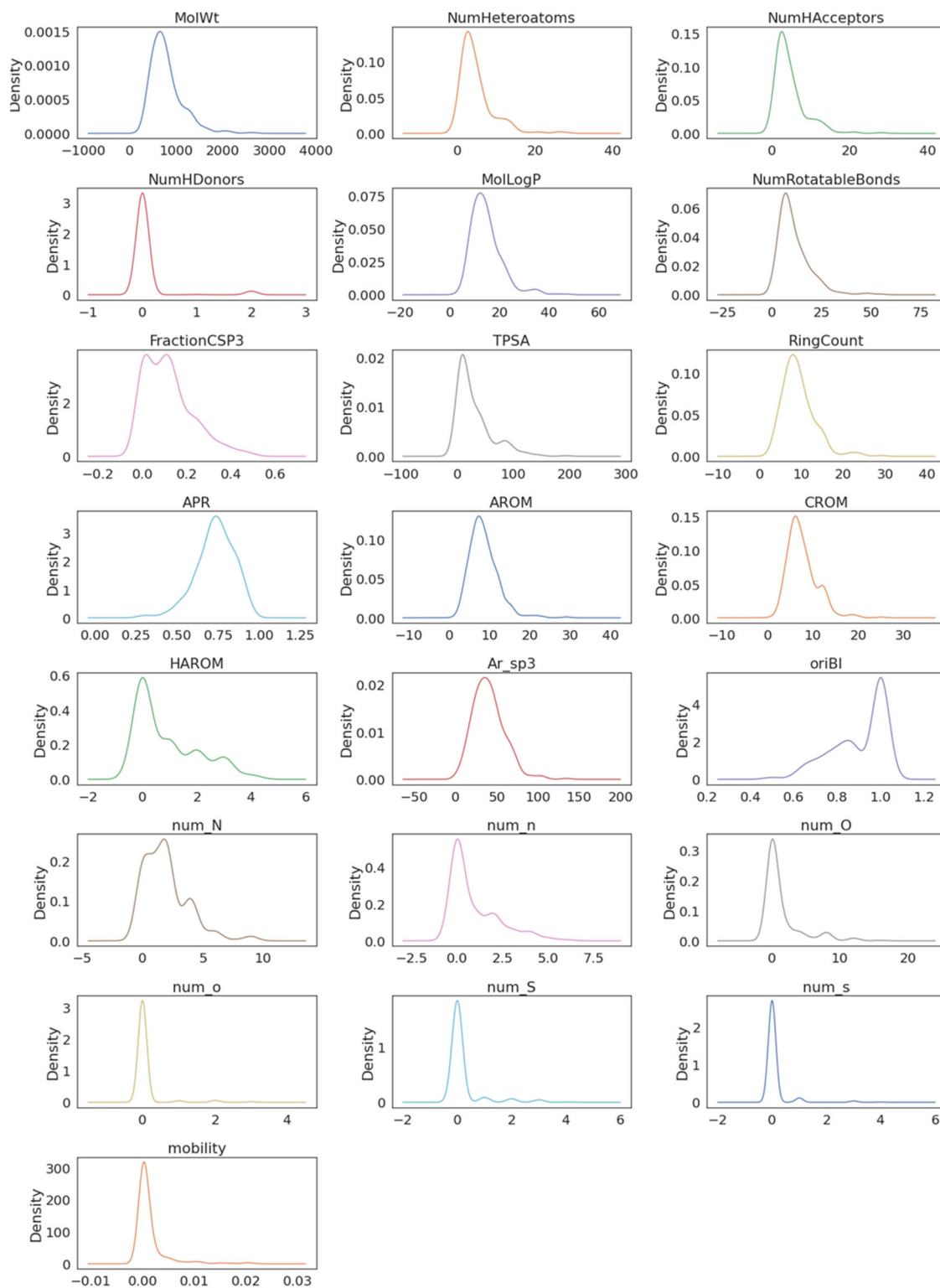


(a)



(b)

