Supplementary Materials for

Natural Selection in a Bangladeshi Population from the Cholera-Endemic Ganges River Delta

Elinor K. Karlsson, Jason B. Harris, Shervin Tabrizi, Atiqur Rahman, Ilya Shylakhter, Nick Patterson, Colm O'Dushlaine, Stephen F. Schaffner, Sameer Gupta, Fahima Chowdhury, Alaullah Sheikh, Ok Sarah Shin, Crystal Ellis, Christine Becker, Lynda M. Stuart, Stephen B. Calderwood, Edward T. Ryan, Firdausi Qadri, Pardis C. Sabeti, Regina C. LaRocque

correspondence to: elinor@broadinstitute.org

This PDF file includes:

Methods
Figs. S1 to S4
Tables S1 to S11

Materials and Methods

Selection analysis

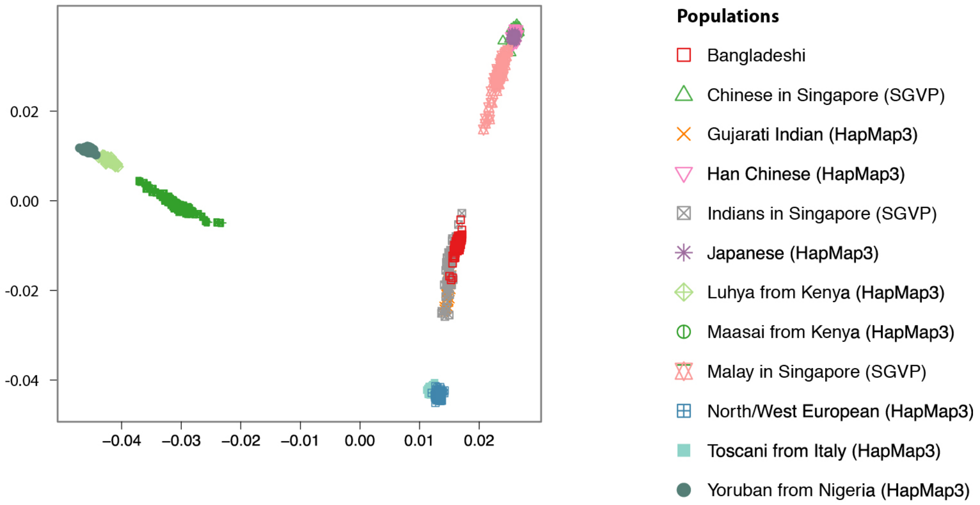
Phased SNP data from the 1000 Genomes Project was from March 2010 release (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/pilot_data/release/2010_03/pilot1/). We tested for signals of natural selection in the BEB, Yoruba (YRI), North/West Europeans (CEU) and East Asians/Han Chinese (CHB+JPT) using CMS_{GW}, which combines posterior probabilities from 5 tests for selection into a composite likelihood score based on empirical distributions from simulations (24). To accommodate four populations, we modified the cross-population component tests to analyze three population pairs. We used likelihood tables developed for comparing highly differentiated populations (YRI, CEU, CHB+JPT), as CMS is robust to large variations in demography (23) and any skew from inaccurate demography will be conservative in BEB, minimizing false positives. Genes were mapped to selected regions using the RefSeqGene gene catalog (UCSC Genome Brower; hg18) and the LINC RNA catalog (http://www.broadinstitute.org/genome_bio/human_lincrnas/) (76-78).

Gene set enrichment testing

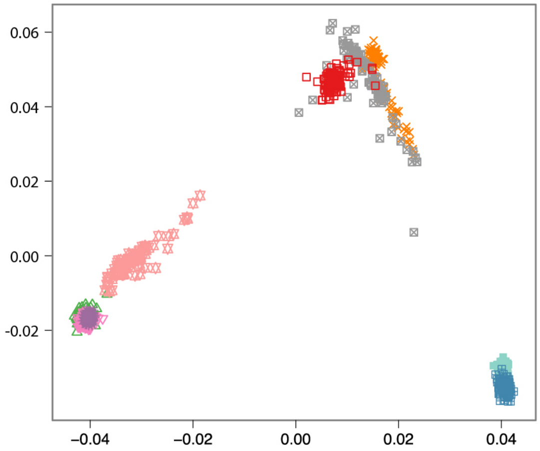
We tested for enrichment in gene set catalogs from MSigDB v3.0 (http://www.broadinstitute.org/gsea/msigdb/download_file.jsp?filePath=/resources/msigdb/3.0/c4.all.v3.0.orig.gmt) and Gene Ontology (downloaded from http://www.ensembl.org/biomart). Blood group gene sets were made for all blood systems in NCBI's Blood Group Antigen Gene Mutation Database (http://www.ncbi.nlm.nih.gov/projects/gv/mhc/xslcgi.cgi?cmd=bgmut/systems). Potassium channel gene sets were generated from categories described in (38). For the IBD analysis, we tested four gene sets based on supplementary table in (60), two each for ulcerative colitis and Crohn's disease: one of all regions and one of regions with pGWAS<1x10-10.

Fig. S1. Principle Component Analysis of Bangladesh, HapMap3 and Singapore populations.

A. Bangladeshi with HapMap3 and SGVP



B. Bangladeshi with non-African HapMap3 and SGVP



C. Bangladeshi with Asian populations from Hapmap3 and SGVP

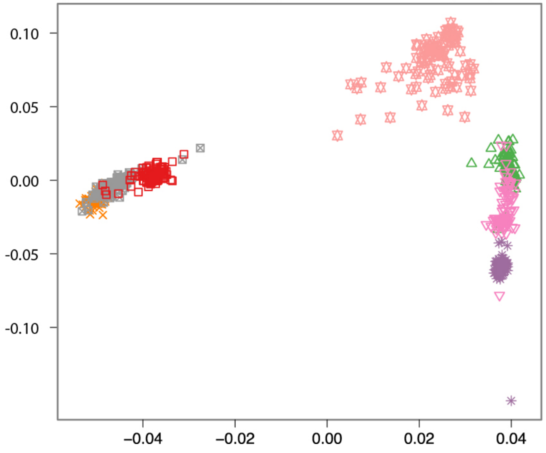


Fig. S2. Comparing *ROLLOFF* fit of models with 1 or 2 admixture events.

ROLLOFF analysis of decay in linkage disequilbrium in BEB, comparing 1 and 2 admixture event models (parental populations from India and East Asia). The fit of a two event admixture model (blue line) to our data (grey crosses) obtains events dated 250 and 43 generations ago, versus 52 generations for the single event model (red dashed line). There is little difference in fit between the two models and we find no statisical support for the more ancient flow.

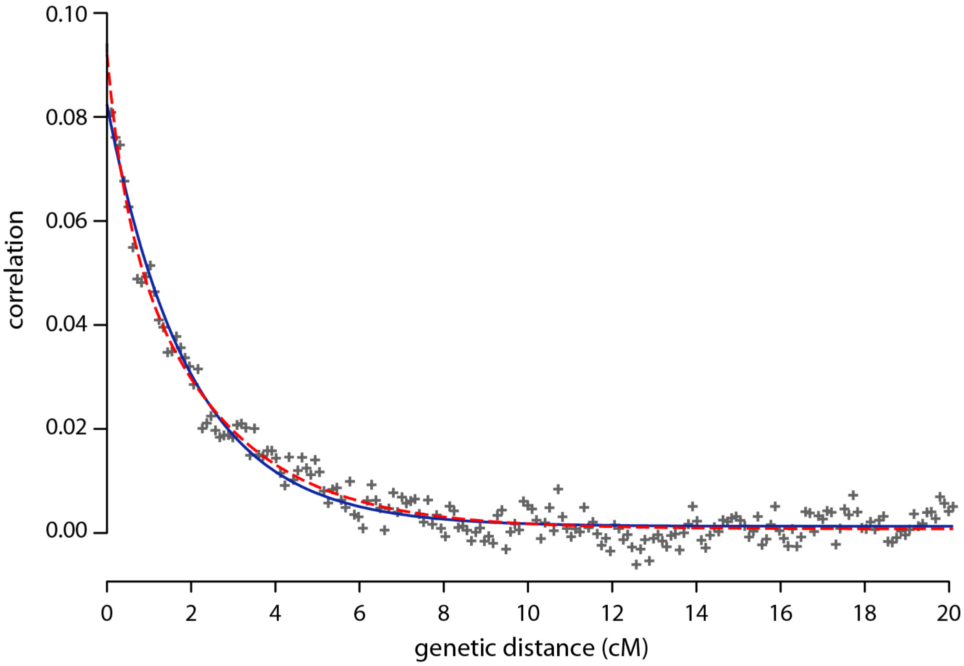


Fig. S3. ADMIXTURE modeling with increasing numbers of ancestral populations

ADMIXTURE maximum likelihood ancestry estimation at variable K (number of ancestral populations). (a) Individual ancestry sorted by population of origin (b) Cross validation error rate.

