

Supplemental Data

A New Statistical Method for Predicting

Classical HLA Alleles from SNP Data

Stephen Leslie, Peter Donnelly, and Gil McVean

Table S1. Performance of HLA-Allele Predictions in the Training Data for YRI at Two-Digit Resolution

Allele	No. Train	No. Test ^a	No. Call ^b	Correct	Sensitivity ^c	Specificity ^d
<i>HLA-A</i>						
1	2	2	0	0	NA	0
2	11	11	11	11	1	1
3	15	14	14	14	1	1
23	13	11	11	11	1	1
24	1	0	0	NA	NA	NA
26	3	3	3	3	1	1
29	2	2	2	2	1	1
30	17	15	15	14	0.933333	0.933333
32	1	0	0	NA	NA	NA
33	16	13	13	12	0.923077	0.923077
34	5	4	4	4	1	1
36	17	15	17	15	0.882353	1
66	2	2	2	2	1	1
68	9	9	9	9	1	1
74	5	4	4	4	1	1
80	1	0	0	NA	NA	NA
<i>HLA-C</i>						
1	1	0	0	NA	NA	NA
2	11	10	10	10	1	1
3	5	5	5	5	1	1
4	30	26	26	26	1	1
5	1	0	0	NA	NA	NA
6	1	0	0	NA	NA	NA
7	23	22	22	22	1	1
8	12	11	11	11	1	1
12	1	0	0	NA	NA	NA
14	1	0	0	NA	NA	NA
15	2	2	2	2	1	1
16	20	16	16	15	0.9375	0.9375
17	10	8	8	7	0.875	0.875
18	2	1	1	1	1	1
<i>HLA-B</i>						
7	9	9	9	9	1	1
8	1	0	0	NA	NA	NA
13	3	3	3	3	1	1
14	3	3	5	3	0.6	1
15	18	17	17	17	1	1
18	6	6	6	6	1	1
35	10	8	3	3	1	0.375

39	2	2	0	0	NA	0
42	10	8	8	8	1	1
44	2	1	1	1	1	1
45	2	2	2	2	1	1
49	6	5	5	5	1	1
51	5	4	1	0	0	0
52	10	8	12	8	0.666667	1
53	19	16	20	15	0.75	0.9375
56	1	0	0	NA	NA	NA
57	4	3	3	3	1	1
58	7	7	7	7	1	1
78	1	0	0	NA	NA	NA
81	1	0	0	NA	NA	NA
<i>HLA-DRB1</i>						
1	6	6	6	6	1	1
3	22	19	20	19	0.95	1
4	1	0	0	NA	NA	NA
7	15	11	11	11	1	1
8	9	9	10	8	0.8	0.888889
9	3	2	2	2	1	1
11	19	15	14	14	1	0.933333
12	2	2	1	1	1	0.5
13	28	27	28	25	0.892857	0.925926
14	3	3	2	2	1	0.666667
15	12	10	10	10	1	1
<i>HLA-DQA1</i>						
1	51	44	44	42	0.954545	0.954545
2	18	17	17	17	1	1
3	7	5	5	5	1	1
4	16	16	16	16	1	1
5	28	26	26	24	0.923077	0.923077
<i>HLA-DQB1</i>						
2	33	30	30	30	1	1
3	23	20	20	20	1	1
4	13	13	13	13	1	1
5	25	20	20	20	1	1
6	26	23	23	23	1	1

^aThe numbers in the training and testing sets can differ because (1) alleles that had been imputed were included in the training, but not testing sets, and (2) singleton alleles were excluded.

^bThe number of times a given allele was predicted.

^cSensitivity represents the number correct divided by the number of calls. Note that calls are reported for a call threshold of 0.

^dSpecificity represents the number correct divided by the number of tests.

Table S2. Performance of HLA-Allele Predictions in the Training Data for CEU at Two-Digit Resolution

Allele	No. Train	No. Test ^a	No. Call ^b	Correct	Sensitivity ^c	Specificity ^d
<i>HLA-A</i>						
1	36	34	32	32	1	0.941176
2	66	63	61	61	1	0.968254
3	17	16	18	16	0.888889	1

American Journal of Human Genetics, Volume 82

11	7	7	7	7	1	1
23	3	3	3	3	1	1
24	17	13	13	13	1	1
25	7	6	8	6	0.75	1
26	4	4	2	2	1	0.5
29	4	4	4	4	1	1
30	2	2	2	2	1	1
31	6	6	6	6	1	1
32	5	5	5	5	1	1
68	6	5	7	5	0.714286	1
<hr/> <i>HLA-C</i> <hr/>						
1	3	3	3	3	1	1
2	4	4	4	4	1	1
3	19	19	19	19	1	1
4	15	12	12	12	1	1
5	21	20	21	20	0.952381	1
6	21	17	17	17	1	1
7	74	68	68	68	1	1
8	3	2	2	2	1	1
12	5	5	5	5	1	1
14	3	1	1	1	1	1
15	5	5	5	5	1	1
16	6	6	5	5	1	0.833333
17	1	0	0	NA	NA	NA
<hr/> <i>HLA-B</i> <hr/>						
7	33	27	28	27	0.964286	1
8	25	23	22	21	0.954545	0.913043
13	3	3	3	3	1	1
14	4	3	3	3	1	1
15	15	14	14	14	1	1
18	7	6	7	6	0.857143	1
27	3	3	3	3	1	1
35	12	9	9	9	1	1
37	2	2	2	2	1	1
38	1	0	0	NA	NA	NA
39	3	3	3	3	1	1
40	10	10	10	10	1	1
41	1	0	0	NA	NA	NA
44	29	28	28	28	1	1
49	2	2	2	2	1	1
50	3	3	3	3	1	1
51	8	6	6	6	1	1
52	3	3	2	2	1	0.666667
55	2	2	0	0	NA	0
56	2	2	4	2	0.5	1
57	11	10	10	10	1	1
58	1	0	0	NA	NA	NA
<hr/> <i>HLA-DRB1</i> <hr/>						
1	16	13	13	13	1	1
3	17	15	13	13	1	0.866667
4	42	39	39	39	1	1
7	23	21	21	21	1	1
8	6	4	6	4	0.666667	1
10	1	0	0	NA	NA	NA

