Participating CARe studies

Atherosclerosis Risk in Communities Study (ARIC)
ARIC is a longitudinal cohort study of atherosclerosis and its clinical sequelae. From 1987 to 1989, a population-based sample of 15,792 men and women aged 45 to 64 years were recruited from 4 US communities (Forsyth County NC, Jackson MS, suburban Minneapolis MN, and Washington County MD). Fibrinogen was measured at the first exam, and the sample for fibrinogen analysis comprised 9406 EA and 2853 AA individuals.

Coronary Artery Risk Development in Young Adults (CARDIA)
CARDIA is a longitudinal study of the evolution of coronary heart disease risk, started in 1985–86 in 5,115 AA and EA men and women, then aged 18–30 years. The CARDIA sample was recruited was recruited at random during 1985–86 primarily from geographically based populations in Birmingham AL, Chicago IL, and Minneapolis MN and, in Oakland, CA, from the membership of the Kaiser-Permanente Health Plan. Examinations after baseline were year 2 (1987–88, n=4624, 90% retention, year 5 (1990–91, n=4352, 85% retention), year 7 (1992–93, n=4086, 80% retention), year 10 (1995–96, n=3950, 79% retention), year 15 (2000–2001, n=3672, 74% retention) and year 20 (2005–06, n=3549, 72% retention). For the present analysis, fibrinogen measured on 1370 EA and 1178 AA participants from the Coronary Artery Risk in Young Adults (CARDIA) Year 7 examination was used.

Cleveland Family Study (CFS)
The Cleveland Family Study (CFS) comprises 2,534 individuals (46% AA) from 352 families examined every 4 years over a 16 year period (1990–2006). The study was begun with the initial aim to quantify the familial aggregation of sleep apnea. Index probands (n=275) were recruited from 3 area sleep centers if they had a confirmed diagnosis of sleep apnea and at least 2 first-degree relatives available to be studied. In the first 5 study years, neighborhood control probands (n=87) with ≥ 2 living relatives available were also recruited. All available first-degree relatives and spouses of the case and control probands were recruited. Blood was sampled and DNA isolated for subjects seen in the last 2 exam cycles (n=1447). The 4th exam included 736 subjects (60% African American), with oversampling subjects who had had a microsatellite genome scan. In the Cleveland Family Study (CFS), fibrinogen was measured on 244 EA and 341 AA at Visit 5.

Cardiovascular Health Study (CHS)
The CHS is a population-based, observational study of risk factors for clinical and subclinical cardiovascular diseases. The study recruited participants 65 years and older from 4 US communities (Forsyth County, North Carolina; Sacramento County, California; Washington County, Maryland; and Pittsburgh, Pennsylvania) in 2 phases: 5201 participants from 1989 to 1990, and 687 (primarily AA participants) from 1992 to 1993. Fibrinogen measured at baseline in the Cardiovascular Health Study (CHS) on 719 AA and 3859 EA was used in this analysis.

Framingham Heart Study (FHS)
The FHS started in 1948 with 5209 randomly ascertained participants from Framingham, MA, who had undergone biannual examinations to investigate cardiovascular disease and its risk factors. In 1971, the offspring cohort (comprised of 5124 children of the original cohort and the

children's spouses) and in 2002, the third generation (consisting of 4095 children of the offspring cohort), were recruited. FHS participants in this study are of European ancestry. In the Framingham Heart Study, fibrinogen measures from 6502 individuals from the parent (10^{th} examination), offspring (5^{th} examination) and generation 3 cohorts were used.

Multi-Ethnic Study of Atherosclerosis (MESA)
MESA is a cohort study of the characteristics of subclinical cardiovascular disease and the risk factors that predict progression to clinically overt cardiovascular disease or progression of subclinical disease. MESA researchers study a diverse, population-based sample of 6,814 asymptomatic men and women aged 45–84: 2,622 white (38%), 1,893 African-American (28%), 1,496 Hispanic (22%) and 803 (12%) of Chinese descent. Participants were recruited from six Field Centers across the United States (Winston-Salem, NC; St. Paul, MN; Chicago, IL; Los Angeles, CA; New York, NY; Baltimore, MD). Fibrinogen was assayed on 2253 EA and 1566 AA at the baseline visit in the Multi-ethnic Study of Atherosclerosis (MESA).

