CORTEX DISCOVERY

Building new foundations for drug development

Cortex Virtual HTS

Cortex Discovery offers *in silico* virtual High Throughput Screening in any therapeutic area using its unique deep learning system.

Your target - Our AI acquires its broad knowledge of biochemistry from our curated database of about 260 millions measurements on more than 2000 properties. If our AI is not acquainted with your target, we just need data from about 1000+ compounds (incl. inactives) to rapidly produce a new extended model that is able to predict it as well. (We don't need to know the identity of a target to predict it, and we do not keep or re-use the specialized model or data.)

Lab-proven accuracy - Our *in silico* predictions have been demonstrated in the lab to have precision and sensitivity *on par with that of repeat experiments*.

Large chemical space - Our system can make an exhaustive search through vast chemical spaces that include millions of purchasable molecules and billions of synthesizable compounds that would be impossible to access otherwise.

Qualified and diverse hits - Cortex's AI returns a high-value diverse shortlist of virtual qualified hits. We can also simultaneously optimize dozens of properties out of the thousands that the AI predicts (target, off-targets, toxicity etc.) according to an advanced filter fine-tuned for your project by our team.

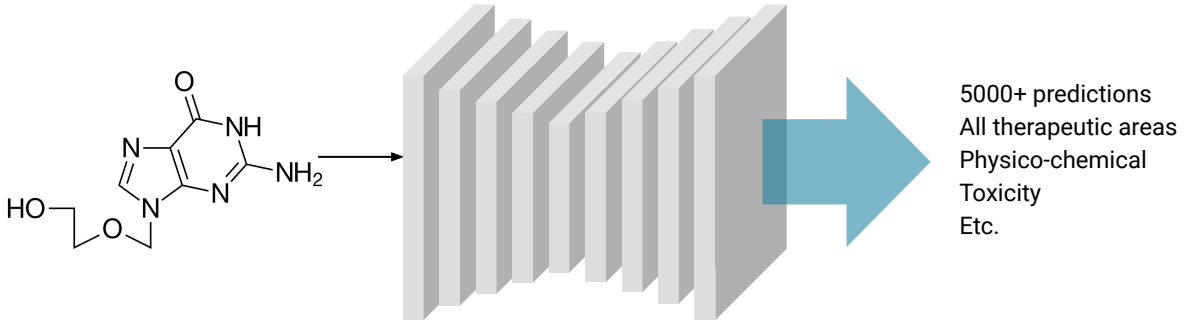
The Technology Behind Cortex Virtual HTS

Our approach emphasizes neural net architectures based on first principles, and training on large and comprehensive datasets. This enables our neural nets to work directly on the raw molecule structures and to gain a global understanding of biochemistry, free of biases or oversimplifications inherent in older methods.

Our tech includes custom in-house implementations of covariant, relational/graph and attention-based neural nets, as well as innovative metrics and visualisations to characterize prediction performances.

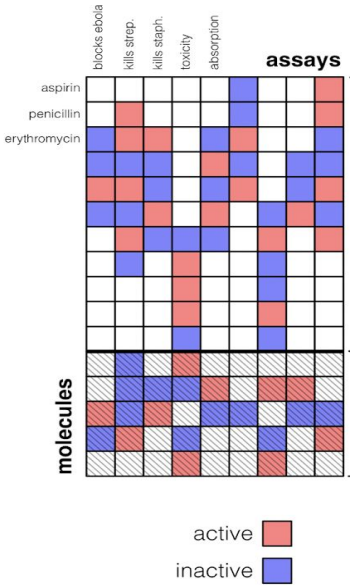
We provide a rigorous statistical characterization of the accuracy and sensitivity of our predictions on held-out portions of the data, and each individual prediction comes with its own a **confidence interval**, so that you know exactly how much to trust and use each result.

Cortex's Deep Graph Neural Net

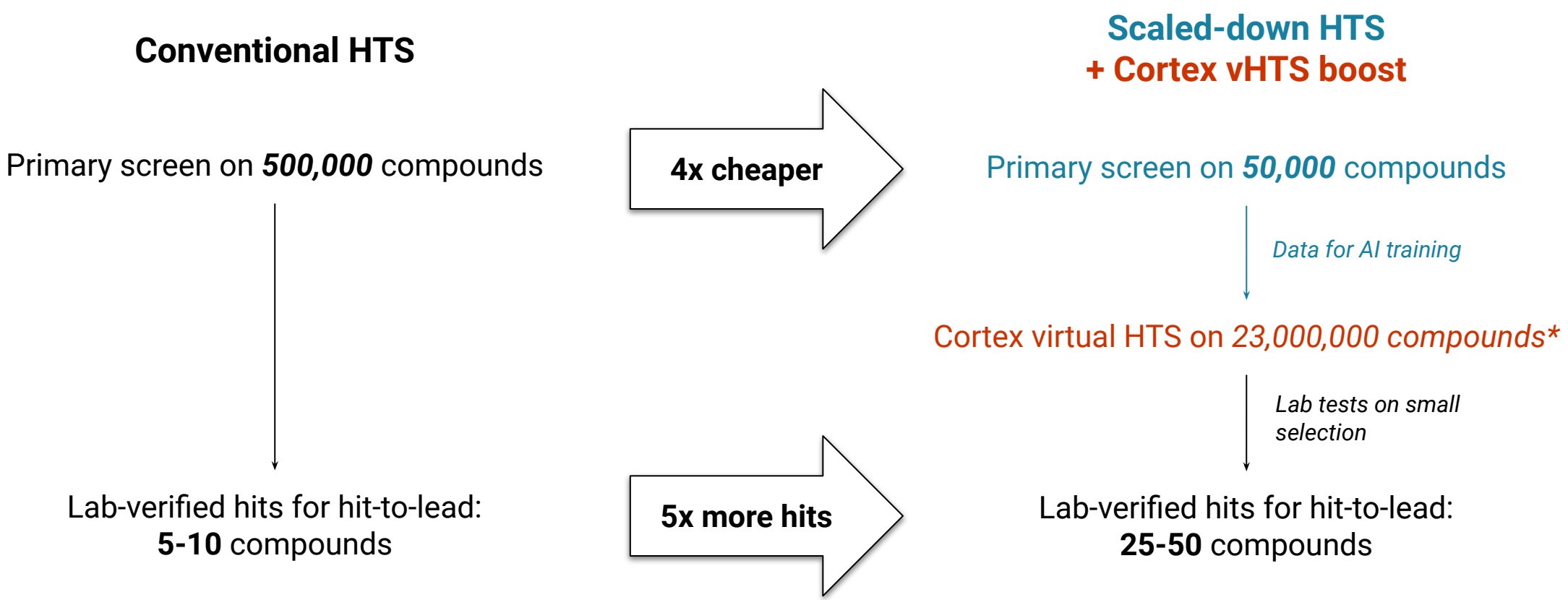


Cortex's training database

- 2000+ assays
- 2,500,000 unique molecules
- 260,000,000 data points
  + project-specific data