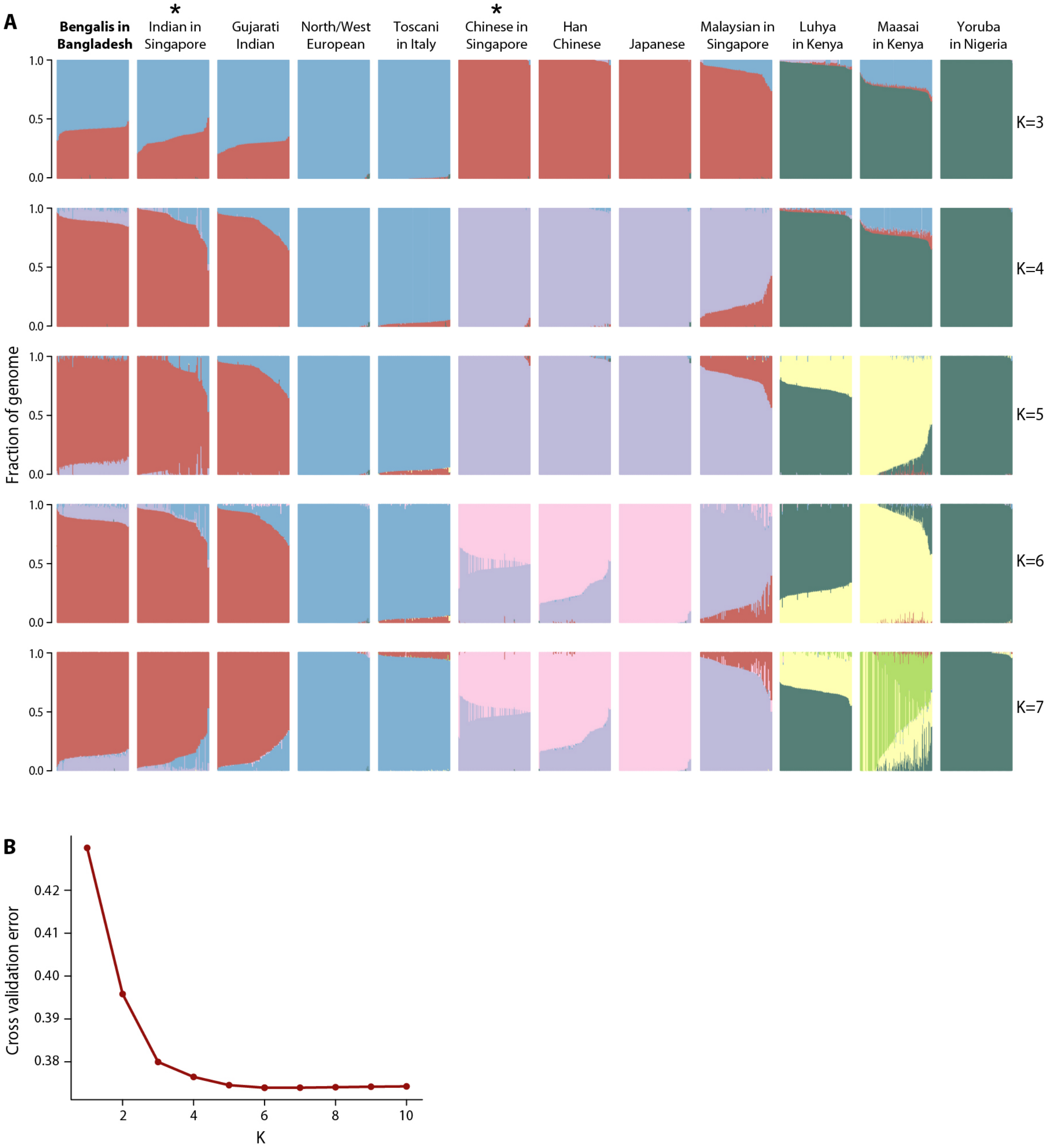


Fig. S4. CMS_{GW} found no selection at CFTR in the BEB.

CMS_{GW} detected no significant evidence for selection at CFTR in the BEB population. There is a strong signal approximately 2.5 Mb from CFTR, peaking at the 5' end of KCND2, a member of the voltage gated, calcium activated potassium channel gene set enriched for selection in the BEB (Table 1). Dashed line marks normalized CMS_{GW} score of 3 (false positive rate < 0.01%).

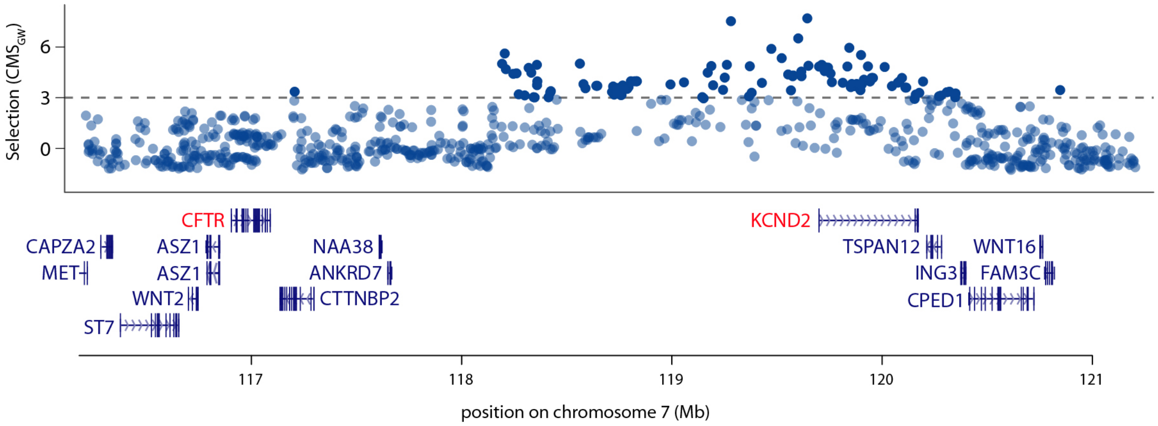


Table S1. Whole Genome SNP Datasets

Whole Genome SNP Datasets. To examine the history of the BEB, and their relationship to other populations, we combined the BEB genome-wide SNP dataset with data from two publically available datasets covering 16 populations: 8 populations from the International Haplotype Map Phase 3 datsaet (HapMap3), 3 from the Singapore Genome Variation Project (SGVP) and 5 from the Human Genome Diversity Project (HGDP). To identify genomic regions with signals of selection using CMS, which includes cross-population tests of selection, we combined the BEB data with dense data for 3 populations from the 1000 Genomes Project (1KG).

				Population Genetics			
Population	**Name**	**Source**	**#**	f3 statistic	Principle Component Analysis	ADMIXTURE (LD pruned)	**CMS selection scan**
Bengalis from Dhaka, Bangladesh	BEB	Illumina 1M Genotyping Array	108	X	X	X	X
Gujarati Indians in Houston, Texas	GIH	HapMap3	88		X	X	
Utah residents with Northern and Western European ancestry from the CEPH collection	CEU	HapMap3	165		X	X	
Japanese in Tokyo, Japan	JPT	HapMap3	86		X	X	
Han Chinese in Beijing, China	CHB	HapMap3	84		X	X	
Luhya in Webuye, Kenya	LWK	HapMap3	90		X	X	
Maasai in Kinyawa, Kenya	MKK	HapMap3	171		X	X	
Toscani in Italia	TSI	HapMap3	88		X	X	
Yoruba in Ibadan, Nigeria	YRI	HapMap3	167		X	X	
Chinese in Singapore	CHS	SGVP	96	X	X	X	
Indians in Singapore	INS	SGVP	83	X	X	X	
Malay in Singapore	MAS	SGVP	89	X	X	X	
Chinese Dai in Yunnan Province, China	Dai	HGDP	10	X			
French Caucasions in Strasbourg, Grenoble, Marseille, Lille, Nantes, and Paris	French	HGDP	28	X			
Chinese Han in Shanghai, China	Han	HGDP	44	X			
Pathan muslims in Hyderabad, Andra Pradesh, India.	Pathan	HGDP	22	X			
Sindhi tribal group in Southern Pakistan	Sindhi	HGDP	24	X			
CEPH: Utah residents with Northern and Western European ancestry	CEU	1KG	60				X
Japanese in Tokyo, Japan and Han Chinese in Beijing, China	JPT + CHB	1KG	60				X
Yoruba in Ibadan, Nigeria	YRI	1KG	60				X

Number of populations	9	12	12	4
Number of SNPs	98,429	567,213	121,920	795,517

Table S2. F_{ST} between Bangladesh and HapMap3 populations

Population differentiation, measured as Fst, between Bangladesh and 8 HapMap3 populations

