

## Evaluating the utility of brightfield image data for mechanism of action prediction

**Table S2.** Comparison of the results for three model runs with different seeds when site-level normalization was applied to the data.

(a) Mean and standard deviation of the Macro-F1 scores on the test sets for the five data splits.

	BF <sub>site</sub>	FL <sub>site</sub>
Split 1	0.661±0.004	<b>0.644±0.022</b>
Split 2	<b>0.772±0.003</b>	0.750±0.010
Split 3	0.657±0.002	<b>0.668±0.007</b>
Split 4	<b>0.673±0.006</b>	0.650±0.008
Split 5	<b>0.706±0.004</b>	0.691±0.002

(b) Mean and standard deviation of the F1 scores per MoA on the test sets.

	BF <sub>site</sub>	FL <sub>site</sub>
ATPase-i	0.657±0.007	<b>0.677±0.004</b>
AuroraK-i	<b>0.714±0.001</b>	0.667±0.003
HDAC-i	0.766±0.002	<b>0.786±0.003</b>
HSP-i	<b>0.751±0.004</b>	0.692±0.008
JAK-i	0.378±0.025	<b>0.407±0.022</b>
PARP-i	<b>0.781±0.007</b>	0.735±0.012
Prot.Synth.-i	0.517±0.008	<b>0.634±0.010</b>
Ret.Rec.Ag	0.785±0.013	<b>0.791±0.016</b>
Topo.-i	<b>0.687±0.011</b>	0.643±0.007
Tub.Pol.-i	0.858±0.006	<b>0.874±0.006</b>
DMSO	<b>0.846±0.007</b>	0.684±0.007
Macro average	<b>0.704±0.001</b>	0.690±0.005