Supplementary Information

Title: Graph-Guided Multi-Task Sparse Learning Model: a Method for Identifying Antigenic Variants of Influenza A(H3N2) Virus

Running title: Multi-task Learning Model For Influenza Vaccine Strain Selection

Authors: Han, L.[1,2], Li, L.[1], Wen, F.[1], Zhong, L.[1], Zhang, T.[2] and Wan, X.-F. [1*]

Affiliations: [1]Department of Basic Science, College of Veterinary Medicine, Mississippi State University, Mississippi State, Mississippi, USA; [2]Tencent AI Laboratory, Shenzhen, China

*To whom correspondence should be addressed: Dr. Xiu-Feng Wan, wan@cvm.msstate.edu

Supplementary Tables

Table S1. Numerical evaluation of the prediction performance of the multi-task learning (MTL) methods (including the GG-MTSL, $l_{1,2}$ norm regularized MTL and $l_{1,\infty}$ norm regularized MTL), and the single task learning methods (including Lasso and Ridge regressions). The running time, rooted mean square error (RMSE), accuracy, sensitivity, and specificity are evaluated.

Method	RMSE	Accuracy, %	Sensitivity, %	Specificity, %
GG-MTSL (training time: 2 minutes)				
2009	0.5495	78.32	44.90	95.99
2010	0.5959	68.73	52.11	94.80
2011	0.9994	80.90	84.78	66.17
2012	1.4209	97.07	97.04	100.0
2013	0.8613	98.20	97.85	78.26
2014	0.9092	93.15	94.26	88.28
2015	1.1416	81.58	78.00	93.36
2016	0.8457	86.44	89.59	79.92
Average	0.9154	85.55	79.82	87.10
$l_{1,\infty}$ norm regularized MTL (training time: 32 minutes)				
2009	1.2662	72.90	57.27	88.24
2010	1.2679	72.85	63.13	82.05
2011	1.1008	95.50	95.86	95.32
2012	1.189	91.50	78.64	97.25
2013	1.1265	85.15	67.87	94.83
2014	1.1024	85.32	70.35	93.70
2015	1.1028	85.27	70.06	93.70
2016	1.1019	85.20	69.73	93.70
Average	1.1572	84.21	71.61	92.35
$l_{1,2}$ norm regularized MTL (training time: 29 minutes)				
2009	1.2612	72.94	56.38	89.18
2010	1.2636	73.06	64.06	81.58
2011	1.1865	92.42	95.98	90.68
2012	1.1676	91.01	78.64	96.54
2013	1.1195	85.21	68.23	94.72
2014	1.0984	85.39	70.19	93.90
2015	1.0983	85.34	69.90	93.90
2016	1.0978	85.30	69.75	93.84
Average	1.1616	83.83	71.64	91.79
Lasso (training time: 1.5 minutes)				
2009	1.4428	69.62	41.76	96.94
2010	1.4721	68.21	37.16	97.61
2011	2.0900	95.69	96.30	95.39
2012	2.0737	91.69	78.68	97.51
2013	1.6881	84.22	59.69	97.96
2014	1.4662	84.59	64.02	96.10
2015	1.4588	84.53	63.66	96.11
2016	1.4550	84.53	63.50	96.07
Average	1.6433	82.89	63.10	96.71
Ridge (training time: 0.5 minutes)				
2009	2.1098	53.42	5.94	100
2010	2.1185	51.36	0.00	100
2011	1.6329	96.08	94.72	96.75
2012	1.8990	92.08	77.96	98.39
2013	2.0767	84.48	59.14	98.66
2014	1.9285	84.51	59.23	98.66
2015	1.9299	84.45	58.83	98.66
2016	1.9220	84.41	58.44	98.66
Average	1.9521	78.85	51.78	98.72

Table S2. The selected residues and their scores computed by the product of the local coefficients, $\overline{w}^{local}$, and the PIMA score (i.e., the antigenic distance contribution in terms of units) for 50 different tasks performed on influenza A(H3N2) viruses collected from 1968 through July 2016.