Loci and SNP selection for the IBC array
The IBC array was developed using SNP and linkage disequilibrium information from the HapMap as well as resequencing data from SeattleSNPs and National Institute of Environmental Health Sciences (NIEHS) SNPs [1]. A cosmopolitan tagging approach was employed to produce the optimal union of SNPs to tag the loci of interest for these reference populations at predefined density criteria: down to minor allele frequency (MAF) of 2% and r2 of 0.8 for priority one loci and down to MAF of 5% and r2 of 0.5 for priority 2 loci, and all non-synonymous SNPs with MAF > 1% across the reference populations17. This approach is particularly informative for African Americans as (a) integration of information from SeattleSNPs and NIEHS databases affords greater representation of African and European chromosomes than the CEPH and Yoruba HapMap chromosomes and is thus likely to produce more accurate representation of more frequent African SNPs; (b) the cosmopolitan approach in the priority 1 and 2 loci affords detection of SNPs that may be present within these bounds for the other broader populations but may also detect representation of these same SNPs at lower frequencies in these populations.

For high priority loci, SNPs of specific interest were selected, such as those derived from previously published studies and non-synonymous SNPs with minor allele frequency >1%. Of the fibrinogen-associated genes analyzed in detail, *FGA*, *FGB*, *FGG*, *IL1RN*, *IL6R*, and *IL5* were priority level 1 genes, while *CPS1*, *PCCB*, *NLRP3*, *SLC22A5*, and *IL3* were priority level 2 genes. Additional SNP genotypes in the region of chromosome 5q31 containing the *IRF1* gene were imputed using the HapMap. *CD300LF* is not represented on the IBC array and SNPs in this region were not imputed. *CPS1* was chosen as a low priority hypertension candidate gene. *SLC22A5*, a key carnitine transport gene, was chosen as an essential hypertension candidate gene derived from the literature. *PCCB* was chosen as a component in a lipoprotein metabolism pathway. *NLRP3* was requested by a number of investigators as a candidate gene of interest. *IL13* was chosen as a component of 3 CVD related pathways.

Genotyping quality control
Several quality control (QC) procedures were performed on the genotype data, separately for each cohort. Samples were excluded for sex mismatch, duplicate discordance, or call rate <95%.

For each set of duplicates or monozygotic twins, data from the sample with the highest genotyping call rate was retained. SNPs were removed for call rate <95% or HWE $p<10^{-5}$ in EA (there was no HWE filter for AA), or when monomorphic. This resulted in available data on 47,539 working SNPs.

Principal components estimation

Principal components were calculated using EIGENSTRAT [2] on the cleaned CARe IBC genotype data. HapMap populations (CEU, YRI, CHB+JPT) were used as reference (seed) populations. The main 10 principal components were used as covariates in the regression analyses.

Imputation and association analysis of untyped SNPs

Imputation of untyped and missing SNP genotypes was performed within chromosomal regions containing one or more typed SNPs significantly associated with fibrinogen using MACH 1.0.16 (http://www.sph.umich.edu/csg/abecasis/MACH/). For the European samples, phased haplotypes from the CEU founders of HapMap 2 were used as reference. For the African Americans, a combined CEU+YRI reference panel was created that includes SNPs segregating in both CEU and YRI, as well as SNPs segregating in one panel and monomorphic and non-missing in the other. Imputation for the IBC array was performed in two steps. First, individuals with pedigree relatedness or cryptic relatedness ($\hat{\pi} > 0.05$) were filtered. A subset of individuals was randomly extracted from each panel and used to generate recombination and error rate estimates for the corresponding sample. Second, these rates were used to impute all sample individuals across the entire reference panel. Imputation results were filtered at an $\hat{r}^2$ threshold of 0.5 and a minor allele frequency threshold of 0.01. For imputed genotypes, we used dosage information (*i.e.* a value between 0.0 – 2.0 calculated using the probability of each of the three possible genotypes) in the regression model implemented in PLINK (for cohorts with unrelated individuals) or the Maximum Likelihood Estimation (MLE) routines implemented in R V.2.9.2 (for cohorts with related individuals).

