Supplementary Note

S1. Encoding of *HP* structural variants

HP structural variation was typed with droplet-digital PCR (ddPCR)[1] (**Supplementary Fig 4., Fig. 2b**) across four populations: CEU, TSI, IBS, YRI and was encoded as a series of bi-allelic variants in each reference panel. The structural boundary breakpoints (A-E) were coded as "1" to indicate the presence of the boundary and "0" to indicate its absence. The locations of boundaries A-E are indicated in **Supplementary Figure 4**. Additionally, each common *HP* subtype was encoded as a bi-allelic variant. The following table shows the encoding used for the four common *HP* structural haplotypes.

Subtype Encoding	HP1S	HP1F	HP2FS	HP2SS
breakpoint_A	0	1	1	0
breakpoint_B	0	1	0	0
breakpoint_C	0	0	0	1
breakpoint_D	0	0	1	1
breakpoint_E (CNV)	0	0	1	1
HP1S	1	0	0	0
HP1F	0	1	0	0
HP2FS	0	0	1	0
HP2SS	0	0	0	1

S2. Evaluating the efficacy of imputation for *HP* structural variants

The above *HP* structural variation was combined with SNP genotype data from the 1000 Genomes Project[2] (Illuina OMNI array) and HapMap3[3] (Illumina 1M and Affymetrix 6.0 arrays). These panels used different subsets of the unrelated individuals typed in our study based on the availability of SNP genotypes in each data set.

Each SNP genotype dataset went through quality control prior to being combined with structural variants. SNPs in each panel with a genotyping rate below 90% or that defied Hardy-Weinberg Equilibrium ($p<0.001$) were excluded. Individuals with genotyping rate <90% were also omitted. SNPs assayed in HapMap3 from the region chr16:69628379-71655164 (hg18) were used, and SNPs assayed in the 1000 Genomes project with the Omni 2.5 SNP array from the region chr16:71088193-73097663 (hg19) were used. We discarded all SNP genotypes within the CNV region chr16:72090310-72093744 (hg19) from all panels and replaced these with the encoded structural alleles.

Reference panels were composed of unphased and unrelated individuals. Separate reference panels were created and tested for European and African populations due to differences in SNP haplotype backgrounds for *HP* subtypes (**Figure 4)**.

To evaluate the utility of these reference panels as an imputation resource, we performed a series of “leave one out” trials. For each trial, we constructed a new panel of unphased markers where we set the genotypes for the nine markers for the structural alleles to “missing” for a single individual. We then used Beagle to phase the resulting panel and impute the missing markers. Beagle was run with default parameters plus two additional settings (nsamples=15, niterations=15). European samples were also run with the maxwindow=2000 setting as the best tag SNPs for the HP2SS haplotype are far away. By forcing Beagle to re-phase the panel in each iteration, we avoided any potential problems with data from the left-out individual impacting the phasing of the reference panel. The results of this analysis are shown in **Table 1** and **Supplementary Tables 2-5**.

S3. Imputation of *HP* structural variation into cholesterol cohorts

A phased reference panel composed of 1000 Genomes and HapMap3 SNP genotypes from the Illumina OMNI, Illumina 1M, and Affymetrix 6.0 arrays, (QC discussed in S1) and *HP* structural alleles (based on ddPCR) was created for imputation. Only the CEU, TSI and IBS populations were used for the reference panel. In the case of disagreements among SNP calls for an individual between the arrays, the Illumina OMNI genotype was used. The imputation into the cholesterol cohorts (below) was run with default Beagle parameters plus two additional settings (nsamples=25, niterations=50). In addition to the encoded *HP* structural variants described above, variants were added to the reference panel to indicate the haplotype background of each allele (A, B, C, D as in **Fig. 1**).

S4. Description of cholesterol cohorts used in the association study

Study Cohorts

The influence of *HP* structural alleles on lipid traits was assessed in six cohorts retrieved from dbGaP (see Acknowledgements): ARIC, CHS, CARDIA, FUSION, MESA and NFBC66. Individuals from each cohort were selected for measurement of lipid traits: TC, LDL, HDL, and TRIGS, and availability of relevant covariates: type 2 diabetes diagnosis, cholesterol lowering medication/statin use, age and gender. Analysis were performed using baseline lipid measurements for cohorts with longitudinal follow-up.

Atherosclerosis Risk in Communities (ARIC)

The participants of ARIC [4] were recruited from 4 communities beginning in 1987, Forsyth County, NC; Jackson, MS; the northwest suburbs of Minneapolis, MN; and Washington County, MD [dbGaP phs000280.v3.p1]. Whole genome genotyping was performed using the Affymetrix 6.0 platform and genotype calling was performed using the Birdseed calling algorithm [5]. 8,592 European-American subjects were included in the analyses.

Cardiovascular Health Study (CHS)

The participants of CHS[6] were recruited from Forsyth County, NC; Sacramento County, CA; Washington County, MD; and Pittsburgh, PA [dbGaP phs000287.v4.p1]. Whole genome genotyping was performed using the Illumina

HumanCNV370v1 platform and genotype calling was performed using Illumina GenomeStudio software. 3,220 European-Americans subjects were used for the analyses.

Coronary Artery Risk Development in Young Adults (CARDIA)

The participants of CARDIA[7]
were recruited from four communities in the United States: Birmingham, AL; Chicago, IL; Minneapolis, MN; and Oakland, CA [dbGaP phs000285.v3.p2]. Whole genome genotyping was performed using the Affymetrix 6.0 platform and genotype calling was performed using Birdseed calling algorithm. 1,168 European-Americans subjects were used for the analyses.

The Finland-United States Investigation of NIDDM Genetics (FUSION)

The participants of the FUSION[8] were recruited from throughout Finland as part of a type 2 diabetes case-control research studies. Whole genome genotyping was performed using the Illumina HumanHap300v1.1 platform. Analyses were performed by type 2 diabetes case-control status of 1,687 Finnish European subjects, consisting of 910 with diabetes and 777 without evidence of diabetes.

Multi-Ethnic Study of Atherosclerosis (MESA)

The participants of the MESA[9] were recruited from six regions in the United States: Baltimore City and Baltimore County, Maryland; Chicago, Illinois; Forsyth County, North Carolina; Los Angeles County, California; New York, New York; and St. Paul, Minnesota.

Whole genome genotyping was performed using the Affymetrix 6.0 platform and 2,358 European-American subjects were used for the analyses.

Northern Finland Birth Cohort 1966 (NFBC1966)

The participants of the NFBC66[10] were recruited among women at maternity health clinics in northern Finland in 1966. Whole genome genotyping was performed using the HumanCNV370-Quadv3 core platform. A total of 5,263 Finnish European ancestry subjects were used for the analyses.

Cohort	Country of Origin	N Samples	Mean Age, years ± SD	Sex, % Female	Total Cholesterol (TC), mg/dL ± SD	LDL cholesterol, mg/dL ± SD	HDL Cholesterol, mg/dL ± SD	Triglycerides (TC), mg/dL ± SD	Statin Use (%)	T2D: N (%)
ARIC	US	8592	54.33 ± 5.68	4344 (51.7%)	214.8 ± 40.1	137.7 ± 37.2	50.2 ± 16.5	137.6 ± 89.8	300 (3.6%)	702 (8.4%)
CARDIA	US	1168	25.6 ± 3.3	611 (51.3%)	175.7 ± 32.0	108.2 ±29.8	51.9 ± 13.0	78.1 ± 56.5	0 (0%)	8 (0.7%)
CHS	US	3220	72.3 ± 5.4	1937 (60.2%)	216.0 ± 39.0	133.6 ± 35.5	55.0 ± 15.9	140.6 ± 74.6	140 (4.4%)	677 (21%)
FUSION T2D Cases	Finland	910	62.8 (7.6)	382 (42.0%)	221.2 ± 48.5	137.9 ± 39.5	45.3 ± 12.9	201.9 ± 159.6	n/a	910 (100%)
FUSION T2D Controls	Finland	777	62.7 (7.4)	369 (47.5.0%)	223.3 ± 37.6	141.9 ± 34.3	58.1 ± 16.3	119.3 ± 60.0	n/a	0 (0%)
MESA	US	2358	62.7 ± 10.2	1320 (52.3%)	196.0 ± 35.3	117.1 ± 30.3	52.4 ± 15.8	133.0 ± 90.1	462 (18.3%)	151 (6.0%)
NFBC66	Finland	5263	31.0 ± 0	2610 (52.7%)	196.7 ± 39.01	116.5 ± 34.5	60.4 ± 14.8	105.2 ± 64.0	n/a	10 (0.2%)