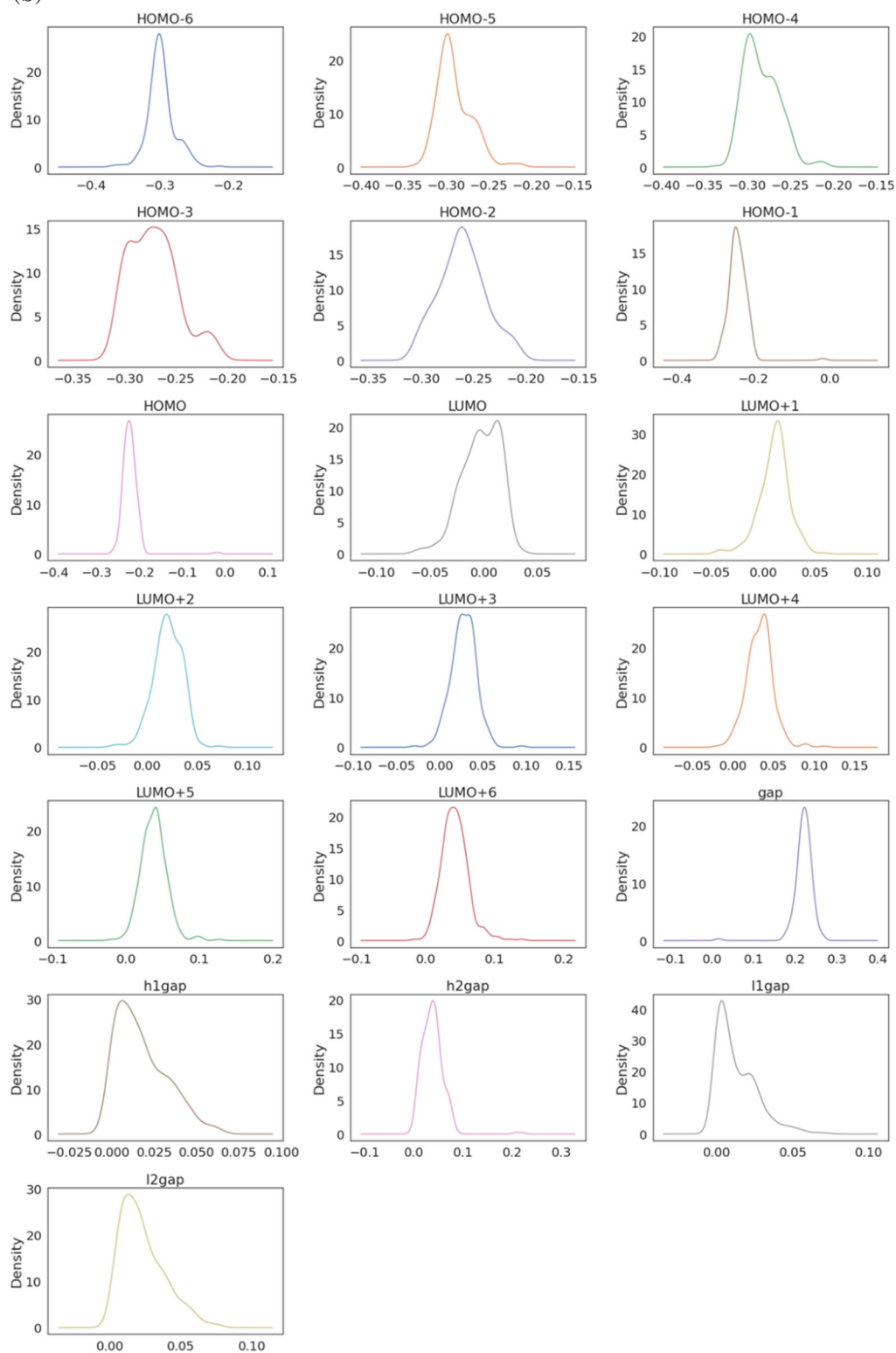


Figure S1. Kernel density examination of the explanatory descriptors.

(a) Descriptors related to the structural features. (b) Descriptors related to the orbital energies.

