Genotoxicity Predictions for Rapid Compound Screening: A Case Study for Accurate Classification using ML
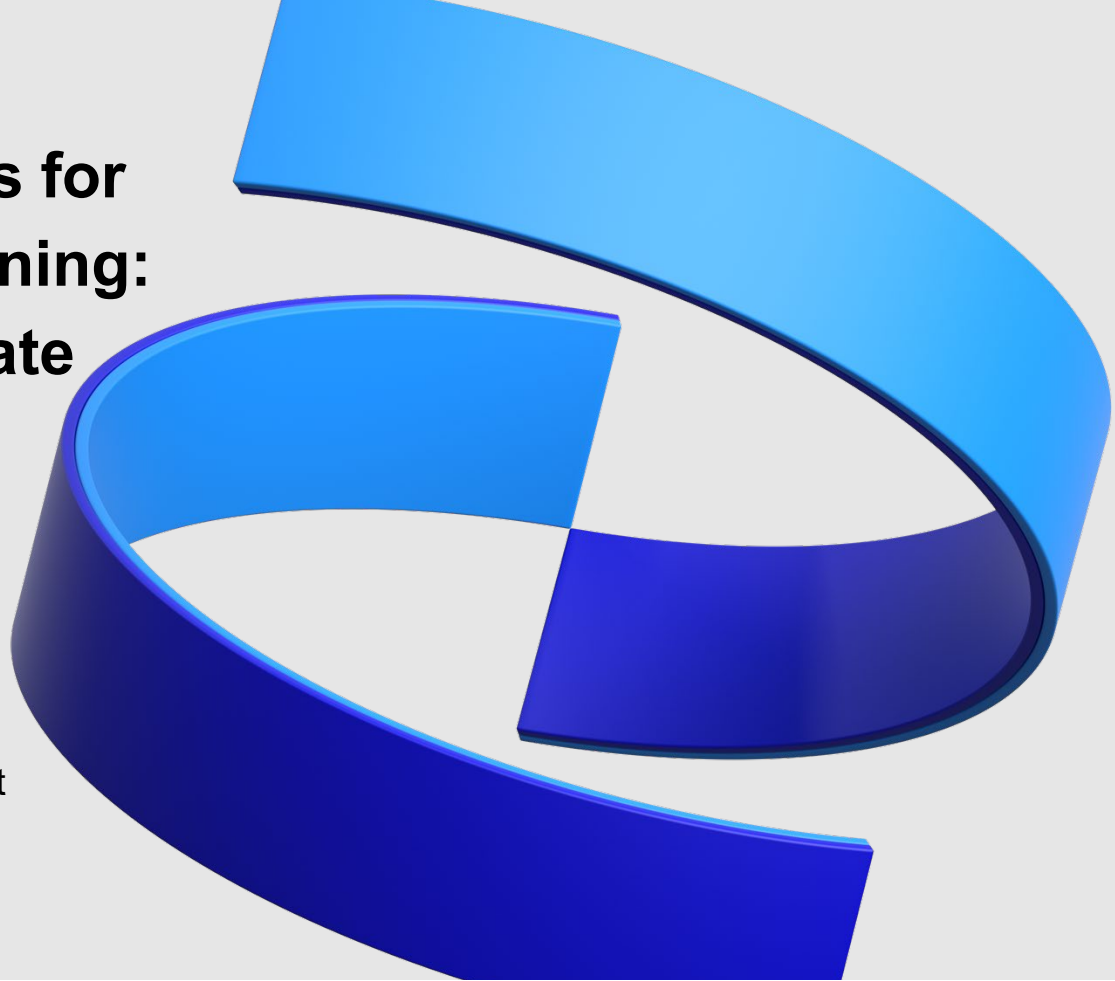
Seda Arat

Wen Sun, Maik Schuler, Matt Martin

Pfizer Drug Safety Research & Development

GTA 2023 Annual Meeting
May 4, 2023

Disclosure

The authors are employees of and have equity interests in Pfizer Inc. All research was funded by Pfizer.

iScreen - a Comprehensive, High-Throughput Imaging Platform for Genotoxicity

- 96-well format
- 2 time points (24h and 4h)
- 22 doses of each compound
- 35 fields, 4 z-stacks
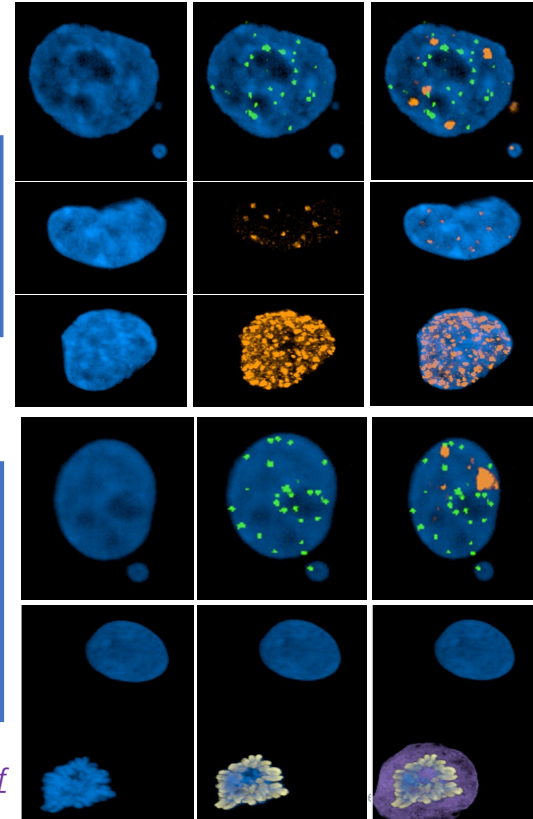- >9000 cells analyzed/condition
- 6 different antibodies

