



# Multilocus Heterozygosity and Coronary Heart Disease: Nested Case-Control Studies in Men and Women

## Citation

Mukamal, Kenneth J., Majken K. Jensen, Tune H. Pers, Jennifer K. Pai, Peter Kraft, and Eric B. Rimm. 2015. "Multilocus Heterozygosity and Coronary Heart Disease: Nested Case-Control Studies in Men and Women." PLoS ONE 10 (5): e0124847. doi:10.1371/journal.pone.0124847. <http://dx.doi.org/10.1371/journal.pone.0124847>.

## Published Version

[doi:10.1371/journal.pone.0124847](https://doi.org/10.1371/journal.pone.0124847)

## Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:16120881>

## Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

## Share Your Story

The Harvard community has made this article openly available.  
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

RESEARCH ARTICLE

# Multilocus Heterozygosity and Coronary Heart Disease: Nested Case-Control Studies in Men and Women

Kenneth J. Mukamal<sup>1\*</sup>, Majken K. Jensen<sup>2</sup>, Tune H. Pers<sup>3,4,5,6,7</sup>, Jennifer K. Pai<sup>2,8</sup>, Peter Kraft<sup>9,10</sup>, Eric B. Rimm<sup>2,8,9</sup>

**1** Department of Medicine, Beth Israel Deaconess Medical Center, Boston, Massachusetts, United States of America, **2** Department of Nutrition, Harvard School of Public Health, Boston, Massachusetts, United States of America, **3** Center for Biological Sequence Analysis, Technical University of Denmark, Lyngby, Denmark, **4** Division of Genetics, Children's Hospital, Boston, Massachusetts, United States of America, **5** Division of Endocrinology, Children's Hospital, Boston, Massachusetts, United States of America, **6** Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts, United States of America, **7** Institute of Preventive Medicine, Copenhagen University Hospital, Copenhagen, Denmark, **8** Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, United States of America, **9** Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, United States of America, **10** Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts, United States of America

\* [kmukamal@bidmc.harvard.edu](mailto:kmukamal@bidmc.harvard.edu)



OPEN ACCESS

**Citation:** Mukamal KJ, Jensen MK, Pers TH, Pai JK, Kraft P, Rimm EB (2015) Multilocus Heterozygosity and Coronary Heart Disease: Nested Case-Control Studies in Men and Women. PLoS ONE 10(5): e0124847. doi:10.1371/journal.pone.0124847

**Academic Editor:** Gualtiero I. Colombo, Centro Cardiologico Monzino IRCCS, ITALY

**Received:** June 24, 2014

**Accepted:** March 19, 2015

**Published:** May 13, 2015

**Copyright:** © 2015 Mukamal et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** Data are from the Nurses Health Study and Health Professionals Follow-up Study whose authors may be contacted at [nhspermission@channing.harvard.edu](mailto:nhspermission@channing.harvard.edu). Procedures to obtain access are described at [http://www.channing.harvard.edu/nhs/?page\\_id=471](http://www.channing.harvard.edu/nhs/?page_id=471).

**Funding:** This study was supported by grants HL35464, CA55075, and CA87969 from the National Institutes of Health, Bethesda, Maryland (EBR), with additional support for genotyping from Merck/Rosetta Research Laboratories, North Wales, Pennsylvania. The funders had no role in study design, data

## Abstract

### Background

Generalized allelic heterozygosity has been proposed to improve reproductive fitness and has been associated with higher blood pressure, but its association with chronic disease is not well characterized.

### Methods

Using the Affymetrix Genome-Wide Human 6.0 array, we performed whole genome scans in parallel case-control studies of coronary heart disease (CHD) nested in the Health Professionals Follow-up Study and Nurses' Health Study. We examined ~700,000 single nucleotide polymorphisms (SNPs) in 435 men with incident CHD and 878 matched controls and 435 women with incident CHD with 931 matched controls. We examined the relationship of genome-wide heterozygosity with risk of incident of CHD and with baseline levels of cardiovascular risk factors.

### Results

In both cohorts, approximately 227650 (SD 2000) SNPs were heterozygous. The number of heterozygous SNPs was not related to risk of CHD in either men or women (adjusted odds ratios per 2000 heterozygous SNPs 1.01 [95% confidence interval, 0.91-1.13] in women and 0.94 [0.84-1.06] in men). We also found no consistent associations of genome-wide

collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

heterozygosity with levels of lipids, inflammatory markers, adhesion molecules, homocysteine, adiponectin, or body-mass index.

