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

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

[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

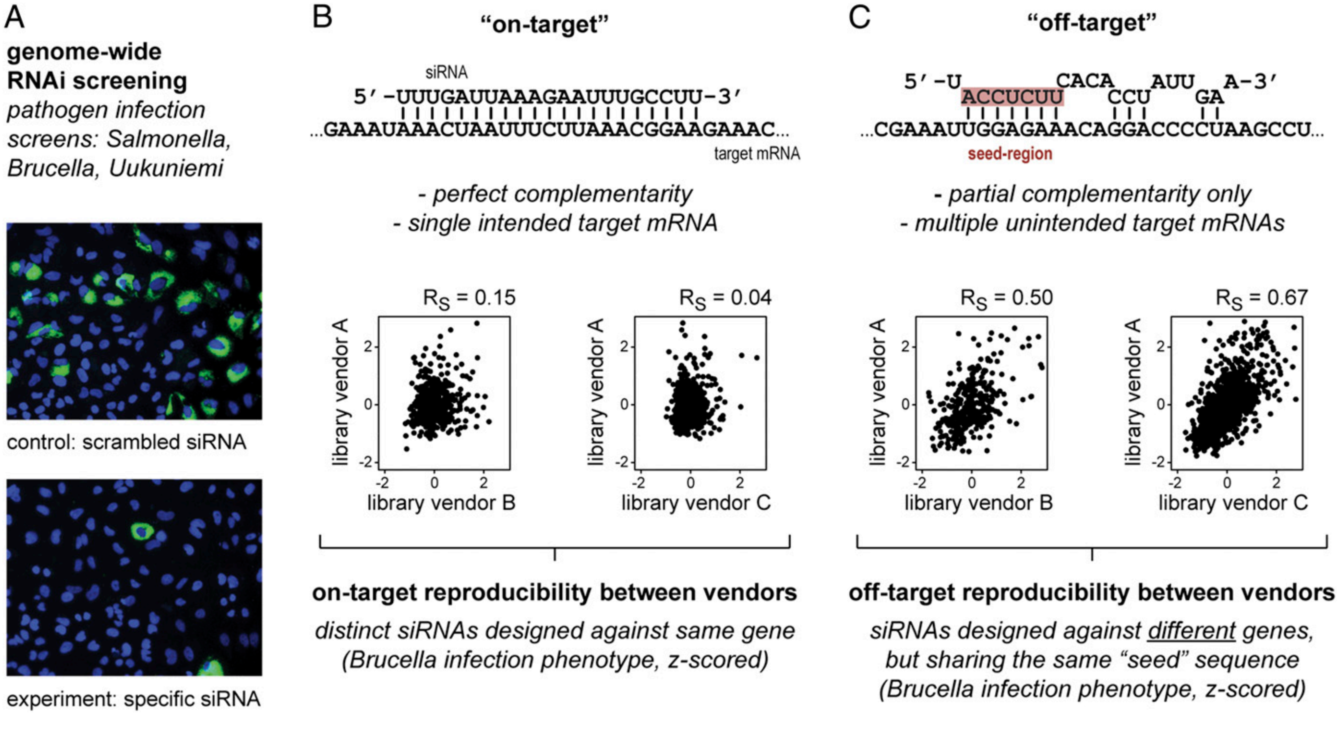


Fig. 1. Off-target effects in RNAi pathogen infection screens. (*A*) Experimental setup. HeLa cells were screened for host factors required for pathogen entry. Microscopy images from two separate wells of a typical perturbation experiment are shown (DAPI-stained HeLa cell nuclei in blue; successful pathogen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

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).