Genotype Quality control and 1000 Genomes Imputation

A stringent quality control protocol was applied to study genome wide genotype data retrieved from dbGaP. Samples that had ambiguous or incorrect gender were filtered out. SNPs that have > 5% missing rate were filtered out. Samples that have > 2% missing SNPs were removed. SNPs that have minor allele frequencies < 1% were dropped. Samples that have extreme heterozygosity (+/- 4 standard deviations) were removed. The SNP annotations for chromosome and base-pair positions were set to the coordinates of hg19 (GRCh37) using the liftOver tool[11]. Pairwise IBD/IBS were calculated and individuals that have excessive matching with other individuals. For all other cohorts, the threshold was set at PI_HAT > 0.05. Principal components analysis was performed using SMARTPCA[12] and each study was projected onto HapMap version 3[13]. Samples that do not cluster with the expected HapMap population were removed. SNPs that have excessive plate-effects ($P<1x10^{-7}$) were dropped. SNPs that have excessive deviation from Hardy-Weinberg equilibrium ($P<1x10^{-7}$) were also dropped. The QC protocol was performed using PLINK[14] (v1.07) and custom R and PERL scripts.

The imputation for all cohorts was performed with a 1000 Genomes version 1 phase 3 reference panel containing 379 Europeans, 246 Africans and African-Americans, 286 Asians and 181 Latin Americans. The imputation panel consists of approximately 22 million variants (SNPs and indels). For the X-chromosome, only the non-pseudo autosomal region was imputed. The phasing and imputation

were done separately for males and females for the X-chromosome. The genotypes were phased using SHAPEIT2[15] (version 2.644) and imputed using IMPUTE2[16] (version 2.3).

Cohort	Ancestry	Array Vender	Array	N Subjects	N SNPs	Genome Build
ARIC	EA	Affymetrix	Affy6.0	8,592	595,125	hg19 1k+
CARDIA	EA	Affymetrix	Affy6.0	1,168	588,503	hg19 1k+
CHS	EA	Illumina	HumanCNV370v1	3,220	310,108	hg19 1k+
FUSION	EA (Finnish)	Illumina	HumanHap300v1	1,687	312,032	hg19 1k+
MESA	EA	Affymetrix	Affy6.0	2,358	700,473	hg19 1k+
NFBC66	EA (Finnish)	Illumina	HumanCNV370v1	5,263	328,401	hg19 1k+

S5. Association of *HP* structural variants to cholesterol levels

The association between imputed *HP* structural variants and the four lipids traits (total cholesterol (TC), low-density lipoprotein (LDL, high density lipoprotein (HDL), and triglycerides (TRIGS)) was performed in 6 studies of 22,288 individuals of European ancestry. Each lipid trait was regressed on age and gender, and inverse-normal transformed prior to analysis. Linear regression was performed to test the association between imputed structural variants or SNPs in the locus and lipid trait, assuming an additive genetic model using PLINK (v1.07). The imputed *HP* structural variants and genotypes were analyzed as dosages to account for imputation uncertainty, poorly imputed variants discarded (INFO<0.4). All analyses were adjusted for 10 study specific principal components. Study specific results were combined inverse-variance fixed effects meta-analysis method implemented in METAL[17]. Conditional analyses were performed for *HP* structural variants and SNPs in the region by adjusting for HP2 and the top lipid GWAS signal. Additional conditional analysis were performed for the four HP2 SNP-haplotypes adjusted each other and with inclusion of the top lipid GWAS signal (rs2000999). Sensitivity and specificity analyses were

performed to assess the influence of type 2 diabetes (adjustment or removal of samples) and cholesterol lowering/statin medication use (recalculating values, adjustment or removal of samples). All analyses were performed using baseline lipid measurements for cohorts with longitudinal follow-up.

Supplementary Tables

Supplementary Table 1. Neandertal, Denisova and ancient Ethiopian coverage of HP2FS haplotype. This table lists the position of each base that defines the HP2FS haplotype along with the number of mapped reads and uniquely mapped reads for the neandertal[18] and denisova[19] high-coverage genomes. Only uniquely mapped reads are shown for the ancient Ethiopian (Mota)[20] as the published bam file was filtered for a mapping quality of 30. The specific variant and CNV copy for each base is indicated on the right side of the table. Parts of the *HPR* gene conversion region in the *HP* gene lack uniquely-mapped reads due to high homology with the *HPR* gene.
(continues on next page)

			Denisova		Neandertal		Mota (ancient Ethiopian)		
Chromosome	Position	Base	Uniquely mapped reads	Mapped reads	Uniquely mapped reads	Mapped reads	Uniquely mapped reads	Variant	CNV copy
chr16	72091102	T	22	33	21	29	6	*HPR* gene conversion region	HP2FS-Left
chr16	72091103	G	23	35	18	26	6		
chr16	72091168	C	23	27	43	51	0		
chr16	72091169	T	22	27	42	51	0		
chr16	72091178	G	18	25	35	55	0		
chr16	72091180	C	17	25	34	57	0		
chr16	72091183	C	17	25	31	57	0		
chr16	72091185	T	16	24	29	57	0		
chr16	72091191	C	13	28	24	56	0		
chr16	72091192	T	12	27	24	57	0		
chr16	72091195	T	10	25	22	56	0		
chr16	72091201	T	7	22	18	59	0		
chr16	72091207	A	6	21	16	59	0		
chr16	72091216	C	4	22	12	61	0		
chr16	72091238	T	1	30	6	65	0		
chr16	72091245	G	1	36	4	64	0		
chr16	72091246	C	1	37	4	65	0		
chr16	72091258	T	0	40	3	68	0		
chr16	72091271	T	0	43	3	64	0		
chr16	72091295	A	0	35	2	61	0		
chr16	72091307	T	0	32	0	66	0		
chr16	72091308	G	0	32	0	63	0		
chr16	72091311	A	0	31	0	68	0		
chr16	72091382	G	0	19	0	40	0		
chr16	72091383	A	0	20	0	41	0		
chr16	72091384	G	0	20	0	39	0		
chr16	72091385	C	0	19	0	40	0		
chr16	72091386	A	0	21	0	43	0		
chr16	72091387	C	0	21	0	43	0		
chr16	72091388	T	0	21	0	47	0		
chr16	72091565	G	17	21	35	38	2	Highly diverged region Form L	
chr16	72091593	C	20	22	46	48	4		
chr16	72091598	G	21	22	47	48	4		
chr16	72091599	G	21	22	48	49	4		
chr16	72091602	C	21	21	48	49	5		
chr16	72091603	T	19	19	49	50	5		
chr16	72091614	C	21	21	50	51	4		
chr16	72091648	A	29	29	46	47	3		
chr16	72091808	C	34	34	49	50	13		
chr16	72091819	A	30	30	51	52	13		
chr16	72092029	C	18	18	64	65	17	HP2 breakpoint	
chr16	72092030	G	18	18	62	63	17		
chr16	72092031	C	18	18	62	62	17		
chr16	72092032	G	18	18	58	58	17		
chr16	72092033	A	18	18	60	60	17		
chr 16	72092034	G	18	18	60	60	17		
chr16	72092826	C	23	23	46	46	19	Ancestral *HP* bases	HP2FS-Right
chr16	72092827	A	23	23	45	45	19		
chr16	72092892	T	25	25	48	48	18		
chr16	72092893	C	23	23	48	48	19		
chr16	72092902	T	25	25	48	48	18		
chr16	72092904	T	24	24	48	48	18		
chr16	72092907	T	24	24	50	50	18		
chr16	72092909	C	23	23	50	50	17		
chr16	72092917	C	26	26	48	48	16		
chr16	72092923	C	25	25	51	51	16		
chr16	72092929	G	25	25	51	51	19		
chr16	72092938	T	21	21	61	61	20		
chr16	72092960	A	22	22	69	69	18		
chr16	72092967	C	22	22	68	68	24		
chr16	72092968	T	24	24	70	70	24		
chr16	72092980	C	28	28	68	68	21		
chr16	72092993	C	29	29	70	70	17		
chr16	72093017	G	28	28	56	56	11		
chr16	72093029	A	35	35	58	58	10		
chr16	72093030	C	35	35	58	58	9		
chr16	72093026	A	36	36	60	60	9		
chr16	72093097	G	17	22	34	42	7		
chr16	72093280	A	29	30	40	40	9	Highly diverged region Form R	
chr16	72093308	T	33	33	39	39	13		
chr16	72093313	T	31	31	37	37	14		
chr16	72093314	A	31	31	39	39	14		
chr16	72093317	T	33	33	41	41	16		
chr16	72093318	C	34	34	41	41	15		
chr16	72093329	T	28	28	38	38	11		
chr16	72093363	G	13	25	25	35	6		
chr16	72093523	A	38	38	37	37	11		
chr16	72093534	G	33	33	41	41	11		

Supplementary Table 2. Imputation and tag SNP correlations for structural variants using the Illumina Omni2.5 array. The correlation between SNPs and HP1/HP2 is increased in Africans due to a high-frequency recombinant haplotype (**Supplementary Figure 8**).

