RESEARCH ARTICLE

REVISED Machine learning models identify molecules active against the Ebola virus *in vitro* [version 3; referees: 2 approved]

Sean Ekins[1-3], Joel S. Freundlich[4], Alex M. Clark[5], Manu Anantpadma[6], Robert A. Davey[6], Peter Madrid[7]

[1]Collaborations in Chemistry, Fuquay-Varina, NC, 27526, USA
[2]Collaborations Pharmaceuticals Inc, Fuquay-Varina, NC, 27526, USA
[3]Collaborative Drug Discovery, Burlingame, CA, 94010, USA
[4]Departments of Pharmacology & Physiology and Medicine, Center for Emerging and Reemerging Pathogens, UMDNJ, New Jersey Medical School, Newark, NJ, 07103, USA
[5]Molecular Materials Informatics, Inc., Montreal, 94025, Canada
[6]Texas Biomedical Research Institute, San Antonio, TX, 78227, USA
[7]SRI International, Menlo Park, CA, 94025, USA

v3 **First published:** 20 Oct 2015, **4**:1091 (doi: 10.12688/f1000research.7217.1)
Second version: 04 Jan 2016, **4**:1091 (doi: 10.12688/f1000research.7217.2)
Latest published: 17 Jan 2017, **4**:1091 (doi: 10.12688/f1000research.7217.3)

Abstract

The search for small molecule inhibitors of Ebola virus (EBOV) has led to several high throughput screens over the past 3 years. These have identified a range of FDA-approved active pharmaceutical ingredients (APIs) with anti-EBOV activity *in vitro* and several of which are also active in a mouse infection model. There are millions of additional commercially-available molecules that could be screened for potential activities as anti-EBOV compounds. One way to prioritize compounds for testing is to generate computational models based on the high throughput screening data and then virtually screen compound libraries. In the current study, we have generated Bayesian machine learning models with viral pseudotype entry assay and the EBOV replication assay data. We have validated the models internally and externally. We have also used these models to computationally score the MicroSource library of drugs to select those likely to be potential inhibitors. Three of the highest scoring molecules that were not in the model training sets, quinacrine, pyronaridine and tilorone, were tested *in vitro* and had EC_{50} values of 350, 420 and 230 nM, respectively. Pyronaridine is a component of a combination therapy for malaria that was recently approved by the European Medicines Agency, which may make it more readily accessible for clinical testing. Like other known antimalarial drugs active against EBOV, it shares the 4-aminoquinoline scaffold. Tilorone, is an investigational antiviral agent that has shown a broad array of biological activities including cell growth inhibition in cancer cells, antifibrotic properties, α7 nicotinic receptor agonist activity, radioprotective activity and activation of hypoxia inducible factor-1. Quinacrine is an antimalarial but also has use as an anthelmintic. Our results suggest data sets with less than 1,000 molecules can produce validated machine learning models that can in turn be utilized to identify novel EBOV inhibitors *in vitro*.

Open Peer Review

Referee Status: ✓ ✓

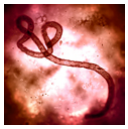
	Invited Referees 1	2
REVISED version 3 published 17 Jan 2017	✓	
REVISED version 2 published 04 Jan 2016		
version 1 published 20 Oct 2015	✓ report	✓ report

1 **Sanja Glisic**, University of Belgrade, Serbia

2 **Sandeep Chakraborty**, University of California, USA

Discuss this article

Comments (0)



This article is included in the Ebola collection.

This article is included in the Machine learning: life sciences collection.

Corresponding author: Sean Ekins (ekinssean@yahoo.com)

Competing interests: S.E. works for Collaborations in Chemistry, and Collaborations Pharmaceuticals, Inc. and S.E. and A.M.C. consult for Collaborative Drug Discovery Inc.

How to cite this article: Ekins S, Freundlich JS, Clark AM *et al.* **Machine learning models identify molecules active against the Ebola virus *in vitro* [version 3; referees: 2 approved]** *F1000Research* 2017, **4**:1091 (doi: 10.12688/f1000research.7217.3)

Grant information: The author(s) declared that no grants were involved in supporting this work.

First published: 20 Oct 2015, **4**:1091 (doi: 10.12688/f1000research.7217.1)

REVISED **Amendments from Version 2**

It has been brought to our attention that there was an error in the pyronaridine 2D structure drawing in Table 3 (missing a nitrogen from bottom right ring). This has now been corrected.

See referee reports

Introduction

In 2014, the outbreak of the Ebola virus (EBOV) in West Africa highlighted the need for broad-spectrum antiviral drugs for this and other emerging viruses[1]. Several groups had previously performed high throughput screens (HTS) and identified FDA approved drugs (amodiaquine, chloroquine, clomiphene and toremifene) with *in vitro* growth inhibitory activities against EBOV[2,3]. It appears none of these molecules were tried during the epidemic in Africa[4], likely due to the lack of efficacy data in higher order species. We have previously summarized the numerous small molecules described in the literature as possessing antiviral activity that could be further evaluated for their potential EBOV activity alongside the few new antivirals. We have found that there is considerable prior knowledge regarding these small molecules possessing activity against EBOV *in vitro* or in animal models[5–8], and this includes a number of accessible FDA-approved drugs[2,3,9]. Another recent study has shown three approved ion channel blockers (amiodarone, dronedarone, and verapamil) inhibited EBOV cellular entry[9]. The drugs were given at concentrations that would be achieved in human serum, and were effective against several of the filoviruses[9]. None of the FDA approved drugs described in these various studies were designed to target the Ebola virus. For example amodiaquine and chloroquine are well known antimalarials, clomiphene and toremifene are selective estrogen receptor modulators, while amiodarone, dronedarone, and verapamil are anti-arrhythmics[4]. It may or may not be of importance but all of these compounds have a common tertiary amine feature[10,11]. What is important is that they are all orally bioavailable and generally safe for humans at their approved doses. Some have suggested that G-protein-coupled receptors (GPCRs) may play a role in filoviral entry and receptor antagonists could be developed as anti-EBOV therapies[12]. The compounds which are FDA-approved drugs for other diseases[2,3,9] but with activity against EBOV *in vitro* or *in vivo* may represent useful starting points with the advantage that much is known regarding their absorption, distribution, metabolism and excretion (ADME) and toxicity properties. Thus, these repurposed drugs may represent a more advanced starting point for therapeutic development and approval compared with new chemical entities for preventing the spread and mortality associated with EBOV.

Beyond these early stage drugs, there are a number of other compounds that have also been identified as active against EBOV (summarized in a review[13]). A thorough literature search identified 55 molecules suggested to have activity against EBOV *in vitro* and/or *in vivo* which were evaluated from the perspective of an experienced medicinal chemist as well as using simple molecular properties and ultimately 16 were highlighted as desirable[14]. This dataset overlaps to some extent with another review that identified over 60 molecules[15]. Two recent repurposing screens identified 53[16] and 80[17] compounds with antiviral activity which also overlap the earlier screens. Additional studies have identified small number of inhibitors[18,19]. In total there may now be close to several hundred compounds identified with activity against EBOV *in vitro*.

Approaches with more capacity to screen compounds include using computational methods as a filter before *in vitro* testing. Computational models for anti-EBOV activity include one which used the average quasi valence number (AQVN) and the electron-ion interaction potential (EIIP), parameters determining long-range interaction between biological molecules for virtual screening of DrugBank and suggested hundreds of compounds to test[20]. A follow up to this study proposed ibuprofen for testing[21]. Others have also used computational docking studies to propose multi-target inhibitors of VP40, VP35, VP30 and VP24[22], inhibitors of VP40[23] or have suggested molecules to test in the absence of computational approaches[24,25]. We are unaware of any validation of these compounds. A further computational approach used a pharmacophore[26] that was generated from four FDA approved compounds resulting from the two earliest high throughput screens against EBOV[2,3]. This pharmacophore closely matched the receptor-ligand pharmacophores for the EBOV protein 35 (VP35)[5]. Follow-up docking studies suggested that these compounds may also have favorable inhibitory interactions with this receptor. The pharmacophore was further used to screen several compound libraries[27]. We proposed that if we could learn from the many compounds already screened for anti-EBOV activity, we could more efficiently find additional compounds and perhaps understand the key molecular features needed for antiviral activity[14]. We speculated then that Laplacian-corrected Naïve Bayesian classifier models might be useful as they have been for *M. tuberculosis*[28,29] and more recently for *T. cruzi*[30]. To our knowledge machine learning approaches to identify EBOV inhibitors have not been attempted elsewhere. The current study extends the machine learning approach to EBOV and uses both commercially available Bayesian, Support Vector Machines (SVM) and recursive partitioning methods and open source Bayesian software for model generation and compound scoring. We report the identification of three novel EBOV inhibitors with nanomolar EC_{50} values as validation of this approach.

