PREDICTIVE SYSTEMS BIOLOGY APPROACH TO BROAD-SPECTRUM, HOST-DIRECTED DRUG TARGET DISCOVERY IN INFECTIOUS DISEASES

RAMON M. FELCIANO[†*], SINA BAVARI[‡], DANIEL R. RICHARDS[†], JEAN-NOEL BILLAUD[†], TRAVIS WARREN[‡], REKHA PANCHAL[‡], ANDREAS KRÄMER[†]

[†]Ingenuity Systems, Inc
1700 Seaport Blvd, 3rd Floor
Redwood City, CA 94107
email: felciano@ingenuity.com
** Corresponding author*

[‡]USAMRIID
1425 Porter Street
Fort Detrick, MD 21702
email: sina.bavari@us.army.mil

Knowledge of immune system and host-pathogen pathways can inform development of targeted therapies and molecular diagnostics based on a mechanistic understanding of disease pathogenesis and the host response. We investigated the feasibility of rapid target discovery for novel broad-spectrum molecular therapeutics through comprehensive systems biology modeling and analysis of pathogen and host-response pathways and mechanisms. We developed a system to identify and prioritize candidate host targets based on strength of mechanistic evidence characterizing the role of the target in pathogenesis and tractability desiderata that include optimal delivery of new indications through potential repurposing of existing compounds or therapeutics. Empirical validation of predicted targets in cellular and mouse model systems documented an effective target prediction rate of 34%, suggesting that such computational discovery approaches should be part of target discovery efforts in operational clinical or biodefense research initiatives. We describe our target discovery methodology, technical implementation, and experimental results. Our work demonstrates the potential for *in silico* pathway models to enable rapid, systematic identification and prioritization of novel targets against existing or emerging biological threats, thus accelerating drug discovery and medical countermeasures research.

1. Background

New and reemerging infectious diseases pose a growing global health risk across public health concerns and potential bioterrorism threats. Pandemic viruses, resistant bacteria, and technology improvements in bioengineering point to a need for accelerated drug discovery[1]. One approach to this challenge is to use computational techniques to efficiently identify drug targets that may effectively mount a defense against one or more biothreats[2]. Biologically diverse pathogens share common or similar mechanism of infection and pathogenesis, and the host has similarly conserved immune response biology[3–5].

We have previously demonstrated the broad applicability of systems biology analyses to drug discovery and development focused on mammalian disease biology[8–10]. We hypothesize that similar computational characterization of pathogen biology, pathogenesis and host-response genomic pathways across multiple infectious agents can enable systematic identification of targets of intervention that will impact multiple pathogens in a similar manner, and thus serve as *broad-spectrum drug targets* that can be modulated by novel or

repurposed therapeutic modalities[6,7]. To test this hypothesis, we extended our approach to identify and predict host-based pathway mechanisms that, once validated, would have a beneficial therapeutic effect against a given pathogen. Validated host pathways and targets can then form the basis of drug repurposing studies, for example to identify compounds previously approved for other disease indications but that share a host mechanism leveraged by a pathogen of interest. We developed computational drug target identification extensions to Ingenuity's pre-existing systems biology platform, and performed a pilot study to experimentally validate predicted targets against six representative "pilot pathogens": Ebola virus, Marburg virus, Lassa virus, *Yersinia pestis, Francisella tularensis*, and *Bacillus anthracis*.

2. Methods

2.1. *Overview of our drug target discovery approach*

Our approach (Figure 1) centers on computer-based modeling of disease pathways using semantic technology, scientific knowledge bases (KBs) of mammalian biochemistry, and