## Conclusions

In these parallel nested case-control studies, we found no relationship of multilocus heterozygosity with risk of CHD or its major risk factors. Studies in other populations are needed to rule out associations with lower levels of heterozygosity.

## Introduction

Although individual humans share some 99.6% of base pairs with others,<sup>[1]</sup> the size of the human genome dictates that there are nonetheless some 24 million base pairs that differ between individuals. This degree of variation offers substantial opportunity for the manifestation of heterozygote advantage, the phenomenon of biological advantage of mixed over pure breeds.

Multilocus heterozygosity may act in two ways. First, heterozygosity at individual loci may provide an advantage relative to homozygote wild-type or variant alleles, sometimes referred to as overdominance. The best known is variation in genes related to erythrocyte function, many of which confer resistance to malaria.<sup>[2]</sup> In this setting, homozygotes are disadvantaged by either susceptibility to malaria (for wild-type alleles) or clinical syndromes as sickle cell anemia, thalassemia, and glucose-6-phosphate dehydrogenase deficiency. There is also evidence that heterozygosity at specific loci may be associated with perceived attractiveness of humans to members of the opposite sex.<sup>[3,4]</sup> Second, heterozygosity as a manifestation of outbreeding reduces risk of homozygosity in disadvantageous alleles, leading in the most adverse situations to recessive diseases.

An important caveat in studies of human heterozygosity is the limited evidence that heterosis confers protection against chronic disease, which is less likely to exert selection pressure than do acute infectious diseases or attractiveness to mates, yet is the dominant cause of mortality in developed nations. Heterozygosity at selected loci has been associated with extreme longevity, but these have not been replicated and it is uncertain that they exemplify generalized heterosis.<sup>[5,6]</sup> In a genome-wide comparison of populations in the Dalmation Islands,<sup>[7]</sup> Campbell and colleagues found that overall heterozygosity across populations was associated with lower blood pressure and total cholesterol, but individual heterozygosity was not measured, raising the strong possibility that this reflects population stratification. Among individuals in the Framingham Heart Study, multilocus heterozygosity at 706 single nucleotide polymorphisms (SNPs) was associated with higher blood pressure, but not with several other cardiovascular traits measured simultaneously.<sup>[8]</sup> To our knowledge, no genome-wide assessment among individuals has evaluated whether heterozygosity confers an advantage on common chronic diseases among adults.

To elucidate whether heterozygote advantage extends to coronary heart disease (CHD), the most common cause of death in the United States,<sup>[9]</sup> we determined the association of variation in genome-wide single nucleotide polymorphisms (SNPs) with both risk of incident CHD and with several CHD risk factors in two well-characterized parallel cohorts of men and women.

## Materials and Methods

### Study Population and Design

The NHS cohort was established in 1976. The study population consists of 121,700 married female registered nurses aged 30 to 55 years residing in one of 11 larger US states. Women have received follow-up questionnaires biennially to update information on exposures and newly diagnosed illnesses. Since 1980, participants have updated information on diet, alcohol, and vitamin supplements through a food frequency questionnaire every four years.

The HPFS began in 1986, when 51,529 male health professionals 40 to 75 years of age completed the initial 6-page HPFS questionnaire. The population includes 29,683 dentists, 3,745 optometrists, 2,218 osteopathic physicians, 4,185 pharmacists, 1,600 podiatrists, and 10,098 veterinarians. Biennial follow-up has mirrored the NHS.

### Case-Control Sampling

We used data from nested case-control studies of CHD in the NHS and HPFS. Blood samples were requested from all active participants and collected from 32,826 NHS members in 1989–1990 and 18,225 HPFS members in 1993–1994. With the exception of a modestly lower prevalence of smoking, those who returned blood samples did not differ substantially from those who did not in both cohorts; mean ages at blood collection were 64 years in men and 60 in women. Participants underwent local phlebotomy and returned samples to our laboratory via overnight courier. Upon arrival, whole blood samples were centrifuged and stored in cryotubes as plasma, buffy coat, and red blood cells in the vapor phase of liquid nitrogen freezers.

