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

Table S4. Comparison of the test set F1 scores for three different architectures trained on BF and FL when site-level normalization was applied to the data.

(a) Macro-F1 scores on the test sets for the five data splits.

	ResNet-50		DenseNet-169		EfficientNet-B3	
	BF_{site}	FL_{site}	BF_{site}	FL_{site}	BF_{site}	FL_{site}
Split 1	0.662	0.661	**0.677**	0.655	0.638	0.665
Split 2	0.770	0.762	0.776	0.742	0.757	**0.776**
Split 3	0.654	0.677	0.647	0.654	0.630	**0.705**
Split 4	**0.676**	0.645	0.638	0.604	0.625	0.641
Split 5	0.708	0.688	0.690	0.687	0.595	**0.708**

(b) F1 scores per MoA on the test sets.

	ResNet-50		DenseNet-169		EfficientNet-B3	
	BF_{site}	FL_{site}	BF_{site}	FL_{site}	BF_{site}	FL_{site}
ATPase-i	0.650	0.683	0.622	0.648	0.585	**0.702**
AuroraK-i	0.713	0.671	0.693	0.644	**0.724**	0.667
HDAC-i	0.766	0.785	0.749	0.785	0.760	**0.819**
HSP-i	0.756	0.682	**0.767**	0.689	0.748	0.686
JAK-i	0.405	0.429	0.345	0.369	0.255	**0.444**
PARP-i	0.789	0.748	**0.797**	0.734	0.756	0.789
Prot.Synth.-i	0.520	0.646	0.543	0.647	0.455	**0.668**
Ret.Rec.Ag	0.767	0.796	0.790	0.798	0.683	**0.822**
Topo.-i	**0.702**	0.651	0.676	0.616	0.626	0.600
Tub.Pol.-i	0.850	**0.865**	0.856	0.859	0.827	0.818
DMSO	**0.836**	0.691	0.801	0.685	0.732	0.791
Macro average	0.705	0.695	0.695	0.679	0.650	**0.710**