Supplementary Material

Genome-wide comparisons of variation in linkage disequilibrium

Teo et al.

Population pairs compared

We performed a total of 6 sets of comparisons, each between two datasets. The 6 sets of analyses compare:

1. 60 unrelated parents from HapMap CEU against 90 unrelated individuals from HapMap CHB+JPT, using unphased genotype data from the HapMap;
2. 60 unrelated parents from HapMap CEU against 60 unrelated parents from HapMap YRI, using unphased genotype data from the HapMap;
3. 60 unrelated parents from HapMap YRI against 90 unrelated individuals from HapMap CHB+JPT, using unphased genotype data from the HapMap;
4. 60 Jola samples from The Gambia against 60 unrelated parents from HapMap YRI, using unphased genotype data from MalariaGEN and the HapMap respectively;
5. 60 unrelated British samples from the 1958 Birth Cohort against 60 unrelated parents from HapMap CEU, using unphased genotype data from the WTCCC and the HapMap respectively;
6. two simulated datasets of 60 samples each resampled from the 60 unrelated parents in HapMap CEU, meant to investigate an empirical null distribution for varLD scores.

Comparison with imputation diagnostics

Imputation of a target population from The Gambia was performed with the program IMPUTE, using the HapMap YRI panel as reference. The data used belong to the control dataset of a case-control genome-wide study in malaria, and imputation was performed on all 1,382 control individuals. To avoid SNPs with genotyping errors, only SNPs from the Gambian data with minor allele frequencies > 1%, < 5% missingness and HWE $P \geq 10^{-7}$ have been used for imputation. The default buffer region of 250kb on each end of the imputed region was used, and the effective population size N_e was set at 17,469 (a value recommended by the author of IMPUTE for HapMap YRI). All SNPs were mapped to the forward strand prior to imputation and we only considered autosomal chromosomes. We consider a composite metric of imputation diagnostic which is obtained as the product of the call rate and concordance of imputed genotypes. Each imputed call is assigned to the genotype with posterior probability > 0.9, or is otherwise classified as missing. Concordance at each imputed SNP is calculated as the proportion of imputed genotypes that agree with the observed genotypes. As such, this metric is only calculated for SNPs that exist on the Affymetrix platform and have passed our criteria on missingness and minor allele frequencies.

Simulations for sensitivity analysis

We perform a series of simulations in order to assess the sensitivity of varLD at identifying differences in LD. We also investigate the effect of genotyping errors on the varLD signals. Artificial data with differences in patterns of LD is simulated using the program HAPGEN, available online with documentation at:
http://www.stats.ox.ac.uk/~marchini/software/gwas/hapgen.html.

Genotype data spanning a 1Mb region on chromosome 19 was simulated by conditioning on the 120 chromosomes belonging to the 60 unrelated parents from the HapMap CEU panel, and the estimate of the fine-scale recombination rate across this region. Although HAPGEN simulates case-control datasets, we only simulated control data under the setting that parameters for relative risks were all set to 1. While the 1Mb region spans 329 SNPs, our sensitivity analyses only consider a window of 50 SNPs in the center of the region. LD differences across two populations were simulated through the use of different recombination rates in the two populations, while genotyping error is introduced by replacing the simulated genotypes by random assignment conditioned on the allele frequency under the assumption of Hardy-Weinberg equilibrium. We consider a total of 7 scenarios in our sensitivity analysis, each performed over 1000 independent iterations:

(i) 2 populations of 60 individuals each, with identical fine-scale recombination rates corresponding to that reported by the HapMap (see **Supplementary Fig. 6**);

(ii) 2 populations of 60 individuals each, with identical fine-scale recombination rates corresponding to that reported by the HapMap and incorporated genotyping error in the second population at the 1st SNP in the window of 50 SNPs;

(iii) 2 populations of 60 individuals each, with identical fine-scale recombination rates corresponding to that reported by the HapMap and incorporated genotyping error in the second population at the 20th SNP in the window of 50 SNPs;

(iv) 2 populations of 60 individuals each, with identical fine-scale recombination rates corresponding to that reported by the HapMap and incorporated genotyping error in the second population at five chosen SNPs in the window of 50 SNPs. The positions of these five SNPs were arbitrarily determined in the first iteration, but kept fixed subsequently throughout the rest of the iterations (corresponding to the 3rd, 17th, 23rd, 38th, 40th SNP);

(v) 2 populations of 60 individuals each, with identical fine-scale recombination rates corresponding to that reported by the HapMap and incorporated genotyping error in the second population at five consecutive SNPs in the beginning of the window of 50 SNPs;

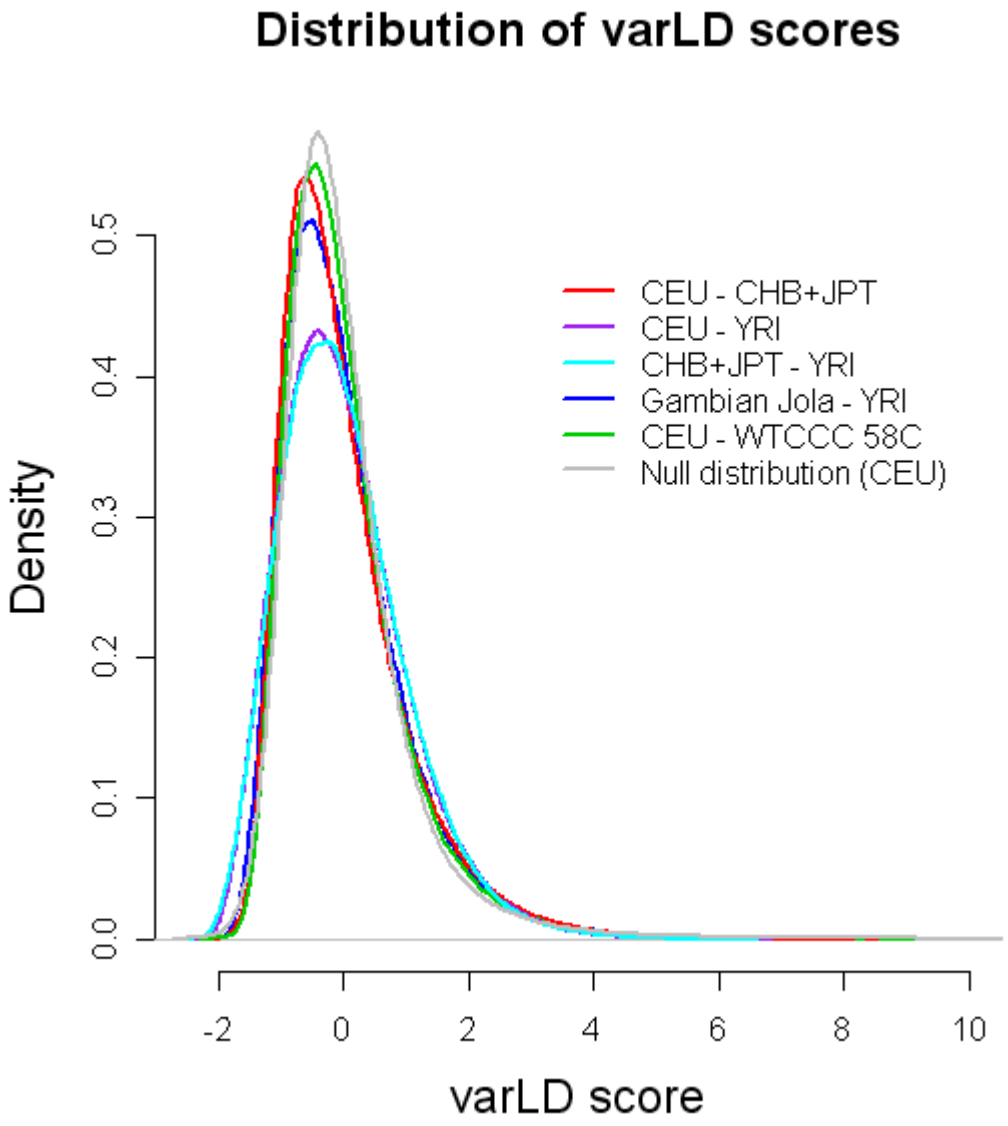
(vi) 2 populations of 60 individuals each. Fine-scale recombination rates corresponding to that reported by the HapMap were used for both populations except that the recombination rates for SNPs 196 to 200 in population 2 were replaced by random draws from the possible blocks of 5 recombination rates across the entire region. Thus the recombination rates between the two populations are exactly identical except between the 196th SNP and 200th SNP inclusive (see **Supplementary Fig. 6**);

(vii) 2 populations of 60 individuals each. Fine-scale recombination rates corresponding to that reported by the HapMap were used for both populations except that the recombination rates for SNPs 181 to 200 in population 2 were replaced by random draws from possible blocks of 30 recombination rates across the entire region. Thus the recombination rates between the two populations are exactly identical except between the 181th SNP and 200th SNP inclusive (see **Supplementary Fig. 6**).