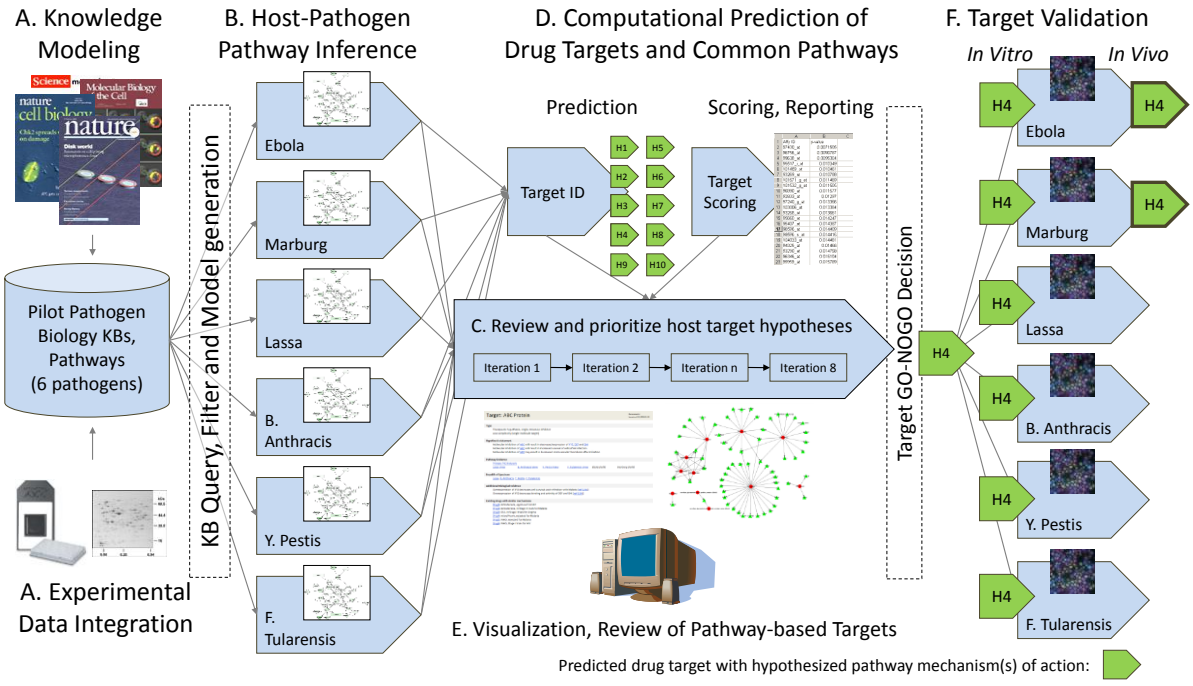

Figure 1. Overview of Ingenuity-USAMRIID predictive systems biology pilot, including knowledge base (KB) construction (A) and host-pathogen pathway model inference (B) for 6 pilot pathogens; multiple rounds ("iterations") of *in silico* target prediction (C) based on suite of target ID algorithms (D); expert review and prioritization of targets using our system prototype (E); and final target selections for *in vitro* and *in* vivo validation at USAMRIID (F). KBs are updated between each iteration. PIC = pathway intervention candidate, i.e. a proposed target centered around the perturbation of a specific pathway of interest.

bioinformatics tools developed by Ingenuity for drug discovery and development and extended herein[10]. We extended existing pathway models of disease biology to bacterial and viral pathogenesis, and developing large-scale, semantically-integrated, knowledge-based models of six pathogens (Ebola virus, Marburg virus, Lassa virus, *Yersinia pestis, Francisella tularensis*, and *Bacillus anthracis*). Specific technology extensions include extending host biomedical ontologies and knowledge models to pathogen biochemistry, pathogenesis staging, and infectious disease; curation and modeling of pathogen-specific pathway content; developing several broad-spectrum target prediction algorithms and target evaluation protocols; and augmenting IPA[11] pathway visualization, filtering and scientific workflows to enable collaborative, team-based broad-spectrum target identification and validation. These extensions, collectively referred to as Pathogen-IPA (P-IPA), were developed as proof-of-concept to demonstrate the feasibility of using computer-based pathway models to accelerate drug target discovery.

2.2. *Knowledge models for target hypothesis generation*

Central to our approach is the notion of *computational hypothesis generation*[12,13], yielding one or more formally-defined "target hypotheses" that relate (1) a host gene or protein and (2) a particular positive or negative impact a drug may have on that target (i.e. "activate" or "inhibit"), and (3) a positive therapeutic effect on one or more clinically-relevant endpoint in hosts infected by each of at least two pathogens. An example of a target hypothesis, rendered computationally to English, is "We hypothesize that inhibition of LAMP2 will counteract the effects of *B. anthracis* and *F. tularensis* (as measured by bacterial uptake studies)". We used P-IPA to computationally characterize the pathogen biology, mechanisms of pathogenesis, and host-response pathways for our 6 pilot pathogens, and use these models to identify and validate one or more such host targets hypotheses.

Table 1. Examples of contextualized pathway findings in our causal reasoning networks, rendered to English syntax through the use of Natural Language Generation algorithms.

Example context	Example of host-pathogen finding(s) in P-IPA causal networks
Attenuated	• Attenuated live *F. tularensis* increases proliferation of human lymphocytes in culture 10-11 months post-treatment.
Virulent	• Decrease of mouse CD45 increases survival of murine-adapted mouse after infection by virulent Ebola virus.
Virulent	• A mutant protein fragment (1-254) (H86K with its Zinc finger domain mutated) from human ZAP protein in Rat2 embryo cells decreases viral replication of Sudan ebolavirus.
Killed or inactivated	• In human neutrophils, killed Marburg virus increases upregulation of human Tlr protein(s) 1 hour post-treatment
Therapeutic (includes vaccine, antiviral, antibacterial	• Oral administration of Salmonella typhimurium-based vector vaccine composed of *Y. pestis* F1 [caf1] protein and of *Y. pestis* V antigen protein increases (by 83 percent) survival of mouse that involves subcutaneous injection of *Y. pestis*.

