Supplemental Data

A Genome-wide Analysis of Admixture in Uyghurs and a High-Density Admixture Map for Disease-Gene Discovery

Shuhua Xu and Li Jin

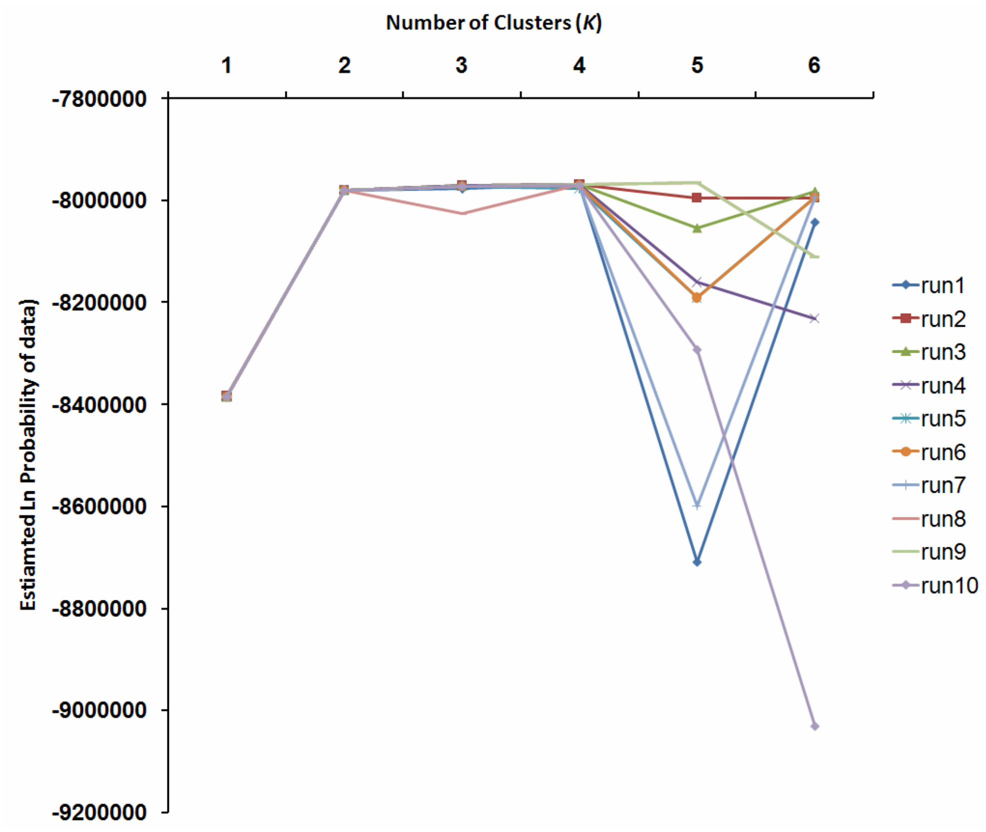

Figure S1 Probability Estimations for the Number of Clusters, with Ten Repeats for Each *K*

The ordinate shows the Ln probability corresponding to the number of clusters (*K*) shown on the abscissa.