	Africans (YRI)	
Structure	Imputation (r^2)	Tag SNP (r^2)
HP1S	0.96	0.96 rs115558487
HP1F	0.92	0.82 rs7198572
HP2FS	0.90	0.76 rs11075922
HP2SS	NA	NA
HP1 vs. HP2	0.90	0.76 rs11075922

Supplementary Table 3. Imputation and tag SNP correlations for structural variants using the HapMap 3 SNPs.

	HapMap 3			
Population	Africans (YRI)		European (CEU, TSI)	
	Imputation (r^2)	Tag SNP (r^2)	Imputation (r^2)	Tag SNP (r^2)
HP1S	1	1 rs16973703	0.90	0.79 rs217181
HP1F	0.94	0.86 rs12928732	0.92	0.79 rs9302635
HP2FS	0.94	0.82 rs6499562	0.83	0.39 rs217181
HP2SS	NA	NA	0.51	0.28 rs1290842
HP1 vs. HP2	0.94	0.82 rs6499562	0.88	0.42 rs217181

Supplementary Table 4. Imputation and tag SNP correlations for structural variants using the Illumina 1M SNPs.

	Illumina 1M			
Population	Africans (YRI)		European (CEU, TSI)	
	Imputation (r^2)	Tag SNP (r^2)	Imputation (r^2)	Tag SNP (r^2)
HP1S	0.97	0.73 rs13331191	0.88	0.34 rs4788597
HP1F	0.90	0.86 rs12928732	0.92	0.79 rs9302635
HP2FS	0.92	0.82 rs6499562	0.83	0.24 rs9302635
HP2SS	NA	NA	0.34	0.28 rs12920842
HP1 vs. HP2	0.92	0.82 rs6499562	0.86	0.28 rs9302635

Supplementary Table 5. Imputation and tag SNP correlations for structural variants using the Affymetrix 6.0 SNPs.

	Affy 6.0			
Population	Africans (YRI)		European (CEU, TSI)	
	Imputation (r^2)	Tag SNP (r^2)	Imputation (r^2)	Tag SNP (r^2)
HP1S	1	0.94 rs16973636	0.93	0.79 rs217181
HP1F	0.92	0.74 rs16970661	0.88	0.79 rs9302635
HP2FS	0.92	0.8 rs8052408	0.85	0.39 rs217181
HP2SS	NA	NA	0.34	0.23 rs12921986
HP1 vs. HP2	0.92	0.8 rs8052408	0.88	0.42 rs217181

Supplementary Table 6. Association of HP2 to total cholesterol levels and LDL cholesterol levels with different covariates.

Trait	MarkerName	A1	A2	Frequency	Effect	StdErr	P-value	Note
TC	HP2	a	b	0.6387	0.072	0.011	2.82E-11	Raw value in mg (inverese-normal)
TC	HP2	a	b	0.6386	0.072	0.011	2.28E-11	Cond on statin use
TC	HP2	a	b	0.6383	0.072	0.011	2.69E-11	Cond on T2D stat
TC	HP2	a	b	0.6383	0.072	0.011	2.06E-11	Cond on statin + T2D stat
TC	HP2	a	b	0.6387	0.071	0.011	4.47E-11	Lipid val adj for statin Use
TC	HP2	a	b	0.6383	0.072	0.011	3.65E-11	Lipid val adj for statin Use cond T2D Stat
TC	HP2	a	b	0.6384	0.074	0.011	2.19E-11	Remove subjects on statin meds
TC	HP2	a	b	0.6419	0.076	0.012	4.25E-11	Remove subjects with T2D
TC	HP2	a	b	0.6415	0.079	0.012	1.02E-11	Remove subjects with T2D or statin use

Trait	MarkerName	A1	A2	Freq1	Effect	StdErr	P-value	Note
LDL	HP2	a	b	0.6393	0.064	0.011	4.26E-09	Raw value in mg (inverese-normal)
LDL	HP2	a	b	0.6393	0.064	0.011	3.84E-09	Cond on statin use
LDL	HP2	a	b	0.6389	0.063	0.011	6.26E-09	Cond on T2D stat
LDL	HP2	a	b	0.6389	0.063	0.011	5.52E-09	Cond on statin + T2D stat
LDL	HP2	a	b	0.6393	0.064	0.011	4.90E-09	Lipid val adj for statin Use
LDL	HP2	a	b	0.639	0.064	0.011	6.22E-09	Lipid val adj for statin Use cond T2D Stat
LDL	HP2	a	b	0.639	0.067	0.011	1.73E-09	Remove subjects on statin meds
LDL	HP2	a	b	0.642	0.065	0.012	2.31E-08	Remove subjects with T2D
LDL	HP2	a	b	0.6417	0.069	0.012	4.10E-09	Remove subjects with T2D or statin use

Trait	MarkerName	A1	A2	Freq1	Effect	StdErr	P-value	Note
HDL	HP2	a	b	0.6386	0.006	0.011	5.70E-01	Raw value in mg (inverese-normal)
HDL	HP2	a	b	0.6385	0.006	0.011	5.54E-01	Cond on statin use
HDL	HP2	a	b	0.638	0.005	0.011	6.66E-01	Cond on T2D stat
HDL	HP2	a	b	0.6379	0.005	0.011	6.44E-01	Cond on statin + T2D stat
HDL	HP2	a	b	0.6386	0.006	0.011	5.70E-01	Lipid val adj for statin Use
HDL	HP2	a	b	0.638	0.005	0.011	6.66E-01	Lipid val adj for statin Use cond T2D Stat

HDL	HP2	a	b	0.6383	0.006	0.011	5.63E-01	Remove subjects on statin meds
HDL	HP2	a	b	0.6418	0.002	0.012	8.46E-01	Remove subjects with T2D
HDL	HP2	a	b	0.6414	0.000	0.012	9.81E-01	Remove subjects with T2D or statin use

Trait	MarkerName	A1	A2	Freq1	Effect	StdErr	P-value	Note
TG	HP2	a	b	0.6386	0.033	0.011	2.25E-03	Raw value in mg (inverese-normal)
TG	HP2	a	b	0.6385	0.033	0.011	2.38E-03	Cond on statin use
TG	HP2	a	b	0.6379	0.036	0.011	7.81E-04	Cond on T2D stat
TG	HP2	a	b	0.6379	0.036	0.011	8.76E-04	Cond on statin + T2D stat
TG	HP2	a	b	0.6386	0.033	0.011	2.25E-03	Lipid val adj for statin Use
TG	HP2	a	b	0.6379	0.036	0.011	7.81E-04	Lipid val adj for statin Use cond T2D Stat
TG	HP2	a	b	0.6384	0.033	0.011	2.74E-03	Remove subjects on statin meds
TG	HP2	a	b	0.6418	0.043	0.012	2.28E-04	Remove subjects with T2D
TG	HP2	a	b	0.6415	0.044	0.012	1.86E-04	Remove subjects with T2D or statin use

Supplementary Table 7. LD between rs2000999 and the CNV on each SNP haplotype background

Haplotype	LD with rs2000999 (r^2)
HP2 on all haplotypes	0.14
HP2 on haplotype A	0.08
HP2 on haplotype B	0.82
HP2 on haplotype C	0.06

Supplementary Table 8. Association of HP structural variation and lipid levels (mg/dL beta estimates). Analyses were performed on untransformed lipid values to generate beta estimates in mg/dL units. Baseline lipid measures were regressed on HP2 CNV, rs2000999 and combined model adjusted for age, sex and 10 study specific PCAs.