Haplotype estimation

Haplotype estimation was performed in R V2.9.2 using the haplo.stats package [3]. Haplotypes were estimated using an expectation maximization (EM) algorithm and a "progressive insertion" method to estimate haplotypes, where groups of loci are inserted into haplotypes of increasing lengths, and pairs of haplotypes are trimmed off when their posterior probabilities fall below a certain threshold. The group size for SNP insertion was set to 4 and pairs of haplotypes per participant with posterior probabilities <0.0001 were trimmed. Haplotype association analysis was performed in SAS V 9.1.3 (SAS Institute, Cary, NC), and haplotypes were modeled as both additive and dominant effects.

Comparison of fibrinogen-associated SNP linkage disequilibirum patterns between EA and AA

By comparing the SNP phenotype association results between EA and AA, in some instances the regional LD patterns suggested further localization of the putative causal variant(s). In the *IL6R* gene, the cluster of 5 SNPs associated with lower fibrinogen in EA (rs7529229, rs4537545, rs7518199, rs4129267, and rs8192284) could be differentiated into 2 distinct clusters among AA.

Only one of these SNP clusters, which include the non-synonymous variant rs8192284 (in addition to rs4129267 and rs7518199) was associated with lower fibrinogen in AA. In the *FGB* gene, the lower extent of LD between rs1800787 and other SNPs in this region among AAs (Figure S2) compared to EAs (Figure S3) may suggest rs1800787 as the causal variant. Based on LD and association patterns in the *FGG* region in African Americans, the other major LD cluster associated with lower fibrinogen levels derived primarily from the *FGG* region could be narrowed to rs2066861, rs2066854, rs2066864, and rs12644950.

REFERENCES

1. Keating BJ, Tischfield S, Murray SS, Bhangale T, Price TS, Glessner JT, Galver L, Barrett JC, Grant SF, Farlow DN, Chandrupatla HR, Hansen M, Ajmal S, Papanicolaou GJ, Guo Y, Li M, Derohannessian S, de Bakker PI, Bailey SD, Montpetit A, Edmondson AC, Taylor K, Gai X, Wang SS, Fornage M, Shaikh T, Groop L, Boehnke M, Hall AS, Hattersley AT, Frackelton E, Patterson N, Chiang CW, Kim CE, Fabsitz RR, Ouwehand W, Price AL, Munroe P, Caulfield M, Drake T, Boerwinkle E, Reich D, Whitehead AS, Cappola TP, Samani NJ, Lusis AJ, Schadt E, Wilson JG, Koenig W, McCarthy MI, Kathiresan S, Gabriel SB, Hakonarson H, Anand SS, Reilly M, Engert JC, Nickerson DA, Rader DJ, Hirschhorn JN, Fitzgerald GA. Concept, design and implementation of a cardiovascular gene-centric 50 k SNP array for large-scale genomic association studies. *PLoS One.* 2008; 3: e3583.

2. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 2006;38:904–9.

3. Sinnwell JP, Schaid DJ, Yu Z. *haplo.stats: Statistical Analysis of Haplotypes with Traits and Covariates when Linkage Phase is Ambiguous*. 2008; R package version 1.3.8:

Table S1. Meta-analysis results of significant SNPs in European Americans

Chr	Coordinate	SNP	Gene	Major Allele	Minor Allele	Beta Fixed	P-value Fixed	Beta random	P-value Random	I^2 for heterogeneity
4	155703465	rs1800787	FGB	C	T	0.151	1.43E-39	0.151	1.43E-39	0
4	155700739	rs4508864	FGB	C	T	0.148	1.65E-38	0.148	1.65E-38	0
4	155707434	rs2227407	FGB	C	T	0.148	1.72E-38	0.148	1.72E-38	0
4	155712869	rs2059503	FGB	A	T	0.146	6.93E-35	0.147	2.34E-29	11.11
4	155708271	rs6056	FGB	C	T	0.145	9.25E-35	0.147	3.97E-30	9.18
4	155711209	rs4220	FGB	G	A	0.146	1.04E-34	0.147	2.26E-29	10.71
4	155729764	rs2070016	FGA	A	G	0.138	2.69E-27	0.139	9.48E-23	11.63
4	155709058	rs6054	FGB	C	T	−0.580	1.20E-17	−0.580	1.20E-17	0
4	155703158	rs1800790	FGB	G	A	0.152	6.25E-11	0.152	6.25E-11	0
5	131895734	rs17690122	IL5	A	G	−0.077	1.79E-10	−0.075	3.22E-06	31.34
4	155746886	rs2066861	FGG	C	T	−0.068	4.18E-10	−0.047	0.23	90.00
4	155756771	rs12644950	FGG	G	A	−0.068	4.68E-10	−0.047	0.23	89.95
4	155754631	rs2066854	FGG	A	T	−0.068	4.94E-10	−0.047	0.23	89.93
4	155757920	rs13130318	FGG	T	G	−0.067	7.64E-10	−0.046	0.24	89.96
2	211248752	rs7422339	CPS1	C	A	−0.060	1.95E-09	−0.060	1.95E-09	0
5	131900282	rs743562	IL5	C	T	−0.054	8.05E-09	−0.054	8.05E-09	0
5	132017035	rs3091307	IL13	A	G	−0.063	1.32E-08	−0.062	7.93E-06	23.27
4	155699884	rs7673587	FGB	C	T	−0.056	1.41E-08	−0.056	1.41E-08	0
5	131913139	rs12652920	IL5	G	C	−0.063	2.61E-08	−0.062	2.40E-06	16.20
4	155711674	rs2227421	FGB	A	C	−0.054	3.04E-08	−0.054	3.04E-08	0
1	245670086	rs4925659	NLRP3	G	A	0.052	5.26E-08	0.053	0.0006	51.13
1	152687402	rs7529229	IL6R	T	C	−0.050	8.28E-08	−0.051	0.003	61.51
5	132011957	rs2158177	IL13	A	G	−0.062	9.00E-08	−0.062	5.01E-07	5.69
1	152685503	rs4537545	IL6R	C	T	−0.049	1.98E-07	−0.050	0.005	62.83
1	245668218	rs12239046	NLRP3	C	T	−0.049	3.26E-07	−0.050	0.0003	38.53
1	152674043	rs7518199	IL6R	A	C	−0.048	3.89E-07	−0.047	0.0003	35.51
1	245666924	rs1539019	NLRP3	C	A	−0.047	7.74E-07	−0.045	0.004	52.40
1	152692888	rs4129267	IL6R	C	T	−0.046	8.41E-07	−0.046	0.004	54.00

Chr	Coordinate	SNP	Gene	Major Allele	Minor Allele	Beta Fixed	P-value Fixed	Beta random	P-value Random	I^2 for heterogeneity
5	131751187	rs2073643	SLC22A5	C	T	−0.046	9.31E-07	−0.046	9.31E-07	0
1	152693594	rs8192284	IL6R	A	C	−0.046	9.34E-07	−0.046	0.004	52.99
3	137485499	rs3821445	PCCB	A	G	0.056	1.19E-06	0.056	1.19E-06	0
2	241877453	rs6752050	HDLBP	T	C	−0.063	1.22E-06	−0.063	2.01E-06	2.24
5	131895601	rs2706399	IL5	G	A	0.044	2.18E-06	0.044	7.01E-06	4.79
2	113588522	rs315921	IL1RN	G	A	−0.057	2.65E-06	−0.057	2.65E-06	0

Table S2. Meta-analysis results of significant SNPs in African Americans

Chrom	Coordinate	SNP	Gene	Major Allele	Minor Allele	Beta Fixed	P-value Fixed	Beta random	P-value Random	I^2 for hetero.
4	155749031	rs2066874	FGG	T	C	−0.333	2.86E-11	−0.314	6.32E-05	49.41
4	155729726	rs2070017	FGG	C	T	−0.163	4.82E-09	−0.163	4.82E-09	0
4	155734845	rs10050257	FGG	T	G	−0.213	7.66E-09	−0.213	7.66E-09	0
4	155746595	rs2066877	FGG	T	C	−0.204	2.15E-08	−0.204	2.15E-08	0
4	155703465	rs1800787	FGB	C	T	0.146	4.66E-07	0.128	0.001659	40.86
4	155709794	rs6058	FGB	G	T	−0.170	9.52E-07	−0.163	0.00561	55.39

Figure S1. The quantile-quantile plots for European Americans and African Americans using the fixed effects meta-analysis results
The Y axis represents the −log(p-value) for the observed p-values, and the X axis represents the −log(p-value) for the expected p-values.