To generate target hypotheses, we built a global network of causal pathway relationships derived from the Ingenuity Knowledge Base (IKB), a large-scale, manually-curated, semantically-structured ontology-based knowledge base of disease biology research findings[14]. A "finding" is single biochemical insight derived from an original experiment, as supported by primary research or review articles, and tied to a specific biomedical investigation and experimental context. The underlying knowledge representation has semantics based on RDFS[15,16], with pathway models similar to BioPAX Level 3 and SBGN[17], and extensions for modeling drugs, vaccines, biomarkers and clinical phenotypes. We extended IKB with 535,599 new findings curated from primary research, focused on host-pathogen interactions for our 6 pathogens, that increasing the IKB size by 5.1% (Table 1).

Updates to IKB findings and pathway models are ongoing. On a weekly basis a series of knowledge transformations post-process IKB findings to generate (infer) causal networks and other data structures optimized for specific algorithmic approaches (Figure 2), similar to [18] but using semantic rather than linguistic dependency graphs. We infer a causal network where nodes represent form-, species- and state-specific molecules: DNA, RNA, protein, complexes, or pathogen particles, including strain-specific forms. Directional edges represent causal dependencies between the biological activity of linked nodes. These cause-effect relationships include gene regulation, activation / inhibition, chemical modification and other interactions, as supported by one or more experimentally-demonstrated findings from IKB. Such findings are classified by implied direction of change (DOC) of the associated

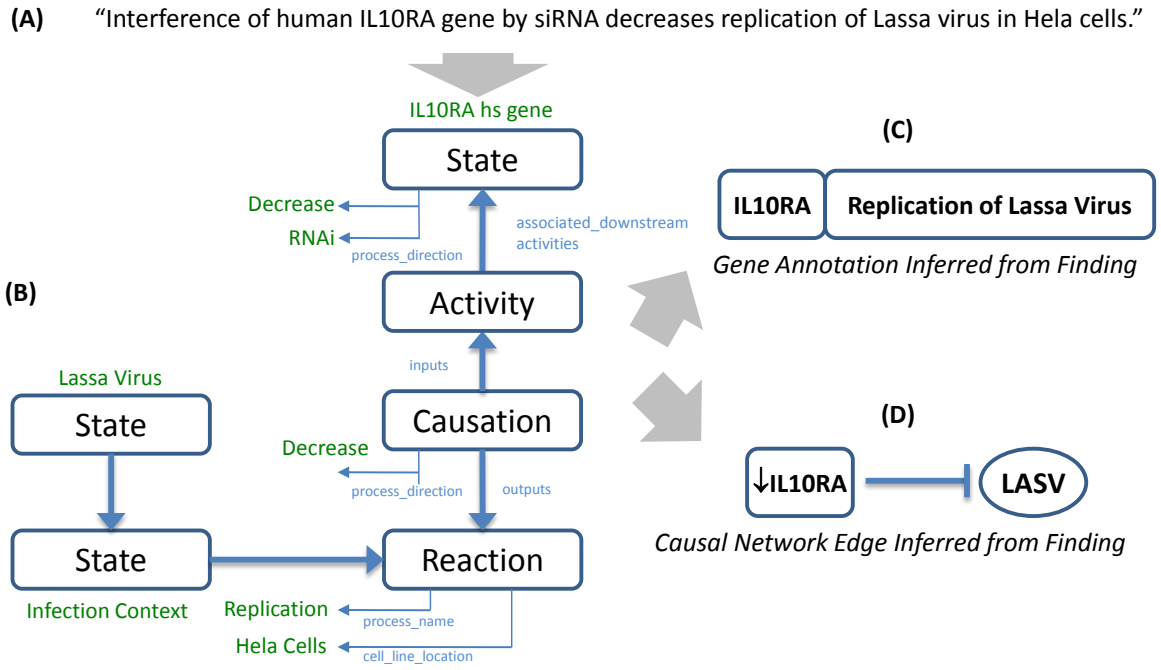

Figure 2. Example causal finding used in our predictive analytics. This example illustrates how a single experimental observation (A) is modeled as a semantic network of interrelated concepts (B), which can then be further transformed into a number of secondary data structures useful for computation, such as gene annotations (C) and causal network relationships (D).

causal effect (*increase*, *decrease*, *affects* or *no-effect*). For example, the finding "In human neutrophils, killed Marburg virus increases upregulation of human Tlr protein(s) 1 hour after initial treatment" (see Table 1) would result in a positive causal regulatory relationship between the pathogen (Marburg virus) and host (Tlr protein). Conflicts are resolved by preferentially assigning a DOC if >85% of findings support it, or a non-directional *affects* annotation that must be manually inspected to resolve the conflict.