MarkerName	Allele1	Allele2	Freq1	FreqSE	MinFreq	MaxFreq	Effect	StdErr	P-value	Direction	HetChiSq	df	P-value
HP2_hdl_mg	a	b	0.6391	0.0281	0.6165	0.7011	0.1249	0.154	4.17E-01	+--++++	3.674	6	0.7206
HP2_ldl_mg	a	b	0.6404	0.0278	0.6194	0.7012	2.1555	0.376	9.93E-09	+++++++	4.904	6	0.5562
HP2_tchol_mg	a	b	0.6386	0.0269	0.6165	0.7011	2.6051	0.4106	2.24E-10	+++-+++	9.324	6	0.1561
HP2_trigs_mg	a	b	0.6439	0.032	0.6165	0.7011	1.6192	0.8361	5.28E-02	++---++	8.729	6	0.1894
rs2000999_hdl_mg	a	b	0.1972	0.0076	0.1837	0.2043	-0.2203	0.1719	2.00E-01	--+----	3.698	6	0.7175
rs2000999_ldl_mg	a	b	0.1973	0.0077	0.1838	0.2043	2.1013	0.4189	5.28E-07	++-++++	8.92	6	0.1781
rs2000999_tchol_mg	a	b	0.1977	0.0074	0.1837	0.2044	2.3835	0.4598	2.17E-07	++-++++	7.444	6	0.2818
rs2000999_trigs_mg	a	b	0.1955	0.0082	0.1837	0.2044	1.9438	0.9161	3.39E-02	++-++++	11.93	6	0.06355
HP2_hdl_mg_ConD_rs2000999	a	b	0.6392	0.0281	0.6165	0.7011	0.234	0.1663	1.59E-01	+--++++	7.655	6	0.2645
HP2_ldl_mg_ConD_rs2000999	a	b	0.6405	0.0278	0.6194	0.7012	1.7058	0.406	2.65E-05	+++++++	1.784	6	0.9384
HP2_tchol_mg_ConD_rs2000999	a	b	0.6387	0.0269	0.6165	0.7011	2.1109	0.4433	1.92E-06	+++-+++	6.744	6	0.3452
HP2_trigs_mg_ConD_rs2000999	a	b	0.644	0.032	0.6165	0.7011	1.064	0.9025	2.38E-01	+++--++	14.405	6	0.02543
rs2000999_hdl_mg_ConD_HP2	a	b	0.1972	0.0076	0.1837	0.2043	-0.3219	0.1856	8.28E-02	-++----	7.636	6	0.266
rs2000999_ldl_mg_ConD_HP2	a	b	0.1973	0.0077	0.1838	0.2043	1.3885	0.4521	2.13E-03	++-++++	6.331	6	0.3872
rs2000999_tchol_mg_ConD_HP2	a	b	0.1977	0.0074	0.1837	0.2044	1.4903	0.4961	2.66E-03	++-++++	5.165	6	0.5228
rs2000999_trigs_mg_ConD_HP2	a	b	0.1955	0.0082	0.1837	0.2044	1.4634	0.9889	1.39E-01	++-+++-	17.556	6	0.00744

Supplementary Table 9. Combinations of diploid genotypes are a unique sum of two haplotypes. Boundary locations (A-E) are shown in Supplementary Fig. 4b. This conversion chart was used to determine subtypes from diploid copy number calls.

Boundary copy number assay	**A**	**B**	**C**	**D**	**E**
HP1S	0	0	0	0	0
HP1S	0	0	0	0	0
Diploid genotype combination	0	0	0	0	0
HP1F	1	1	0	0	0
HP1F	1	1	0	0	0
Diploid genotype combination	2	2	0	0	0
HP2FS	1	0	0	1	1
HP2FS	1	0	0	1	1
Diploid genotype combination	2	0	0	2	2
HP2SS	0	0	1	1	1
HP2SS	0	0	1	1	1
Diploid genotype combination	0	0	2	2	2
HP1S	0	0	0	0	0
HP1F	1	1	0	0	0
Diploid genotype combination	1	1	0	0	0
HP1S	0	0	0	0	0
HP2FS	1	0	0	1	1
Diploid genotype combination	1	0	0	1	1
HP1S	0	0	0	0	0
HP2SS	0	0	1	1	1

Diploid genotype combination	0	0	1	1	1
HP1F	1	1	0	0	0
HP2FS	1	0	0	1	1
Diploid genotype combination	2	1	0	1	1
HP1F	1	1	0	0	0
HP2SS	0	0	1	1	1
Diploid genotype combination	1	1	1	1	1
HP2FS	1	0	0	1	1
HP2SS	0	0	1	1	1
Diploid genotype combination	1	0	1	1	1

Supplementary Table 10. *HP* subtypes are in Hardy-Weinberg Equilibrium (HWE). A chi-squared test was used to calculate p-values for Hardy Weinberg equilibrium in each population.

Population	Genotype	Observed Genotype Counts	X^2	*p* value
CEU	HP1S, HP1S	6		

CEU	HP1S, HP1F	7		
CEU	HP1S, HP2FS	33		
CEU	HP1S, HP2SS	2		
CEU	HP1F, HP1F	1		
CEU	HP1F, HP2FS	24		
CEU	HP1F, HP2SS	0		
CEU	HP2FS, HP2FS	41		
CEU	HP2FS, HP2SS	5		
CEU	HP2SS, HP2SS	0	2.15	0.99
IBS	HP1S, HP1S	6		
IBS	HP1S, HP1F	8		
IBS	HP1S, HP2FS	26		
IBS	HP1S, HP2SS	2		
IBS	HP1F, HP1F	2		
IBS	HP1F, HP2FS	16		
IBS	HP1F, HP2SS	3		
IBS	HP2FS, HP2FS	29		
IBS	HP2FS, HP2SS	2		
IBS	HP2SS, HP2SS	0	4.02	0.91
TSI	HP1S, HP1S	5		
TSI	HP1S, HP1F	4		
TSI	HP1S, HP2FS	21		
TSI	HP1S, HP2SS	2		
TSI	HP1F, HP1F	1		
TSI	HP1F, HP2FS	15		
TSI	HP1F, HP2SS	0		
TSI	HP2FS, HP2FS	36		
TSI	HP2FS, HP2SS	2		
TSI	HP2SS, HP2SS	0	3.37	0.95
YRI	HP1S, HP1S	4		
YRI	HP1S, HP1F	19		
YRI	HP1S, HP2FS	15		
YRI	HP1F, HP1F	19		
YRI	HP1F, HP2FS	43		
YRI	HP2FS, HP2FS	18	0.8	0.98

Supplementary Table 11. *HP* subtypes are consistent with trio transmission.

Haploid structural alleles were determined from diploid copy number calls using Supplementary Table 1. Haplotypes in each trio are denoted as "-T" for transmitted or "-U " for untransmitted.

Source Individual	Relationship	Subtype
NA12146	father-T	HP1S
NA12146	father-U	HP2FS
NA12239	mother-T	HP2FS
NA12239	mother-U	HP2FS
NA12144	father-T	HP1F
NA12144	father-U	HP2FS
NA12145	mother-T	HP1S
NA12145	mother-U	HP2FS
NA07022	father-T	HP1F
NA07022	father-U	HP1F
NA07056	mother-T	HP2FS
NA07056	mother-U	HP1S
NA06993	father-T	HP2FS
NA06993	father-U	HP2FS
NA06985	mother-T	HP2FS
NA06985	mother-U	HP2FS
NA07034	father-T	HP1S
NA07034	father-U	HP2FS
NA07055	mother-T	HP2SS
NA07055	mother-U	HP2FS
NA12056	father-T	HP1S
NA12056	father-U	HP1F
NA12057	mother-T	HP2FS