From within these subcohorts of women and men who provided blood samples, we conducted parallel nested case-control studies of CHD, defined as non-fatal myocardial infarction or fatal CHD. We wrote to participants who reported incident CHD on the follow-up questionnaires to confirm the report and request permission to review medical records. We also sought medical records for deceased participants, whose deaths were identified by families and postal officials and through the National Death Index. Physicians blinded to the participant's questionnaire reports reviewed all medical records. Cases of myocardial infarction and fatal CHD were confirmed through review of medical records, as previously described.[\[10,11\]](#)

In the NHS, we matched women free of cardiovascular disease or cancer in 1990 who sustained an incident MI or fatal CHD through 2004 to two randomly selected controls on the basis of age and smoking using risk-set sampling.[\[12\]](#) In the HPFS, we matched men free of cardiovascular disease in 1994 who developed incident CHD through 2004 to 878 control men. In this study design, a control for an early case may be included again if the person develops CHD during follow-up; after counting such converters only once, the total number of samples sent for genotyping were 1524 women and 1354 men.

### Genotyping

In 2008, DNA was extracted for genotyping at Merck Research Laboratories, North Wales, PA, using the Affymetrix Genome-Wide Human 6.0 array. Quality control criteria used to define unsuccessful genotyping were a call rate <97%, sex mismatch, and Hardy-Weinberg equilibrium  $p < 10^{-4}$  in controls. Analyses based on principal components were conducted to assess self-reported race. Subsequent analysis sample was restricted to subjects of European ancestry. Self-reported "white" samples with substantial similarity to non-European reference samples (the HapMap Yoruba or Asian samples) were excluded. Three eigenvectors were included as covariates to adjust for potential population stratification in the final sample.

Because women with diabetes were genotyped earlier than those without using slightly different platforms, we excluded the ~28,000 SNPs that were not common to the platforms of both diabetic and non-diabetic women. This left a total of 692,794 SNPs for analysis in women and 724,881 in men.

## Measurement of Risk Factor Levels

All biomarkers were measured on samples stored at  $-130^{\circ}\text{C}$  and have been found to be largely unaffected by transport conditions and reproducible within persons over time.<sup>[13,14]</sup> Measurements included lipids (triglycerides, HDL, calculated LDL), inflammatory markers (C-reactive protein, interleukin-6, fibrinogen), homocysteine, adhesion molecules (E-selectin, intercellular adhesion molecule [ICAM]-1, vascular cell adhesion molecule [VCAM]-1) and total and high molecular weight (HMW) adiponectin. Of these, E-selectin, HMW adiponectin, and homocysteine were only measured in women, and IL-6, fibrinogen, homocysteine, E-selectin, ICAM-1, and VCAM-1 were measured only in the initial subset of case-control pairs among women. We also included body-mass index (BMI) based upon self-reported height and weight, which have been found to be reliable within these cohorts.<sup>[15]</sup>

## Statistical Methods

We examined the associations of multilocus heterozygosity with CHD and its risk factors in multiple ways. Because the number of SNPs and their associated genes were identical across genotyped individuals, we used both the overall number of heterozygous loci and the proportion of loci that were heterozygous; the latter yielded essentially identical results and are not shown here. We conducted three SNP-level analyses, evaluating all autosomal SNPs, a subset of those within the boundaries of all human protein-coding genes and their 70kb upstream and 20kb upstream flanking regions ( $n = 562,289$ ), and only nonsynonymous SNPs ( $n = 4,760$ ). Gene coordinates and nonsynonymous SNPs were retrieved from the Ensembl database version 56.37a.

For incident CHD, we performed unconditional logistic regression with adjustment for matching factors in all analyses; this approach yielded results similar to conditional logistic regression with more consistent model convergence. We report odds ratios per 2000 SNPs for analyses of all SNPs and SNPs within genes and per 25 SNPs for nonsynonymous SNPs; these units reflect ~1 standard deviation in each measure.

Analyses of risk factors were conducted with mixed models, including both cases and controls and adjusting for age, smoking, eigenvectors, and case-control status; this approach provides unbiased estimates when genetic risk factors are not strongly associated with case-control status.<sup>[16]</sup> Information on allelic dosage was extracted for all SNPs using PLINK,<sup>[17]</sup> and statistical analysis were performed in SAS version 9 (SAS Institute Inc., Cary, NC).

## Ethics Statement

The study protocol, including genotyping, was approved by the institutional review boards of the Brigham and Women's Hospital and the Harvard School of Public Health. The completion of the self-administered questionnaire was considered to imply informed consent; participants who provided blood samples additionally provided written informed consent.