Large Scale vHTS Example: a 20x Cost Reduction per Hit



6 months
Cost to HTS ~ € 1M
Value of hits ~ € 1M

2 months
Cost to HTS ~ € 250k
Value of hits ~ € 5M

The Value of Precision

Cortex specializes on making *accurate* predictions, quantified here using AUC: a measure robust to differences in scoring system, choice of decision threshold, or test set composition.

Each increment in accuracy translates to manifold more real hits.

	Virtual HTS prediction accuracy (AUC*)	**Verified Hits** each worth €50k-200k
	90%	8-16
Value used in previous slide's estimate →	**95%**	**25-50**
	97%	45-90
	98%	65-130
	99%	100-200

Cortex prediction accuracies are on par with experiments, with AUCs typically between 90% and 99% (measured on held-out data or external dataset).

Cortex's system excels on highly selective datasets, and is able to predict novel hits.

*__AUC__: a standard robust measure of predictions accuracy (probability that an active compound is scored higher than an inactive one).

Cortex's Compounds Library: An Organised Chemical Space

Cortex's molecule libraries and tools enables fast, broad, and cost-effective hit discovery

All relevant physical and virtual compound libraries are instantly available for virtual screening

- Exhaustive screening of **23M purchasable compounds** from reliable providers such as Enamine, LifeChem, ChemDiv, and more.
- Combinatorial ligand structure optimization in virtual libraries defined by building blocks and reactions.
    - E.g., **50B compounds** from the REAL Enamine virtual database.
- Our powerful **neural fingerprints** help map, organize and search through arbitrary chemical spaces.

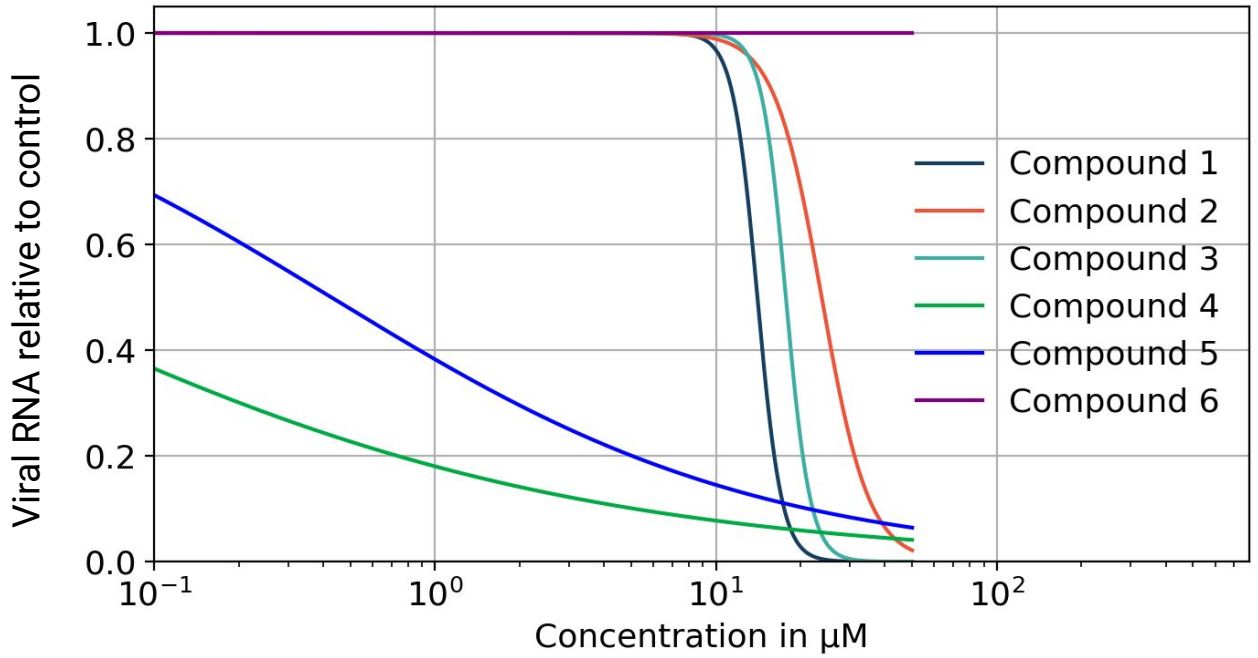
Our libraries are organized by potential compound approval speed

- **FDA-approved natural products:** direct inclusion in nutraceutical and cosmetic products.
- **Natural compounds:** a significant proportion will lead to fast clinical studies that will not require FDA approval.
- **Existing or experimental drugs:** drug repurposing bypasses most preclinical and clinical phase I experiments.
- **Physical libraries:** faster drug discovery - quick order with the option to choose the library containing most hits
- **Virtual libraries:** faster drug discovery - with the guarantee that hits are synthesizable in a short timeframe.

In vitro confirmation of new COVID19 drug predictions

Cortex's predicted *new* anti-COVID19 drug candidates whose activity was confirmed in vitro.

5/6 predicted hits showed activity at non-toxic doses, of which two are highly potent and verified in multiple cell types*.