Task 1		Task 2		Task 3		Task 4		Task 5	
1968–1979		1969–1980		1970–1981		1971–1982		1972–1983	
Residue	Score	Residue	Score	Residue	Score	Residue	Score	Residue	Score
145	0.1092	145	0.1075	145	0.1079	145	0.1061	137	0.1111
137	0.0709	137	0.0867	137	0.0955	137	0.1061	145	0.0995
278	0.0551	278	0.0642	278	0.0671	278	0.0706	278	0.0754
193	0.0478	193	0.05	193	0.0509	158	0.0567	158	0.0566
83	0.0254	158	0.0351	158	0.0422	193	0.0513	193	0.0442
155	0.0238	83	0.0271	62	0.028	62	0.0374	62	0.0424
78	0.0237	62	0.0226	83	0.0274	83	0.0256	172	0.0297
275	0.0218	189	0.0213	189	0.0229	172	0.0255	189	0.026
158	0.021	155	0.0198	155	0.0161	189	0.0241	160	0.0247
189	0.018	275	0.0164	172	0.0151	160	0.0217	83	0.0206
62	0.0164	78	0.0164	260	0.0145	260	0.0195	260	0.0193
122	0.0145	174	0.013	174	0.0141	54	0.0163	54	0.0185
174	0.0119	260	0.0121	160	0.0122	174	0.0145	133	0.0174
159	0.0107	122	0.012	275	0.012	133	0.0139	174	0.017
144	0.0081	172	0.0107	53	0.0115	53	0.0127	53	0.0152
53	0.0077	159	0.0104	159	0.0106	155	0.0111	156	0.0102
		53	0.0102	54	0.0098	159	0.0093	213	0.0101
		160	0.0084	122	0.0098			50	0.008
		54	0.0071	78	0.0097				
		217	0.0065						

Task 6		Task 7		Task 8		Task 9		Task 10	
1973–1984		1974–1985		1975–1986		1976–1987		1977–1988	
Residue	Score	Residue	Score	Residue	Score	Residue	Score	Residue	Score
137	0.1169	137	0.1097	137	0.0787	124	0.0685	124	0.1215
145	0.0869	145	0.0654	144	0.0445	144	0.0666	144	0.0852
278	0.0713	158	0.0568	62	0.0444	159	0.0496	159	0.082
158	0.0597	278	0.0533	158	0.0435	163	0.0469	163	0.0748
62	0.0481	62	0.0503	172	0.0335	137	0.0429	173	0.0424
172	0.0345	172	0.0367	159	0.0314	173	0.0367	248	0.0354
193	0.0338	160	0.0261	173	0.0289	248	0.0351	155	0.0247
160	0.0277	144	0.025	248	0.0286	62	0.0257	156	0.0231
189	0.0248	133	0.0236	124	0.026	156	0.0256	137	0.0209
54	0.0214	193	0.023	160	0.0232	158	0.0243	186	0.0191
133	0.0209	54	0.0227	133	0.0231	172	0.0229	133	0.0167
260	0.0203	173	0.0196	163	0.023	133	0.0197	213	0.0143
174	0.0197	248	0.0196	174	0.0217	174	0.0194	189	0.0113
53	0.0163	174	0.0195	54	0.0206	160	0.0167	158	0.0096
83	0.0128	260	0.0192	145	0.0204	54	0.0139	172	0.0094
156	0.0128	189	0.0191	156	0.0199	193	0.0125	193	0.0089
213	0.0125	156	0.017	193	0.0197	262	0.0115	174	0.0086
262	0.0097	159	0.0159	260	0.0145	213	0.011	62	0.0085
		53	0.0148	278	0.0132	260	0.0078	160	0.0084
		213	0.0144	213	0.0122	307	0.0077	307	0.0079
		124	0.0122	262	0.0119	155	0.0075	262	0.0071
						67	0.0058	219	0.0068
						138	0.0047	54	0.0056
						189	0.0045	138	0.0055
						217	0.0032	67	0.0054
						53	0.0031		
						186	0.0031		
						50	0.0029		

Task 11		Task 12		Task 13		Task 14		Task 15	
1978–1989		1979–1990		1980–1991		1981–1992		1982–1993	
Residue	Score	Residue	Score	Residue	Score	Residue	Score	Residue	Score
124	0.1588	124	0.1428	193	0.113	193	0.1978	193	0.1884
159	0.0846	159	0.075	124	0.0987	121	0.1166	135	0.1195
155	0.0628	155	0.0604	135	0.0857	135	0.0784	189	0.0791
163	0.06	82	0.0581	121	0.079	124	0.0451	121	0.0662
82	0.0547	83	0.0581	145	0.0581	145	0.0416	158	0.0461