2.3. *Predictive algorithms for drug target identification*

We identified several *target identification strategies*, each motivated by a specific aspect of pathogenesis that could form basis of a therapeutic strategy and formalized algorithmically to explore the associated hypothesis space using models of host pathobiology pathways. Based on this analysis we developed a general framework for hypothesis generation algorithms, and implemented two complementary approaches for identifying candidate broad-spectrum therapeutic targets, as described in [19] (see supporting materials).

First, we observed that individual host proteins may be regulated in similar ways by multiple pathogens, suggesting an important shared regulatory influence by the pathogen on host proteins. Reversing this regulatory effect may thus therapeutically benefit the host. Our *Commonalities algorithm* seeks to reverse the polarity of multiple pathogens' similar, direct regulatory effect on a single common host protein, hopefully countering the associated pathogenic impact.

We further observed that multiple host proteins may be similarly regulated by a given pathogen. Rather than pursue a complex "drug cocktail" to target multiple components of this genomic signature, we hypothesize that such panels of host markers may share common upstream regulatory partners. Our second *Upstream Regulators algorithm* thus seeks to identify optimal targets that are upstream of directly affected host molecules, and can serve as a single target more easily modulated by a novel or repurposed drug.

Every target hypothesis generated by these algorithms is supported by a (proposed) pathway mechanism that aggregates immunological evidence and a logical rationale for selecting the target. Hypotheses were further cross-referenced and annotated existing drugs that are either FDA-approved or in various stages of clinical trials for other indications[14,20,21]. Availability of compounds against a protein target was not used to generate hypotheses, but served as a "tie breaker" between otherwise biologically compelling targets when prioritizing our final target list for experimental validation.

2.4. *Experimental design for target validation studies*

To assess the effectiveness of our approach, we performed two-phase *in vitro* and *in vivo* validation studies against our predicted host targets. All validation studies were performed by the Bavari lab at the USAMRIID research facilities, using established protocols for working with our pilot pathogens.

For viral *in vitro* studies, Hela cells were selected as a well-established infection model. Two main experimental approaches were used for validating targets against Ebola, Marburg

and Lassa: high content image (HCI) analysis and quantitative real time-PCR (qRT-PCR). Both these assays measure viral replication as the relevant biological endpoint. Inhibition or activation of each targets were achieved by transfection of specific siRNA or transfection of cDNA specific to that target, respectively. For bacterial studies we used three specific types of assays: (1) phagocytosis/bacterial uptake, a HCI assay that measures phagocytosis/ bacterial uptake by the macrophages; (2) fluorescent antibodies specific to pathogen protein(s) used to detect the pathogen that has attached to (and thus phagocytosis by) the host cell; and (3) a Live/Dead assay that measures cytotoxicity.

In vivo studies were designed to further validate inhibition-based targets at the USAMRIID research facilities, based on protocols previously designed in the Bavari lab. To knock down target expression, we used antisense phosphomorpholino oligonucleotides (PMO) inhibition technology (GeneTools, LLC. , Philomath, Oregon). Groups of 10 mice were used: one group per target received target-specific PMOs, and a control group receiving either standard non-specific PMOs, or phosphate buffered saline. All animals received PMOs intraperitoneal (i.p.) or intranasal (i.n.) 4 times (-24h, -4h, 24h, 48h) at 100 to 150 μg per injection per mouse. Mice were challenged i.p. at day 0 with the corresponding lethal dose. For one set of *F. tularensis* experiments, the bacterial challenge was performed using intranasal administration to evaluate survival/protection using a different route of infection.

3. Results

3.1. *Target prediction and prioritization*

We used P-IPA to generate a target pipeline of 490 host proteins whose activation or inhibition was predicted to have a beneficial therapeutic impact against at least two of our 6 pilot pathogens. Through iterative review and filtering using the P-IPA tool suite we prioritized this pipeline to identify the most promising targets and select them for target validation. Target hypotheses were reviewed and prioritized in P-IPA based on:

(a) *Broad-spectrum potential.* Selected host targets must be predicted to impair at least 2 of the 6 pilot pathogens.