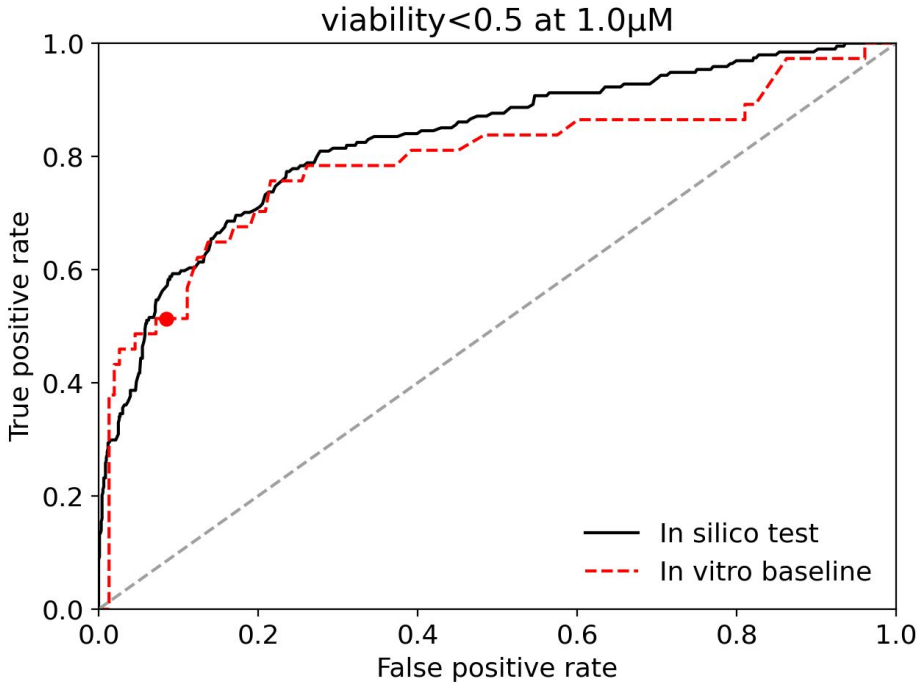
AARHUS UNIVERSITY

STANFORD UNIVERSITY

*Collaboration with **Dr. Olagnier, Aarhus University**. Another in vitro validation with a different method on different novel antiviral compounds was successfully conducted with a laboratory in **Stanford University**, identifying two more highly potent compounds.

Cancer Cell Viability Prediction Validation with High-Content Imaging

- Training set data with **cancer cell viability** from high-content imaging data
- Prediction tested on withheld compounds (unpublished data hidden from us and only known by our collaborator*)



viability threshold	*reproducibility AUC* (%)*	*predictions AUC* (%)*
0.3	87.3 ±1.8	**85.1** ±0.8
0.5	78.0 ±1.9	**81.3** ±0.8

* Pr. Kaylene Simpson, Head of the Victorian Centre for Functional Genomics, Peter MacCallum Cancer Centre, Australia.

VCFG
Victorian Centre For Functional Genomics

*__AUC__: a standard robust measure of predictions accuracy (probability that an active compound is scored higher than an inactive one).

HIV Predictions Validated in Molecular Assay

External *in vitro* validation of predicted GP41 (HIV-1) inhibitors.

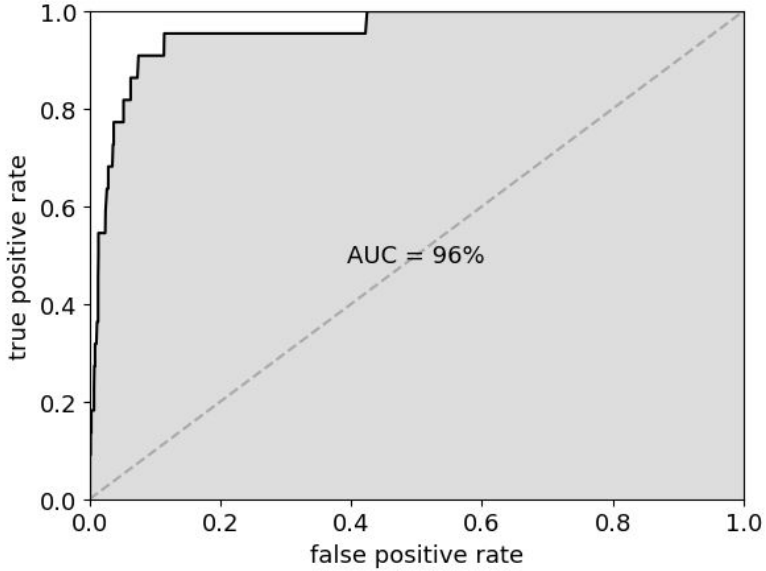
- Our neural net was trained on a PubChem assay with 0.3% activity rate.
- We picked the available small library (1950 new compounds) with most predicted hits.
- We *committed* our ranked top 10% predictions prior to the experiment for independent confirmation by NCMM and SINTEF.

➔ 22 hits found in library

➔ 20/22 hits in our top 10% predictions *

➔ 2 hits in our top 3 predictions

➔ AUC ** = **95.6** ±1.9 (%)



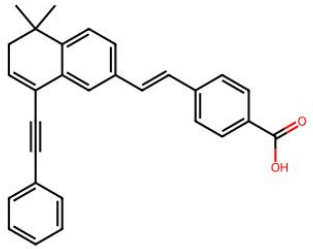
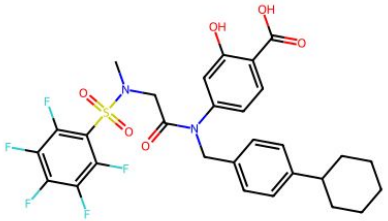
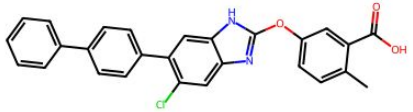
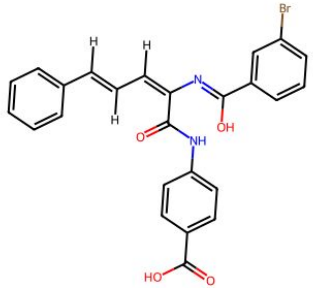
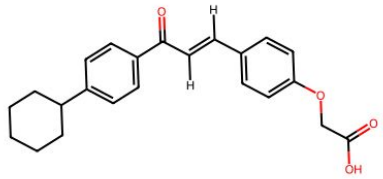
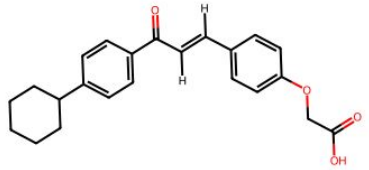
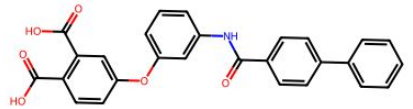
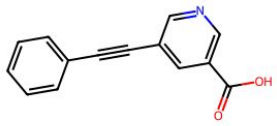
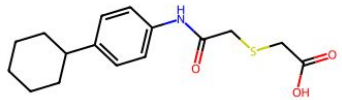
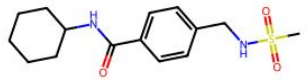
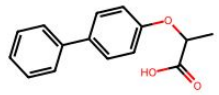
NCMM
Centre for Molecular Medicine Norway
Nordic EMBL partnership for Molecular Medicine