83	0.0545	131	0.0581	159	0.0521	158	0.0382	226	0.0411
131	0.0545	163	0.0513	155	0.0484	155	0.0298	133	0.0409
144	0.0542	144	0.0445	82	0.0449	156	0.0273	262	0.0384
186	0.0399	186	0.0359	83	0.0449	159	0.0252	219	0.0381
156	0.0299	156	0.0358	131	0.0449	82	0.0247	156	0.0374
189	0.0274	135	0.0296	163	0.0324	83	0.0247	172	0.0249
299	0.0236	189	0.0264	156	0.0287	131	0.0247	155	0.0232
135	0.0226	299	0.0251	144	0.024	172	0.0245	214	0.0229
173	0.0183	193	0.0244	186	0.0225	45	0.0234	45	0.0197
248	0.013	173	0.0151	189	0.0219	189	0.0197	124	0.0196
193	0.0098	145	0.014	299	0.0194	262	0.0193	82	0.0144
145	0.0093	121	0.0128	45	0.0164	163	0.0159	83	0.0144
219	0.0063	248	0.0103	158	0.0144	214	0.0135	131	0.0144
213	0.0044	45	0.009	173	0.0113	78	0.0132	159	0.0143
138	0.0042	262	0.0076	262	0.0108	299	0.0114	145	0.0123
142	0.0042	219	0.0057	138	0.009	133	0.0112	190	0.0121
262	0.0037	142	0.0043	248	0.0072	226	0.0106	299	0.012
307	0.0028	138	0.0041	219	0.0047			138	0.0118
121	0.0025	226	0.003	172	0.0044			163	0.0112
214	0.0024	213	0.0029	142	0.0043			78	0.0105
88	0.0022	122	0.0028	226	0.0036			196	0.0094
45	0.0021	88	0.0027	196	0.0035			276	0.0084
188	0.002	214	0.0024	78	0.0027			248	0.0064
67	0.0018	307	0.0022	214	0.0026			186	0.0047
				88	0.0024				
				188	0.0022				
				122	0.0022				
				307	0.0014				

Task 16		Task 17		Task 18		Task 19		Task 20	
1983–1994		1984–1995		1985–1996		1986–1997		1987–1998	
Residue	Score	Residue	Score	Residue	Score	Residue	Score	Residue	Score
193	0.1705	135	0.1515	135	0.1724	135	0.167	135	0.1698
135	0.1317	193	0.1494	193	0.1367	193	0.1278	193	0.1202
189	0.0795	189	0.072	226	0.0715	262	0.0972	262	0.101
219	0.0687	226	0.0645	262	0.0662	226	0.0808	226	0.0844
121	0.0586	219	0.0576	124	0.0644	124	0.0737	124	0.0806
158	0.0463	121	0.055	189	0.0618	158	0.0721	158	0.0791
226	0.0453	124	0.051	121	0.0503	278	0.0599	196	0.0692
133	0.0426	158	0.0481	219	0.0502	196	0.0541	278	0.062
262	0.038	190	0.048	190	0.0501	190	0.0447	190	0.046
156	0.0337	262	0.0442	158	0.0486	219	0.0443	276	0.0449
172	0.0235	133	0.0363	278	0.0458	276	0.044	219	0.0423
214	0.0233	278	0.0305	133	0.0356	121	0.0372	121	0.0366
45	0.0226	172	0.0295	172	0.0232	133	0.0348	133	0.0356
124	0.0193	156	0.0256	45	0.0228	142	0.0234	142	0.0283
190	0.0176	45	0.0225	145	0.0215	189	0.0229	45	0.0243
155	0.0157	214	0.0184	276	0.0196	45	0.0217	62	0.0218
145	0.0121	145	0.0178	214	0.0177	145	0.019	145	0.0171
82	0.0108	276	0.0174	156	0.0158	214	0.0162	189	0.0169
131	0.0108	155	0.0124	196	0.0119	156	0.0155	214	0.0157
83	0.0108	138	0.0123	138	0.0118	62	0.0144	138	0.0142
138	0.0104	196	0.0097	155	0.0092	138	0.0129	275	0.0137
276	0.0098	131	0.0085	50	0.0078	275	0.0093	156	0.013
299	0.0096	83	0.0085	131	0.0071	172	0.0076	172	0.0111
196	0.0091	82	0.0085	83	0.0071	198	0.007	144	0.0096
78	0.0088	299	0.0078	82	0.0071	201	0.0065	198	0.0068
159	0.007	201	0.0066	198	0.0071	50	0.0056	201	0.0064
201	0.0056	78	0.0058	299	0.0068	144	0.0051	50	0.0055
163	0.0055	159	0.0055	201	0.0067	131	0.0046	75	0.0041
75	0.0045	122	0.0054	122	0.0066	83	0.0046	299	0.0036
278	0.0045	75	0.0049	75	0.0044	82	0.0046	131	0.0035
248	0.0043	163	0.0034	142	0.0044	299	0.0046	83	0.0035
		53	0.003	53	0.0027	122	0.0045		

