Supplemental Data

Figure S1

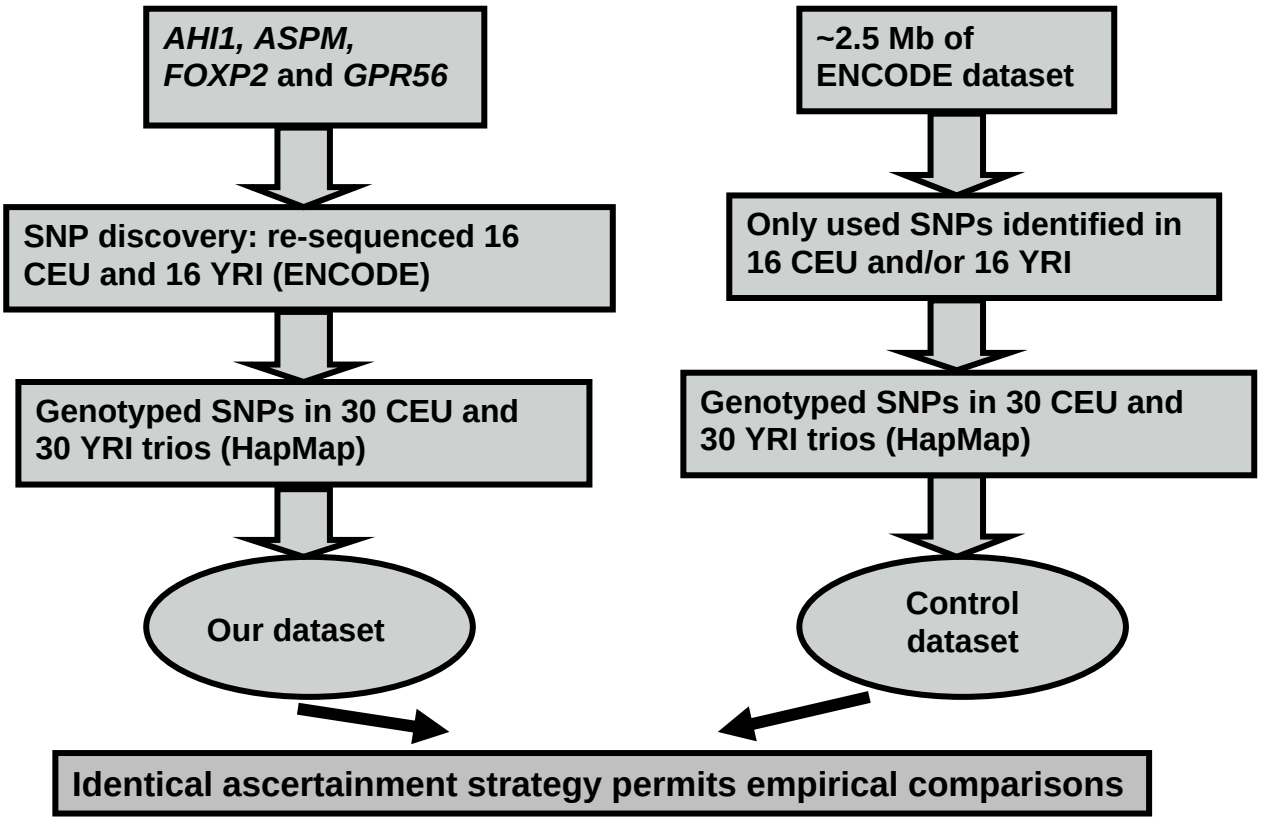


Figure S1: Strategy used to match the ascertainment between SNPs collected from the re-sequencing and genotyping of the genic regions in this study, and SNPs from the ~2.5 Mb of ENCODE data. In our study, we newly discovered and genotyped SNPs in the regions of interest, and then re-curated the ENCODE data so that the SNP ascertainment from these two sources are maximally comparable. This permits us to use ENCODE as control data to carry out various empirical statistical tests.

