BMC Genetics

BioMed Central

Research article

Patterns of population differentiation of candidate genes for cardiovascular disease

Iftikhar J Kullo* and Keyue Ding

Address: Division of Cardiovascular Diseases, Mayo Clinic, Rochester MN, USA

Email: Iftikhar J Kullo* - kullo.iftikhar@mayo.edu; Keyue Ding - ding.keyue@mayo.edu

* Corresponding author

Published: 12 July 2007

BMC Genetics 2007, **8**:48 doi:10.1186/1471-2156-8-48

Received: 16 April 2007
Accepted: 12 July 2007

This article is available from: http://www.biomedcentral.com/1471-2156/8/48

Abstract

Background: The basis for ethnic differences in cardiovascular disease (CVD) susceptibility is not fully understood. We investigated patterns of population differentiation (F_{ST}) of a set of genes in etiologic pathways of CVD among 3 ethnic groups: Yoruba in Nigeria (YRI), Utah residents with European ancestry (CEU), and Han Chinese (CHB) + Japanese (JPT). We identified 37 pathways implicated in CVD based on the PANTHER classification and 416 genes in these pathways were further studied; these genes belonged to 6 biological processes (apoptosis, blood circulation and gas exchange, blood clotting, homeostasis, immune response, and lipoprotein metabolism). Genotype data were obtained from the HapMap database.

Results: We calculated F_{ST} for 15,559 common SNPs (minor allele frequency ≥ 0.10 in at least one population) in genes that co-segregated among the populations, as well as an average-weighted F_{ST} for each gene. SNPs were classified as putatively functional (non-synonymous and untranslated regions) or non-functional (intronic and synonymous sites). Mean F_{ST} values for common putatively functional variants were significantly higher than F_{ST} values for nonfunctional variants. A significant variation in F_{ST} was also seen based on biological processes; the processes of 'apoptosis' and 'lipoprotein metabolism' showed an excess of genes with high F_{ST}. Thus, putative functional SNPs in genes in etiologic pathways for CVD show greater population differentiation than non-functional SNPs and a significant variance of F_{ST} values was noted among pairwise population comparisons for different biological processes.

Conclusion: These results suggest a possible basis for varying susceptibility to CVD among ethnic groups.

Background

The human population is not homogeneous in terms of disease susceptibility and substantial differences in susceptibility to common chronic diseases such as cardiovascular disease (CVD), are present between self-identified ancestral/ethnic groups [1,2]. Significant differences in CVD prevalence were noted in the Seven Countries Study [3]. In the United States, African-Americans have a higher prevalence of hypertension [4] and hypertensive heart disease and significantly greater cardiovascular morbidity and mortality than Whites [5], whereas Japanese-Americans are less prone to CVD than Whites [6]. Differences in cardiovascular 'intermediate' phenotypes also occur among populations; for example, plasma lipid levels differ significantly between African-Americans and non-Hispanic whites, and plasma levels of C-reactive protein vary

substantially between people of different ethnic origins [7]. The basis for ethnic differences in CVD susceptibility is not fully understood but it is likely that in addition to environmental factors, genetic factors contribute either by determining type or severity of risk factors, as well as the susceptibility to environmental/lifestyle risk factors [8-10].

Since different populations are subject to distinct environments, natural selection may produce population-specific allele frequencies. If a functional genetic variant exhibits significantly different pattern of geographic variation compared to a neutral variant, this may be indicative of different selective pressures across populations [11]. For instance, a given genetic variation may be adaptive under a local environmental stressor, which would increase the allele frequencies of this selected locus in a particular population and lead to a greater level of population differentiation [12,13]. A recent study suggested that differential susceptibility to hypertension may be due to differential exposure to selection pressures during the out-of-Africa expansion [14].