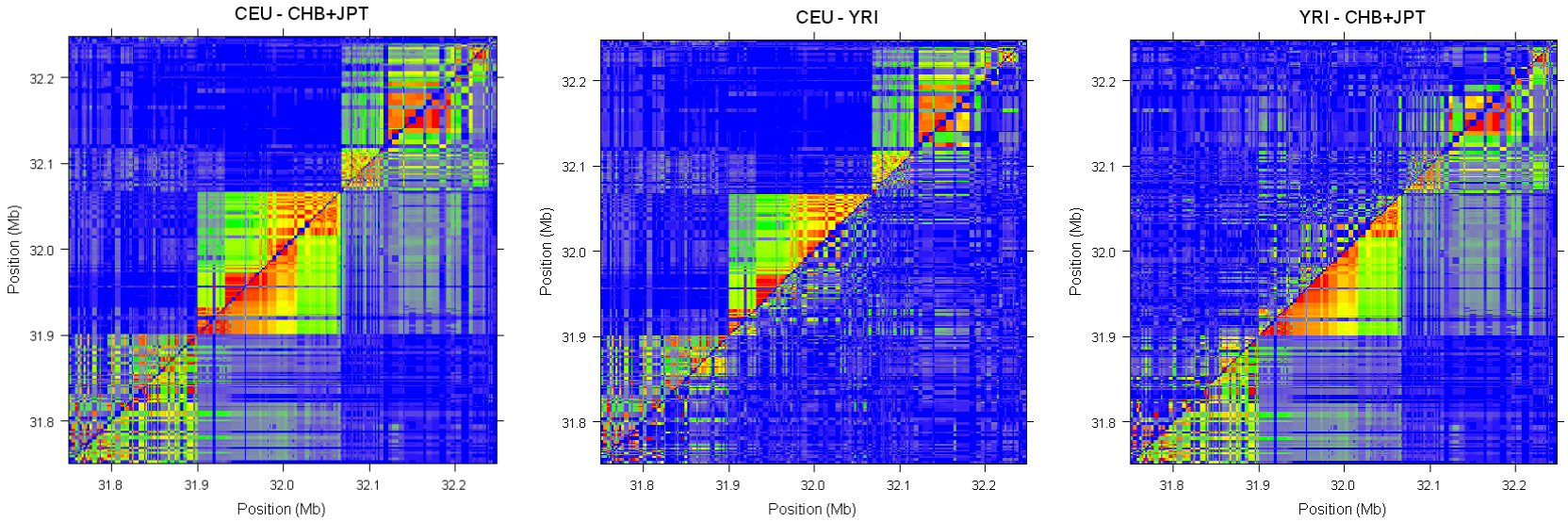
As the simulations effectively compare the CEU population against itself, we standardize the varLD signals using the mean and standard deviation obtained in our genome-wide comparison of the two

simulated CEU datasets. We calculated the proportion of simulations out of 1000 iterations where the standardized varLD score is greater than the 90^{th}, 95^{th} and 99^{th} quantiles of the genome-wide distribution of standardized varLD scores obtained in the comparisons of the two simulated CEU datasets. This is defined as the positive rate, and the false positive rate can be obtained under scenario (i) with two populations of 60 individuals simulated while conditioning on identical fine-scale recombination rates. The positive rates obtained for scenarios (ii) – (v) represent the sensitivity of varLD to different extent of genotyping errors. Scenarios (vi) and (vii) are meant to represent varying extent of LD differences between two populations.

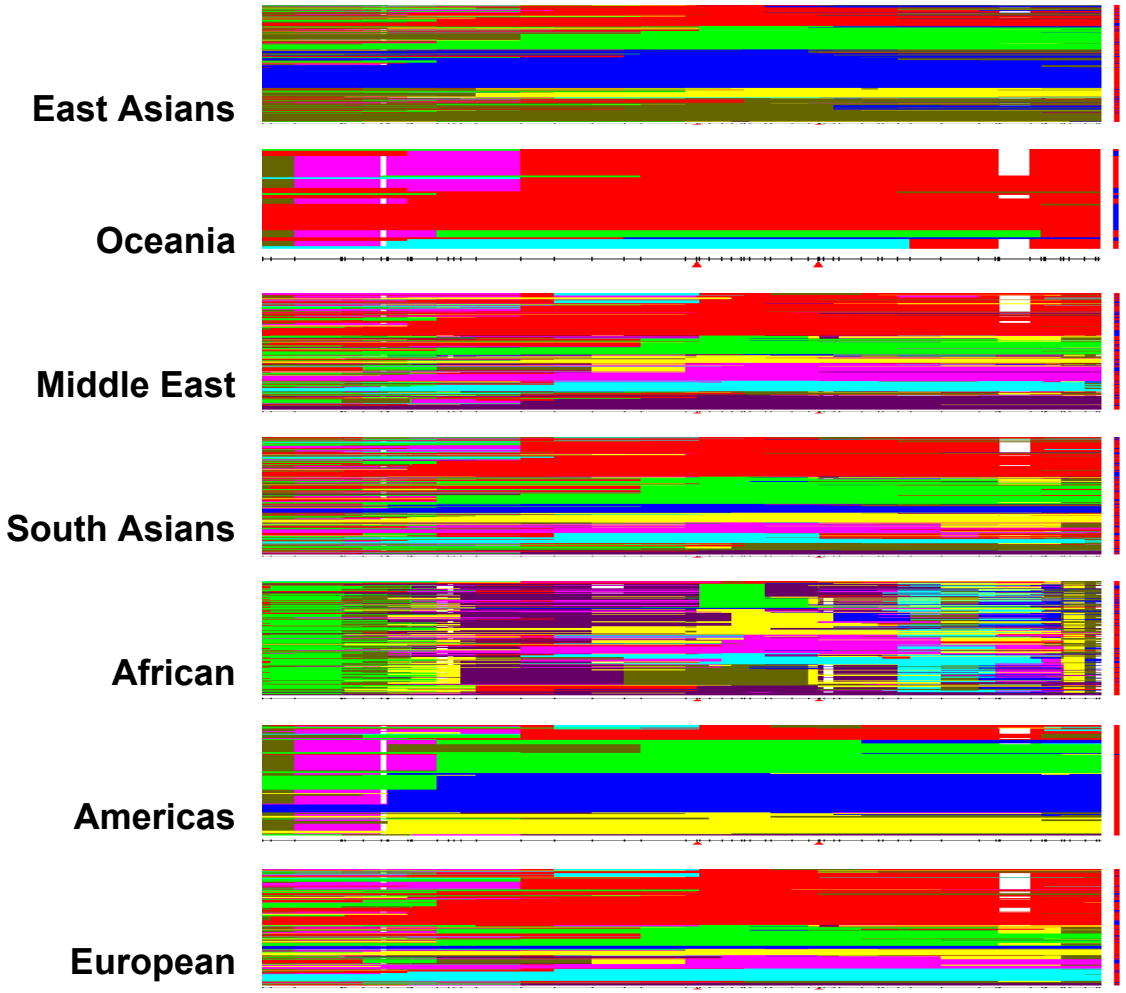
Supplementary Figure 1. Density plot of the distribution of standardized varLD scores in each of the six population-pair comparisons.