SINTEF

* Probability of this happening by chance (p-value from random predictions): 1 in 10^{18}

****AUC**: a standard robust measure of predictions accuracy (probability that an active compound is scored higher than an inactive one).

Cortex's AI Predicts Novel Hits

Predicted active				
Closest active in training set*				
	20% similar	*21% similar*	*17% similar*	*25% similar*
Closest inactive in training set*				
	27% similar	*27% similar*	*26% similar*	*26% similar*

* Using Tanimoto similarity of Morgan fingerprints

Predictions as Accurate as *In Vitro* Assays

Our predictions can even be more accurate than the data we learned from. This is because our neural network is capable of discovering general laws governing the activity of chemicals, which it uses to identify outliers.

***In vitro* AUC**: measurements from an assay A are compared to another assay B (reproducibility).
Cortex AUC: our neural net trained on A made predictions for new molecules tested in B.

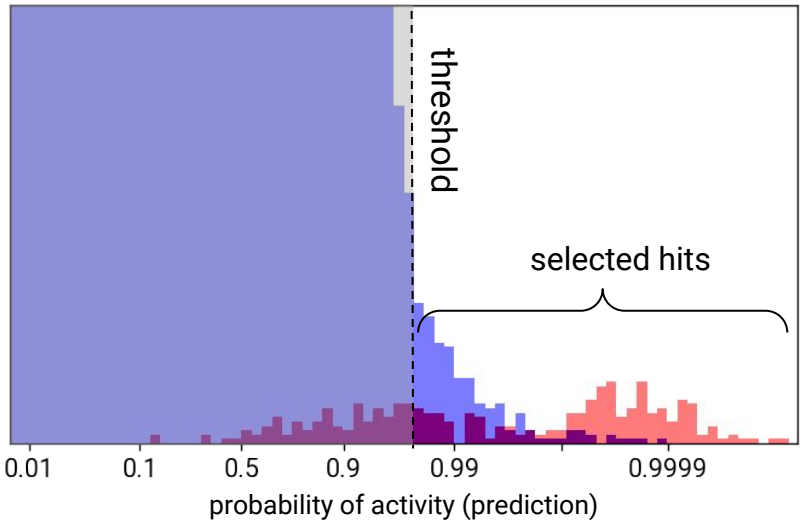
			***In vitro* accuracy** *(Exp. reproducibility)*	***In silico* accuracy** (Cortex)
Target	**Assay A**	**Assay B**	**AUC* %**	**AUC* %**
CYP 2C9 inhib.	AID777	AID1645842	71.8 ±7.9	**90.0** ±0.6
CYP 2C9 inhib.	AID883	AID1645842	81.1 ±2.5	**88.2** ±0.7
CYP 2D6 inhib.	AID891	AID1645840	**93.7** ±1.2	89.3 ±0.6
CYP 3A4 inhib.	AID884	AID1645841	89.0 ±1.2	**90.3** ±0.6
genotoxicity	AID504466	AID493106	91.7 ±4.0	**98.5** ±1.1
luciferase inhib.	AID588342	AID624030	**99.1** ±0.8	96.0 ±1.3

*__AUC__: a standard robust measure of predictions accuracy (probability that an active compound is scored higher than an inactive one).

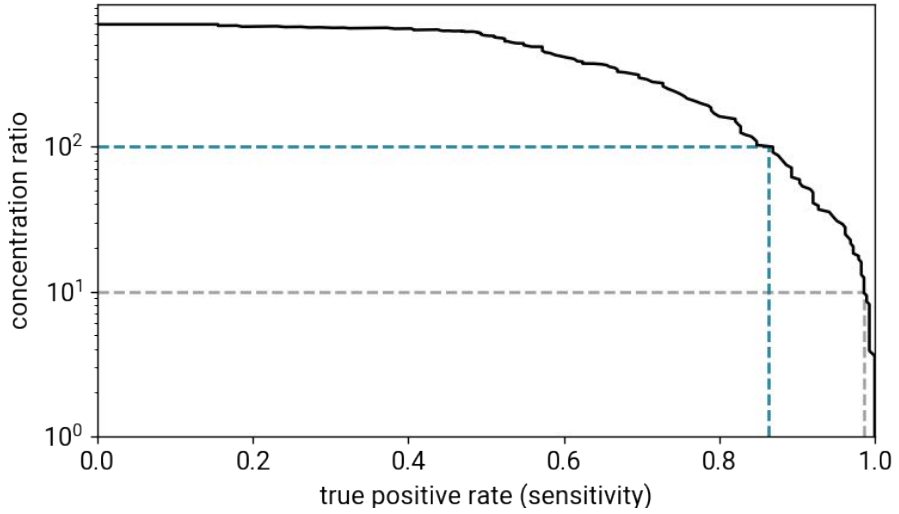
Cortex Virtual HTS: Hit Identification

Our system takes advantage of large quantities of negative examples, producing models with demonstrably high enrichment power. In this example, our virtual screening produces a 10x smaller sample of leads containing 99% of all new active compounds in the test population (or 100x for 86%).

