SUPPLEMENTARY MATERIALS

Table 1: Recombination rate estimates and summaries of polymorphism data for the 84 genic regions surveyed by SeattleSNPs.

Genic regions	bps total	bps seq.	S AfAm	S CEPH	ρ_{LD} AfAm	ρ_{LD} CEPH	c_{map} (cM/Mb)
BDKRB2	15152	14050	61	49	0.000575	0.000206	2.29
BF	9956	9956	31	25	0.000448	0.000179	0.23
CCR2	10073	10073	41	22	0.001493	0.000137	0.33
CD36	30180	29600	167	80	0.001259	0.000314	1.18
CRF	10451	9560	38	24	0.003243	0.000151	0.57
CRP	6836	6715	29	12	0.000406	0	1.53
CSF2	5992	5992	30	17	0.002389	0.000757	0.63
CSF3	5527	5527	28	17	0.000453	0.000125	1.11
CSF3R	18843	18443	99	52	0.000824	0.000597	1.57
CYP4F2	23567	18701	144	83	0.000688	0.000102	1.27
DCN	35378	34947	139	40	0.000074	0.000043	0.79
F10	29488	25406	98	63	0.002588	0.001609	2.83
F12	10616	10616	44	22	0.003060	0.001858	2.92
F13A1	178952	29536	195	130	0.002676	0.001161	2.37
F2	22128	20407	90	57	0.000135	0.000002	1.24
F2R	24758	24216	94	52	0.000629	0.000298	1.45
F2RL1	18351	18351	98	43	0.000400	0.000069	1.45
F2RL2	9273	9273	61	37	0.005576	0.001451	1.45
F2RL3	11828	10214	42	25	0.001506	0.000270	1.63
F3	17222	16114	61	26	0.001114	0.000350	0.48
F7	18381	11924	44	30	0.002939	0.000610	2.83
FGA	9946	9946	23	17	0.000540	0.000381	0.92
FGB	11956	11604	45	36	0.000053	0.000003	0.92
FGG	10170	10170	21	12	0.000159	0.000210	0.92

FGL2	6382	6382	19	6	0.002905	-	1.13
FSBP	9846	9846	31	26	0.000299	0.000144	0.98
GP1BA	6241	6241	27	21	0.000984	0.000327	2.61
IFNG	7665	7665	28	13	0.000220	0.000418	1.56
IGF2	9013	5251	22	10	0.006010	-	0.56
IL10	7879	7879	26	24	0.000211	0.000291	1.4
IL12A	11792	11330	52	29	0.000803	0.000110	0.47
IL12B	15648	14902	52	33	0.000136	0.000119	0.86
IL13	6919	6919	27	16	0.010907	0.001769	0.23
IL17B	9077	9077	32	22	0.000207	0.000145	1.79
IL19	10998	10998	43	24	0.000470	0.000079	1.4
IL1A	18446	17849	78	50	0.000134	0	0.99
IL1B	17447	17447	51	35	0.001451	0.000658	0.99
IL1R1	27864	27864	134	85	0.001596	0.000398	0.32
IL1R2	23160	23160	188	101	0.000403	0.000043	0.79
IL2	6752	6752	20	10	0.002830	-	0.64
IL20	6634	6634	25	17	0.009734	0.001993	1.4
IL21R	26391	25844	116	76	0.001980	0.001472	1.31
IL22	8393	8393	46	28	0.001229	0.000342	1.15
IL24	11114	10628	41	24	0.000734	0.000077	1.4
IL2RB	26469	26029	147	100	0.003114	0.001053	2.36
IL3	6387	6387	27	9	0.001609	-	0.63
IL4	22845	22845	105	56	0.001045	0.000470	0.23
IL4R	26550	25917	179	118	0.001562	0.000508	1.31
IL5	5186	5186	16	3	0.000800	-	0.23
IL6	8019	7526	41	26	0.001837	0.000435	1.1
IL8	7156	7035	35	9	0.000691	-	0.86
IL9	6676	6676	28	14	0.000704	0.000104	0.78
ITGA2	105904	71521	348	283	0.000167	0.000151	0.64
JAK3	21323	19067	113	56	0.005576	0.001451	2.07

