Supplementary material for: Population Differentiation as a Test for Selective Sweeps

Hua Chen[1,2,*], Nick Patterson[2] and David Reich[1,2,*]

[1]Department of Genetics, Harvard Medical School, Boston MA 02115

[2]Broad Institute of MIT and Harvard, Cambridge MA 02142

* To whom correspondence should be addressed (Department of Genetics, Harvard Medical School, NRB, 77 Avenue Louis Pasteur, Boston MA 02115; Hua Chen (hchen@genetics.med.harvard.edu) or David Reich (reich@genetics.med.harvard.edu).

1. Four demographic histories used in the simulation.

1. Two populations with constant sizes of $10,000$ and split times of 700 generations ago (meant to be similar to a potential split time for Asian and European populations).

2. Two populations with constant sizes of $10,000$ and split times of $3,000$ generations ago (around the lower bound of the split time for African and Eurasian populations).

3. The Schaffner model that has been used to fit genetic variation patterns across continentally diverse populations in Schaffner et al. (2005).

4. Out-of-Africa bottleneck model with intensity of 0.20 and split time $3,000$ generations ago.

Table S1: The top 40 regions of the human genome based on XP-CLR test in the ASN-YRI comparison

Chr#	Positions (Mb)			max XP-CLR	Genes	XP-EHH	iHS(ASN)	iHS(YRI)	Study
2	72.33	-	73.03	733.7	CYP26B1,SPR,EMX1,SFXN5	0.0002	0.0374	0.0023	[1],[2]
8	10.77	-	11.17	673.6	PINX1,XKR6,C8orf15,MTMR9,AMAC1L2	0.0004	<0.0001	0.0030	[1]
5	117.38	-	117.77	493.8		0.0013	0.0709	0.0092	[1],[3]
22	44.95	-	45.14	448.2	FLJ27365,PPARA,LOC150383,PKDREJ,FLJ20699,GTSE1,TRMU,CELSR1	0.0002	0.0088	0.0048	[1]
8	134.92	-	135	426.1		0.0047	0.1122	0.0403	
2	177.26	-	177.7	412.8	HNRPA3	0.0000	0.0147	0.0041	[1],[2]
11	96.69	-	96.79	394.7		0.0054	0.0563	0.0053	
7	135.24	-	135.34	374.7	MTPN	0.0010	0.1167	0.0091	
11	60.6	-	60.77	373.5	CD6,CD5,VPS37C,PGA3,PGA4,PGA5,VWCE,DDB1,DAK	0.0030	0.0412	0.0185	[2]
5	116.92	-	117.03	372.6		0.0464	0.2349	0.0066	
4	159.83	-	160.16	349.2	RXFP1,LOC201725,ETFDH,PPID	0.0009	0.0406	0.0039	
4	158.85	-	159.12	348.6		0.0002	0.0001	0.0144	[1],[2]
10	22.59	-	22.8	344.9	COMMD3,BMI1,SPAG6,PIP5K2A	0.0003	0.1616	0.0012	[1],[2]
2	121.39	-	121.44	341.1	GLI2	0.0089	0.2221	0.0079	
11	6.04	-	6.14	340.5	OR52L1,OR56A4,OR56A1,OR56B4,OR52B2,OR52W1,C11orf42,C11orf56,CNGA4	0.0217	0.0375	0.0417	
16	46.44	-	47.04	339.2	ABCC12,ABCC11,LONP2,SIAH1,N4BP1	0.0078	0.1544	0.0266	[2],[3]
16	65.61	-	66.35	324.6	CDH16,RRAD,FAM96B,CES2,FLJ21736,FLJ37464,CBFB,C16orf70,MGC4655 TRADD,FBXL8,HSF4,NOL3,LOC653319,EXOC3L,E2F4,ELMO3,LRRC29 HSPC171,FHOD1,SLC9A5,PLEKHG4,LRRC36,CGI38,ZDHHC1,HSD11B2 ATP6V0D1,AGRP,FAM65A,CTCF,RLTPR,ACD,PARD6A,C16orf48,LOC388284 GFOD2,RANBP10,TSNAXIP1,CENPT,THAP11,NUTF2	0.0129	0.2124	0.0086	[2],[3]
18	36.61	-	36.7	323.1		0.2036	0.3439	0.0079	
12	23.23	-	23.37	311.0		0.0480	0.0271	0.0049	
10	107.14	-	107.28	307.9		0.0152	0.0088	0.0342	[2]
22	44.79	-	44.84	303.4	WNT7B,C22orf26,FLJ27365,PPARA	0.0031	0.1647	0.0536	
18	65.73	-	66.02	301.1	DOK6,CD226,SOCS6	0.0211	0.0962	0.0147	[3]
15	43.93	-	44.06	301.0		0.0024	0.0333	0.0043	
1	106.58	-	106.67	298.7		0.0122	0.2030	0.0023	[2]
4	41.78	-	41.99	297.1	SLC30A9,CCDC4	0.0020	0.0002	0.0015	[1],[2]
8	11.27	-	11.31	293.3	MTMR9,AMAC1L2,C8orf13,BLK	0.0021	0.1399	0.0150	[2]
3	71.44	-	71.72	293.0	EIF4E3	0.0068	0.0090	0.0001	
14	68.51	-	68.62	292.9	ACTN1,WDR22	0.0106	0.0556	0.0081	
17	70.7	-	70.82	291.2	SLC16A5,ARMC7,NT5C,HN1,SUMO2,NUP85,GGA3,MRPS7 MIF4GD,SLC25A19,GRB2	0.0304	0.0725	0.0317	
3	109.02	-	109.16	288.3	BBX,CD47	0.0068	0.0241	0.0081	[1]
3	136.18	-	136.22	287.1	EPHB1	0.0075	0.0268	0.0174	[2]
11	78.25	-	78.35	286.8		0.0382	0.0268	0.0174	
7	111.75	-	111.94	283.4	ZNF277P,IFRD1,FLJ39575	0.0245	0.0048	0.0131	[2]
20	55.27	-	55.42	282.1	BMP7,SPO11,RAE1,RBM38,HMG1L1,CTCFL	0.0240	0.0760	0.0069	[2]
18	30.42	-	30.75	280.5	DTNA	0.0034	0.0797	0.0100	[2]
7	54.87	-	54.92	278.2	SEC61G	0.0049	0.0100	0.0715	
6	13.46	-	13.48	276.6	PHACTR1,TBC1D7,GFOD1,C6orf114	0.1712	0.1021	0.0726	
11	3.07	-	3.09	275.1	NAP1L4,CARS,OSBPL5	0.0170	0.0224	0.0285	
3	25.8	-	26.37	272.9	NGLY1,OXSM	0.0005	0.0034	0.0122	[1],[2]
4	135.71	-	136.01	271.0		0.0006	0.1928	0.0157	