Selective lethality to parasite Giardia Lamblia
(predicted activity scores on test set)*

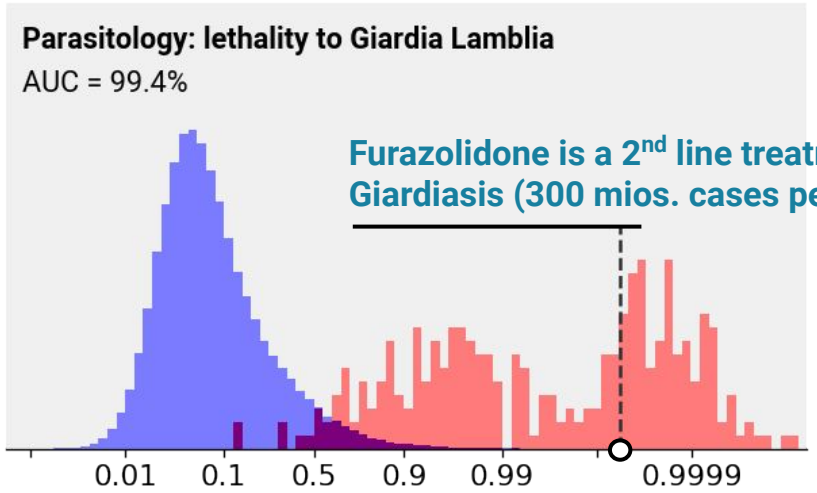


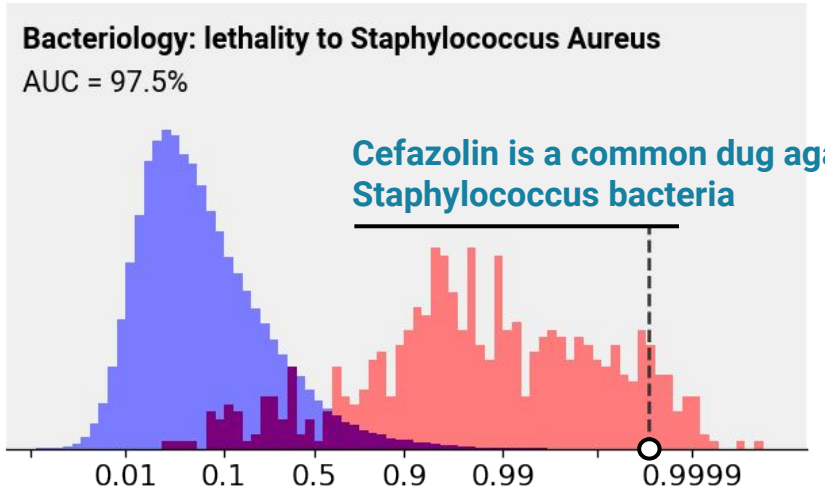
Hit enrichment and sensitivity
for all decision thresholds

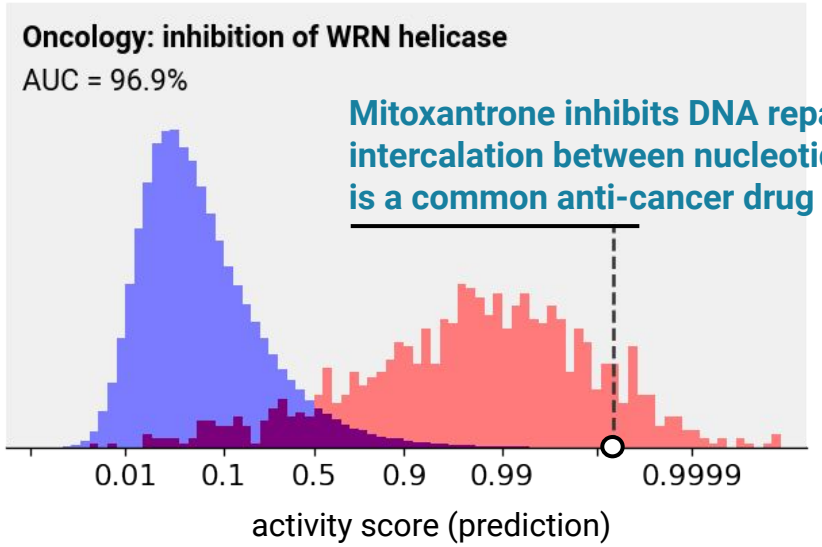


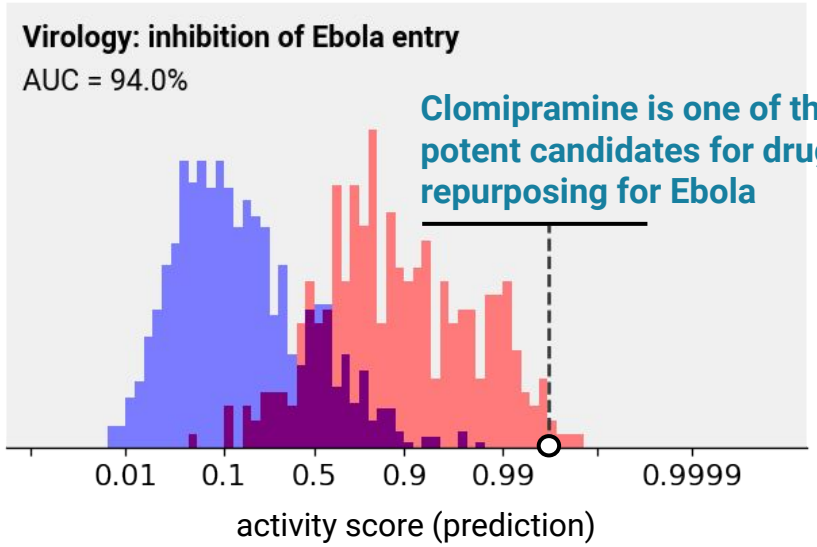
* **Blue**: inactive compounds, **Red**: active compounds

Cortex Virtual HTS: Hit Identification









* **Blue**: inactive compounds, **Red**: active compounds

A wide scope of applications

Our technology is not bound to any specific therapeutic area and its performance only depends on the amount and quality of the training data.