10 Endpoints
- Micronucleus (MN)
- FISH
- H2AXhigh
- H2AX spots
- Mitotic index
- H3 dephosphorylation
- Cytotoxicity
- polyploidy
- PARP
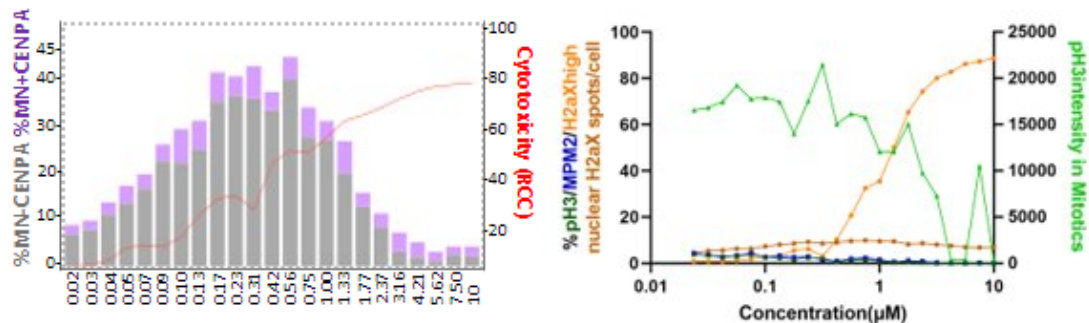- cell cycle

Clastogens
- DNA (MN)
- CENPA (-)
- H2AX (+)

Aneugens
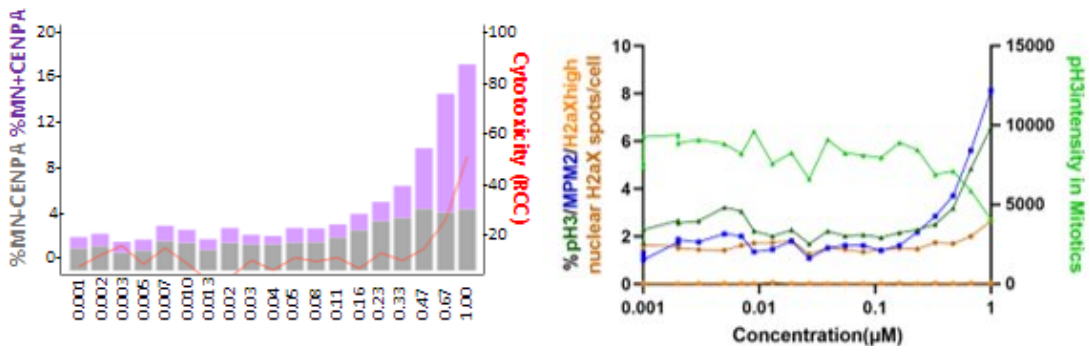- DNA (MN)
- CENPA (+)
- H2AX (-)
- MPM-2
- P- H3

Sun, X. et. al (2022) A new imaging platform (iScreen) allows for the concurrent assessment of micronucleus induction and genotoxic mode of action in human A375 cells