Methods

Chemicals and materials

Quinacrine hydrochloride, pyronaridine tetraphosphate, and tilorone dihydrochloride (BOC Sciences, Shirley, NY), bafilomycin A1, and chloroquine diphosphate (Sigma, St. Louis, MO) were dissolved in either DMSO or water as 10 mM stock solutions and were stored at -20°C. The nucleus staining dye, Hoechst 33342, CellMask Deep™ Red cytoplasmic/nuclear stain, NHS-Alexa-488 dye, the Dual-Glo® Luciferase Assay System and CytoTox 96™ assay kit were purchased from Promega (Promega, Madison, WI). The modified MTT assay Cell Counting Kit 8 was procured from Dojindo Molecular Technologies (Dojindo Molecular Technologies, Gaithersburg, MD). The 96-well high-content imaging plates were obtained from BD (BD Biosciences, Franklin Lakes, NJ) and 96-well white-walled tissue culture plates were from Corning (Corning Life Sciences, MA). The Opera QEHS confocal imaging reader, Acapella™ and Definiens™ image analysis packages were purchased from PerkinElmer (PerkinElmer, USA). Image acquisition was done using Nikon TI eclipse high content imaging enabled microscope running NIS elements high content imaging software (version 4.30.02).

Machine learning

868 molecules from the viral pseudotype entry assay and the EBOV replication assay from a recent publication[3,31] were made available as an sdf file[3]. Salts were stripped and duplicates removed using Discovery Studio 4.1 (Biovia, San Diego, CA)[32–36]. For each assay, compounds with IC_{50} values less than 50 µM were selected as actives. All other compounds were classed as inactives. Models were generated using a standard protocol with the following molecular descriptors: molecular function class fingerprints of maximum diameter 6 (FCFP_6)[37], AlogP[37a], molecular weight, number of rotatable bonds, number of rings, number of aromatic rings, number of hydrogen bond acceptors, number of hydrogen bond donors, and molecular fractional polar surface area. Models were validated using five-fold cross validation (leave out 20% of the database). Bayesian, Support Vector Machine and Recursive Partitioning Forest and single tree models built with the same molecular descriptors in Discovery Studio were compared. For SVM models, we calculated interpretable descriptors in Discovery Studio and then used Pipeline Pilot to generate the FCFP_6 descriptors followed by integration with R[38]. RP Forest and RP Single Tree models used the standard protocol in Discovery Studio. In the case of RP Forest models, ten trees were created with bagging. Bagging is short for "Bootstrap AGgregation". For each tree, a bootstrap sample of the original data is taken, and this sample is used to grow the tree. RP Single Trees had a minimum of ten samples per node and a maximum tree depth of 20. In all cases, 5-fold cross validation or leave out 50% × 100 fold cross validation was used to calculate the Receiver Operator Curve (ROC) for the models generated[28,29].

Open Bayesian models

Open Bayesian models for the Ebola datasets were developed using open source software[39–41] and loaded into the Mobile Molecular Data Sheet (MMDS (http://molmatinf.com/)) and then the two models were used to score the three compounds selected by the earlier models. These two models are also openly accessible (http://molsync.com/ebola/) and can be uploaded into MMDS in order to score molecules of interest.

Pharmacophore mapping

Pyronaridine was mapped to the recently published pharmacophore[26] derived from Ebola *in vitro* inhibitors amodiaquine, chloroquine, clomiphene and toremifene in Discovery Studio Vers 4.1 and a fit score was generated.

In vitro testing

Recombinant, infectious Ebola virus encoding green fluorescent protein (GFP) was used for testing efficacy of compounds and was originally provided by Dr. Heinz Feldmann, Rocky Mountain Laboratories. The strain that was used has the GFP gene inserted between the VP30 and VP24 genes. All viral infections were done in the BSL-4 lab at Texas Biomedical Research Institute. Briefly, 4,000 HeLa cells per well were grown overnight in 384-well tissue culture plates, the volume of DMEM (Fisher scientific, Cat#MT10017CV) culture medium supplemented with 10% fetal bovine serum (Gemini Bio-Products, Cat#100106) was 25 µL. On the day of assay, test drugs were diluted to 1 mM concentration in complete medium. 25 µL of this mixture was added to the cells already containing 25 µL medium to achieve a concentration of 500 µM. All treatments were done in triplicates. 25 µL of medium was removed from the first wells and added to the next well. This type of serial dilution was done 12 times and treated cells were then incubated at 37°C in a humidified CO_2 incubator for 1 hour. Final concentrations of 250, 125, 62.5, 31.25, 15.62, 7.81, 3.9, 1.9, 0.97, 0.48, 0.24 and 0.12 µM were achieved upon addition of 25 µL of infection mix containing Ebola-GFP virus, Bafilomycin at a final concentration of 10 nM was used as a positive control drug. Infections were done to achieve a MOI of 0.05 to 0.15. Infected cells were incubated for 24 hours. 24 hours post-infection cells were fixed by immersing the plates in formalin for 24 hours at 4°C. Fixed plates were decontaminated and brought out of the BSL-4. Formalin from fixed plates was decanted and plates were washed thrice with PBS. EBOV-infected cells were stained for nuclei using Hoechst at 1:50,000 dilution and plates were imaged. Nuclei (blue) and infected cells (green) were counted using CellProfiler software (Broad Institute)- Version 2.1.1. Total number of nuclei (blue) was used as a proxy for cell numbers and a loss of cell number was assumed to reflect cytotoxicity. Concentrations where total cell numbers were 20% less than the control were rejected from the analysis.