11	12	12	12	12	1	1
12	1	0	0	NA	NA	NA
13	14	12	12	10	0.833333	0.833333
14	7	6	6	6	1	1
15	37	33	33	32	0.969697	0.969697
16	4	3	3	3	1	1
<hr/>						
<i>HLA-DQA1</i>						
1	77	75	76	75	0.986842	1
2	24	23	23	23	1	1
3	41	41	41	41	1	1
4	6	4	4	4	1	1
5	32	31	30	30	1	0.967742
<hr/>						
<i>HLA-DQB1</i>						
2	28	27	26	26	1	0.962963
3	68	68	68	68	1	1
4	6	4	4	4	1	1
5	28	26	26	26	1	1
6	50	49	50	49	0.98	1

^aThe numbers in the training and testing sets can differ because (1) alleles that had been imputed were included in the training, but not testing sets and (2) singleton alleles were excluded.

^bThe number of times a given allele was predicted.

^cSensitivity represents the number correct divided by the number of calls. Note that calls are reported for a call threshold of 0.

^dSpecificity represents the number correct divided by the number of tests.

Table S3. Performance of HLA-Allele Predictions in the Training Data for YRI at Four-Digit Resolution

Allele	No. Train	No. Test ^a	No. Call ^b	Correct	Sensitivity ^c	Specificity ^d
<hr/>						
<i>HLA-A</i>						
101	2	2	0	0	NA	0
201	5	5	5	5	1	1
202	3	3	3	3	1	1
205	2	2	2	2	1	1
211	1	0	0	NA	NA	NA
301	15	14	14	14	1	1
2301	13	11	11	11	1	1
2424	1	0	0	NA	NA	NA
2601	3	3	3	3	1	1
2902	2	2	2	2	1	1
3001	11	9	9	8	0.888889	0.888889
3002	6	6	6	6	1	1
3201	1	0	0	NA	NA	NA
3301	3	3	3	3	1	1
3303	13	10	10	9	0.9	0.9
3402	5	4	4	4	1	1
3601	17	15	17	15	0.882353	1
6602	2	2	2	2	1	1
6801	2	2	2	2	1	1
6802	7	7	7	7	1	1
7401	5	4	4	4	1	1
8001	1	0	0	NA	NA	NA

American Journal of Human Genetics, Volume 82

<i>HLA-C</i>						
102	1	0	0	NA	NA	NA
202	11	10	10	10	1	1
302	5	5	5	5	1	1
401	30	26	26	26	1	1
501	1	0	0	NA	NA	NA
602	1	0	0	NA	NA	NA
701	16	15	15	15	1	1
702	7	7	7	7	1	1
802	6	5	4	4	1	0.8
804	6	6	7	6	0.857143	1
1203	1	0	0	NA	NA	NA
1402	1	0	0	NA	NA	NA
1505	2	2	2	2	1	1
1601	20	16	16	15	0.9375	0.9375
1701	10	8	8	7	0.875	0.875
1801	2	1	1	1	1	1
<i>HLA-B</i>						
702	9	9	9	9	1	1
801	1	0	0	NA	NA	NA
1302	3	3	3	3	1	1
1401	1	0	0	NA	NA	NA
1402	1	0	0	NA	NA	NA
1403	1	0	0	NA	NA	NA
1503	7	6	6	6	1	1
1510	10	10	10	10	1	1
1516	1	0	0	NA	NA	NA
1801	6	6	6	6	1	1
3501	10	8	3	3	1	0.375
3903	1	0	0	NA	NA	NA
3910	1	0	0	NA	NA	NA
4201	9	7	7	7	1	1
4202	1	0	0	NA	NA	NA
4403	2	1	1	1	1	1
4501	2	2	2	2	1	1
4901	6	5	5	5	1	1
5101	5	4	1	1	1	0.25
5201	10	8	11	8	0.727273	1
5301	19	16	21	16	0.761905	1
5601	1	0	0	NA	NA	NA
5702	1	0	0	NA	NA	NA
5703	3	2	2	2	1	1
5801	7	7	7	7	1	1
7801	1	0	0	NA	NA	NA
8101	1	0	0	NA	NA	NA
<i>HLA-DRBI</i>						
101	2	2	3	2	0.666667	1
102	4	4	4	4	1	1
301	11	9	10	9	0.9	1
302	11	10	11	10	0.909091	1
403	1	0	0	NA	NA	NA
701	15	11	11	11	1	1
804	9	9	10	8	0.8	0.888889
901	3	2	2	2	1	1

American Journal of Human Genetics, Volume 82

1101	7	6	5	5	1	0.833333
1102	3	3	3	3	1	1
1104	9	6	5	5	1	0.833333
1201	2	2	0	0	NA	0
1301	10	10	8	8	1	0.8
1302	7	6	6	5	0.833333	0.833333
1303	9	9	10	9	0.9	1
1327	1	0	0	NA	NA	NA
1335	1	0	1	NA	0	NA
1401	3	3	3	3	1	1
1501	1	0	0	NA	NA	NA
1503	11	9	9	9	1	1

HLA-DQA1

101	12	11	12	11	0.916667	1
102	29	24	22	20	0.909091	0.833333
103	10	9	10	9	0.9	1
201	18	17	17	17	1	1
301	7	5	5	5	1	1
401	16	16	16	16	1	1
501	28	26	26	24	0.923077	0.923077

HLA-DQB1

201	33	30	30	30	1	1
301	21	19	19	19	1	1
302	2	1	1	1	1	1
402	13	13	13	13	1	1
501	15	12	12	12	1	1
502	7	6	6	6	1	1
503	3	2	2	2	1	1
602	12	10	11	10	0.909091	1
603	7	6	6	6	1	1
604	2	2	1	1	1	0.5
605	5	5	5	4	0.8	0.8

^aThe numbers in the training and testing sets can differ because (1) alleles that had been imputed were included in the training, but not testing sets and (2) singleton alleles were excluded.

