Electronic Supplementary Material (ESI) for RSC Advances.

Supporting Information

Computational approaches for the prediction of the selective uptake of magnetofluorescent nanoparticles into human cells

Ester Papa, Jean Pierre Doucet, Annick Doucet-Panaye

Figures S1 – S15

Tables S1 – S10

Figure S1 A-E: Plot of experimental vs. predicted values for the best logPaCa2 model in five different splitting (yellow dots=training set; blue dots=prediction set).

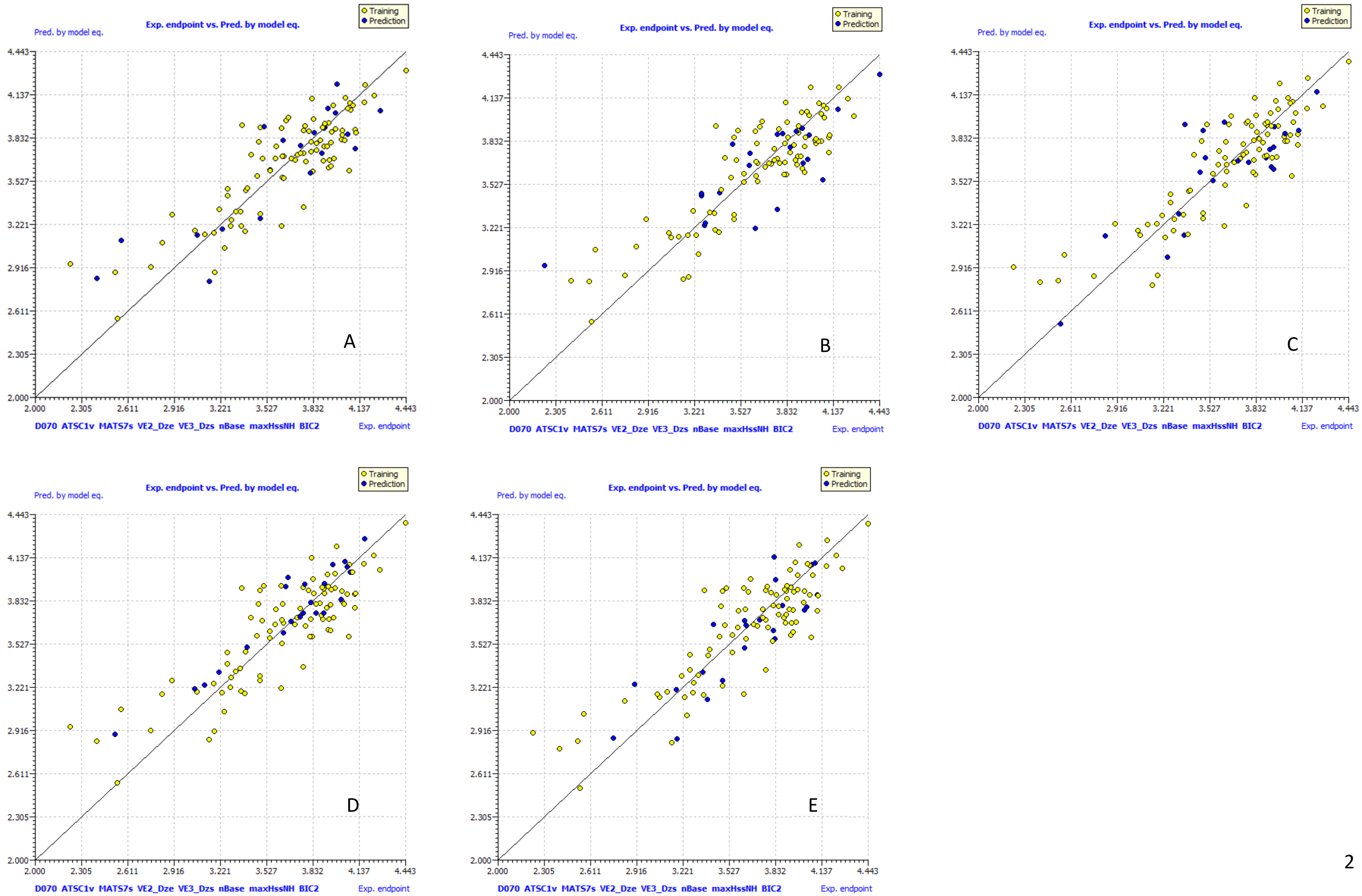


Figure S2. Comparison of the predictive performances of linear and non-linear approaches applied to predict logPaCa2, tested on five external validation sets on the basis of Mean Absolute Errors (MAE) and Root Mean Squared Errors (RMSE). Combined=calculated on averaged predictions.

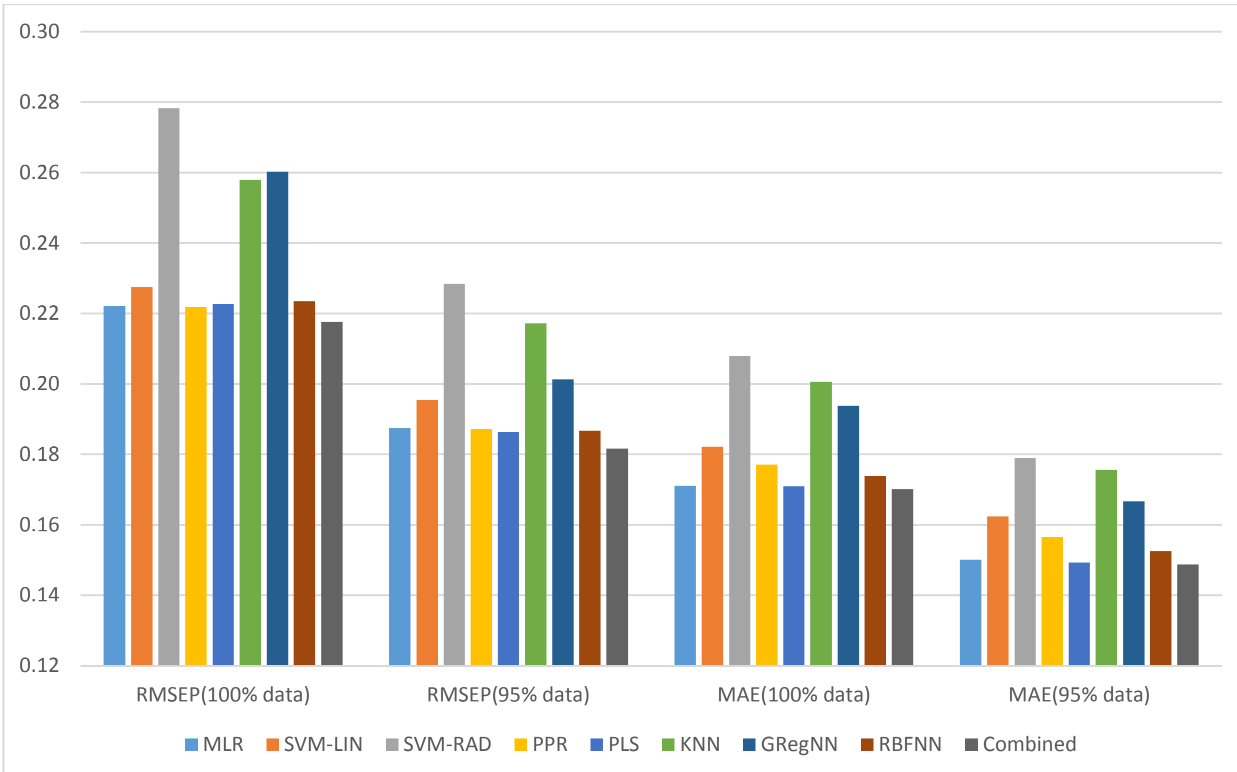


Figure S3 A-E: Plot of hat diagonal values vs. standardized residuals for the best logPaCa2 model in five different splitting (yellow dots=training set; blue dots=prediction set).

Figure S4 A-B: Plots of the first 2 principal components calculated on residuals in prediction from 8 different modelling techniques applied to the logPaCa2 response. Figure A: score plot (1=anhydrides; 2=amines; 3=aminoacids). Figure B: loading plot.

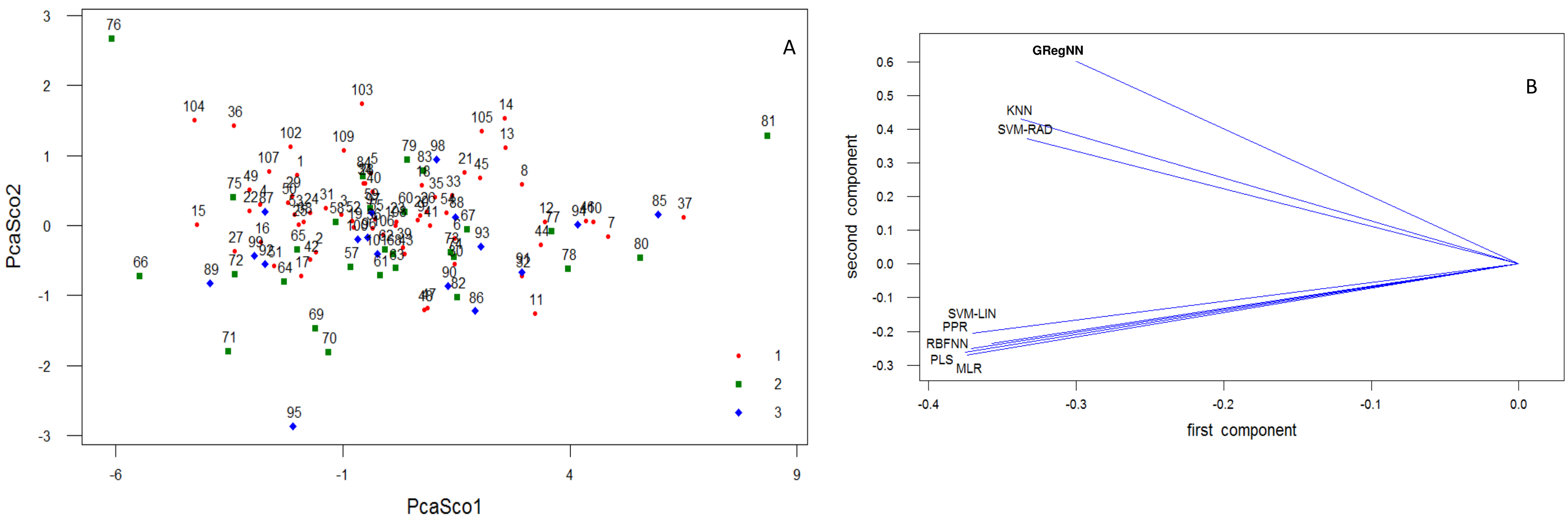


Figure S5 A-E: Plot of experimental vs. predicted values for the best logHUVEC model in five different splittings. (yellow dots=training set; blue dots=prediction set).

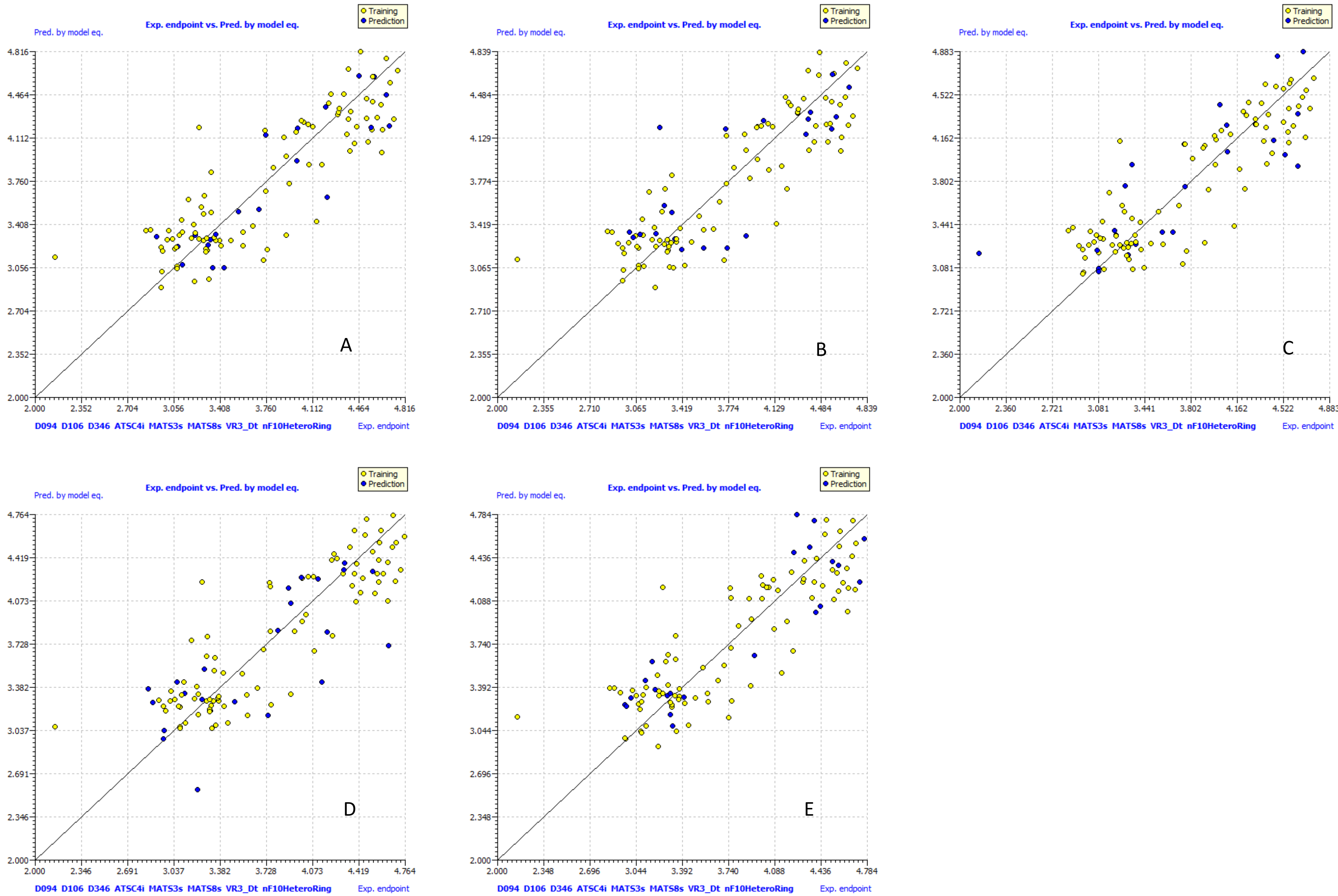


Figure S6 A-E: Plot of hat diagonal values vs. standardized residuals for the best logHUVEC model in five different splittings. (yellow dots=training set; blue dots=prediction set).