Note that the studies referenced are: top 14 signals from JPB+CHB-YRI comparison of [1] (Sabeti et al. 2007), top 200 signals of [2] (Frazer et al 2007) and top 20 signals from Asian populations of [3] (Carlson et al. 2005)

Table S2: Top 40 regions of the human genome based on the XP-CLR test in the northern-southern European comparison

Chr#	Positions (Mb)			max XP-CLR	Genes
2	135.12	-	136.68	330.57	LCT,R3HDM1,UBXD2,MCM6,DARS,CXCR4,TMEM163 ACMSD,CCNT2,YSK4,RAB3GAP1,R3HDM1
21	37.95	-	38.01	109.45	KCNJ6
12	43.5	-	43.76	92.01	NELL2,DBX2,PLEKHA9
14	30.36	-	30.42	87.18	SCFD1,COCH,STRN3
7	110.01	-	110.06	83.95	IMMP2L
3	183.42	-	183.5	83.37	
12	100.44	-	100.52	82.77	SPIC,MYBPC1,CHPT1
20	24.85	-	24.97	81.35	CST7,C20orf3,ACSS1,VSX1
9	6.86	-	6.91	80.09	JMJD2C
10	43.43	-	43.51	79.44	ZNF239,ZNF485,ZNF32
3	10.84	-	10.93	77.48	SLC6A11,SLC6A1
3	171.06	-	171.14	75.62	ARPM1,MYNN,LRRC34,LOC344657,LRRC31 SAMD7,TLOC1
12	4.08	-	4.14	75.52	
3	9.35	-	9.51	74.98	SRGAP3,THUMPD3,SETD5,LHFPL4
1	208.59	-	208.7	72.74	HHAT
1	59.62	-	59.7	71.74	FLJ10986
1	19.92	-	19.99	71.66	C1orf151,NBL1,HTR6,TMCO4,RNF186
17	52.11	-	52.16	70.67	NOG
15	34.53	-	34.6	69.42	C15orf41
6	28.67	-	28.86	68.05	GPX6,GPX5,ZNF452
1	88.57	-	88.7	65.24	
18	42.26	-	42.29	65.20	RNF165,LOXHD1
10	124.8	-	124.89	64.56	C10orf88,C10orf89,IKZF5,ACADSB,HMX2,BUB3
7	151.49	-	151.64	62.42	GALNT11,MLL3
2	162.82	-	162.93	61.01	FAP,IFIH1,GCA,KCNH7
19	43.6	-	43.68	60.61	YIF1B,KCNK6,C19orf15,PSMD8,GGN,SPRED3 FAM98C,RASGRP4,RYR1,MAP4K1
3	171.42	-	171.54	60.56	PHC3,PRKCI,SKIL,CLDN11
11	1.18	-	1.37	60.34	MUC2,TOLLIP,BRSK2,HCCA2
2	111.01	-	111.31	59.18	RGPD5,RGPD7,BUB1,ACOXL
11	115.57	-	115.61	59.08	
4	114.74	-	114.89	58.07	CAMK2D
1	162.95	-	163.03	58.02	PBX1
4	30.7	-	30.87	57.17	PCDH7
6	160.44	-	160.55	56.99	IGF2R,SLC22A1,SLC22A2
11	70.56	-	70.62	56.00	
15	25.96	-	26.32	55.81	OCA2,HERC2,GOLGA8G,FLJ32679
3	27.49	-	27.58	55.75	SLC4A7
4	125.35	-	125.56	54.75	
4	90.86	-	90.99	53.85	SNCA,SNCA,MMRN1
8	34.81	-	34.91	53.51	