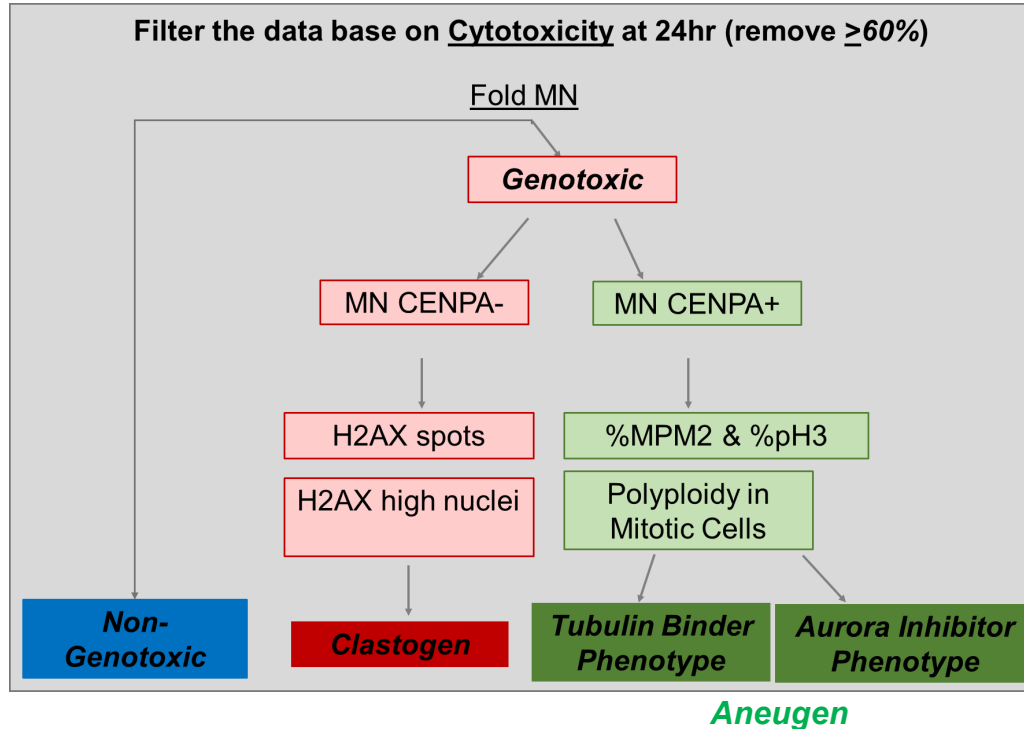# How Does iScreen Output Look Like?
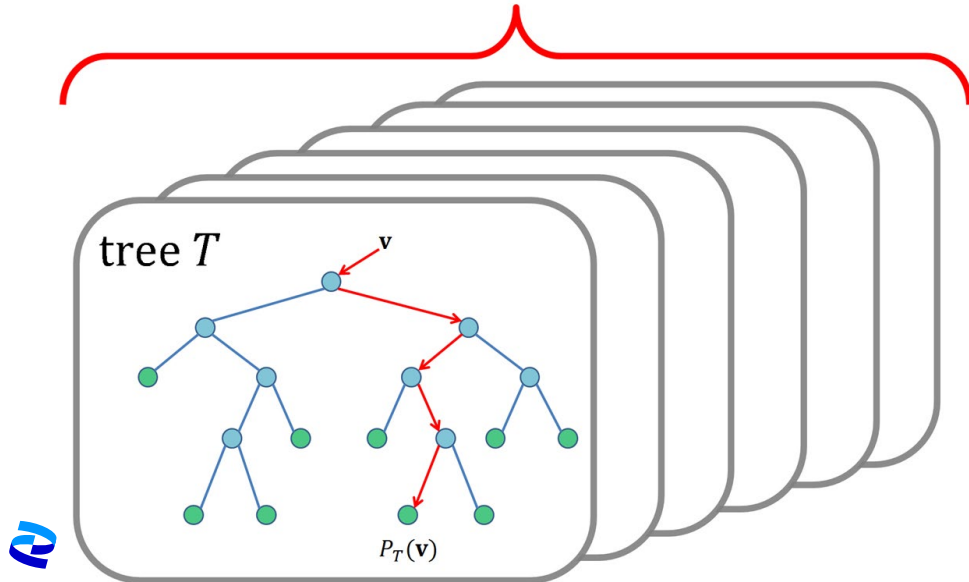

Clastogen (Etoposide)

Aneugen (Kw-2449)