^bThe number of times a given allele was predicted.

^cSensitivity represents the number correct divided by the number of calls. Note that calls are reported for a call threshold of 0.

^dSpecificity represents the number correct divided by the number of tests.

Table S4. Performance of HLA-Allele Predictions in the Training Data for CEU at Four-Digit Resolution

Allele	No. Train	No. Test ^a	No. Call ^b	Correct	Sensitivity ^c	Specificity ^d
<i>HLA-A</i>						
101	34	32	31	31	1	0.96875
201	59	56	56	56	1	1
205	2	2	2	2	1	1
206	1	0	0	NA	NA	NA
301	17	16	17	16	0.941176	1
1101	7	7	7	7	1	1
2301	3	3	3	3	1	1
2402	16	12	12	12	1	1

American Journal of Human Genetics, Volume 82

2403	1	0	0	NA	NA	NA
2501	7	6	8	6	0.75	1
2601	4	4	2	2	1	0.5
2902	3	3	3	3	1	1
3001	1	0	0	NA	NA	NA
3002	1	0	0	NA	NA	NA
3101	6	6	6	6	1	1
3201	5	5	5	5	1	1
6801	6	5	5	5	1	1
<hr/>						
<i>HLA-C</i>						
<hr/>						
102	3	3	3	3	1	1
202	4	4	4	4	1	1
303	11	11	10	10	1	0.909091
304	7	7	8	7	0.875	1
401	15	12	12	12	1	1
501	21	20	21	20	0.952381	1
602	21	17	17	17	1	1
701	32	30	30	29	0.966667	0.966667
702	37	33	33	32	0.969697	0.969697
704	5	5	5	5	1	1
802	3	2	2	2	1	1
1202	2	2	2	2	1	1
1203	3	3	3	3	1	1
1402	3	1	1	1	1	1
1502	5	5	5	5	1	1
1601	6	6	5	5	1	0.833333
1701	1	0	0	NA	NA	NA
<hr/>						
<i>HLA-B</i>						
<hr/>						
702	33	27	28	27	0.964286	1
801	25	23	21	21	1	0.913043
1302	3	3	3	3	1	1
1401	1	0	0	NA	NA	NA
1402	2	1	0	0	NA	0
1501	13	12	12	12	1	1
1510	1	0	0	NA	NA	NA
1801	7	6	7	6	0.857143	1
2705	1	0	0	NA	NA	NA
3501	9	6	7	6	0.857143	1
3503	2	2	0	0	NA	0
3701	2	2	2	2	1	1
3801	1	0	0	NA	NA	NA
3905	1	0	0	NA	NA	NA
3906	2	2	3	2	0.666667	1
4001	7	7	7	7	1	1
4002	3	3	3	3	1	1
4101	1	0	0	NA	NA	NA
4402	22	21	21	21	1	1
4403	4	4	4	4	1	1
4404	1	0	0	NA	NA	NA
4405	1	0	0	NA	NA	NA
4901	2	2	2	2	1	1
5001	3	3	3	3	1	1
5101	5	3	4	3	0.75	1
5201	3	3	2	2	1	0.666667

American Journal of Human Genetics, Volume 82

5601	2	2	3	2	0.666667	1
5701	11	10	10	10	1	1
5801	1	0	0	NA	NA	NA
<i>HLA-DRB1</i>						
101	16	13	13	13	1	1
301	17	15	13	13	1	0.866667
401	16	16	11	10	0.909091	0.625
402	1	0	0	NA	NA	NA
403	3	3	1	1	1	0.333333
404	19	16	24	16	0.666667	1
405	1	0	0	NA	NA	NA
407	2	2	1	1	1	0.5
701	18	16	16	16	1	1
801	6	4	4	4	1	1
1001	1	0	0	NA	NA	NA
1101	9	9	11	9	0.818182	1
1104	1	0	0	NA	NA	NA
1103	2	2	0	0	NA	0
1201	1	0	0	NA	NA	NA
1301	8	6	7	6	0.857143	1
1302	4	4	4	4	1	1
1303	2	2	2	2	1	1
1401	6	5	5	5	1	1
1404	1	0	0	NA	NA	NA
1501	36	32	33	32	0.969697	1
1502	1	0	0	NA	NA	NA
1601	4	3	3	3	1	1
<i>HLA-DQA1</i>						
101	24	23	23	23	1	1
102	43	42	43	42	0.976744	1
103	10	10	10	10	1	1
201	24	23	23	23	1	1
301	41	41	41	41	1	1
401	6	4	4	4	1	1
501	32	31	30	30	1	0.967742
<i>HLA-DQB1</i>						
201	28	27	26	26	1	0.962963
301	30	30	30	30	1	1
302	26	26	26	26	1	1
303	11	11	11	11	1	1
305	1	0	0	NA	NA	NA
402	6	4	4	4	1	1
501	17	17	17	17	1	1
502	4	3	3	3	1	1
503	7	6	6	6	1	1
601	1	0	0	NA	NA	NA
602	36	35	36	35	0.972222	1
603	9	9	9	9	1	1
604	3	3	3	3	1	1
609	1	0	0	NA	NA	NA

^aThe numbers in the training and testing sets can differ because (1) alleles that had been imputed were included in the training, but not testing sets and (2) singleton alleles were excluded.

