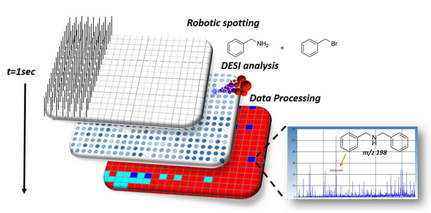
Detection of Molecules Interfering with High-Throughput Screening Assay Technologies

IUPAC Conference 2019

Laurianne David[a,b] ✉, J.W.M. Nissink[c], N. Sturm[d], J. Walsh[e], I. Feierberg[f], J. Bajorath[b] , O. Engkvist[a]

[a] Hit Discovery, Discovery Sciences, R&D BioPharmaceuticals, AstraZeneca, Gothenburg, Sweden, laurianne.david1@gmail.com, [b] Department of Life Science Informatics, B-IT, LIMES Program Unit Chemical Biology and Medicinal Chemistry, Rheinische Friedrich-Wilhelms-Universität Bonn, Germany [c] Chemistry, Oncology, R&D BioPharmaceuticals, AstraZeneca, Cambridge, CB4 OWG, UK, [d] Data Science and AI, Drug Safety & Metabolism, R&D BioPharmaceuticals, AstraZeneca, Gothenburg, Sweden [e] Hit Discovery, Discovery Sciences, R&D BioPharmaceuticals, AstraZeneca, Cambridge, UK, [f] Hit Discovery, Discovery Sciences, R&D BioPharmaceuticals, AstraZeneca, Boston, US



- A known **issue in High-Throughput Screening** (HTS) is the presence of **frequent-hitters**[1] (FHs). FHs include problematic (e.g., reactive, impure or aggregating) compounds, promiscuous compounds, and **compounds that are interfering with the assay technology** (**CIATs**) and are therefore false actives across many assays. Such interference may arise through e.g., inhibition of a coupled enzyme in the assay or fluorescence quenching properties. FHs are often investigated in follow-up studies at the expense of other experiments, thus impeding research and wasting time and resources.
- Herein we present a new **machine-learning model** based on compound's structural properties and aiming to predict CIATs and NCIATs (non-interfering compounds). We compare this model to two already published methods, the Binomial Survivor Function Score[2] (BSF) and the PAINS substructure filters[3]. These models are applied to three popular HTS technologies (AlphaScreen, FRET and TRF).

Assay Cascade

Primary Assay
- Technology A
- Target A
- Single concentration
- **Aim: identify potential hit for follow-up analysis**

Confirmatory dose response Assay
- Technology A
- Target A
- Multiple concentrations
- Aim: analyse hit potency

Orthogonal Assay
- Technology B
- Target A
- Single concentration
- Aim: identify true active compounds

Counter-screen Assay
- Technology A
- Target B or no target
- Single or multiple concentrations
- **Aim: identify CIATs**

Methods of CIATs Identification

Statistically-based model: Binomial Survivor Function (BSF)

Compounds	Technology	# Time tested (N)	# Time active (a)	Hit Rate over all assays (h)
XX0		12	12	
XX1	**AlphaScreen**	3	2	0.01
XX2		14	1	

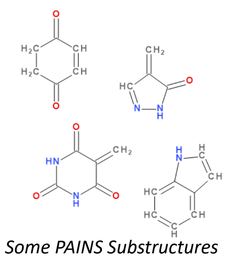
$$pBSF = -\log \sum_{a=A}^{N} \binom{N}{a} h^a (1-h)^{N-a}$$