	Bangladesh	Gujarati Indian	Toscani from Italy	North/West European	Han Chinese	Japanese	Maasai from Kenya	Luhya from Kenya
Gujarati Indian	0.0047							
Toscani from Italy	0.0409	0.0345						
North/West European	0.0418	0.0351	0.0038					
Han Chinese	0.0594	0.0763	0.1112	0.1105				
Japanese	0.0614	0.0777	0.1132	0.1119	0.0068			
Maasai from Kenya	0.0953	0.0958	0.0990	0.1045	0.1431	0.1444		
Luhya from Kenya	0.1329	0.1344	0.1435	0.1477	0.1780	0.1797	0.0170	
Yoruban from Nigeria	0.1441	0.1458	0.1555	0.1594	0.1872	0.1890	0.0271	0.0078

Table S3. 3 Population Test

f_3 statistics

Pop 1	Pop 2	Pop 3	f3 score	Std Error	Z
CHS	**INS**	**BEB**	**-0.005469**	**0.000113**	**-48.43**
Han	INS	BEB	-0.005506	0.000114	-48.28
Dai	INS	BEB	-0.005586	0.000122	-45.76
CHS	Sindhi	BEB	-0.007959	0.000206	-38.69
Han	Sindhi	BEB	-0.00803	0.000208	-38.64
Dai	Sindhi	BEB	-0.008023	0.000216	-37.09
CHS	Pathan	BEB	-0.007558	0.000214	-35.36
Han	Pathan	BEB	-0.007595	0.000216	-35.11
Dai	Pathan	BEB	-0.007951	0.000231	-34.46
Dai	French	BEB	-0.00945	0.000347	-27.22
CHS	French	BEB	-0.008851	0.000332	-26.69
French	Han	BEB	-0.008816	0.000338	-26.07
French	INS	BEB	0.003412	0.000116	29.33
INS	Pathan	BEB	0.002677	0.000077	34.59
INS	Sindhi	BEB	0.002716	0.000076	35.93
Pathan	Sindhi	BEB	0.008658	0.000149	58.00
French	Sindhi	BEB	0.01372	0.000227	60.32
French	Pathan	BEB	0.016954	0.000262	64.74
Dai	Han	BEB	0.056847	0.000553	102.77
CHS	Dai	BEB	0.05719	0.000543	105.40
CHS	Han	BEB	0.059387	0.000541	109.76

Table S4. Ancestry analysis with ADMIXTURE.

	Cluster 1		**Cluster 2**		**Cluster 3**		**Cluster 4**		**Cluster 5**	
Population	*avg*	*std. dev*	*avg*	*std. dev*	*avg*	*std. dev*	*avg*	*std. dev*	*avg*	*std. dev*
Bangladesh	0.897	0.024	0.093	0.026	0.009	0.013	0.000	0.002	0.000	0.000
North/West European	0.001	0.002	0.000	0.002	0.998	0.006	0.001	0.004	0.000	0.000
Han Chinese	0.001	0.004	0.995	0.010	0.004	0.008	0.000	0.003	0.000	0.000
Chinese in Singapore	0.002	0.009	0.998	0.010	0.000	0.001	0.000	0.002	0.000	0.000
Gujarati Indian	0.876	0.084	0.001	0.004	0.123	0.084	0.000	0.001	0.000	0.000
Indians in Singapore	0.906	0.095	0.017	0.036	0.077	0.094	0.000	0.000	0.000	0.001
Japanese	0.000	0.000	0.999	0.007	0.001	0.003	0.000	0.004	0.000	0.000
Luhya from Kenya	0.000	0.000	0.001	0.003	0.002	0.008	0.734	0.033	0.263	0.036
Malay in Singapore	0.176	0.070	0.822	0.072	0.001	0.005	0.000	0.001	0.000	0.002
Maasai from Kenya	0.008	0.017	0.000	0.001	0.003	0.008	0.078	0.094	0.911	0.097
Toscani from Italy	0.035	0.012	0.000	0.002	0.963	0.012	0.000	0.001	0.002	0.006
Yoruban from Nigeria	0.000	0.001	0.000	0.001	0.000	0.003	0.999	0.004	0.000	0.002

Table S5. Candidate selected regions in BEB population

#	Location (hg18)	Size (kb)	Max CMS score	# Genes	Genes
1	chr2:96204472-96450166	246	12.66	7	STARD7, LOC285033, TMEM127, CIAO1, ASCC3L1, KIAA1754L, NCAPH
2	chr1:246860501-247003688	143	9.38	4	OR2T35, OR2T27, OR14I1, LOC646627
3	chr20:47858175-48083964	226	9.30	4	SLC9A8, SPATA2, RNF114, SNAI1
4	chr10:45284677-45384677	100	9.02	1	MARCH8
5	chr16:30312959-31301270	988	8.97	40	ZNF553, ZNF771, XTP3TPA, SEPHS2, ITGAL, ZNF768, ZNF747, ZNF764, ZNF688, ZNF785, ZNF689, PRR14, FBRS, SRCAP, PHKG2, LOC90835, RNF40, ZNF629, BCL7C, CTF1, FBXL19, ORAI3, SETD1A, HSD3B7, STX1B, STX4, ZNF668, ZNF646, POL3S, VKORC1, BCKDK, MYST1, PRSS8, PRSS36, FUS, PYCARD, TRIM72, PYDC1, ITGAM, ITGAX
6	chr8:8110844-8211017	100	8.77	0	*FLJ10661*
7	chr2:72556447-72934657	378	8.30	1	EXOC6B
8	chr2:163254213-163942575	688	7.74	1	KCNH7
9	chr7:119120356-120210059	1090	7.70	1	KCND2
10	chr14:62443764-63110864	667	7.57	4	KCNH5, RHOJ, GPHB5, PPP2R5E
11	chr7:65591573-65901445	310	7.52	2	KCTD7, RABGEF1
12	chr11:129367548-129564791	197	7.37	3	PRDM10, APLP2, ST14
13	chr17:56212699-56662837	450	7.32	2	BCAS3, LOC729617
14	chr2:97581923-97702667	121	7.20	3	COX5B, ACTR1B, ZAP70
15	chr1:35196775-36378774	1182	7.16	17	ZMYM6, ZMYM1, SFPQ, ZMYM4, KIAA0319L, NCDN, TFAP2E, PSMB2, C1orf216, CLSPN, EIF2C4, EIF2C1, EIF2C3, TEKT2, ADPRHL2, COL8A2, TRAPPC3
16	chr22:45116361-45222092	106	7.02	2	TRMU, CELSR1
17	chr5:117332265-117961872	630	6.98	0	
18	chr13:56776369-56906916	131	6.96	0	
19	chr13:56441720-56653368	212	6.88	1	FLJ40296
20	chr2:96476699-96981116	504	6.80	11	LINCR, ARID5A, KIAA1310, FER1L5, LMAN2L, CNNM4, CNNM3, ANKRD23, ANKRD39, SEMA4C, LOC51252
21	chr2:60957829-61241691	284	6.75	4	REL, PUS10, PEX13, KIAA1841
22	chr7:65192057-65327158	135	6.74	3	ASL, RCP9, TPST1
23	chr4:41500330-41897307	397	6.73	4	TMEM33, WDR21B, SLC30A9, CCDC4
24	chr2:97787810-98033835	246	6.72	1	TMEM131
25	chr3:48649603-49077789	428	6.46	13	CELSR3, NCKIPSD, IHPK2, PRKAR2A, SLC25A20, C3orf71, ARIH2, PH-4, WDR6, DALRD3, C3orf60, IMPDH2, QRICH1
26	chr1:205127559-205227559	100	6.45	4	IL24, FAIM3, PIGR, FCAMR
27	chr10:22313652-22450452	137	6.45	1	DNAJC1
28	chr15:81373168-81591756	219	6.43	5	HOMER2, FAM103A1, C15orf40, BTBD1, TM6SF1