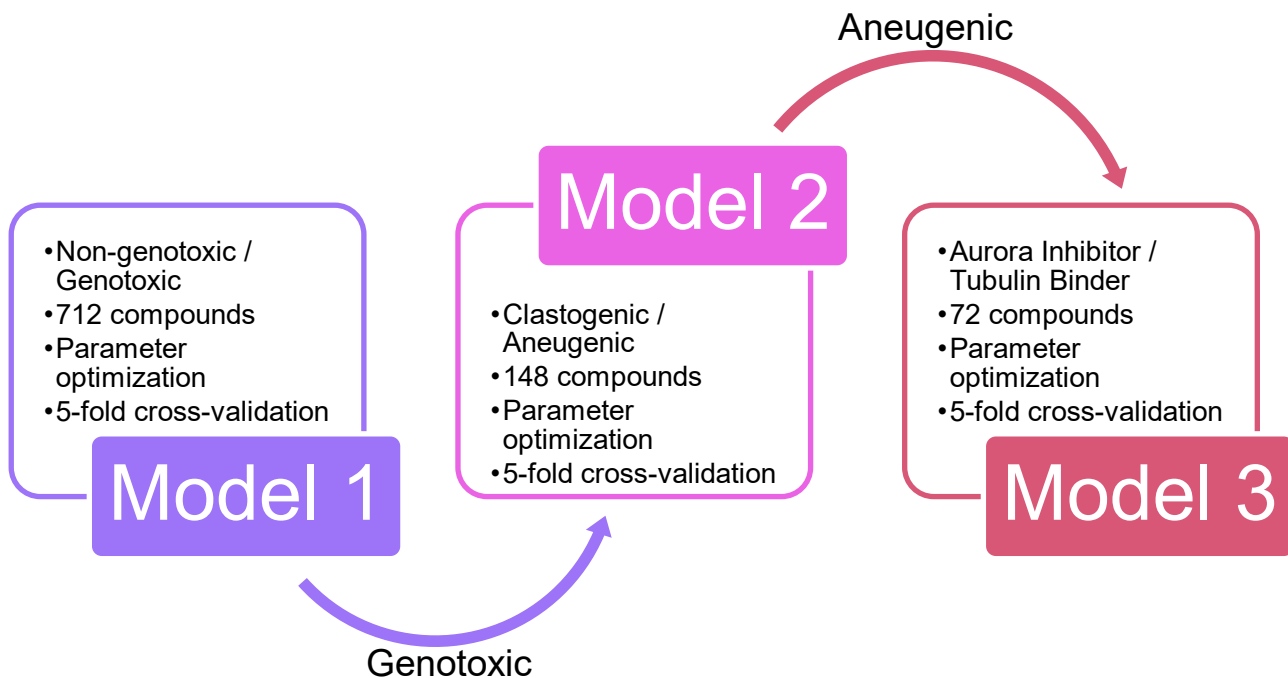# Decision Tree for Compound Classification based on Regulatory Recommendations
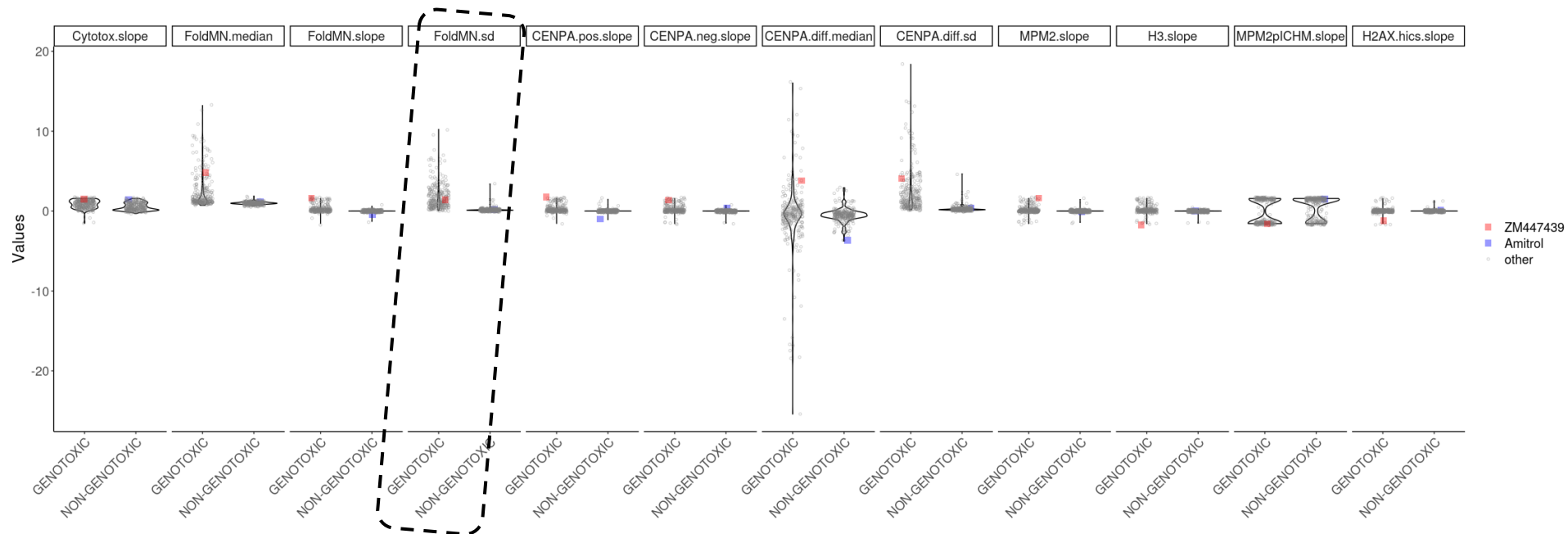
How does Random Forest work?

- Estimating the # of Jelly Beans to the best of your ability

- "Wisdom of the Crowds"

tree T $\mathbf{v}$

$P_T(\mathbf{v})$

https://iq.opengenus.org/random-decision-forest/

Tiered-Random Forest Modeling Approach

Model 2

Model 1
- Non-genotoxic / Genotoxic
- 712 compounds
- Parameter optimization
- 5-fold cross-validation

- Clastogenic / Aneugenic
- 148 compounds
- Parameter optimization
- 5-fold cross-validation

Aneugenic

Model 3
- Aurora Inhibitor / Tubulin Binder
- 72 compounds
- Parameter optimization
- 5-fold cross-validation