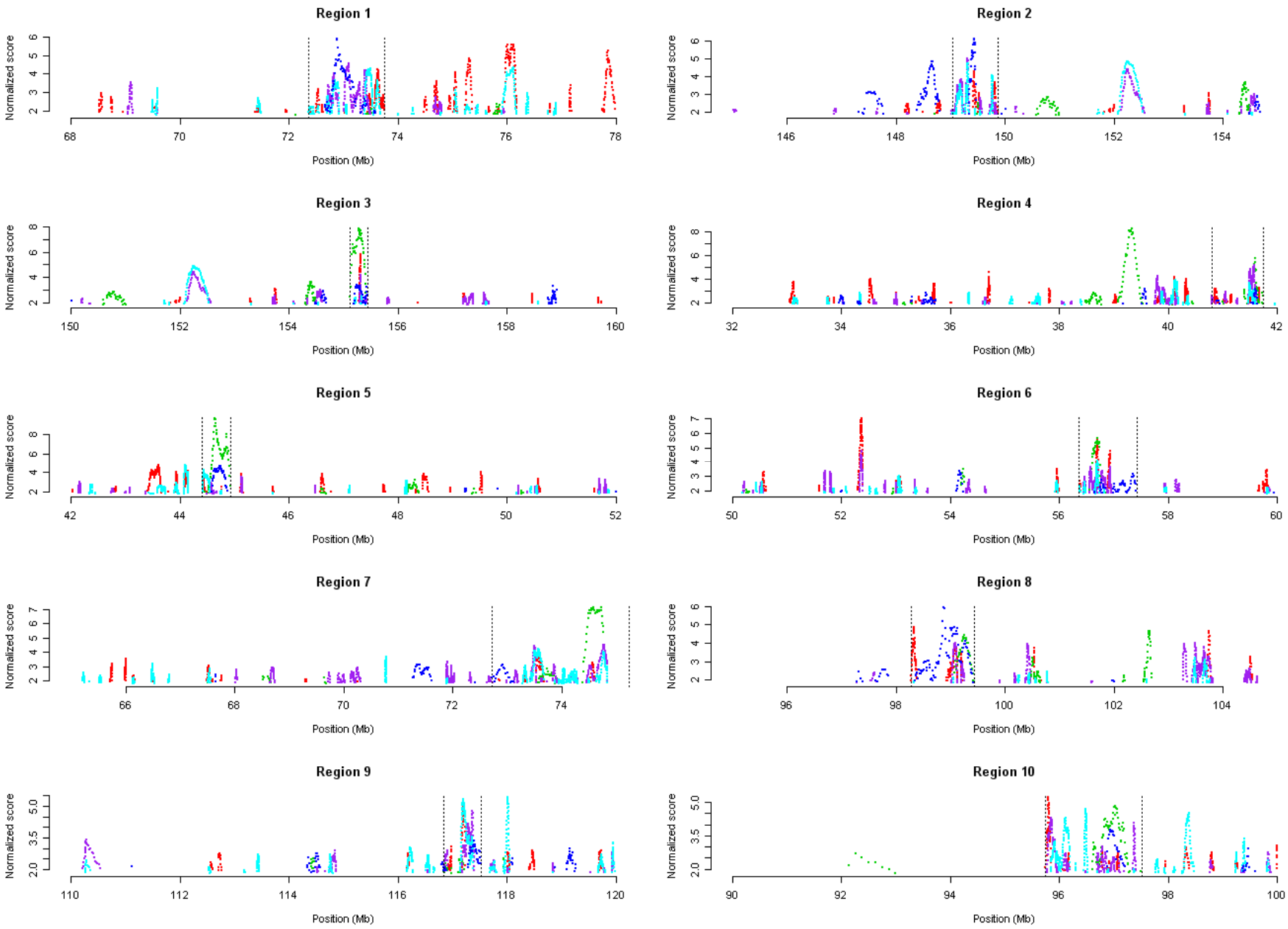
Supplementary Figure 2. Heatmap representations of LD at the *NRG1* gene on chromosome 8 between pairs of populations in HapMap. The upper left and lower right triangles of each plot correspond to the LD in a region for each of two populations respectively as measured by the pairwise r^2 metric, with the left panel comparing HapMap Europeans with HapMap Asians, the middle panel comparing HapMap Europeans with HapMap Africans, and the right panel comparing HapMap Africans with HapMap Asians.

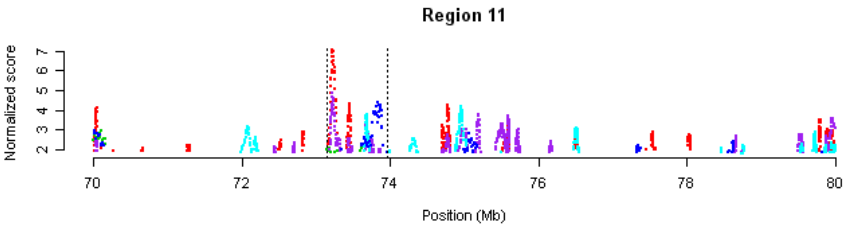


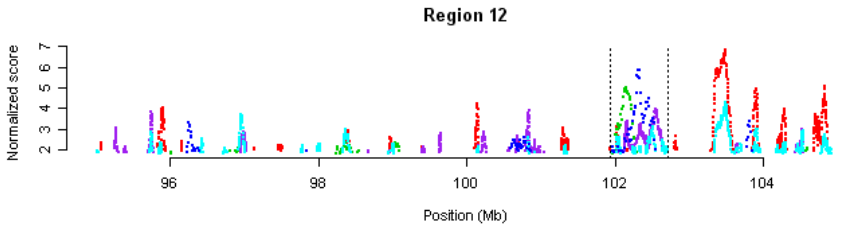
Supplementary Figure 3. An illustration of the haplotype diversity at the *NRG1* gene across the samples in the Human Genome Diversity Project (HGDP), categorized according to continents. Each colour refers to a template haplotype that is most common in a specific continent, and each chromosome in the HGDP is mapped as a mosaic of the seven template haplotypes. The figure is generated using the online genome browser by the Pritchard Lab, available at http://hgdp.uchicago.edu/cgi-bin/gbrowse/HGDP.

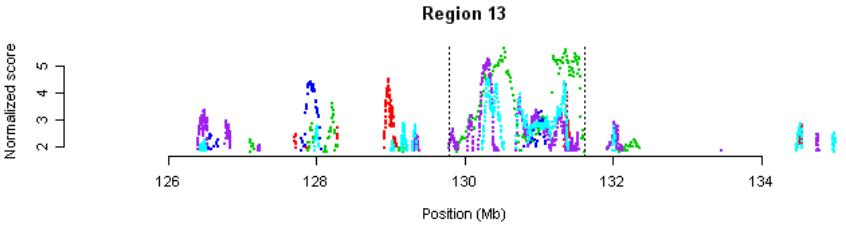


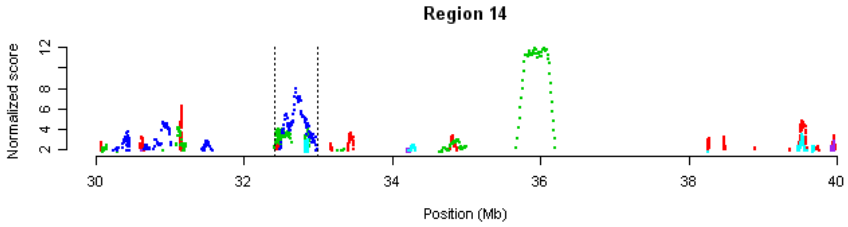
Supplementary Figure 4. Standardized varLD scores in the top 20 candidate regions identified with significant LD differences between all five sets of population comparisons as reported in Table 1 in the main text. Only scores above their respective 95th quantiles are shown, and points in red correspond to LD comparisons between HapMap Europeans (CEU) and HapMap East Asians (CHB+ JPT); points in purple between CEU and HapMap Africans (YRI); points in cyan between CHB+JPT and YRI; points in green between two European populations (HapMap CEU and WTCCC 58C); points in blue between two African populations (HapMap YRI and Gambian Jola). Dotted lines in each panel designate the start and end positions of each identified region.