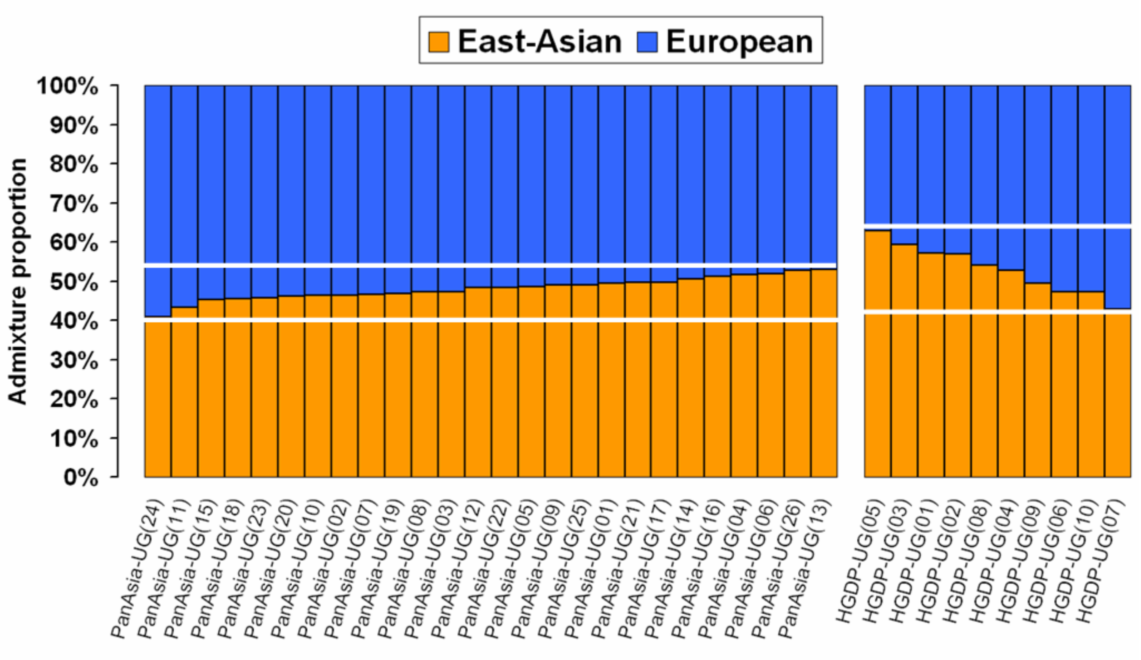

Figure S2 Summary Plot of Individual Admixture Proportions

The results of individual admixture proportions estimated from 19,934 SNPs. Each individual is represented by a single vertical line broken into 2 colored segments, with lengths proportional to each of the 2 inferred clusters, orange indicates East-Asian ancestry proportion, and blue indicates European ancestry proportion. The ordinate indicates the proportion unit. The invividual IDs are presented at the bottom of the plot.