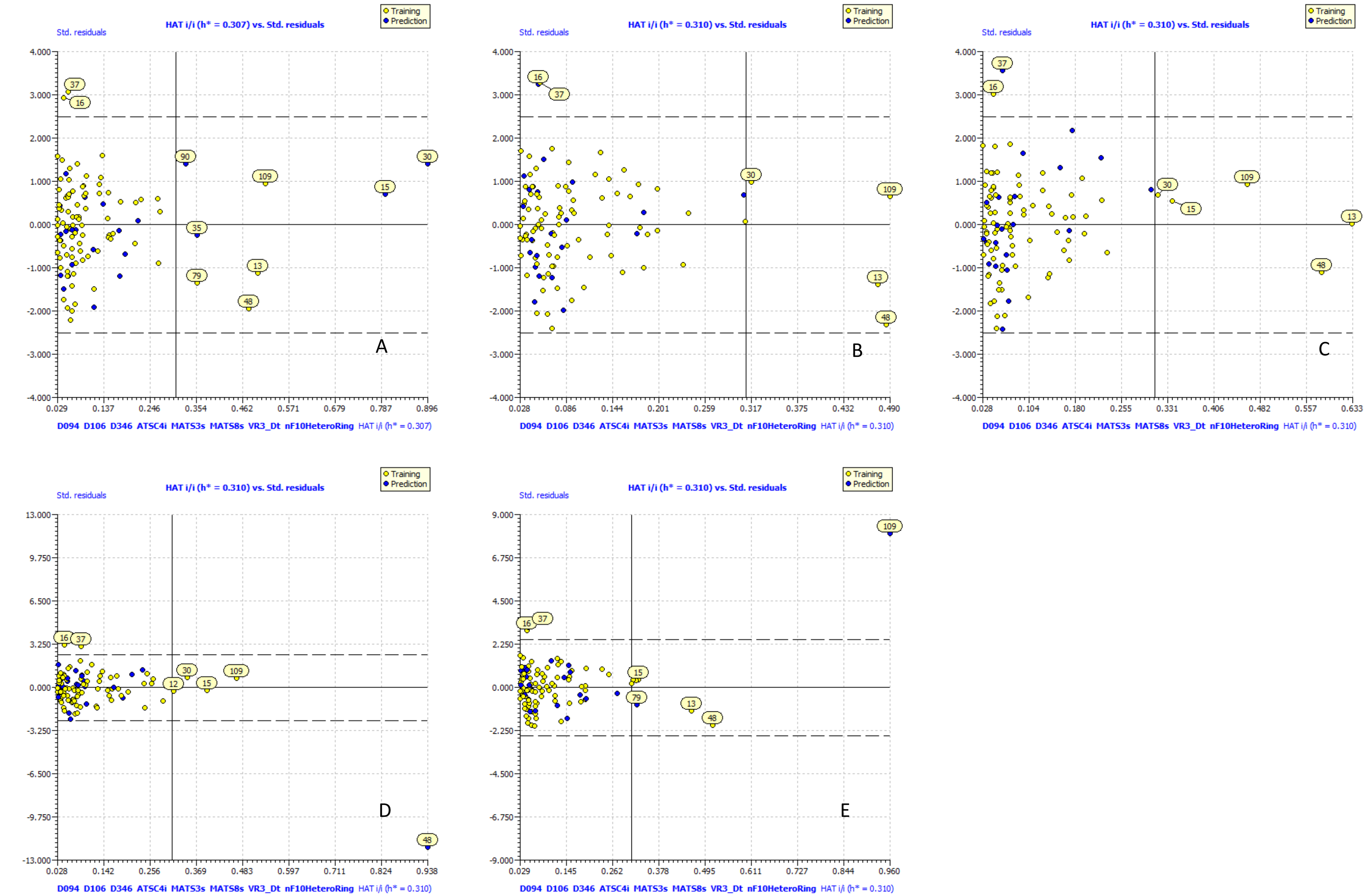


Figure S7. Comparison of the predictive performances of linear and non-linear approaches applied to predict logHUVEC, tested on five external validation sets on the basis of Mean Absolute Errors (MAE) and Root Mean Squared Errors (RMSE). Combined=calculated on averaged predictions.

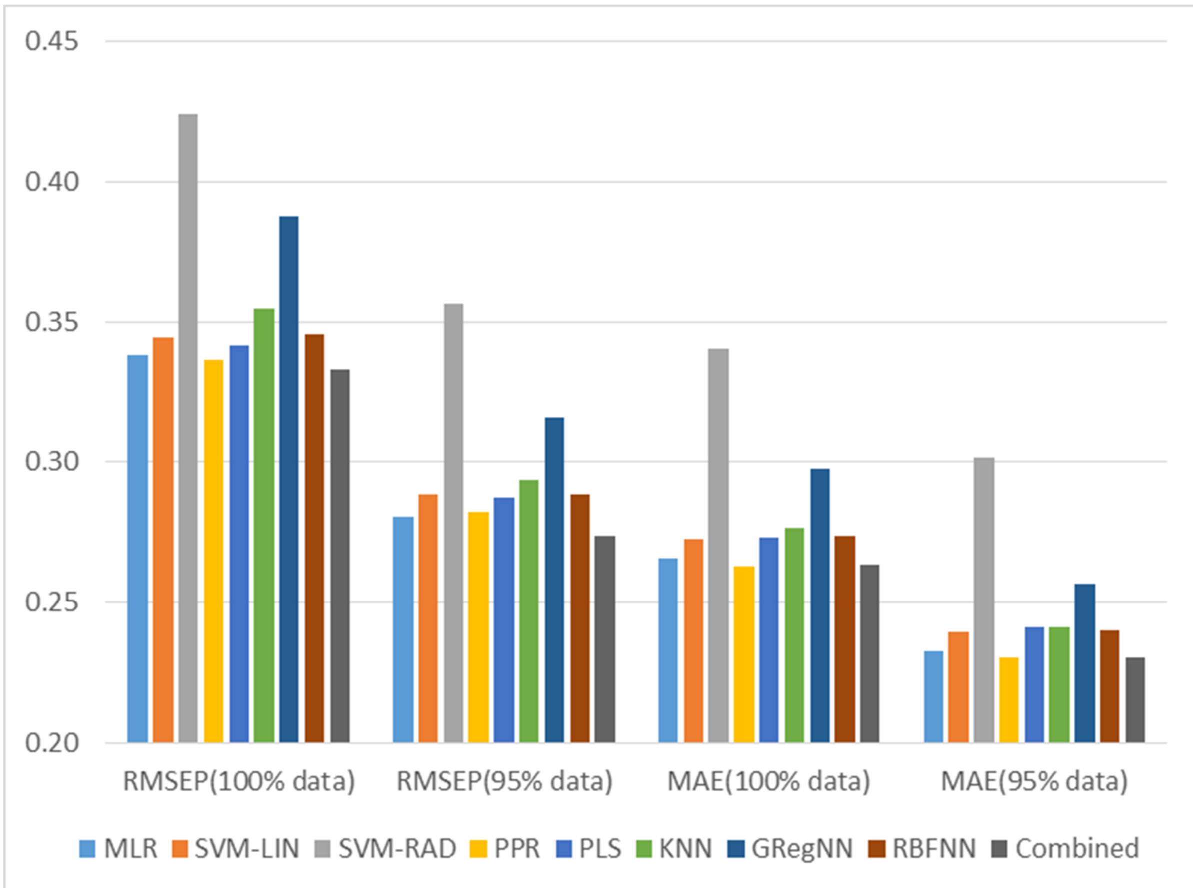


Figure S8 A-B: Plots of the first 2 principal components calculated on residuals in prediction from 8 different modelling techniques applied to the log HUVEC response. Figure A: score plot (1=anhydrides; 2=amines; 3=aminoacids). Figure B: loading plot.

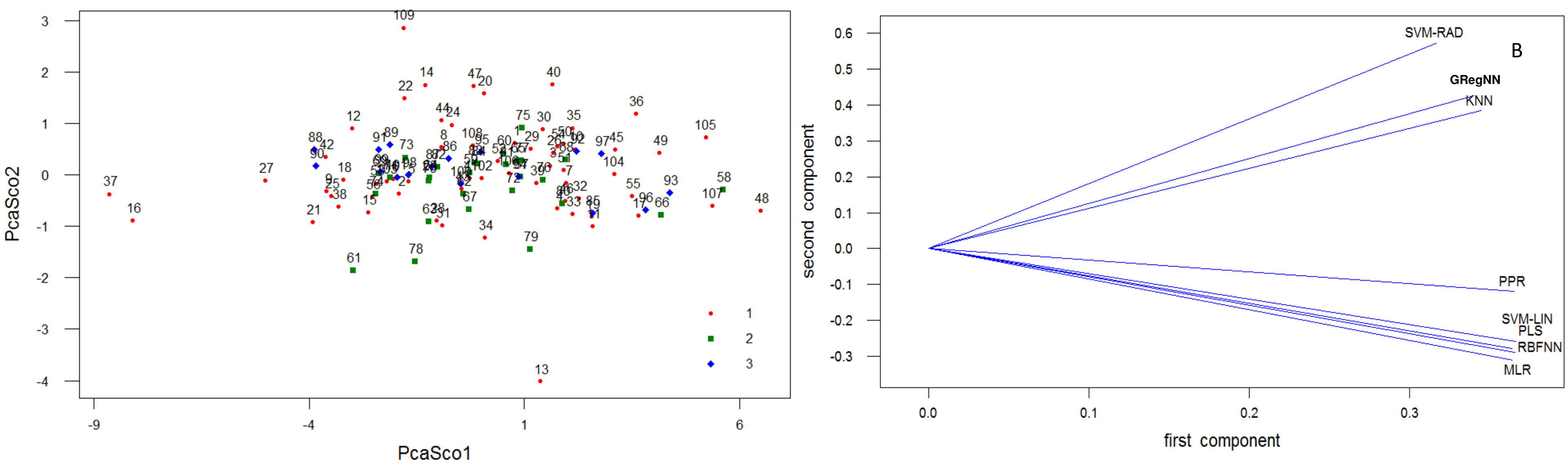


Figure S9 A-B: Plots of the first 2 principal components calculated on log-transformed uptake values in five human cell types. Figure A: score plot. Figure B: loading plot.

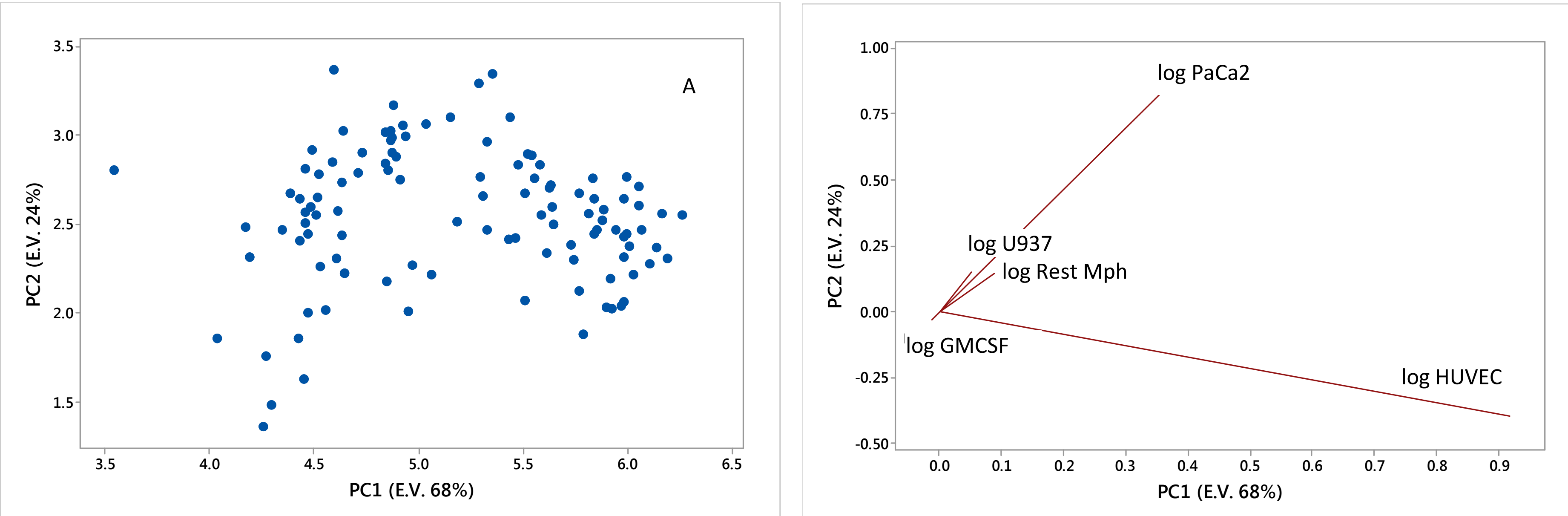


Figure S10: Loading plot calculated on the first two rotated factors calculated on log-transformed uptake values in five human cells.