Table S3: Enrichment of genes under selective sweeps in GO categories

GO category	Function	CEU-YRI	ASN-YRI	CEU-ASN	NEU-SEU
GO:0042981	regulation of apoptosis	0.00003	0.01900	0.0041000	-
GO:0001569	pattern of blood vessels	0.00041	0.02500	-	-
GO:0007156	homophilic cell adhesion	0.00063	-	-	-
GO:0002520	immune system development	0.00065	-	-	-
GO:0045934	negative regulation of nucleobase, nucleoside	0.00110	-	0.0010000	-
GO:0004672	protein kinase activity	0.00360	0.00660	0.0001000	-
GO:0007420	brain development	0.00540	-	-	-
GO:0042127	regulation of cell proliferation	-	0.00110	-	-
GO:0001655	urogenital system development	-	0.00240	-	-
GO:0060047	heart contraction	-	0.00990	-	-
GO:0009966	regulation of signal transduction	-	0.00032	-	-
GO:0030323	respiratory tube development	-	0.00900	0.0009500	-
GO:0002541	inflammatory response	-	-	**0.0000057***	-
GO:0006464	protein modification process	-	-	**0.0000130***	-
GO:0045941	positive regulation of transcription	-	-	0.0026000	-
GO:0008202	steroid metabolic process	-	-	0.0019000	0.00580
GO:0007517	muscle development	-	-	0.0200000	-
GO:0019724	B cell medicated immunity	-	-	-	0.00039
GO:0003708	retinoic acid receptor activity	-	-	-	0.00041
GO:0032787	monocarboxylic acid metabolic process	-	-	-	0.00770
GO:0007399	nervous system development	-	-	-	0.00570

* indicates P values that are significant after correcting for multiple testing using Benjamini-Hochberg correction (Benjamini and Hochberg 1995).

"-" indicates P values that are greater than 5%.

Table S4: Genes related to type-1 and type-2 diabetes (T1D and T2D), and their empirical significance according to the XP-CLR test

Genes	NSEU	NEU-YRI	ASN-YRI	CEU-ASN
T1D				
PTPN22	-	-	-	-
IFIH1	-	0.02350	-	0.00930
CTLA4	-	-	-	-
IL2RA	-	-	-	-
IGF2	-	-	-	-
ERBB3	-	-	-	-
PTPN2	-	-	-	-
PRKCQ	-	-	-	-
CTSH	-	-	-	-
PRM3, TNP2	-	-	-	-
KIAA0350	-	-	-	-
SH2B3	-	-	-	0.00430
INS	-	-	-	-
CD25	-	-	-	-
T2D				
JAZF1	-	-	-	-
CDC123, CAMK1D	-	0.00970	-	0.00790
TSPAN8, LGR5	-	-	-	-
THADA	-	-	0.04500	-
ADAMTS9	-	-	-	0.05360
PPARG	-	-	0.02610	-
IGF2BP2	-	-	-	-
TCF7L2	-	-	0.03860	-
FTO	-	-	-	-
NOTCH2	-	-	-	-
CDKAL1	-	-	-	-
CDKN2A	-	-	-	-
HHEX	-	-	-	-
KCNJ11	-	-	-	-
KCNQ1	-	-	-	-

Notes: "-" indicates empirical P values that are greater than 5%.

Figure Legends

Figure S1: The distribution of the inferred locations of the selected mutants in simulations with $s = 0.005$. The true locations of the selected mutants are at the middle of the region. The whole length of the region is 500Kb.

Figure S2: (a)-(b) A comparison of XP-CLR scores, calculated from simulations of an ascertainment scheme in which discovery was initially carried out in 50 individuals followed by requiring the minor allele frequency to be >5%. (a) Constant population size model with divergence time of $T = 700$ generations ago. (b) Constant population size model with divergence time of $T = 3000$ generations ago. Note that the XP-CLR scores in the figures were normalized. (c)-(d) A comparison of XP-CLR scores, calculated from simulations of models assuming constant recombination rates with those including recombination hotspots or mis-specified recombination rates. (c) The estimated recombination rate is $1/2$ of the true recombination rates. (d) The estimated recombination rate is 2 times the true recombination rates. Note that the XP-CLR scores in the figures were normalized.

Figure S3: Plot of XP-CLR scores along chromosomes in Northern-Southern European populations. Red line indicates the genome-wide 1% cutoff level.

Figure S4: Plot of XP-CLR scores along chromosomes in CEU-YRI population comparisons. Red line indicates a 1% genome-wide cutoff level.

Figure S5: Plot of XP-CLR scores along chromosomes in JPT+CHB-YRI populations. Red line indicates the genome-wide 1% cutoff level.

Figure S6: Plot of XP-CLR scores along chromosomes in CEU-JPT+CHB populations. Red line indicates the genome-wide 1% cutoff level.

Figure S7: The haplotype structure around *NRXN3* in CEU (upper panel) and YRI (lower panel) samples. Each row represents one haplotype. The two colors correspond to the two alleles of SNPs.

Figure S8: The haplotype structure around *TSLP* in CEU (upper panel) and YRI (lower panel) samples. Each row represents one haplotype. The two colors correspond to the two alleles of SNPs.