^bThe number of times a given allele was predicted.

^cSensitivity represents the number correct divided by the number of calls. Note that calls are reported for a call threshold of 0.

^dSpecificity represents the number correct divided by the number of tests.

Table S5. Performance of HLA-Allele Predictions in the Validation Data (1958 Birth Cohort) at Two-Digit Resolution with the Affymetrix 500K SNP Data

Allele	No. Database ^a	No. Test	No. Call	Correct	Sensitivity ^b	Specificity ^c
<i>HLA-A</i>						
1	34	322	321	315	0.98	0.98
2	64	541	558	532	0.95	0.98
3	16	249	246	246	1.00	0.99
11	7	118	110	108	0.98	0.92
23	3	27	47	27	0.57	1.00
24	13	128	121	116	0.96	0.91
25	6	27	45	23	0.51	0.85
26	4	43	42	25	0.60	0.58
29	4	76	83	74	0.89	0.97
30	2	47	28	20	0.71	0.43
31	6	49	60	47	0.78	0.96
32	5	73	73	70	0.96	0.96
33	0	12	0	0	NA	0.00
34	0	1	0	0	NA	0.00
66	0	5	0	0	NA	0.00
68	5	73	58	58	1.00	0.79
74	0	1	0	0	NA	0.00
<i>HLA-B</i>						
7	27	243	228	223	0.98	0.92
8	23	225	222	217	0.98	0.96
13	3	25	55	25	0.45	1.00
14	3	72	84	68	0.81	0.94
15	14	146	168	133	0.79	0.91
18	6	65	58	52	0.90	0.80
27	3	74	28	24	0.86	0.32
35	10	108	99	82	0.83	0.76
37	2	28	28	23	0.82	0.82
38	1	11	25	10	0.40	0.91
39	3	32	33	19	0.58	0.59
40	10	122	162	122	0.75	1.00
41	1	9	7	7	1.00	0.78
44	28	286	281	272	0.97	0.95
45	0	14	0	0	NA	0.00
47	0	4	0	0	NA	0.00
49	2	17	26	14	0.54	0.82
50	3	16	33	15	0.45	0.94
51	6	65	43	36	0.84	0.55
52	3	14	29	13	0.45	0.93
53	0	2	0	0	NA	0.00
55	2	41	0	0	NA	0.00
56	2	8	18	7	0.39	0.88
57	10	70	67	65	0.97	0.93
58	1	11	14	5	0.36	0.45
<i>HLA-DRB1</i>						

1	13	207	175	173	0.99	0.84
3	16	231	218	202	0.93	0.87
4	39	392	463	392	0.85	1.00
7	21	253	253	226	0.89	0.89
8	4	43	38	37	0.97	0.86
9	0	30	0	0	NA	0.00
10	1	11	10	3	0.30	0.27
11	12	119	125	102	0.82	0.86
12	1	29	33	23	0.70	0.79
13	12	176	131	118	0.90	0.67
14	6	36	74	28	0.38	0.78
15	33	258	245	235	0.96	0.91
16	3	13	33	8	0.24	0.62

HLA-DQB1

2	28	397	345	333	0.97	0.84
3	68	658	625	592	0.95	0.90
4	4	47	34	34	1.00	0.72
5	26	261	314	235	0.75	0.90
6	49	411	456	399	0.88	0.97

^aOnly the CEU population data was used in predicting HLA alleles in the 1958 birth cohort data.

^bSensitivity represents the number correct divided by the number of calls. Note that calls are reported for a call threshold of 0.

^cSpecificity represents the number correct divided by the number of tests.

Table S6. Performance of HLA-Allele Predictions in the Validation Data (1958 Birth Cohort) at Two-Digit Resolution with the Illumina 550K and 15K SNP Data

Allele	No. Database ^a	No. Test	No. Call	Correct	Sensitivity ^b	Specificity ^c
<i>HLA-A</i>						
1	34	322	322	317	0.98	0.98
2	64	541	574	541	0.94	1.00
3	16	249	233	233	1.00	0.94
11	7	118	112	108	0.96	0.92
23	3	27	28	25	0.89	0.93
24	13	128	121	120	0.99	0.94
25	6	27	39	24	0.62	0.89
26	4	43	36	28	0.78	0.65
29	4	76	72	71	0.99	0.93
30	2	47	53	44	0.83	0.94
31	6	49	59	47	0.80	0.96
32	5	73	75	73	0.97	1.00
33	0	12	0	0	NA	0.00
34	0	1	0	0	NA	0.00
66	0	5	0	0	NA	0.00
68	5	73	68	67	0.99	0.92
74	0	1	0	0	NA	0.00
<i>HLA-B</i>						
7	27	243	264	231	0.88	0.95
8	23	225	221	219	0.99	0.97
13	3	25	37	24	0.65	0.96
14	3	72	25	23	0.92	0.32
15	14	146	140	138	0.99	0.95
18	6	65	70	62	0.89	0.95