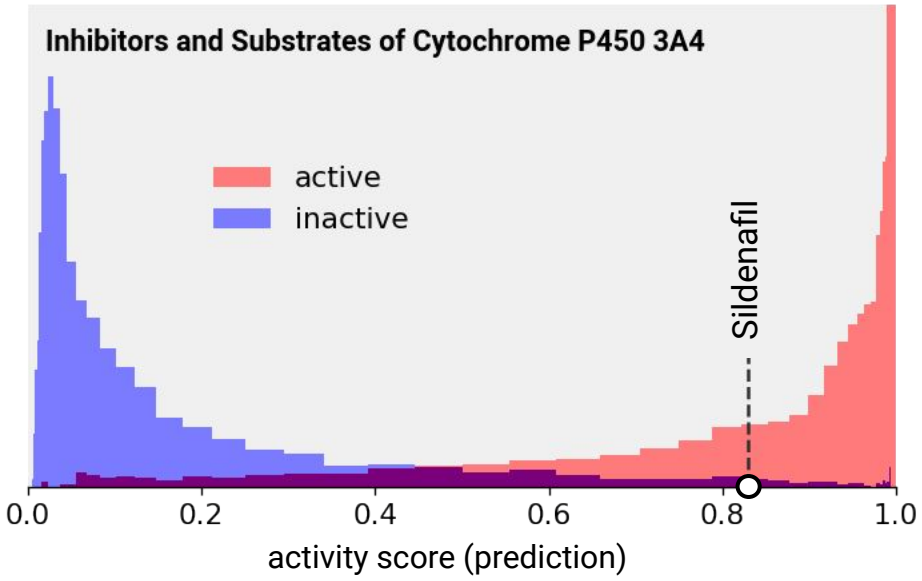
As examples, our current data gives us particularly high accuracy in the following areas:

- **Virology**: SARS-CoV-2, MERS, HIV, HCV, Influenza, Ebola, Dengue, Zika, West Nile
- **Bacteriology**: Salmonella, Staphylococcus, Tuberculosis, E. Coli
- **Parasitology**: Malaria, Chagas (Trypanosoma cruzi), Giardiasis (Giardia lamblia)
- **Pathology**: Cancer (general mechanisms), Myeloid Leukemia, Alzheimer, Parkinson, Pompe disease, Metabolic disorders, Diabetes and Cardiovascular disease

Cortex Virtual ADMET: Hit-to-Lead

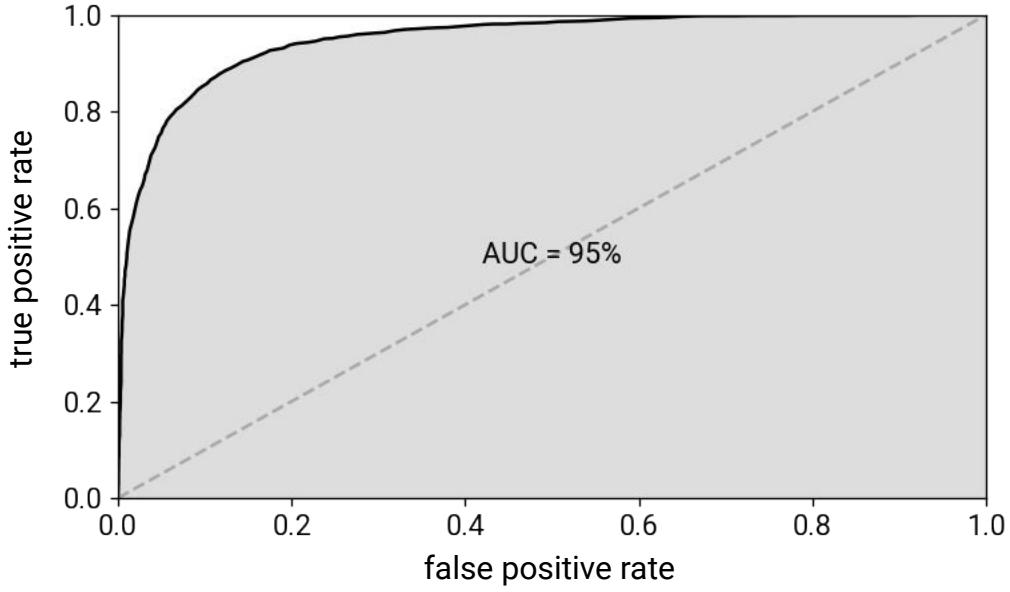
Our system encompasses cytochrome interaction predictions among other ADMET properties. Knowing which enzyme metabolizes a drug may be vital for prioritizing Hit-to-Lead activities.

Predicted activity scores on test set*



Our models are rigorously tested on held out data. Here it correctly predicts that Viagra is metabolized by CYP3A4.

Predictive power: area under ROC curve (AUC)