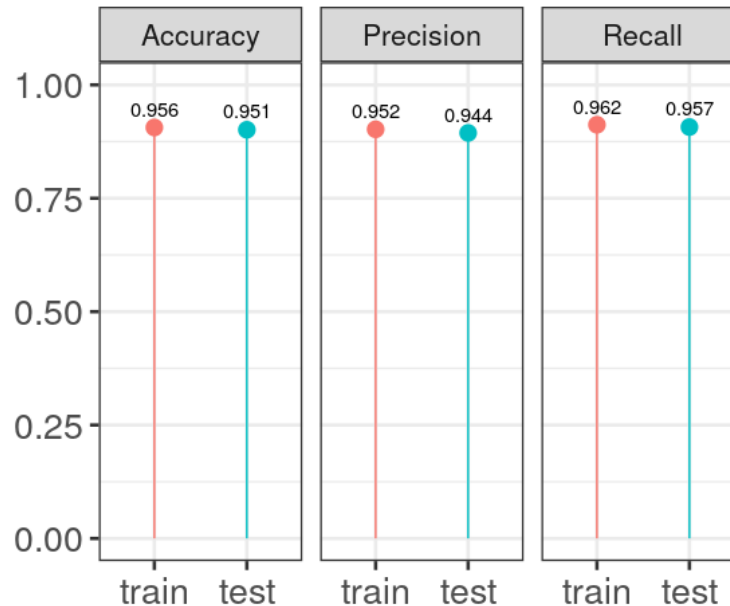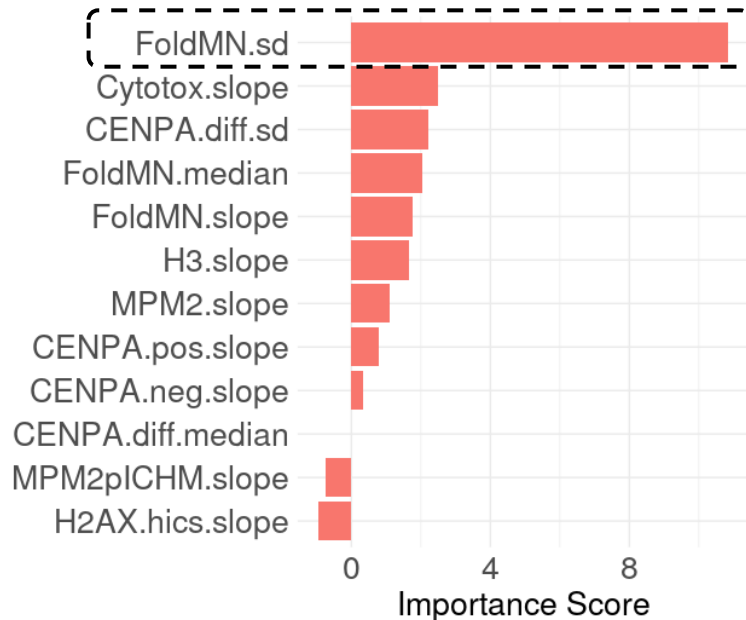Genotoxic
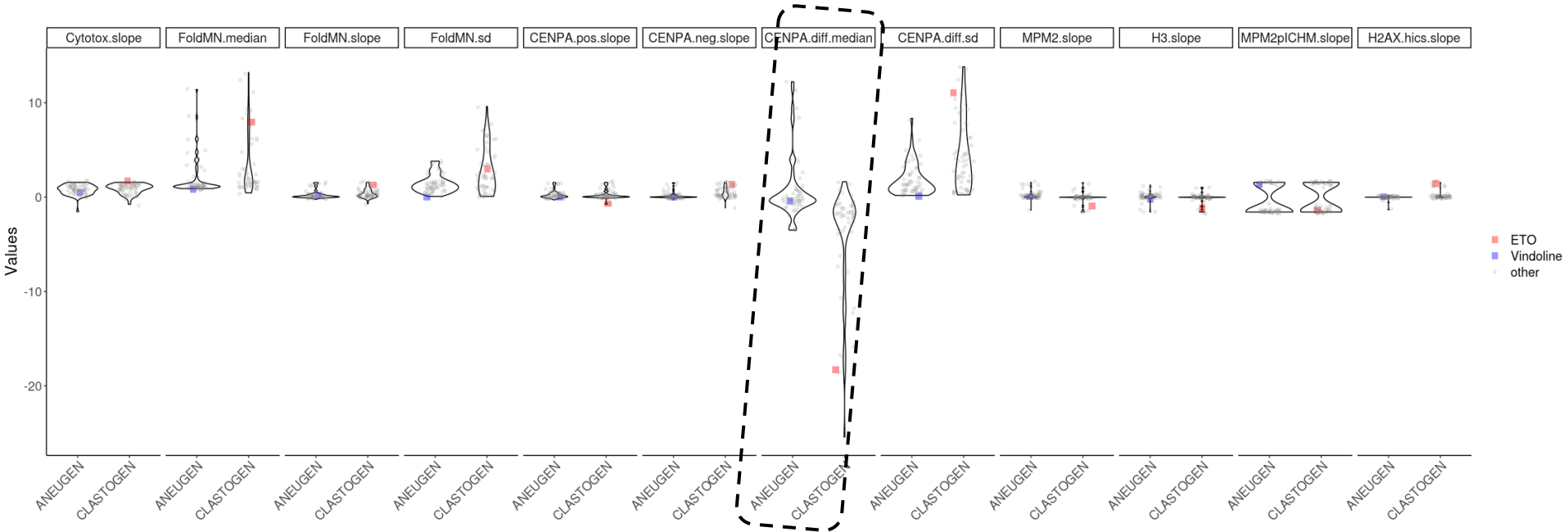
Classification 1: Nongenotoxic vs. Genotoxic Data

In Model 1, FoldMN** has higher predictive value to differentiate genotoxicants from non-genotoxicants