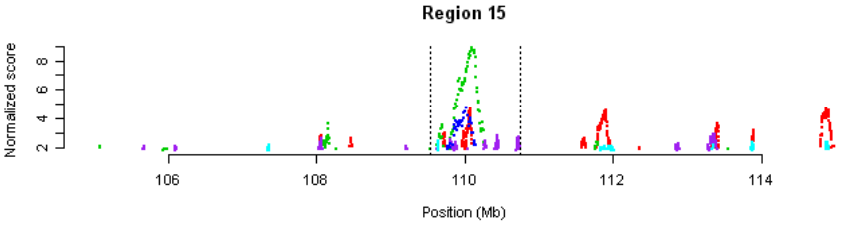

Region 1
Normalized score
Position (Mb)
Region 2
Normalized score
Position (Mb)
Region 3
Normalized score
Position (Mb)
Region 4
Normalized score
Position (Mb)
Region 5
Normalized score
Position (Mb)
Region 6
Normalized score
Position (Mb)
Region 7
Normalized score
Position (Mb)
Region 8
Normalized score
Position (Mb)
Region 9
Normalized score
Position (Mb)
Region 10
Normalized score
Position (Mb)

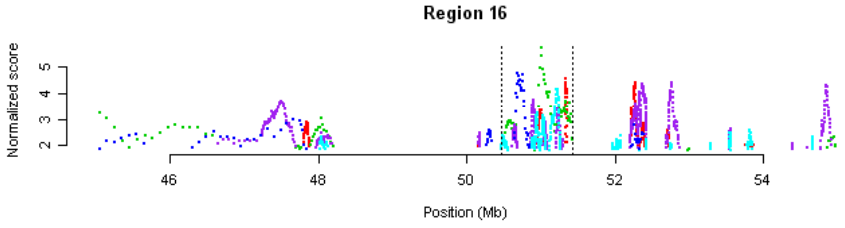

Region 11
Normalized score
2 3 4 5 6 7
70 72 74 76 78 80
Position (Mb)

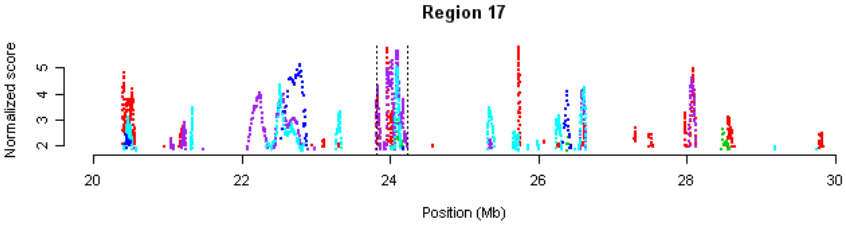

Region 12
Normalized score
2 3 4 5 6 7
96 98 100 102 104
Position (Mb)

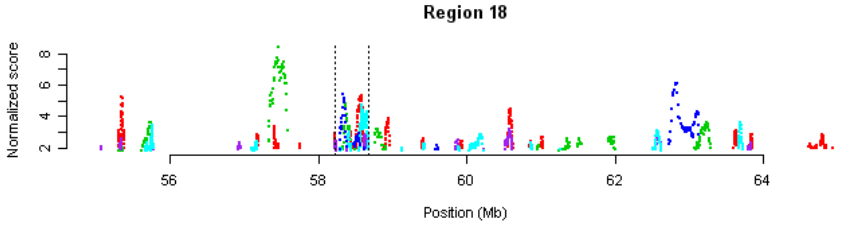

Region 13
Normalized score
2 3 4 5
126 128 130 132 134
Position (Mb)

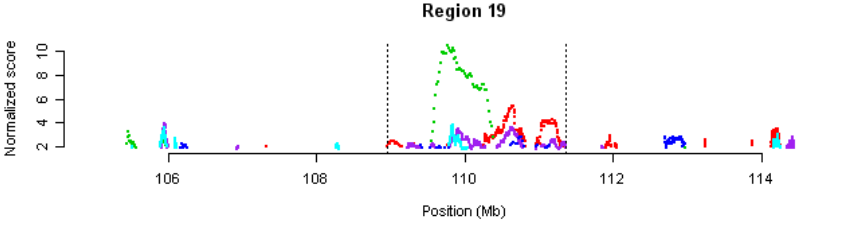

Region 14
Normalized score
2 4 6 8 12
30 32 34 36 38 40
Position (Mb)

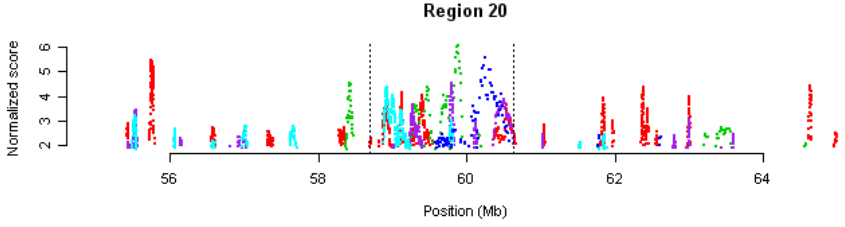

Region 15
Normalized score
2 4 6 8
106 108 110 112 114
Position (Mb)


Region 16
Normalized score
2 3 4 5
46 48 50 52 54
Position (Mb)

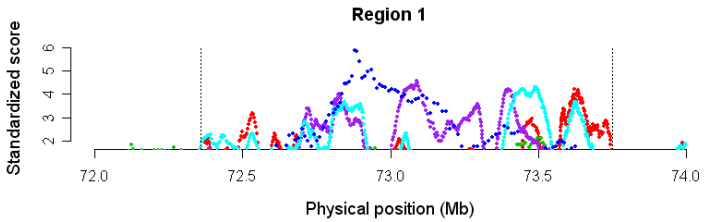

Region 17
Normalized score
2 3 4 5
20 22 24 26 28 30
Position (Mb)

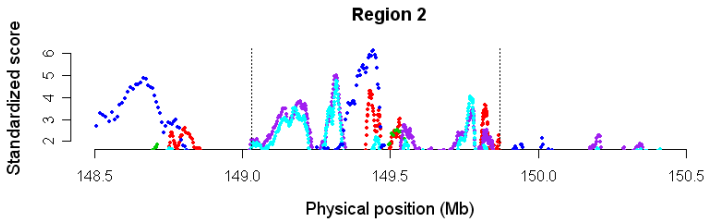

Region 18
Normalized score
2 4 6 8
56 58 60 62 64
Position (Mb)

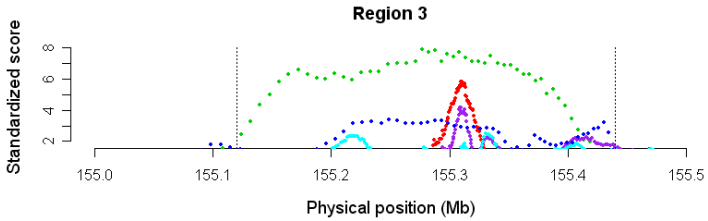

Region 19
Normalized score
2 4 6 8 10
106 108 110 112 114
Position (Mb)

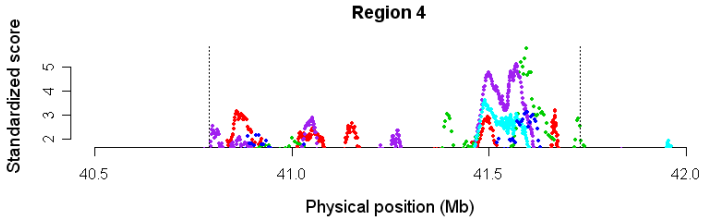

Region 20
Normalized score
2 3 4 5 6
56 58 60 62 64
Position (Mb)

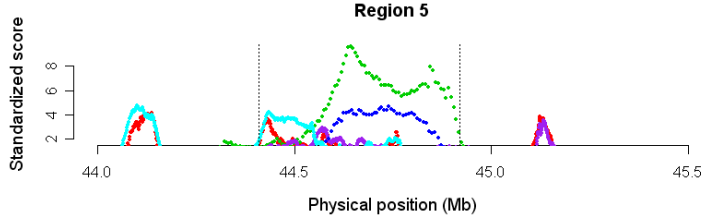
Supplementary Figure 5. Standardized varLD scores across different population pairs in the top 20 candidate regions undergoing positive selection from Sabeti et al. as reported in Table 2 in the main text. Only scores above their respective 95th quantiles are shown in a non-gray colour, and points in red correspond to LD comparisons between HapMap Europeans (CEU) and HapMap Asians (CHB+JPT); points in purple between CEU and HapMap Africans (YRI); points in cyan between CHB+JPT and YRI. Dotted lines in each panel designate the approximate start and end positions each region.