NA12057	mother-U	HP2FS
NA07357	father-U	HP2FS
NA07357	father-T	HP2SS
NA07345	mother-T	HP2FS
NA07345	mother-U	HP1S
NA11881	father-T	HP2FS
NA11881	father-U	HP2FS
NA11882	mother-T	HP1S
NA11882	mother-U	HP1F
NA11839	father-T	HP2FS
NA11839	father-U	HP1S
NA11840	mother-T	HP2SS
NA11840	mother-U	HP2FS
NA11831	father-T	HP1F
NA11831	father-U	HP2FS
NA11832	mother-T	HP2FS
NA11832	mother-U	HP2FS
NA11829	father-T	HP2FS
NA11829	father-U	HP2FS
NA11830	mother-T	HP1F
NA11830	mother-U	HP2FS
NA12716	father-T	HP2FS
NA12716	father-U	HP2FS
NA12717	mother-T	HP2SS
NA12717	mother-U	HP1S
NA11994	father-T	HP1S
NA11994	father-U	HP1S
NA11995	mother-T	HP2FS
NA11995	mother-U	HP1S
NA11992	father-T	HP1S
NA11992	father-U	HP2FS

NA11993	mother-T	HP2FS
NA11993	mother-U	HP2FS
NA12264	father-T	HP2FS
NA12264	father-U	HP1F
NA12234	mother-T	HP2FS
NA12234	mother-U	HP2FS
NA12155	father-T	HP2FS
NA12155	father-U	HP2FS
NA12156	mother-T	HP1S
NA12156	mother-U	HP1F
NA12154	father-T	HP1S
NA12154	father-U	HP1S
NA12236	mother-T	HP2SS
NA12236	mother-U	HP2FS
NA12248	father-T	HP2FS
NA12248	father-U	HP2FS
NA12249	mother-T	HP1S
NA12249	mother-U	HP2FS
NA12005	father-T	HP2FS
NA12005	father-U	HP1F
NA12006	mother-T	HP1S
NA12006	mother-U	HP2FS
NA12003	father-T	HP2FS
NA12003	father-U	HP2FS
NA12004	mother-T	HP2FS
NA12004	mother-U	HP2FS
NA12750	father-T	HP1S
NA12750	father-U	HP1S
NA12751	mother-T	HP1S
NA12751	mother-U	HP2FS
NA12762	father-T	HP2FS

NA12762	father-U	HP2FS
NA12763	mother-T	HP2FS
NA12763	mother-U	HP2FS
NA12760	father-T	HP1F
NA12760	father-U	HP2FS
NA12761	mother-T	HP2FS
NA12761	mother-U	HP1S
NA12814	father-T	HP2FS
NA12814	father-U	HP2FS
NA12815	mother-T	HP2FS
NA12815	mother-U	HP1S
NA12812	father-T	HP2FS
NA12812	father-U	HP2FS
NA12813	mother-T	HP2FS
NA12813	mother-U	HP2FS
NA12874	father-T	HP1S
NA12874	father-U	HP1S
NA12875	mother-T	HP2FS
NA12875	mother-U	HP2FS
NA12872	father-T	HP2FS
NA12872	father-U	HP1S
NA12873	mother-T	HP2FS
NA12873	mother-U	HP2FS
NA12891	father-T	HP2FS
NA12891	father-U	HP2FS
NA12892	mother-T	HP2FS
NA12892	mother-U	HP2FS
NA06984	father-T	HP2FS
NA06984	father-U	HP1F
NA06989	mother-T	HP2FS
NA06989	mother-U	HP2FS

NA12342	father-T	HP1F
NA12342	father-U	HP2FS
NA12343	mother-T	HP2FS
NA12343	mother-U	HP2FS
NA12340	father-T	HP1S
NA12340	father-U	HP2FS
NA12341	mother-T	HP2FS
NA12341	mother-U	HP1F
NA07347	father-T	HP2FS
NA07347	father-U	HP1F
NA07346	mother-T	HP2SS
NA07346	mother-U	HP1S
NA12546	father-T	HP2FS
NA12546	father-U	HP2FS
NA12489	mother-T	HP1F
NA12489	mother-U	HP2FS
NA12399	father-T	HP1F
NA12399	father-U	HP2FS
NA12400	mother-T	HP1S
NA12400	mother-U	HP2FS
NA12414	mother-T	HP2FS
NA12414	mother-U	HP1F
NA12413	father-T	HP2FS
NA12413	father-U	HP1S
NA11893	father-T	HP2FS
NA11893	father-U	HP2FS
NA11894	mother-T	HP1S
NA11894	mother-U	HP2FS
NA11891	father-T	HP1S
NA11891	father-U	HP1F
NA11892	mother-T	HP2FS

NA11892	mother-U	HP1F
NA12274	father-T	HP1S
NA12274	father-U	HP2FS
NA12275	mother-T	HP2FS
NA12275	mother-U	HP2FS
NA12272	father-T	HP1F
NA12272	father-U	HP2FS
NA12273	mother-T	HP1S
NA12273	mother-U	HP2FS
NA12286	father-T	HP1F
NA12286	father-U	HP2FS
NA12287	mother-T	HP1F
NA12287	mother-U	HP1S
NA11919	father-T	HP2FS
NA11919	father-U	HP2FS
NA11920	mother-T	HP2FS
NA11920	mother-U	HP1F
NA11917	father-T	HP1S
NA11917	father-U	HP2FS
NA11918	mother-T	HP2FS
NA11918	mother-U	HP2FS
NA11930	father-T	HP2FS
NA11930	father-U	HP2FS
NA11931	mother-T	HP2FS
NA11931	mother-U	HP2FS
NA12748	father-T	HP2FS
NA12748	father-U	HP1S
NA12749	mother-T	HP2FS
NA12749	mother-U	HP1F
NA12777	father-T	HP2FS
NA12777	father-U	HP2FS

NA12778	mother-T	HP2SS
NA12778	mother-U	HP2FS
NA12775	father-T	HP1S
NA12775	father-U	HP1S
NA12776	mother-T	HP2FS
NA12776	mother-U	HP2FS
NA12829	father-T	HP2FS
NA12829	father-U	HP2FS
NA12830	mother-T	HP2FS
NA12830	mother-U	HP1S
NA12827	father-T	HP1F
NA12827	father-U	HP2FS
NA12828	mother-T	HP2FS
NA12828	mother-U	HP2FS
NA12842	father-T	HP2FS
NA12842	father-U	HP2FS
NA12843	mother-T	HP1S
NA12843	mother-U	HP2FS
NA12889	father-T	HP2FS
NA12889	father-U	HP2FS
NA12890	mother-T	HP1F
NA12890	mother-U	HP2FS
NA12347	father-T	HP1F
NA12347	father-U	HP1S
NA12348	mother-T	HP1S
NA12348	mother-U	HP1S
NA06986	father-T	HP2FS
NA06986	father-U	HP2FS
NA07045	mother-T	HP1S
NA07045	mother-U	HP2FS
NA07435	father-T	HP1S

NA07435	father-U	HP2FS
NA07037	mother-T	HP2FS
NA07037	mother-U	HP1F
NA07051	father-T	HP2FS
NA07051	father-U	HP2FS
NA07031	mother-T	HP2FS
NA07031	mother-U	HP1S
HG01500	father-T	HP2FS
HG01500	father-U	HP1S
HG01501	mother-T	HP2FS
HG01501	mother-U	HP2FS
HG01503	father-T	HP1S
HG01503	father-U	HP2FS
HG01504	mother-T	HP1F
HG01504	mother-U	HP1S
HG01506	father-T	HP1F
HG01506	father-U	HP2FS
HG01507	mother-T	HP2FS
HG01507	mother-U	HP2FS
HG01509	father-T	HP2FS
HG01509	father-U	HP2FS
HG01510	mother-T	HP2FS
HG01510	mother-U	HP1F
HG01512	father-T	HP1S
HG01512	father-U	HP1S
HG01513	mother-T	HP2FS
HG01513	mother-U	HP1S
HG01515	father-T	HP2FS
HG01515	father-U	HP1F
HG01516	mother-T	HP2FS
HG01516	mother-U	HP2FS

HG01518	father-T	HP2FS
HG01518	father-U	HP2FS
HG01519	mother-T	HP2FS
HG01519	mother-U	HP1S
HG01521	father-T	HP2FS
HG01521	father-U	HP1F
HG01522	mother-T	HP1F
HG01522	mother-U	HP1S
HG01524	father-T	HP2FS
HG01524	father-U	HP2SS
HG01525	mother-T	HP1S
HG01525	mother-U	HP1F
HG01527	father-T	HP1S
HG01527	father-U	HP1S
HG01528	mother-T	HP1S
HG01528	mother-U	HP1S
HG01530	father-T	HP1F
HG01530	father-U	HP2SS
HG01531	mother-T	HP2FS
HG01531	mother-U	HP1F
HG01536	father-T	HP1F
HG01536	father-U	HP2FS
HG01537	mother-T	HP2FS
HG01537	mother-U	HP1S
HG01631	father-T	HP1S
HG01631	father-U	HP2SS
HG01632	mother-T	HP2FS
HG01632	mother-U	HP2FS
HG01630	father-T	HP2FS
HG01630	father-U	HP2FS
HG01628	mother-T	HP2FS