(b) *Contextual consistency of pathway evidence.* Targets must be supported by a pathway mechanism consistent with existing research data as well as the clinically relevant disease context (e.g. virulent rather than attenuated pathogen strains)

(c) *De novo experimental evidence.* As special case of (b), we re-integrated our *in vitro* experimental results into IKB as "new but unpublished findings" to facilitate *in vivo* target prioritization,.

(d) *Availability of animal models.* Targets must be testable in a mouse system used by a reference animal model for 5 of our pathogens (Ebola, Marburg, *B. anthracis, F. tularensis,* and *Y. pestis*). To the best of our knowledge, there are no well-validated mouse models for Lassa virus.

(e) *Clinically-relevant endpoints.* Target validity should be confirmed against clinically-relevant endpoints (e.g. improved host survival, reduced viral load, etc).

(f) *Operational tractability.* Host targets were tested using of antisense-based intervention across all experiments evaluating loss-of-function or inhibition-based targets, as permitted by schedule and budget constraints that determined the total number of targets we could test.

We selected 28 target hypotheses (16 inhibition-based targets and 14 activation targets) for Phase 1 *in vitro* validation. In Phase 2, 12 targets were selected for *in vivo* testing, including 8 inhibition targets validated *in vitro* (DUSP1, HSP90B1, LAMP1, SERPIN5, SERPINE2, SMAD3, AP3D1, IL10RA), and 4 new targets selected based on new curated findings highlighted in updated prediction runs (BTRC, HGS, PDCD6IP, PPARA).

3.2. *Example broad-spectrum pathway hypothesis and host drug target*

By way of illustrating our approach, we describe one target prediction in detail (Figure 3). Pathogens may similarly activate or inhibit the function several host proteins. Rather than target these commonly-regulated host proteins individually, the *Upstream Regulator algorithm* treats them as protein signature, and tries to identify a single, additional, host protein that could counter or reverse the impact of the pathogen's effect on this signature. In this example, Ebola and Marburg viruses have been reported to inhibit a number of common host proteins (F2, PROC, PLAU, KLKB1, and C1S). IKB findings (and their underlying research publications) further indicate that SERPINE2 represses the activity of the same proteins. Thus, the algorithmically-generated hypothesis is that both viruses build upon the naturally-occurring suppressive effect of SERPINE2 in the host, and that by removing this effect, we may effectively "pull the rug out" from these viruses and potentially slow pathogenesis by making them work harder. Significantly, our hypothesis re-uses findings from cancer and cardiovascular molecular studies that characterize SERPINE2's effect on the other host proteins include results, as SERPINE2 was previously unassociated with viral hemorrhagic fever infection.

3.3. *Classification of broad-spectrum target validation results*

We formalized our performance evaluation developing a classification framework for target validity that partitioning targets based on whether our experimentation demonstrated a desired effect or lack of effect, and whether that effect was deemed to be clearly demonstrated or whether additional studies were needed to confirm the effect. We used a 5 category scale: *clearly-validated*, *possibly-validated*, *not-tested*, *possibly-not-validated*, and *clearly-not-validated*. For *in vitro* assays, we use 30% reduction in viral load or bacterial uptake as a baseline threshold for a *clearly-validated* classification, adjusted to pathogen-specific thresholds if they exist for a specific virus or bacterium. For *in vivo* assays, target validity was defined as a minimum level of protection conveyed to infected mice, consistent with screening practices. Our baseline threshold was >40% survival in mice after 9 to 22 days (depending on the pathogen) and twice (2x) the standard control survival rate, replicated twice with 10+ mice per experiment. Two other target categories—*possibly-validated*, *possibly-not-validated*—demonstrated lesser phenotypic effect or were not