Table S1. Comparison of R^2 and RMSE values evaluated by different fingerprints using the MO-embedding explanatory descriptors.

model	finger-print	radius	bit length	R^2		RMSE		5-fold mean R^2
				test	train	test	train	
1	FCFP	2	1024	0.741	0.912	0.554	0.360	0.516
2	FCFP	2	2049	0.743	0.919	0.553	0.346	0.518
3	FCFP	3	2048	0.756	0.920	0.596	0.336	0.521
4	ECFP	2	1024	0.742	0.914	0.554	0.356	0.529
5	ECFP	2	2048	0.712	0.893	0.657	0.388	0.530
6	ECFP	3	2048	0.722	0.907	0.636	0.363	0.526

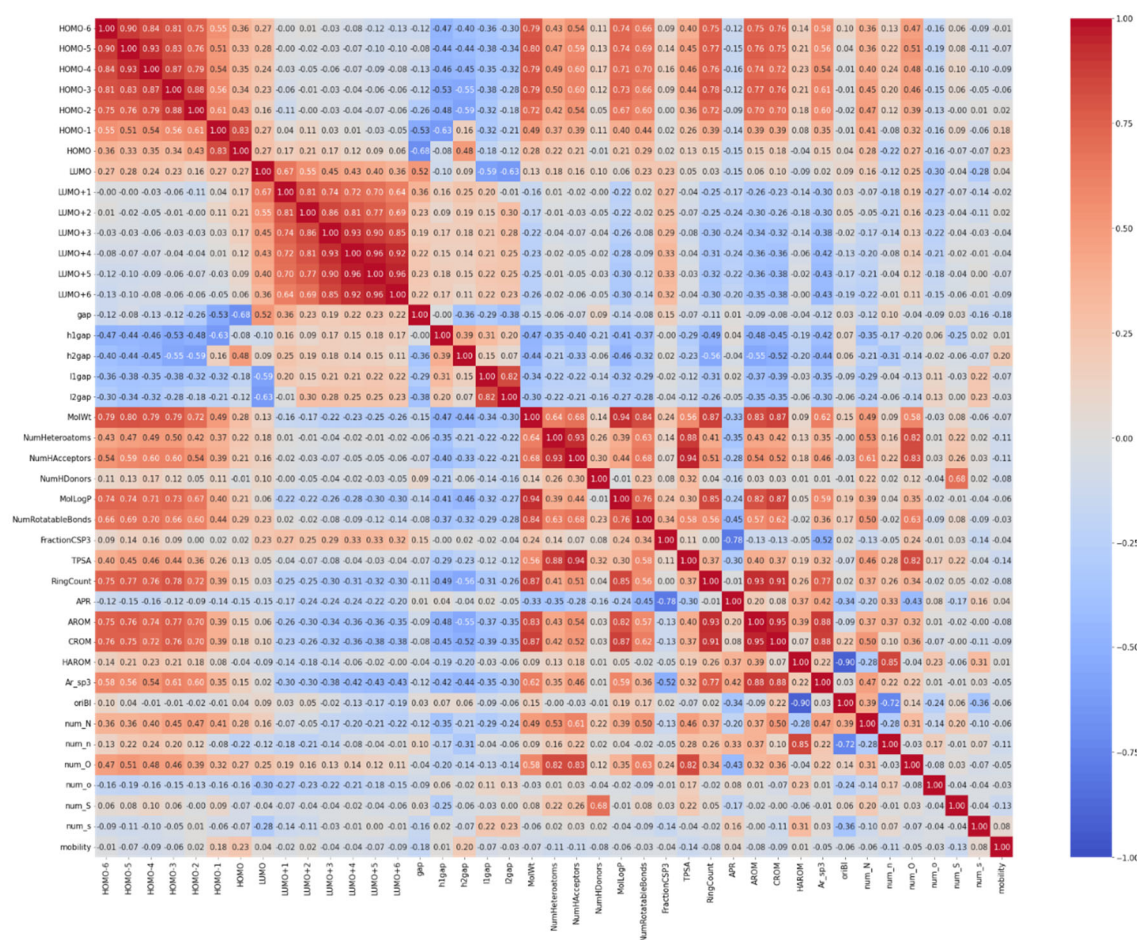


Figure S2. Heatmap of correlation coefficients for all pairs of explanatory descriptors except for fingerprints.