29	chr11:87911055-88223280	312	6.43	1	GRM5
30	chr3:87812428-88370387	558	6.42	4	HTR1F, CGGBP1, ZNF654, C3orf38
31	chr11:66560041-67041927	482	6.41	21	SYT12, RHOD, FBXL11, ADRBK1, ANKRD13D, SSH3, POLD4, CLCF1, RAD9A, PPP1CA, TBC1D10C, KIAA1394, RPS6KB2, PTPRCAP, CORO1B, GPR152, CABP4, TMEM134, AIP, PITPNM1, CDK2AP2
32	chr12:55794772-56187449	393	6.41	11	LRP1, NXPH4, SHMT2, NDUFA4L2, STAC3, R3HDM2, INHBC, INHBE, GLI1, ARHGAP9, MARS
33	chr2:205884483-206042496	158	6.35	1	PARD3B
34	chr6:105237478-105587627	350	6.21	2	HACE1, LIN28B
35	chr16:31622282-31775062	153	6.17	0	
36	chr2:72171698-72390351	219	6.16	2	CYP26B1, EXOC6B
37	chr8:99560993-100619301	1058	6.14	3	STK3, OSR2, VPS13B
38	chr17:59873629-60064061	190	6.06	5	C17orf60, POLG2, DDX5, CCDC45, SMURF2
39	chr2:74441365-74643702	202	6.05	18	DCTN1, WDR54, RTKN, ZNHIT4, WBP1, GCS1, MRPL53, CCDC142, TTC31, LBX2, PCGF1, TLX2, DQX1, AUP1, HTRA2, LOXL3, DOK1, C2orf65
40	chr17:60723634-60909402	186	6.04	0	
41	chr15:46163776-46282679	119	6.01	2	SLC24A5, MYEF2
42	chr3:142211453-142490751	279	5.83	2	SPSB4, ACPL2
43	chr1:165476379-165658590	182	5.79	1	POU2F1
44	chr4:106703606-107031286	328	5.74	3	FLJ20184, INTS12, GSTCD
45	chr8:66849249-67123775	275	5.70	2	PDE7A, DNAJC5B
46	chr6:131753370-131901651	148	5.69	0	
47	chr4:29727693-29948439	221	5.68	0	
48	chr6:73745382-74032917	288	5.67	3	KCNQ5, LOC100129128, KHDC1
49	chr2:82761461-82987356	226	5.66	0	
50	chr4:29336821-29558398	222	5.65	0	
51	chr10:118088143-118261246	173	5.65	2	C10orf96, PNLIPRP3
52	chr11:122671246-122799152	128	5.63	0	
53	chr10:130932072-131087143	155	5.62	0	
54	chr7:73664429-73827987	164	5.62	2	GTF2I, NCF1
55	chr13:105802626-105906075	103	5.61	0	
56	chr7:118155039-118409989	255	5.61	0	
57	chr6:105618813-105799829	181	5.60	3	LIN28B, BVES, POPDC3
58	chr7:111175434-111355649	180	5.60	1	DOCK4
59	chr11:66031159-66257760	227	5.54	11	DPP3, BBS1, ZDHHC24, ACTN3, CTSF, CCDC87, CCS, RBM14, RBM4, RBM4B, SPTBN2
60	chr19:11395725-11567995	172	5.53	8	CCDC151, PRKCSH, ELAVL3, ZNF653, ECSIT, CNN1, ELOF1, ACP5
61	chr17:55510921-55769951	259	5.53	3	HEATR6, CA4, USP32
62	chr7:148319871-148502995	183	5.51	4	PDIA4, ZNF786, ZNF425, ZNF398
63	chr1:191292717-191514030	221	5.49	5	UCHL5, TROVE2, GLRX2, CDC73, B3GALT2
64	chr12:87733005-87835702	103	5.48	0	

65	chr17:50954728-51293371	339	5.45	2	TMEM100, PCTP
66	chr15:81648149-81778085	130	5.42	2	HDGFRP3, BNC1
67	chr1:831808-958247	126	5.41	7	SAMD11, NOC2L, KLHL17, PLEKHN1, HES4, ISG15, AGRN
68	chr15:28888490-29104439	216	5.40	3	MTMR15, MTMR10, TRPM1
69	chr12:48010827-48407359	397	5.32	9	TROAP, C1QL4, DNAJC22, SPATS2, KCNH3, MCRS1, C12orf25, PRPF40B, FMNL3
70	chr9:95381811-95552915	171	5.28	1	PHF2
71	chr15:70521862-70672069	150	5.24	1	ARIH1
72	chr2:9290239-9689347	399	5.23	6	DDEF2, ITGB1BP1, CPSF3, IAH1, ADAM17, YWHAQ
73	chr13:31010903-31110903	100	5.22	0	
74	chr12:20927650-21027650	100	5.19	1	SLCO1B3
75	chr3:168770464-169143547	373	5.19	3	WDR49, PDCD10, SERPINI1
76	chr2:237651106-237828782	178	5.19	1	COPS8
77	chr15:69912174-70024573	112	5.18	1	MYO9A
78	chr1:6234024-6410129	176	5.17	4	GPR153, ACOT7, HES2, ESPN
79	chr8:106772233-106956416	184	5.17	1	ZFPM2
80	chr10:134957945-135142863	185	5.17	9	TUBGCP2, ZNF511, CALY, PRAP1, C10orf125, ECHS1, PAOX, MTG1, SPRN
81	chr1:26030965-26172646	142	5.14	5	FAM54B, C1orf135, PAQR7, STMN1, PAFAH2
82	chr11:66268824-66415567	147	5.14	4	C11orf80, RCE1, PC, LRFN4
83	chr2:197503694-197779960	276	5.14	1	ANKRD44
84	chr18:22014765-22160306	146	5.13	2	PSMA8, TAF4B
85	chr17:17478246-17960954	483	5.13	8	RAI1, SREBF1, TOM1L2, LRRC48, ATPAF2, C17orf39, DRG2, MYO15A
86	chr2:190352420-190730503	378	5.08	4	ORMDL1, PMS1, MSTN, MGC13057
87	chr19:62920241-63020241	100	5.07	4	ZNF671, ZNF776, ZNF586, ZNF552
88	chr21:16749665-16967157	217	5.07	1	C21orf34
89	chr14:61795612-61955845	160	5.07	0	
90	chr20:9441622-9574643	133	5.06	2	C20orf103, PAK7
91	chr20:53080729-53221407	141	5.03	0	
92	chr7:142353741-142588076	234	4.98	4	KEL, OR9A2, OR6V1, PIP
93	chr2:205419045-205599057	180	4.96	1	PARD3B
94	chr9:95125501-95354732	229	4.95	3	C9orf129, FAM120AOS, FAM120A
95	chr10:96367941-96620192	252	4.95	2	CYP2C18, CYP2C19
96	chr3:112262597-112423795	161	4.95	1	PVRL3
97	chr2:125891457-126138821	247	4.94	0	
98	chr13:40891670-40991670	100	4.92	1	C13orf15
99	chr6:31703901-31815903	112	4.92	16	BAT2, BAT3, APOM, C6orf47, BAT4, CSNK2B, LY6G5B, LY6G5C, BAT5, LY6G6F, LY6G6D, LY6G6C, C6orf25, DDAH2, CLIC1, MSH5
100	chr18:38146306-38256332	110	4.91	0	
101	chr3:50822404-50946281	124	4.90	1	DOCK3
102	chr3:198725887-198830355	104	4.90	1	BDH1
103	chr10:22469942-22628180	158	4.89	0	
104	chr8:134987966-135274918	287	4.88	0	
105	chr16:34094548-34194548	100	4.87	0	
106	chr13:91173009-91438176	265	4.86	1	GPC5

107	chr10:23143356-23337383	194	4.85	1	ARMC3
108	chr1:103167085-103320617	154	4.85	1	COL11A1
109	chr2:190839830-190988591	149	4.84	3	HIBCH, INPP1, FLJ20160
110	chr16:85999328-86099328	100	4.84	1	ZCCHC14
111	chr12:125532855-125747401	215	4.81	0	
112	chr3:199159015-199278595	120	4.80	3	IQCG, RPL35A, LMLN
113	chr2:158925179-159108381	183	4.80	2	CCDC148, PKP4
114	chr12:81799661-81996865	197	4.78	1	TMTC2
115	chr6:101036227-101352801	317	4.76	1	ASCC3
116	chr21:30003605-30103605	100	4.75	1	GRIK1
117	chr16:27471237-27690930	220	4.74	1	KIAA0556
118	chr1:232733129-232854961	122	4.74	1	IRF2BP2
119	chr12:87335813-87636557	301	4.71	1	KITLG
120	chr2:219185249-219313817	129	4.71	6	PLCD4, ZNF142, BCS1L, RNF25, STK36, TTLL4
121	chr16:46883278-46983278	100	4.70	2	LONP2, SIAH1
122	chr11:121017926-121238991	221	4.69	0	
123	chr17:36911424-37068570	157	4.69	7	KRT13, KRT15, KRT19, KRT9, KRT14, KRT16, KRT17
124	chr17:46560973-46712763	152	4.69	5	NME1, NME1-NME2, NME2, MBTD1, UTP18
125	chr8:107183154-107368735	186	4.68	1	OXR1
126	chr3:88405026-88505026	100	4.67	0	
127	chr15:58458349-58632099	174	4.64	3	ANXA2, NARG2, RORA
128	chr2:213045634-213249128	203	4.64	1	ERBB4
129	chr16:45091599-45191599	100	4.63	1	SHCBP1
130	chr4:171685791-172094831	409	4.63	0	
131	chr16:46765699-46865699	100	4.62	2	ABCC11, LONP2
132	chr17:55779097-55996815	218	4.61	3	USP32, C17orf64, APPBP2
133	chr16:49929755-50081366	152	4.60	0	
134	chr16:34895632-34995632	100	4.59	0	
135	chr2:88026347-88175133	149	4.58	2	KRCC1, SMYD1
136	chr9:122799711-123110846	311	4.58	4	C5, CEP110, RAB14, GSN
137	chr3:106206655-106339543	133	4.58	0	
138	chr15:27077239-27177239	100	4.57	1	APBA2
139	chr2:125708515-125837346	129	4.56	0	
140	chr1:67139020-67309707	171	4.56	3	WDR78, MIER1, SLC35D1
141	chr12:86937582-87147672	210	4.56	4	C12orf50, C12orf29, CEP290, TMTC3
142	chr4:80136407-80445408	309	4.54	0	
143	chr8:71606476-71773595	167	4.54	3	TRAM1, LACTB2, XKR9
144	chr9:10954145-11061845	108	4.53	0	
145	chr1:55296952-55452792	156	4.52	2	PCSK9, USP24
146	chr11:88240385-88359000	119	4.52	1	GRM5
147	chr20:38422242-38536905	115	4.51	0	
148	chr1:64098164-64256265	158	4.51	1	ROR1
149	chr2:73489360-73803227	314	4.50	3	ALMS1, NAT8, NAT8B
150	chr16:79497360-79687142	190	4.50	5	C16orf61, CENPN, ATMIN, C16orf46, GCSH
151	chr3:166926418-167239549	313	4.47	1	BCHE
152	chr20:29940726-30118633	178	4.47	5	TTLL9, PDRG1, XKR7, C20orf160, HCK