Figure S2

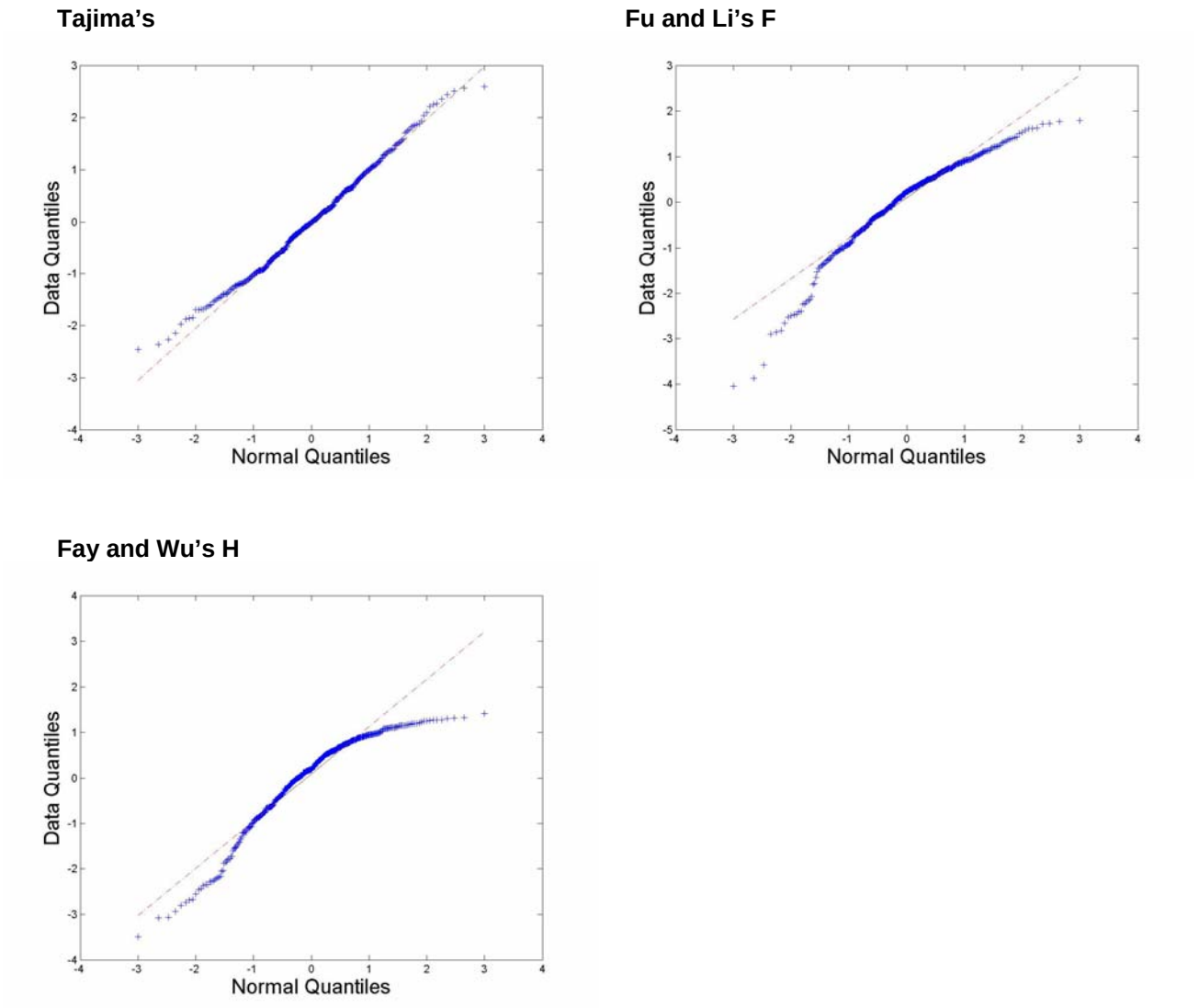
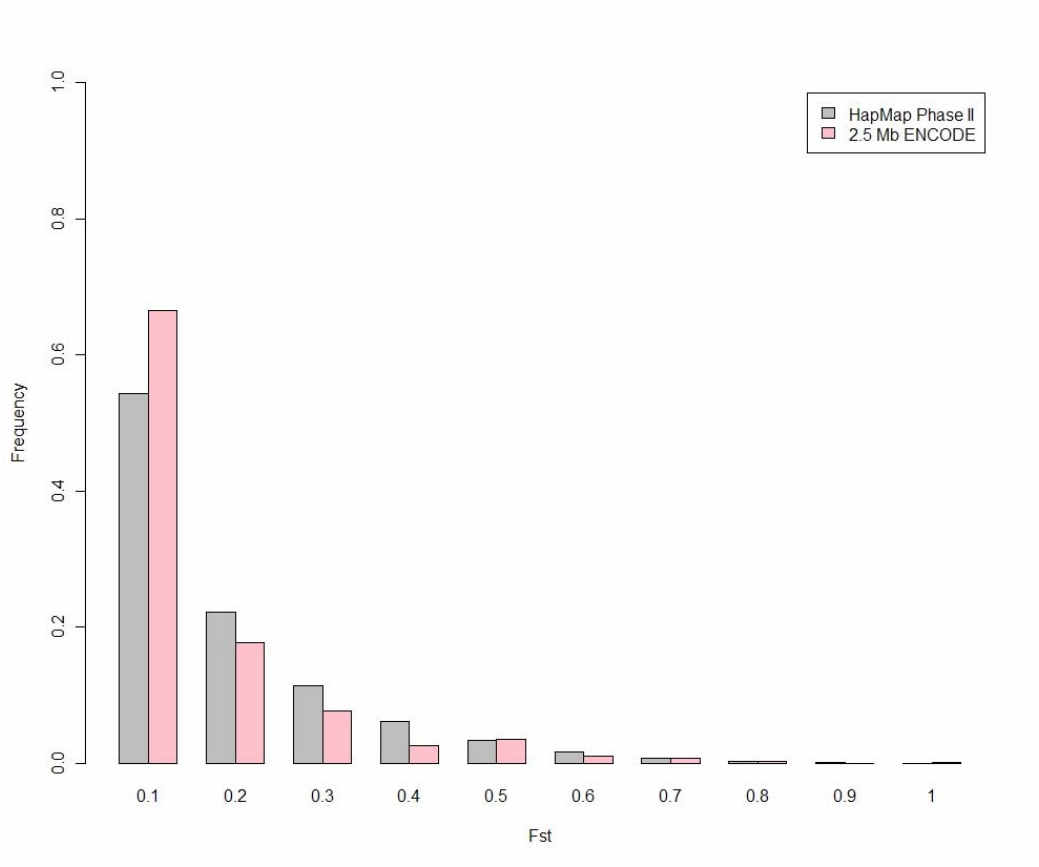