Natural selection alters the amount of differentiation between or among populations within a species so that a measure quantifying the differences in allele frequencies among human populations from diverse geographical regions – the F_{ST} statistic – has been used to test for evidence of selection. F_{ST} is a measure of the correlation between alleles in subpopulations relative to the alleles in the total population [15,16]. It is expected that local adaptation will lead to an increase in F_{ST}, when comparing populations under different environmental pressures [17]. Multilocus scans in the human genome, using either single nucleotide polymorphisms (SNPs) [18] or microsatellite markers [19], that compare different populations for several loci, can identify genomic regions carrying a variant that results in a local adaptation [20]. Recently, Ryan *et al*. [21] investigated population differentiation among different functional classes of immunologically important genes and found significantly increased F_{ST} in individual nonsynonymous SNPs of the intercellular adhesion molecule 1 (*ICAM1*) and Toll-like receptors (*TLR*) genes.

Population differentiation has particular relevance for studies of genetic susceptibility to complex diseases since many of the genes that are known to have been affected by natural selection are medically important [22]. Loci with an increased F_{ST} should be considered high priority candidate genes for association studies of complex diseases as well as the study of local adaptation to environmental conditions. An example – provided by Rockman *et al*. [23] – is the increased frequency of high-expression allele (5T) of *MMP3* due to positive selection in Europe but not elsewhere (i.e., a significant differentiation was noted between populations). This variant is associated with reduced arterial stiffness, resulting in lower CVD risk. Understanding genotypic difference among ethnic groups for these genes in relevant biological pathways will provide insights into ethnic differences in complex diseases that may be useful in the prevention and treatment of such diseases [1].

In the present study, using genotype data for three populations from HapMap [24] – Yoruban Africans (YRI), European Whites (CEU), and East Asians (CHB + JPT) – we investigated differences in the distribution of common variants (minor allele frequency ≥ 0.10) of 364 genes in etiologic pathways for CVD, and assessed patterns of population differentiation of these genes in the various biological processes underlying CVD. Our goal was to identify loci with high levels of population differentiation, as a step towards understanding the genetic basis of ethnic differences in cardiovascular risk.

Results

Genotype data for CVD candidate genes

Genotype data for 35,369 SNPs in 405 of 416 genes in etiologic pathways of CVD were available from HapMap; these included 24,391 SNPs in YRI, 22,751 SNPs in CEU, and 20,965 SNPs in CHB+JPT, respectively. Pairwise population comparisons of common variants (MAF ≥ 0.10) showed that a sizeable fraction (79% – 89%) of variants common in one population were present in another population. To avoid ascertainment bias, we focused on the co-segregating SNPs among three populations that had a minor allele frequency (MAF) ≥ 0.10 in at least one population and were in Hardy-Weinberg equilibrium, as suggested by Weir *et al*. [25]. In all, 15,559 such SNPs from 364 genes (87.5%) were identified and classified as putative functional SNPs [5' untranslated regions (5' UTR); coding (nonsynonymous) sites; and 3' untranslated regions (3' UTR)] and non-functional SNPs [coding (synonymous) sites; and intronic sites], based on NCBI SNP database [26] or UCSC genome browsers [27] (Table 1).

Table 1: The number of co-segregating SNPs for various genomic regions in three population samples from HapMap data set

Class	Number
5' flanking regions	1,224
Introns	13,601
Synonymous sites	217
Non-synonymous sites	140
3' flanking regions	377
Total	15,559