## Results

Following cleaning of genotyping data, we included 435 women with incident CHD and 931 control women and 435 men with incident CHD and 878 control men. [Table 1](#) demonstrates

**Table 1. Mean ( $\pm$ SD) numbers of heterozygous loci among CHD cases and controls and odds ratios (95% confidence interval) for CHD associated with heterozygous loci.**

	Women			Men		
	Cases (n = 435)	Controls (n = 931)	p-value	Cases (n = 435)	Controls (n = 878)	p-value
Number of heterozygous loci	227650 $\pm$ 1869	227642 $\pm$ 2240	0.94	227349 $\pm$ 2184	227451 $\pm$ 1860	0.40
Number of heterozygous loci within genes	141601 $\pm$ 1314	141623 $\pm$ 1471	0.78	142581 $\pm$ 1452	142633 $\pm$ 1276	0.53
Number of heterozygous nonsynonymous SNPs	894 $\pm$ 27	895 $\pm$ 27	0.58	914 $\pm$ 27	912 $\pm$ 27	0.23
<b>Odds ratio of CHD per:</b>						
2000 heterozyg. loci	1.01 (0.91–1.13)	Referent		0.94 (0.84–1.06)	Referent	
2000 heterozygous loci within genes	0.99 (0.84–1.17)	Referent		0.94 (0.79–1.12)	Referent	
25 heterozygous nonsynonymous SNPs	0.97 (0.87–1.08)	Referent		1.07 (0.96–1.19)	Referent	

Odds ratios adjusted for age, smoking, and population eigenvectors.

doi:10.1371/journal.pone.0124847.t001

the associations of genome-wide heterozygosity with risk of CHD among women and men. In both sexes, we observed no relationships of CHD with overall number of heterozygous loci, number of heterozygous loci within genes, or number of heterozygous nonsynonymous loci. These results were essentially identical when the proportions of heterozygous SNPs overall, within genes, or at nonsynonymous loci were used, or when we restricted to individuals below 65 years of age at blood draw.

Table 2 shows the relationship of heterozygosity with levels of selected CHD risk factors. There were only two modestly significant relationships in men (LDL among all SNPs and SNPs within genes and fibrinogen among nonsynonymous SNPs;  $p = 0.03$ – $0.04$ ) and none that were consistent in both sexes, regardless of which measure of heterozygosity was used. One possible exception was the association of overall heterozygosity and heterozygosity within genes with E-selectin ( $p = 0.03$ ), which was only measured in women and hence could not be confirmed or refuted in men.

**Table 2. Association of CHD risk factors with overall heterozygous loci, loci within genes, and nonsynonymous SNPs.**

	Women/Men	Women			Men		
		All	Within Genes	Nonsynonymous	All	Within Genes	Nonsynonymous
HDL (mg/dL)	1198/1287	-0.7 $\pm$ 0.4	-0.9 $\pm$ 0.6	-0.7 $\pm$ 0.4	0.3 $\pm$ 0.3	0.5 $\pm$ 0.5	-0.08 $\pm$ 0.3
LDL (mg/dL)	1181/1286	0.3 $\pm$ 1.0	0.9 $\pm$ 1.5	1.0 $\pm$ 1.0	-1.9 $\pm$ 0.9	-2.7 $\pm$ 1.3	-0.6 $\pm$ 0.8
Triglycerides (mg/dL)	1131/1287	1 $\pm$ 2	2 $\pm$ 3	3 $\pm$ 2	1 $\pm$ 3	3 $\pm$ 4	-2 $\pm$ 3
CRP (mg/dL)	1185/1287	-0.02 $\pm$ 0.02	-0.01 $\pm$ 0.02	-0.03 $\pm$ 0.02	-0.01 $\pm$ 0.02	-0.01 $\pm$ 0.02	0.001 $\pm$ 0.01
IL-6 (pg/mL)	616/762	-0.4 $\pm$ 0.4	-0.5 $\pm$ 0.5	0.03 $\pm$ 0.4	0.03 $\pm$ 0.5	0.3 $\pm$ 0.7	0.5 $\pm$ 0.4
Fibrinogen (mg/dL)	211/762	-0.2 $\pm$ 7	3 $\pm$ 10	-2 $\pm$ 5	-2 $\pm$ 3	-3 $\pm$ 4	5 $\pm$ 3
Homocysteine ( $\mu$ mol/L)	644	-0.1 $\pm$ 0.2	-0.07 $\pm$ 0.3	-0.03 $\pm$ 0.2			
E-selectin (ng/mL)	640	1.5 $\pm$ 0.8	2.7 $\pm$ 1.2	0.4 $\pm$ 0.8			
ICAM (ng/mL)	211/762	-0.8 $\pm$ 8	-2 $\pm$ 11	4 $\pm$ 6	-2 $\pm$ 4	-0.1 $\pm$ 5	-4 $\pm$ 3
VCAM (ng/mL)	209/762	5 $\pm$ 11	-0.3 $\pm$ 16	-3 $\pm$ 8	0.4 $\pm$ 11	-4 $\pm$ 17	-4 $\pm$ 10
Total adiponectin (ng/mL)	1209/762	-20 $\pm$ 106	-54 $\pm$ 159	117 $\pm$ 107	-194 $\pm$ 318	-94 $\pm$ 465	-532 $\pm$ 290
HMW adiponectin (ng/mL)	1209	-3 $\pm$ 85	1 $\pm$ 126	85 $\pm$ 85			
BMI (kg/m <sup>2</sup> )	1323/1304	0.1 $\pm$ 0.1	0.2 $\pm$ 0.2	0.1 $\pm$ 0.1	0.2 $\pm$ 0.1	0.2 $\pm$ 0.1	0.02 $\pm$ 0.08