Figure S9: Plot of XP-CLR scores in the *OCA2-HERC2* gene region of chromosome 15. *OCA2* and *HERC2* are marked with bars and lines. *HERC2* is shaded with yellow color. The three red dots denote the position of SNPs associated with hair color and iris color in former GWA studies.
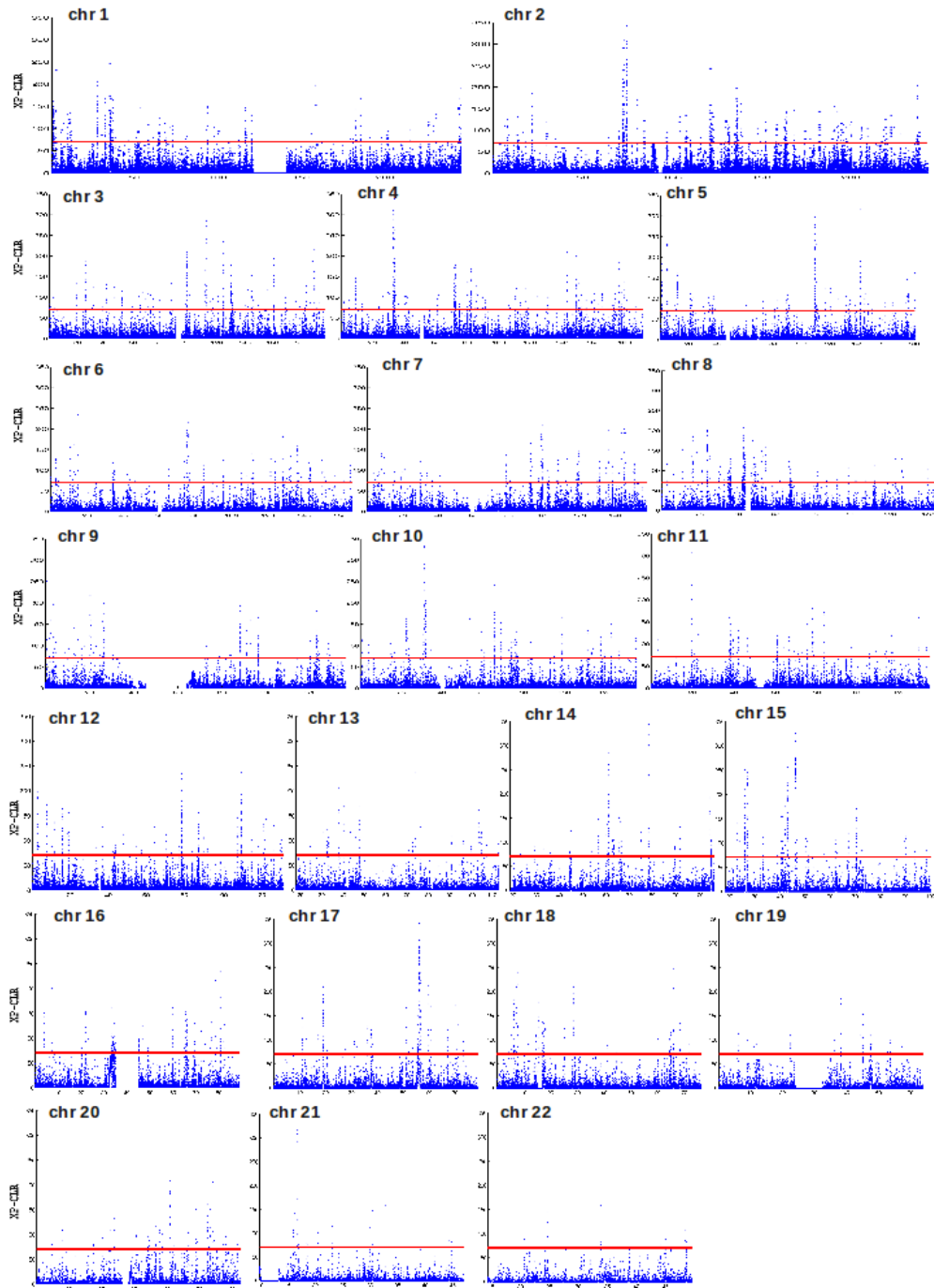