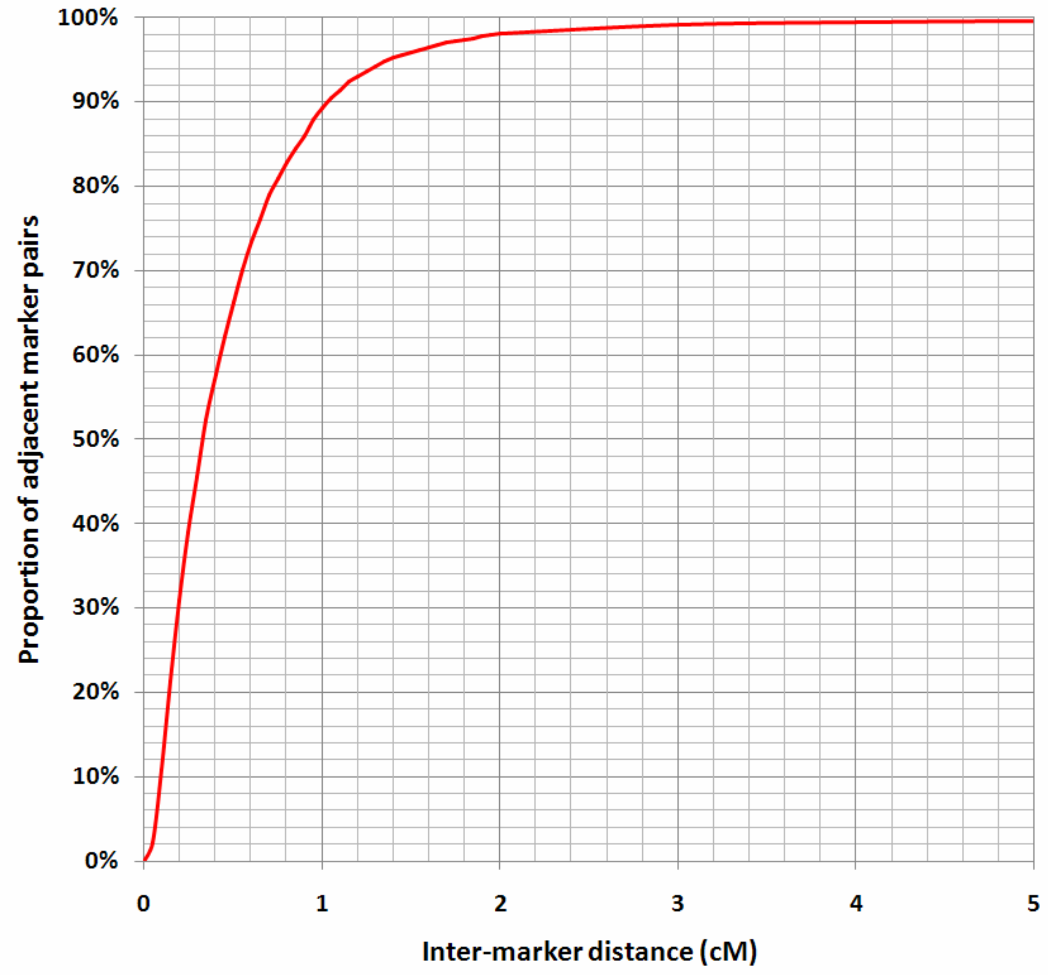

Figure S3 The Distribution of Intermarker Distance

The average intermarker distance is 0.50 cM and the median is 0.33 cM, with 66.3% of adjacent marker pairs having intermarker distance less than 0.5 cM, 89.3% less than 1 cM, and 98% less than 2 cM.

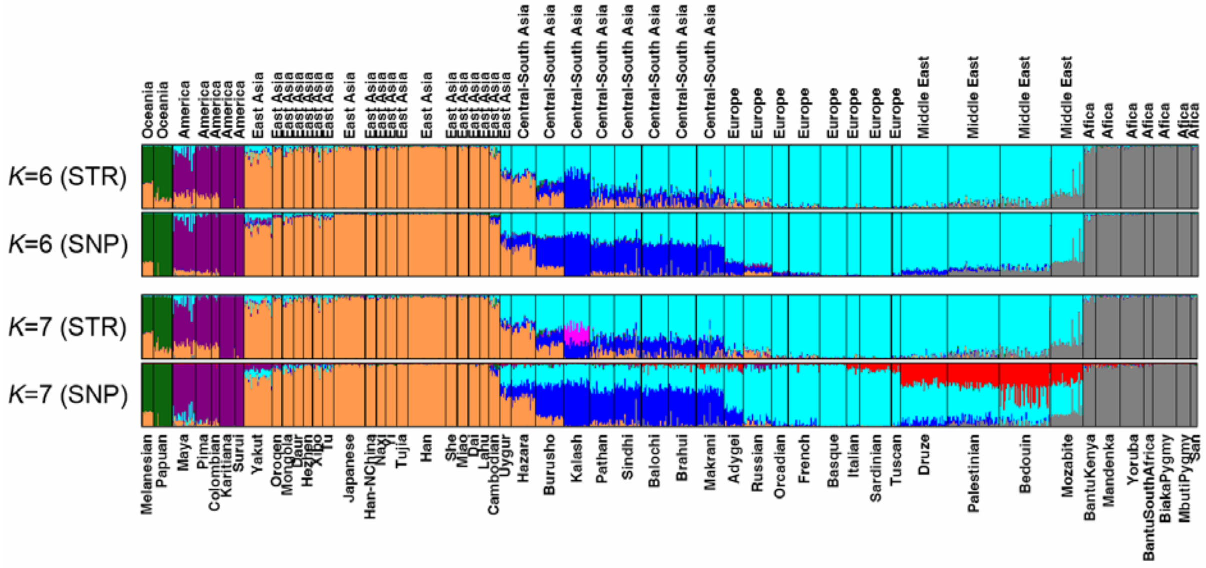

Figure S4 Summary Plot of Individual Admixture Proportions

STRUCTURE analysis was performed with STR and SNP data separately in 940 individuals. The results of K=6 and K=7 are displayed, which match that of Li et al.[1]