All estimates adjusted for age, smoking, population eigenvectors, and case-control status.

doi:10.1371/journal.pone.0124847.t002



## Discussion

In this analysis of two prospective case-control studies, we found no consistent evidence of heterozygosity advantage in risk for CHD. The number or proportion of heterozygous loci was unrelated to either incident CHD or to circulating CHD risk factors.

The hypothesis of overdominance associated with heterozygosity at individual loci is attractive and has clear support in a few well-described instances. Certainly, the best recognized of these is the selection pressure exerted by *P. falciparum*, which infects erythrocytes. The widespread prevalence of *falciparum* malaria appears to have selected for hemoglobin variants that, while disadvantageous on their own, appear to mitigate the consequences of plasmodial infection.[2] Another proposed manifestation of overdominance may exist in perceived attractiveness, even in non-human primate species.[18] For example, in one study, women rated the faces of men heterozygous at three major histocompatibility complex loci as more attractive than the faces of homozygotes, regardless of the level of genetic similarity between female raters and male subjects.[19]

At the same time, evidence that generalized heterozygosity provides an advantage for risk of chronic disease, and particularly for CHD, is sparse. The Framingham Heart Study reported a positive association of heterozygosity at 706 loci with blood pressure and two related echocardiographic parameters among approximately 1000 participants, but no association with any circulating risk factors or with BMI; no association with risk of CHD *per se* was tested.[8] Because we did not have standardized blood pressure or echocardiographic measurement in our cohorts, we cannot directly replicate their findings, but our results do confirm the lack of association with a large number of putative risk factors and extend these results to clinical CHD endpoints. Why blood pressure might differ from other risk factors in its association with heterosis is necessarily speculative, but it is striking that the observed association in Framingham was positive—that is, greater heterozygosity conferred a phenotype of *higher* blood pressure that might well be protective in the setting of acute primordial stressors like trauma or dehydration, but could be harmful for conditions like CHD that manifest in late adulthood. Thus, we find no clear evidence that multilocus heterozygosity is associated with a lower risk of CHD, the most common cause of mortality in developed nations. Its associations with cancer, dementia, and other chronic diseases remain to be tested, although recent work suggests it may influence schizophrenia.[20]

We found few significant relationships—about the number expected from chance alone—and unsurprisingly these generally did not appear consistent across cohorts. The exception of E-selectin is interesting, as its genetic variation is strongly related to ABO blood group genotype[21] and hence could also be subjected to plasmodial selection pressure.[22] However, there appears to be little ancestral selection pressure in erythrocyte traits among populations of European ancestry and hence even this possible association clearly requires confirmation in other cohorts.

Two differences of our study with more common GWAS designs warrant mention. First, although we performed standard chip-based genotyping as in other GWAS designs, we used a nested case-control design for efficiency within the very large parent cohorts. This has the advantage of preserving power to detect associations with CHD with efficiency and produces unbiased and reasonably precise estimates of relative risk comparable to what would be achieved with a GWAS of the full cohorts of NHS or HPFS (because we genotyped all of the cases that would have been present in the full cohort). While power to examine associations on biomarkers was not as efficiently maintained (because the case-control design explicitly tests a single case definition or endpoint), even those associations had ample precision in our analyses because of their continuous distribution.