Figure S1
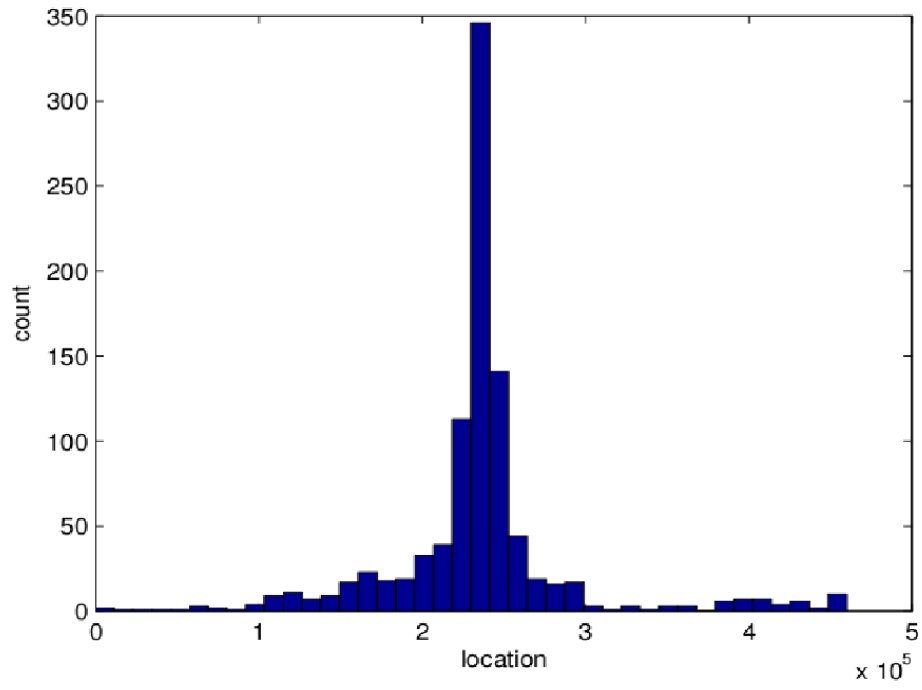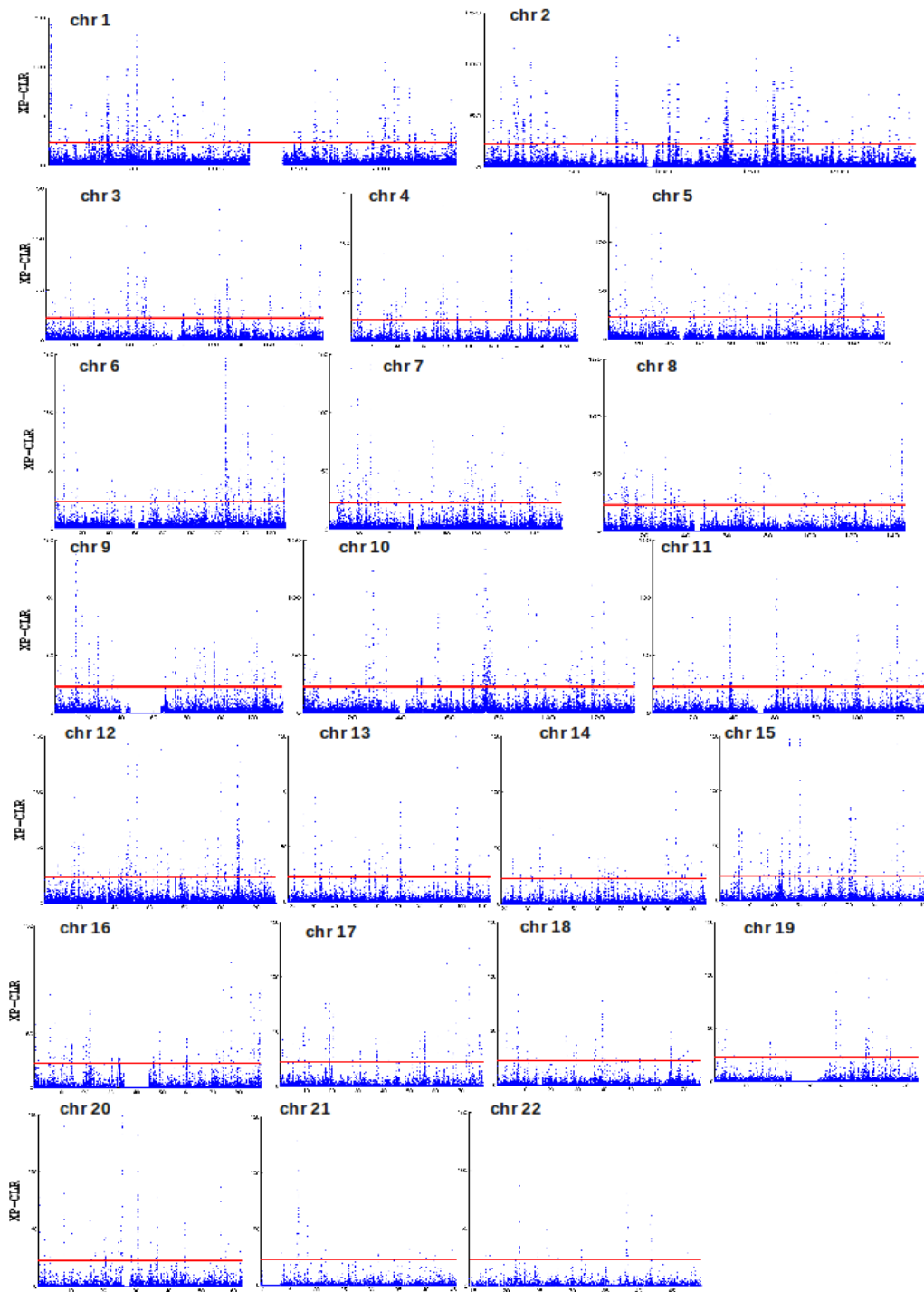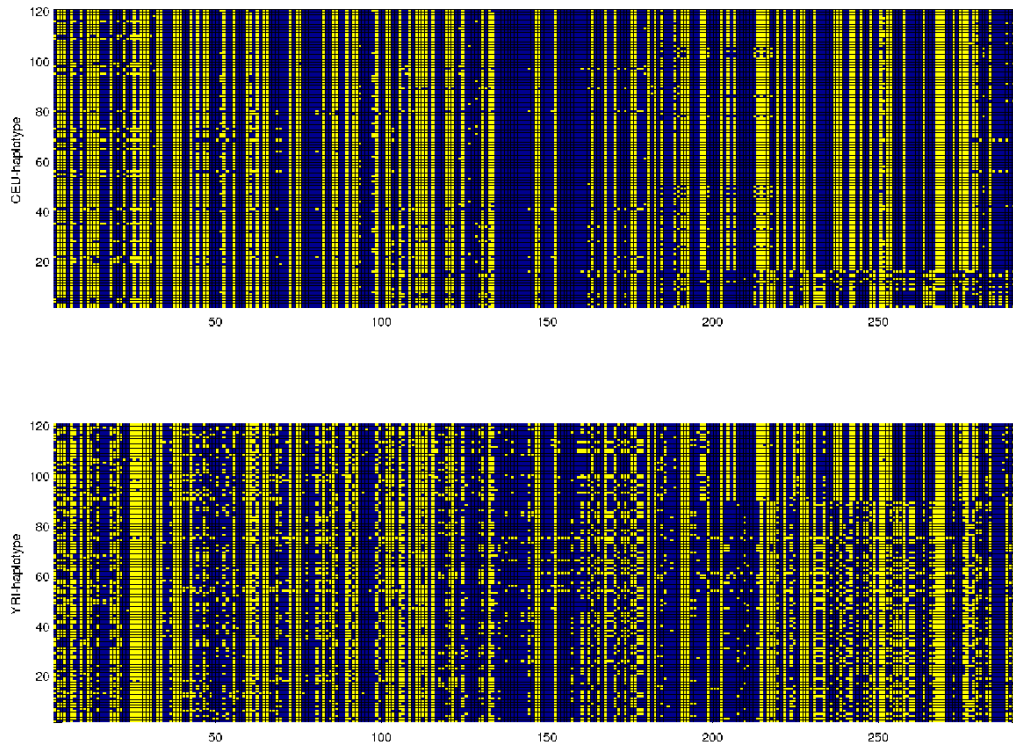