27	3	74	45	37	0.82	0.50
35	10	108	102	96	0.94	0.89
37	2	28	27	21	0.78	0.75
38	1	11	18	10	0.56	0.91
39	3	32	77	24	0.31	0.75
40	10	122	151	121	0.80	0.99
41	1	9	9	7	0.78	0.78
44	28	286	278	267	0.96	0.93
45	0	14	0	0	NA	0.00
47	0	4	0	0	NA	0.00
49	2	17	31	15	0.48	0.88
50	3	16	20	15	0.75	0.94
51	6	65	46	45	0.98	0.69
52	3	14	30	14	0.47	1.00
53	0	2	0	0	NA	0.00
55	2	41	0	0	NA	0.00
56	2	8	44	7	0.16	0.88
57	10	70	66	65	0.98	0.93
58	1	11	7	6	0.86	0.55
HLA-DRB1						
1	13	207	178	176	0.99	0.85
3	16	231	240	209	0.87	0.90
4	39	392	470	392	0.83	1.00
7	21	253	248	230	0.93	0.91
8	4	43	38	38	1.00	0.88
9	0	30	0	0	NA	0.00
10	1	11	4	0	0.00	0.00
11	12	119	125	111	0.89	0.93
12	1	29	32	20	0.63	0.69
13	12	176	128	108	0.84	0.61
14	6	36	70	25	0.36	0.69
15	33	258	244	240	0.98	0.93
16	3	13	21	12	0.57	0.92
HLA-DQB1						
2	28	397	386	380	0.98	0.96
3	68	658	655	621	0.95	0.94
4	4	47	44	42	0.95	0.89
5	26	261	278	228	0.82	0.87
6	49	411	411	380	0.92	0.92

^aOnly the CEU population data was used in predicting HLA alleles in the 1958 birth cohort data.

^bSensitivity represents the number correct divided by the number of calls. Note that calls are reported for a call threshold of 0.

^cSpecificity represents the number correct divided by the number of tests.

Table S7. Performance of HLA-Allele Predictions in the Validation Data (1958 Birth Cohort) at Four-Digit Resolution with the Affymetrix 500K SNP Data

Allele	No. Database ^a	No. Test	No. Call	Correct	Sensitivity ^b	Specificity ^c
HLA-A						
101	32	230	231	226	0.98	0.98
201	57	0	16	0	0.00	NA
202	0	1	0	0	NA	0.00
205	2	15	19	13	0.68	0.87

American Journal of Human Genetics, Volume 82

206	1	1	6	0	0.00	0.00
301	16	175	173	172	0.99	0.98
302	0	1	0	0	NA	0.00
1101	7	84	75	74	0.99	0.88
1119	0	1	0	0	NA	0.00
2301	3	19	38	19	0.50	1.00
2402	12	83	69	68	0.99	0.82
2403	1	1	8	0	0.00	0.00
2501	6	18	28	16	0.57	0.89
2601	4	23	20	13	0.65	0.57
2608	0	2	0	0	NA	0.00
2902	3	55	58	53	0.91	0.96
3001	1	17	22	16	0.73	0.94
3002	1	16	0	0	NA	0.00
3004	0	1	0	0	NA	0.00
3101	6	35	34	33	0.97	0.94
3201	5	39	39	37	0.95	0.95
3402	0	1	0	0	NA	0.00
6601	0	5	0	0	NA	0.00
6801	5	49	40	40	1.00	0.82
6802	0	3	0	0	NA	0.00
7406	0	1	0	0	NA	0.00
<hr/>						
<i>HLA-B</i>						
702	27	239	224	220	0.98	0.92
801	23	224	221	216	0.98	0.96
1302	3	25	55	25	0.45	1.00
1401	1	28	26	23	0.88	0.82
1402	1	43	57	43	0.75	1.00
1501	12	118	141	107	0.76	0.91
1503	0	2	0	0	NA	0.00
1507	0	1	0	0	NA	0.00
1510	1	0	10	0	0.00	NA
1516	0	3	0	0	NA	0.00
1517	0	4	0	0	NA	0.00
1518	0	6	0	0	NA	0.00
1801	6	62	55	50	0.91	0.81
2701	0	1	0	0	NA	0.00
2702	0	2	0	0	NA	0.00
2705	1	71	28	23	0.82	0.32
3501	7	62	64	51	0.80	0.82
3502	0	5	0	0	NA	0.00
3503	2	8	11	7	0.64	0.88
3508	0	4	0	0	NA	0.00
3701	2	28	28	23	0.82	0.82
3801	1	11	24	10	0.42	0.91
3901	0	15	0	0	NA	0.00
3905	1	0	1	0	0.00	NA
3906	2	15	31	15	0.48	1.00
4001	7	112	109	107	0.98	0.96
4002	3	9	49	9	0.18	1.00
4101	1	4	5	2	0.40	0.50
4102	0	3	0	0	NA	0.00
4402	21	188	178	174	0.98	0.93
4403	4	85	83	76	0.92	0.89
4404	1	3	5	1	0.20	0.33
4405	1	1	6	1	0.17	1.00

American Journal of Human Genetics, Volume 82

4501	0	14	0	0	NA	0.00
4701	0	4	0	0	NA	0.00
4901	2	17	26	14	0.54	0.82
5001	3	16	33	15	0.45	0.94
5101	3	58	41	34	0.83	0.59
5201	3	14	26	13	0.50	0.93
5301	0	2	0	0	NA	0.00
5501	0	41	0	0	NA	0.00
5601	2	8	18	7	0.39	0.88
5701	10	63	61	59	0.97	0.94
5801	1	11	14	5	0.36	0.45

HLA-DRB1

101	13	4	28	3	0.11	0.75
102	0	5	0	0	NA	0.00
103	0	23	0	0	NA	0.00
104	0	1	0	0	NA	0.00
301	16	169	151	145	0.96	0.86
401	16	133	156	111	0.71	0.83
402	1	4	5	0	0.00	0.00
403	3	43	9	3	0.33	0.07
404	16	60	86	44	0.51	0.73
405	1	6	8	2	0.25	0.33
407	2	3	13	0	0.00	0.00
701	16	177	172	159	0.92	0.90
801	4	19	18	18	1.00	0.95
804	0	1	0	0	NA	0.00
901	0	18	0	0	NA	0.00
1001	1	4	5	1	0.20	0.25
1101	9	23	38	21	0.55	0.91
1102	0	4	0	0	NA	0.00
1103	2	6	4	1	0.25	0.17
1104	1	6	0	0	NA	0.00
1201	1	17	17	12	0.71	0.71
1301	6	20	10	8	0.80	0.40
1302	4	13	12	9	0.75	0.69
1303	2	12	12	11	0.92	0.92
1401	5	27	42	21	0.50	0.78
1404	1	0	4	0	0.00	NA
1501	32	26	25	24	0.96	0.92
1502	1	2	4	1	0.25	0.50
1601	3	7	15	4	0.27	0.57
1602	0	1	0	0	NA	0.00