Second, we did not conduct an unbiased effort to identify single, specific loci that influence risk of CHD or its risk factors, as occurs in most GWAS. Rather, we tested a specific, global hypothesis about the benefit of widespread variation across the genome. Had our hypothesis been proven correct, the natural next step would be to search for specific loci at which that hypothesis was borne out, more similar to standard GWAS efforts. As a consequence of our single underlying hypothesis, however, our statistical analyses use conventional levels of statistical significance and not the extreme thresholds more familiar in GWAS.

Our study has strengths, but important limitations as well. In both studies, participants who provided blood were middle aged or older, and hence we did not study cases of CHD that occurred early in life, which may be most closely associated with genetic factors, although our findings in those below 65 years of age were also null. We also studied a geographically diverse sample but occupationally homogeneous sample, in whom ascertainment of CHD is reliable but who may not generalize readily to the full US population. We performed genome-wide scans, and hence had an unbiased and comprehensive assessment of variation in SNPs, but we cannot exclude the possibility that heterozygosity at specific loci (e.g., major histocompatibility complex) or in other forms of genetic variation (e.g., copy number variants) favorably influence CHD or its risk factors. Likewise, we studied CHD and its risk factors but cannot rule out associations of heterozygosity with other chronic diseases nor with overall survival.

Power is an important concern in studies of heterozygosity. Number or proportion of heterozygous SNPs are an imperfect estimate of inbreeding,[23] potentially requiring large samples to overcome. Further, we were limited to case-control samples of moderate size with narrow ranges in heterozygosity, illustrated by the small standard deviation in number of heterozygous loci relative to the number genotyped. Nonetheless, our point estimates demonstrated no association whatsoever with CHD risk, and our study compares favorably in size to the Framingham Heart Study, which showed a significant association with blood pressure. Studies in more inbred populations are needed to rule out associations with lower levels of heterozygosity; at the extreme, consanguinity, which maximally suppresses heterozygosity, has been associated with CHD.[24] Equally, studies in populations in other ethnicities are warranted, as populations of African descent in particular are apt to exhibit variability at substantially larger numbers of loci.

In conclusion, in two cohort studies, genome-wide measures of heterozygosity were not associated with risk of incident CHD nor with CHD risk factors. If generalized heterozygosity provides any fitness advantage in humans, it seems unlikely to do so for chronic disease phenotypes like CHD at the levels of outbreeding seen in populations like these.

## Author Contributions

Conceived and designed the experiments: KJM MKJ EBR PK. Performed the experiments: MKJ EBR PK JKP. Analyzed the data: MKJ. Contributed reagents/materials/analysis tools: EBR PK THP. Wrote the paper: KJM. Revised the paper critically: MKJ EBR THP JKP PK.

## References

1. Kidd JM, Cooper GM, Donahue WF, Hayden HS, Samps N, Graves T, et al. (2008) Mapping and sequencing of structural variation from eight human genomes. *Nature* 453: 56–64. doi: [10.1038/nature06862](https://doi.org/10.1038/nature06862) PMID: [18451855](https://pubmed.ncbi.nlm.nih.gov/18451855/)
2. Richer J, Chudley AE (2005) The hemoglobinopathies and malaria. *Clin Genet* 68: 332–336. PMID: [16143020](https://pubmed.ncbi.nlm.nih.gov/16143020/)
3. Lie HC, Rhodes G, Simmons LW (2008) Genetic diversity revealed in human faces. *Evolution* 62: 2473–2486. doi: [10.1111/j.1558-5646.2008.00478.x](https://doi.org/10.1111/j.1558-5646.2008.00478.x) PMID: [18691260](https://pubmed.ncbi.nlm.nih.gov/18691260/)
4. Lewis MB (2010) Why are mixed-race people perceived as more attractive? *Perception* 39: 136–138. PMID: [20301855](https://pubmed.ncbi.nlm.nih.gov/20301855/)