Table S1. Summary of SNP-Screen Results for Excluded SNPs with F_{ST} > 0.05 between EAS Populations

	Mongola	JPT	Japanese	Han	CHB
Mongola					
JPT	14475				
Japanese	14000	2049			
Han	13245	4627	4995		
CHB	12426	3036	3957	1752	
Han-NChina	11499	10276	10430	8163	7757

Table S2. Summary of SNP-Screen Results for Excluded SNPs with F_{ST} > 0.05 between EUR Populations

	Basque	Sardinian	Italian	Tuscan	French	Orcadian	CEU	Adygei
Basque								
Sardinian	8682							
Italian	9087	8391						
Tuscan	11157	10786	9682					
French	6666	7022	6426	9120				
Orcadian	10438	11060	9866	11068	6678			
CEU	5909	7054	6460	8926	2375	5596		
Adygei	13189	13629	10668	11769	9911	1970	9090	
Russian	10912	12513	9315	11531	6798	8683	11158	6667

Table S3. Pairwise F_{ST} between Populations

	JPT	Japanese	Mongola	CHB	Han N.China	Han	PanAsia UG	HGDP UG	Basque	Sardinian	Italian	Tuscan	French	Orcadian	CEU	Adygei
Japanese	0.0003															
Mongola	0.0083	0.0075														
CHB	0.0069	0.0063	0.0050													
Han-N.China	0.0070	0.0066	0.0030	0.0007												
Han	0.0081	0.0078	0.0075	0.0008	0.0023											
CN-UG	0.0392	0.0385	0.0240	0.0368	0.0302	0.0383										
Uygur	0.0351	0.0340	0.0194	0.0322	0.0266	0.0340	0.0009									
Basque	0.1207	0.1201	0.0997	0.1193	0.1082	0.1190	0.0395	0.0428								
Sardinian	0.1232	0.1228	0.1035	0.1208	0.1119	0.1216	0.0423	0.0466	0.0134							
Italian	0.1149	0.1147	0.0953	0.1133	0.1041	0.1142	0.0309	0.0358	0.0076	0.0069						
Tuscan	0.1135	0.1156	0.0941	0.1119	0.1046	0.1148	0.0281	0.0334	0.0093	0.0073	0.0011					
French	0.1108	0.1091	0.0900	0.1083	0.0994	0.1086	0.0292	0.0338	0.0070	0.0089	0.0015	0.0024				
Orcadian	0.1175	0.1162	0.0954	0.1137	0.1047	0.1149	0.0328	0.0372	0.0129	0.0170	0.0081	0.0092	0.0051			
CEU	0.1093	0.1075	0.0888	0.1076	0.0977	0.1078	0.0297	0.0338	0.0090	0.0126	0.0037	0.0042	0.0012	0.0048		
Adygei	0.1003	0.0994	0.0790	0.0992	0.0880	0.0991	0.0215	0.0260	0.0181	0.0180	0.0082	0.0070	0.0091	0.0144	0.0107	
Russian	0.0985	0.0976	0.0769	0.0958	0.0868	0.0969	0.0229	0.0271	0.0143	0.0200	0.0088	0.0090	0.0057	0.0085	0.0050	0.0116

Table S4. Statistics of Principle-Component Analysis in 17 Population Samples with the Use of 19,934 SNPs

Principle Component	Eigenvalue	Percent Eigenvalue[a]	TW statistic	TW p value
PC1	36.97	71.6%	247.096	0
PC2	2.32	4.5%	187.364	0
PC3	1.82	3.5%	98.895	8.09×10^{-287}
PC4	1.74	3.4%	87.799	2.95×10^{-240}
PC5	1.62	3.1%	65.712	2.93×10^{-156}
PC6	1.61	3.1%	65.005	8.97×10^{-154}
PC7	1.51	2.9%	44.966	2.72×10^{-89}
PC8	1.37	2.7%	14.804	2.32×10^{-18}
PC9	1.34	2.6%	7.041	1.22 × 10-7
PC10	1.33	2.6%	6.060	1.67 × 10-6
PC11	1.29	-	-1.900	0.697

[a] The percent Eigenvalue is the percentage of the total variance in the first ten PCs.