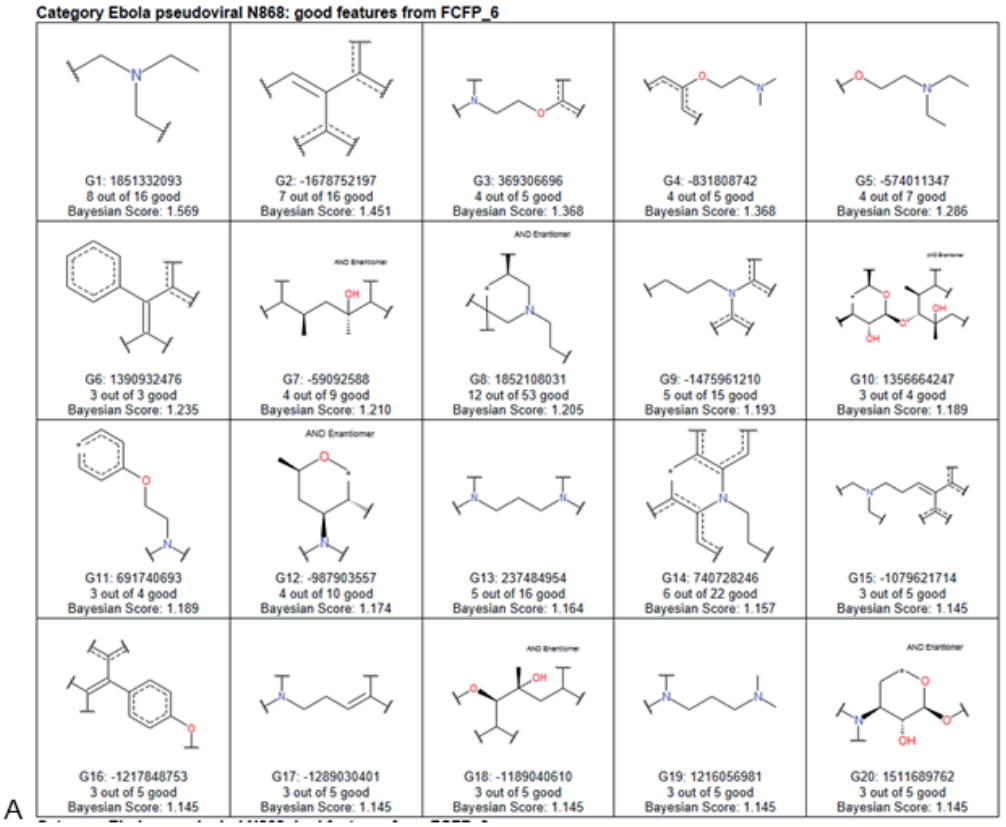
Results

Machine learning

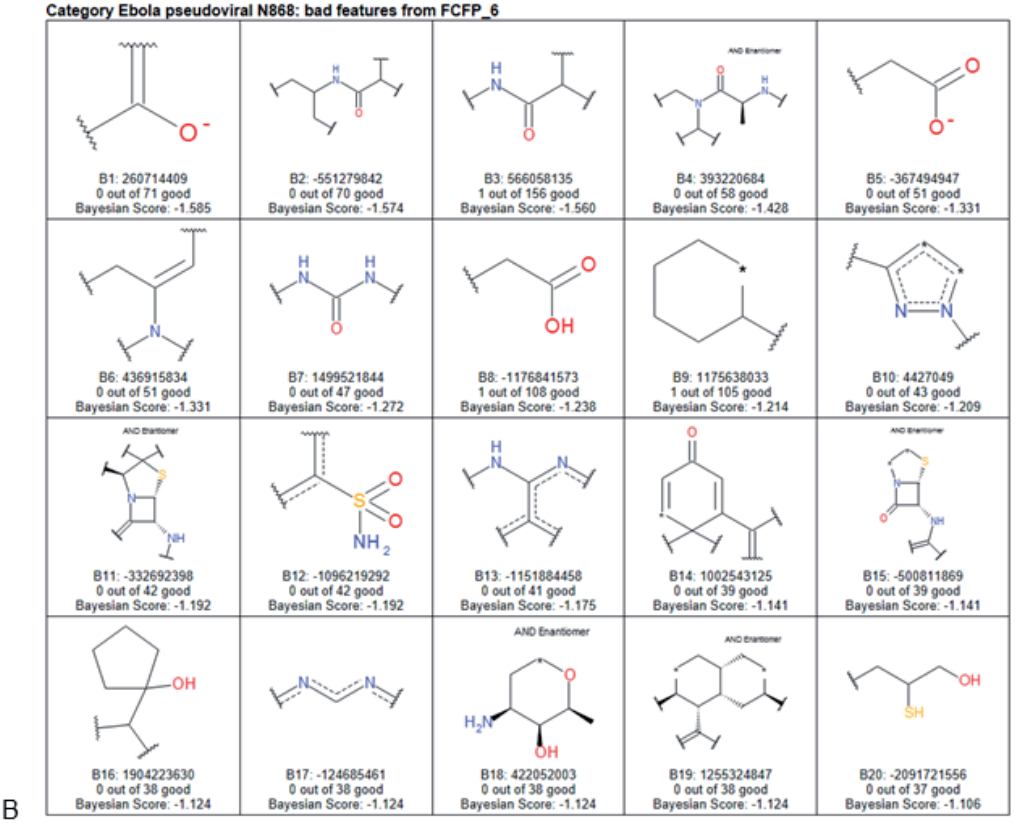
Using 5-fold cross validation the Bayesian approach (Data S1 and Data S2) performed the best for the EBOV replication data and was equivalent for the RP Forest approach (Table 1) and was better than SVM (Data S3 and Data S4) for the pseudotype data. The Open Bayesian models had ROC scores slightly lower than the Bayesian models built with Discovery Studio. A more exhaustive cross validation for the Bayesian models is the 'leave out 50% repeated randomly 100 times' which produced ROC values greater than 0.8 and were comparable to the 5-fold cross validation data. This indicated the models are stable. For the EBOV pseudotype assay, alkoxyethylamino was a common feature amongst active compounds in the training set, as were 1,3-diaminopropyl and saturated six-member heterocycles with an oxygen and perhaps an additional heteroatom in the ring (Figure 1A). Training set inactives commonly featured carboxylic acids, N,N'-disubstituted ureas, secondary and tertiary amides, pyrazoles, aromatic sulfonamides, tertiary cyclopentanols, 1,2-mercaptoethanol, and penams (Figure 1B). For the replication assay training set, active features included piperazine, phenothiazine, tertiary amines, and alkoxyethylamino (Figure 2A).

Table 1. Machine learning model cross validation Receiver Operator Curve (ROC) statistics.

Models (training set 868 compounds)	RP Forest (Out of bag ROC)	RP Single Tree (With 5 fold cross validation ROC)	SVM (with 5 fold cross validation ROC)	Bayesian (with 5 fold cross validation ROC)	Bayesian (leave out 50% × 100 ROC)	Open Bayesian (with 5 fold cross validation ROC)
Ebola replication (actives = 20)	0.70	0.78	0.73	0.86	0.86	0.82
Ebola Pseudotype (actives = 41)	0.85	0.81	0.76	0.85	0.82	0.82

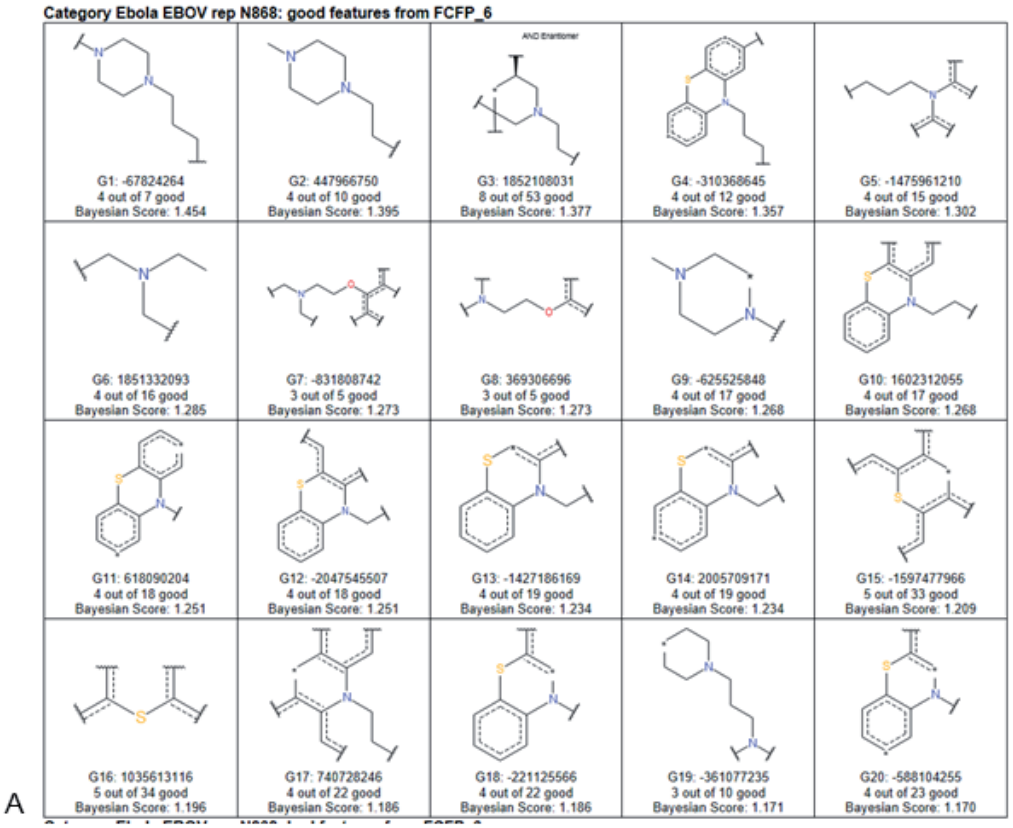


A



B

Figure 1. A. Active and **B**. Inactive features for the Discovery Studio pseudotype Bayesian model.