Figure S2: Quantile-quantile plots comparing the three allele frequency spectrum selection statistics considered in this study with a normal distribution. The data quantiles are determined by a leave-one-out cross-validation procedure, in which we left out each window in the ENCODE control data set in turn, and calculated its number of standard deviations from the mean based on all other windows in the control data (Materials and Methods). Tajima's D closely follows a normal distribution (P = 0.61 for rejection of normality by a χ^2 goodness-of-fit test). The other two statistics violate a normal distribution ($P=3.9\times10^{-7}$ and $P=1.6\times10^{-13}$ for Fu and Li's F and Fay and Wu's H). Positive values of these two statistics do not deviate as many standard deviations from the mean as expected from a normal distribution, suggesting that a normality assumption is conservative for positive values of these statistics.

Figure S3

a



b

	HapMap Phase II	2.5 Mb ENCODE
Median	0.088	0.057
Mean	0.13	0.10
75th percentile	0.19	0.13
90th percentile	0.33	0.26
99th percentile	0.60	0.61
99.9th percentile	0.78	0.73

Figure S3: Analysis of the F_{ST} distributions of the HapMap Phase II data and the ENCODE SNPs. (a) The F_{ST} distribution of SNPs in HapMap Phase II (grey) is right-skewed relative to the ENCODE data (pink). (b) Numerical comparison of F_{ST} distributions between HapMap and ENCODE.