Patterns of F_{ST} of CVD candidate genes in three populations

Distribution of pairwise F_{ST} values

We first compared the distribution of F_{ST} values of 15,559 SNPs among the three populations. The mean F_{ST} for YRI vs. CEU was 0.139, for YRI vs. CHB + JPT, 0.158, and for CEU vs. CHB + JPT, 0.095 (Figure 1a). Thus, F_{ST} between Africans and East Asians was slightly higher than the genome-wide average of F_{ST} (0.10 ~0.15, i.e., the background F_{ST}) previously noted to be present between sub-Saharan Africans, Northern Europeans, and East Asians [28-31]. The distribution of single-locus estimates of F_{ST} values between two populations has an approximate χ^2 distribution (Figure 1b). The distribution of F_{ST} between pairwise populations was significantly different (Kolmogorov-Smirnov test, $P < 10^{-16}$). There was a higher proportion of low (< 0.10) pairwise F_{ST} in CEU vs. CHB + JPT (63.5%), compared with YRI vs. CEU (52.5%) or CHB + JPT (49.0%). Larger F_{ST} values (> 0.2) for common SNPs were observed in YRI vs. CEU (25.6%) and YRI vs. CHB + JPT (30.0%), but less often in CEU vs. CHB + JPT (15.4%).

We randomly selected 15,559 SNPs from the data generated by coalescent simulations (one MB region, 1,000 times, see Methods), which matched the characteristics of the observed data in terms of sample size, average F_{ST} (Figure 1c), and MAF (i.e., ≥ 0.10 at least in one population).

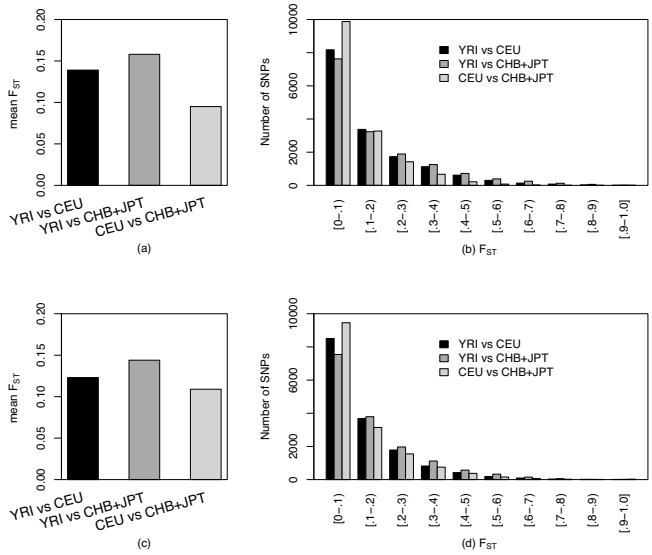

Figure 1

(a) Mean F_{ST} in pairwise population comparisons in the observed data. (b) Distribution of F_{ST} values in HapMap data set. (c) Mean F_{ST} in pairwise population comparisons in the simulated data. (d) Distribution of F_{ST} values in the simulated data set. YRI, Yoruba in Ibadan, Nigeria; CEU, Utah residents with ancestry from northern and western Europe, CHB, Han Chinese in Beijing, China, and JPT, Japanese in Tokyo, Japan

The simulated distribution of pairwise F_{ST} (Figure 1d) was significantly different compared with the observed HapMap data (Kolmogorov-Smirnov test, $P < 10^{-10}$). An excess of high-F_{ST} values were present in the HapMap data for the set of genes in the present study, consistent with action of either genetic drift or natural selection and local adaptation leading to an increase in allele frequencies for the selected locus in a particular population [18]. It should be noted that the simulated F_{ST} underestimated the F_{ST} from the empirical data, consistent with the previous findings [32] and indicating that simulation did not perfectly predict F_{ST}.

Distribution of pairwise F_{ST} based on SNP functional classification

To assess differences in the distribution of combined F_{ST} values according to different categories of SNPs, we plotted the correlation between mean pairwise F_{ST} and MAF according to the five different SNPs categories (Figure 2). The mean F_{ST} values for SNPs of different categories conditioned on MAF are listed in Table S2 (see additional file 1). Using analysis of variance (ANOVA), we found that the pairwise mean F_{ST} values in CEU vs. CHB + JPT varied significantly among different SNP categories ($P = 0.019$) by analysis of variance, but not in YRI vs. CEU ($P = 0.273$) and YRI vs. CHB + JPT ($P = 0.124$) (see Methods). In addition, pairwise mean F_{ST} values between any two populations differed with MAF ($P < 2.2 \times 10^{-16}$), and there was a significant interaction of logarithm transformed MAF × category ($P < 3.3 \times 10^{-5}$), indicating that the effect of SNP category was modified by MAF.