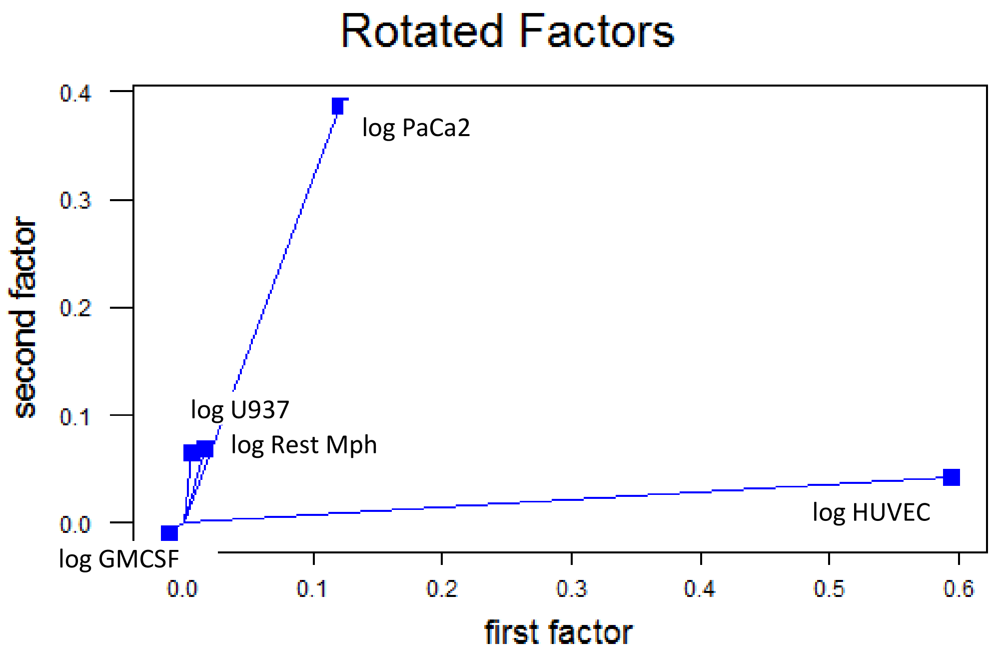
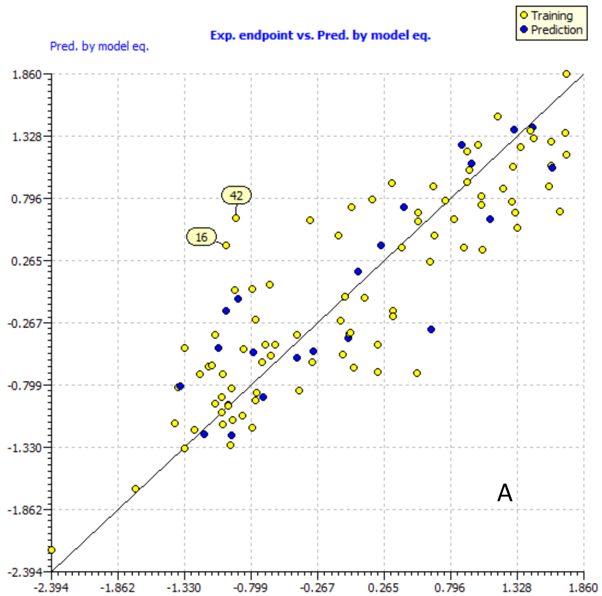


Figure S11 A-B: Plot of experimental vs. predicted values for the best F1 model (A); Plot of hat diagonal values vs. standardized residuals for the best F1 model (B) (yellow dots=training set; blue dots=prediction set).



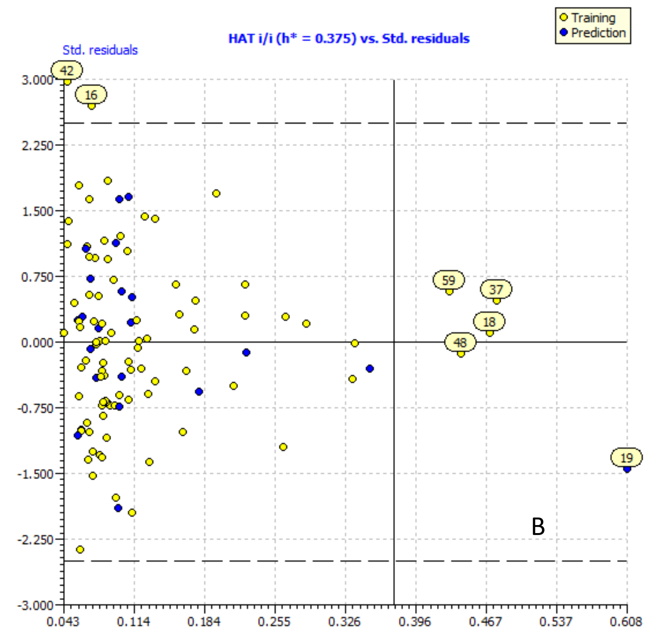
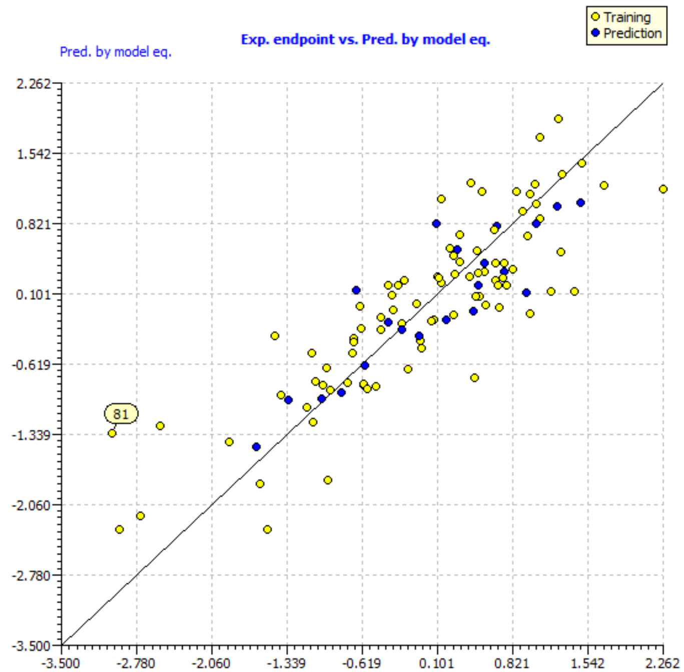


Figure S12 A-B: Plot of experimental vs. predicted values for the best F2 model (A); Plot of hat diagonal values vs. standardized residuals for the best F1 model (B) (yellow dots=training set; blue dots=prediction set).



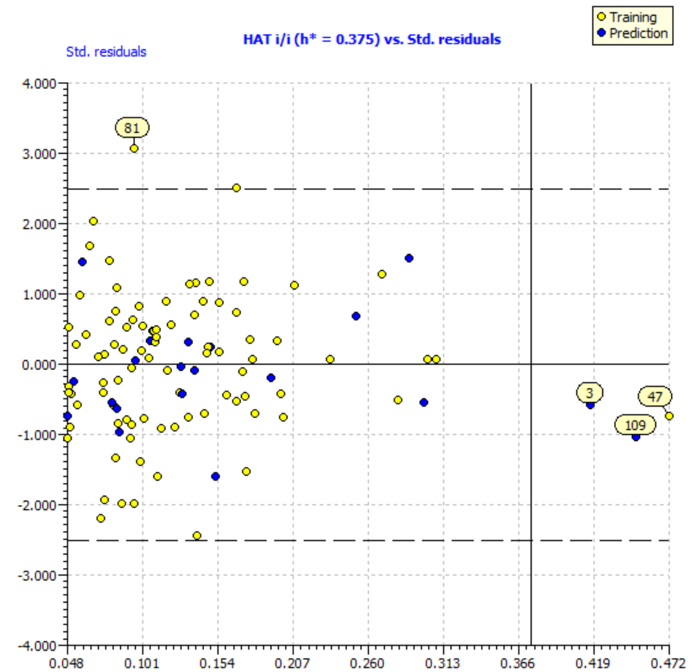


Figure S13: Rotated score plot of the two rotated factors F1 and F2 for 109 NPs and 28 surface substituents (predicted values). (Class= chemical class, i.e. 1=anhydrides; 2=amines; 3=aminoacids).

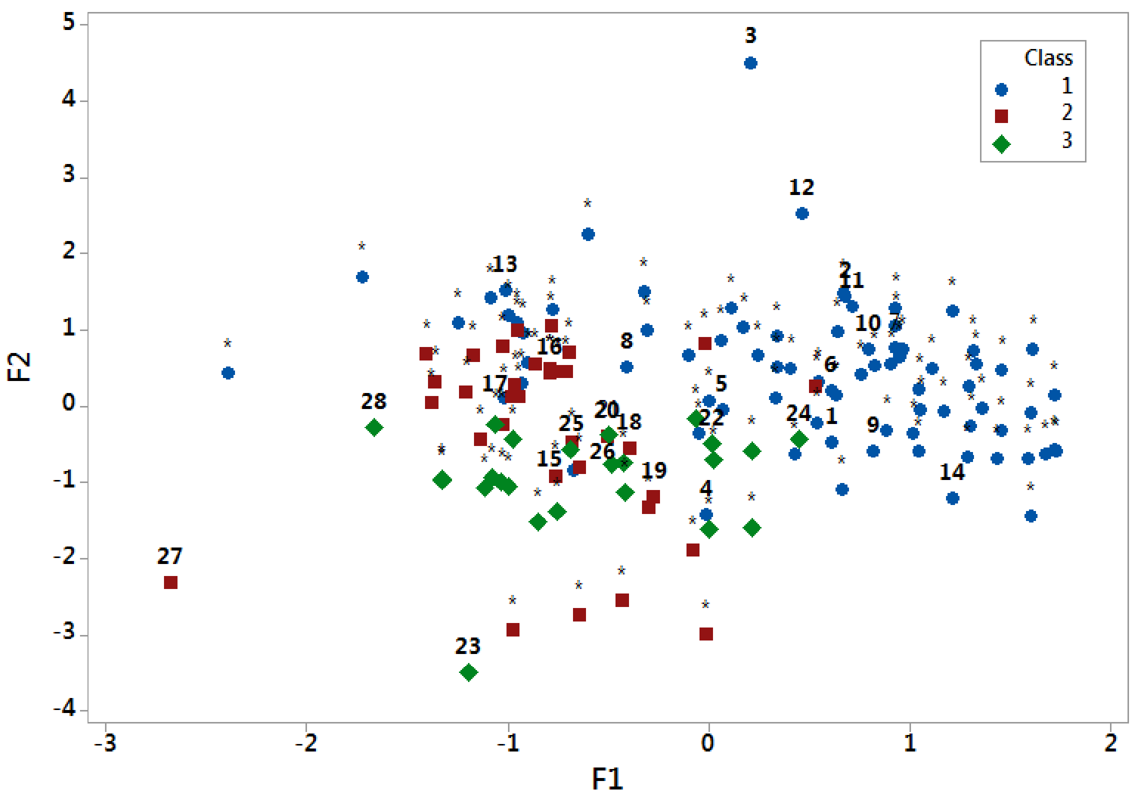


Figure S14 A-C: Plot of experimental vs. predicted values for F1 Full model (i.e. calibrated on empirical F1 values available for all the 109 NPs) (A). Plot of hat diagonal values vs. standardized residuals for F1 Full model (B). Plot of hat diagonal values vs. predicted values for F1 Full model (C) (yellow dots=training set; res dots= 28 chemicals with unknown uptake).