Figure S4

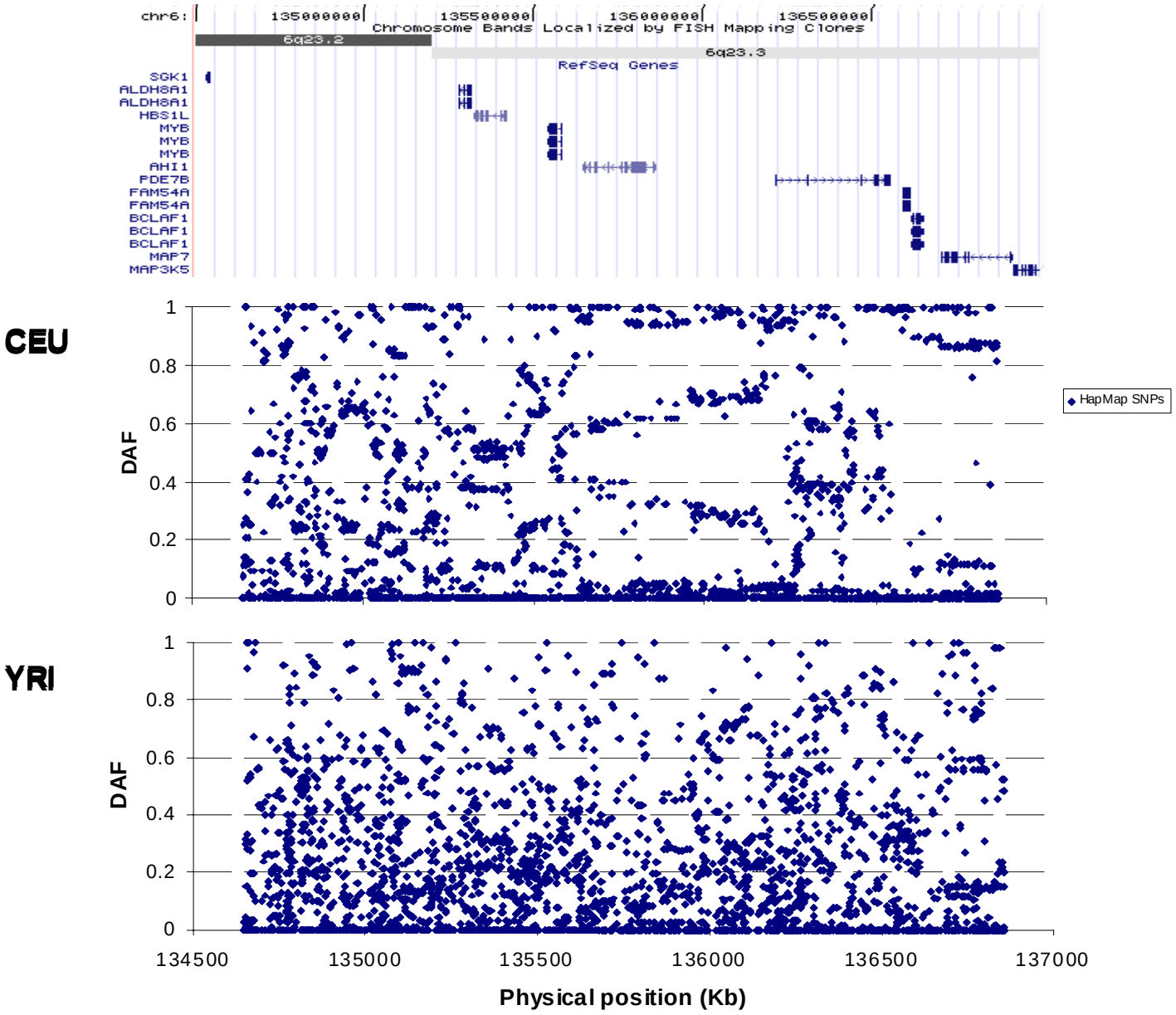


Figure S4: Derived allele frequencies of HapMap SNPs within the 2 Mb region centered near *AHI1* are plotted by their physical positions from NCBI Build 36 (the two bottom panels). Additional genes located in this ~2Mb span are shown in the top panel, using a screenshot from the UCSC genome browser (http://genome.ucsc.edu).

Table S1. SNPs with strong LRH scores (REHH values exceeding the 99th percentile)

CEU

Genic region	Core SNP	Physical position (hg16)	Core allele	Core SNP allele frequency	EHH	REHH	ln(REHH) deviation	P-value
AHI1	rs6908428	135774276	A	0.62	0.10	8.02	2.48	0.0066
AHI1	rs12206850	135778378	T	0.76	0.07	6.97	2.68	0.0037
AHI1	rs9373131	135780494	T	0.70	0.08	10.16	2.83	0.0023
AHI1	rs4526212	135785201	C	0.76	0.07	6.99	2.68	0.0036
GPR56	rs3848270	57446991	T	0.55	0.13	13.25	2.78	0.0028
GPR56	rs3916059	57447303	G	0.58	0.12	11.66	2.77	0.0028
GPR56	rs6499906	57447754	T	0.56	0.13	13.38	2.92	0.0018
GPR56	rs3760061	57448857	A	0.43	0.20	14.46	2.70	0.0035
GPR56	rs1975630	57449159	G	0.51	0.16	18.08	3.10	0.0010
GPR56	rs935740	57451818	G	0.52	0.16	17.42	3.06	0.0011
GPR56	rs935742	57451841	C	0.39	0.24	11.33	2.33	0.0099
GPR56	Broad10422842	57461700	T	0.63	0.10	8.48	2.54	0.0056