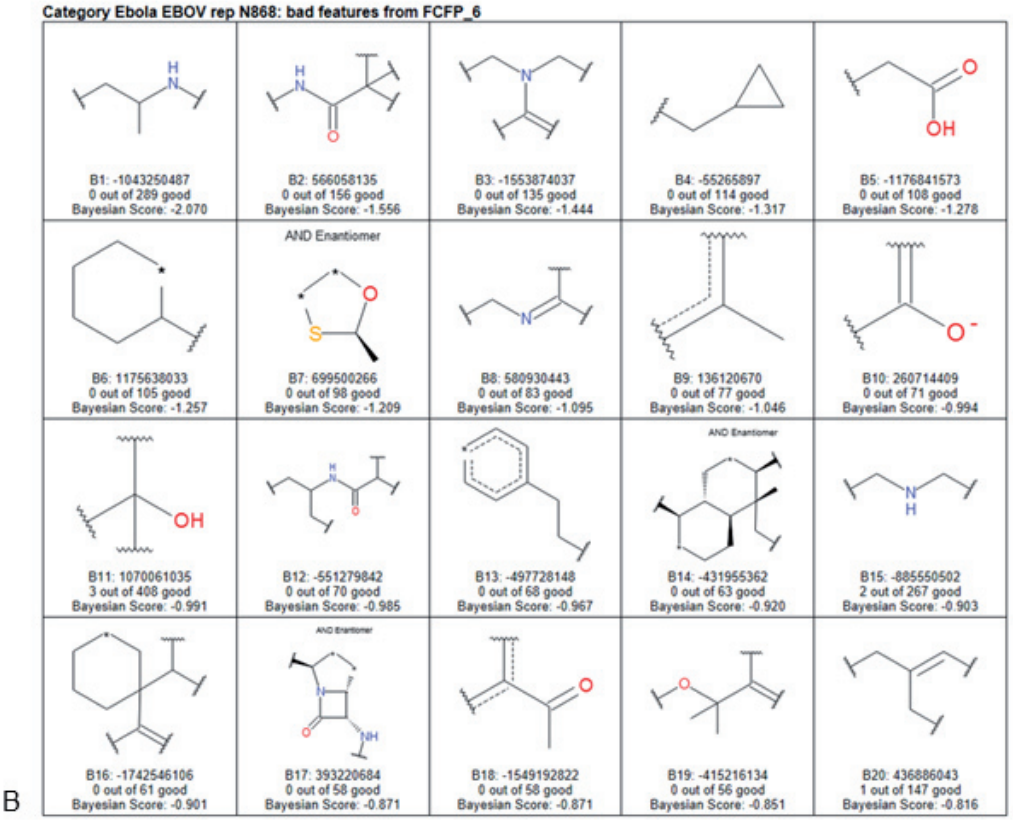


Figure 2. **A**. Active and **B**. Inactive features for the Discovery Studio EBOV replication model.

Inactive features included secondary amides, disubstituted amines, cyclopropylmethyl, carboxylic acids, 1,3-oxathiolanes, tertiary alcohols, phenethyl, and penams (Figure 2B). An actives feature common between both assays/models was alkoxyethylamino. Inactives features in common between both were carboxylic acids, secondary amides, penams and tertiary alcohols, which may relate to properties which prevent the molecules from accessing cellular sites of viral activity.

The MicroSource Spectrum set of 2320 compounds was then scored with both Bayesian models (Data S5). Predicted actives were quantified as to their chemical similarity, or distance, from molecules in the training set. When excluding compounds in the training set (as well as antipsychotics and other less desirable CNS active compounds), those scoring highly were considered most interesting and included the antiviral tilorone, the antimalarials quinacrine and pyronaridine (Figure 3). Perhaps not surprisingly, tertiary amines scored particularly well. These molecules were also scored with the open Bayesian models (Data S6) and all replication models scored the compounds highly (values close to or greater than 1). None of these three compounds has been described in recent reviews of small molecules with activity against EBOV[14–16], to our knowledge.

Pharmacophore

The MicroSource set had previously been screened with the published Ebola common feature pharmacophore[26,27], using the van der Waals surface of amodiaquine (which was more potent than chloroquine[3]) to limit the number of hits retrieved[42–44]. Two of the three selected – compounds quinacrine (fit score 2.59) and tilorone (fit score 3.65) – were retrieved previously. We therefore used the ligand pharmacophore mapping to map pyronaridine to the pharmacophore without the van der Waals surface (Figure 4, Fit score of 3.60 suggested this was a good match to pharmacophore features).

In vitro testing

The three selected compounds were tested *in vitro* alongside the positive control chloroquine which gave an expected dose response curve (Figure 5, Table 2). Quinacrine, pyronaridine and tilorone, were tested *in vitro* and had EC_{50} values of 350, 420 and 230 nM, respectively which were lower than for chloroquine 4.0 µM. Several images created in this study illustrate the results of this high content screen (Data S7).

Discussion

Our recent work on neglected diseases has shown that we can learn from existing assay datasets. Specifically we have previously analyzed large datasets for *Mycobacterium tuberculosis* to build machine learning models that use single point data, dose-response data[43,45], combine bioactivity and cytotoxicity data (e.g. Vero, HepG2 or other model mammalian cells)[28,29,46] or combinations of these sets[47,48]. These models in turn have been validated

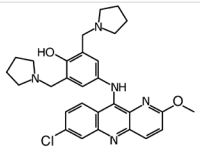
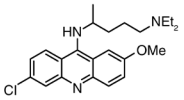
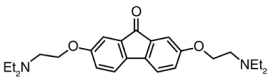
	Pyronaridine	Quinacrine	Tilorone
			
Discovery Studio Replication model score	23.62	29.73	20.90
Discovery Studio Pseudovirus model score	17.16	22.25	17.73
Open Bayesian Replication model score	1.01	1.63	1.31
Open Bayesian Pseudovirus model score	0.72	1.28	1.17

Figure 3. Molecules scoring well with the Ebola Bayesian models. For comparison, chloroquine scored 31.38 in the replication Discovery Studio Bayesian model, 24.55 in the Discovery Studio Pseudovirus Bayesian model, 1.63 in the Open Bayesian Replication model and 0.51 in the Open Bayesian Pseudovirus model.

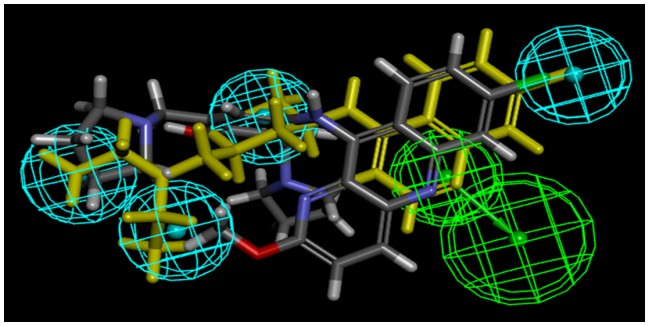


Figure 4. Pyronaridine mapped to a previously published pharmacophore based on compounds active against Ebola virus *in vitro*. Fit score of 3.60 (Chloroquine (yellow) = 4.21).

Table 2. Effect of drug treatment on infection with Ebola-GFP (n=3 per compound). The cytotoxicity of compounds are represented as a 50% cytotoxicity concentration (CC_{50}) estimated by the lowest concentration of drug that produced ≥ 50% loss in cell number by nuclei counting.

Compound	EC_{50} (μM) [95% CI]	Cytotoxicity CC_{50} (μM)
Chloroquine	4.0 [1.0–15]	250
Pyronaridine	0.42 [0.31–0.56]	3.1
Quinacrine	0.35 [0.28–0.44]	6.2
Tilorone	0.23 [0.09–0.62]	6.2

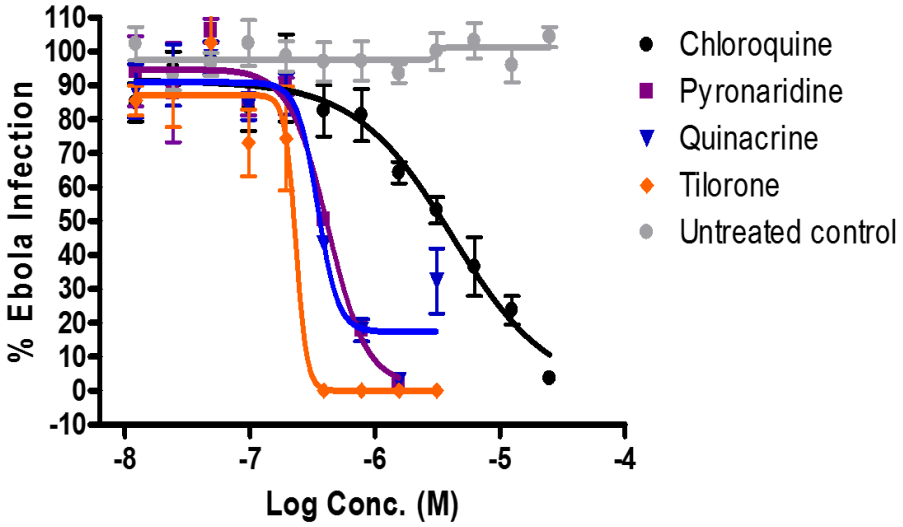


Figure 5. Effect of drug treatment on infection with Ebola-GFP. Cells were treated and then challenged with Ebola virus encoding GFP. Infection efficiency was calculated as infected cells (expressing GFP)/total cells and normalized to infection efficiency seen in the untreated control. Shown is one representative experiment where each point is the average of 3 independent measurements of infection +/- standard deviation. Dose response curves were fitted by non-linear regression.