KLK1	10094	9922	63	46	0.001583	0.000348	3.32
LTA	5033	5033	20	19	0.002322	0.000984	0.46
LTB	4412	4412	16	7	0.001174	-	0.46
MC1R	7046	6545	36	22	0.001156	0.000470	1.29
MMP3	13003	11903	50	35	0.000230	0.000023	0.36
MMP9	11224	11049	48	26	0.000367	0.000104	1.57
NOS3	25106	23307	102	54	0.002444	0.000613	1.8
PLAU	10075	9274	30	23	0.000189	0.000254	0.41
PON1	29370	29052	174	120	0.001140	0.001161	0.63
PPARG	83894	40608	134	166	0.000124	0.000074	0.75
PROC	13870	12877	52	39	0.000920	0.000334	1.19
PROCR	7199	6968	14	13	0.000043	0	0.55
PROZ	16418	14366	86	46	0.001714	0.000052	2.83
SCYA2	9174	9070	38	25	0.001224	0.000518	2.43
SELE	14019	13892	86	70	0.000292	0.000186	1.41
SELP	44381	43454	254	271	0.001259	0.000600	1.41
SERPINA5	8145	7806	61	41	0.001562	0.001156	2.02
SERPINC1	16330	15208	43	27	0.000527	0.000105	0.62
SERPINE1	14544	13208	84	48	0.000562	0.000241	0.7
SFTPB	11807	11094	49	18	0.001011	0.001520	1.05
SMP1	25269	23549	89	39	0.000196	0.000151	0.72
STAT4	52796	28360	94	52	0.000864	0.000540	1.18
STAT6	18913	18791	53	22	0.001340	0.000215	0.64
THBD	8532	7254	24	14	0.001829	0.001816	0.96
TNF	4830	4830	21	12	0.006224	0	0.46
TNFAIP1	15259	14331	53	19	0.000205	0.000050	0.61
TNFRSF1A	17350	16207	26	63	0.001612	0.000312	2.54
VCAM1	22868	22868	102	39	0.004475	0.000912	0.73
VEGF	15442	15442	62	48	0.003748	0.001391	1.61
VTN	5559	5559	28	15	0.000525	0	0.61

Following the name of each gene are: "bps total", the total length of the region surveyed; "bps seq.", the number of base pairs sequenced; S, the number of segregating sites; ρ_{LD}, HUDSON's (2001) estimate of the population crossing-over rate per base pair (for $f=0$); and c_{map}, an estimate of the crossing-over rate per base pair per generation from KONG et al. (2002). ρ_{LD} was only estimated from data sets with more than ten segregating sites.

Table 2: The performance of estimators of f and ρ, based on the African-American data.

PROFILE	$\hat{f}=0$	$\hat{f}=1$	$\hat{f}=2$	$\hat{f}=3$	$\hat{f}=4$	$\hat{f}=5$	$\hat{f}=6$	$\hat{f}=7$	$\hat{f}=8$	Med. $\hat{\rho}$
$f=0$	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.00068
$f=1$	0.090	0.900	0.005	0.000	0.000	0.005	0.000	0.000	0.000	0.00068
$f=2$	0.000	0.295	0.670	0.025	0.005	0.000	0.000	0.000	0.005	0.00070
$f=3$	0.000	0.000	0.475	0.430	0.085	0.010	0.000	0.000	0.000	0.00071
$f=4$	0.000	0.000	0.030	0.465	0.400	0.090	0.015	0.000	0.000	0.00070
JOINT	$\hat{f}=0$	$\hat{f}=1$	$\hat{f}=2$	$\hat{f}=3$	$\hat{f}=4$	$\hat{f}=5$	$\hat{f}=6$	$\hat{f}=7$	$\hat{f}=8$	Med. $\hat{\rho}$
$f=0$	0.730	0.265	0.005	0.000	0.000	0.000	0.000	0.000	0.000	0.00060
$f=1$	0.000	0.510	0.440	0.045	0.000	0.000	0.000	0.005	0.000	0.00050
$f=2$	0.000	0.015	0.490	0.360	0.110	0.020	0.000	0.000	0.005	0.00050
$f=3$	0.000	0.000	0.060	0.400	0.355	0.120	0.050	0.010	0.005	0.00050
$f=4$	0.000	0.000	0.000	0.075	0.300	0.335	0.160	0.100	0.030	0.00050