153	chr14:68552144-68697749	146	4.47	1	WDR22
154	chr13:44533690-44732263	199	4.46	2	GTF2F2, KCTD4
155	chr4:33318325-33492559	174	4.45	0	
156	chr2:229052479-229239730	187	4.45	0	
157	chr9:125381089-125552528	171	4.45	1	DENND1A
158	chr6:136668671-137001357	333	4.44	2	MAP7, MAP3K5
159	chr2:158207988-158364390	156	4.43	1	ACVR1
160	chr6:118228907-118339350	110	4.39	1	SLC35F1
161	chr5:170280848-170623067	342	4.39	1	RANBP17
162	chr3:176856725-176995567	139	4.38	1	NAALADL2
163	chr6:79579015-79892326	313	4.38	2	IRAK1BP1, PHIP
164	chr3:44202343-44493069	291	4.37	2	C3orf23, ZNF445
165	chr4:170027542-170127542	100	4.37	1	PALLD
166	chr15:70290476-70415126	125	4.37	3	PKM2, PARP6, BRUNOL6
167	chr15:33567373-33667373	100	4.34	1	ATPBD4
168	chr7:36976683-37080082	103	4.34	1	ELMO1
169	chr10:83991355-84355739	364	4.34	1	NRG3
170	chr6:129261105-129446944	186	4.33	1	LAMA2
171	chr15:89552911-89652911	100	4.33	1	SV2B
172	chr15:41438222-41666300	228	4.32	5	ZSCAN29, TUBGCP4, TP53BP1, MAP1A, HISPPD2A
173	chr17:71298737-71503745	205	4.32	9	UNK, UNC13D, WBP2, TRIM47, TRIM65, MRPL38, FBF1, ACOX1, LOC100134934
174	chr2:157816760-157920279	104	4.32	2	GALNT5, ERMN
175	chr14:56687476-57040198	353	4.30	4	EXOC5, C14orf108, NAT12, C14orf105
176	chr1:93778867-94154851	376	4.29	4	FNBP1L, BCAR3, DNTTIP2, GCLM
177	chr2:121119134-121239271	120	4.28	0	
178	chr15:73448936-73689839	241	4.28	3	SIN3A, PTPN9, SNUPN
179	chr17:59007879-59138788	131	4.27	5	WDR68, CCDC44, MAP3K3, LIMD2, LYK5
180	chr4:33499980-33684195	184	4.26	0	
181	chr2:192920099-193020099	100	4.25	0	
182	chr1:233362909-233555310	192	4.24	2	RBM34, ARID4B
183	chr11:60470408-60570408	100	4.22	2	SLC15A3, CD6
184	chr6:34252844-34383909	131	4.22	3	HMGA1, C6orf1, NUDT3
185	chr2:82565084-82721859	157	4.21	0	
186	chr2:22259026-22403595	145	4.20	0	
187	chr4:12010385-12121961	112	4.17	0	
188	chr7:106753657-106867115	113	4.17	1	COG5
189	chr1:54274463-54374463	100	4.16	2	TMEM59, C1orf83
190	chr1:233866562-234087074	221	4.14	2	GNG4, LYST
191	chr11:133528364-133677671	149	4.14	5	NCAPD3, VPS26B, THYN1, ACAD8, GLB1L3
192	chr18:64760381-64945268	185	4.11	1	CCDC102B
193	chr4:114419346-114570349	151	4.10	1	ANK2
194	chr2:183438653-183571657	133	4.07	2	FRZB, NCKAP1
195	chr5:109940006-110089526	150	4.07	1	FLJ43080
196	chr17:56091407-56191407	100	4.06	2	PPM1D, BCAS3
197	chr1:64965446-65065446	100	4.06	1	RAVER2
198	chr6:34675122-34880880	206	4.04	3	C6orf106, SNRPC, UHRF1BP1
199	chr2:98152695-98345361	193	4.04	2	VWA3B, CNGA3

200	chr1:93075257-93175257	100	4.04	2	RPL5, FAM69A
201	chr8:71788747-71929582	141	4.03	1	XKR9
202	chr16:46236842-46424754	188	4.03	1	PHKB
203	chr12:65737418-65892386	155	4.03	0	
204	chr4:123581338-123726077	145	4.02	1	IL2
205	chr2:94888725-94988725	100	4.02	1	TEKT4
206	chr9:90820420-90933940	114	4.02	1	SHC3
207	chr1:246296107-246396107	100	4.02	2	OR2L13, OR2M5
208	chr18:65894434-66044084	150	4.01	1	RTTN
209	chr12:107674927-107779318	104	4.01	1	SSH1
210	chr7:86141717-86262469	121	4.01	1	GRM3
211	chr4:12139855-12245691	106	4.01	0	
212	chr3:31257160-31357160	100	4.00	0	
213	chr3:50628862-50797668	169	4.00	2	MAPKAPK3, DOCK3
214	chr10:21974224-22074224	100	4.00	1	MLLT10
215	chr5:99762015-99977675	216	4.00	1	FAM174A
216	chr21:37503499-37619014	116	3.99	1	DSCR3
217	chr7:118665467-118883416	218	3.97	0	
218	chr11:120721267-120821267	100	3.96	0	
219	chr3:108823269-108937391	114	3.96	1	BBX
220	chr1:153198332-153306921	109	3.94	10	PYGO2, SHC1, CKS1B, FLAD1, LENEP, ZBTB7B, DCST2, DCST1, ADAM15, EFNA4
221	chr3:190084486-190227443	143	3.94	0	
222	chr5:24698962-24836371	137	3.93	0	
223	chr2:21894744-22002427	108	3.93	0	
224	chr5:134071831-134171831	100	3.93	3	SEC24A, CAMLG, DDX46
225	chr11:22547793-22647793	100	3.92	2	FANCF, GAS2
226	chr8:72021632-72167150	146	3.92	0	
227	chr4:70563870-70700911	137	3.91	1	SULT1B1
228	chr9:32630443-32798898	168	3.91	1	TMEM215
229	chr12:55560763-55668008	107	3.88	2	SDR-O, RDH16
230	chr17:4724989-4916850	192	3.87	12	MINK1, CHRNE, GP1BA, SLC25A11, RNF167, PFN1, ENO3, SPAG7, CAMTA2, INCA1, KIF1C, GPR172B
231	chr4:113347053-113447053	100	3.87	3	C4orf16, TIFA, ALPK1
232	chr6:7034815-7166406	132	3.87	1	RREB1
233	chr3:49121136-49361236	240	3.87	8	USP19, LAMB2, CCDC71, KLHDC8B, LOC646498, CCDC36, C3orf62, USP4
234	chr1:116397280-116513106	116	3.86	2	SLC22A15, C1orf161
235	chr11:76655222-76755222	100	3.86	2	GDPD4, PAK1
236	chr12:18704919-18804919	100	3.86	2	PLCZ1, CAPZA3
237	chr3:76034636-76192189	158	3.85	0	
238	chr2:191934437-192037572	103	3.85	1	MYO1B
239	chr4:151481457-151585884	104	3.84	1	LRBA
240	chr4:38882719-39070335	188	3.84	2	WDR19, RFC1
241	chr2:81671643-81824573	153	3.83	0	
242	chr1:92825740-92961962	136	3.83	1	EVI5
243	chr5:21089799-21201056	111	3.83	0	

