Laboratories at NICD Johannesburg, South Africa Martin P. Grobusch, M.D. and Biozentrum, University of Basel Laboratory: Prof. Dr. Jean Pieters
2007 –2009	**Master of Science in Molecular Biology**, grade 6 (6 = best) Major in Molecular Microbiology Focal Area Infection Biology, Biozentrum, University of Basel Laboratory: Prof. Dr. Urs Jenal
2004 – 2007	**Bachelor of Science in Biology**, grade 5.2 (6 = best) Major in Molecular Biology University of Basel
2000 - 2003	**Grammar school** Gymnasium Liestal, BL

Publications and Scientific Contributions

Publications

Franceschini A*, Meier R*, **Casanova A***, Kreibich S*, Daga N, Andritschke D, Dilling S, Ramo P, Emmenlauer M, Kaufmann A, Conde-Alvarez R, Low SH, Pelkmans L, Helenius A, Hardt WD, Dehio C, von Mering C (2014) Specific inhibition of diverse pathogens in human cells by synthetic microRNA-like oligonucleotides inferred from RNAi screens. *Proc Natl Acad Sci U S A* **111:** 4548-4553

Boehm A, Steiner S, Zaehringer F, **Casanova A**, Hamburger F, Ritz D, Keck W, Ackermann M, Schirmer T, Jenal U (2009) Second messenger signalling governs Escherichia coli biofilm induction upon ribosomal stress. *Mol Microbiol* **72:** 1500-1516

Scientific Talks

Biozentrum Symposium, CH (2014)
Systems Biology of Infection Symposium, CH (2013)
PhD Retreat Biozentrum, CH (2013)
PhD Retreat Biozentrum, CH (2012)