HG01628	mother-U	HP1S
HG01625	father-T	HP2FS
HG01625	father-U	HP1F
HG01626	mother-T	HP2FS
HG01626	mother-U	HP2FS
HG01624	father-T	HP2FS
HG01624	father-U	HP2FS
HG01623	mother-T	HP1S
HG01623	mother-U	HP2FS
HG01619	father-T	HP2FS
HG01619	father-U	HP1S
HG01620	mother-T	HP2FS
HG01620	mother-U	HP2FS
HG01617	father-T	HP1S
HG01617	father-U	HP2FS
HG01618	mother-T	HP2FS
HG01618	mother-U	HP2FS
HG01615	father-T	HP2FS
HG01615	father-U	HP1S
HG01613	mother-T	HP2FS
HG01613	mother-U	HP1S
HG01610	father-T	HP1S
HG01610	father-U	HP2FS
HG01612	mother-T	HP2FS
HG01612	mother-U	HP1F
HG01608	father-T	HP1F
HG01608	father-U	HP1F
HG01607	mother-T	HP1S
HG01607	mother-U	HP2SS
HG01603	father-T	HP2FS
HG01603	father-U	HP1S

HG01602	mother-T	HP2FS
HG01602	mother-U	HP2FS
HG01672	father-T	HP1S
HG01672	father-U	HP1S
HG01670	mother-T	HP1F
HG01670	mother-U	HP1S
HG01675	father-T	HP1S
HG01675	father-U	HP2FS
HG01673	mother-T	HP1F
HG01673	mother-U	HP2FS
HG01678	father-T	HP2FS
HG01678	father-U	HP2FS
HG01676	mother-T	HP2FS
HG01676	mother-U	HP2FS
HG01680	father-T	HP1F
HG01680	father-U	HP1S
HG01679	motheR-T	HP1F
HG01679	mother-U	HP2SS
HG01682	father-T	HP2FS
HG01682	father-U	HP2FS
HG01684	mother-T	HP2FS
HG01684	mother-U	HP2FS
HG01686	father-T	HP2FS
HG01686	father-U	HP2FS
HG01685	mother-T	HP2FS
HG01685	mother-U	HP2FS
HG01694	father-T	HP2FS
HG01694	father-U	HP2FS
HG01695	mother-T	HP1F
HG01695	mother-U	HP2FS
HG01699	father-T	HP2FS

HG01699	father-U	HP2FS
HG01697	mother-T	HP2FS
HG01697	mother-U	HP2FS
HG01700	father-T	HP2FS
HG01700	father-U	HP1F
HG01702	mother-T	HP1F
HG01702	mother-U	HP1S
HG01705	father-T	HP1F
HG01705	father-U	HP2FS
HG01704	mother-T	HP2FS
HG01704	mother-U	HP1S
HG01708	father-T	HP1F
HG01708	father-U	HP2SS
HG01707	mother-T	HP1S
HG01707	mother-U	HP1S
HG01709	father-T	HP1S
HG01709	father-U	HP1F
HG01710	mother-T	HP2FS
HG01710	mother-U	HP2FS
HG01747	father-T	HP2FS
HG01747	father-U	HP2FS
HG01746	mother-T	HP2FS
HG01746	mother-U	HP2FS
HG01756	father-T	HP2FS
HG01756	father-U	HP1F
HG01757	mother-T	HP1F
HG01757	mother-U	HP1F
HG01771	father-T	HP1F
HG01771	father-U	HP1S
HG01770	mother-T	HP1S
HG01770	mother-U	HP2FS

HG01775	father-T	HP2FS
HG01775	father-U	HP1S
HG01773	mother-T	HP1S
HG01773	mother-U	HP1S
HG01777	father-T	HP2FS
HG01777	father-U	HP2FS
HG01776	mother-T	HP2FS
HG01776	mother-U	HP2FS
HG01781	father-T	HP2FS
HG01781	father-U	HP1S
HG01779	mother-T	HP2FS
HG01779	mother-U	HP1S
HG01783	father-T	HP2FS
HG01783	father-U	HP2FS
HG01784	mother-T	HP1S
HG01784	mother-U	HP2FS
HG02224	father-T	HP2FS
HG02224	father-U	HP1S
HG02223	mother-T	HP2FS
HG02223	mother-U	HP1F
HG02236	father-T	HP2FS
HG02236	father-U	HP1S
HG02235	mother-T	HP2SS
HG02235	mother-U	HP2FS
HG02238	father-T	HP2FS
HG02238	father-U	HP2FS
HG02239	mother-T	HP2FS
HG02239	mother-U	HP1S
NA18501	father-T	HP1F
NA18501	father-U	HP1F
NA18502	mother-T	HP1S

NA18502	mother-U	HP2FS
NA18504	father-T	HP2FS
NA18504	father-U	HP1S
NA18505	mother-T	HP2FS
NA18505	mother-U	HP1F
NA18859	father-T	HP2FS
NA18859	father-U	HP1S
NA18858	mother-T	HP2FS
NA18858	mother-U	HP1F
NA18522	father-T	HP2FS
NA18522	father-U	HP1F
NA18523	mother-T	HP2FS
NA18523	mother-U	HP2FS
NA18871	father-T	HP1F
NA18871	father-U	HP1F
NA18870	mother-T	HP1F
NA18870	mother-U	HP2FS
NA18853	father-T	HP1F
NA18853	father-U	HP1F
NA18852	mother-T	HP1F
NA18852	mother-U	HP2FS
NA18856	father-T	HP1F
NA18856	father-U	HP2FS
NA18855	mother-T	HP1F
NA18855	mother-U	HP1S
NA18862	father-T	HP2FS
NA18862	father-U	HP2FS
NA18861	mother-T	HP1S
NA18861	mother-U	HP1F
NA18913	father-T	HP1F
NA18913	father-U	HP2FS

NA18912	mother-T	HP1F
NA18912	mother-U	HP2FS
NA19101	father-T	HP2FS
NA19101	father-U	HP1S
NA19102	mother-U	HP2FS
NA19102	mother-U	HP2FS
NA19138	father-T	HP1F
NA19138	father-U	HP1S
NA19137	mother-T	HP1F
NA19137	mother-U	HP1F
NA19200	father-T	HP1S
NA19200	father-U	HP1F
NA19201	mother-T	HP1F
NA19201	mother-U	HP2FS
NA19171	father-T	HP1S
NA19171	father-U	HP1F
NA19172	mother-T	HP2FS
NA19172	mother-U	HP1F
NA19210	father-T	HP1S
NA19210	father-U	HP2FS
NA19209	mother-T	HP1S
NA19209	mother-U	HP2FS
NA19160	father-T	HP1F
NA19160	father-U	HP1F
NA19159	mother-T	HP2FS
NA19159	mother-U	HP2FS
NA19223	father-T	HP2FS
NA19223	father-U	HP1F
NA19222	mother-T	HP2FS
NA19222	mother-U	HP1F
NA19141	father-T	HP2FS

NA19141	father-U	HP2FS
NA19140	mother-T	HP1S
NA19140	mother-U	HP1S
NA19153	father-T	HP2FS
NA19153	father-U	HP1F
NA19152	mother-T	HP2FS
NA19152	mother-U	HP2FS
NA19144	father-T	HP1F
NA19144	father-U	HP2FS
NA19143	mother-T	HP1S
NA19143	mother-U	HP1F
NA19128	father-T	HP2FS
NA19128	father-U	HP1S
NA19127	mother-T	HP1F
NA19127	mother-U	HP1F
NA19130	father-T	HP1F
NA19130	father-U	HP1F
NA19131	mother-T	HP1F
NA19131	mother-U	HP2FS
NA19098	father-T	HP2FS
NA19098	father-U	HP2FS
NA19099	mother-T	HP2FS
NA19099	mother-U	HP2FS
NA19192	father-T	HP2FS
NA19192	father-U	HP1F
NA19193	mother-T	HP1F
NA19193	mother-U	HP1F
NA19239	father-T	HP1F
NA19239	father-U	HP1F
NA19238	mother-T	HP2FS
NA19238	mother-U	HP2FS