Common, putative functional SNPs (i.e., SNPs in potentially functional genomic elements such as non-synonymous sites, 5' and 3' UTR) had systematically higher mean F_{ST} values than SNPs in nonfunctional genomic elements (i.e., intronic and synonymous sites), although this was limited to SNPs with MAF ≥ 0.30. For example, when comparing YRI vs. CHB + JPT, mean F_{ST} for common SNPs with MAF of 0.35–0.40 in non-synonymous sites (F_{ST} = 0.303) was higher than that in synonymous (F_{ST} = 0.124) and intronic sites (F_{ST} = 0.201) (Figure 2b), although the effect was not statistically significant (pairwise comparison by 'multcomp' library in *R*) due to the limited number of non-synonymous sites. Much higher mean F_{ST} values for non-synonymous SNPs with MAF of 0.45–0.50 were noted in YRI vs. CEU (F_{ST} = 0.505).

Patterns of F_{ST} in biological processes and functional pathways

We compared the distribution of F_{ST} values for 364 genes in various biological processes (= 6) or functional pathways (= 37). The boxplots of F_{ST} values in different biological processes are shown in Figure 3. Significant variation in F_{ST} values was seen among different biological processes ($P \leq 3.1 \times 10^{-15}$, Kruskal-Wallis test).

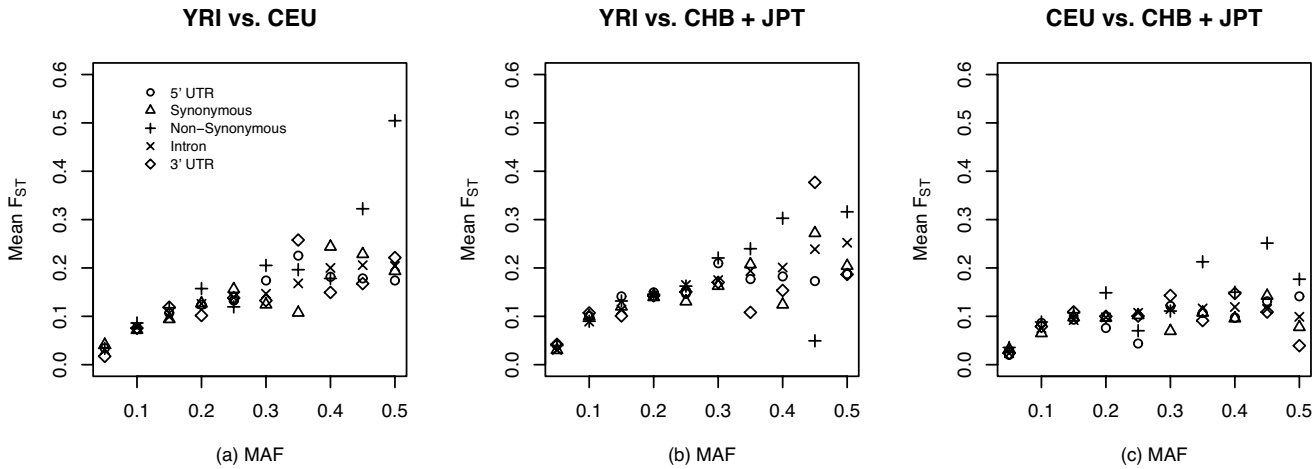

Figure 2
Mean F_{ST} values for SNPs of different categories, conditioned on MAF. YRI, Yoruba in Ibadan, Nigeria; CEU, Utah residents with ancestry from northern and western Europe, CHB, Han Chinese in Beijing, China, and JPT, Japanese in Tokyo, Japan