Compound	Genotoxic	Non-genotoxic
ZM447439	100%	0%
Amitrol	8.7%	91.3%

Classification 2: Clastogen vs. Aneugen Data

In Model 2, CENPA** has higher predictive value to differentiate clastogens from aneugens
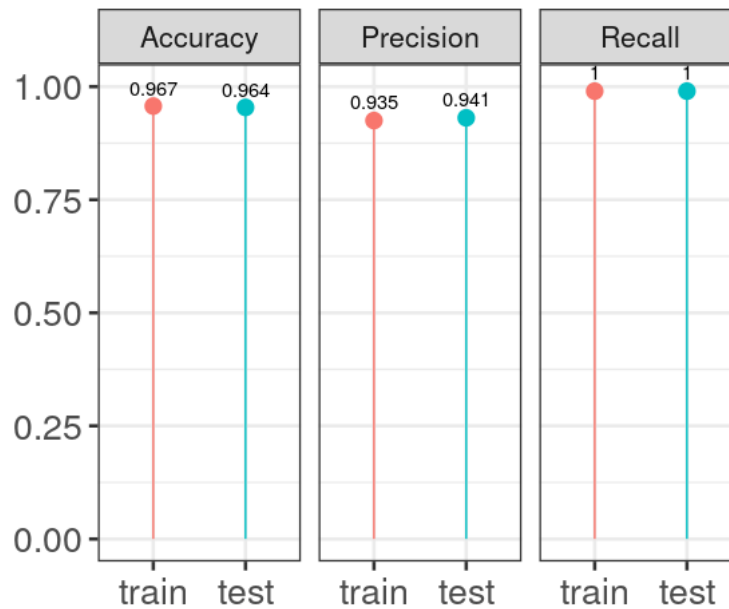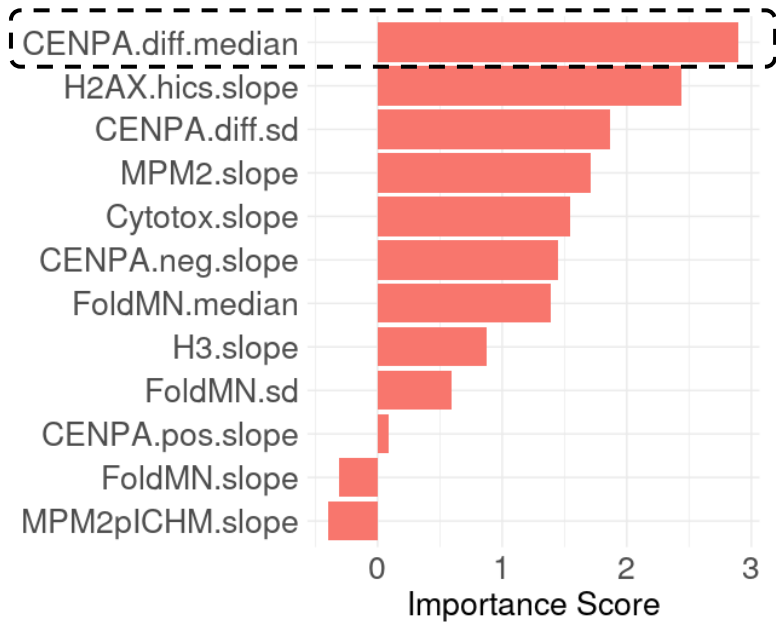
Compound	Clastogen	Aneugen
ETO	**88.2%**	**11.8%**
Vindoline	**5.9%**	**94.1%**