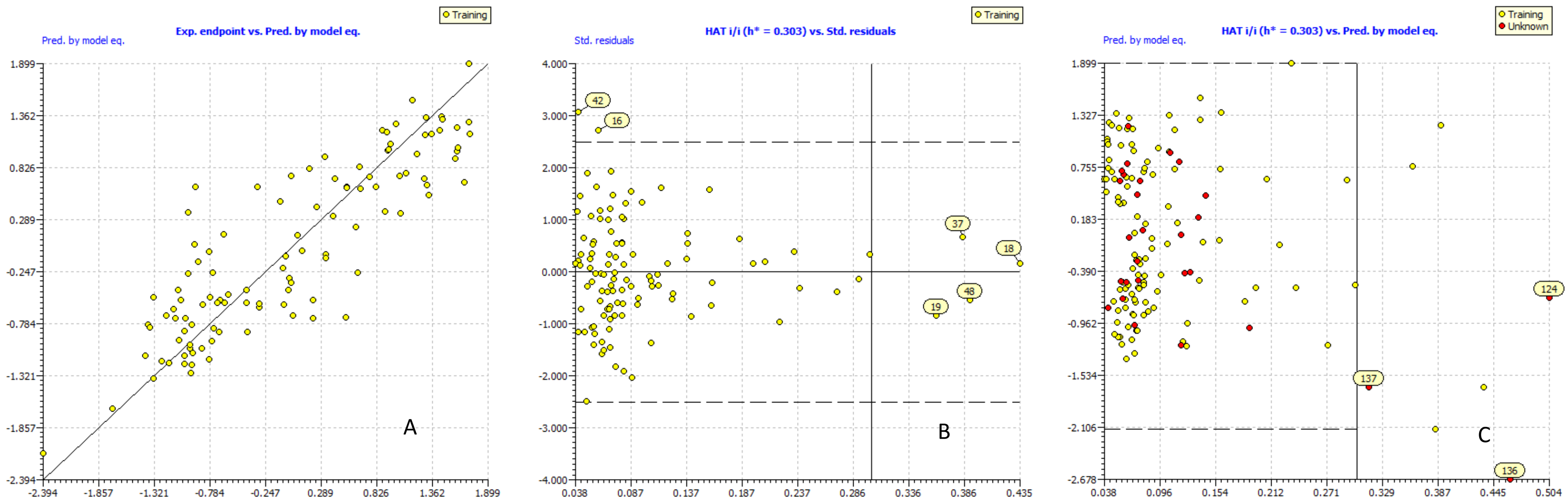


Figure S15 A-C: Plot of experimental vs. predicted values for F2 Full model (i.e. calibrated on empirical F2 values available for all the 109 NPs) (A). Plot of hat diagonal values vs. standardized residuals for F2 Full model (B). Plot of hat diagonal values vs. predicted values for F2 Full model (C) (yellow dots=training set; res dots= 28 chemicals with unknown uptake).

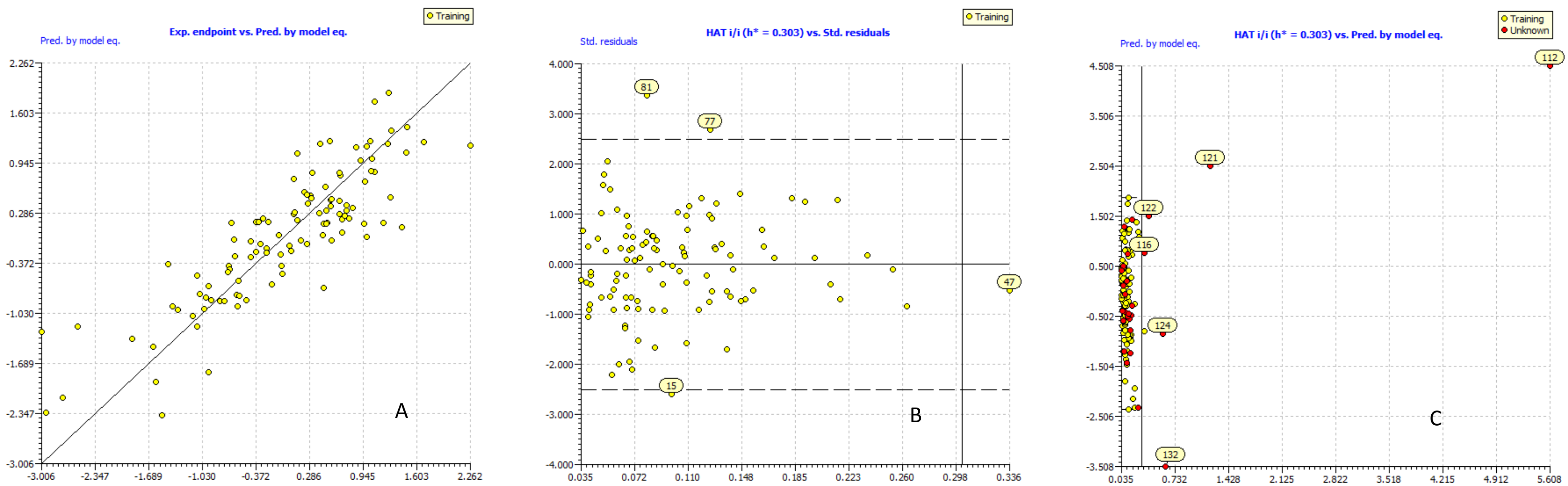


Table S1. List of ID, names and log transformed values of uptake in five human cell types (picomoles/L (pM)), measured for a library of 109 functionalized nanoparticles, in addition to 28 structures (ID nos° 110 – 137, *in Italic*) with unknown cellular uptake (n.a.= not available). Columns with the splittings applied to generate .MLR-OLS models and to calibrate additional linear and non-linear models are also reported (i.e. 1= Training set, 2=Prediction set.)

ID 137	ID 109	Name	log PaCa2 (pM)	log HUVEC (pM)	log GMCSF Mph (pM)	log Rest Mph (pM)	log U937 (pM)	Split M0	Split M1	Split M2	Split M3	Split M4
1	**1**	**trifluoroacetic anhydride**	4.172	4.405	2.915	2.715	3.045	1	2	1	1	1
2	**2**	**chlorodifluoroacetic anhydride**	3.952	4.082	2.856	2.925	3.007	1	1	2	1	1
3	**3**	**pentafluopropionic anhydride**	4.077	4.525	2.893	2.615	2.960	1	1	1	2	1
4	**4**	**dimethysucccinic anhydride**	4.111	3.939	2.825	2.735	2.959	1	1	1	1	2
5	**5**	**maleic anhydride**	3.979	3.998	2.834	2.789	3.023	2	1	1	1	1
6	**6**	**citraconic anhydride**	3.583	4.044	2.973	2.966	3.109	1	2	1	1	1
7	**7**	**2,3-dimethymaleic anhydride**	3.483	4.635	2.843	2.528	2.985	1	1	2	1	1
8	**8**	**hexanoic anhydride**	3.654	3.912	2.898	2.586	3.010	1	1	1	2	1
9	**9**	**methysuccinic anhydride**	3.639	3.168	2.837	2.495	2.998	1	1	1	1	2
10	**10**	**3,3',4,4'-benzophenontetracarboxylic dianhydride**	3.511	4.672	2.848	2.538	3.000	2	1	1	1	1
11	**11**	**4-nitrophthalic anhydride**	3.270	4.568	2.820	2.611	3.015	1	2	1	1	1
12	**12**	**4-bromo-1,8-naphthalic anhydride**	3.626	4.027	2.883	2.604	3.068	1	1	2	1	1
13	**13**	**1,4,5,8-naphtalenetrtracarboxylic anhydride**	3.669	3.212	2.798	2.797	2.955	1	1	1	2	1
14	**14**	**tetrafluorophthalic anhydride**	3.827	4.386	2.802	2.637	3.004	1	1	1	1	2
15	**15**	**3-nitro-1,8-naphthalic anhydride**	4.112	3.219	2.884	2.652	3.076	2	1	1	1	1
16	**16**	**3-hydroxyphthalic anhydride**	3.972	3.247	2.826	2.636	2.959	1	2	1	1	1
17	**17**	**endo-bicyclo [2,2,2] oct-5-ene-2,3-dicarboxylic anhydride**	3.901	4.539	2.849	2.559	3.012	1	1	2	1	1
18	**18**	**5-chloroisatoic anhydride**	4.176	2.882	2.888	2.701	3.068	1	1	1	2	1
19	**19**	**2-sulfobenzoic acid cyclic anhydride**	3.880	4.396	2.749	2.476	2.983	1	1	1	1	2
20	**20**	**dicholoromaleic anhydride**	3.840	4.468	2.798	2.912	3.009	2	1	1	1	1
21	**21**	**5-acetylmercaptosuccinic anhydride**	3.587	3.754	2.894	2.859	3.023	1	2	1	1	1
22	**22**	**4,5-dicholorophthalic anhydride**	4.117	4.478	2.772	2.898	2.999	1	1	2	1	1
23	**23**	**exo-3,6-epoxy-1,2,3,6-tetrahydrophthalic anhydride**	3.819	4.113	2.822	2.709	2.981	1	1	1	2	1
24	**24**	**bicyclo [2,2,2] oct-7-ene-2,3,5,6-tetracarboxylic anhydride**	3.628	4.354	2.849	2.761	2.996	1	1	1	1	2
25	**25**	**cis-1,2,3,6-tetrahydrophthalic anhydride**	3.891	3.760	2.918	2.752	3.034	2	1	1	1	1
26	**26**	**diphenic anhydride**	3.772	4.701	2.774	2.661	2.993	1	2	1	1	1
27	**27**	**4-nitro-1,8-naphthalic anhydride**	3.932	3.343	2.839	2.841	3.075	1	1	2	1	1
28	**28**	**1,2,3,4-cyclobutanetetra carboxylic dianhydride**	3.768	3.816	2.903	2.838	3.070	1	1	1	2	1
29	**29**	**lauric anhydride**	3.821	4.523	2.847	2.903	3.024	1	1	1	1	2
30	**30**	**1,2,4-benzenetricarboxylic anhydride**	3.550	4.215	2.792	3.049	3.048	2	1	1	1	1

31	**31**	**4-methylphthalic anhydride**	3.979	4.305	2.763	2.808	3.020	1	2	1	1	1
32	**32**	**3-nitrophthalic anhydride**	3.498	4.449	2.793	2.621	2.965	1	1	2	1	1
33	**33**	**homophthalic anhydride**	3.777	4.182	2.829	2.808	2.961	1	1	1	2	1
34	**34**	**glutaric anhydride**	4.074	3.407	2.823	2.694	3.004	1	1	1	1	2
35	**35**	**ethylenediamintetraacetic anhydride**	3.929	3.996	2.830	2.610	2.936	2	1	1	1	1
36	**36**	**isatoic anhydride**	4.443	3.587	2.987	2.759	3.124	1	2	1	1	1
37	**37**	**N-methylisatoic anhydride**	3.363	2.152	2.710	2.007	2.853	1	1	2	1	1
38	**38**	**3-methylglutaric anhydride**	3.906	3.063	2.826	2.551	2.980	1	1	1	2	1
39	**39**	**3,4,5,6-tetrahydrophthalic anhydride**	3.727	4.569	2.832	2.856	2.998	1	1	1	1	2
40	**40**	**diacetyl-L-tartaric anhydride**	3.909	4.584	2.854	2.803	3.035	2	1	1	1	1
41	**41**	**Tetra bromophthalic anhydride**	3.804	4.574	2.853	2.718	3.068	1	2	1	1	1
42	**42**	**1,2-cyclohexanedicarboxy anhydride**	3.925	3.289	2.845	2.685	2.988	1	1	2	1	1
43	**43**	**1-cyclopentene-1,2-dicarboxy anhydride**	3.690	4.309	2.826	2.815	3.016	1	1	1	2	1
44	**44**	**Iodoacetic anhydride**	3.423	4.236	2.521	2.284	2.798	1	1	1	1	2
45	**45**	**Chloroacetic anhydride**	3.633	4.559	2.794	2.574	2.995	2	1	1	1	1
46	**46**	**Hexachloro-5-norborenene anhydride**	3.474	4.604	2.815	2.592	2.960	1	2	1	1	1
47	**47**	**Palmitic anhydride**	3.550	4.678	2.738	2.489	2.967	1	1	2	1	1
48	**48**	**4-amino-1,8-naphthalic anhydride**	3.640	4.644	2.756	2.905	2.972	1	1	1	2	1
49	**49**	**Decanoic ahd**	4.026	4.732	2.989	2.962	3.116	1	1	1	1	2
50	**50**	**3,3'-Tetramethyleneglutaric anhydride**	4.058	3.707	2.959	2.881	3.086	2	1	1	1	1
51	**51**	**Cis-5-norborne-endo-2,3-dicarboxylic anhydride**	3.939	4.373	2.900	2.826	3.010	1	2	1	1	1
52	**52**	**1,8-naphthalic anhydride**	3.955	4.087	2.821	2.790	3.040	1	1	2	1	1
53	**53**	**2-phenylglutaric anhydride**	4.018	3.898	2.850	2.783	3.018	1	1	1	2	1
54	**54**	**Tetrachlorophthalic anhydride**	3.833	4.764	2.863	2.638	3.044	1	1	1	1	2
55	**55**	**3,6-dicholorphthalic anhydride**	3.904	4.695	2.867	2.532	2.947	2	1	1	1	1
56	**56**	**2,2-dimethylglutaric anhydride**	3.937	3.284	2.899	2.745	2.968	1	2	1	1	1
57	**57**	**Amylamine**	3.783	3.313	2.851	2.761	2.951	1	1	2	1	1
58	**58**	**1,3-dimethylbutylamine**	3.853	4.142	2.880	2.835	3.035	1	1	1	2	1
59	**59**	**D-Glucosamine**	3.355	3.281	2.790	2.589	2.886	1	1	1	1	2
60	**60**	**Hexylamine**	3.748	3.337	2.939	2.767	3.001	2	1	1	1	1
61	**61**	**Tert-butylamine**	3.856	3.342	2.981	2.698	2.969	1	2	1	1	1
62	**62**	**isobutylamine**	3.716	3.082	2.855	2.620	2.923	1	1	2	1	1
63	**63**	**dimethylpropylamine**	3.748	2.966	2.746	2.608	2.881	1	1	1	2	1
64	**64**	**isoamylamine**	3.828	3.191	2.777	2.590	2.944	1	1	1	1	2
65	**65**	**1-ethylpropylamine**	3.815	3.316	2.766	2.908	3.033	2	1	1	1	1
66	**66**	**Tert-amylamine**	4.068	3.911	2.848	2.965	3.017	1	2	1	1	1
67	**67**	**Ethylenediamine**	3.462	3.084	2.689	2.468	2.907	1	1	2	1	1
68	**68**	**pentadecylamine**	4.061	3.493	2.963	3.055	3.260	1	1	1	2	1
69	**69**	**1,3-diaminopropane**	3.486	2.960	2.660	2.693	2.727	1	1	1	1	2
70	**70**	**1,4-diaminobutane**	3.484	3.086	2.705	2.690	2.932	2	1	1	1	1
71	**71**	**Hexamethylenediamine**	3.623	3.047	2.748	2.600	2.926	1	2	1	1	1
72	**72**	**2-Ethylhexylamine**	3.950	3.303	2.950	2.899	3.279	1	1	2	1	1
73	**73**	**1-hexadecylamine**	3.967	3.268	2.892	2.673	3.069	1	1	1	2	1
74	**74**	**2-aminoheptane**	3.629	2.975	2.776	2.657	2.889	1	1	1	1	2
75	**75**	**1-tetradecylamine**	4.274	3.548	2.988	3.013	3.441	2	1	1	1	1
76	**76**	**Diethylenetriamine**	3.772	3.414	2.501	2.651	2.729	1	2	1	1	1
77	**77**	**1-adamantanemethylamine**	2.837	3.375	2.979	2.932	3.092	1	1	2	1	1
78	**78**	**3-hydroxytyramine**	2.530	2.962	2.860	2.741	2.923	1	1	1	2	1