NA18486	father-T	HP1F
NA18486	father-U	HP1F
NA18488	mother-T	HP1F
NA18488	mother-U	HP1F
NA18487	father-T	HP2FS
NA18487	father-U	HP1F
NA18489	mother-T	HP1F
NA18489	mother-U	HP1S
NA18498	father-T	HP1F
NA18498	father-U	HP1F
NA18499	mother-T	HP2FS
NA18499	mother-U	HP2FS
NA19107	father-T	HP1F
NA19107	father-U	HP2FS
NA19108	mother-T	HP2FS
NA19108	mother-U	HP1S
NA18868	father-T	HP1F
NA18868	father-U	HP2FS
NA18867	mother-T	HP1F
NA18867	mother-U	HP1F
NA18519	father-T	HP1F
NA18519	father-U	HP2FS
NA18520	mother-T	HP1F
NA18520	mother-U	HP1F
NA18874	father-T	HP1F
NA18874	father-U	HP1F
NA18873	mother-T	HP1F
NA18873	mother-U	HP2FS
NA18910	father-T	HP1F
NA18910	father-U	HP1F
NA18909	mother-T	HP1S

NA18909	mother-U	HP1F
NA18917	father-T	HP2FS
NA18917	father-U	HP2FS
NA18916	mother-T	HP1S
NA18916	mother-U	HP1F
NA18923	father-T	HP2FS
NA18923	father-U	HP2FS
NA18924	mother-T	HP1F
NA18924	mother-U	HP1F
NA19198	father-T	HP1S
NA19198	father-U	HP1F
NA19197	mother-T	HP2FS
NA19197	mother-U	HP2FS
NA18934	father-T	HP1S
NA18934	father-U	HP1S
NA18933	mother-T	HP1F
NA18933	mother-U	HP1S
NA19178	father-T	HP1F
NA19178	father-U	HP1S
NA19179	mother-T	HP1F
NA19179	mother-U	HP1F
NA19184	father-T	HP2FS
NA19184	father-U	HP1F
NA19185	mother-T	HP1S
NA19185	mother-U	HP1S
NA19096	father-T	HP2FS
NA19096	father-U	HP2FS
NA19095	mother-T	HP1S
NA19095	mother-U	HP1F
NA19175	father-T	HP1S
NA19175	father-U	HP2FS

NA19176	mother-T	HP1F
NA19176	mother-U	HP1F
NA19226	father-T	HP2FS
NA19226	father-U	HP1S
NA19225	mother-T	HP1F
NA19225	mother-U	HP2FS
NA19121	father-T	HP1F
NA19121	father-U	HP2FS
NA19122	mother-T	HP1F
NA19122	mother-U	HP1S
NA19150	father-T	HP1S
NA19150	father-U	HP1F
NA19149	mother-T	HP1S
NA19149	mother-U	HP1F
NA19113	father-T	HP2FS
NA19113	father-U	HP1F
NA19114	mother-T	HP1S
NA19114	mother-U	HP2FS
NA19256	father-T	HP1F
NA19256	father-U	HP1F
NA19257	mother-T	HP2FS
NA19257	mother-U	HP1S
NA19117	father-T	HP2FS
NA19117	father-U	HP1S
NA19118	mother-T	HP1S
NA19118	mother-U	HP1S
NA19189	father-T	HP2FS
NA19189	father-U	HP2FS
NA19190	mother-T	HP2FS
NA19190	mother-U	HP2FS
NA19236	father-T	HP2FS

NA19236	father-U	HP1F
NA19235	mother-T	HP1S
NA19235	mother-U	HP1F
NA19248	father-T	HP2FS
NA19248	father-U	HP1F
NA19247	mother-T	HP1S
NA19247	mother-U	HP2FS

Supplementary Table 12. Rare *HP* triplication subtype and base pair differences that disrupt boundary assays. The table below indicates the sample that did not match one of four structural alleles (likely *HP* triplication) as well as individuals containing single base pair mutations that interfere with the boundary breakpoint assays as shown in Supplementary Fig. 4b. The remaining 580 haplotypes typed in this study contained the standard boundaries as typed by the assays in Supplementary Fig. 4b and were consistent with one of the four common subtypes.

Low Frequency Variant	**Carriers**	**Evidence**
Likely triplication Copy number 3 at HP2 (breakpoint assay E)	NA20806	triplicate ddPCR measurements
T to C mutation at chr16: 72091054 which disrupts breakpoint assay A on HP2FS haplotype	NA20524, NA20803, HG01783, HG01782, NA12892	Sanger sequencing
T to C mutation at chr16: 72091102 which disrupts breakpoint assay A on HP2FS haplotype	HG01686, NA12873, NA12864	Sanger sequencing

Supplementary Table 13. Phasing recent deletion alleles with SNP haplotypes using Drop-Phase[21]. The CNV was phased with nucleotide polymorphisms on either side of the CNV for each individual carrying a recent deletion allele. The numbers indicate the linkage statistic for each allele. The order in which the nucleotides are listed corresponds to the order of the numbers beneath each individual. For example, individual NA20502 has a linkage statistic of 105 with the A allele and 0 with the T allele at rs113811215.

SNP \\ Individual	NA11843	NA18502	NA20502	NA20786
rs113811215 A/T	NA	N/A	105/0	NA
rs12917999 C/T	0/182	NA	NA	80/0
rs34682685 A/G	NA	NA	0/139	160/0
rs7189591 G/T	191/0	NA	NA	0/151
rs138230207 A/G	NA	57/0	NA	NA

Supplementary Table 14. Primer Sequences. The hg19 genome build has the HP2FS subtype. If a given assay will amplify HP2FS sequence, the genomic coordinates of the amplicon are provided.

Technology	Assay	Sequence (5' to 3')	Targeted sequence (hg19 coordinates)
Sanger sequencing	HPregion_F	CCAGGGCCAAAGTTTGTAGA	chr16:72088426-72095328
	HPregion_R	TCCTAATTCACTTTGAGGAGGC	
	HPregion_F	CCAGGGCCAAAGTTTGTAGA	chr16:72088426-72089409
	HPregion_int1_R	TCTGCCAGTTCACCATCTCA	
	HP_FI1_F	GCTTCCTGGCATCCTGAGTA	chr16:72088835-72089704
	HP_FI1_R	CCCTCCAGGAAAGAGAAACC	
	HPregion_int1_F	AAGTAGTGCCCGAATGGTTG	chr16:72089359-72090383
	HPregion_int2int4_R	CCCAGTGAACCGTGAAAAGT	
	HP_FI2_F	GAGCTGTCATTGCCCTCCT	chr16:72090068-72091063
	HP_FI2_R	TCACAGCCAAGCAGTACAGG	
	HPregion_int2_F	AGAAATGAGGGGAGCTTGC	chr16:72090281-72091443
	HPregion_int3int6_R	ACTGGAAGGCTGTGCCTCTA	
	HP_FI3_F	GTCCTGGCACTGCTCTAAGG	chr16:72090860-72091767
	HP_FI3_R	AACGAGAAGCAGGAGTTCCA	
	HPregion_int3int6_F	TCCTGAATGTGAAGCAGGTG	chr16:72093071-72094116
	HPregion_ex7_R	TGGTGGGAAACCATCTTAGC	