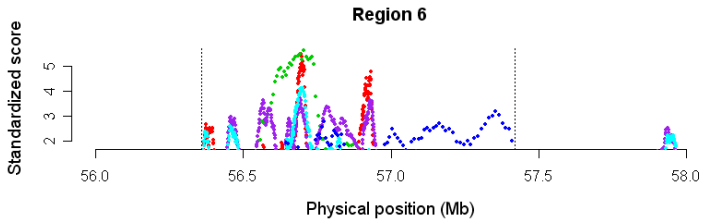

Region 1
Standardized score
2 3 4 5 6
72.0 72.5 73.0 73.5 74.0
Physical position (Mb)

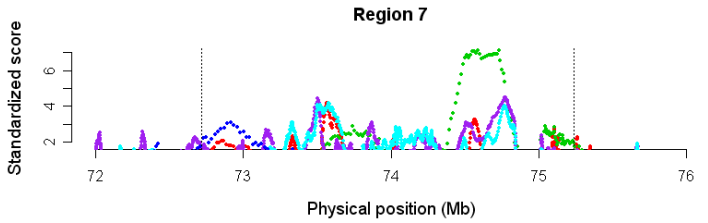

Region 2
Standardized score
2 3 4 5 6
148.5 149.0 149.5 150.0 150.5
Physical position (Mb)

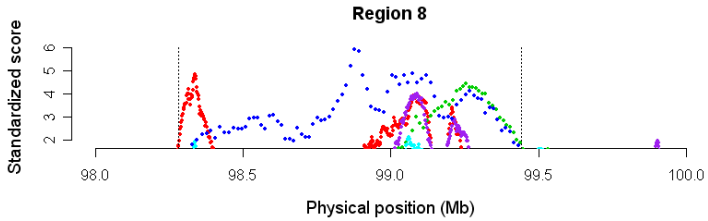

Region 3
Standardized score
2 4 6 8
155.0 155.1 155.2 155.3 155.4 155.5
Physical position (Mb)

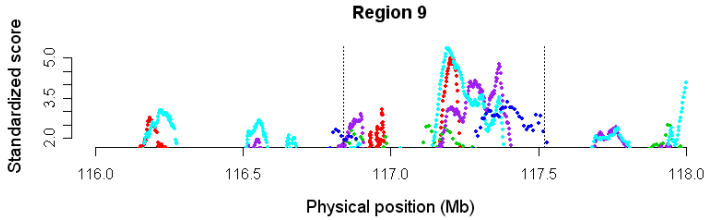

Region 4
Standardized score
2 3 4 5
40.5 41.0 41.5 42.0
Physical position (Mb)

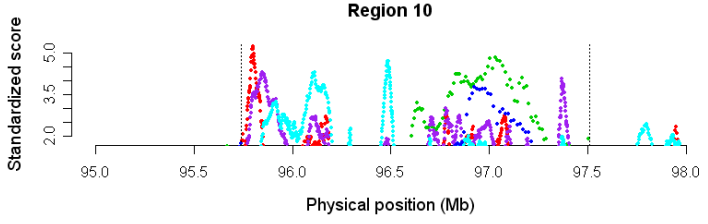

Region 5
Standardized score
2 4 6 8
44.0 44.5 45.0 45.5
Physical position (Mb)

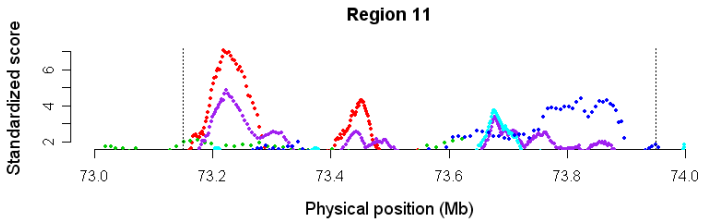

Region 6
Standardized score
2 3 4 5
56.0 56.5 57.0 57.5 58.0
Physical position (Mb)

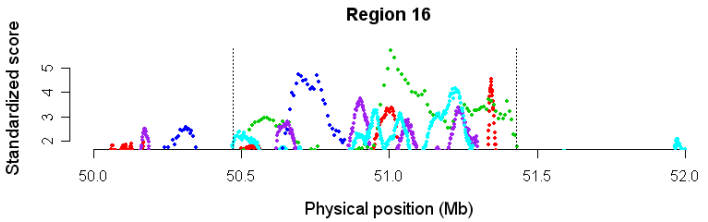

Region 7
Standardized score
2 4 6
72 73 74 75 76
Physical position (Mb)

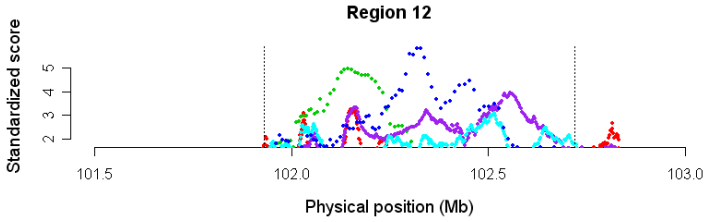

Region 8
Standardized score
2 3 4 5 6
98.0 98.5 99.0 99.5 100.0
Physical position (Mb)

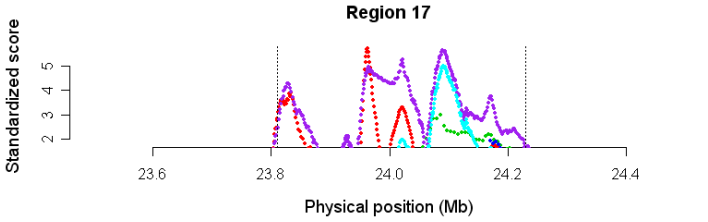

Region 9
Standardized score
2.0 3.5 5.0
116.0 116.5 117.0 117.5 118.0
Physical position (Mb)

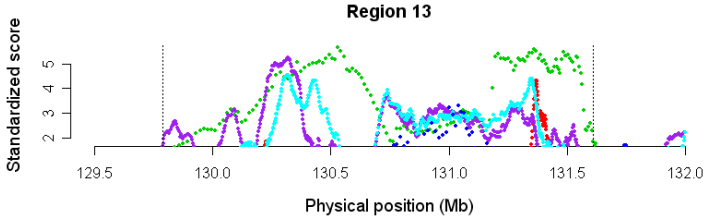

Region 10
Standardized score
2.0 3.5 5.0
95.0 95.5 96.0 96.5 97.0 97.5 98.0
Physical position (Mb)

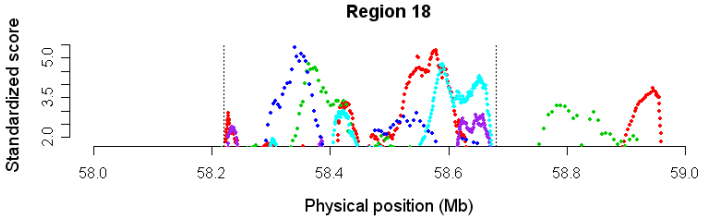

Region 11
Standardized score
2 4 6
73.0 73.2 73.4 73.6 73.8 74.0
Physical position (Mb)