79	**79**	**Tyramine**	2.766	3.323	3.044	2.493	2.767	1	1	1	1	2
80	**80**	**Spermine**	2.407	3.351	2.985	2.368	2.574	2	1	1	1	1
81	**81**	**N,N'-Bis(2-aminoethyl)-1,3-propanediamine**	2.234	3.366	2.955	2.447	2.749	1	2	1	1	1
82	**82**	**Pentaethylenehexamine**	2.541	3.207	2.810	2.881	2.733	1	1	2	1	1
83	**83**	**3-Noradamantanamine**	3.118	3.248	2.954	2.868	2.804	1	1	1	2	1
84	**84**	**2-adamantanamine**	3.180	3.305	3.076	2.843	2.934	1	1	1	1	2
85	**85**	**Gly**	2.569	3.439	3.141	2.472	2.731	2	1	1	1	1
86	**86**	**L-Phenylalanine Methyl Ester**	3.391	3.219	2.956	2.643	3.399	1	2	1	1	1
87	**87**	**L-Ser**	3.357	3.074	2.808	2.700	3.063	1	1	2	1	1
88	**88**	**L-Thr**	3.215	2.845	2.790	2.485	3.003	1	1	1	2	1
89	**89**	**L-Trp**	3.186	3.115	2.970	2.606	3.250	1	1	1	1	2
90	**90**	**L-Tyr**	3.070	2.929	2.886	2.482	2.954	2	1	1	1	1
91	**91**	**L-Val**	3.271	3.017	3.056	2.434	3.077	1	2	1	1	1
92	**92**	**L-Lys**	3.249	3.582	2.835	2.457	2.693	1	1	2	1	1
93	**93**	**4-cholorophenyl alanine**	3.055	3.741	2.904	2.689	2.964	1	1	1	2	1
94	**94**	**L-Ala**	2.905	3.303	3.008	2.542	2.810	1	1	1	1	2
95	**95**	**L-Arg**	3.149	3.374	3.062	2.664	2.927	2	1	1	1	1
96	**96**	**L-Asp**	3.288	3.767	3.141	2.686	3.028	1	2	1	1	1
97	**97**	**L-Gln**	3.322	3.661	2.919	2.592	2.989	1	1	2	1	1
98	**98**	**L-Glu**	3.399	3.120	2.832	2.755	2.829	1	1	1	2	1
99	**99**	**L-His**	3.382	3.009	2.818	2.961	3.021	1	1	1	1	2
100	**100**	**L-Met**	3.232	3.123	2.963	2.914	2.973	2	1	1	1	1
101	**101**	**L-Phe**	3.294	3.097	2.928	2.714	3.122	1	2	1	1	1
102	**102**	**succinic anhydride**	4.235	3.758	2.895	2.913	3.029	1	1	2	1	1
103	**103**	**acetic anhydride**	4.047	3.991	2.956	2.899	3.012	1	1	1	2	1
104	**104**	**Itaconic anhydride**	4.038	4.434	2.950	2.926	2.923	1	1	1	1	2
105	**105**	**Diglycolic anhydride**	3.991	4.226	2.957	2.820	2.976	2	1	1	1	1
106	**106**	**phthalic anhydride**	3.896	4.388	2.905	2.821	2.927	1	2	1	1	1
107	**107**	**cis-aconitic anhydride**	4.026	4.639	2.939	2.884	3.018	1	1	2	1	1
108	**108**	**diflorophthalic anhydride**	3.910	4.315	2.803	2.766	2.939	1	1	1	2	1
109	**109**	**diethylenetriaminepentaacetic dianhydride**	4.097	4.256	3.072	2.829	2.984	1	1	1	1	2
110	**1**	***(2-Dodecen-1-yl)succinic anhydride***	n.a.	n.a.	n.a.	n.a.	n.a.	2	2	2	2	2
111	**2**	***Propionic Anhydride***	n.a.	n.a.	n.a.	n.a.	n.a.	2	2	2	2	2
112	**3**	***trymethylacetic anhydride***	n.a.	n.a.	n.a.	n.a.	n.a.	2	2	2	2	2
113	**4**	***trychloroacetic anhydride***	n.a.	n.a.	n.a.	n.a.	n.a.	2	2	2	2	2
114	**5**	***Hexahydro 4 methylphtalic anhydride***	n.a.	n.a.	n.a.	n.a.	n.a.	2	2	2	2	2
115	**6**	***Crotonic Anhydride***	n.a.	n.a.	n.a.	n.a.	n.a.	2	2	2	2	2
116	**7**	***heptafluorobutyric anhydride***	n.a.	n.a.	n.a.	n.a.	n.a.	2	2	2	2	2
117	**8**	***methyl-5-norborne-2,3-di anhydrde***	n.a.	n.a.	n.a.	n.a.	n.a.	2	2	2	2	2
118	**9**	***dichloro acetic anhydride***	n.a.	n.a.	n.a.	n.a.	n.a.	2	2	2	2	2
119	**10**	***valeric anhydride***	n.a.	n.a.	n.a.	n.a.	n.a.	2	2	2	2	2
120	**11**	***butyric anhydride***	n.a.	n.a.	n.a.	n.a.	n.a.	2	2	2	2	2
121	**12**	***Isobutyric anhydride***	n.a.	n.a.	n.a.	n.a.	n.a.	2	2	2	2	2
122	**13**	***Hexafluoroglutaric anhydride***	n.a.	n.a.	n.a.	n.a.	n.a.	2	2	2	2	2
123	**14**	***benzoic anhydride***	n.a.	n.a.	n.a.	n.a.	n.a.	2	2	2	2	2
124	**15**	***3,4,9,10-perylenetetracarboxylic dianhydride***	n.a.	n.a.	n.a.	n.a.	n.a.	2	2	2	2	2
125	**16**	***sec-butylamine***	n.a.	n.a.	n.a.	n.a.	n.a.	2	2	2	2	2
126	**17**	***Undecylamine***	n.a.	n.a.	n.a.	n.a.	n.a.	2	2	2	2	2
127	**18**	***Bis(hexamethylene)-ethylenediamine***	n.a.	n.a.	n.a.	n.a.	n.a.	2	2	2	2	2

128	**19**	***N,N'-Bis(3-aminopropyl)-ethylenediamine***	n.a.	n.a.	n.a.	n.a.	n.a.	2	2	2	2	2
129	**20**	***Spermidine***	n.a.	n.a.	n.a.	n.a.	n.a.	2	2	2	2	2
130	**21**	***N-Acetyl-L-arginine***	n.a.	n.a.	n.a.	n.a.	n.a.	2	2	2	2	2
131	**22**	***Boc-arg-OH***	n.a.	n.a.	n.a.	n.a.	n.a.	2	2	2	2	2
132	**23**	***N-2,4-DNP-L-arginine***	n.a.	n.a.	n.a.	n.a.	n.a.	2	2	2	2	2
133	**24**	***2-imino-l-imidazoliacetic acid***	n.a.	n.a.	n.a.	n.a.	n.a.	2	2	2	2	2
134	**25**	***Allo-octopine***	n.a.	n.a.	n.a.	n.a.	n.a.	2	2	2	2	2
135	**26**	***D-arginine***	n.a.	n.a.	n.a.	n.a.	n.a.	2	2	2	2	2
136	**27**	***N-Ethyl-N-(2-hydroxyethyl)-4-(4-nitrophenylazo)aniline***	n.a.	n.a.	n.a.	n.a.	n.a.	2	2	2	2	2
137	**28**	***3-Mercapto-2-methylpropionyl-L-proline***	n.a.	n.a.	n.a.	n.a.	n.a.	2	2	2	2	2

Table S2. Statistical parameters calculated for the best MLR-OLS models developed for PaCa2 and HUVEC cells.

(*) Parameters calculated with the software Xternal Validation Plus available online at http://dtclab.webs.com/software-tools.

PaCa2	**R^2**	**Q^2**	**Q^2lmo**	**CCC training**	**R^2Y Scrambling**	**RMSE training**	**MAE Training**	**R^2ext**	**Average Q^2ext**	**CCC ext**	**RMSE ext * (100% data)**	**MAE ext * (100% data)**	**RMSE ext * (95% data)**	**MAE ext * (95% data)**	**MAE+3*SD ext * (95% data)**	**Prediction Quality ***
M0	0.73	0.67	0.64	0.85	0.09	0.21	0.16	0.75	0.72	0.85	0.24	0.19	0.20	0.15	0.53	Good
M1	0.77	0.71	0.72	0.87	0.09	0.2	0.16	0.65	0.62	0.75	0.27	0.20	0.20	0.16	0.53	Good
M2	0.76	0.68	0.63	0.86	0.09	0.21	0.16	0.66	0.66	0.81	0.24	0.20	0.20	0.17	0.49	Good
M3	0.71	0.64	0.58	0.83	0.09	0.23	0.18	0.88	0.86	0.91	0.15	0.12	0.12	0.09	0.31	Good
M4	0.74	0.67	0.63	0.85	0.09	0.22	0.17	0.75	0.75	0.86	0.19	0.16	0.18	0.14	0.46	Good
Full Model	0.74	0.69	0.66	0.85	0.07	0.21	0.17	-	-	-	-	-	-	-	-	-
HUVEC	**R^2**	**Q^2**	**Q^2lmo**	**CCC training**	**R^2Y Scrambling**	**RMSE training**	**MAE Training**	**R^2ext**	**Average Q^2ext**	**CCC ext**	**RMSE ext * (100% data)**	**MAE ext * (100% data)**	**RMSE ext * (95% data)**	**MAE ext * (95% data)**	**MAE+3*SD ext * (95% data)**	**Prediction Quality ***
M0	0.74	0.66	0.65	0.85	0.09	0.31	0.25	0.8	0.8	0.88	0.26	0.21	0.21	0.17	0.56	Good
M1	0.77	0.7	0.68	0.87	0.09	0.29	0.23	0.63	0.64	0.79	0.35	0.28	0.26	0.23	0.64	Good
M2	0.76	0.72	0.7	0.87	0.09	0.29	0.23	0.66	0.63	0.8	0.38	0.29	0.29	0.23	0.75	Good
M3	0.77	0.72	0.7	0.87	0.09	0.29	0.23	0.56	0.57	0.74	0.38	0.29	0.30	0.24	0.81	Good
M4	0.72	0.65	0.63	0.84	0.09	0.31	0.24	0.8	0.79	0.89	0.29	0.26	0.26	0.23	0.58	Good
Full Model	0.75	0.7	0.68	0.86	0.07	0.3	0.24	-	-	-	-	-	-	-	-	-

Table S3. Regression coefficients calculated for PaCa2 and HUVEC MLR-OLS QSAR models on the five external prediction set and on the full dataset. Variables are listed in decreasing order of importance according to standardized coefficients.

PaCa2	**M0**	**M1**	**M2**	**M3**	**M4**	**Full Model**	**HUVEC**	**M0**	**M1**	**M2**	**M3**	**M4**	**Full Model**
Intercept and descriptors	**Regr. Coeff.**	**Regr. Coeff.**	**Regr. Coeff.**	**Regr. Coeff.**	**Regr. Coeff.**	**Regr. Coeff.**	**Intercept and descriptors**	**Regr. Coeff.**	**Regr. Coeff.**	**Regr. Coeff.**	**Regr. Coeff.**	**Regr. Coeff.**	**Regr. Coeff.**
Intercept	4.478	4.422	4.596	4.563	4.583	4.583	**Intercept**	3.937	3.975	3.889	3.982	3.866	3.929
nBase	-0.231	-0.226	-0.240	-0.233	-0.230	-0.230	**D106**	5.991	6.220	5.674	6.433	5.324	5.897
VE3_Dzs	0.010	0.011	0.011	0.011	0.012	0.012	**nF10HeteroRing**	-0.316	-0.331	-0.244	-0.419	-0.315	-0.316
ATSC1v	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002	**D346**	0.003	0.003	0.004	0.003	0.003	0.003
BIC2	-0.963	-0.860	-1.080	-1.058	-1.122	-1.122	**VR3_Dt**	0.016	0.015	0.019	0.015	0.017	0.016
VE2_Dze	-12.072	-14.492	-15.049	-13.484	-13.826	-13.826	**D094**	-47.645	-42.304	-34.515	-38.486	-44.557	-40.769
D070	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	**MATS3s**	-0.482	-0.479	-0.421	-0.484	-0.666	-0.510
maxHssNH	0.639	0.622	0.669	0.694	0.712	0.712	**ATSC4i**	-0.010	-0.009	-0.008	-0.009	-0.014	-0.009
MATS7s	-0.087	-0.080	-0.103	-0.100	-0.106	-0.106	**MATS8s**	-0.225	-0.276	-0.275	-0.210	-0.202	-0.243

Table S4. Residuals calculated for logPaCa2 QSAR models developed by eight different methodologies, and on the basis of combined (i.e. averaged) predictions.