Figure S2. A pairwise linkage disequilibrium (LD) plot for single nucleotide polymorphisms (SNPs) in the *FGA*, *FGB*, and *FGG* loci in European Americans
White squares indicate no LD, gray squares indicate low to moderate LD, and black squares indicate high LD between pairs of SNPs. Numbers within each square are r^2 values of LD. Haplotype blocks are designated by blacks lines.

Figure S3. A pairwise linkage disequilibrium (LD) plot for single nucleotide polymorphisms (SNPs) in the *FGA*, *FGB*, and *FGG* loci in African Americans
White squares indicate no LD, gray squares indicate low to moderate LD, and black squares indicate high LD between pairs of SNPs. Numbers within each square are r^2 values of LD. Haplotype blocks are designated by black lines.

Figure S1. QQ plot of meta-analysis P-value distributions in European Americans and African Americans

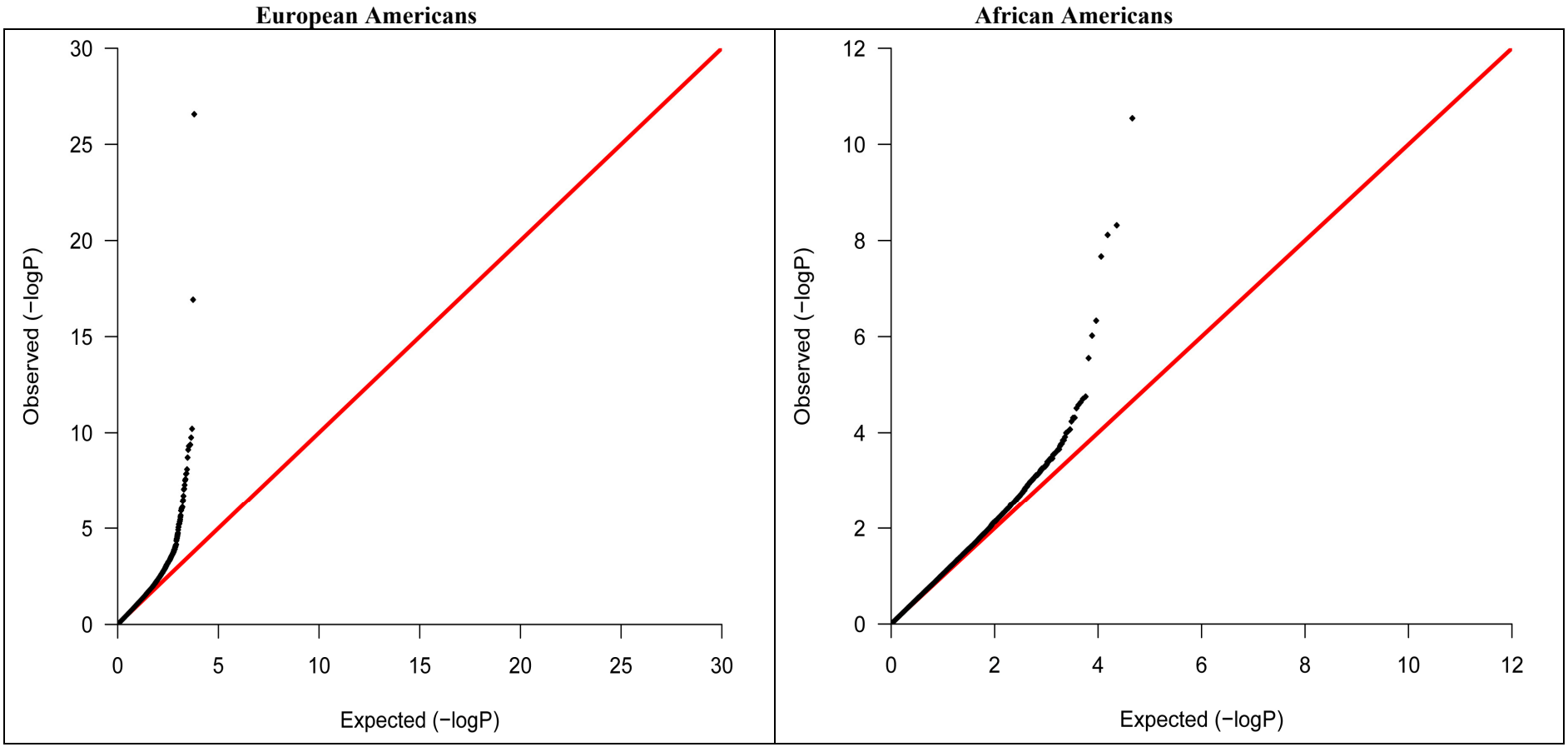

Figure S2. Linkage disequilibrium plot (r^2) for *FGA*, *FGB*, and *FGG* SNPs in European Americans