HLA-DQB1

201	28	0	13	0	0.00	NA
203	0	1	0	0	NA	0.00
301	30	270	252	211	0.84	0.78
302	26	167	145	128	0.88	0.77
303	11	77	63	49	0.78	0.64
305	1	0	18	0	0.00	NA
402	4	39	28	28	1.00	0.72
501	17	168	148	142	0.96	0.85
502	3	10	33	2	0.06	0.20
503	6	28	51	23	0.45	0.82
504	0	1	0	0	NA	0.00
601	1	8	12	0	0.00	0.00

American Journal of Human Genetics, Volume 82

602	35	195	172	164	0.95	0.84
603	9	64	92	51	0.55	0.80
604	3	42	40	34	0.85	0.81
609	1	18	21	5	0.24	0.28

^aOnly the CEU population data was used in predicting HLA alleles in the 1958 birth cohort data.

^bSensitivity represents the number correct divided by the number of calls. Note that calls are reported for a call threshold of 0.

^cSpecificity represents the number correct divided by the number of tests.

Table S8. Performance of HLA-Allele Predictions in the VALIDATION data (1958 Birth Cohort) at Four-Digit Resolution with the Illumina 550K and 15K SNP Data

Allele	No. Database ^a	No. Test	No. Call	Correct	Sensitivity ^b	Specificity ^c
<i>HLA-A</i>						
101	32	230	232	227	0.98	0.99
201	57	0	19	0	0.00	NA
202	0	1	0	0	NA	0.00
205	2	15	17	14	0.82	0.93
206	1	1	2	0	0.00	0.00
301	16	175	163	162	0.99	0.93
302	0	1	0	0	NA	0.00
1101	7	84	79	77	0.97	0.92
1119	0	1	0	0	NA	0.00
2301	3	19	21	18	0.86	0.95
2402	12	83	72	70	0.97	0.84
2403	1	1	7	0	0.00	0.00
2501	6	18	25	16	0.64	0.89
2601	4	23	23	16	0.70	0.70
2608	0	2	0	0	NA	0.00
2902	3	55	53	52	0.98	0.95
3001	1	17	18	17	0.94	1.00
3002	1	16	23	14	0.61	0.88
3004	0	1	0	0	NA	0.00
3101	6	35	33	33	1.00	0.94
3201	5	39	40	39	0.98	1.00
3402	0	1	0	0	NA	0.00
6601	0	5	0	0	NA	0.00
6801	5	49	49	45	0.92	0.92
6802	0	3	0	0	NA	0.00
7406	0	1	0	0	NA	0.00
<i>HLA-B</i>						
702	27	239	259	228	0.88	0.95
801	23	224	220	218	0.99	0.97
1302	3	25	37	24	0.65	0.96
1401	1	28	23	23	1.00	0.82
1402	1	43	1	0	0.00	0.00
1501	12	118	119	110	0.92	0.93
1503	0	2	0	0	NA	0.00
1507	0	1	0	0	NA	0.00
1510	1	0	10	0	0.00	NA
1516	0	3	0	0	NA	0.00
1517	0	4	0	0	NA	0.00
1518	0	6	0	0	NA	0.00
1801	6	62	67	59	0.88	0.95

American Journal of Human Genetics, Volume 82

2701	0	1	0	0	NA	0.00
2702	0	2	0	0	NA	0.00
2705	1	71	45	35	0.78	0.49
3501	7	62	64	50	0.78	0.81
3502	0	5	0	0	NA	0.00
3503	2	8	10	5	0.50	0.63
3508	0	4	0	0	NA	0.00
3701	2	28	27	21	0.78	0.75
3801	1	11	18	10	0.56	0.91
3901	0	15	0	0	NA	0.00
3905	1	0	1	0	0.00	NA
3906	2	15	72	15	0.21	1.00
4001	7	112	109	108	0.99	0.96
4002	3	9	41	9	0.22	1.00
4101	1	4	7	2	0.29	0.50
4102	0	3	0	0	NA	0.00
4402	21	188	175	173	0.99	0.92
4403	4	85	91	77	0.85	0.91
4404	1	3	2	0	0.00	0.00
4405	1	1	1	0	0.00	0.00
4501	0	14	0	0	NA	0.00
4701	0	4	0	0	NA	0.00
4901	2	17	30	15	0.50	0.88
5001	3	16	20	15	0.75	0.94
5101	3	58	41	40	0.98	0.69
5201	3	14	29	14	0.48	1.00
5301	0	2	0	0	NA	0.00
5501	0	41	0	0	NA	0.00
5601	2	8	44	7	0.16	0.88
5701	10	63	60	59	0.98	0.94
5801	1	11	7	6	0.86	0.55
<hr/>						
<i>HLA-DRB1</i>						
101	13	4	26	3	0.12	0.75
102	0	5	0	0	NA	0.00
103	0	23	0	0	NA	0.00
104	0	1	0	0	NA	0.00
301	16	169	152	147	0.97	0.87
401	16	133	143	114	0.80	0.86
402	1	4	19	3	0.16	0.75
403	3	43	18	3	0.17	0.07
404	16	60	85	46	0.54	0.77
405	1	6	2	0	0.00	0.00
407	2	3	13	0	0.00	0.00
701	16	177	169	161	0.95	0.91
801	4	19	17	17	1.00	0.89
804	0	1	0	0	NA	0.00
901	0	18	0	0	NA	0.00
1001	1	4	3	0	0.00	0.00
1101	9	23	35	20	0.57	0.87
1102	0	4	0	0	NA	0.00
1103	2	6	4	1	0.25	0.17
1104	1	6	5	0	0.00	0.00
1201	1	17	15	11	0.73	0.65
1301	6	20	13	11	0.85	0.55
1302	4	13	12	9	0.75	0.69
1303	2	12	17	10	0.59	0.83