with additional non-overlapping datasets, demonstrating that it is possible to use publically accessible data to find novel *in vitro* active antituberculars. We have also applied the same approach recently to identify a molecule with *in vitro* and *in vivo* activity against *T. cruzi*[30]. In the current study we found that different machine learning methods produced similar 5-fold cross validation data, although the Bayesian models had ROC values consistently above 0.80, which is preferable. One of the issues with computational models is that they are rarely accessible to others due to the commercial software licensing requirements. We have previously showed that models built with open source tools can produce validation statistics comparable to commercial modeling tools[49]. We recently made "function class fingerprints of maximum diameter 6" (FCFP6) and "extended connectivity (ECFP6) fingerprints," open source and have described their implementation with the Chemistry Development Kit (CDK)[50] components[41]. In addition we described an open source Bayesian algorithm that can be used with these descriptors[39,40]. One way to make such models more accessible is to use mobile devices for their delivery and we have developed cheminformatics mobile apps[41,51–55]. Several of these apps combine Bayesian models and open source fingerprint descriptors to enable models that can be used within a mobile app (TB Mobile, MMDS, Approved Drugs and MolPrime). This enables a scientist to select a molecule and score it with models. In the current study we used the same training sets for the anti-EBOV activity using replication and pseudotype screening data to build open source models that we can share with the community (http://molsync.com/ebola/).

The Bayesian models allowed us to select three compounds from the MicroSource compound library that scored highly and were not in the model training sets. The Open Bayesian models also scored the three hits favorably, which bodes well for screening other

compounds of interest. Two of these molecules had also been identified with our earlier pharmacophore model which may be indicative of binding to VP35[26]. When tested *in vitro* the three compounds possessed EC_{50} values 230–420 nM, much lower than the positive control chloroquine (EC_{50} 4.0 μM) used in this study and identified previously[3]. Tilorone is an investigational agent that has been known for over 40 years as an antiviral[56] and is an inducer of interferon in mice[57]. It has been shown to possess a broad array of biological activities including cell growth inhibition in PC3 CDK5dn prostate cancer cells (IC_{50} 8–12 μM)[58], inhibition of Primase DnaG from *Bacillus anthracis* (IC_{50} 7.1 μM)[59], in a mouse model of pulmonary fibrosis it decreased lung hydroxyproline content and the expression of collagen genes[60], α7 nicotinic receptor (nAChR) agonist activity (K_i 56 nM)[61], activated human alpha7 nAChR with an EC_{50} value of 2.5 μM[62], radioprotective activity[63], potent modulation of HIF-mediated gene expression in neurons with neuroprotective properties[64] and induction of the accumulation of glycosaminoglycans, delay infectious prion clearance, and prolong prion disease incubation time[65]. Quinacrine is an old antimalarial drug now more widely used as an antiprotozoal for the treatment of giardiasis[66] and as an anthelmintic. Pyronaridine is a potent antimalarial (IC_{50} 13.5 nM)[67], has activity against *Babesi*a spp.[68], is active *in vitro* (EC_{50} 225 nM) and *in vivo* (85.2% efficacy 4 days treatment at 50 mg/kg) against *T. cruzi*[30] and is a P-glycoprotein inhibitor[69]. Pyronaridine is used in combination with artesunate in the European Medicines Agency approved Pyramax[70] which has performed well in clinical trials for malaria[71]. As this molecule has already been approved this may have a more direct path to clinical testing if it is found to be active in standard animal models infected with the Ebola Virus.

As stated before in perspectives by us[72] and others[2,16,20,73], the fact that approved drugs may be repurposed for other diseases should not be viewed as a negative aspect of the small molecules, belying undesirable target promiscuity[74]. Instead, we prefer to reference recently published crystallographic analyses[75] demonstrating that small molecules may bind multiple proteins in different types of binding sites and with distinct conformations to ultimately facilitate molecular repurposing. While it would be most desirable to repurpose an approved drug and, thus, catapult a discovery effort into a Phase II trial, one should not ignore the significance of utilizing the discovery of a new use for an old drug to seed efforts in the lead optimization phase[76]. Such an expedited program would be expected to have a high probability of producing novel small molecules, closely related to or inspired by the drug, with the opportunity to translate quickly to clinical trials.

In summary, this study has added to the previous work that identified several FDA approved compounds active against EBOV *in vitro*. Future work may include identification of targets using computational or experimental approaches. We propose that these three molecules may warrant further evaluation *in vivo* as they are significantly more active than chloroquine. Larger scale virtual screening could be performed on the millions of commercially available molecules or more complete sets of approved and older no longer used drugs than have already been screened. These computational efforts can then prioritize molecules for testing. Such an approach may be a useful way to leverage the HTS data that has already been developed at great cost. In this study we have focused on just the data from a single group[3,31] but it may also be possible to combine this with the data from the other high throughput screens[2,16,17] to provide a much larger training set. There is also the opportunity to apply many different computational approaches beyond those described here to identify whole cell active compounds against EBOV. Ultimately, we should be able to identify additional compounds that could be immediately useful to treat patients with the disease while we await the approval of a vaccine.

Data availability

Supplemental data contains results from Bayesian models and SVM models as well as the output of predictions with Bayesian models and open Bayesian models.

The training sets used in the models are available as SDF files (http://molsync.com/ebola/).

Author contributions

Conceived and designed the experiments: S.E., M.A., R.A.D., P.B.M

Performed the experiments: S.E., M.A., R.A.D., P.B.M

Analyzed the data: S.E., J.S.F., M.A., R.A.D., P.B.M

Contributed reagents/materials/analysis tools: S.E., A.M.C., M.A., R.A.D., P.B.M

Wrote the manuscript: S.E., J.S.F., A.M.C., M.A., R.A.D., P.B.M

All authors have seen and agreed to the final content of the manuscript.

Competing interests

S.E. works for Collaborations in Chemistry, and Collaborations Pharmaceuticals, Inc. and S.E. and A.M.C. consult for Collaborative Drug Discovery Inc.

Grant information

The author(s) declared that no grants were involved in supporting this work.

Acknowledgments

SE kindly acknowledges Biovia for kindly providing Discovery Studio and Dr. Megan Coffee and Dr. Christopher Southan for initially stimulating interest in this topic.

Supplementary materials

Supplemental data S1–S4 and S6, S7.

Supplemental data S1. Pseudotype Bayesian model, Supplemental data S2. EBOV replication Bayesian, Supplemental Data 3. SVM output file for Pseudotype model, Supplemental Data S4. SVM output file for EBOV replication model, Supplemental Data S6. Predictions for Ebola activity using Open Bayesian models in the MMDS app, Supplemental Data S7. High content screening images illustrating inhibition of Ebola and cytotoxic concentration.

http://dx.doi.org/10.5256/f1000research.7217.s110159

Click here to access the data.

Supplemental data S5.

MicroSource predictions with Bayesian models xl file.

Click here to access the data.