ID109	Res GRegNN	Res RBFNN	Res SVM-LIN	Res SVM-RAD	Res PPR	Res PLS	Res KNN	Res MLR	Res Combi.
1	0.24	0.09	0.10	0.26	0.17	0.12	0.25	0.11	0.17
2	0.07	0.12	0.12	0.07	0.13	0.18	0.08	0.18	0.12
3	-0.09	0.08	-0.04	0.39	0.18	0.04	0.07	0.04	0.08
4	0.32	0.19	0.18	0.17	0.26	0.22	0.27	0.23	0.23
5	0.18	-0.02	-0.06	0.02	0.08	-0.03	0.11	-0.03	0.03
6	-0.13	-0.12	-0.15	-0.10	-0.13	-0.09	-0.25	-0.08	-0.13
7	-0.39	-0.43	-0.35	-0.59	-0.39	-0.40	-0.41	-0.40	-0.42
8	-0.18	-0.30	-0.33	-0.19	-0.20	-0.30	-0.20	-0.28	-0.25
9	-0.02	-0.02	-0.10	-0.01	-0.12	-0.05	-0.15	-0.03	-0.06
10	-0.30	-0.43	-0.42	-0.33	-0.24	-0.40	-0.56	-0.40	-0.39
11	-0.55	-0.24	-0.23	-0.35	-0.16	-0.23	-0.42	-0.18	-0.30
12	-0.24	-0.32	-0.24	-0.40	-0.25	-0.32	-0.30	-0.32	-0.30
13	-0.16	-0.32	-0.34	0.12	-0.14	-0.32	-0.19	-0.33	-0.21
14	-0.05	-0.32	-0.28	-0.09	-0.30	-0.30	0.02	-0.32	-0.21
15	0.37	0.36	0.34	0.35	0.33	0.36	0.32	0.35	0.35
16	0.15	0.25	0.25	0.30	0.17	0.25	0.16	0.27	0.23
17	0.07	0.18	0.26	0.06	0.13	0.21	0.03	0.21	0.14
18	-0.27	-0.02	-0.21	-0.09	-0.12	-0.13	0.39	-0.10	-0.07
19	0.07	0.05	0.07	0.07	0.05	0.07	-0.01	0.08	0.06
20	0.07	-0.05	-0.03	-0.13	-0.17	-0.03	-0.16	-0.03	-0.07
21	-0.08	-0.16	-0.22	-0.27	-0.24	-0.16	0.14	-0.16	-0.14
22	0.26	0.21	0.28	0.19	0.24	0.23	0.34	0.23	0.25
23	0.00	-0.02	-0.03	-0.07	-0.04	-0.01	-0.01	-0.01	-0.02
24	0.30	0.21	0.21	0.12	0.03	0.10	0.00	0.13	0.14
25	0.18	0.15	0.17	0.08	0.16	0.16	0.18	0.16	0.16
26	-0.08	-0.08	-0.10	-0.02	-0.06	-0.09	-0.08	-0.11	-0.08
27	0.10	0.30	0.30	0.39	0.24	0.30	0.25	0.30	0.27
28	0.02	-0.01	-0.01	0.27	-0.04	0.02	0.07	0.02	0.04
29	0.27	0.19	0.10	0.27	0.10	0.16	0.12	0.19	0.18
30	-0.14	-0.09	-0.07	-0.15	-0.18	-0.06	-0.36	-0.06	-0.14
31	0.13	0.12	0.07	0.17	0.07	0.12	0.11	0.10	0.11
32	-0.33	-0.19	-0.17	-0.43	-0.27	-0.20	-0.34	-0.20	-0.27
33	-0.04	-0.16	-0.19	-0.08	-0.02	-0.18	-0.13	-0.18	-0.12
34	0.19	0.02	0.03	0.08	0.04	-0.01	0.03	-0.02	0.04
35	0.07	-0.08	-0.16	-0.12	-0.07	-0.11	-0.14	-0.11	-0.09
36	0.27	0.23	0.14	0.52	0.28	0.14	0.64	0.14	0.30
37	-0.51	-0.54	-0.53	-0.70	-0.52	-0.58	-0.50	-0.57	-0.56
38	0.19	0.14	0.16	0.12	0.15	0.15	0.13	0.16	0.15
39	-0.06	-0.02	-0.05	-0.09	-0.04	0.01	-0.11	0.03	-0.04
40	0.06	-0.03	-0.06	0.08	0.08	-0.01	0.11	0.00	0.03

41	-0.07	-0.07	-0.09	-0.07	-0.08	-0.08	-0.15	-0.08	-0.09
42	0.11	0.15	0.22	-0.01	0.13	0.18	0.08	0.17	0.13
43	-0.10	-0.02	0.01	-0.02	-0.05	0.00	-0.16	0.00	-0.04
44	-0.37	-0.29	-0.24	-0.21	-0.33	-0.26	-0.41	-0.24	-0.29
45	-0.17	-0.22	-0.23	0.00	-0.27	-0.19	-0.10	-0.19	-0.17
46	-0.37	-0.36	-0.36	-0.38	-0.44	-0.35	-0.39	-0.34	-0.37
47	-0.28	0.12	-0.23	-0.10	0.00	0.01	-0.30	0.02	-0.10
48	-0.18	0.03	0.03	-0.31	-0.12	0.04	-0.26	0.03	-0.09
49	0.30	0.22	0.19	0.41	0.20	0.22	0.25	0.25	0.26
50	0.30	0.17	0.21	0.07	0.06	0.20	0.23	0.19	0.18
51	0.10	0.24	0.24	0.19	0.16	0.26	0.11	0.26	0.20
52	0.10	0.02	0.12	0.00	0.08	0.05	0.06	0.04	0.06
53	0.20	0.17	0.14	0.15	0.15	0.17	0.19	0.17	0.17
54	-0.05	-0.12	-0.10	-0.20	-0.08	-0.14	-0.09	-0.15	-0.12
55	0.12	0.00	-0.03	-0.01	0.12	0.00	-0.07	0.00	0.02
56	0.08	-0.03	-0.02	-0.09	0.19	0.00	0.01	0.01	0.02
57	-0.01	0.12	0.10	-0.02	0.03	0.11	-0.03	0.12	0.05
58	0.07	0.11	0.09	0.10	0.02	0.10	0.10	0.11	0.09
59	0.07	0.04	0.04	-0.03	-0.06	0.04	0.09	0.02	0.03
60	-0.02	-0.03	-0.06	0.00	-0.09	-0.03	-0.04	-0.03	-0.04
61	-0.08	0.10	0.06	0.00	-0.07	0.09	-0.17	0.07	0.00
62	-0.07	0.03	0.03	-0.04	-0.04	0.04	-0.05	0.04	-0.01
63	-0.06	0.02	0.05	-0.14	-0.02	0.03	-0.13	0.03	-0.03
64	0.08	0.23	0.25	0.16	0.16	0.26	0.01	0.26	0.18
65	0.05	0.21	0.19	0.17	0.04	0.22	0.15	0.22	0.16
66	0.30	0.50	0.45	0.44	0.46	0.53	0.37	0.51	0.44
67	-0.20	-0.10	-0.15	-0.12	-0.24	-0.13	-0.18	-0.13	-0.15
68	-0.02	0.05	0.01	-0.18	0.01	0.01	-0.06	-0.02	-0.02
69	-0.14	0.19	0.23	0.00	0.17	0.23	-0.04	0.22	0.11
70	-0.24	0.19	0.19	-0.06	0.20	0.22	-0.07	0.22	0.08
71	0.03	0.36	0.38	0.08	0.41	0.40	0.06	0.40	0.27
72	0.19	0.31	0.33	0.25	0.27	0.33	0.13	0.34	0.27
73	-0.14	-0.05	-0.10	-0.28	-0.09	-0.10	-0.15	-0.13	-0.13
74	-0.19	-0.12	-0.10	-0.24	-0.15	-0.06	-0.17	-0.07	-0.14
75	0.42	0.26	0.24	0.27	0.31	0.26	0.26	0.25	0.28
76	**1.35**	0.36	0.41	0.37	0.28	0.43	0.64	0.42	0.53
77	-0.31	-0.30	-0.30	-0.37	-0.30	-0.28	-0.33	-0.31	-0.31
78	-0.29	-0.27	-0.34	-0.52	-0.11	-0.36	-0.53	-0.36	-0.35
79	0.18	-0.12	-0.22	0.02	0.13	-0.12	-0.02	-0.10	-0.03
80	-0.47	-0.48	-0.47	-0.61	-0.35	-0.45	-0.60	-0.44	-0.48
81	-0.18	-0.81	-0.75	-0.91	-0.74	-0.73	-0.75	-0.72	-0.70
82	0.18	0.26	-0.41	-0.68	-0.12	0.01	-0.47	0.02	-0.15
83	0.05	-0.08	-0.09	-0.02	-0.19	-0.09	0.01	-0.12	-0.07
84	0.14	0.03	0.01	0.13	-0.01	0.01	0.10	-0.02	0.05
85	-0.64	-0.48	-0.52	-0.59	-0.48	-0.56	-0.27	-0.54	-0.51

86	-0.37	-0.10	-0.11	-0.18	-0.12	-0.08	-0.42	-0.08	-0.18
87	0.20	0.22	0.14	0.36	0.26	0.20	0.20	0.21	0.22
88	-0.15	-0.12	-0.15	-0.12	-0.17	-0.13	-0.11	-0.12	-0.13
89	0.07	0.32	0.27	0.33	0.55	0.33	0.30	0.33	0.31
90	-0.41	-0.09	-0.08	-0.15	-0.05	-0.07	-0.11	-0.08	-0.13
91	-0.38	-0.22	-0.24	-0.27	-0.22	-0.20	-0.38	-0.19	-0.26
92	0.08	0.23	0.16	0.31	0.34	0.26	0.09	0.25	0.21
93	-0.30	-0.15	-0.17	-0.18	-0.18	-0.15	-0.19	-0.16	-0.18
94	-0.44	-0.36	-0.41	-0.25	-0.33	-0.36	-0.38	-0.34	-0.36
95	-0.29	0.29	0.30	-0.50	0.44	0.30	0.15	0.32	0.13
96	-0.04	0.04	0.02	0.01	0.08	0.02	0.04	0.05	0.03
97	0.02	0.04	-0.01	0.04	-0.01	0.02	0.06	0.03	0.02
98	-0.03	-0.11	-0.13	-0.09	-0.28	-0.11	0.15	-0.11	-0.09
99	0.11	0.27	0.24	0.27	0.31	0.24	0.18	0.24	0.23
100	-0.15	0.07	0.05	0.16	0.07	0.03	0.10	0.04	0.05
101	-0.11	0.03	0.02	0.06	0.04	0.04	-0.06	0.04	0.01
102	0.37	0.07	0.17	0.28	0.18	0.08	0.27	0.08	0.19
103	0.23	-0.06	-0.09	0.18	0.00	-0.07	0.36	-0.07	0.06
104	0.36	0.31	0.28	0.75	0.20	0.23	0.59	0.24	0.37
105	0.17	-0.21	-0.26	-0.12	-0.31	-0.22	-0.16	-0.22	-0.17
106	0.02	-0.01	-0.02	-0.04	0.12	0.00	-0.06	-0.01	0.00
107	0.35	0.16	0.23	0.24	0.15	0.16	0.31	0.16	0.22
108	0.06	-0.05	-0.07	-0.01	0.15	-0.05	-0.16	-0.05	-0.02
109	0.17	0.10	0.03	0.47	-0.11	0.03	0.03	0.00	0.09

Table S5. List of compounds out of the AD of the logPaCa2 model according to the Standardized Descriptors Based Approach (K. Roy et al.. Chemom. Intell. Lab. Sys.2015) and to the Leverage Based Approach (values highlighted in red).