Task 21		Task 22		Task 23		Task 24		Task 25	
1988–1999		1989–2000		1990–2001		1991–2002		1992–2003	
Residue	Score	Residue	Score	Residue	Score	Residue	Score	Residue	Score
135	0.1688	135	0.1777	135	0.2186	144	0.3527	144	0.5225
193	0.1118	262	0.1098	124	0.1309	156	0.2657	156	0.4492
262	0.1066	124	0.1043	262	0.1112	135	0.2123	124	0.2354
124	0.0909	193	0.0954	144	0.1109	124	0.1874	50	0.1514

226	0.0872	226	0.0898	196	0.0974	196	0.1027	135	0.1411
158	0.081	196	0.0808	226	0.0926	158	0.1019	158	0.1357
196	0.0806	158	0.0786	158	0.0816	262	0.0921	186	0.1285
278	0.0588	144	0.0635	156	0.0599	226	0.0902	196	0.1276
276	0.0507	276	0.0517	142	0.0462	276	0.0807	226	0.1232
190	0.0474	278	0.0501	276	0.0458	137	0.0676	276	0.1068
219	0.0405	133	0.0474	278	0.0455	62	0.0648	57	0.0978
121	0.0404	190	0.0433	121	0.0402	50	0.0602	137	0.092
133	0.038	121	0.0428	219	0.0382	142	0.0552	133	0.0915
142	0.0375	142	0.0391	137	0.0374	121	0.0524	145	0.0743
144	0.0316	219	0.0379	62	0.0355	57	0.0515	62	0.0736
62	0.0264	62	0.0265	190	0.0343	186	0.0506	121	0.0676
45	0.0223	137	0.0249	133	0.0325	189	0.0473	189	0.0444
137	0.0178	45	0.0218	57	0.0243	219	0.0394	278	0.0385
145	0.0176	57	0.0171	45	0.0226	172	0.0305	131	0.0344
189	0.0168	145	0.0169	192	0.0209	133	0.0289	262	0.0331
214	0.0147	192	0.0162	193	0.0193	275	0.0265	155	0.0305
138	0.0133	189	0.0147	275	0.0171	155	0.0247	142	0.0301
57	0.0109	214	0.0132	138	0.0143	190	0.023	219	0.0295
275	0.0104	138	0.0131	145	0.0107	75	0.0215	275	0.0288
172	0.0089	156	0.0129	214	0.0086	138	0.0172	83	0.0273
192	0.0084	275	0.0129	5	0.0077	45	0.0166	172	0.0242
198	0.0065	198	0.0072	198	0.0076	131	0.0145	227	0.0236
50	0.0063	50	0.0067	75	0.0075	199	0.0118	138	0.0226
201	0.0062	201	0.0061	199	0.0059	278	0.0117		
156	0.0057	75	0.0055	201	0.0059	145	0.0102		
75	0.0046	5	0.0043			83	0.0102		
122	0.0033	172	0.0041						