Is CIAT when pBSF > 2

Compound	pBSF	Is CIAT
XX0	24	Yes
XX1	3.5	Yes
XX2	0.88	No

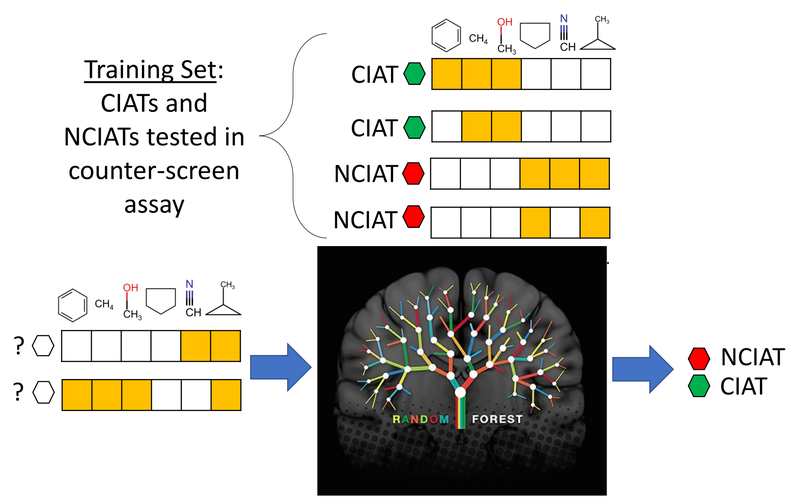
Well known structurally-based model: PAINS

- PAINS refers to '*Pan-Assay Interference Compounds*'
- In 2003, J. Baell and his team established a HTS library of 100,000 compounds.
- Some compounds were found to be active in a high number of AlphaScreen assays, however further analysis showed these activities to be false.
- The structure of these compounds were investigated and recurrent substructures were identified and termed PAINS.
- Nowadays, PAINS filters are routinely applied in drug discovery. However, PAINS-compounds are not all unworthy of development and the filters should be applied carefully.

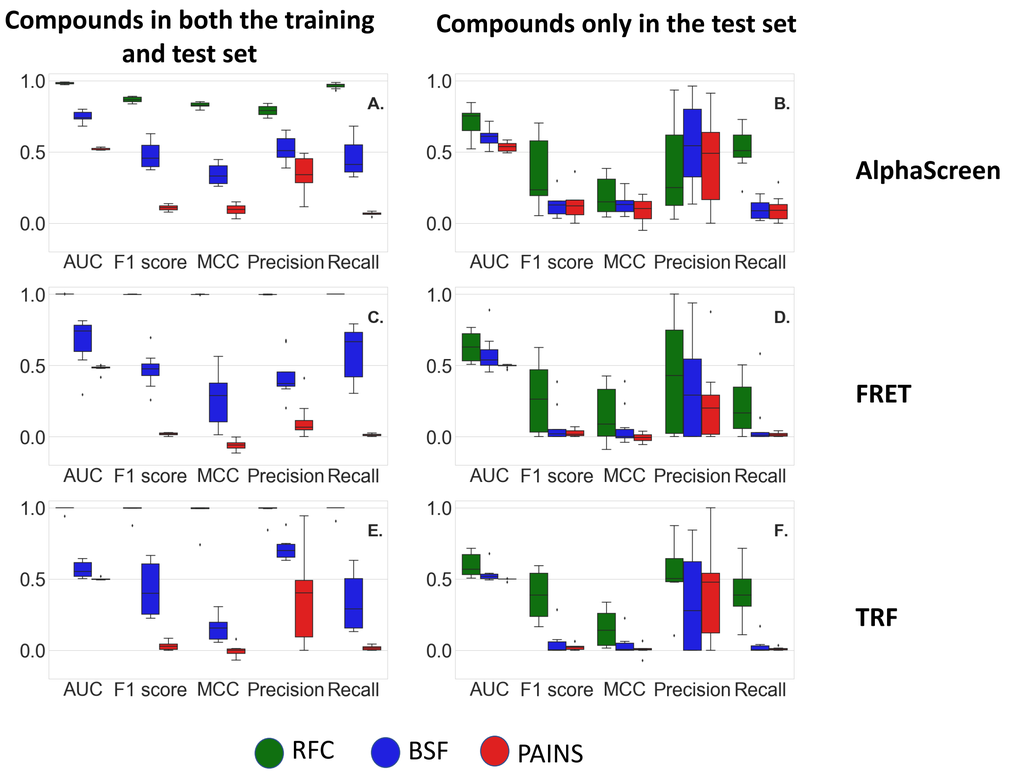


Some PAINS Substructures

Structurally-based model: Random Forest Classifier (RFC)