244	chr18:9057562-9218306	161	3.82	2	NDUFV2, ANKRD12
245	chr13:19255353-19390879	136	3.80	1	ZMYM5
246	chr15:43096138-43196138	100	3.80	3	SORD, DUOX2, DUOXA2
247	chr3:50159130-50275841	117	3.78	4	SEMA3F, GNAT1, SLC38A3, GNAI2
248	chr2:123277111-123377111	100	3.77	0	
249	chr10:510172-774229	264	3.77	1	DIP2C
250	chr8:100844697-101001323	157	3.75	2	VPS13B, COX6C
251	chr6:15259758-15367176	107	3.74	1	JARID2
252	chr5:11778395-11904562	126	3.74	1	CTNND2
253	chr5:30561813-30715403	154	3.73	0	
254	chr15:61652432-61832476	180	3.71	3	USP3, FBXL22, HERC1
255	chr5:111192852-111292852	100	3.71	0	
256	chr12:78400289-78560786	160	3.70	1	PAWR
257	chr6:126646936-127062588	416	3.68	1	C6orf173
258	chr1:178599787-178711103	111	3.68	1	ACBD6
259	chr5:99559919-99698857	139	3.67	0	
260	chr2:63997751-64143531	146	3.67	1	VPS54
261	chr12:42491863-42614756	123	3.67	1	TMEM117
262	chr11:60782100-60986304	204	3.66	8	VWCE, DDB1, DAK, CYBASC3, TMEM138, TMEM216, FLJ12529, C11orf79
263	chr10:86943502-87111961	168	3.66	0	
264	chr10:65531300-65661430	130	3.64	0	
265	chr3:38418216-38557515	139	3.64	3	XYLB, ACVR2B, ENDOGL1
266	chr6:58625121-58725121	100	3.63	0	
267	chr10:24098364-24213025	115	3.62	1	KIAA1217
268	chr11:91680137-91805571	125	3.62	1	FAT3
269	chr16:2081735-2185557	104	3.62	4	PKD1, RAB26, TRAF7, CASKIN1
270	chr2:73860072-73962402	102	3.59	3	DUSP11, FLJ43987, STAMBP
271	chr2:17703186-17849261	146	3.59	2	SMC6, GEN1
272	chr20:35382039-35482039	100	3.59	1	SRC
273	chr1:146183191-146298908	116	3.59	0	
274	chr2:148559224-148681014	122	3.57	0	
275	chr9:130084933-130226839	142	3.57	6	C9orf119, TRUB2, COQ4, SLC27A4, URM1, CERCAM
276	chr11:20341684-20488999	147	3.57	2	HTATIP2, PRMT3
277	chr14:106059588-106159588	100	3.56	0	
278	chr12:12171446-12293333	122	3.54	2	BCL2L14, LRP6
279	chr2:21693084-21793084	100	3.54	0	
280	chr16:60649093-60749093	100	3.53	0	
281	chr3:49984343-50140249	156	3.52	2	RBM6, RBM5
282	chr19:44458940-44603354	144	3.52	8	IL29, LRFN1, GMFG, SAMD4B, PAF1, MED29, ZFP36, PLEKHG2
283	chr17:53374925-53490256	115	3.51	2	VEZF1, SFRS1
284	chr20:7035757-7140997	105	3.48	0	
285	chr2:190184469-190320953	136	3.47	3	ASNSD1, ANKAR, OSGEPL1
286	chr6:77092051-77203728	112	3.47	0	
287	chr10:59251617-59351617	100	3.46	0	
288	chr22:27333003-27433003	100	3.44	1	CHEK2
289	chr20:30948926-31048926	100	3.39	1	SPAG4L

290	chr2:177707009-177844073	137	3.37	2	HNRNPA3, NFE2L2
291	chr7:120220572-120399549	179	3.36	2	TSPAN12, ING3
292	chr1:191573307-191700187	127	3.33	0	
293	chr13:19031017-19161340	130	3.29	1	MPHOSPH8
294	chr16:14435797-14566760	131	3.27	1	PARN
295	chr4:119132533-119269939	137	3.27	1	NDST3
296	chr1:92554157-92713125	159	3.27	2	RPAP2, GFI1
297	chr1:148102207-148202207	100	3.27	8	HIST2H2BE, HIST2H2AC, HIST2H2AB, BOLA1, SV2A, SF3B4, MTMR11, OTUD7B
298	chr6:101707195-101825358	118	3.25	0	
299	chr7:110897132-111017780	121	3.23	1	IMMP2L
300	chr7:138669713-138769713	100	3.19	2	C7orf55, LUC7L2
301	chr1:210518780-210622483	104	3.11	2	PPP2R5A, TMEM206
302	chr15:28758580-28858580	100	3.11	0	
303	chr16:45855281-45955281	100	3.11	1	ITFG1
304	chr16:46448606-46548606	100	3.10	0	
305	chr1:25510488-25610488	100	3.00	3	RHD, TMEM50A, RHCE

Table S6. Gene content of $\mathrm{CMS}_{\mathrm{GW}}$ regions

Gene counts for candidate regions of natural selection found by CMS

Population	*# Regions*	*Median region size*	*# RefSeq Genes (avg)*	*# RefSeq Genes (sd)*	*0 Genes*	*1 Gene*	*LINC*	*only LINC**
BEB	305	149	2.33	3.65	72	95	91	29
North/West European	233	144	2.01	2.51	55	76	67	23
Japanese and Han Chinese	254	128	1.53	2.05	79	92	71	25
Yoruban from Nigeria	130	125	2.07	2.21	23	45	26	9

* Regions with at least one LINC and no RefSeq genes

Table S7. INRICH gene set enrichment analysis

All INRICH gene sets with $p_{set} < 0.05$ in any of the four CMS populations

	BEB	BEB			CEU			JPT/CHB			YRI		
TARGET	rank	p_{set}	p_{exp}	#	p_{set}	p_{exp}	#	p_{set}	p_{exp}	#	p_{set}	p_{exp}	#
1. Molecular Signatures Database MSigDB.c4 (851 gene sets with > 10 genes)													
MORF_IKBKG (110 genes)	1	0.00005	0.017	15									
module_85 (47 genes)								**0.00009**	**0.027**	**7**			
module_259 (42 genes)								0.00018	0.050	6			
module_199 (53 genes)								0.00019	0.053	7			
MORF_FDXR (198 genes)	2	0.00024	0.074	18				0.02660	0.997	9			
module_514 (10 genes)								0.00028	0.075	3			
MORF_PHB (108 genes)	3	0.00067	0.183	12									
MORF_ORC1L (56 genes)	4	0.00121	0.313	9									
MORF_PPP1CA (144 genes)					0.00164	0.366	10						
GNF2_RFC3 (38 genes)											0.00213	0.402	4
MORF_RAD21 (154 genes)					0.00208	0.443	11						
module_102 (15 genes)											0.00246	0.451	3
MORF_PML (122 genes)	5	0.00209	0.458	12									
GCM_PPM1D (22 genes)	6	0.00219	0.470	5									
module_286 (35 genes)								0.00252	0.485	4			
MORF_CSNK2B (250 genes)					0.00263	0.514	14						
GNF2_MYD88 (54 genes)								0.00319	0.561	5			
module_93 (174 genes)								0.00354	0.600	9			
module_350 (55 genes)					0.00349	0.611	5						
MORF_DDX11 (134 genes)	7	0.00342	0.625	13									
2. Gene Ontology gene sets (2,430 gene sets with > 10 genes)													
GO:0005925 Focal adhesion (101 genes)								0.00062	0.388	9			
GO:0005516 Calmodulin binding (134 genes)											0.00142	0.630	8
GO:0042438 Melanin biosynthetic process (10 genes)					0.00149	0.697	3						
GO:0030672 Synaptic vesicle membrane (43 genes)											0.00178	0.704	4
GO:0034660 NcRNA metabolic process (22 genes)											0.00194	0.734	3
GO:0045777 Positive regulation of blood pressure (21 genes)								0.00177	0.764	3			
GO:0007243 Intracellular protein kinase cascade (88 genes)	1	0.00175	0.772	11									
GO:0008076 Voltage-gated potassium											0.00221	0.775	6

channel complex (89 genes)													
GO:0051593 Response to folic acid (10 genes)	2	0.00198	0.808	3									
GO:0042787 Protein ubiquitination involved in ubiquitin-dependent protein catabolic process (45 genes)					0.00225	0.838	7				0.01911	1.000	4
GO:0016070 RNA metabolic process (249 genes)											0.00305	0.865	10
GO:0016747 Transferase activity, transferring acyl groups other than amino-acyl groups (45 genes)								0.00281	0.886	5			
GO:0000387 Spliceosomal snRNP assembly (25 genes)											0.00332	0.887	3
GO:0004812 Aminoacyl-tRNA ligase activity (37 genes)											0.00333	0.888	4
GO:0005903 Brush border (26 genes)					0.00299	0.904	4						
GO:0048066 Developmental pigmentation (15 genes)					0.00393	0.953	3						
GO:0003779 Actin binding (259 genes)					0.00411	0.958	16						
GO:0016881 Acid-amino acid ligase activity (71 genes)					0.00422	0.962	8						
GO:0007626 Locomotory behavior (72 genes)	3	0.00411	0.964	8									
GO:0006464 Protein modification process (139 genes)					0.00501	0.977	11				0.02329	1.000	6