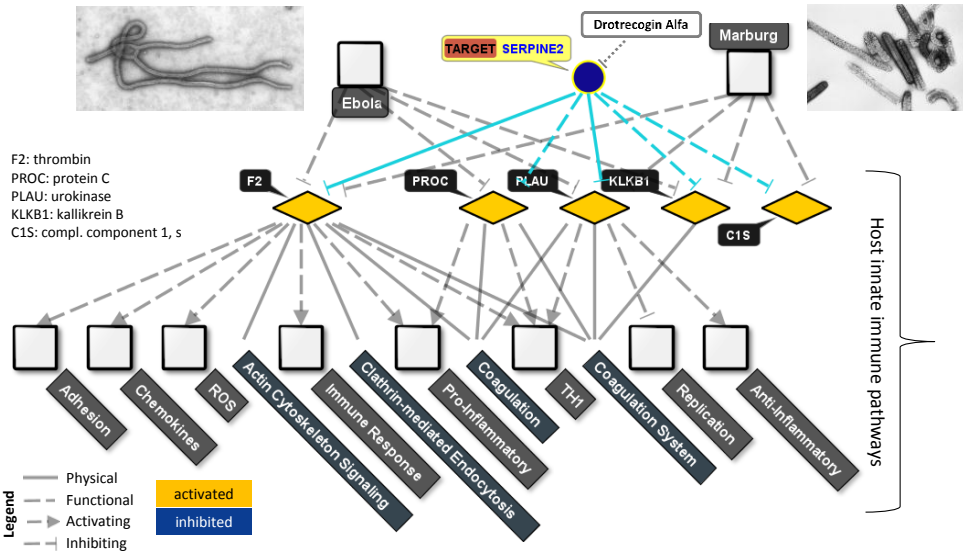

Figure 3. Example of target (SERPINE2, blue node) hypothesis identified by the upstream regulators algorithm as playing a common role in pathogenesis of Ebola virus and Marburg virus, and a drug (Drotrecogin Alfa / Xigris™, Elli Lily)) that may be repositioning for this indication. This drug target hypothesis is grounded in signature of host proteins (yellow nodes) that are commonly downregulated by Ebola and Marburg infection. SERPINE2 is further linked to relevant immune functions, including ones found in viral hemorrhagic fever infection (e.g. coagulation pathways). SERPINE2 was validated *in vitro* and *in vivo* to have the predicted effect on systems infected by the Ebola and Marburg viruses.

Figure 4. Pipeline of prioritized inhibition-based target hypotheses, with our 16 initially-selected inhibition-based targets. *In vitro* and *in vivo* validation results color code the hypothesis arrows based on success or failure classification. For example, the top-left target is AP3D1, which was predicted to have a beneficial effect under Ebola and Marburg infection if the target was inhibited, shown as two down arrows. Knock-down *in vitro* screens and *in vivo* studies confirmed these predictions (filled circles, green). Pipeline visualization is interactive and updated dynamically as new target hypotheses and validation results are integrated.

replicated across multiple experiments, thus requiring additional study to conclusively rule them in or out as drug targets. Commercial availability or maturity of a given compound through the FDA approval process was presented but not used as a validation criterion.

3.4. *Validation of drug target predictions*

We analyzed the performance of our method using both *in vitro* and *in vivo* experimental data by aggregating, discretizing and classifying this hypothesis-specific target validation data into the classifications, described in section 3.3. Briefly, in vitro validation experiments in Phase I demonstrated that 24 of 28 predicted targets resulted in hits against at least one pilot pathogen, Moreover, 22 hits are broad-spectrum (2 + pathogens) target candidates. For example, SERPINB5 showed clear or partial impact against 4 of our 6 pathogens. From this panel of prioritized targets, 11 of these 12 tested targets showed effect against at least 1 pathogen in mice, and 5 clearly inhibit 2+ pathogens (broad-spectrum). Additional targets showed promise, but require additional work to confirm. Inhibition-based targets in Figure 4 have the greatest potential for drug repurposing with compound inhibitors.

4. DISCUSSION

Based on this analysis, 34% directly predicted targets we tested were validated in mouse models, which we believe to a very promising yield. This lower bound (34% for *in vivo*) is a conservative performance assessment, treating only *clearly-valid* results as successes. Performance increases if one includes targets that showed some promising effect but not sufficient to meet our threshold, although this requires additional experimentation to confirm. Table 2 summarizes our findings as predictive success rate, across activation- and inhibition-hypotheses and *in vitro* and *in vivo* results. SERPINB5 is our strongest validated target, clearly validated against *B. Anthracis*, Ebola virus and Marburg virus, and may further show impact against *F. Tularensis* and *Y. Pestis*, although further studies may be required to optimize dosing to confirm this. As our top-ranking target, we believe SERPINB5 is worthy of further investigation to assess mechanism of action.

Table 2. Topline performance of computational target predictions based on *in vitro* and *in vivo* experimental results, across all prioritized, tested hypotheses

Success rate	N (# tested target hypotheses)	Lower bound (*clearly-validated*)	Upper bound (*clearly-validated* + *possibly-validated*)
In vitro	81	27%	46%
In vivo	32	34%	50%

The measured endpoint across these experiments was percentage survival post-infection and treatment. Specifically, we measure the number of mice (out of a total of 10 per group) that survived following PMO treatment and challenge with the corresponding pathogen. For example, 50% survival rate indicates that 5 of 10 mice survived after treatment. In addition

to percentage survival, we factored in the number of independent experiments performed, the number of replicates for a sample test, the difference relative to baseline threshold from the standard control, and non-measurable expert evaluation for a given sample. In some cases we were not able to perform identical replicate experiments for a given pathogen.