Task 26		Task 27		Task 28		Task 29		Task 30	
1993–2004		1994–2005		1995–2006		1996–2007		1997–2008	
Residue	Score	Residue	Score	Residue	Score	Residue	Score	Residue	Score
156	0.3983	159	0.3261	196	0.3374	196	0.3924	196	0.6679
144	0.3105	156	0.3025	159	0.3089	145	0.2905	193	0.3139
159	0.2857	196	0.2185	193	0.2827	227	0.2901	227	0.2801
196	0.2139	226	0.1874	156	0.2234	159	0.2649	145	0.2474
189	0.1785	145	0.1785	145	0.2081	193	0.2462	138	0.2306
186	0.1784	189	0.1629	227	0.1945	156	0.1628	159	0.2033
57	0.1712	193	0.1615	225	0.1837	138	0.1546	156	0.1142
226	0.1705	57	0.1585	226	0.1762	50	0.1158	226	0.0883
50	0.1629	225	0.1433	50	0.1509	226	0.1102	50	0.079
145	0.1562	137	0.1391	276	0.1312	276	0.0821	276	0.0611
137	0.1399	144	0.134	57	0.1184	25	0.0726	25	0.0585
158	0.1332	50	0.1307	189	0.1069	144	0.068	186	0.0517
135	0.1115	186	0.1062	138	0.1002	142	0.065	144	0.0429
124	0.1015	276	0.0848	144	0.0958	57	0.0631	225	0.0406
121	0.0979	121	0.0794	142	0.0817	186	0.0615	142	0.0388
225	0.0761	155	0.0633	25	0.0757	189	0.0595	83	0.0342
155	0.062	75	0.0617	186	0.072	225	0.0554	57	0.0301
83	0.0617	158	0.0611	137	0.0623	83	0.0391	199	0.0279
142	0.0531	227	0.0563	188	0.0537	188	0.035	189	0.0274
25	0.0528	62	0.0558	202	0.0484	202	0.0333	173	0.0239
227	0.0499	25	0.055	199	0.0361	199	0.0321	202	0.0211
62	0.0482	188	0.0472	62	0.0345	137	0.028	158	0.0142
75	0.0464	138	0.0435	155	0.0268	158	0.0204	45	0.0138
278	0.0457	202	0.0401	75	0.0267	62	0.0192	188	0.0124
138	0.0431	199	0.0385	83	0.0226	45	0.0162		
131	0.0423	131	0.0352	121	0.0198	173	0.013		
172	0.0416	192	0.0339	135	0.0185	133	0.0101		
199	0.0288	83	0.0304	131	0.0185	131	0.0098		
188	0.0254	172	0.0295	45	0.0169	219	0.0091		
192	0.0237			133	0.0153	53	0.009		
				173	0.0141				
				158	0.0137				
				192	0.0128				
				275	0.0117				

Task 31		Task 32		Task 33		Task 34		Task 35	
1998–2009		1999–2010		2000–2011		2001–2012		2002–2013	
Residue	Score	Residue	Score	Residue	Score	Residue	Score	Residue	Score
196	0.7354	196	0.7453	196	0.8015	196	0.8148	196	1.0755
193	0.3251	193	0.3192	193	0.3276	193	0.3268	193	0.3142

227	0.2742	227	0.2723	227	0.2679	227	0.2667	138	0.2437
138	0.2599	138	0.2516	145	0.2368	145	0.2349	227	0.2377
145	0.2324	145	0.2385	138	0.2118	138	0.2109	145	0.2015
159	0.175	159	0.1712	159	0.1636	159	0.1585	159	0.107
156	0.0836	156	0.0747	156	0.0695	226	0.0671	225	0.0567
226	0.0652	50	0.0594	226	0.0662	156	0.065	188	0.0482
50	0.0613	226	0.0592	225	0.0591	225	0.0572	156	0.0437
225	0.0603	225	0.056	50	0.0542	50	0.052	226	0.0412
186	0.0451	186	0.0406	142	0.0353	142	0.0337	50	0.0394
25	0.0434	142	0.0381	186	0.0329	186	0.03	173	0.0364
142	0.0325	25	0.037	25	0.0313	173	0.0293	186	0.0307
173	0.0278	173	0.0267	173	0.0287	25	0.0277	25	0.0181
83	0.0268	83	0.0242	199	0.0248	199	0.0207	199	0.0165
276	0.0265	199	0.0223	83	0.0193	276	0.0173	142	0.0154
199	0.0257	189	0.0211	276	0.0174	83	0.0166	276	0.0152
189	0.0223	276	0.0177	189	0.0151	189	0.0154	144	0.0143
144	0.0212	202	0.0149	202	0.0145	188	0.0143	112	0.0122
202	0.0151	45	0.0094	188	0.013	202	0.0136		
57	0.0106	144	0.0088						
112	0.0106								
45	0.0086								
53	0.0083								