1401	5	27	45	20	0.44	0.74
1404	1	0	0	0	NA	NA
1501	32	26	26	25	0.96	0.96
1502	1	2	5	2	0.40	1.00
1601	3	7	10	7	0.70	1.00
1602	0	1	0	0	NA	0.00
<hr/>						
<i>HLA-DQB1</i>						
201	28	0	7	0	0.00	NA
203	0	1	0	0	NA	0.00
301	30	270	253	250	0.99	0.93
302	26	167	189	164	0.87	0.98
303	11	77	53	50	0.94	0.65
305	1	0	8	0	0.00	NA
402	4	39	38	36	0.95	0.92
501	17	168	148	136	0.92	0.81
502	3	10	13	10	0.77	1.00
503	6	28	56	25	0.45	0.89
504	0	1	0	0	NA	0.00
601	1	8	13	6	0.46	0.75
602	35	195	181	176	0.97	0.90
603	9	64	67	42	0.63	0.66
604	3	42	44	34	0.77	0.81
609	1	18	18	15	0.83	0.83

^aOnly the CEU population data was used in predicting HLA alleles in the 1958 birth cohort data.

^bSensitivity represents the number correct divided by the number of calls. Note that calls are reported for a call threshold of 0.

^cSpecificity represents the number correct divided by the number of tests.

Table S9. Minimal SNP Prediction Sets in CEU and YRI

rsID	Position	<i>HLA-A</i>	<i>HLA-C</i>	<i>HLA-B</i>	<i>HLA-DRB1</i>	<i>HLA-DQA1</i>	<i>HLA-DQB1</i>
	NCBI Build 34 (bp)						
rs1632933	29905944	-/+ ^a	-/-	-/-	-/-	-/-	-/-
rs885928	29949566	+/-	-/-	-/-	-/-	-/-	-/-
rs2844821	29950109	+/-	-/-	-/-	-/-	-/-	-/-
rs2508042	29994117	+/-	-/-	-/-	-/-	-/-	-/-
rs2157680	30018226	+/-	-/-	-/-	-/-	-/-	-/-
rs3823339	30018808	+/-	-/-	-/-	-/-	-/-	-/-
rs3823343	30018951	+/+	-/-	-/-	-/-	-/-	-/-
rs1061235	30019138	+/-	-/-	-/-	-/-	-/-	-/-
rs2499	30019382	+/-	-/-	-/-	-/-	-/-	-/-
rs6921314	30020060	+/-	-/-	-/-	-/-	-/-	-/-
rs2735097	30021141	-/+	-/-	-/-	-/-	-/-	-/-
rs7745413	30021309	+/+	-/-	-/-	-/-	-/-	-/-
rs1632880	30023450	+/-	-/-	-/-	-/-	-/-	-/-
rs7747114	30025614	+/-	-/-	-/-	-/-	-/-	-/-
rs1655912	30026305	+/-	-/-	-/-	-/-	-/-	-/-
rs3893538	30030261	+/-	-/-	-/-	-/-	-/-	-/-
rs2256543	30043671	+/+	-/-	-/-	-/-	-/-	-/-
rs6457116	30043863	+/-	-/-	-/-	-/-	-/-	-/-
rs7754245	30044316	-/+	-/-	-/-	-/-	-/-	-/-
rs2523969	30044615	+/-	-/-	-/-	-/-	-/-	-/-
rs435766	30045689	+/-	-/-	-/-	-/-	-/-	-/-