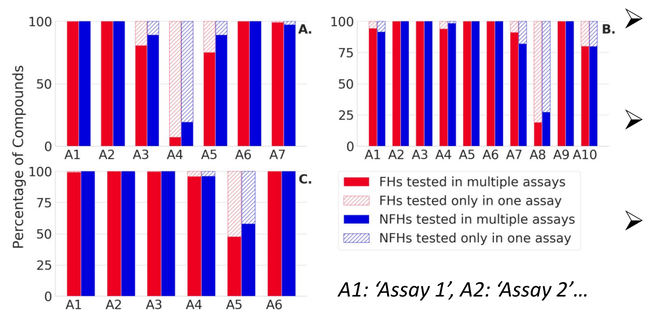


General Performance of the Models



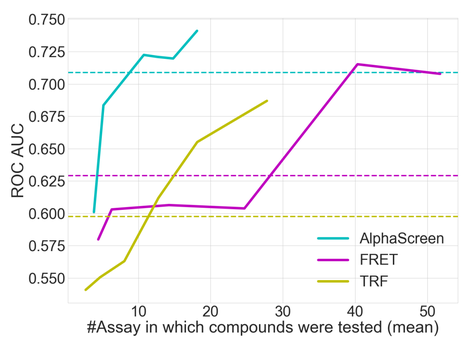
- Models were applied on primary assays that were followed-up by a counter-screen assay, i.e. 7, 10 and 6 primary assays for AlphaScreen, FRET and TRF respectively.
- RFC performance was controlled by cross-validation and label-randomization.
- RFC performed better than the other two models (mean AUC of 0.70, 0.62 and 0.57 for RFC, BSF and PAINS, respectively) and had a higher recall (i.e., it identifies more CIATs).
- PAINS filters performed poorly, although their performance increased for AlphaScreen. This was expected as the filters were developed based on AlphaScreen data.

Behavior Analysis of RFC and BSF



- RFC is based on structural information and can infer predictions for compounds that were never tested before
- BSF is based on historical HTS data, it therefore requires a compound to have been tested at least once to predict it.
- In our data, a fair amount of compounds were tested only once. In a prospective analysis scenario, only RFC would be able to predict them.

- BSF performance increases with the number of data it considers.
- To investigate how well BSF can perform compared to RFC, BSF was calculated based on an increasing number of primary assays.
- It was observed that, to outperform RFC, BSF needed the compounds to be tested in an average of 10 assays for AlphaScreen and TRF and 30 assays for FRET.



Straight lines shows the performance of BSF for an increasing number of assays. Dashed lines represent the performance of RFC for each technology.

Conclusions

- Frequent-hitters and especially CIATs currently slow the process of drug discovery.
- Models such as BSF and the PAINS filters can identify CIATs to some extent.
- Our RFC model is based on compounds that were experimentally validated as CIATs or NCIATs. It can be applied to any technology, contrary to the PAINS filters. It can also predict the behavior of new compounds, which is something BSF cannot do.
- The good performance of RFC shows that there is indeed a relation between interference behavior and chemical structures. PAINS-like filters could be developed for each HTS technology.

References

1. O. Roche et al., *J. Med. Chem*., **2002**, 45 (1), 137-142.
2. J.W.M. Nissink, S. Blackburn, *Future Med. Chem*., **2014**, 6 (10), 1113-26
3.. J. B. Baell, G. A. Holloway, *J. Med. Chem*. **2010**, 53, 2719-2740

This work was supported financially by the **European Union's Horizon 2020 research and innovation program** under the Marie Skłodowska-Curie grant agreement No. 676434, "Big Data in Chemistry" (**"BIGCHEM",** http://bigchem.eu).