Classification 3: Aurora Inhibitor vs. Tubulin Binder Data
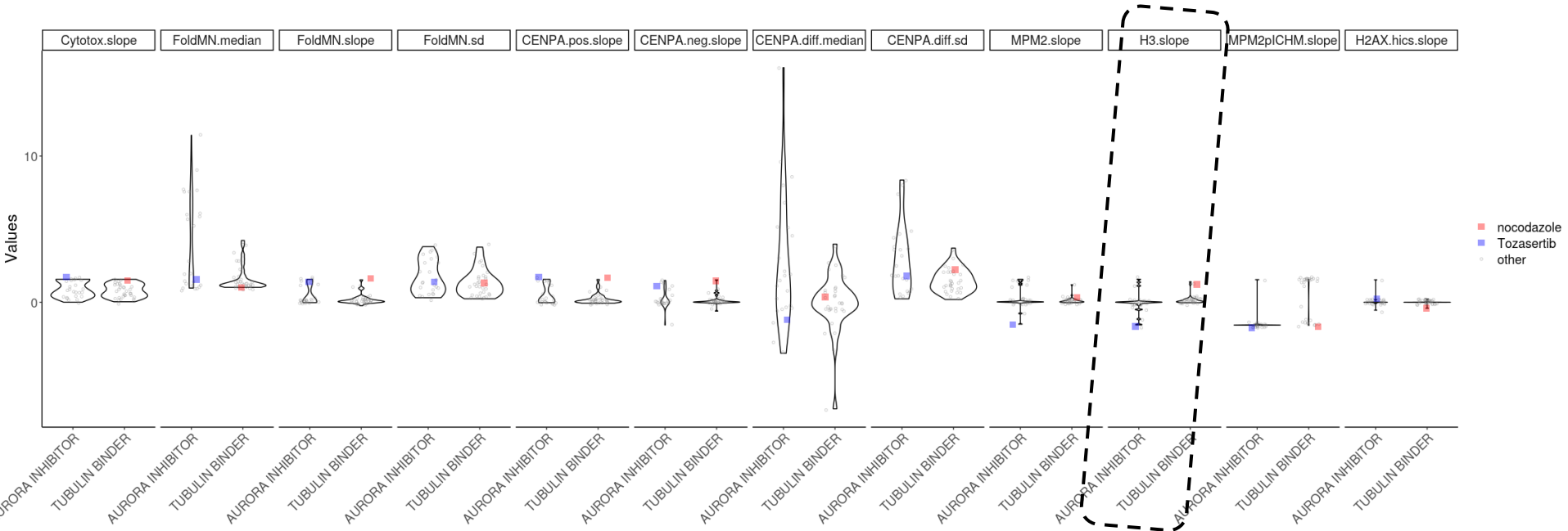
In Model 3, MPM2 and H3 have higher predictive value to differentiate aurora inhibitors from tubulin binders
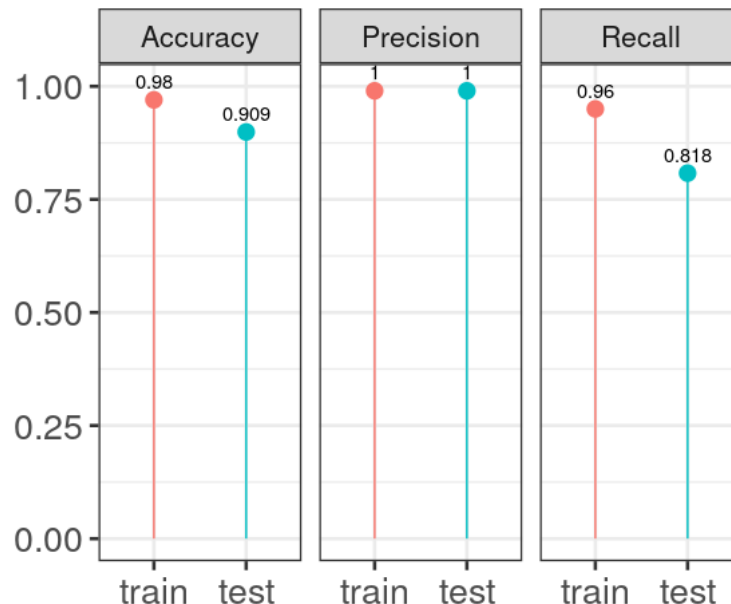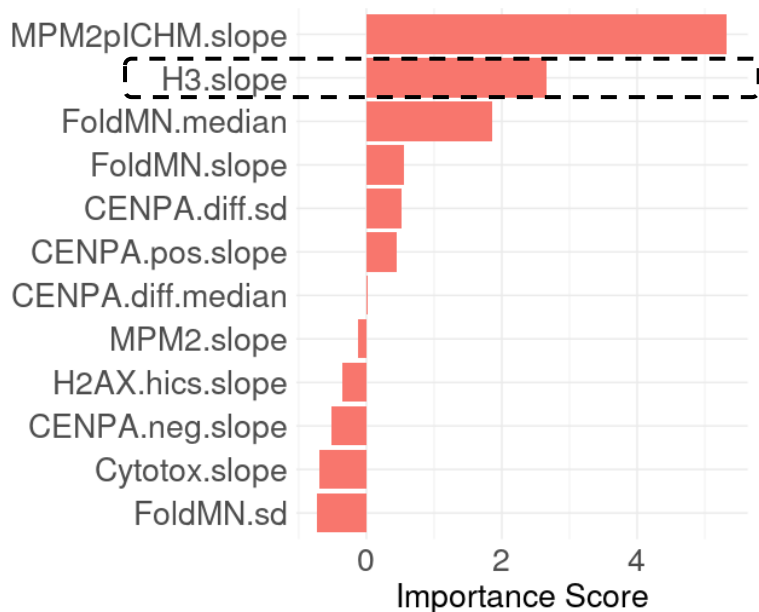
Compound	Aurora Inh	Tubulin Bin
nocodazole	**28.6%**	**71.4%**
Tozasertib	**95.2%**	**4.8%**

Our Models Predicts Compounds with "mixed" MoA in the literature as genotoxic with mixed/inconclusive MoA

Compound	Non-genotoxic	Genotoxic	Clastogen	Aneugen	Aurora In	Tubulin Bi
Calyculin A	26%	**74%**	53%	47%	48%	52%
Doxorubicin	0%	**100%**	59%	41%	38%	62%
Trichostatin A	0%	**100%**	59%	41%	**81%**	19%

MoA: Mechanism of Action

Conclusion

- iScreen - a Comprehensive, High-Throughput Imaging Platform for Genotoxicity

- In this study, 1500+ compounds were used to generate a tiered-random forest modeling approach

- Our models have high predictive power, even on compounds with mixed MoA

Model 1
- Non-genotoxic / Genotoxic
- 712 compounds
- Parameter optimization
- 5-fold cross-validation

Model 2
- Clastogenic / Aneugenic
- 148 compounds
- Parameter optimization
- 5-fold cross-validation

Model 3
- Aurora Inhibitor / Tubulin Binder
- 72 compounds
- Parameter optimization
- 5-fold cross-validation