5. DeVry CG, Clarke S (1999) Polymorphic forms of the protein L-isoaspartate (D-aspartate) O-methyltransferase involved in the repair of age-damaged proteins. *J Hum Genet* 44: 275–288. PMID: [10496068](#)
6. Barbieri M, Bonafe M, Rizzo MR, Ragno E, Olivieri F, Marchegiani F, et al. (2004) Gender specific association of genetic variation in peroxisome proliferator-activated receptor (PPAR)gamma-2 with longevity. *Exp Gerontol* 39: 1095–1100. PMID: [15236769](#)
7. Campbell H, Carothers AD, Rudan I, Hayward C, Biloglav Z, Barac L, et al. (2007) Effects of genome-wide heterozygosity on a range of biomedically relevant human quantitative traits. *Hum Mol Genet* 16: 233–241. PMID: [17220173](#)
8. Govindaraju DR, Larson MG, Yin X, Benjamin EJ, Rao MB, Vasan RS (2009) Association between SNP heterozygosity and quantitative traits in the Framingham Heart Study. *Ann Hum Genet* 73: 465–473. doi: [10.1111/j.1469-1809.2009.00514.x](#) PMID: [19523151](#)
9. Heron M, Hoyert DL, Murphy SL, Xu J, Kochanek KD, Tejada-Vera B (2009) Deaths: final data for 2006. *Natl Vital Stat Rep* 57: 1–134. PMID: [19788058](#)
10. Rimm EB, Giovannucci EL, Willett WC, Colditz GA, Ascherio A, Rosner B, et al. (1991) Prospective study of alcohol consumption and risk of coronary disease in men. *Lancet* 338: 464–468. PMID: [1678444](#)
11. Stampfer MJ, Willett WC, Colditz GA, Rosner B, Speizer FE, Hennekens CH (1985) A prospective study of postmenopausal estrogen therapy and coronary heart disease. *N Engl J Med* 313: 1044–1049. PMID: [4047106](#)
12. Prentice RL, Breslow NE (1978) Retrospective studies and failure time models. *Biometrika* 65: 153–158.
13. Pai JK, Curhan GC, Cannuscio CC, Rifai N, Ridker PM, Rimm EB (2002) Stability of novel plasma markers associated with cardiovascular disease: processing within 36 hours of specimen collection. *Clin Chem* 48: 1781–1784. PMID: [12324497](#)
14. Pischon T, Hankinson SE, Hotamisligil GS, Rifai N, Rimm EB (2003) Leisure-time physical activity and reduced plasma levels of obesity-related inflammatory markers. *Obes Res* 11: 1055–1064. PMID: [12972675](#)
15. Rimm EB, Stampfer MJ, Colditz GA, Chute CG, Litin LB, Willett WC (1990) Validity of self-reported waist and hip circumferences in men and women. *Epidemiology* 1: 466–473. PMID: [2090285](#)
16. Monsees GM, Tamimi RM, Kraft P (2009) Genome-wide association scans for secondary traits using case-control samples. *Genet Epidemiol* 33: 717–728. doi: [10.1002/gepi.20424](#) PMID: [19365863](#)
17. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81: 559–575. PMID: [17701901](#)
18. Charpentier MJ, Prugnolle F, Gimenez O, Widdig A (2008) Genetic heterozygosity and sociality in a primate species. *Behav Genet* 38: 151–158. doi: [10.1007/s10519-008-9191-6](#) PMID: [18293079](#)
19. Roberts SC, Little AC, Gosling LM, Perrett DI, Carter V, Jones BC, et al. (2005) MHC-heterozygosity and human facial attractiveness. *Evolution and Human Behavior* 26: 213–226.
20. Keller MC, Simonson MA, Ripke S, Neale BM, Gejman PV, Howrigan DP, et al. (2012) Runs of homozygosity implicate autozygosity as a schizophrenia risk factor. *PLoS genetics* 8: e1002656. doi: [10.1371/journal.pgen.1002656](#) PMID: [22511889](#)
21. Qi L, Cornelis MC, Kraft P, Jensen M, van Dam RM, Sun Q, et al. (2010) Genetic variants in ABO blood group region, plasma soluble E-selectin levels and risk of type 2 diabetes. *Hum Mol Genet* 19: 1856–1862. doi: [10.1093/hmg/ddq057](#) PMID: [20147318](#)
22. Cserti CM, Dzik WH (2007) The ABO blood group system and Plasmodium falciparum malaria. *Blood* 110: 2250–2258. PMID: [17502454](#)
23. Keller MC, Visscher PM, Goddard ME (2011) Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. *Genetics* 189: 237–249. doi: [10.1534/genetics.111.130922](#) PMID: [21705750](#)
24. Shami SA, Qaisar R, Bittles AH (1991) Consanguinity and adult morbidity in Pakistan. *Lancet* 338: 954. PMID: [1681304](#)