3. Gene Ontology gene sets (2,430 gene sets with > 10 genes) in regions with CMS > 5

GO:0007626 Locomotory behavior (72 genes)	1	0.00004	0.027	7									
GO:0042438 Melanin biosynthetic process (10 genes)					0.00016	0.109	3						
GO:0048066 Developmental pigmentation (15 genes)					0.00062	0.370	3						
GO:0042572 Retinol metabolic process (12 genes)								0.00213	0.729	2			
GO:0045727 Positive regulation of translation (29 genes)											0.00270	0.796	3
GO:0001523 Retinoid metabolic process (11 genes)								0.00260	0.797	2			
GO:0005844 Polysome (18 genes)					0.00305	0.866	3						
GO:0008286 Insulin receptor signaling pathway (131 genes)					0.00325	0.879	7						
GO:0032420 Stereocilium (16 genes)	2	0.00338	0.889	3									

GO:0006813 Potassium ion transport (153 genes)	3	0.00376	0.915	8									
GO:0019841 Retinol binding (12 genes)								0.00414	0.922	2			
GO:0042800 Histone methyltransferase activity (H3-K4 specific) (11 genes)								0.00440	0.934	2			
GO:0008630 DNA damage response, signal transduction resulting in induction of apoptosis (21 genes)					0.00466	0.944	3						
GO:0042254 Ribosome biogenesis (29 genes)											0.00548	0.962	2
GO:0005249 Voltage-gated potassium channel activity (81 genes)	4	0.00533	0.968	6									
GO:0007266 Rho protein signal transduction (39 genes)	5	0.00555	0.973	4							0.03034	1.000	2
GO:0005096 GTPase activator activity (162 genes)											0.00618	0.974	5
GO:0008236 Serine-type peptidase activity (58 genes)	6	0.00785	0.992	4									
GO:0034660 NcRNA metabolic process (22 genes)											0.00936	0.993	2
GO:0016070 RNA metabolic process (249 genes)											0.00948	0.993	6

4. Potassium channel genes

6/7TM-1P EAG, ERG, ELK (8 genes)	1	0.03170	0.137	3									
6/7TM-1P (47 genes)	2	0.04636	0.177	5									

5. Potassium channel genes in regions with CMS > 5

6/7TM-1P (47 genes)	1	0.00165	0.008	5									
6/7TM-1P EAG, ERG, ELK (8 genes)	2	0.00519	0.030	3									

6. Blood systems

Kell and Kell complex genes (8 genes)	1	0.00444	0.007	3				0.02527	0.029	2			

Table S8. Custom gene sets tested with INRICH

Gene set	N genes	Genes (autosomal)
1. Potassium channel gene sets		
2TM-1P	15	KCNJ1, KCNJ2, KCNJ3, KCNJ4, KCNJ5, KCNJ6, KCNJ8, KCNJ9, KCNJ10, KCNJ11, KCNJ12, KCNJ13, KCNJ14, KCNJ15, KCNJ16
6/7TM-1P (voltage gated K+ channels)	47	KCNA1, KCNA2, KCNA3, KCNA4, KCNA5, KCNA6, KCNA7, KCNA10, KCNB1, KCNC1, KCNC2, KCNC3, KCNC4, KCND2, KCND3, KCNF1, KCNG1, KCNH1, KCNH2, KCNMA1, KCNN1, KCNN2, KCNN3, KCNN4, KCNQ1, KCNQ2, KCNQ3, KCNS1, KCNS2, KCNS3, KCNQ4, KCNB2, KCNH4, KCNH3, KCNG2, KCNV1, KCNH5, KCNQ5, KCNT1, KCNH6, KCNH7, KCNG4, KCNH8, KCNU1, KCNV2, KCNG3, KCNT2
6/7TM-1P EAG, ERG, ELK	8	KCNH1, KCNH2, KCNH4, KCNH3, KCNH5, KCNH6, KCNH7, KCNH8
6/7TM-1P KCNQ	5	KCNQ1, KCNQ2, KCNQ3, KCNQ4, KCNQ5
6/7TM-1P Modifier	10	KCNF1, KCNG1, KCNS1, KCNS2, KCNS3, KCNG2, KCNV1, KCNG4, KCNV2, KCNG3
6/7TM-1P SK/IK (calcium activated)	4	KCNN1, KCNN2, KCNN3, KCNN4
6/7TM-1P Shab	2	KCNB1, KCNB2
6/7TM-1P Shaker	8	KCNA1, KCNA2, KCNA3, KCNA4, KCNA5, KCNA6, KCNA7, KCNA10
6/7TM-1P Shal	2	KCND2, KCND3
6/7TM-1P Shaw	4	KCNC1, KCNC2, KCNC3, KCNC4
6/7TM-1P Slo (calcium activated)	4	KCNMA1, KCNT1, KCNU1, KCNT2
ATM-2P	15	KCNK1, KCNK2, KCNK3, KCNK5, KCNK6, KCNK7, KCNK4, KCNK9, KCNK10, KCNK13, KCNK12, KCNK15, KCNK16, KCNK17, KCNK18
ATM-2P acid sensitive	8	KCNK2, KCNK3, KCNK5, KCNK4, KCNK9, KCNK10, KCNK16, KCNK17
accessory subunits	11	KCNE1, KCNMB1, KCNAB1, KCNAB2, KCNAB3, KCNE2, KCNE3, KCNMB2, KCNE4, KCNMB3, KCNMB4
accessory subunits KCNA associated	3	KCNAB1, KCNAB2, KCNAB3
accessory subunits KCNMA1-associated	4	KCNMB1, KCNMB2, KCNMB3, KCNMB4
accessory subunits KCNQ1-associated	4	KCNE1, KCNE2, KCNE3, KCNE4
2. Blood group systems		
ABO	1	ABO
Chido/Rodgers	4	C4A, C4A, C4B, C4B
Colton	1	AQP1
Cromer	1	CD55
Diego	1	SLC4A1
Dombrock	1	ART4
Duffy	1	DARC
FY	1	DARC
Gerbich	1	GYPC
Gill	1	AQP3
Globoside	1	B3GALNT1
H	2	FUT1, FUT2

I	1	GCNT2
Indian	1	CD44
JK	1	SLC14A1
John Milton Hagen	1	SEMA7A
Junior	1	ABCG2
Kell	1	KEL
Kell and Kell complex genes	8	KEL, XKR3, XKR4, XKR5, XKR6, XKR7, XKR8, XKR9
Kidd	2	SLC14A1, SLC1A1
Knops	1	CR1
Lan	1	ABCB6
Lewis	4	FUT2, FUT3, FUT6, FUT7
Lutheran	2	BCAM, KLF1
Landsteiner-Wiener	1	ICAM4
MNS	3	GYPA, GYPB, GYPE
OK	1	BSG
P1PK and Globoside	2	A4GALT, B3GALNT1
P1PK	1	A4GALT
RAPH	1	CD151
RH	5	RHAG, RHBG, RHCE, RHCG, RHD
Rh-associated glycoprotein	1	RHAG
Rodgers	2	C4A, C4A
Scianna	1	ERMAP
T/Tn	1	C1GALT1
YT (Cartwright)	1	ACHE

Table S9. Regions included in association study.