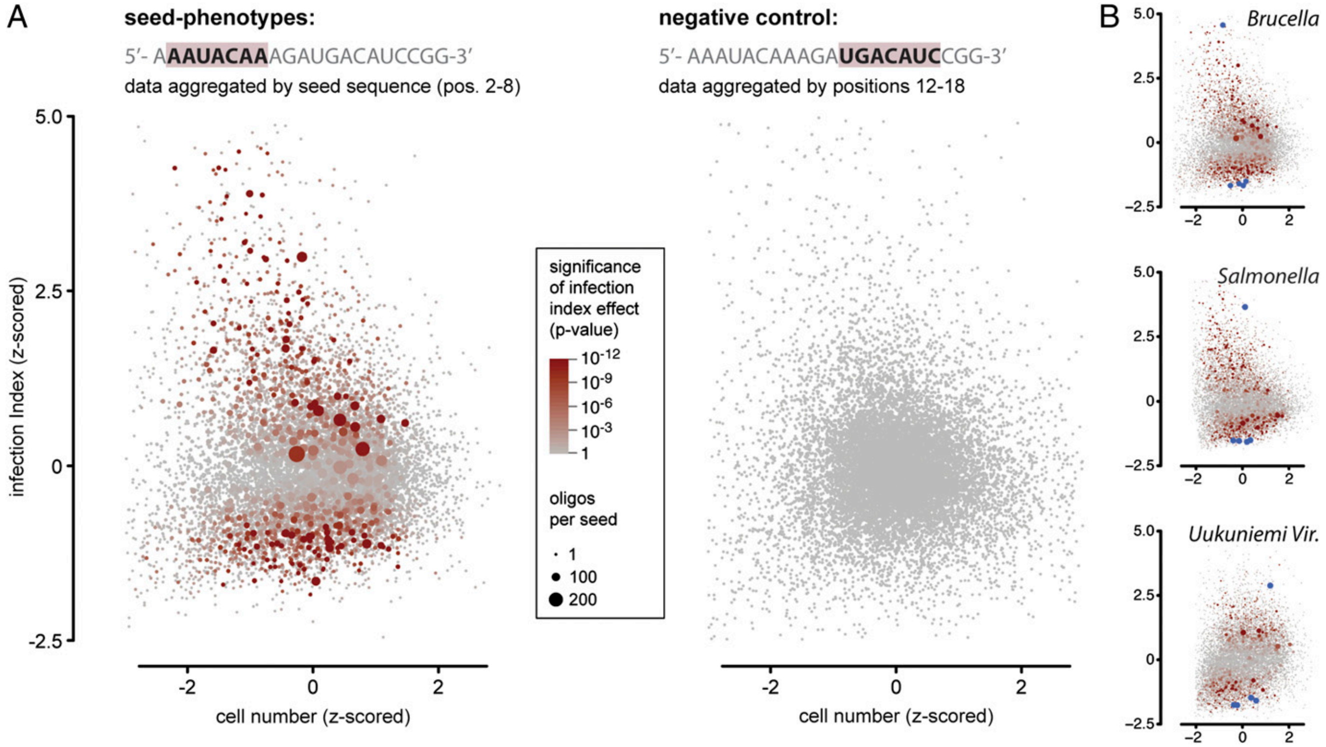


Fig. 2. Genome-wide screening data aggregated by shared seed sequences. (*A*) Visualization of the entire genome-wide data of the infection screen for *B. abortus*, aggregated by the seed sequences found in the various siRNA oligos. Each data point represents one heptameric seed sequence, showing the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