		Standardised Descriptors-Based Approach						Leverage-Based Approach					
No.	**Name**	**Full Model**	**M0**	**M1**	**M2**	**M3**	**M4**	**Full HAT i/i (h*=0.25)**	**M0 HAT i/i (h*=0.31)**	**M1 HAT i/i (h*=0.31)**	**M2 HAT i/i (h*=0.31)**	**M3 HAT i/i (h*=0.31)**	**M4 HAT i/i (h*=0.31)**
3	**pentafluopropionic anhydride**	Outlier	Outlier	Outlier	Outlier	Outside AD	Outlier	0.3148	0.3452	0.3855	0.3827	0.5114	0.4025
11	**4-nitrophthalic anhydride**	-	-	-	-	-	Outlier	0.226	0.2398	0.3303	0.2591	0.2499	0.411
18	**5-chloroisatoic anhydride**	Outlier	Outlier	Outlier	Outlier	Outside AD	Outlier	0.384	0.4294	0.7258	0.3939	0.6371	0.3966
24	**bicyclo [2,2,2] oct-7-ene-2,3,5,6-tetracarboxylic anhydride**	Outlier	Outlier	Outlier	Outlier	Outlier	Outside AD	0.4831	0.5151	0.6657	0.524	0.5447	0.9707
29	**lauric anhydride**	Outlier	Outlier	-	Outlier	Outlier	Outside AD	0.1924	0.2163	0.2006	0.2995	0.2264	0.3108
36	**isatoic anhydride**	Outlier	Outlier	Outside AD	Outlier	Outlier	Outlier	0.4087	0.4501	0.7895	0.4205	0.6683	0.4246
47	**palmitic anhydride**	Outlier	Outlier	Outlier	Outside AD	Outlier	Outlier	0.3099	0.3456	0.3233	0.4901	0.3658	0.4986
82	**pentaethylenehexamine**	Outlier	Outlier	Outlier	Outside AD	Outlier	Outlier	0.2935	0.404	0.3947	0.4638	0.3075	0.3259
95	**L-Arg**	Outlier	Outside AD	Outlier	Outlier	Outlier	Outlier	0.2351	0.3458	0.3438	0.2841	0.2976	0.2692
109	**diethylenetriaminepentaacetic dianhydride**	-	-	-	-	-	-	0.3031	0.3749	0.3307	0.3577	0.3819	0.5338

Table S6 Correlations among descriptors, response and additional descriptors used for the interpretation of logPaCa2 and logHUVEC models. Correlations > +/- 0.5 are highlighted in green.

	D094	***D106***	***D346***	***ATSC4i***	***MATS3s***	***MATS8s***	***VR3_Dt***	***nF10HeteroRing***	***D070***	***ATSC1v***	***MATS7s***	***VE2_Dze***	***VE3_Dzs***	***nBase***	***maxHssNH***	***BIC2***	***XLogP***	***nHBDon***	***logPaCa2***	***logHUVEC***
D094	1																			
D106	0.27	1																		
D346	-0.15	0.18	1																	
ATSC4i	-0.05	-0.11	-0.14	1																
MATS3s	-0.17	-0.29	0.09	-0.15	1															
MATS8s	-0.13	-0.01	0.15	0.15	0.08	1														
VR3_Dt	0.11	0.38	0.27	-0.23	0.04	-0.05	1													
nF10HeteroRing	-0.13	0.16	0.30	-0.09	0.10	0.15	0.14	1												
D070	-0.31	-0.36	-0.03	0.42	0.01	0.01	-0.50	0.06	1											
ATSC1v	-0.17	-0.19	0.10	0.04	0.25	0.13	-0.36	0.10	0.25	1										
MATS7s	0.03	0.08	-0.16	-0.09	0.03	-0.16	0.12	-0.09	-0.05	0.02	1									
VE2_Dze	0.18	-0.14	-0.13	0.15	0.15	-0.09	-0.36	-0.17	0.12	0.32	-0.07	1								
VE3_Dzs	-0.18	-0.15	-0.08	0.04	-0.01	0.03	-0.60	0.05	0.49	0.44	-0.02	0.39	1							
nBase	-0.22	-0.70	-0.22	-0.04	0.25	-0.01	-0.20	-0.19	0.28	-0.08	0.02	-0.05	-0.11	1						
maxHssNH	-0.12	-0.20	-0.07	0.02	0.09	0.10	-0.09	0.11	0.15	-0.02	-0.24	-0.17	-0.08	0.40	1					
BIC2	-0.08	-0.12	-0.08	0.08	0.06	0.12	-0.53	-0.03	0.12	0.47	-0.16	0.49	0.53	-0.08	0.08	1				
XLogP	0.08	0.29	0.23	0.00	0.09	-0.01	0.54	0.11	-0.18	-0.26	-0.06	-0.31	-0.51	-0.37	-0.12	-0.61	1			
nHBDon	-0.20	-0.74	-0.16	0.10	0.31	0.10	-0.35	-0.17	0.21	0.30	-0.05	0.18	0.02	0.81	0.37	0.30	-0.49	1		
logPaCa2	0.11	0.49	0.10	-0.15	-0.26	0.07	0.24	0.13	-0.26	-0.40	-0.15	-0.33	0.04	-0.56	-0.07	-0.27	0.35	-0.68	1	
logHUVEC	-0.02	0.67	0.38	-0.26	-0.26	-0.15	0.50	-0.09	-0.35	-0.22	0.08	-0.23	-0.25	-0.44	-0.19	-0.26	0.32	-0.54	0.36	1

Table S7. Residuals calculated for logHUVEC QSAR models developed by eight different methodologies, and on the basis of combined (i.e. averaged) predictions.

ID109	Res GRegNN	Res RBFNN	Res SVM-LIN	Res SVM-RAD	Res PPR	Res PLS	Res KNN	Res MLR	Res Combi.
1	0.24	0.01	0.07	0.41	-0.06	0.16	-0.01	0.06	0.11
2	-0.25	-0.19	-0.21	-0.31	-0.26	-0.18	-0.41	-0.19	-0.25
3	0.19	0.13	0.24	0.46	0.17	0.22	0.04	0.22	0.21
4	-0.19	0.29	0.31	0.34	0.26	0.31	0.12	0.29	0.21
5	-0.21	-0.16	-0.15	-0.33	-0.29	-0.23	-0.19	-0.20	-0.22
6	-0.26	-0.21	-0.20	-0.33	-0.46	-0.20	-0.23	-0.23	-0.26
7	0.35	0.26	0.27	0.16	0.22	0.28	0.16	0.27	0.25
8	0.02	-0.21	-0.15	-0.05	-0.17	-0.15	-0.06	-0.15	-0.11
9	-0.66	-0.46	-0.34	-0.49	-0.43	-0.44	-0.46	-0.43	-0.46
10	0.27	0.31	0.22	0.48	0.14	0.27	0.37	0.20	0.28
11	0.17	0.37	0.37	0.06	0.50	0.48	0.17	0.36	0.31
12	-0.27	-0.43	-0.50	0.07	-0.46	-0.40	-0.48	-0.41	-0.36
13	-0.51	0.77	0.37	-0.77	0.13	0.65	-0.36	0.65	0.12
14	0.11	-0.35	-0.34	0.26	-0.17	-0.34	0.04	-0.35	-0.14
15	-0.57	-0.30	-0.43	-0.65	-0.11	-0.38	-0.25	-0.11	-0.35
16	-1.15	-0.97	-0.95	-1.38	-0.88	-0.89	-1.15	-0.97	-1.04
17	0.32	0.53	0.55	0.28	0.50	0.53	0.36	0.51	0.45
18	-0.27	-0.45	-0.46	-0.79	-0.36	-0.38	-0.25	-0.38	-0.42
19	0.17	0.41	0.36	0.24	0.38	0.44	0.11	0.40	0.31
20	0.29	-0.11	-0.06	0.40	-0.13	-0.18	0.20	-0.15	0.03
21	-0.62	-0.44	-0.38	-0.80	-0.31	-0.43	-0.66	-0.45	-0.51
22	0.04	-0.38	-0.42	0.10	-0.20	-0.36	-0.06	-0.37	-0.21
23	-0.17	-0.13	-0.16	-0.15	-0.16	-0.14	-0.22	-0.14	-0.16
24	0.04	-0.16	-0.22	0.18	-0.16	-0.15	0.04	-0.17	-0.07
25	-0.54	-0.37	-0.42	-0.54	-0.49	-0.34	-0.51	-0.38	-0.45
26	0.27	0.12	0.18	0.33	0.20	0.22	0.26	0.15	0.22
27	-0.81	-0.61	-0.65	-0.61	-0.64	-0.57	-0.61	-0.60	-0.64
28	-0.39	-0.08	-0.05	-0.19	-0.02	-0.02	-0.36	-0.02	-0.14
29	0.16	0.10	0.15	0.37	0.09	0.06	0.17	0.11	0.15
30	0.10	0.30	0.00	0.35	0.03	0.29	0.59	-0.15	0.19
31	-0.22	-0.03	-0.04	-0.43	-0.16	0.10	-0.24	-0.03	-0.13
32	0.23	0.33	0.34	0.11	0.29	0.33	0.27	0.31	0.28
33	-0.07	0.35	0.28	0.20	0.36	0.35	0.23	0.35	0.26
34	-0.13	0.10	0.25	-0.32	0.10	0.07	-0.22	0.09	-0.01
35	0.73	0.17	0.22	0.20	0.20	0.23	0.39	0.06	0.28
36	0.91	0.34	0.30	0.73	0.40	0.34	0.40	0.36	0.47
37	-1.13	-1.07	-1.04	-1.42	-1.07	-1.04	-1.03	-1.05	-1.11
38	-0.56	-0.43	-0.35	-0.67	-0.28	-0.37	-0.46	-0.37	-0.43
39	0.25	0.19	0.15	0.10	0.14	0.19	0.04	0.19	0.16
40	0.41	0.07	0.11	0.68	0.05	0.07	0.51	-0.03	0.23
41	-0.09	-0.08	-0.08	-0.12	0.05	0.06	0.03	-0.08	-0.04

42	-0.50	-0.49	-0.47	-0.43	-0.52	-0.48	-0.29	-0.48	-0.46
43	-0.09	-0.02	-0.06	-0.15	-0.06	-0.02	-0.11	-0.02	-0.07
44	-0.01	-0.23	-0.07	0.13	-0.30	-0.22	0.10	-0.25	-0.10
45	0.52	0.38	0.46	0.43	0.31	0.20	0.52	0.36	0.40
46	0.18	0.30	0.30	0.07	0.12	0.38	0.25	0.29	0.24
47	0.12	-0.16	-0.21	0.50	-0.01	-0.23	0.20	-0.21	0.00
48	1.15	0.89	0.87	0.66	0.75	0.93	0.35	0.93	0.82
49	0.58	0.46	0.54	0.69	0.48	0.44	0.55	0.49	0.53
50	0.33	0.15	0.09	0.40	0.49	0.09	0.25	0.18	0.25
51	0.18	0.22	0.21	0.51	0.34	0.25	0.04	0.21	0.25
52	-0.09	0.02	-0.01	0.15	0.04	0.05	0.20	0.04	0.05
53	-0.39	-0.24	-0.33	-0.29	-0.28	-0.28	-0.39	-0.28	-0.31
54	0.39	0.17	0.20	0.35	0.19	0.14	0.23	0.17	0.23
55	0.40	0.49	0.44	0.34	0.42	0.54	0.38	0.48	0.44
56	-0.46	-0.25	-0.33	-0.18	-0.25	-0.20	-0.64	-0.29	-0.32
57	0.07	0.13	0.14	0.20	0.09	0.12	0.04	0.12	0.11
58	0.82	0.68	0.74	0.42	0.73	0.71	0.77	0.71	0.70
59	0.02	-0.03	-0.10	-0.13	0.02	-0.04	0.00	-0.04	-0.04
60	0.10	-0.01	0.04	0.14	0.08	-0.01	0.16	0.05	0.07
61	-0.77	-0.14	-0.03	-0.59	-0.14	-0.45	-0.92	-0.18	-0.40
62	-0.15	0.03	0.01	-0.10	-0.17	0.01	-0.11	0.00	-0.06
63	-0.26	0.00	-0.13	-0.50	-0.12	-0.07	-0.24	-0.07	-0.17
64	-0.12	-0.19	-0.16	-0.29	-0.01	-0.18	-0.14	-0.18	-0.16
65	0.08	0.08	0.07	0.21	0.10	0.09	0.16	0.07	0.11
66	0.47	0.59	0.64	0.10	0.65	0.46	0.59	0.58	0.51
67	-0.15	0.06	0.02	-0.18	-0.03	0.04	-0.17	0.03	-0.05
68	0.26	0.23	0.19	0.35	0.25	0.22	0.28	0.22	0.25
69	-0.33	-0.29	-0.29	-0.31	-0.33	-0.30	-0.29	-0.29	-0.30
70	-0.16	-0.15	-0.14	-0.18	-0.13	-0.14	-0.22	-0.15	-0.16
71	-0.21	-0.25	-0.31	-0.43	-0.19	-0.30	-0.24	-0.27	-0.27
72	0.06	0.13	0.13	0.01	0.08	0.13	0.03	0.12	0.09
73	-0.04	-0.28	-0.27	-0.45	-0.24	-0.26	-0.02	-0.26	-0.23
74	-0.34	-0.27	-0.25	-0.47	-0.34	-0.27	-0.35	-0.27	-0.32
75	0.30	0.00	0.07	0.23	0.05	0.01	0.33	0.03	0.13
76	0.17	0.19	0.17	0.16	0.22	0.18	0.14	0.20	0.18
77	0.12	0.10	0.12	0.23	0.05	0.09	0.14	0.10	0.12
78	-0.35	-0.09	-0.05	-0.84	-0.12	-0.01	-0.35	-0.01	-0.23
79	0.27	0.25	0.40	-0.48	0.23	0.19	-0.16	0.24	0.12
80	0.13	0.34	0.31	0.09	0.13	0.31	0.21	0.29	0.23
81	0.10	0.08	0.02	0.09	0.13	-0.01	0.11	0.06	0.07
82	-0.04	-0.18	-0.17	-0.32	0.01	-0.19	0.01	-0.19	-0.13
83	0.00	-0.05	-0.05	-0.05	-0.03	-0.04	0.09	-0.04	-0.02
84	0.03	-0.04	-0.06	0.01	0.05	-0.05	0.01	-0.04	-0.01
85	0.17	0.34	0.39	0.16	0.53	0.34	0.19	0.38	0.31
86	-0.03	-0.13	-0.18	-0.06	-0.05	-0.11	-0.07	-0.13	-0.10