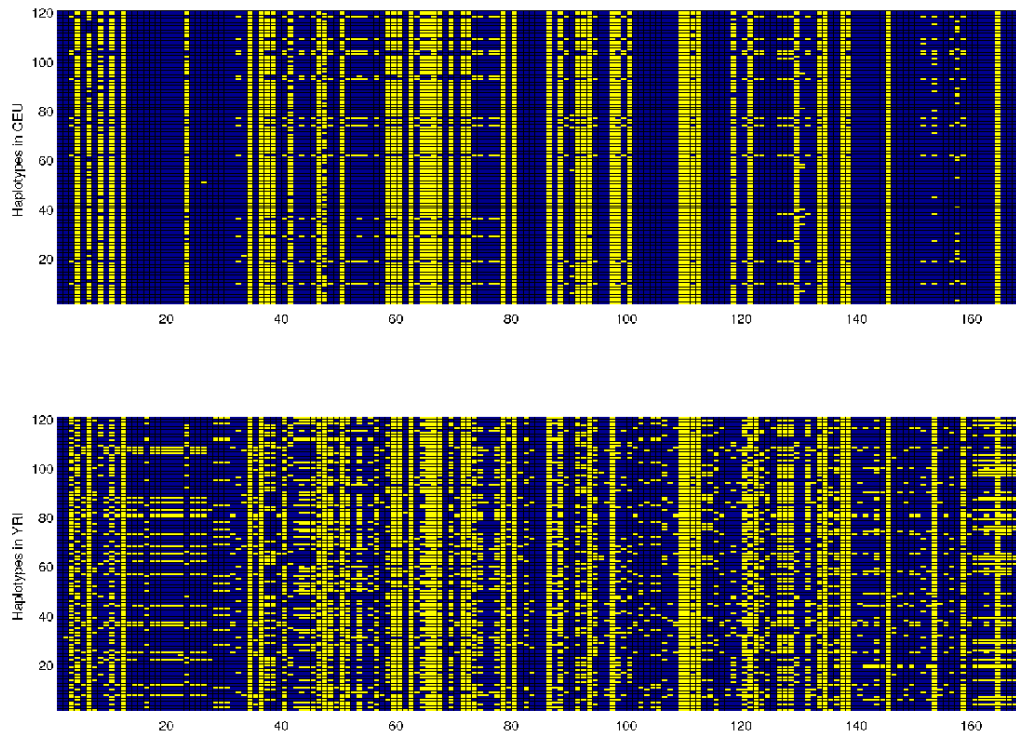
Figure S2
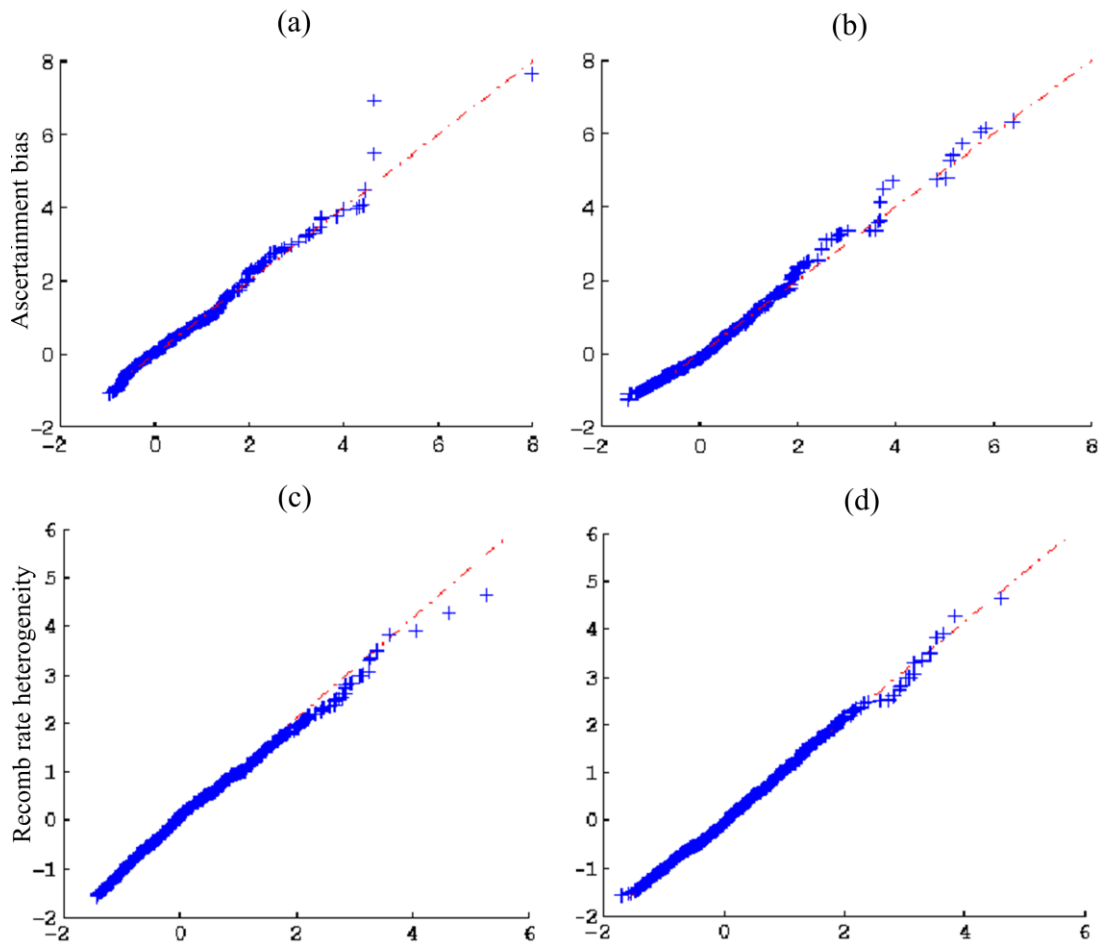
Figure S3
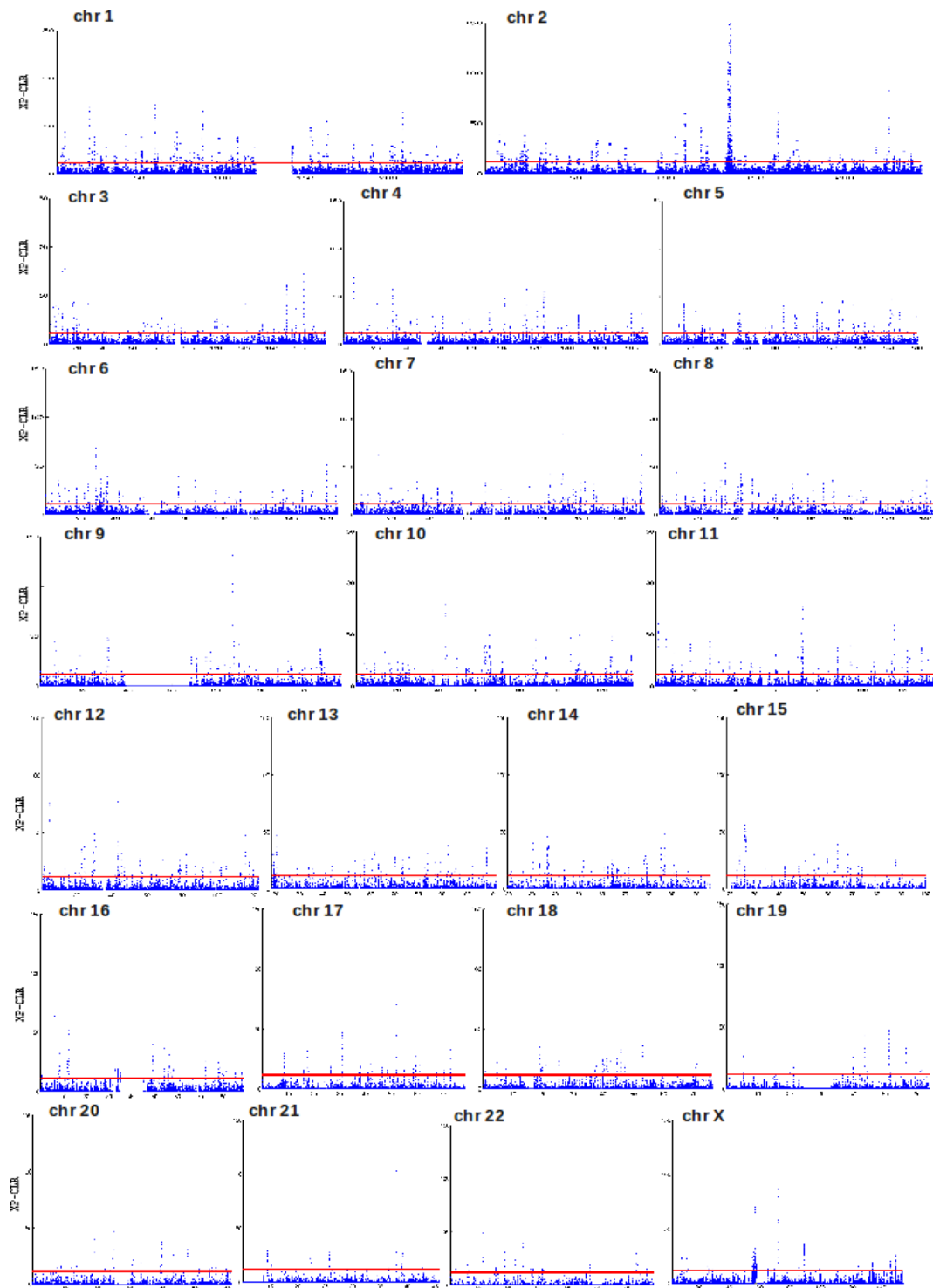
Figure S4