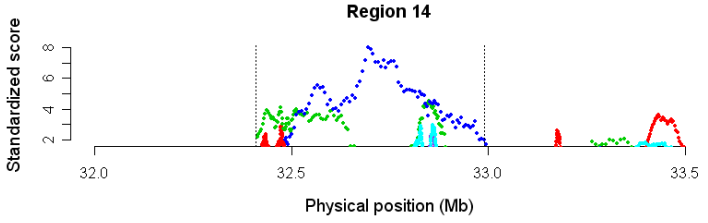

Region 16
Standardized score
2 3 4 5
50.0 50.5 51.0 51.5 52.0
Physical position (Mb)

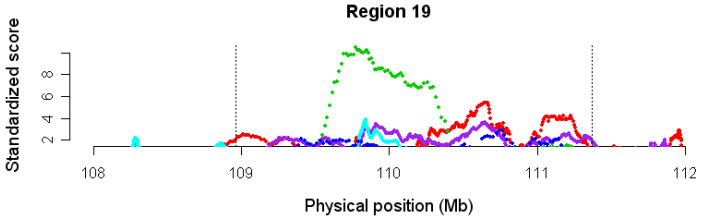

Region 12
Standardized score
2 3 4 5
101.5 102.0 102.5 103.0
Physical position (Mb)

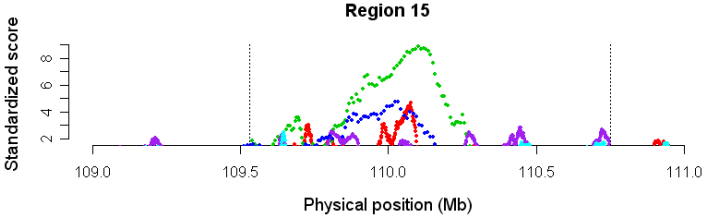

Region 17
Standardized score
2 3 4 5
23.6 23.8 24.0 24.2 24.4
Physical position (Mb)

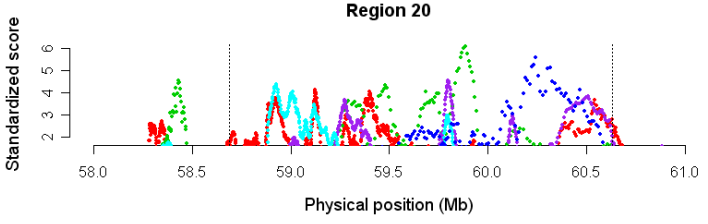

Region 13
Standardized score
2 3 4 5
129.5 130.0 130.5 131.0 131.5 132.0
Physical position (Mb)


Region 18
Standardized score
2.0 3.5 5.0
58.0 58.2 58.4 58.6 58.8 59.0
Physical position (Mb)

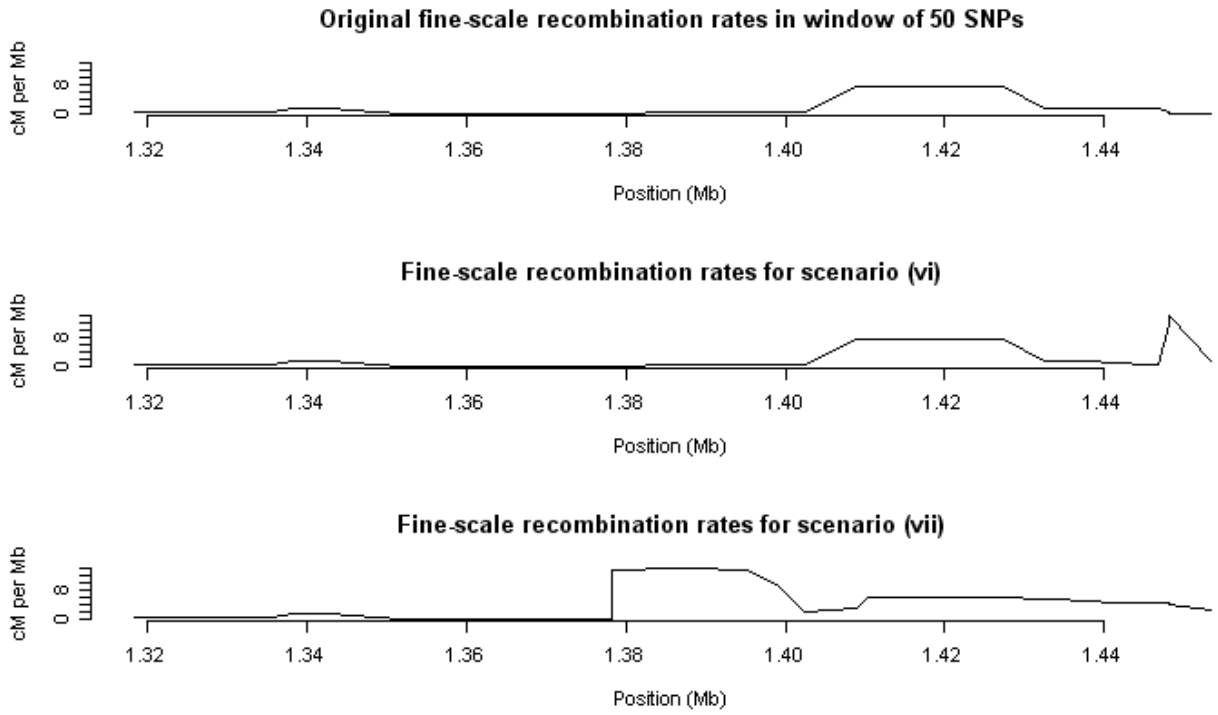

Region 14
Standardized score
2 4 6 8
32.0 32.5 33.0 33.5
Physical position (Mb)


Region 19
Standardized score
2 4 6 8
108 109 110 111 112
Physical position (Mb)


Region 15
Standardized score
2 4 6 8
109.0 109.5 110.0 110.5 111.0
Physical position (Mb)


Region 20
Standardized score
2 3 4 5 6
58.0 58.5 59.0 59.5 60.0 60.5 61.0
Physical position (Mb)

Supplementary Figure 6. Fine-scale recombination rates in the window of 50 SNPs used in a simulation exercise to investigate the sensitivity of varLD. The top panel shows the original fine-scale recombination rates for 50 SNPs; the middle panel shows the fine-scale recombination rates for the same 50 SNPs, except that the rates for the last 5 SNPs are different; the bottom panel shows the fine-scale recombination rates for the same 50 SNPs, except that the rates for the last 20 SNPs are different.



Supplementary Table 1. Top 20 candidate regions with overlapping signals of LD differences between CEU and CHB+JPT.

Region	Chr: start – end (Mb, HG17)	Genes in region	CNV region	Details (type[1], population[2])
1	chr1: 102.80 – 103.30	*COL11A1*	No	-
2	chr1: 106.53 – 106.99	-	Yes	Deletions, HapMap samples
3	chr1: 201.44 – 201.56	*NFASC*	No	-
4	chr2: 52.30 – 52.62	-	No	-
5	chr2: 149.82 – 150.50	*C2orf25*	No	-
6	chr3: 99.99 – 100.23	*ST3GAL6*, *DCBLD2*	No	-
7	chr3: 109.63 – 109.98	*KIAA1524*, *DZIP3*, *RETNLB*	No	-
8	chr4: 26.99 – 27.42	-	No	-
9	chr4: 73.17 – 73.47	*GPR74*	No	-
10	chr5: 103.29 – 103.60	-	Yes	Deletions, HapMap samples and HapMap-CEU
11	chr7: 86.23 – 86.39	*KIAA1324L*	No	-
12	chr9: 10.83 – 11.42	-	Yes	Deletions, German and HapMap samples
13	chr10: 111.42 – 111.48	-	No	-
14	chr11: 82.93 – 83.66	*DLG2*	Yes	Copy number differences, French and HapMap samples
15	chr11: 100.11 – 100.60	*PGR*	No	-
16	chr11: 131.36 – 131.44	*HNT*	Yes	Insertions, Chinese samples
17	chr13: 87.91 – 88.36	-	Yes	Copy number differences, HapMap and 36 diverse human samples
18	chr13: 106.22 – 106.56	-	Yes	Insertions, CEPH and Japanese samples
19	chr15: 46.04 – 46.67	*SLC24A5*, *DUT*, *MYEF2*, *SLC12A1*, *FBN1*	Yes	Deletions, HapMap samples, HapMap-CEU and 36 diverse human samples
20	chr21: 24.31 – 24.34	-	No	-

[1] Copy number differences refer to the occurrence of both insertions and deletions.