Genotoxic

Aneugenic
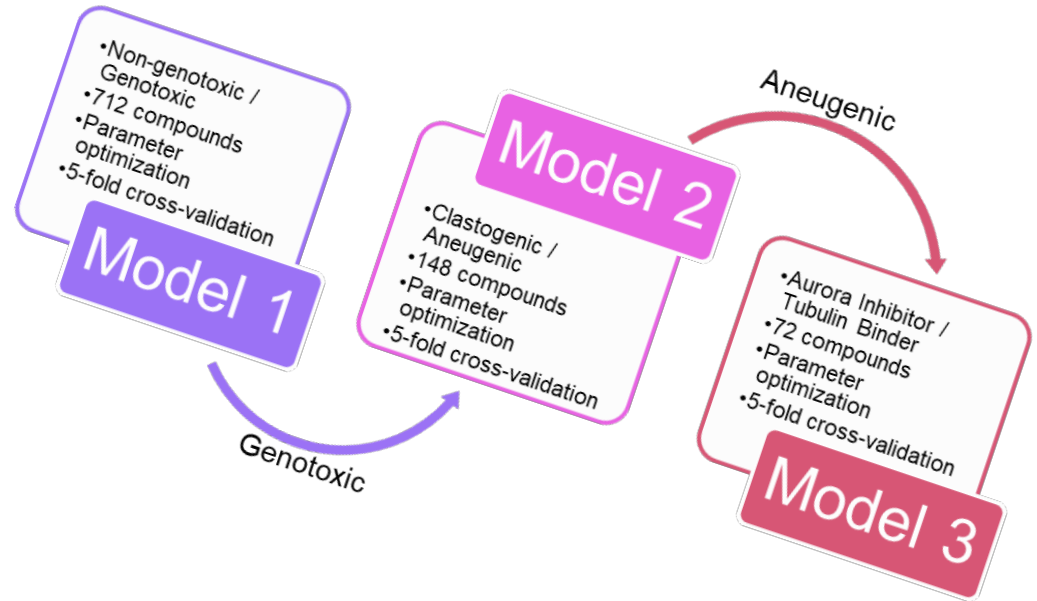
We prospectively monitor our model predictions for genotoxicity risk assessment and regular updates (1600+ compounds)

Supplementary Materials
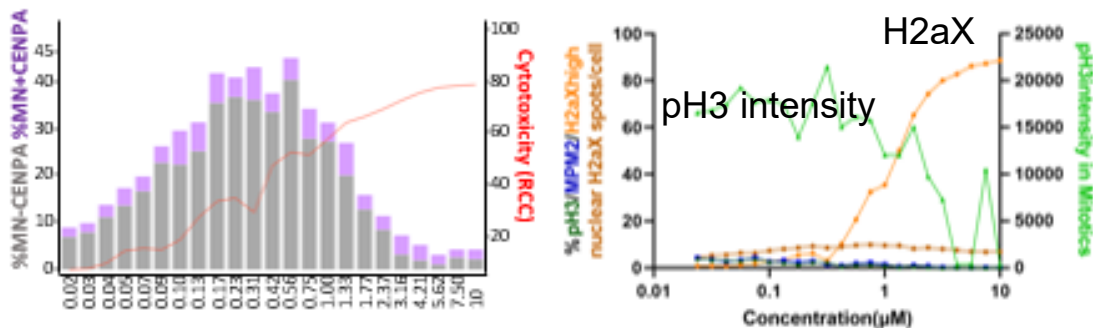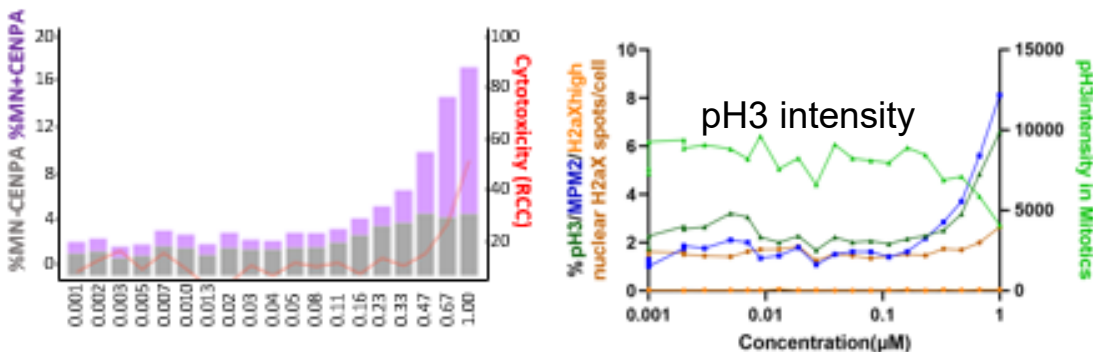
How Does iScreen Output Look Like?

antibody	endpoint
FoldMN	MicroNucleus by DNA stain
MN-CENPA	No centromere region
MN+CENPA	centromere region
MPM2	mitotic proteins
pH3	phosphorylated histone H3
H2AX	double strand break
MPM2pICHM	intensity of pH3

Clastogen (Etoposide)

H2aX

pH3 intensity

Aneugen (Kw-2449)

pH3 intensity

Genotoxicity Markers

antibody	endpoint	note
FoldMN	MN by DNA stain	detects genotoxic compounds
MN-CENPA	No centromere region	MN without CENPA contains fragment of chromosome, direct DNA damaging (clastogens)
MN+CENPA	centromere region	MN with CENPA contains full chromosome, compound affect mitotic apparatus, non-DNA damaging (aneugens)
MPM2	mitotic proteins	shows mitotic arrest by aneugens
pH3	phosphorylated histone H3	shows mitotic arrest by aneugens, in addition dephosphorylation is a marker for aurora inhibitors
H2AX	double strand break	H2AX high: cells had high level of damage; H2AX foci: individual double strand break sites
MPM2pICHM	intensity of pH3	shows pH3 dephosphorylation progression, using actual intensity instead of thresholded cutoff

FoldMN: Fold change in MicroNucleus relative to DMSO
CENPA: Centromere protein A