Table S2. Comparison of R^2 and RMSE values evaluated by different number of descriptors, Nd, extracted by applying the threshold to the importance in RF using the MO-embedding explanatory descriptors.

model	Nd	importance	R^2		RMSE		5-fold mean R^2
			test	train	test	train	
1	917	0.0	0.742	0.914	0.554	0.356	0.529
2	131	0.001	0.753	0.912	0.585	0.352	0.5334
3	85	0.002	0.734	0.891	0.606	0.395	0.5335
4	61	0.003	0.741	0.868	0.555	0.442	0.514

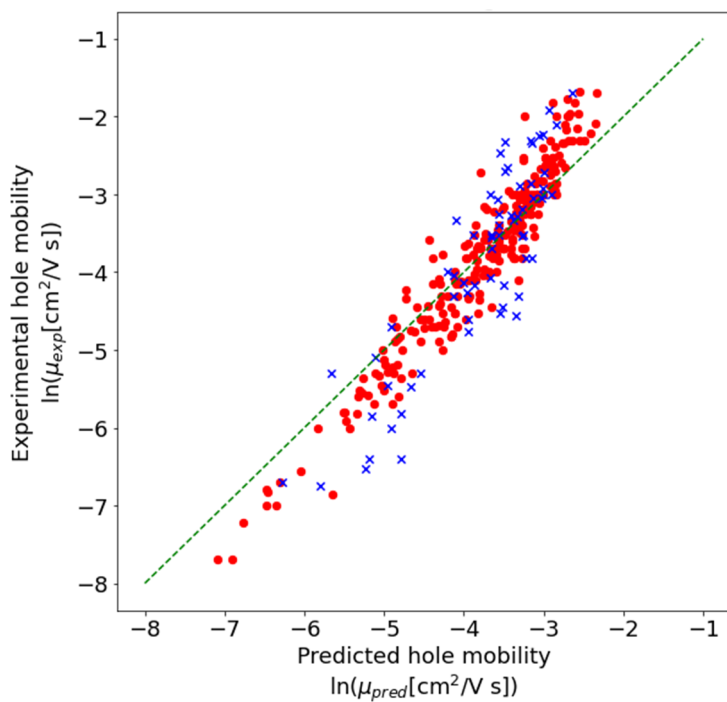


Figure S3. . Correlation plots between predicted and experimental hole mobilities for training and testing datasets using MO-embedding model. The red circle and blue cross-plots denote the training and testing datasets, respectively.

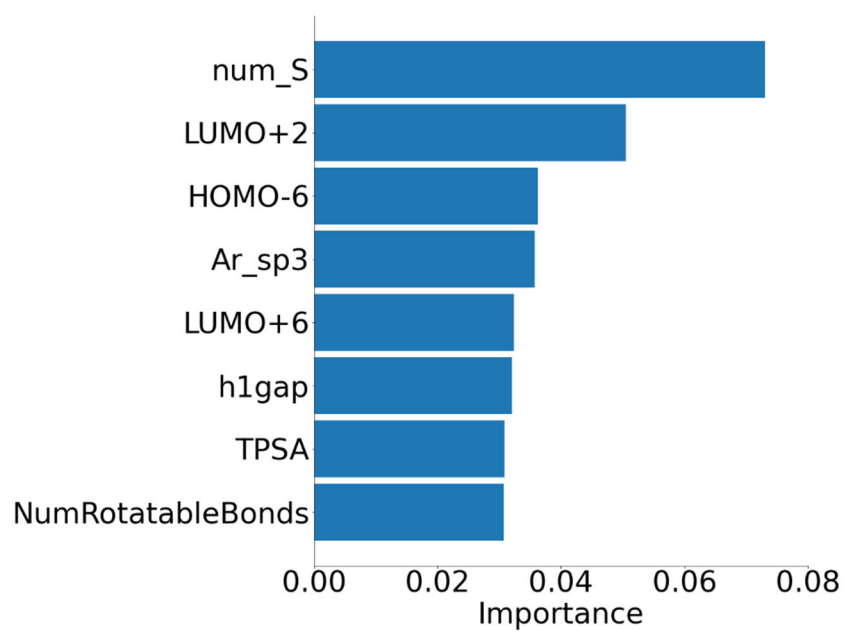


Figure S4. Importance of top 5 descriptors (importance > 0.05) for the MO-embedding model. The significances of these indexes are listed in Table 1.