87	-0.17	-0.15	-0.14	-0.05	-0.19	-0.15	-0.17	-0.15	-0.14
88	-0.40	-0.57	-0.54	-0.49	-0.48	-0.53	-0.37	-0.53	-0.49
89	-0.15	-0.34	-0.34	-0.24	-0.20	-0.36	-0.15	-0.34	-0.27
90	-0.31	-0.58	-0.49	-0.75	-0.49	-0.60	-0.34	-0.38	-0.49
91	-0.23	-0.34	-0.38	-0.25	-0.27	-0.36	-0.22	-0.34	-0.30
92	0.35	0.23	0.22	0.35	0.35	0.20	0.37	0.20	0.28
93	0.53	0.61	0.52	0.41	0.58	0.58	0.55	0.58	0.55
94	0.04	0.13	0.12	0.18	0.07	0.12	0.07	0.13	0.11
95	0.13	-0.11	-0.03	-0.03	-0.01	-0.11	0.12	0.04	0.00
96	0.46	0.53	0.56	0.11	0.57	0.46	0.51	0.54	0.47
97	0.41	0.26	0.29	0.37	0.45	0.29	0.48	0.28	0.35
98	-0.13	-0.24	-0.21	-0.37	-0.20	-0.22	-0.17	-0.22	-0.22
99	-0.26	-0.30	-0.32	-0.32	-0.25	-0.31	-0.33	-0.30	-0.30
100	-0.10	0.02	-0.03	-0.05	-0.31	0.01	-0.09	0.04	-0.06
101	-0.16	-0.23	-0.30	-0.34	-0.17	-0.25	-0.32	-0.24	-0.25
102	-0.07	-0.02	0.08	-0.02	-0.01	-0.01	0.03	0.00	0.00
103	-0.45	-0.30	-0.13	-0.29	-0.32	-0.28	-0.21	-0.28	-0.28
104	0.25	0.38	0.42	0.49	0.33	0.41	0.44	0.38	0.39
105	0.81	0.55	0.58	0.79	0.81	0.44	0.76	0.59	0.67
106	0.06	0.10	0.11	0.20	-0.10	0.17	0.03	0.10	0.08
107	0.51	0.72	0.80	0.60	0.69	0.72	0.61	0.71	0.67
108	0.10	-0.08	-0.03	0.10	-0.14	-0.07	0.03	-0.07	-0.02
109	0.49	-0.44	-0.56	0.35	-0.31	-0.53	0.00	-0.53	-0.19

Table S8. List of compounds out of the AD of the logHUVEC model according to the Standardized Descriptors Based Approach (K. Roy et al.. Chemom. Intell. Lab. Sys.2015) and to the Leverage Based Approach (values highlighted in red).

		Standardised Descriptors-Based Approach						Leverage-Based Approach					
No.	Name	Full Model	M0	M1	M2	M3	M4	Full HAT i/i (h*=0.25)	M0 HAT i/i (h*=0.31)	M1 HAT i/i (h*=0.31)	M2 HAT i/i (h*=0.31)	M3 HAT i/i (h*=0.31)	M4 HAT i/i (h*=0.31)
12	4-bromo-1,8-naphthalic anhydride	-	-	-	-	Outlier	-	0.1426	0.1492	0.1497	0.2225	0.3151	0.1536
13	1,4,5,8-naphtalenetrtracarboxylic anhydride	Outlier	Outlier	Outlier	Outlier	Outside AD	Outlier	0.4478	0.4991	0.4758	0.6331	1.024	0.4618
15	3-nitro-1,8-naphthalic anhydride	Outlier	Outside AD	Outlier	Outlier	Outlier	Outlier	0.2922	0.7967	0.3098	0.3382	0.3966	0.327
30	1,2,4-benzenetricarboxylic anhydride	Outlier	Outside AD	-	-	-	Outlier	0.2909	0.8957	0.3173	0.3157	0.3486	0.3215
35	ethylenediamintetra acetic anhydride	-	Outside AD	-	-	-	-	0.1798	0.3571	0.1996	0.1949	0.2025	0.3138
47	Palmitic anhydride	-	-	-	-	-	-	0.2162	0.2695	0.239	0.3046	0.2596	0.3111
48	4-amino-1,8-naphthalic anhydride	Outlier	Outlier	Outlier	Outlier	Outside AD	Outlier	0.4437	0.4767	0.486	0.5832	0.9381	0.5146
79	Tyramine	-	Outlier	-	-	-	-	0.2138	0.3574	0.2321	0.2327	0.2885	0.324
90	L-Tyr	-	Outside AD	-	-	-	-	0.1143	0.33	0.122	0.1262	0.1399	0.133
109	diethylenetriaminepentaacetic dianhydride	Outlier	Outlier	Outlier	Outlier	Outlier	Outside AD	0.4233	0.5157	0.4901	0.4611	0.4686	0.9601

Table S9. Rotated score values for the first two principal factors (F1-F2) generated by Factorial Analysis starting from experimental values of uptake in different cell typologies for 109 nanoparticles, and scores predicted by QSAR for additional 28 structures (data reported in Italic). (split columns: 1=Training set; 2=Prediction set).

ID 137	Chemical Class	Factor1	Factor2	Split Factor1	Split Factor2
1	1	0.93	1.29	1	1
2	1	0.34	0.51	1	1
3	1	1.22	1.25	1	2
4	1	0.11	1.28	1	1
5	1	0.25	0.67	2	2
6	1	0.43	-0.64	2	1
7	1	1.67	-0.63	1	1
8	1	0.33	0.10	1	1
9	1	-0.93	0.30	1	2
10	1	1.73	-0.60	1	1
11	1	1.60	-1.45	1	1
12	1	0.54	-0.22	1	1
13	1	-1.02	0.10	1	2
14	1	1.05	0.23	1	1
15	1	-1.09	1.42	1	1
16	1	-0.99	1.20	1	1
17	1	1.33	0.55	1	1
18	1	-1.72	1.70	1	1
19	1	1.11	0.49	2	1
20	1	1.05	-0.06	1	1
21	1	-0.05	-0.37	1	2
22	1	0.96	0.75	2	2
23	1	0.54	0.32	1	1
24	1	1.02	-0.36	1	1
25	1	-0.10	0.66	1	1
26	1	1.60	-0.09	1	1
27	1	-0.90	0.56	2	1
28	1	0.00	0.07	1	1
29	1	1.17	-0.07	1	2
30	1	0.66	-1.09	1	1
31	1	0.76	0.41	1	1
32	1	1.29	-0.67	1	2
33	1	0.63	0.14	1	1
34	1	-0.78	1.27	2	1
35	1	0.34	0.92	1	1
36	1	-0.60	2.26	1	1
37	1	-2.39	0.42	1	1
38	1	-1.25	1.09	1	1
39	1	1.31	-0.28	2	1
40	1	1.29	0.26	1	1
41	1	1.36	-0.04	1	1
42	1	-0.92	0.96	1	2
43	1	0.89	-0.32	2	1
44	1	1.05	-0.60	1	1
45	1	1.45	-0.33	2	1
46	1	1.59	-0.70	1	1

47	1	1.72	-0.57	1	1
48	1	1.43	-0.70	1	1
49	1	1.46	0.47	1	1
50	1	-0.31	1.00	1	1
51	1	0.91	0.56	1	1
52	1	0.41	0.50	1	2
53	1	0.06	0.87	2	1
54	1	1.72	0.14	1	1
55	1	1.61	0.74	2	1
56	1	-0.95	1.09	2	1
57	2	-0.86	0.56	1	2
58	2	0.53	0.26	1	1
59	2	-0.68	-0.48	1	1
60	2	-0.79	0.49	1	1
61	2	-0.79	1.05	1	2
62	2	-1.17	0.66	2	1
63	2	-1.40	0.69	1	1
64	2	-1.02	0.77	1	1
65	2	-0.95	0.11	1	1
66	2	-0.02	0.82	2	1
67	2	-1.03	-0.24	1	1
68	2	-0.75	0.45	1	2
69	2	-1.38	0.05	1	1
70	2	-1.13	-0.43	1	1
71	2	-1.21	0.19	1	2
72	2	-0.96	0.27	1	1
73	2	-0.95	0.99	1	1
74	2	-1.36	0.32	2	1
75	2	-0.70	0.69	2	1
76	2	-0.71	0.46	1	1
77	2	-0.43	-2.55	2	1
78	2	-0.97	-2.94	1	1
79	2	-0.30	-1.33	2	2
80	2	-0.08	-1.89	1	1
81	2	-0.01	-3.01	1	1
82	2	-0.64	-2.74	1	1
83	2	-0.76	-0.92	1	1
84	2	-0.64	-0.81	1	2
85	3	0.00	-1.63	1	2
86	3	-0.76	-1.40	1	1
87	3	-1.08	-0.94	1	1
88	3	-1.33	-0.95	1	1
89	3	-0.85	-1.52	1	1
90	3	-1.11	-1.09	1	1
91	3	-0.97	-0.44	1	1
92	3	-0.06	-0.17	1	1
93	3	0.22	-1.59	1	1
94	3	-0.42	-1.14	1	1
95	3	-0.43	-0.75	1	1
96	3	0.22	-0.59	1	2
97	3	0.02	-0.71	1	1
98	3	-1.06	-0.24	2	2
99	3	-1.33	-0.99	1	1
100	3	-1.03	-1.01	1	2
101	3	-0.99	-1.06	2	1

102	1	-0.32	1.49	1	1
103	1	0.18	1.03	1	1
104	1	0.93	1.06	1	1
105	1	0.64	0.97	2	1
106	1	0.95	0.65	1	1
107	1	1.32	0.73	1	1
108	1	0.82	0.54	1	1
109	1	0.67	1.48	1	2
110	*1*	*0.61*	*-0.48*	*2*	*2*
111	*1*	*0.68*	*1.44*	*2*	*2*
112	*1*	*0.21*	*4.51*	*2*	*2*
113	*1*	*-0.01*	*-1.43*	*2*	*2*
114	*1*	*0.07*	*-0.05*	*2*	*2*
115	*1*	*0.61*	*0.20*	*2*	*2*
116	*1*	*0.93*	*0.77*	*2*	*2*
117	*1*	*-0.41*	*0.51*	*2*	*2*
118	*1*	*0.82*	*-0.59*	*2*	*2*
119	*1*	*0.80*	*0.75*	*2*	*2*
120	*1*	*0.72*	*1.30*	*2*	*2*
121	*1*	*0.46*	*2.52*	*2*	*2*
122	*1*	*-1.01*	*1.52*	*2*	*2*
123	*1*	*1.21*	*-1.22*	*2*	*2*
124	*1*	*-0.67*	*-0.85*	*2*	*2*
125	*2*	*-0.79*	*0.43*	*2*	*2*
126	*2*	*-0.98*	*0.12*	*2*	*2*
127	*2*	*-0.39*	*-0.55*	*2*	*2*
128	*2*	*-0.27*	*-1.20*	*2*	*2*
129	*2*	*-0.50*	*-0.40*	*2*	*2*
130	*3*	*-0.50*	*-0.38*	*2*	*2*
131	*3*	*0.01*	*-0.50*	*2*	*2*
132	*3*	*-1.19*	*-3.51*	*2*	*2*
133	*3*	*0.45*	*-0.44*	*2*	*2*
134	*3*	*-0.68*	*-0.57*	*2*	*2*
135	*3*	*-0.48*	*-0.77*	*2*	*2*
136	*2*	*-2.68*	*-2.32*	*2*	*2*
137	*3*	*-1.66*	*-0.28*	*2*	*2*

Table S10. Setting parameters for models different from MLR.

Method	Optimization	Parameters PaCa2	Parameters HUVEC
GRegNN	user-driven *	radius of Gaussian r=0.125	radius of Gaussian r=0.155
RBFNN	user-driven *	radius of Gaussian s=4.4	radius of Gaussian s=2.7
kNN	software-based (grid search)**	k=4	k=8
SVMRadial	software-based (grid search)**	C=2; g=0.19; e= 0.1 (default value)	C=1; g=0.29; e= 0.1 (default value)
SVM-Linear	software-based (grid search)**	C =2; e= 0.1 (default value)	C =0.25; e= 0.1 (default value)
PLS	software-based (grid search)**	n° of components = 5	n° of components = 2
PPR	software-based (grid search)**	n° of projections =1	n° of projections =1

* r or s values which minimize RMSEcv (leave-one-out). Calculations performed in MATLAB (M.J.L. Orr, MATLAB Routines for Subset Selection and Ridge Regression in Linear Neural Networks, Centre for Cognitive Science, Edinburgh University, Edinburgh, U.K., 1996.).

** values which minimize RMSEcv using five-fold cross validation repeated 3 times. Calculations performed in R, Caret Package (M. Kuhn, Building predictive models in R using the caret package. J. Stat. Soft. 28 (2008), 1-26.;
R.D.C. Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2014. Available at: http://www.R-project.org)