Interestingly, *in vivo* results out-performed *in vitro* (34% vs. 27%), which may be attributable the limited applicability of cellular assays for modeling host immune biology, as well as the overall lower number of tests run in animal studies relative to our in vitro studies. In addition, the kinetics of each *in vivo* experiment is dependent on each pathogen, and we occasionally observed off-target effects with scrambled PMOs that enabled some increased survival on its own and which we could not control for. This suggests the need for additional research into effective, low-cost alternatives to animal and clinical studies for drug target validation studies[22].

4.1. *Contributions*

We have demonstrated the use of causal network analysis to effectively identify valid drug target hypotheses for a complex disease indication, with a good success rate as demonstrated experimentally through animal studies. To the best of our knowledge, such predictive causal analytics have not been validated to this extent in a host-directed infectious disease context or across multiple viral and bacterial agents. Further, our novel *upstream regulators algorithm* successfully identified previously unassociated valid protein targets based on the predicted propagation of net regulatory effects on the host-pathogen interface. We propose that causal network analysis can extend to previous target identification approaches[7,23] by identifying valid, functionally important targets not identifiable through study of direct host-pathogen interactions alone.

We attribute part of our success the *accuracy and contextual detail of the underlying causal network*, which in turn is based on semantically-normalized IKB content. In particular, IKB findings are (a) manually modeled by experts to ensure accurate representation of the underlying biology[24]; (b) always supported by experimental evidence (no predicted or inferred data); and (c) annotated in sufficient biological and experimental detail to allow finding inclusion or exclusion based on contextual fit to the pathogen in question. We suggest that such normalized, contextualized, experimentally-grounded network datasets can improve the quality of any causal network analyses by driving the algorithm directly (as is our case), or by serving as a high-quality training set for learning-based approaches[25].

Finally, we developed a *framework for rapid, team-based, computational target discovery* to run multiple target ID algorithms in parallel, formalize their predictive outputs and supporting evidence as hypothesized mechanism of action for a novel drug target, and review and prioritize the targets using interactive, collaborative pathway tools. In addition to supporting rapid, evidence-based generation of target lists for medical countermeasures, we believe this model can be extended to include targets identified experimentally e.g. via screening approaches, as well as expert suggested hypotheses[26], thus potentially helping unify computational and experimental target identification approaches.

Our methodology can be applied to any disease where a body of host pathway knowledge has been experimentally characterized and can be modeled as causal, regulatory network relationships. For novel or emerging pathogens that are as of yet unstudied, evolutionary mapping using next-generation sequencing would allow a similar approach using host-pathogen pathway knowledge from closely-related evolutionary neighbors, although some loss of performance should be expected. Finally, a drug repurposing use case could be directly supported by automatically filtering or prioritizing hypotheses anchored by a specific drug or drug class. This would, in turn, highlight candidate compounds for use in target validation studies.

5. Conclusion and future work

Our scientific objective was to identify broad spectrum countermeasures to viral and intracellular biothreats. We have described and evaluated a novel target discovery methodology that is: *host-directed* and *broad-spectrum* in biological focus; *unbiased* in its consideration of prior target association with the disease of interest; *computationally-enabled* by formal models of *disease pathways* and *host-pathogen mechanisms*; and delivers *testable, evidence-based target hypotheses* suitable for experimental validation in rapid response scenario. Our empirical results validate this approach and, more generally, for the use of causal analysis for the discovery of novel drug targets. While our "pathogen and mechanism first" approach focuses primarily on broad-spectrum therapeutics, we believe this approach is readily adaptable to single-spectrum (i.e. against only one pathogen) target identification scenarios as well as other disease areas. We suggest that systems biology pathway models are sufficiently mature to be used alongside traditional screening-based approaches in most applied drug discovery initiatives.

5.1. *Acknowledgments*

This work is funded by Ingenuity Systems and the Defense Threat Reduction Agency under award W81XWH-08-2-0002, for which the U.S. Army Medical Research Acquisition, 820 Chandler Street, Fort Detrick MD 21702-5014 is the awarding and administering acquisition office. We thank COL Joseph Palma, COL George Christopher, COL George Korch, COL John Skvorak, Dr. Eric Van Gieson, Dr. Connie Smalljohn, and the Ingenuity and USAMRIID research communities for their support throughout this investigation. Portions of this work are available from Ingenuity at http://www.ingenuity.com/. Support materials for this work, including technical description of the causal algorithms are available at http://pages.ingenuity.com/PSB2013-Felciano-Supporting-Materials.html.