Table S5 Information of 7999 Validated AIMs

Chr.	HapMap SNPs	AIMs I	AIMs II	AIMs III	Start Position	End Position	Total Length
1	302715	20531	4675	592	0.02	300.16	300.1
2	321609	25745	5858	650	0.22	274.13	273.9
3	248946	19273	4482	557	0.00	244.81	244.8
4	238779	15235	3390	516	1.08	232.81	231.7
5	242843	17171	3945	510	0.07	217.12	217.1
6	264747	17092	3733	438	0.00	222.19	222.2
7	207602	14526	3378	482	0.01	197.63	197.6
8	208948	14448	3440	369	0.01	181.33	181.3
9	176904	12719	3061	369	0.00	198.40	198.4
10	205964	14730	3338	384	0.00	183.25	183.2
11	201066	12470	2959	350	0.04	187.41	187.4
12	185475	13423	3080	351	0.00	184.02	184.0
13	153534	9971	2189	238	0.00	138.75	138.7
14	120960	8563	2000	230	0.03	129.94	129.9
15	105076	9033	2292	252	0.01	142.51	142.5
16	107586	6808	1834	233	0.01	134.81	134.8
17	87910	6090	1593	238	0.01	159.42	159.4
18	116573	6121	1539	214	0.00	131.13	131.1
19	55273	3713	1039	176	0.00	117.38	117.4
20	117528	6440	1538	217	0.84	109.45	108.6
21	49188	2993	769	135	0.16	68.11	67.9
22	53543	3346	974	164	0.56	85.03	84.5
X	115535	11466	2438	334	0.04	187.71	187.7
Total	3888304	271907	63544	7999			4024.4

AIMs I: screened from HapMap data, $F_{ST} > 0.25$ between CEU and JPT+CHB.

AIMs II: HapMap AIMs genotyped in HGDP-CEPH panel.

AIMs III: final set of AIMs that validated in HGDP-CEPH panel, with the filtering of intermarker distance.

Table S6. Information of 2750 AIMs for Studying PanAsia-UG Only

Chr.	AIMs	Start Position	End Position	Total Length
1	238	8.8	287.0	278.2
2	268	3.3	257.6	254.3
3	212	2.7	243.5	240.8
4	183	14.7	230.5	215.8
5	205	7.2	216.0	208.8
6	192	0.0	213.7	213.7
7	150	4.1	196.8	192.7
8	135	0.0	173.9	173.9
9	117	0.0	177.7	177.7
10	133	2.4	177.8	175.4
11	96	12.6	186.0	173.4
12	122	2.7	183.6	180.9
13	99	1.5	138.8	137.3
14	77	8.0	124.5	116.5
15	87	0.3	142.5	142.2
16	67	15.2	128.8	113.6
17	53	9.0	157.8	148.8
18	86	0.0	125.6	125.6
19	22	0.0	104.2	104.2
20	53	5.9	109.4	103.5
21	45	1.2	61.1	59.8
22	15	21.6	74.5	52.9
X	95	0.0	180.5	180.5
Total	2750			3770.5

Supplemental Reference

1. Li, J.Z., Absher, D.M., Tang, H., Southwick, A.M., Casto, A.M., Ramachandran, S., Cann, H.M., Barsh, G.S., Feldman, M., Cavalli-Sforza, L.L., et al. (2008). Worldwide human relationships inferred from genome-wide patterns of variation. Science *319*, 1100–1104.