Task 36		Task 37		Task 38		Task 39		Task 40	
2003–2014		2004–2015		2005–2016		1995–2006		1996–2007	
Residue	Score	Residue	Score	Residue	Score	Residue	Score	Residue	Score
196	1.228	196	1.2111	196	0.6069	50	0.2303	140	0.5502
193	0.2522	138	0.2241	138	0.5875	140	0.1809	173	0.3494
227	0.2083	193	0.17	188	0.1396	193	0.1481	50	0.3266
145	0.1952	227	0.1245	50	0.0425	186	0.1111	186	0.0888
138	0.1236	225	0.069	193	0.0364	173	0.0734	156	0.0549
159	0.0697	188	0.0515	144	0.0282	156	0.0409	193	0.0445
188	0.059	145	0.0507	186	0.0227	142	0.0292	138	0.0263
173	0.0358	186	0.0341	145	0.0181	225	0.0206	142	0.0234
50	0.0325	50	0.0328	225	0.0144	138	0.0176	225	0.0169
186	0.0319	144	0.0231	173	0.0111	223	0.0141	223	0.0148
226	0.0259	199	0.0164	78	0.0071	188	0.0136	53	0.0146
144	0.0204	173	0.0154	142	0.0069	199	0.0113	226	0.0098
199	0.0162	112	0.0139	227	0.006	112	0.0075	199	0.0096
156	0.0141	242	0.0123	199	0.0057	67	0.0073	262	0.0089
67	0.0117	67	0.0101	112	0.0054	278	0.005	192	0.0078
189	0.0107	189	0.0087	223	0.0039	145	0.0048	57	0.0069
225	0.0104	156	0.008			196	0.0045	260	0.0035
242	0.0102					226	0.0033	188	0.0032
142	0.0083					260	0.0029		
						242	0.0023		
						122	0.0022		

Task 41		Task 42		Task 43		Task 44		Task 45	
1997–2008		1998–2009		1999–2010		2000–2011		2001–2012	
Residue	Score	Residue	Score	Residue	Score	Residue	Score	Residue	Score
173	0.5471	173	0.8999	158	0.8971	158	0.7814	158	0.6634
140	0.4485	158	0.8527	173	0.6865	173	0.4238	189	0.2762
50	0.2167	189	0.7672	189	0.6296	189	0.4212	173	0.275
186	0.0859	140	0.3221	140	0.2346	140	0.162	145	0.1636
156	0.0813	144	0.1824	144	0.0953	199	0.1251	278	0.1309
138	0.0342	62	0.1239	312	0.0789	312	0.0962	45	0.1254
45	0.0317	156	0.0717	213	0.0713	144	0.0723	53	0.1243
193	0.0273	212	0.0553	156	0.0639	213	0.0707	199	0.1218
225	0.0258	213	0.0445	62	0.0628	280	0.0674	144	0.1197
142	0.0177	45	0.0439	199	0.0615	45	0.0571	140	0.0872
57	0.0169	214	0.0312	280	0.0605	53	0.0505	223	0.0704
53	0.0166	50	0.0213	260	0.0511	156	0.0484	213	0.0645
158	0.0135	225	0.019	45	0.0445	223	0.0482	280	0.0519
223	0.0114	260	0.0182	50	0.0344	212	0.0479	198	0.0512
199	0.0085	138	0.0153	214	0.0221	260	0.038	312	0.0496
192	0.0081	312	0.0129	142	0.021	142	0.036	142	0.0425
262	0.0077	219	0.0128	186	0.0204	145	0.0356	156	0.0393
226	0.0065	145	0.0108			62	0.0322	260	0.038
260	0.0035	199	0.0107			50	0.0298	212	0.0373
		192	0.0086					57	0.0216
								225	0.015