	HP_FI4_F	CCAGGGCTCAAAAATCTCAG	chr16:72093750-72095328
	HPregion_R	TCCTAATTCACTTTGAGGAGGC	
	HPregion_ex7_F	ATCCTGGGTGGACACCTG	chr16:72094052-72095328
	HPregion_R	TCCTAATTCACTTTGAGGAGGC	
	HPregion_F	CCAGGGCCAAAGTTTGTAGA	chr16:72088426-72092042
	HP2Left_R	AATAGAAGCTCGCGAACTGTATTAT	
	HP2Right_F	ATAATACAGTTCGCGAGCTTCTATT	chr16:72092018-72095328
	HPregion_R	TCCTAATTCACTTTGAGGAGGC	
droplet-digital PCR	A_F	GAGCCCTCCCTGTACTGCT	chr16:72091036-72091119
	A_R	AAGCCAAGTGCCTGGACA	
	A_probe	FAM-CCGCCATGACCACAGTGTGTTC	
	B_F	TGCTGAGCACTGAGCACTTA	HP1F only
	B_R	TCCTAATCTGTGGGCGATGT	
	B_probe	FAM-CCAGTGCGGCTTCCTCTGAGC	
	D_F	GGCTGAATCCACTGTCGG	chr16:72091582-72091721
	D_R	TCCATGGATTAAGATTTTATGTCATT	
	D_probe	FAM-TGCCCACAGATCAGGAGAGCCTG	
	E_F	TGGAAATTCCTTTATTGGGATAAT	chr16:72091985-72092074
	E_R	GCTGCTCTGCACATCAATCT	
	E_probe	FAM-CGCGAGCTTCTATTCGGGGTG	
	control_RPP30_F	GATTTGGACCTGCGAGCG	chr10:92631759-92631820
	control_RPP30_R	GCGGCTGTCTCCACAAGT	
	control_RPP30_probe	VIC-CTGACCTGAAGGCTCT	

	control_UCE_F	GCTATAATAGAAGGGGGAAGTCG	
	control_UCE_R	ATTGTGGGCCTGTTTTGAAT	
	control_UCE_probe	VIC-ATTGCGTCGCTCTGAGCCCC	chr12:106976559-106976654
PCR	C_F	TCACTCTGCACGGGGTCT	HP2SS only
	C_R	TCTCCTGATCTGTGGGCAG	

Supplementary Table 15. Population Identifiers.

CEU	Utah residents with Northern and Western European ancestry
IBS	Iberian population in Spain
TSI	Toscani in Italia
YRI	Yoruba in Ibadan, Nigeria

dbGaP Study Acknowledgements

ARIC
The Atherosclerosis Risk in Communities Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C). Funding for GENEVA genotyping was provided by National Human Genome Research Institute grant U01HG004402 (E. Boerwinkle). This manuscript was not prepared in collaboration with investigators of the Atherosclerosis Risk in Communities Study and does not necessarily reflect the opinions or views of the Atherosclerosis Risk in Communities Study.

CARDIA
The Coronary Artery Risk Development in Young Adults Study (CARDIA) is conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with the University of Alabama at Birmingham (N01-HC95095 & N01-HC48047), University of Minnesota (N01-HC48048), Northwestern University (N01-HC48049), and Kaiser Foundation Research Institute (N01-HC48050). Genotyping for the CARDIA GENEVA cohort was supported by grant U01 HG004729 from the National Human Genome Research Institute. This manuscript was not approved by CARDIA. The opinions and conclusions contained in this publication are solely those of the authors, and are not endorsed by CARDIA or the NHLBI and should not be assumed to reflect the opinions or conclusions of either CARDIA, or the NHLBI.

CHS
The Cardiovascular Health Study (CHS) research reported in this article was supported by contract numbers N01-HC-85079, N01-HC-85080, N01-HC-85081, N01-HC-85082, N01-HC-85083, N01-HC-85084, N01-HC-85085, N01-HC-85086, N01-HC-35129, N01 HC-15103, N01 HC-55222, N01-HC-75150, N01-HC-45133, N01-HC-85239 and HHSN268201200036C; grant numbers U01 HL080295 from the National Heart, Lung, and Blood Institute and R01 AG-023629 from the National Institute on Aging, with additional contribution from the National Institute of Neurological Disorders and Stroke. A full list of principal CHS investigators and institutions can be found at http://www.chs-nhlbi.org/pi.htm. Support for the Cardiovascular Health Study Whole Genome Study was provided by NHLBI grant HL087652. Additional support for infrastructure was provided by HL105756 and additional genotyping among the African-American cohort was supported in part by HL085251. DNA handling and genotyping at Cedars-Sinai Medical Center was supported in part by National Center for Research Resources grant UL1RR033176, now at the National Center for Advancing Translational Technologies CTSI grant UL1TR000124; in addition to the National Institute of Diabetes and Digestive and Kidney Diseases grant DK063491 to the Southern California Diabetes Endocrinology Research Center. This manuscript

was not prepared in collaboration with CHS investigators and does not necessarily reflect the opinions or views of CHS, or the NHLBI.

FUSION
The Finland-United States Investigation of NIDDM Genetics (FUSION) Study was conducted by the FUSION Study Investigators and supported by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK). Genotyping for the FUSION study was supported by NIH HHSN268200782096C award. The data from the FUSION reported here were supplied by the dbGaP and NIDDK Central Repositories. This manuscript was not prepared in collaboration with Investigators of the FUSION study and does not necessarily reflect the opinions or views of the FUSION study, the NIDDK Central Repositories, or the NIDDK.

MESA
The Multi-Ethnic Study of Atherosclerosis (MESA) and the MESA SHARe project are conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for MESA is provided by contracts N01-HC-95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-95163, N01-HC-95164, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169 and CTSA UL1-RR-024156. Funding for SHARe genotyping was provided by NHLBI Contract N02-HL-64278. Genotyping was performed at Affymetrix (Santa Clara, California, USA) and the Broad Institute of Harvard and MIT (Boston, Massachusetts, USA) using the Affymetric Genome-Wide Human SNP Array 6.0. This manuscript was not prepared in collaboration with MESA investigators and does not necessarily reflect the opinions or views of MESA, or the NHLBI.

NFBC66
The NFBC1966 Study is conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with the Broad Institute, UCLA, University of Oulu, and the National Institute for Health and Welfare in Finland. This manuscript was not prepared in collaboration with investigators of the NFBC1966 Study and does not necessarily reflect the opinions or views of the NFBC1966 Study Investigators, Broad Institute, UCLA, University of Oulu, National Institute for Health and Welfare in Finland and the NHLBI

References

1. Hindson, B. J. *et al.* High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. *Anal. Chem.* **83,** 8604–8610 (2011).
2. 1000 Genomes Project Consortium *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491,** 56–65 (2012).
3. International HapMap Consortium *et al.* A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449,** 851–861 (2007).
4. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. *Am. J. Epidemiol.* **129,** 687–702 (1989).
5. Korn, J. M. *et al.* Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nat Genet* **40,** 1253–1260 (2008).
6. Fried, L. P. *et al.* Risk factors for 5-year mortality in older adults: the Cardiovascular Health Study. *JAMA* **279,** 585–592 (1998).
7. Gardin, J. M. *et al.* Relationship of cardiovascular risk factors to echocardiographic left ventricular mass in healthy young black and white adult men and women. The CARDIA study. Coronary Artery Risk Development in Young Adults. *Circulation* **92,** 380–387 (1995).
8. Valle, T., Tuomilehto, J., Bergman, R. N. & Ghosh, S. Mapping Genes for NIDDM: Design of the Finland—United States Investigation of NIDDM Genetics (FUSION) Study. *Diabetes* (1998).
9. Bild, D. E., Bluemke, D. A. & Burke, G. L. Multi-ethnic study of atherosclerosis: objectives and design. *American journal of …* (2002).
10. Järvelin, M. R., Sovio, U., King, V., Lauren, L. & Xu, B. Early life factors and blood pressure at age 31 years in the 1966 northern Finland birth cohort. … (2004).
11. Dreszer, T. R. *et al.* The UCSC Genome Browser database: extensions and updates 2011. *Nucleic Acids Res* **40,** D918–23 (2012).
12. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38,** 904–909 (2006).
13. International HapMap Consortium. A haplotype map of the human genome. *Nature* **437,** 1299–1320 (2005).
14. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81,** 559–575 (2007).
15. Delaneau, O., Zagury, J.-F. & Marchini, J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat. Methods* **10,** 5–6 (2013).
16. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G. R. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet* **44,** 955–959 (2012).
17. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26,** 2190–2191

(2010).
18. Prüfer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505,** 43–49 (2014).
19. Meyer, M., Kircher, M., Gansauge, M. T., Li, H. & Racimo, F. A high-coverage genome sequence from an archaic Denisovan individual. *Science* (2012).
20. Llorente, M. G. *et al.* Ancient Ethiopian genome reveals extensive Eurasian admixture throughout the African continent. *Science* (2015). doi:10.1126/science.aad2879
21. Regan, J. F. *et al.* A rapid molecular approach for chromosomal phasing. *PLoS ONE* **10,** e0118270 (2015).