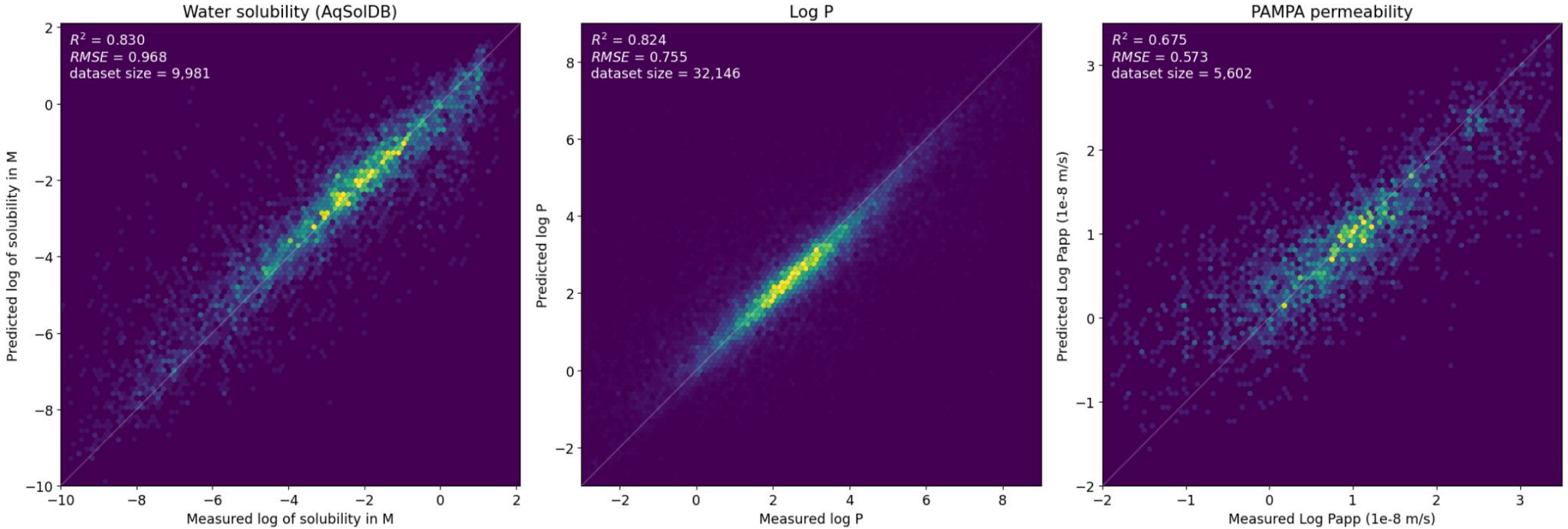


The ROC-AUC metric shows high overall predictive power for the interaction of compounds with CYP3A4.

* **Blue**: inactive compounds, **Red**: active compounds

Cortex Virtual ADMET: Regression Examples

Cumulative point clouds for measured vs predicted values on test sets (multiple folds)



Cortex Virtual ADMET: Current Offering

		AUC*
Physicochemical	Log P hydrophobicity	98%
	Aqueous solubility	96%
	DMSO solubility at 10 mM	83%
ADE	PAMPA permeability	92%
	Caco-2 permeability	88%
	Fraction unbound in plasma	93%
	Half-life in plasma	86%
	Half-life in liver microsome	90%
	Solubility in FaSSIF	82%
Toxicity	Cytotoxic	97%
	Genotoxic	97%
	hERG inhibitor	96%

		AUC*
Metabolism	CYP2A9 activator	93%
	CYP3A4 activator	94%
	CYP1A2 inhibitor	92%
	CYP2C9 inhibitor	93%
	CYP2C19 inhibitor	92%
	CYP2D6 inhibitor	93%
	CYP3A4 inhibitor	96%
	CYP19A1 inhibitor	92%
	AhR activator	94%
	CAR agonist	94%
	CAR antagonist	86%
	PXR activator	94%
	PXR agonist	94%
	GSTO1 inhibitor	94%
	ALDH1A1 inhibitor	87%

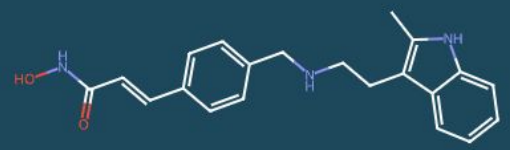
*__AUC__: a standard robust measure of predictions accuracy (probability that an active compound is scored higher than an inactive one).

Cortex Virtual ADMET: Report Example

Cortex Discovery - Model 1.18i

PANOBINOSTAT

O=C(NO)\C=C\c1ccc(cc1)CNCCc3c2ccccc2[nH]c3C



Physicochemical

target		predictions [1]	other info
Log P	log P =	1.8 ± 0.3	
	log P < 5 with prob	0.998	prior = 0.904
Solubility in water	S =	$10^{-3.3 \pm 0.6}$ mol/l	E(S) = 1.5 mmol/l
	S > 10 mg/l with prob	0.945	prior = 0.764

ADE

target		predictions [1]	other info
Solubility in FaSSIF	S =	$10^{1.7 \pm 0.3}$ mg/l	E(S) = 55 mg/l
	S > 10 mg/l with prob	0.943	prior = 0.738
PAMPA permeability	P_{app} =	$10^{0.9 \pm 0.5}$ nm/s	E(P_{app}) = 14 nm/s
	P_{app} < 50 nm/s with prob	0.808	prior = 0.423
Caco-2 A→B permeability	P_{app} =	$10^{1.4 \pm 0.3}$ nm/s	E(P_{app}) = 30 nm/s
Fraction unbound in plasma	f_U =	$10^{-1.2 \pm 0.2}$	E(f_U) = 0.073
	f_U < 0.1 with prob	0.845	prior = 0.563
Half-life in plasma	$t_{½}$ =	$10^{1.7 \pm 0.2}$ hr	E($t_{½}$) = 53 hr
	$t_{½}$ > 15 hr with prob	0.808	prior = 0.246
Half-life in liver microsome	$t_{½}$ =	$10^{1.6 \pm 0.2}$ min	E($t_{½}$) = 42 min

Metabolism

target		predictions [1]	other info	confidence [2]
CYP2D6	antagonist with prob	0.902	prior = 0.417	
	max efficacy =	1.0 ± 0.2	hill slope = 0.61	
	pEC_{50} =	5.2 ± 0.6		
CYP3A4	antagonist with prob	0.852	prior = 0.495	
	max efficacy =	0.8 ± 0.2	hill slope = 1.0	
	pEC_{50} =	4.9 ± 0.5		
CYP19A1	antagonist with prob	0.707	prior = 0.048	
	pEC_{50} =	5.2 ± 0.3		
CAR	inhibitor with prob	0.574	prior = 0.031	
	max efficacy =	0.8 ± 0.3		
	pEC_{50} =	5.9 ± 0.2		
PXR	agonist with prob	0.650	prior = 0.252	

Toxicity

target		predictions [1]	other info	confidence [2]
hERG	inhibitor with prob	0.554	prior = 0.142	
	max efficacy =	0.91 ± 0.14	hill slope = 0.67	
	pEC_{50} =	5.1 ± 0.4		
IEC-6 cells	cytotoxic at 5 μM with prob	0.597	prior = 0.008	
HEK293 cells	cytotoxic with prob	0.878	prior = 0.096	
	max efficacy =	0.98 ± 0.16	hill slope = 1.6	
	pEC_{50} =	4.21 ± 0.17		

[1] Predictions are bayesian posteriors which are bernoulli, normal or huber distributions. Confidence intervals specified to ±σ.