American Journal of Human Genetics, Volume 82

rs2894003	30052418	+/-	-/-	-/-	-/-	-/-	-/-
rs3915798	30058271	-/+	-/-	-/-	-/-	-/-	-/-
rs5875226	30059808	-/+	-/-	-/-	-/-	-/-	-/-
rs9468675	30066556	-/+	-/-	-/-	-/-	-/-	-/-
rs1345274	31296875	-/-	-/+	-/-	-/-	-/-	-/-
rs3869115	31310909	-/-	+/-	-/-	-/-	-/-	-/-
rs3868082	31313896	-/-	+/-	-/-	-/-	-/-	-/-
rs3130425	31323067	-/-	-/-	-/+	-/-	-/-	-/-
rs3130542	31336843	-/-	+/-	-/-	-/-	-/-	-/-
rs2844623	31337305	-/-	+/-	-/-	-/-	-/-	-/-
rs2248902	31338876	-/-	+/-	-/-	-/-	-/-	-/-
rs2524100	31340517	-/-	+/+	-/-	-/-	-/-	-/-
rs2853950	31340923	-/-	+/+	-/-	-/-	-/-	-/-
rs2001181	31341746	-/-	+/+	-/-	-/-	-/-	-/-
rs7383157	31344054	-/-	+/-	-/-	-/-	-/-	-/-
rs2074491	31344644	-/-	+/-	-/-	-/-	-/-	-/-
rs6457358	31344747	-/-	+/-	-/-	-/-	-/-	-/-
rs2524094	31344789	-/-	-/+	-/-	-/-	-/-	-/-
rs2074489	31344876	-/-	+/+	-/-	-/-	-/-	-/-
rs2074488	31345179	-/-	+/-	-/-	-/-	-/-	-/-
rs7759127	31345736	-/-	+/-	-/-	-/-	-/-	-/-
rs6923313	31346118	-/-	-/+	-/-	-/-	-/-	-/-
rs3130696	31348627	-/-	+/+	-/-	-/-	-/-	-/-
rs2524074	31348764	-/-	-/+	-/-	-/-	-/-	-/-
rs2524069	31349532	-/-	-/-	+/-	-/-	-/-	-/-
rs9348859	31354709	-/-	+/-	-/-	-/-	-/-	-/-
rs3873379	31366897	-/-	-/-	-/+	-/-	-/-	-/-
rs2844586	31422213	-/-	-/-	-/+	-/-	-/-	-/-
rs2523619	31422333	-/-	-/-	+/-	-/-	-/-	-/-
rs2523618	31422492	-/-	-/-	+/-	-/-	-/-	-/-
rs2442719	31424730	-/-	-/-	+/-	-/-	-/-	-/-
rs2596503	31425002	-/-	-/-	+/+	-/-	-/-	-/-
rs2596501	31425403	-/-	-/-	-/+	-/-	-/-	-/-
rs1058026	31425877	-/-	-/-	+/+	-/-	-/-	-/-
rs2769	31426074	-/-	-/-	+/+	-/-	-/-	-/-
rs3819300	31426495	-/-	-/-	+/+	-/-	-/-	-/-
rs3819299	31426559	-/-	-/-	+/+	-/-	-/-	-/-
rs3819294	31426679	-/-	-/-	+/+	-/-	-/-	-/-
rs2523608	31426751	-/-	-/-	+/+	-/-	-/-	-/-
rs3819285	31426934	-/-	-/-	+/-	-/-	-/-	-/-
rs3819284	31426959	-/-	-/-	+/+	-/-	-/-	-/-
rs1051488	31427103	-/-	-/-	+/-	-/-	-/-	-/-
rs2394985	31427747	-/-	-/-	-/+	-/-	-/-	-/-
rs3998357	31431253	-/-	-/-	-/+	-/-	-/-	-/-
rs2596438	31444042	-/-	-/-	-/+	-/-	-/-	-/-
rs2507987	31447204	-/-	-/-	+/-	-/-	-/-	-/-
rs9266791	31463126	-/-	-/-	+/-	-/-	-/-	-/-
rs7751505	31464441	-/-	-/-	+/-	-/-	-/-	-/-
rs2256028	31483398	-/-	-/-	+/-	-/-	-/-	-/-
rs9268516	32450662	-/-	-/-	-/-	-/+	-/-	-/-
rs2395175	32476248	-/-	-/-	-/-	+/-	+/-	-/-
rs2395178	32476584	-/-	-/-	-/-	-/+	-/-	-/-
rs2239804	32482746	-/-	-/-	-/-	+/-	-/-	-/-
rs3135388	32484240	-/-	-/-	-/-	+/-	-/-	-/-

American Journal of Human Genetics, Volume 82

rs7452076	32486270	-/-	-/-	-/-	-/-	-/-	+/-
rs2395185	32504360	-/-	-/-	-/-	+/+	-/-	-/-
rs2187823	32510706	-/-	-/-	-/-	-/+	-/-	-/-
rs5026743	32511162	-/-	-/-	-/-	-/+	-/-	-/-
rs6901541	32513462	-/-	-/-	-/-	-/-	+/+	-/-
rs5020946	32521277	-/-	-/-	-/-	+/-	-/-	-/-
rs6932266	32532340	-/-	-/-	-/-	-/+	-/-	-/-
rs2894266	32536120	-/-	-/-	-/-	+/+	-/-	-/-
rs11754052	32563120	-/-	-/-	-/-	+/-	-/-	-/-
rs7753605	32612642	-/-	-/-	-/-	-/-	+/-	+/-
rs7451962	32633927	-/-	-/-	-/-	+/+	+/+	-/-
rs2097432	32642277	-/-	-/-	-/-	+/-	-/-	-/-
rs3129763	32642431	-/-	-/-	-/-	+/-	-/-	-/-
rs4639334	32653366	-/-	-/-	-/-	+/-	+/+	-/-
rs5875382	32653817	-/-	-/-	-/-	-/+	-/+	+/+
rs2040410	32653850	-/-	-/-	-/-	-/+	-/-	-/-
rs2187668	32657047	-/-	-/-	-/-	-/-	+/+	-/-
rs6927022	32663617	-/-	-/-	-/-	-/-	+/+	-/-
rs6906021	32672361	-/-	-/-	-/-	-/-	-/-	+/+
rs1064173	32673520	-/-	-/-	-/-	-/-	-/-	+/-
rs1063355	32673754	-/-	-/-	-/-	-/+	-/+	-/+
rs2854276	32674218	-/-	-/-	-/-	-/-	-/-	+/-
rs3134996	32683288	-/-	-/-	-/-	-/-	-/-	+/+
rs4947342	32699490	-/-	-/-	-/-	-/-	-/-	+/-
rs5875390	32703793	-/-	-/-	-/-	+/-	-/-	-/-
rs6457617	32710272	-/-	-/-	-/-	-/-	-/-	+/-
rs2647012	32710879	-/-	-/-	-/-	-/-	-/-	-/+
rs2647045	32714501	-/-	-/-	-/-	-/-	-/-	-/+
rs7764856	32727018	-/-	-/-	-/-	+/-	+/+	-/-
rs3998159	32728398	-/-	-/-	-/-	-/-	-/+	-/-
rs3997854	32729294	-/-	-/-	-/-	-/-	-/-	-/+
Mean prediction accuracy (%)		98	98	96	91	99	99

^aPositive and negative signs indicate whether a SNP is used to predict alleles at a given HLA locus in CEU and YRI (CEU/YRI).