[2] CEPH/Chinese/Japanese/Yoruba samples: Kidd et al. (2008); German: Pinto et al. (2007); HapMap samples: Conrad et al. (2006), McCarroll et al. (2006), Redon et al. (2006), Pinto et al. (2007); French: de Smith et al. (2007); Canadian Ontario controls: Zogopoulos et al. (2007); HapMap-CEU: Wang et al. (2008); 36 diverse human samples: Mills et al. (2006).

Supplementary Table 2. Top 20 candidate regions with overlapping signals of LD differences between CEU and YRI.

Region	Chr: start – end (Mb, HG17)	Genes in region	CNV region	Details (type[1], population[2])
1	chr1: 105.57 – 105.69	-	Yes	Copy number differences, HapMap-CEU
2	chr3: 40.34 – 40.70	*ENTPD3*, *RPL14*, *ZNF619*, *ZNF620*, *ZNF621*	Yes	Deletions, HapMap-CEU
3	chr4: 17.02 – 17.37	*QDPR*, *LAP3*, *MED28*	Yes	Deletions, HapMap samples and 36 diverse human samples
4	chr5: 88.33 – 88.47	-	No	-
5	chr5: 129.79 – 131.54	*H1NT1*, *ACSL6*, *CDC42SE2*, *IL3*, *RAPGEF6*, *KIAA1961*	Yes	Deletions, HapMap samples
6	chr8: 64.59 – 64.84	-	No	-
7	chr8: 91.02 – 92.21	*NBN*, *DECR1*, *CALB1*, *TMEM64*, *TMEM55A*, *EFCBP1*, *OTUD6B*	Yes	Copy number differences, German, HapMap samples, 36 diverse human samples
8	chr8: 92.55 – 93.24	*RUNX1T1*	No	-
9	chr8: 102.75 – 102.95	*GRHL2*, *NCALD*	No	-
10	chr9: 34.77 – 34.92	-	No	-
11	chr10: 23.81 – 24.23	-	No	-
12	chr10: 111.03 – 111.80	*XPNPEP1*, *ADD3*	No	-
13	chr10: 121.95 – 122.18	-	Yes	Deletions, 36 diverse human samples
14	chr11: 77.59 – 78.34	*GAB2*	Yes	Copy number differences, French and HapMap samples
15	chr12: 23.31 – 23.52	-	No	-
16	chr13: 44.57 – 44.75	*GTF2F2*, *KCTD4*	No	-
17	chr14: 78.51 – 78.66	*NRXN3*	No	-
18	chr15: 55.90 – 56.00	-	No	-
19	chr17: 56.22 – 56.63	*BCAS3*	No	-
20	chr19: 34.49 – 34.56	-	No	-

[1] Copy number differences refer to the occurrence of both insertions and deletions.

[2] CEPH/Chinese/Japanese/Yoruba samples: Kidd et al. (2008); German: Pinto et al. (2007); HapMap samples: Conrad et al. (2006), McCarroll et al. (2006), Redon et al. (2006), Pinto et al. (2007); French: de Smith et al. (2007); Canadian Ontario controls: Zogopoulos et al. (2007); HapMap-CEU: Wang et al. (2008); 36 diverse human samples: Mills et al. (2006).

Supplementary Table 3. Top 20 candidate regions with overlapping signals of LD differences between CHB+JPT and YRI.

Region	Chr: start – end (Mb, HG17)	Genes in region	CNV region	Details (type[1], population[2])
1	chr1: 106.35 – 107.29	-	Yes	Copy number differences, HapMap samples
2	chr2: 117.14 – 117.38	-	Yes	Deletions, HapMap samples
3	chr2: 117.69 – 118.05	-	Yes	Deletions, HapMap samples
4	chr2: 215.20 – 215.64	*BARD1*, *ABCA12*	No	-
5	chr4: 84.71 – 84.91	*HEL308*, *MRPS18C*	No	-
6	chr5: 19.04 – 19.25	-	Yes	Deletions, Yoruba and HapMap samples
7	chr5: 26.54 – 27.77	*CDH9*	Yes	Copy number differences, CEPH and Yoruba samples, HapMap-CEU
8	chr5: 145.36 – 145.55	*SH3RF2*, *PLAC8L1*, *LARS*	No	-
9	chr5: 159.93 – 160.00	-	No	-
10	chr7: 3.61 – 4.19	*SDK1*	Yes	Copy number differences, CEPH, Yoruba and HapMap samples
11	chr7: 107.99 – 108.80	-	Yes	Insertions, HapMap samples
12	chr7: 154.91 – 155.23	*RMB33*, *SHH*	Yes	Deletions, Yoruba-Trios and 36 diverse human samples
13	chr8: 85.34 – 86.62	*LRRCC1*, *E2F5*, *CA13*, *CA1*, *CA2*, *CA3*	Yes	Copy number differences, Yoruba and HapMap samples, 36 diverse human samples
14	chr8: 111.57 – 112.30	-	No	-
15	chr9: 10.81 – 11.64	-	No	-
16	chr10: 10.02 – 10.34	-	Yes	Deletions, HapMap samples
17	chr10: 24.02 – 24.14	-	No	-
18	chr11: 77.58 – 78.36	*GAB2*	Yes	Copy number differences, French and HapMap samples
19	chr14: 31.10 – 31.45	*NUBPL*	No	-
20	chr19: 34.50 – 34.55	-	No	-

[1] Copy number differences refer to the occurrence of both insertions and deletions.

[2] CEPH/Chinese/Japanese/Yoruba samples: Kidd et al. (2008); German: Pinto et al. (2007); HapMap samples: Conrad et al. (2006), McCarroll et al. (2006), Redon et al. (2006), Pinto et al. (2007); French: de Smith et al. (2007); Canadian Ontario controls: Zogopoulos et al. (2007); HapMap-CEU: Wang et al. (2008); 36 diverse human samples: Mills et al. (2006); CEU-Trios/Yoruba-Trios: Conrad et al. (2006).

Supplementary Table 4. Top 20 candidate regions with overlapping signals of LD differences between Gambian Jola and YRI.