CMS region	CMS score	region	# SNPs *	# var. SNPs **	# repl. SNP[†]	reason	name (fig 3)	Genes tested for association
305	3.00	chr1:25510488..25610488	8	4		blood group genes	1(1)	RHCE, RHD
26	6.45	chr1:205127559..205227559	23	18	1 / 5	immune related genes	1(2)	FAIM3, FCAMR, IL24, PIGR
72	5.23	chr2:9290239..9689347	24	18	4 / 22	immune related genes	2(1)	ADAM17, IAH1
21	6.75	chr2:60957829..61241691	14	10	3 / 8	immune related genes	2(2)	REL
7	8.30	chr2:72556447..72934657	1	1		among top 10 CMS regions	2(3)	EXOC6B
39	6.05	chr2:74441365..74643702	7	4		MORF IKBKG genes	2(4)	DOK1, LOXL3, PCGF1
1	12.65	chr2:96204472..96450166	65	38	6 / 27	among top 10 CMS regions	2(5)	CIAO1, LOC285033, NCAPH, SNRNP200, STARD7, TMEM127
20	6.80	chr2:96476699..96981116	8	5		immune related genes	2(6)	LMAN2L
8	7.74	chr2:163254213..163942575	20	19	4 / 12	among top 10 CMS regions	2(7)	KCNH7
204	4.02	chr4:123581338..123726077	4	3		immune related genes	4	IL2
99	4.92	chr6:31703901..31815903	27	25		immune related genes	6	BAT2, BAT3, BAT4, BAT5, CSNK2B, LY6G5B, LY6G5C, LY6G6C, LY6G6D
11	7.52	chr7:65591573..65901445	4	4	1	potassium channel gene(s)	7(1)	KCTD7, RABGEF1
9	7.70	chr7:119120356..120210059	2	1		among top 10 CMS regions	7(2)	KCND2
92	4.98	chr7:142353741..142588076	6	2		blood group genes	7(3)	KEL
62	5.51	chr7:148319871..148502995	19	13		immune related genes	7(4)	PDIA4
6	8.77	chr8:8110844..8211017	1	1		among top 10 CMS regions	8(1)	*no genes*
143	4.54	chr8:71606476..71773595	5	4	1 / 3	blood group genes	8(2)	XKR9
136	4.58	chr9:122799711..123110846	14	13		immune related genes	9	C5
4	9.02	chr10:45284677..45384677	33	22		among top 10 CMS regions	10	MARCH8

183	4.22	chr11:60470408..60570408	4	4	2	immune related genes	11(1)	CD6
59	5.54	chr11:66031159..66257760	4	4		immune related genes	11(2)	CTSF
31	6.41	chr11:66560041..67041927	15	7	3 / 13	MORF IKBKG genes	11(3)	PITPNM1, RPS6KB2
10	7.57	chr14:62443764..63110864	35	32		among top 10 CMS regions	14	KCNH5
5	8.97	chr16:30312959..31301270	79	54	5 / 43	among top 10 CMS regions	16(1)	BCKDK, CTF1, ITGAL, ITGAM, ITGAX, MYST1, ORAI3, PYCARD, PYDC1, SEPHS2, TRIM72, ZNF768, ZNF785
131	4.62	chr16:46765699..46865699	5	2		immune related genes	16(2)	ABCC11
60	5.53	chr19:11395725..11567995	9	5		MORF IKBKG genes	19	PRKCSH
152	4.47	chr20:29940726..30118633	4	4		blood group genes	20(1)	XKR7
3	9.30	chr20:47858175..48083964	36	34	2 / 5	among top 10 CMS regions	20(2)	SLC9A8, SNAI1, SPATA2
		random selected SNPs	19	19	11			
		TOTAL	495	370	55 / 204			

* SNPs with genotyping rates > 75% across individuals with genotyping rates > 75%
** SNPs with minor allele frequency > 1%
† replication SNPs (genotyped / genotyped + imputed)

Table S10. Association study results

		Association #1 (all cases)				Association #1 (severe cases)			Replication	
region	*chr*	*SNP*	*position*	p_{DFAM}	p_{TDT}	*SNP*	*position*	p_{DFAM}	*haplotype block*	*p*
1(1)	1	rs2765572	25529815	0.1577	0.3173	rs2765572	25529815	0.2987		
1(2)	1	rs1856746	205210045	0.3228	0.5637	rs11119925	205202614	0.1094	chr1:205199163-205199163	0.1115
2(1)	2	rs6432017	9611326	0.1684	0.2393	rs6432017	9611326	0.1274	chr2:9547467-9614823	0.0806
2(2)	2	rs9309331	60995119	0.1263	0.2482	rs842620	60980129	0.1127	chr2:60966056-60987297	0.1251
2(3)	2	rs4852873	72606447	0.7166	0.1573	rs4852873	72606447	0.7175		
2(4)	2	rs3806607	74589122	0.1097	0.4795	rs13416172	74635379	0.3061		
2(5)	2	rs62153901	96347686	**0.0015**	**0.0278**	rs62153901	96347686	**0.0026**	chr2:96254472-96397667	**0.0064**
2(6)	2	rs10189727	96750553	**0.0299**	0.0881	rs10189727	96750553	0.1775		
2(7)	2	rs10199249	163279790	0.0822	**0.0495**	rs1273201	163344431	**0.0194**	chr2:163273563-163304213	**0.0409**
4(1)	4	rs2069778	123595585	0.1233	0.6547	rs2069776	123591426	0.0761		
6(1)	6	rs9469042	31790997	**0.0187**	**0.0254**	rs9267532	31747958	0.1282		
7(1)	7	rs1267818	65744459	0.1573	0.2008	rs1638735	65733173	0.0927		
7(2)	7	rs4730954	119704394	0.7371	0.7815	rs4730954	119704394	0.4511		
7(3)	7	rs3757853	142360597	0.4405	0.5127	rs3757853	142360597	0.4740		
7(4)	7	rs56932055	148347253	0.3687	0.5637	rs6464933	148354807	0.3940		
8(1)	8	rs951702	8161017	0.9410	1.0000	rs951702	8161017	0.9309		
8(2)	8	rs6472538	71749525	0.0740	0.1088	rs6472538	71749525	0.4075	chr8:71749525-71764563	0.2230
9(1)	9	rs2300934	122848784	**0.0273**	0.2623	rs2300934	122848784	**0.0240**	chr9:122797008-122848784	**0.0027**
10(1)	10	rs1906284	45297609	0.0585	0.2393	rs1906284	45297609	0.0574	chr10:45297609-45441052	0.0889
11(1)	11	rs3019561	60500429	0.0820	0.1655	rs2074229	60539684	0.3248		
11(2)	11	rs2242663	66091884	0.6450	0.8084	rs13897	66087745	0.5275		
11(3)	11	rs10274	66959402	0.2120	**0.0412**	rs10274	66959402	**0.0486**	chr11:66956282-66991927	**0.0061**
14(1)	14	rs6573480	62613774	**0.0452**	0.0736	rs1951805	62637907	**0.0356**	chr14:62449574-62476691	0.0618
16(1)	16	rs13337037	31386212	0.0690	0.0881	rs10871453	30444419	0.0852	chr16:30918684-31055049	**0.0005**
16(4)	16	rs12443685	46783980	0.6170	0.6374	rs12443685	46783980	0.2622		
19(1)	19	rs313624	11413910	0.0693	0.0833	rs313624	11413910	0.0959		
20(1)	20	rs754568	30033526	0.3282	0.7963	rs6121302	30031338	0.2779		
20(2)	20	rs73271169	47940104	0.1239	**0.0184**	rs6020178	48037347	0.1811	chr20:47921264-47921264	0.5238

Table S11. Overlap between IBD GWAS loci and BEB selected regions

A. Summary of overlap between BEB CMS regions and IBD GWAS regions (*60*)

Set	# GWAS regions	# GWAS genes	overlapping selected regions	p_{set}	p_{exp}
Crohn's Disease	64	502	8	0.42	0.60
Crohn's Disease (p<1e-10)	25	248	6	0.12	0.25
Ulcerative colitis	38	301	8	0.046	0.084
Ulcerative colitis (p<1e-10)	17	227	7	0.017	**0.049**

B. Regions of overlap between BEB CMS regions and IBD GWAS regions

CMS region	Max CMS score	**Ulcerative colitis**		**Crohn's disease**		Genes
		region	*p*	*region*	*p*	
chr2:60957829-61241691	6.75	chr2:60760000-61867442 (rs7608910)	1.7E-14	chr2:60770000-61740000 (rs10181042)	6.6E-09	REL, PEX13, PUS10, KIAA1841
chr3:48649603-49077789	6.46	chr3:48140000-51770000 (rs9822268)	1.6E-17	chr3:48140000-51770000 (rs9822268)	1.5E-16	CELSR3, NCKIPSD, PRKAR2A, SLC25A20, ARIH2, DALRD3, WDR6, IMPDH2, QRICH1
chr3:50628862-50946281	4.90					MAPKAPK3, DOCK3
chr3:49121136-49361236	3.87					USP19, LAMB2, CCDC71, KLHDC8B, LOC646498, CCDC36, C3orf62, USP4
chr3:50159130-50275841	3.78					SEMA3F, GNAT1, SLC38A3, GNAI2
chr3:49984343-50140249	3.52					RBM6, RBM5
chr6:31703901-31815903	4.92	chr6:31490696-33008579 (rs9268853)	1.4E-55	chr6:31490000-32980000 (rs1799964)	4.0E-11	BAT2, BAT3, APOM, C6orf47, BAT4, CSNK2B, LY6G5B, LY6G5C, BAT5, LY6G6F, LY6G6D, LY6G6C, C6orf25, DDAH2, CLIC1, MSH5
chr4:123581338-123726077	4.02	chr4:123201764-123784752 (rs17388568)	9.5E-07	*ns*		IL2
chr1:153198332-153306921	3.94	*ns*		chr1:153240000-154390000 (rs3180018)	1.3E-09	ZBTB7B, DCST2, DCST1, ADAM15, EFNA4