								214	0.0139
								192	0.0133
Task 46		Task 47		Task 48		Task 49		Task 50	
2002–2013		2003–2014		2004–2015		2005–2016		2012–2016	
Residue	Score	Residue	Score	Residue	Score	Residue	Score	Residue	Score
158	0.6048	158	0.6048	158	0.5993	158	0.5911	159	0.3456
145	0.3163	145	0.3163	145	0.3171	145	0.319	311	0.3311
173	0.2223	173	0.2223	173	0.2192	173	0.2143	138	0.2337
189	0.211	189	0.211	189	0.2069	278	0.2072	160	0.1881
278	0.2072	278	0.2072	278	0.2064	189	0.2063	142	0.1612
45	0.1381	45	0.1381	45	0.1386	45	0.1391	5	0.1285
144	0.115	144	0.115	144	0.1152	144	0.1156	144	0.1188
199	0.1005	199	0.1005	199	0.0997	199	0.1005	225	0.1163
53	0.0958	53	0.0958	53	0.0948	53	0.0949	62	0.0604
198	0.095	198	0.095	198	0.0874	198	0.087	214	0.0449
280	0.076	280	0.076	280	0.0783	280	0.0769	31	0.0404
140	0.0731	140	0.0731	140	0.0717	140	0.0682	83	0.0354
142	0.0543	142	0.0543	142	0.0544	142	0.0546	122	0.0316
213	0.0535	213	0.0535	213	0.0541	213	0.0542	53	0.0312
223	0.0483	223	0.0482	223	0.0403	223	0.0393	124	0.0279
156	0.0287	156	0.0287	156	0.0276	156	0.0273	140	0.0243
260	0.0274	260	0.0274	260	0.0261	260	0.0267	156	0.0109
57	0.0221	57	0.0221	57	0.0221	57	0.0219	78	0.0104
230	0.0209	230	0.0209	230	0.0212	230	0.021	198	0.0087
312	0.0179			312	0.019	312	0.0189	219	0.0085
225	0.0173			225	0.018	225	0.0169	193	0.0084
62	0.0159			62	0.0155	62	0.0145		
				214	0.013	214	0.013		
				192	0.0124	192	0.0124		
				193	0.0097	212	0.01		
				78	0.0091	78	0.0092		
				212	0.0084				

Supplementary Figures

Figure S1. Overview of sequence and serologic data used in this study of influenza A(H3N2) virus variants. A) Distribution of HA sequences in each year, 1968-2016; B) HI titer data generated by using turkey erythrocytes without neuraminidase inhibitor (left panel), guinea pig erythrocytes without neuraminidase inhibitor (middle panel), and guinea pig erythrocytes with neuraminidase inhibitor (right panel). Most entries in these matrices are missing values (indicated by dark blue). The effective HI titers mainly locate close to the diagonal region (the antigen and sera are sorted according to the time). The $\log_2$(HI titers) are shown in colors, from red (high titer) to blue (missing [Miss] values).

Figure S2. Optimization of parameters in the machine learning models. (**A**) The model performance versus the parameter $\boldsymbol{\lambda}$ in matrix completion. (**B**) The model performance versus the parameter $\boldsymbol{\phi}$ in multi-task feature learning. (**C**) The model performance versus the parameter $\boldsymbol{\alpha}$ in multi-task feature learning. (**D**) The average number of selected residues per task versus the parameter $\boldsymbol{\phi}$. (**E**) The number of the union of the selected residues in all tasks versus the parameter $\boldsymbol{\phi}$. RMSE, rooted mean square error.

Figure S3. The 3D structure of the HA protein with identified key mutations colored. Different colors indicate that the identified key mutations belong to different antibody binding sites (ABS) A-E and regions outside the ABS.

Figure S4. Clusters of influenza A(H3N2) virus antigenic variants in antigenic cartography. The antigenic clusters were generated by using spectral clustering. Due to the imbalance of the virus samples, the entire datasets were split into three parts with respect to the time periods 1968-1992 (**A**), 1992-2009 (**B**), and 2009-2016 (**C**). A total of 16 antigenic clusters were generated: HK68, EN72, VI75, TX77, SI87, BE89, BE92, WU95, SY97, FU02, CA04, BR07, PE09, TX12, SWZ13, and HK14. Red crosses indicate the position of pandemic influenza strain A/Hong Kong/1968(H3N2) and vaccine strains determined by World Health Organization collaborative centers; markers with white face and black edge indicate the center of each cluster.