References

1. Ekins S, Southan C, Coffee M: **Finding small molecules for the 'next Ebola' [version 2; referees: 2 approved].** *F1000Res.* 2015; **4**: 58.
PubMed Abstract | Publisher Full Text | Free Full Text
2. Johansen LM, Brannan JM, Delos SE, *et al.*: **FDA-approved selective estrogen receptor modulators inhibit Ebola virus infection.** *Sci Transl Med.* 2013; **5**(190): 190ra79.
PubMed Abstract | Publisher Full Text | Free Full Text
3. Madrid PB, Chopra S, Manger ID, *et al.*: **A systematic screen of FDA-approved drugs for inhibitors of biological threat agents.** *PLoS One.* 2013; **8**(4): e60579.
PubMed Abstract | Publisher Full Text | Free Full Text
4. Ekins S, Coffee M: **FDA approved drugs as potential Ebola treatments [version 2; referees: 2 approved].** *F1000Res.* 2015; **4**: 48.
PubMed Abstract | Publisher Full Text | Free Full Text
5. Brown CS, Lee MS, Leung DW, *et al.*: ***In silico* derived small molecules bind the filovirus VP35 protein and inhibit its polymerase cofactor activity.** *J Mol Biol.* 2014; **426**(10): 2045–58.
PubMed Abstract | Publisher Full Text | Free Full Text
6. Han Z, Lu J, Liu Y, *et al.*: **Small-molecule probes targeting the viral PPxY-host Nedd4 interface block egress of a broad range of RNA viruses.** *J Virol.* 2014; **88**(13): 7294–306.
PubMed Abstract | Publisher Full Text | Free Full Text
7. Opsenica I, Burnett JC, Gussio R, *et al.*: **A chemotype that inhibits three unrelated pathogenic targets: the botulinum neurotoxin serotype A light chain, *P. falciparum* malaria, and the Ebola filovirus.** *J Med Chem.* 2011; **54**(5): 1157–69.
PubMed Abstract | Publisher Full Text | Free Full Text
8. Johnson JC, Martinez O, Honko AN, *et al.*: **Pyridinyl imidazole inhibitors of p38 MAP kinase impair viral entry and reduce cytokine induction by *Zaire ebolavirus* in human dendritic cells.** *Antiviral Res.* 2014; **107**: 102–9.
PubMed Abstract | Publisher Full Text | Free Full Text
9. Gehring G, Rohrmann K, Atenchong N, *et al.*: **The clinically approved drugs amiodarone, dronedarone and verapamil inhibit filovirus cell entry.** *J Antimicrob Chemother.* 2014; **69**(8): 2123–31.
PubMed Abstract | Publisher Full Text
10. Kazmi F, Hensley T, Pope C, *et al.*: **Lysosomal sequestration (trapping) of lipophilic amine (cationic amphiphilic) drugs in immortalized human hepatocytes (Fa2N-4 cells).** *Drug Metab Dispos.* 2013; **41**(4): 897–905.
PubMed Abstract | Publisher Full Text | Free Full Text
11. Nadanaciva S, Lu S, Gebhard DF, *et al.*: **A high content screening assay for identifying lysosomotropic compounds.** *Toxicol In Vitro.* 2011; **25**(3): 715–23.
PubMed Abstract | Publisher Full Text
12. Cheng H, Lear-Rooney CM, Johansen L, *et al.*: **Inhibition of Ebola and Marburg Virus Entry by G Protein-Coupled Receptor Antagonists.** *J Virol.* 2015; **89**(19): 9932–8.
PubMed Abstract | Publisher Full Text | Free Full Text
13. De Clercq E: **Ebola virus (EBOV) infection: Therapeutic strategies.** *Biochem Pharmacol.* 2015; **93**(1): 1–10.
PubMed Abstract | Publisher Full Text
14. Litterman N, Lipinski C, Ekins S: **Small molecules with antiviral activity against the Ebola virus [version 1; referees: 2 approved].** *F1000Res.* 2015; **4**: 38.
PubMed Abstract | Publisher Full Text | Free Full Text
15. Picazo E, Giordanetto F: **Small molecule inhibitors of ebola virus infection.** *Drug Discov Today.* 2014; **20**(2): 277–86.
PubMed Abstract | Publisher Full Text
16. Kouznetsova J, Sun W, Martínez-Romero C, *et al.*: **Identification of 53 compounds that block Ebola virus-like particle entry via a repurposing screen of approved drugs.** *Emerg Microbes Infect.* 2014; **3**(12): e84.
PubMed Abstract | Publisher Full Text | Free Full Text
17. Johansen LM, DeWald LE, Shoemaker CJ, *et al.*: **A screen of approved drugs and molecular probes identifies therapeutics with anti-Ebola virus activity.** *Sci Transl Med.* 2015; **7**(290): 290ra89.
PubMed Abstract | Publisher Full Text
18. Basu A, Mills DM, Mitchell D, *et al.*: **Novel Small Molecule Entry Inhibitors of Ebola Virus.** *J Infect Dis.* 2015; **212**(Suppl 2): S425–34.
PubMed Abstract | Publisher Full Text | Free Full Text
19. Long J, Wright E, Molesti E, *et al.*: **Antiviral therapies against Ebola and other emerging viral diseases using existing medicines that block virus entry [version 2; referees: 2 approved].** *F1000Res.* 2015; **4**: 30.
PubMed Abstract | Publisher Full Text | Free Full Text
20. Veljkovic V, Loiseau PM, Figadere B, *et al.*: **Virtual screen for repurposing approved and experimental drugs for candidate inhibitors of EBOLA virus infection [version 1; referees: 2 approved].** *F1000Res.* 2015; **4**: 34.
PubMed Abstract | Publisher Full Text | Free Full Text
21. Veljkovic V, Goeijenbier M, Glisic S, *et al.*: ***In silico* analysis suggests repurposing of ibuprofen for prevention and treatment of EBOLA virus disease [version 1; referees: 2 approved].** *F1000Res.* 2015; **4**: 104.
PubMed Abstract | Publisher Full Text | Free Full Text
22. Raj U, Varadwaj PK: **Flavonoids as Multi-target Inhibitors for Proteins Associated with Ebola Virus: *In Silico* Discovery Using Virtual Screening and Molecular Docking Studies.** *Interdiscip Sci.* 2015; 1–10.
PubMed Abstract | Publisher Full Text
23. Abazari D, Moghtadaei M, Behvarmanesh A, *et al.*: **Molecular docking based screening of predicted potential inhibitors for VP40 from Ebola virus.** *Bioinformation.* 2015; **11**(5): 243–7.
PubMed Abstract | Publisher Full Text | Free Full Text
24. Nishimura H, Yamaya M: **A Synthetic Serine Protease Inhibitor, Nafamostat Mesilate, Is a Drug Potentially Applicable to the Treatment of Ebola Virus Disease.** *Tohoku J Exp Med.* 2015; **237**(1): 45–50.
PubMed Abstract | Publisher Full Text
25. De Clercq E: **Curious (Old and New) Antiviral Nucleoside Analogues with Intriguing Therapeutic Potential.** *Curr Med Chem.* 2015; **22**(34): 3866–80.
PubMed Abstract | Publisher Full Text
26. Ekins S, Freundlich JS, Coffee M: **A common feature pharmacophore for FDA-approved drugs inhibiting the Ebola virus [version 2; referees: 2 approved].** *F1000Res.* 2014; **3**: 277.
PubMed Abstract | Publisher Full Text | Free Full Text
27. Ekins S: **A pharmacophore for of Ebola active compounds - predictions searching Microsource library.** *Figshare.* 2014.
Publisher Full Text
28. Ekins S, Reynolds RC, Franzblau SG, *et al.*: **Enhancing hit identification in *Mycobacterium tuberculosis* drug discovery using validated dual-event Bayesian models.** *PLoS One.* 2013; **8**(5): e63240.
PubMed Abstract | Publisher Full Text | Free Full Text
29. Ekins S, Reynolds RC, Kim H, *et al.*: **Bayesian models leveraging bioactivity and cytotoxicity information for drug discovery.** *Chem Biol.* 2013; **20**(3): 370–378.
PubMed Abstract | Publisher Full Text | Free Full Text
30. Ekins S, de Siqueira-Neto JL, McCall LI, *et al.*: **Machine Learning Models and**

Pathway Genome Data Base for *Trypanosoma cruzi* Drug Discovery. *PLoS Negl Trop Dis.* 2015; **9**(6): e0003878.
PubMed Abstract | Publisher Full Text | Free Full Text

31. Madrid PB, Panchal RG, Warren TK, *et al.*: **Evaluation of Ebola Virus Inhibitors for Drug repurposing.** *ACS Infect Dis.* 2015; **1**(7): 317–326.
Publisher Full Text

32. Prathipati P, Ma NL, Keller TH: **Global Bayesian models for the prioritization of antitubercular agents.** *J Chem Inf Model.* 2008; **48**(12): 2362–70.
PubMed Abstract | Publisher Full Text

33. Bender A, Scheiber J, Glick M, *et al.*: **Analysis of pharmacology data and the prediction of adverse drug reactions and off-target effects from chemical structure.** *ChemMedChem.* 2007; **2**(6): 861–873.
PubMed Abstract | Publisher Full Text

34. Klon AE, Lowrie JF, Diller DJ: **Improved naïve Bayesian modeling of numerical data for absorption, distribution, metabolism and excretion (ADME) property prediction.** *J Chem Inf Model.* 2006; **46**(5): 1945–56.
PubMed Abstract | Publisher Full Text

35. Hassan M, Brown RD, Varma-O'brien S, *et al.*: **Cheminformatics analysis and learning in a data pipelining environment.** *Mol Divers.* 2006; **10**(3): 283–99.
PubMed Abstract | Publisher Full Text