References

1. Berns, K. I. *et al.* Public health and biosecurity. Adaptations of avian flu virus are a cause for concern. *Science (New York, N.Y.)* **335**, 660–1 (2012).
2. Valdivia-Granda, W. A. Bioinformatics for biodefense: challenges and opportunities. *Biosecurity and bioterrorism : biodefense strategy, practice, and science* **8**, 69–77 (2010).

3. Fauci, A. S. Emerging and re-emerging infectious diseases: influenza as a prototype of the host-pathogen balancing act. *Cell* **124**, 665–70 (2006).
4. Cohen, O. J., Kinter, A. & Fauci, A. S. Host factors in the pathogenesis of HIV disease. *Immunological reviews* **159**, 31–48 (1997).
5. Shurtleff, A. C., Nguyen, T. L., Kingery, D. A. & Bavari, S. Therapeutics for filovirus infection: traditional approaches and progress towards in silico drug design. *Expert opinion on drug discovery* (2012).
6. Bowick, G. C. & Barrett, A. D. T. Comparative pathogenesis and systems biology for biodefense virus vaccine development. *Journal of biomedicine & biotechnology* **2010**, 236528 (2010).
7. Kash, J. C. *et al.* Genomic analysis of increased host immune and cell death responses induced by 1918 influenza virus. **443**, 578–581 (2006).
8. Felciano, R. M. Knowledge-based computational pathways analysis for integrative drug discovery. *American Institute of Chemical Engineers (AIChE) Annual Meeting* (2005).
9. Ficenec, D. *et al.* Computational knowledge integration in biopharmaceutical research. *Brief Bioinform* **4**, 260–278 (2003).
10. Calvano, S. E. *et al.* A network-based analysis of systemic inflammation in humans. *Nature* **437**, 1032–1037 (2005).
11. Ingenuity Systems Inc Ingenuity Pathway Analysis Software. *Ingenuity Technical Documentation* (2011).at <http://www.ingenuity.com/products/ipa/>
12. Callahan, A., Dumontier, M. & Shah, N. H. HyQue: evaluating hypotheses using Semantic Web technologies. *Journal of biomedical semantics* **2 Suppl 2**, S3 (2011).
13. Tran, N., Baral, C., Nagaraj, V. J. & Joshi, L. Knowledge-based framework for hypothesis formation in biochemical networks. *Bioinformatics (Oxford, England)* **21 Suppl 2**, ii213–9 (2005).
14. Ingenuity Systems Inc Ingenuity Knowledge Base. *Ingenuity Technical Documentation* (2011).at <http://www.ingenuity.com/science/knowledgebase/>
15. Ruttenberg, A., Rees, J. A., Samwald, M. & Marshall, M. S. Life sciences on the Semantic Web: the Neurocommons and beyond. *Briefings in bioinformatics* **10**, 193–204 (2009).
16. Lam, H. Y. K. *et al.* AlzPharm: integration of neurodegeneration data using RDF. *BMC bioinformatics* **8 Suppl 3**, S4 (2007).
17. Le Novère, N. *et al.* Meeting report from the first meetings of the Computational Modeling in Biology Network (COMBINE). *Standards in genomic sciences* **5**, 230–42 (2011).
18. Percha, B., Garten, Y. & Altman, R. B. Discovery and explanation of drug-drug interactions via text mining. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing* 410–21 (2012).
19. Felciano, R. M. *et al.* Systems Biology Approaches to Target and Mechanism Discovery in Infectious Diseases. *2010 CBD S&T Conference Conference Proceedings* (2010).
20. Wishart, D. S. *et al.* DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic acids research* **34**, D668–72 (2006).
21. Zhu, F. *et al.* Update of TTD: Therapeutic Target Database. *Nucleic acids research* **38**, D787–91 (2010).
22. Huh, D. *et al.* Reconstituting organ-level lung functions on a chip. *Science (New York, N.Y.)* **328**, 1662–8 (2010).
23. König, R. *et al.* Human host factors required for influenza virus replication. *Nature* **463**, 813–7 (2010).
24. Cho, R. J., Chen, R. O., Felciano, R. M., Richards, D. R. & Norman, P. US Patent US6741986: Method and system for performing information extraction and quality control for a knowledgebase. (2001).
25. Novershtern, N., Regev, A. & Friedman, N. Physical Module Networks: an integrative approach for reconstructing transcription regulation. *Bioinformatics (Oxford, England)* **27**, i177–85 (2011).
26. Clark, T. & Kinoshita, J. Alzforum and SWAN: the present and future of scientific web communities. *Briefings in bioinformatics* **8**, 163–71 (2007).