Figure S5
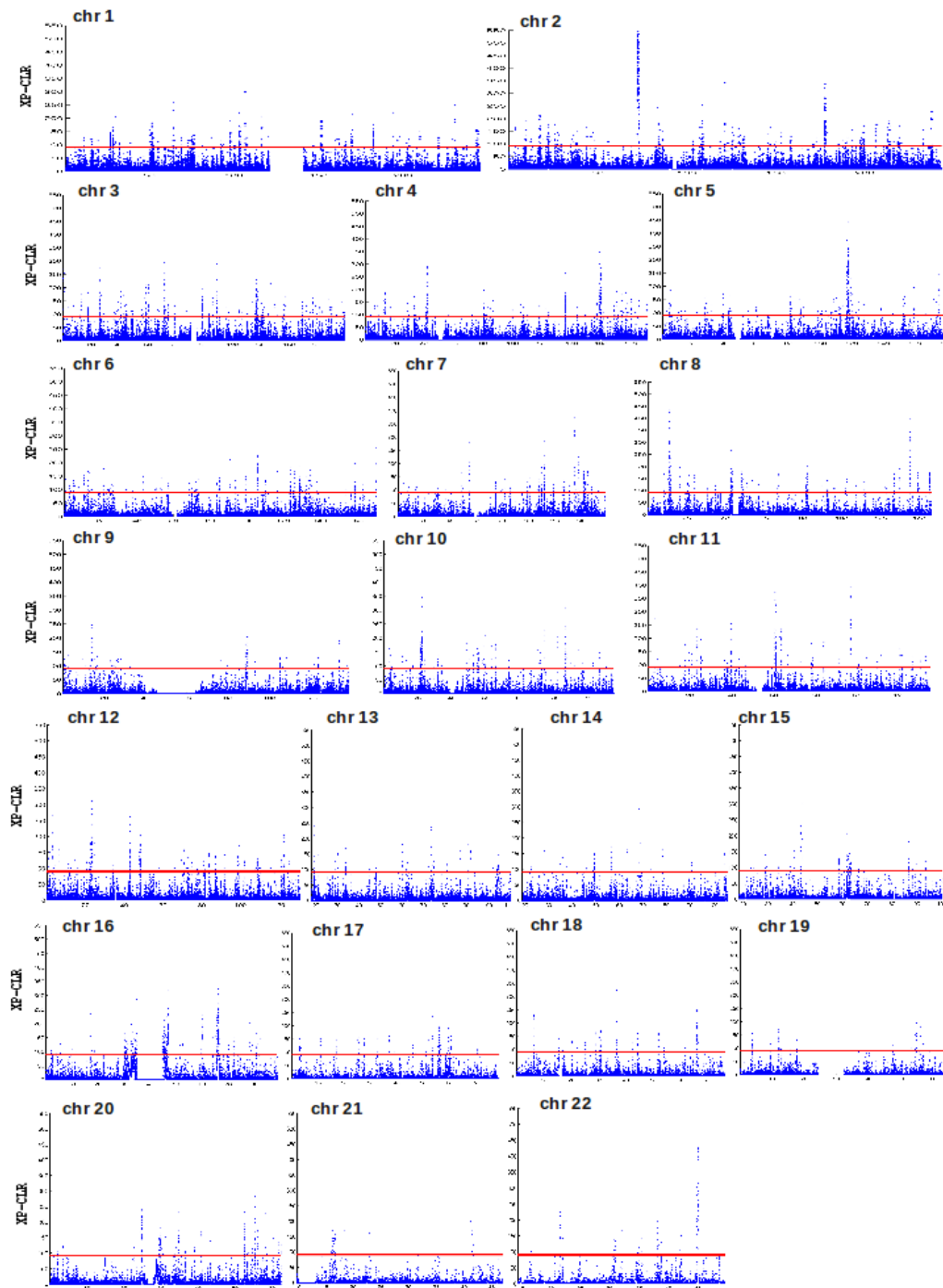
Figure S6

Figure S7

Figure S8

Figure S9