36. Rogers D, Brown RD, Hahn M: **Using extended-connectivity fingerprints with Laplacian-modified Bayesian analysis in high-throughput screening follow-up.** *J Biomol Screen.* 2005; **10**(7): 682–6.
PubMed Abstract | Publisher Full Text

37. Jones DR, Ekins S, Li L, *et al.*: **Computational approaches that predict metabolic intermediate complex formation with CYP3A4 (+b_3).** *Drug Metab Dispos.* 2007; **35**(9): 1466–75.
PubMed Abstract | Publisher Full Text

37a. Ghose AK, Viswanadhan VN, Wendoloski JJ: **Prediction of hydrophobic (lipophilic) properties of small organic molecules using fragmental methods: an analysis of ALOGP and CLOGP methods.** *J Phys Chem.* 1998; **102**(21): 3762–3772.
Publisher Full Text

38. Anon. **R.** Available from: http://www.r-project.org/
Reference Source

39. Clark AM, Ekins S: **Open Source Bayesian Models. 2. Mining a "Big Dataset" To Create and Validate Models with ChEMBL.** *J Chem Inf Model.* 2015; **55**(6): 1246–1260.
PubMed Abstract | Publisher Full Text

40. Clark AM, Dole K, Coulon-Spektor A, *et al.*: **Open Source Bayesian Models. 1. Application to ADME/Tox and Drug Discovery Datasets.** *J Chem Inf Model.* 2015; **55**(6): 1231–1245.
PubMed Abstract | Publisher Full Text | Free Full Text

41. Clark AM, Sarker M, Ekins S: **New target prediction and visualization tools incorporating open source molecular fingerprints for TB Mobile 2.0.** *J Cheminform.* 2014; **6**: 38.
PubMed Abstract | Publisher Full Text | Free Full Text

42. Lamichhane G, Freundlich JS, Ekins S, *et al.*: **Essential metabolites of *Mycobacterium tuberculosis* and their mimics.** *MBio.* 2011; **2**(1): e00301–10.
PubMed Abstract | Publisher Full Text | Free Full Text

43. Ekins S, Bradford J, Dole K, *et al.*: **A collaborative database and computational models for tuberculosis drug discovery.** *Mol Biosyst.* 2010; **6**(5): 840–851.
PubMed Abstract | Publisher Full Text

44. Zheng X, Ekins S, Raufman JP, *et al.*: **Computational models for drug inhibition of the human apical sodium-dependent bile acid transporter.** *Mol Pharm.* 2009; **6**(5): 1591–1603.
PubMed Abstract | Publisher Full Text | Free Full Text

45. Ekins S, Kaneko T, Lipinski CA, *et al.*: **Analysis and hit filtering of a very large library of compounds screened against *Mycobacterium tuberculosis*.** *Mol Biosyst.* 2010; **6**(11): 2316–2324.
PubMed Abstract

46. Ekins S, Casey AC, Roberts D, *et al.*: **Bayesian models for screening and TB Mobile for target inference with *Mycobacterium tuberculosis*.** *Tuberculosis (Edinb).* 2014; **94**(2): 162–9.
PubMed Abstract | Publisher Full Text | Free Full Text

47. Ekins S, Freundlich JS, Reynolds RC: **Are bigger data sets better for machine learning? Fusing single-point and dual-event dose response data for *Mycobacterium tuberculosis*.** *J Chem Inf Model.* 2014; **54**(7): 2157–65.
PubMed Abstract | Publisher Full Text

48. Ekins S, Freundlich JS, Hobrath JV, *et al.*: **Combining computational methods for hit to lead optimization in *Mycobacterium tuberculosis* drug discovery.** *Pharm Res.* 2014; **31**(2): 414–35.
PubMed Abstract | Publisher Full Text | Free Full Text

49. Gupta RR, Gifford EM, Liston T, *et al.*: **Using open source computational tools for predicting human metabolic stability and additional absorption, distribution, metabolism, excretion, and toxicity properties.** *Drug Metab Dispos.* 2010; **38**(11): 2083–2090.
PubMed Abstract | Publisher Full Text

50. Steinbeck C, Han Y, Kuhn S, *et al.*: **The Chemistry Development Kit (CDK): an open-source Java library for Chemo- and Bioinformatics.** *J Chem Inf Comput Sci.* 2003; **43**(2): 493–500.
PubMed Abstract | Publisher Full Text

51. Ekins S, Clark AM, Williams AJ: **Incorporating Green Chemistry Concepts into Mobile Chemistry Applications and Their Potential Uses.** *ACS Sustain Chem Eng.* 2013; **1**(1): 8–13.
Publisher Full Text

52. Ekins S, Clark AM, Sarker M: **TB Mobile: a mobile app for anti-tuberculosis molecules with known targets.** *J Cheminform.* 2013; **5**(1): 13.
PubMed Abstract | Publisher Full Text | Free Full Text

53. Clark AM, Williams AJ, Ekins S: **Cheminformatics workflows using mobile apps.** *Chem-Bio Informatics J.* 2013; **13**: 1–18.
Publisher Full Text

54. Ekins S, Clark AM, Williams AJ: **Open Drug Discovery Teams: A Chemistry Mobile App for Collaboration.** *Mol Inform.* 2012; **31**(8): 585–597.
PubMed Abstract | Publisher Full Text | Free Full Text

55. Williams AJ, Ekins S, Clark AM, *et al.*: **Mobile apps for chemistry in the world of drug discovery.** *Drug Discov Today.* 2011; **16**(21–22): 928–39.
PubMed Abstract | Publisher Full Text

56. Krueger RE, Mayer GD: **Tilorone hydrochloride: an orally active antiviral agent.** *Science.* 1970; **169**(3951): 1213–4.
PubMed Abstract | Publisher Full Text

57. Stringfellow DA: **Comparation interferon- inducing and antiviral properties of 2-amino-5-bromo-6-methyl-4-pyrimidinol (U-25,166), tilorone hydrochloride, and polyinosinic-polycytidylic acid.** *Antimicrob Agents Chemother.* 1977; **11**(6): 984–92.
PubMed Abstract | Publisher Full Text | Free Full Text

58. Wissing MD, Dadon T, Kim E, *et al.*: **Small-molecule screening of PC3 prostate cancer cells identifies tilorone dihydrochloride to selectively inhibit cell growth based on cyclin-dependent kinase 5 expression.** *Oncol Rep.* 2014; **32**(1): 419–24.
PubMed Abstract | Publisher Full Text | Free Full Text

59. Biswas T, Green KD, Garneau-Tsodikova S, *et al.*: **Discovery of inhibitors of *Bacillus anthracis* primase DnaG.** *Biochemistry.* 2013; **52**(39): 6905–10.
PubMed Abstract | Publisher Full Text

60. Leppäranta O, Tikkanen JM, Bespalov MM, *et al.*: **Bone morphogenetic protein-inducer tilorone identified by high-throughput screening is antifibrotic *in vivo*.** *Am J Respir Cell Mol Biol.* 2013; **48**(4): 448–55.
PubMed Abstract | Publisher Full Text

61. Schrimpf MR, Sippy KB, Briggs CA, *et al.*: **SAR of α7 nicotinic receptor agonists derived from tilorone: exploration of a novel nicotinic pharmacophore.** *Bioorg Med Chem Lett.* 2012; **22**(4): 1633–8.
PubMed Abstract | Publisher Full Text

62. Briggs CA, Schrimpf MR, Anderson DJ, *et al.*: **alpha7 nicotinic acetylcholine receptor agonist properties of tilorone and related tricyclic analogues.** *Br J Pharmacol.* 2008; **153**(5): 1054–61.
PubMed Abstract | Publisher Full Text | Free Full Text

63. Kim K, Damoiseaux R, Norris AJ, *et al.*: **High throughput screening of small molecule libraries for modifiers of radiation responses.** *Int J Radiat Biol.* 2011; **87**(8): 839–45.
PubMed Abstract | Publisher Full Text | Free Full Text

64. Ratan RR, Siddiq A, Aminova L, *et al.*: **Small molecule activation of adaptive gene expression: tilorone or its analogs are novel potent activators of hypoxia inducible factor-1 that provide prophylaxis against stroke and spinal cord injury.** *Ann N Y Acad Sci.* 2008; **1147**: 383–94.
PubMed Abstract | Publisher Full Text | Free Full Text