YRI

Genic region	Core SNP	Physical position (hg16)	Core allele	Core SNP allele frequency	EHH	REHH	ln(REHH) deviation	P-value
ASPM	rs10733087*	194344041	T	0.73	0.07	11.75	3.94	4E-05
ASPM	rs10754213*	194344184	T	0.73	0.07	11.75	3.94	4E-05
ASPM	Broad10420993*	194346295	A	0.78	0.06	11.20	4.07	2E-05
ASPM	rs10754214*	194348151	A	0.78	0.07	11.69	4.13	2E-05
ASPM	Broad10421115*	194348187	A	0.78	0.07	11.69	4.13	2E-05
ASPM	rs7542397	194384979	G	0.75	0.06	3.41	2.42	0.0077
ASPM	rs877897	194392234	G	0.73	0.07	4.07	2.64	0.0041
ASPM	Broad10422499	194399081	A	0.68	0.09	4.00	2.40	0.0082
ASPM	Broad10422568	194400792	G	0.86	0.06	4.09	3.15	0.0008
ASPM	Broad10422590	194401332	G	0.86	0.06	4.09	3.15	0.0008
ASPM	Broad10422625	194401832	C	0.68	0.09	4.32	2.49	0.0063
FOXP2	rs11764740	113724755	A	0.83	0.06	4.21	3.06	0.0011
FOXP2	rs17137091	113744289	G	0.48	0.15	8.78	2.55	0.0054
FOXP2	rs10953762	113744976	G	0.49	0.15	8.76	2.54	0.0055
FOXP2	rs10953763	113745009	T	0.49	0.15	8.76	2.54	0.0055
FOXP2	rs7801324	113745232	G	0.49	0.15	8.76	2.54	0.0055
FOXP2	rs11760306*	113762137	C	0.79	0.06	18.88	4.71	1E-06
FOXP2	rs7788346	113767746	T	0.40	0.21	10.50	2.34	0.0095
FOXP2	rs7780785	113771252	C	0.41	0.20	9.75	2.48	0.0066
FOXP2	rs10279820*	113789585	A	0.78	0.07	11.73	4.13	2E-05
FOXP2	rs10085693	113791012	C	0.74	0.07	4.61	2.80	0.0026
FOXP2	rs10085404	113791060	A	0.77	0.07	3.51	2.66	0.0039
FOXP2	rs6954257	113794929	A	0.78	0.07	3.04	2.48	0.0065
FOXP2	rs6953053	113800150	C	0.78	0.06	2.88	2.42	0.0078
FOXP2	rs6953086	113800237	A	0.73	0.07	5.00	2.90	0.0019
FOXP2	rs1456024	113817400	A	0.75	0.07	3.73	2.54	0.0056
FOXP2	rs7795372*	113844716	A	0.76	0.07	14.18	4.36	6E-06
FOXP2	rs11765385	113848362	G	0.90	0.05	3.06	2.81	0.0025
FOXP2	rs936146	113848908	C	0.87	0.06	3.38	2.93	0.0017
GPR56	rs1466132	57445339	C	0.88	0.05	4.60	3.28	0.0005
GPR56	rs3848270	57446991	A	0.78	0.07	3.53	2.67	0.0038
GPR56	rs3916059	57447303	A	0.79	0.06	3.19	2.54	0.0055
GPR56	rs3760061	57448857	G	0.82	0.06	3.32	2.77	0.0028

*****SNP markers with REHH values that exceeded 99.9th percentile when compared to the empirical distribution obtained from all SNPs in HapMap Phase II.

Note: After log transformation of REHH, the ln(REHH) for SNPs in HapMap Phase II is approximately normally distributed(2). We compared ln(REHH) for each tested SNP listed in the table to the ln(REHH) distribution in the respective allele frequency bin, and calculated the number of standard deviation from the empirical mean (i.e. ln(REHH) deviation) as a metric of statistical significance. A given region is considered as a significant candidate region only when at least 10% of the SNPs within the region have "ln(REHH) deviation" above 3.92 (~99.9th percentile). None of the four genes crossed the threshold, although the SNPs with strong P-values are intriguing.