Region	Chr: start – end (Mb, HG17)	Genes in region	CNV region	Details (type[1], population[2])
1	chr1: 93.45 – 94.72	*CCDC18*, *DR1*, *FNBP1L*, *BCAR3*, *DNTTIP2*, *GCLM*, *ABCA4*, *ARHGAP29*, *ABCD3*, *F3*	No	-
2	chr1: 149.34 – 149.47	*LCE2C*, *LCE2D*, *LCE3A*, *LCE3B*, *LCE3C*, *LCE3D*, *LCE3E*	Yes	Copy number differences, multiple global populations (CEPH, German, Japanese, Yoruba and HapMap samples)
3	chr1: 194.83 – 195.18	*NEK7*	No	-
4	chr2: 161.10 – 162.57	*PBX1*, *LMX1A*, *RXRG*, *LRRC52*, *MGST3*, *ALDH9A1*, *TMC01*, *UCK2*	Yes	Deletions, HapMap-CEU and CEU-Trios
5	chr2: 194.35 – 195.07	-	Yes	Deletions, Yoruba and HapMap samples, 36 diverse human samples
6	chr4: 139.06 – 139.88	*SLC7A11*	Yes	Copy number differences, Yoruba samples
7	chr6: 25.93 – 26.43	*SLC17A2*, *TRIM38*, *HFE*, *RPS10P1*, *H3F3AP1*, *HIST1H* cluster (*~1A*, *3A*, *4A*, *4B*, *3B*, *2AB*, *2BB*, *3C*, *1C*, *4C*, *1T*, *2BC*, *2AC*, *1E*, *2BD*, *2BF*, *1D*, *4F*, *4G*, *3F*, *2BH*, *4H*, *PS2*, *PS1*, *2APS4*)	Yes	Deletions, Yoruba samples
8	chr6: 32.48 – 32.99	*HLA-DRA*, *HLA-DRB5*, *HLA-DRB1*, *HLA-DQA1*, *HLA-DOB*, *HLA-DQA2*, *HLA-DQB2*, *TAP2*, *PSMB8*, *TAP1*, *PSMB9*	Yes	Copy number differences, French and HapMap samples, YRI-Trios
9	chr6: 48.87 – 49.46	-	No	-
10	chr7: 124.00 – 124.47	*POT1*	No	-
11	chr8: 136.97 – 137.26	-	Yes	Deletions, Yoruba samples
12	chr9: 15.43 – 16.01	*SNAPC3*, *PSIP1*, *C9orf93*	Yes	Deletions, multiple global populations (CEPH, French, German, Yoruba, HapMap samples)
13	chr9: 21.65 – 21.96	*MTAP*	No	-
14	chr10: 31.13 – 31.45	-	No	-
15	chr10: 62.73 – 63.14	*TMEM26*	No	-
16	chr11: 47.99 – 52.56	*PTPRJ*, *FOLH1*, *OR4* cluster (*~B1*, *X2*, *X1*, *S1*, *C3*, *A47*, *C13*, *C12*, *A5*, *C46*)	Yes	Copy number differences, multiple global populations (CEPH, Canadian, French, German, Chinese, HapMap samples)
17	chr11: 84.30 – 84.92	*DLG2*	Yes	Deletions, HapMap samples
18	chr12: 30.23 – 30.44	-	No	-
19	chr18: 63.67 – 63.87	-	No	-
20	chr19: 34.39 – 34.81	*UQCRFS1*	Yes	Deletions, Chinese and HapMap samples

[1] Copy number differences refer to the occurrence of both insertions and deletions.

[2] CEPH/Chinese/Japanese/Yoruba samples: Kidd et al. (2008); German: Pinto et al. (2007); HapMap samples: Conrad et al. (2006), McCarroll et al. (2006), Redon et al. (2006), Pinto et al. (2007); French: de Smith et al. (2007); Canadian Ontario controls: Zogopoulos et al. (2007); HapMap-CEU: Wang et al. (2008); 36 diverse human samples: Mills et al. (2006); CEU-Trios/Yoruba-Trios: Conrad et al. (2006).

Supplementary Table 5. Top 20 candidate regions with overlapping signals of LD differences between CEU and WTCCC-58C.

Region	Chr: start – end (Mb, HG17)	Genes in region	CNV region	Details (type[1], population[2])
1	chr1: 155.12 – 155.42	*CD1E*, *OR10T2*, *OR10K2*, *OR10K1*, *OR10R2*, *OR6Y1*, *OR10Z1*, *OR10X1*, *SPTA1*	No	-
2	chr1: 165.79 – 166.06	*ATP1B1*, *NME7*	Yes	Copy number differences, French and HapMap-CEU
3	chr1: 185.99 – 186.40	-	Yes	Deletions, Canadian, German, HapMap samples and HapMap-CEU
4	chr2: 39.08 – 39.53	*RHBDL2*, *C1orf108*, *NDUFS5*, *MACF1*	No	-
5	chr2: 44.46 – 44.92	*SLC3A1*, *PREPL*, *C2orf34*	No	-
6	chr2: 74.38 – 75.24	-	Yes	Copy number differences, German and HapMap samples
7	chr3: 134.14 – 134.34	-	No	-
8	chr4: 119.04 – 119.83	*NDST3*, *PRSS12*	Yes	Deletions, CEPH and Yoruba samples, 36 diverse human samples
9	chr6: 35.67 – 36.19	*FKBP5*, *C6orf81*, *UNQ3045*, *CLPS*, *LHFPL5*, *SRPK1*, *SLC26A8*, *MAPK14*	Yes	Copy number differences, multiple global populations (CEPH, French, Chinese, Yoruba, HapMap-CEU)
10	chr6: 109.54 – 110.26	*C6orf182*, *CD164*, *ZBTB24*, *PPIL6*, *KIAA0274*, *SMPD2*, *MICAL1*, *C6orf199*	Yes	Deletions, HapMap samples
11	chr7: 83.59 – 84.60	*SEMA3D*	Yes	Deletions, German and HapMap samples
12	chr7: 86.10 – 86.52	*GRM3*, *KIAA1324L*, *DMTF1*, *C7orf23*	No	-
13	chr8: 58.03 – 58.42	*IMPAD1*	Yes	Copy number differences, multiple global populations (CEPH, French, Chinese, Japanese, Yoruba, HapMap samples, 36 diverse human samples)
14	chr8: 124.25 – 124.65	*C8orf76*, *ZHX1*, *ATAD2*, *C8orf32*, *FBXO32*	No	-
15	chr10: 57.33 – 57.58	-	No	-
16	chr10: 86.52 – 87.06	-	No	-
17	chr11: 13.90 – 14.07	*SPON1*	No	-
18	chr12: 109.53 – 110.39	*PPP1CC*, *MYL2*, *CUTL2*, *LNK*, *ATXN2*	Yes	Deletions, CEPH and HapMap samples
19	chr18: 28.80 – 29.27	*C18orf34*	Yes	Null genotypes, CEPH and HapMap samples
20	chr21: 37.77 – 38.65	*DYRK1A*, *KCNJ6*, *DSCR4*, *DSCR8*, *DSCR10*, *KCNJ15*	Yes	Deletions, French

[1] Copy number differences refer to the occurrence of both insertions and deletions.

[2] CEPH/Chinese/Japanese/Yoruba samples: Kidd et al. (2008); German: Pinto et al. (2007); HapMap samples: Conrad et al. (2006), McCarroll et al. (2006), Redon et al. (2006), Pinto et al. (2007); French: de Smith et al. (2007); Canadian Ontario controls: Zogopoulos et al. (2007); HapMap-CEU: Wang et al. (2008); 36 diverse human samples: Mills et al. (2006); CEU-Trios/Yoruba-Trios: Conrad et al. (2006).

Supplementary Table 6. Proportions of simulated datasets out of 1000 with standardized varLD scores greater than the corresponding quantile values.

Scenario	Positive rates		
	> 90th quantile	> 95th quantile	> 99th quantile
(i) Identical populations	0.346	0.167	0.026
Genotyping errors			
(ii) 1st SNP in pop 2	0.491	0.257	0.029
(iii) 20th SNP in pop 2	0.439	0.238	0.029
(iv) 5 random SNPs in pop 2	0.744	0.552	0.088
(v) 5 consecutive SNPs in pop 2	0.992	0.957	0.555
Differences in recombination rates			
(vi) SNPs 46 – 50	0.597	0.395	0.075
(vii) SNPs 31 – 50	0.651	0.414	0.070

ADDITIONAL SUPPLEMENTARY FILES

Supplementary File 1. (identified_region_CEU_JPT+CHB.xls)
Contains identified regions with LD dissimilarities between HapMap CEU and HapMap CHB+JPT populations.

Supplementary File 2. (identified_region_CEU_YRI.xls)
Contains identified regions with LD dissimilarities between HapMap CEU and HapMap YRI populations.

Supplementary File 3. (identified_region_ JPT+CHB_YRI.xls)
Contains identified regions with LD dissimilarities between HapMap CHB+JPT and HapMap YRI populations.

Supplementary File 4. (identified_region_CEU_58C.xls)
Contains identified regions with LD dissimilarities between HapMap CEU and WTCCC 58C populations.

Supplementary File 5. (identified_region_Jola_YRI.xls)
Contains identified regions with LD dissimilarities between HapMap CEU and HapMap CHB+JPT populations.

Supplementary File 6. (varLD_overlapping_regions.xls)
Contains identified regions with LD dissimilarities between all five pairs of population comparisons.