65. Mayer-Sonnenfeld T, Avrahami D, Friedman-Levi Y, *et al.*: **Chemically induced accumulation of GAGs delays PrP^{Sc} clearance but prolongs prion disease incubation time.** *Cell Mol Neurobiol.* 2008; **28**(7): 1005–15.
PubMed Abstract | Publisher Full Text

66. Wolfe MS: **Giardiasis.** *Clin Microbiol Rev.* 1992; **5**(1): 93–100.
PubMed Abstract | Free Full Text

67. Okombo J, Kiara SM, Mwai L, *et al.*: **Baseline *in vitro* activities of the antimalarials pyronaridine and methylene blue against *Plasmodium falciparum* isolates from Kenya.** *Antimicrob Agents Chemother.* 2012; **56**(2): 1105–7.
PubMed Abstract | Publisher Full Text | Free Full Text

68. Rizk MA, El-Sayed SA, Terkawi MA, *et al.*: **Optimization of a Fluorescence-Based Assay for Large-Scale Drug Screening against *Babesia* and *Theileria* Parasites.** *PLoS One.* 2015; **10**(4): e0125276.
PubMed Abstract | Publisher Full Text | Free Full Text

69. Qi J, Wang S, Liu G, *et al.*: **Pyronaridine, a novel modulator of P-glycoprotein-mediated multidrug resistance in tumor cells *in vitro* and *in vivo*.** *Biochem Biophys Res Commun.* 2004; **319**(4): 1124–31.
PubMed Abstract | Publisher Full Text

70. Anon. **Pyramax® (*pyronaridine artesunate*).** Available from: http://www.mmv.org/access-delivery/access-portfolio/pyramax®-pyronaridine-artesunate
Reference Source

71. Poravuth Y, Socheat D, Rueangweerayut R, *et al.*: **Pyronaridine-artesunate versus chloroquine in patients with acute *Plasmodium vivax* malaria: a randomized,**

double-blind, non-inferiority trial. *PLoS One.* 2011; **6**(1): e14501.
PubMed Abstract | Publisher Full Text | Free Full Text

72. Ekins S, Williams AJ, Krasowski MD, *et al.*: ***In silico* repositioning of approved drugs for rare and neglected diseases.** *Drug Discov Today.* 2011; **16**(7–8): 298–310.
PubMed Abstract | Publisher Full Text

73. Martínez-Romero C, García-Sastre A: **Against the clock towards new Ebola virus therapies.** *Virus Res.* 2015; pii: S0168-1702(15)00236-1.
PubMed Abstract | Publisher Full Text

74. Seidler J, McGovern SL, Doman TN, *et al.*: **Identification and prediction of promiscuous aggregating inhibitors among known drugs.** *J Med Chem.* 2003; **46**(21): 4477–4486.
PubMed Abstract | Publisher Full Text

75. Barelier S, Sterling T, O'Meara MJ, *et al.*: **The recognition of identical ligands by unrelated proteins.** *ACS Chem Biol.* 2015.
PubMed Abstract | Publisher Full Text

76. Ekins S, Williams AJ: **Finding promiscuous old drugs for new uses.** *Pharm Res.* 2011; **28**(8): 1785–1791.
PubMed Abstract | Publisher Full Text

Open Peer Review

Current Referee Status: ✔ ✔

Version 3

Referee Report 18 January 2017

doi:10.5256/f1000research.11510.r19459

✔ **Sanja Glisic**
Center for Multidisciplinary Research, Vinča Institute of Nuclear Sciences, University of Belgrade, Belgrade, Serbia

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Referee Report 03 December 2015

doi:10.5256/f1000research.7773.r11173

✔ **Sandeep Chakraborty**
Plant Sciences Department, University of California, Davis, CA, USA

Ekins *et al.* have presented a crisp and lucid manuscript on a very relevant topic. They have presented a methodology that implements machine learning techniques to learn from known active and inactive compounds (an ever increasing set, that will tend to provide improved results as time goes by), and score a larger set of compounds (MicroSource Spectrum set of 2320 compounds). The *in silico* methodology described here provides an excellent method to quickly screen known compounds for possible therapies (against Ebola in particular), and other viruses in general. Finally, they demonstrate (*in vitro*) the increased effectiveness of three compounds - the antiviral tilorone and the antimalarials quinacrine and pyronaridine - in comparison to the known active chloroquine (albeit at a higher cytotoxicity) in inhibiting viral infection of HeLa cells. Further, their efforts in ensuring open-access to such tools is commendable as the next pathogen caused humanitarian crisis looms in some nations.

The biggest open question is how good are the molecular descriptors, and how much of this is serendipitous. For example, I find it hard to believe that molecular weight and the number of rotatable bonds can be good predictors of drug-protein interactions (although I may be wrong). I have been investigating promiscuous ligand protein interactions for a couple of years on a molecular basis (http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0032011,

http://f1000research.com/articles/2-286/v3). One interesting example (unpublished) is suramin used in the treatment of African sleeping sickness (African trypanosomiasis) and river blindness (onchocerciasis), infections caused by parasites. Suramin binds eight non-homologous proteins in the PDB database, through different parts of the molecule and in binding sites that share little similarity in residues involved. Also, the molecule (in addition to the protein) undergoes conformational changes, underscoring the difficulty of computational methods to model such interactions. In the face of such data, the m/c learning models appear too simplistic.

Also, it is not completely clear why the 23, 31 and 34th compound was chosen from the Table S5, which is ordered on column H (all three have amines, don't the previous ones have it?).

Some minor comments:

1. It would be an interesting case study to evaluate how favipiravir (which I understand is yet to be clinically approved in the US, but approved in Japan) and BCX4430 would rank through the m/c learning methodology.

2. It would be good to have a set of images, and the corresponding nuclei counts obtained from CellProfiler. Is there a way to quantify the color green as a measure of viral infection?

3. AlogP as a molecular descriptor has not been explained (page 3).

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 19 Dec 2015

Sean Ekins, Collaborations in Chemistry, USA

We thank the reviewer for their constructive feedback.

The approach we have taken uses FCFP_6 fingerprints as well as 8 interpretable descriptors, and therefore the models do not depend on molecular weight and rotatable bond number. On the whole this approach has been remarkable useful for predicting whole cell activity as we described for mycobacterium tuberculosis, T. Cruzi and now Ebola. In all cases we are not considering a single target. This suggests the machine learning approach and descriptors used can discern active molecule features from those that are inactive and identify new molecules.

The three compounds were chosen as those above them in the list were either compounds in the training set or antipsychotics and other CNS acting compounds etc. which were deemed less desirable.

The model could certainly be used to predict additional molecules. These two suggested by the reviewer are structurally distinct from any of the actives in our current training set. We didn't identify any of the classical antiviral polymerase-looking compounds in our screening against Ebola. However we have previously collated and described many other diverse compounds active against Ebola *in vitro* as described in this manuscript. Perhaps the next step would be to utilize all of the different HTS screening data to build a combined model that considers this structural diversity and may overcome limitations in the current models.

Response: we have now added a new figure with cell images (S7).

Response: AlogP is a widely used measure of hydrophobicity and we have refered to its use along with the other descriptors in the previous machine learning papers.

Competing Interests: No competing interests were disclosed.

Referee Report 10 November 2015

doi:10.5256/f1000research.7773.r10995

Sanja Glisic
Center for Multidisciplinary Research, Vinča Institute of Nuclear Sciences, University of Belgrade, Belgrade, Serbia

Ekins and his colleagues by using machine learning models and molecular modeling have successfully identified from collection of 2320 compounds 3 promising anti Ebola compounds with *in vitro* nanomolar activity. It is a perfect example which confirms suitability of *in silico* approaches in selection of molecules against Ebola virus.

This result will be strengthened with suggestion of possible therapeutic target(s) of selected candidate drugs by using resources of curated chemistry-to-protein relationships. Such information could help in further improvement of proposed therapeutic molecules, as well as for selection of some other candidates for Ebola drugs.

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 19 Dec 2015

Sean Ekins, Collaborations in Chemistry, USA

We thank the reviewer for their constructive feedback

Response: We are unsure which resources the reviewer is referring to and how these would help identify the antiviral target. We have tried to extensively describe the known activities of the three compounds against various targets outside of viruses. In addition we have previously suggested such antimalarials with Ebola activity may dock into VP35. Preliminary docking results suggest pyronaridine may dock into the same site which is also indicated by the pharmacophore provided already (fig 4). We have added the statement “Future work may include identification of targets using computational or experimental approaches.” to the discussion.

Competing Interests: No competing interests were disclosed.