200 simulated sets of the 84 loci were generated for a given f value, conditional on $\rho = \rho_{map}$; the median ρ_{map} is 0.00065. For each locus, other parameters were chosen to mimic what is observed in the African-American sample. Shown here is the distribution of estimates of f, $\hat{f}$, and the median of our estimates of ρ, $\hat{\rho}$, obtained using either the profile (where $\hat{\rho}$ = the median of $\hat{\rho}_{f*}$) or the joint estimation method (where $\hat{\rho} = \hat{\rho}^+$; see Methods for details).

Table 3: The performance of estimators of f and ρ for a set of 50 loci collected according to the study design of FRISSE et $al.$ (2001).

PROFILE	$\hat{f}=0$	$\hat{f}=1$	$\hat{f}=2$	$\hat{f}=3$	$\hat{f}=4$	$\hat{f}=5$	$\hat{f}=6$	$\hat{f}=7$	$\hat{f}=8$	Med. $\hat{\rho}$
$f=0$	0.94	0.05	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00080
$f=1$	0.56	0.37	0.04	0.02	0.01	0.00	0.00	0.00	0.00	0.00095
$f=2$	0.28	0.49	0.12	0.05	0.01	0.01	0.04	0.00	0.00	0.00090
$f=3$	0.12	0.36	0.22	0.21	0.03	0.03	0.02	0.01	0.00	0.00090
$f=4$	0.06	0.22	0.20	0.19	0.08	0.15	0.07	0.03	0.00	0.00090
JOINT	$\hat{f}=0$	$\hat{f}=1$	$\hat{f}=2$	$\hat{f}=3$	$\hat{f}=4$	$\hat{f}=5$	$\hat{f}=6$	$\hat{f}=7$	$\hat{f}=8$	Med. $\hat{\rho}$
$f=0$	0.74	0.22	0.03	0.01	0.00	0.00	0.00	0.00	0.00	0.00060
$f=1$	0.11	0.53	0.20	0.09	0.06	0.00	0.01	0.00	0.00	0.00060
$f=2$	0.01	0.23	0.35	0.16	0.14	0.04	0.03	0.02	0.02	0.00060
$f=3$	0.00	0.06	0.24	0.29	0.22	0.03	0.07	0.04	0.05	0.00060
$f=4$	0.00	0.04	0.21	0.19	0.13	0.11	0.10	0.09	0.13	0.00060

100 simulated sets of 50 loci were generated for a given f value, with ρ drawn from a Gamma distribution with mean 0.00070 (see Methods). For each locus, other parameters were chosen to mimic data collected by FRISSE et $al.$ (2001): specifically, there were 15 pairs of chromosomes, each locus consisted in two segments of 1000 bps separated by 8000 bps and $\theta = 0.0012$ per bp.

Shown here is the distribution of estimates of f, $\hat{f}$, and the median of our estimates of ρ, $\hat{\rho}$, obtained using either the profile (where $\hat{\rho} = $ the median of $\hat{\rho}_{f*}$) or the joint estimation method (where $\hat{\rho} = \hat{\rho}^{+}$; see Methods for details).

Table 4: Estimates of f and ρ from the SeattleSNPs data, given a mean conversion tract length, L.

L	60	250	500	1000	60	250	500	1000
Estimate of f		PROFILE				JOINT		
African American	5.00	1.50	1.00	0.50	10.00	10.00	5.75	3.25
CEPH	2.00	0.50	0.25	0.00	10.00	4.00	2.25	1.25
Estimate of ρ ($\times 10^{-4}$)								
African American	8.5	8.0	8.0	8.0	5.0	3.0	3.0	3.0
CEPH	3.0	3.0	3.0	3.0	2.0	2.0	2.0	2.0

Shown is the estimate of f, $\hat{f}$, and the estimate of ρ, $\hat{\rho}$, using either the profile or joint estimation method (see Methods for details). We estimated f on a grid from 0 to 10 with increments of 0.25) and ρ on a grid from 0 to 0.011 (with increments of 0.0001).