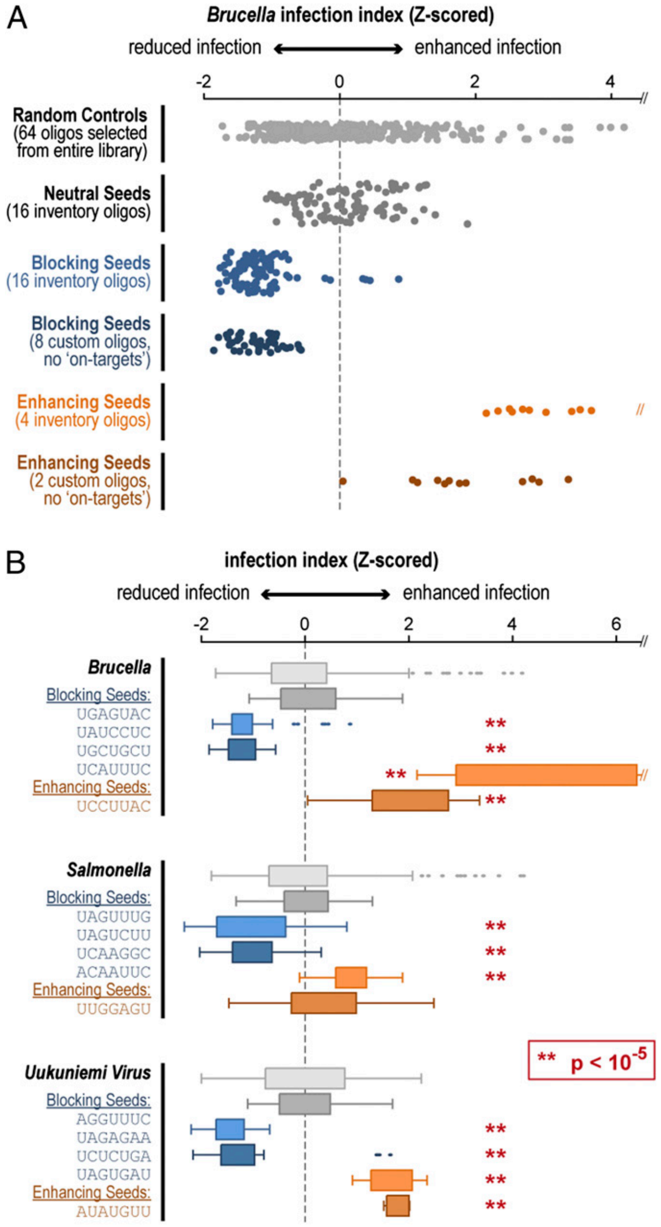
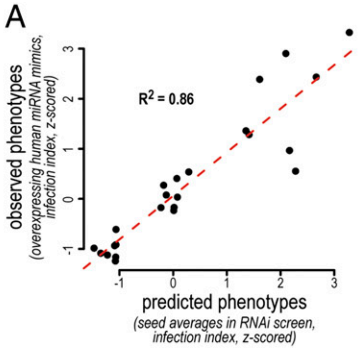


Fig. 3. Experimental confirmation of predicted seed phenotypes. (*A*) Detailed phenotypes measured for *B. abortus* (six replicates per oligo). (*B*) Summary of phenotypes measured for each of the three pathogens. The siRNA oligos predicted to block infection are shown in blue (dark blue for those that 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.

MICROBIOLOGY



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

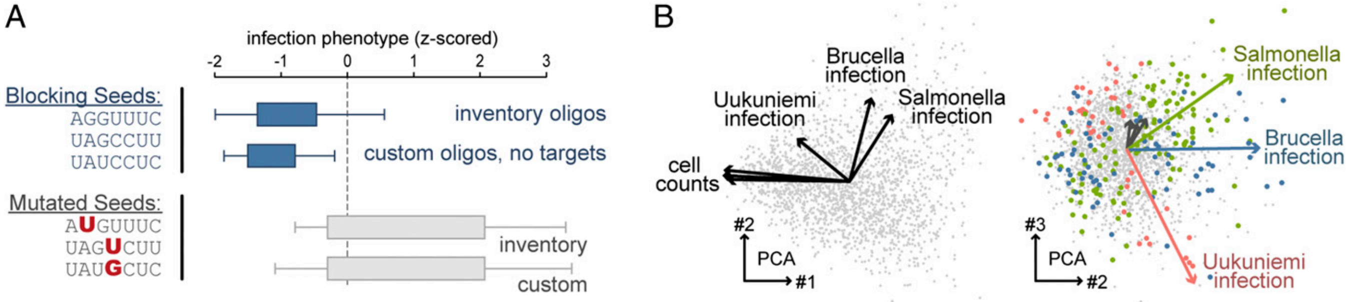


Fig. 5. Specificity of seed effects. (*A*) Effects of single-point mutations located in the seed regions. For each of the three pathogens, one seed was chosen that was predicted to block infection (data shown in blue). Shown in gray are data for the corresponding seeds that have been mutated at one position.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