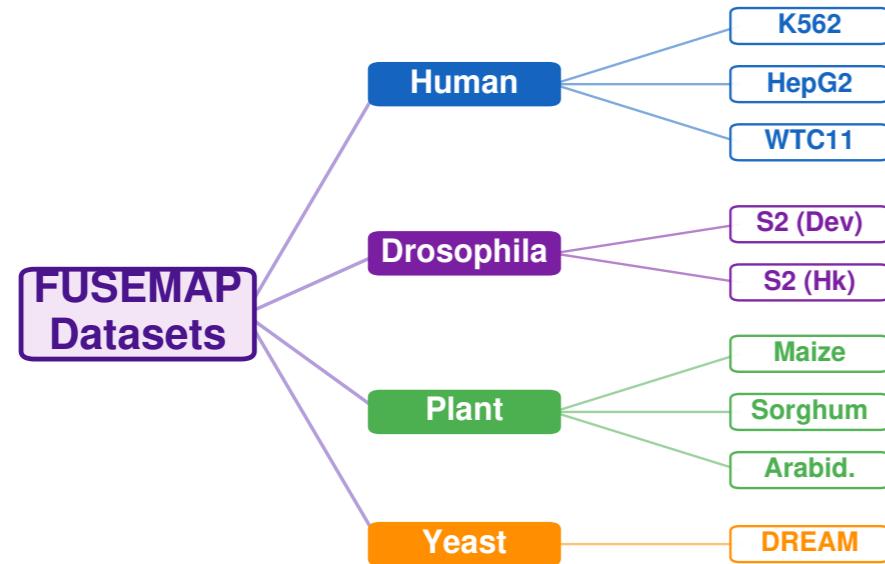


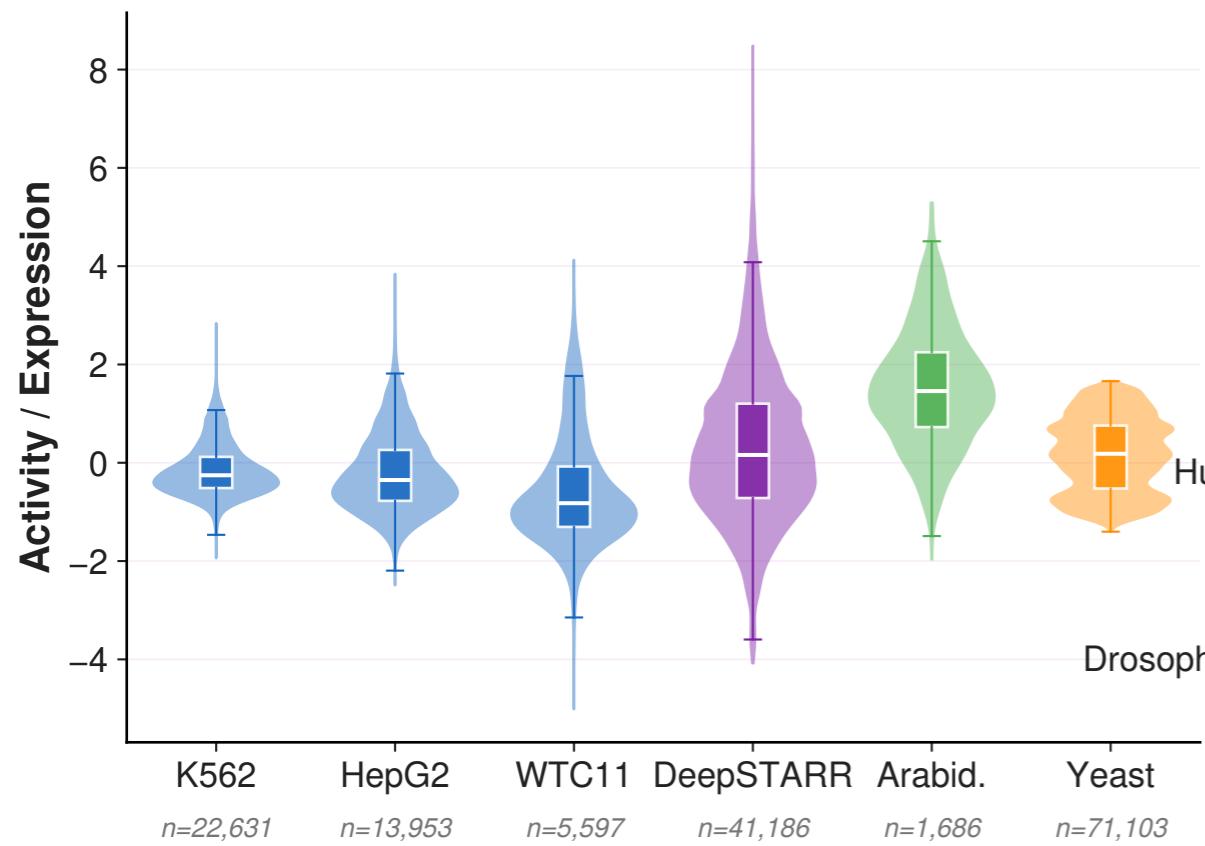
A Cross-Kingdom Dataset Coverage



B Dataset Summary (9 Datasets, 4 Kingdoms)

Dataset	Species	Assay	Train	Val	Test	Len
K562	Human	lentiMPRA	158,377	22,613	22,631	230
HepG2	Human	lentiMPRA	97,925	13,997	13,953	230
WTC11	Human	lentiMPRA	39,201	5,596	5,597	230
DeepSTARR	Drosophila	STARR-seq	352,009	40,570	41,186	249
Maize	Plant	STARR-seq	24,209	3,458	3,001	170
Sorghum	Plant	STARR-seq	19,502	2,785	2,467	170
Arabid.	Plant	STARR-seq	13,169	1,881	1,686	170
Yeast	Yeast	FACS-seq	~6.7M	33,696	71,103	110

C Activity Distributions (Test Sets)



D Input Sequence Lengths