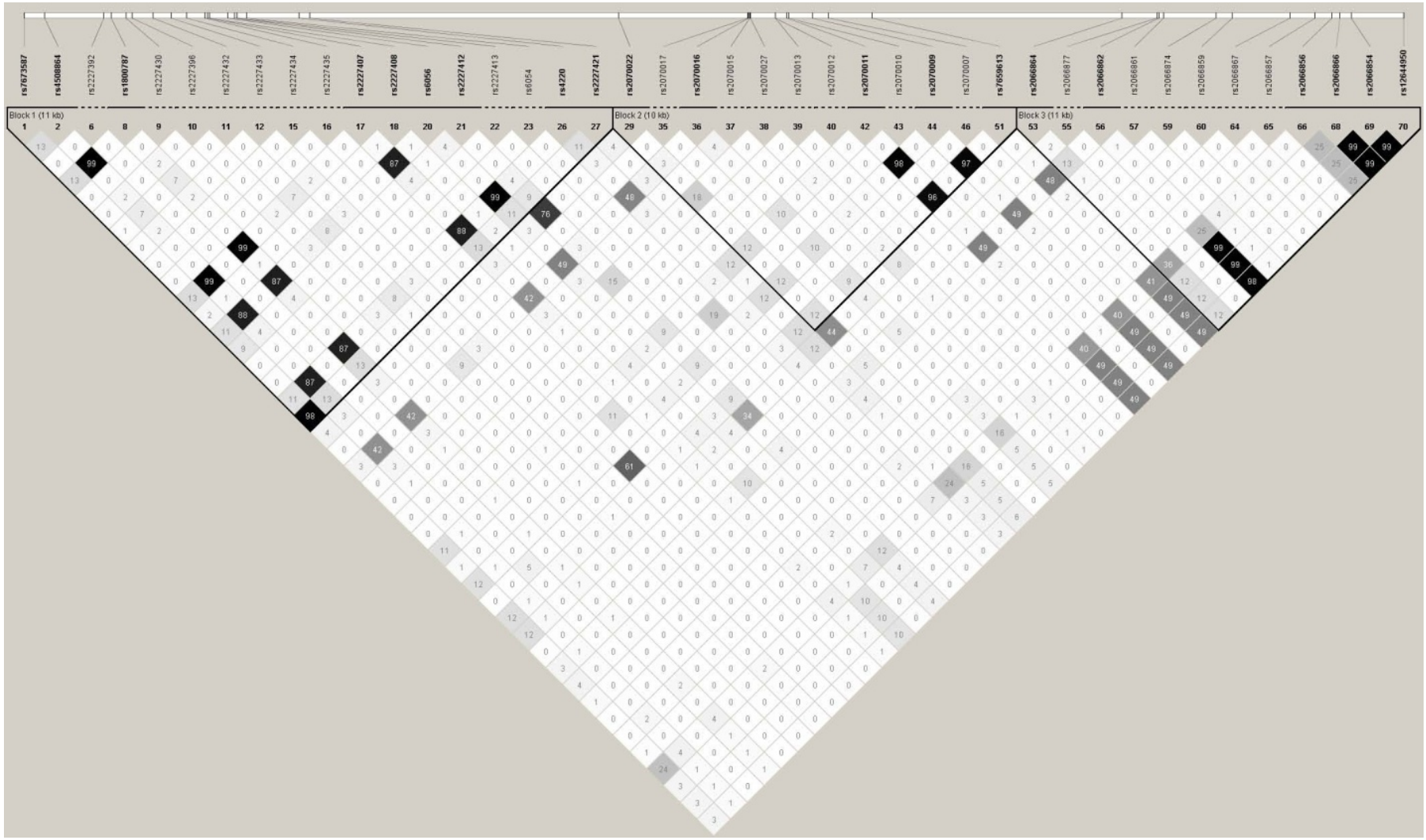

Figure S3. Linkage disequilibrium plot (r^2) for *FGA*, *FGB*, and *FGG* SNPs in African Americans