[2] Based on posterior entropy for prior of 0.5

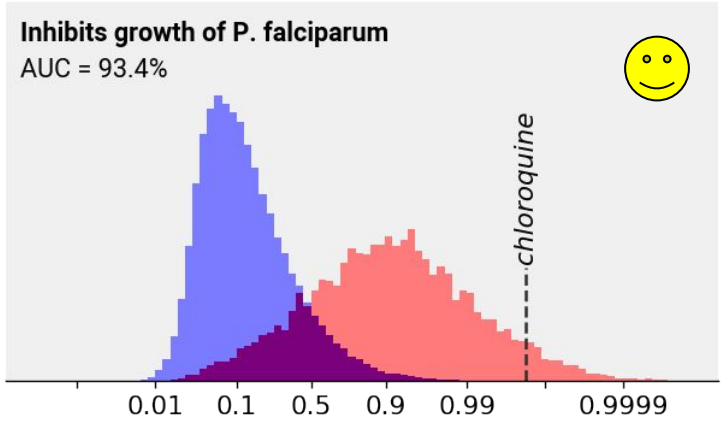
Cortex Virtual Off-Target: Lead Optimization

The solution to avoiding unwanted side-effects

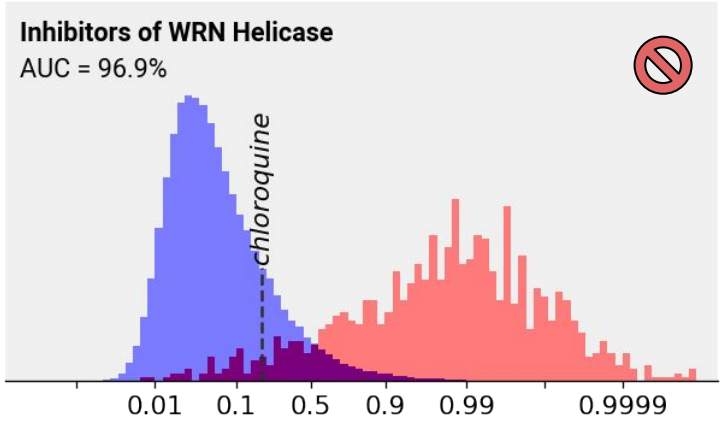
Ideally a drug's activity should be selective for a chosen therapeutic target, with no off-target activity potentially leading to adverse effects.

This example shows our system correctly predicts that Chloroquine has antimalarial activity and no significant off target activity in key systems such as DNA repair, the nervous system, and liver function.

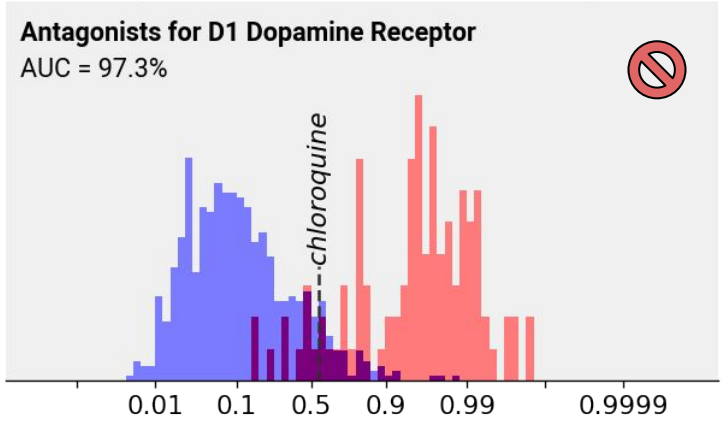
Target: anti-malarial activity?



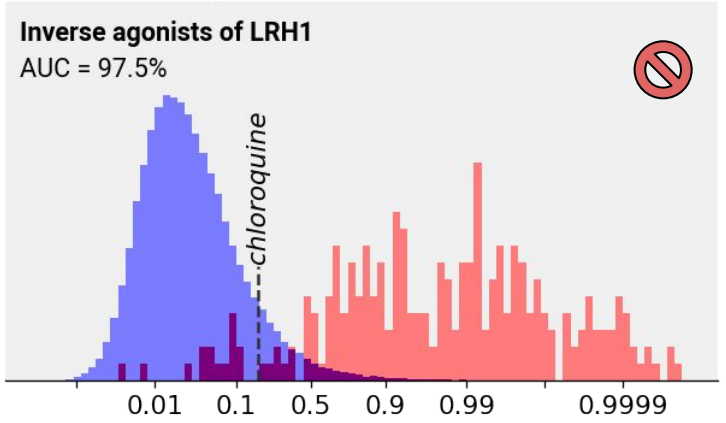
Off-target: affects human DNA repair?



Off-target: affects human nervous system?



Off-target: affects liver function?



* **Blue**: inactive compounds, **Red**: active compounds

The Cortex Team

Dr. Ivan de Weber, PhD
Co-founder & CEO

Scientist and entrepreneur behind pioneering concepts in ageing genetics, Ivan's quest for faster drug development led to the creation of Cortex Discovery.

Dr. Cédric Bény, PhD
Co-founder & CTO

Cédric applied his expertise from 20 years of research experience on complex quantum systems and information theory to build Cortex' AI technology.

Dr. Siham Ceballos, PhD, MBA
Business development advisor

Siham brings 20 years of leadership experience in the Biotech and Pharma industry, including Pfizer, Celgene, Novartis, Biogen, Alnylam, and Rejuveron Life Sciences.

Dr. Vladimir Chorošajev, PhD
Machine learning research

With a vast research expertise in open quantum systems and theoretical spectroscopy, Vladimir has an extensive industry experience in deep learning R&D.

Dr. Gaspar Pinto, PhD
Computational chemistry research

Directing projects ranging from computational chemistry to protein engineering, Gaspar has more than 10 years of research experience at the interface between chemistry and biology.

Collaboration Options

1. Fee for service

Virtual HTS and virtual ADMET.

2. Milestone-based partnerships for hit discovery and lead optimization

Prediction of target activation, pharmacological properties, off-target effects, and screen of virtual hit analogs.

3. Pipeline co-development with shared IP

Cortex virtual hits pipeline to be validated *in vitro* by partner, and subsequent lead generation & optimization processes through the collaboration.

CORTEX DISCOVERY

contact@cortexdiscovery.com