In general, mean F_{ST} for each biological process was significantly higher between Africans and non-Africans (especially between Africans and East Asians), in comparison with that between non-Africans ($P < 0.05$, pairwise t test) (Figure 3). The patterns of F_{ST} suggested differential local factors operating on the selected biological processes among populations. For example, for genes in the 'blood circulation & gas exchange' pathway, the mean F_{ST} value was 0.129 in YRI vs. CEU and 0.158 in YRI vs. CHB + JPT, but significantly lower (F_{ST} = 0.062) in CEU vs. CHB + JPT. A similar pattern was also noted in the 'lipoprotein metabolism' genes (YRI vs. CEU: 0.144; YRI vs. CHB + JPT: 0.165; and CEU vs. CHB + JPT: 0.082).

Significant variation in F_{ST} was also noted among the 37 functional pathways, (data not shown; $P < 2.2 \times 10^{-16}$, Kruskal-Wallis test). As expected, the mean F_{ST} for each functional pathway was significantly higher between Africans and non-Africans, than between non-Africans. Most strikingly, genes in the 'Insulin/IGF-mitogen activated protein kinase kinase/MAP kinase cascade' pathway showed a relatively high F_{ST} in all pairwise population comparisons (YRI vs. CEU: 0.194, YRI vs. CHB + JPT: 0.174 and CEU vs. CHB + JPT: 0.138). Also, a relatively high F_{ST} between Africans and non-Africans was noted in the 'interleukin signaling pathway' genes (YRI vs. CEU: 0.192 and YRI vs. CHB + JPT: 0.210).

Signatures of local adaptation

The HapMap data provides a genome-wide empirical distribution of F_{ST} against which significance of F_{ST} values can be evaluated, rather than based on theoretical computer simulations [33]. SNPs distant from genes are good candidates for neutral mutations since genes and their regulatory elements are more likely to be under selection than non-coding DNA [34]. We acquired the empirical 'neutral' distribution of F_{ST} values from 289 intergenic regions across the autosomal genome (14,792 SNPs) and 17 intergenic regions across the X chromosome (372 SNPs) without considering the effect of MAF. For autosomal chromosomes, the 95% upper limits of F_{ST} values were: YRI vs. CEU (= 0.602); YRI vs. CHB + JPT (= 0.640); and CEU vs. CHB + JPT (= 0.466) (Figure 4). The 95% upper limits of F_{ST} values were higher for chromosome X – 0.729 (YRI vs. CEU), 0.828 (YRI vs. CHB + JPT), and 0.707 (CEU vs. CHB + JPT), respectively.

We first calculated the significance level of F_{ST} for each SNP locus. A small fraction of SNP loci showed a significantly higher F_{ST} ($P < 0.05$) based on the empirical 'neutral' distribution of F_{ST} values – 238 SNPs (1.53%) in YRI vs. CEU, 325 (2.09%) in YRI vs. CHB + JPT, and 164 (1.05%) in CEU vs. CHB + JPT, respectively. The number of genes with at least one unusually high F_{ST} value according to biological process is shown in Table 2. The biological processes of 'apoptosis' and 'lipoprotein metabolism' showed an excess of genes (31.3% and 34.0%) with a significantly higher F_{ST} (Table 2, Figure 5).

We also calculated a weighted-average F_{ST}, combining information over loci [35] that summarizes the levels of interpopulation differentiation in each gene. Genes with a significantly higher weighted-average F_{ST} are shown in Table 3. In total, there were signatures of local adaptation in nine genes (2.5%) – four genes in YRI vs. CEU and three genes in YRI vs. CHB + JPT, and three genes in CEU vs. CHB + JPT. Most of the genes are involved in 'immune response' (*GRB2*, *IKBKB*, *IL4*, *IL6*) and 'apoptosis'

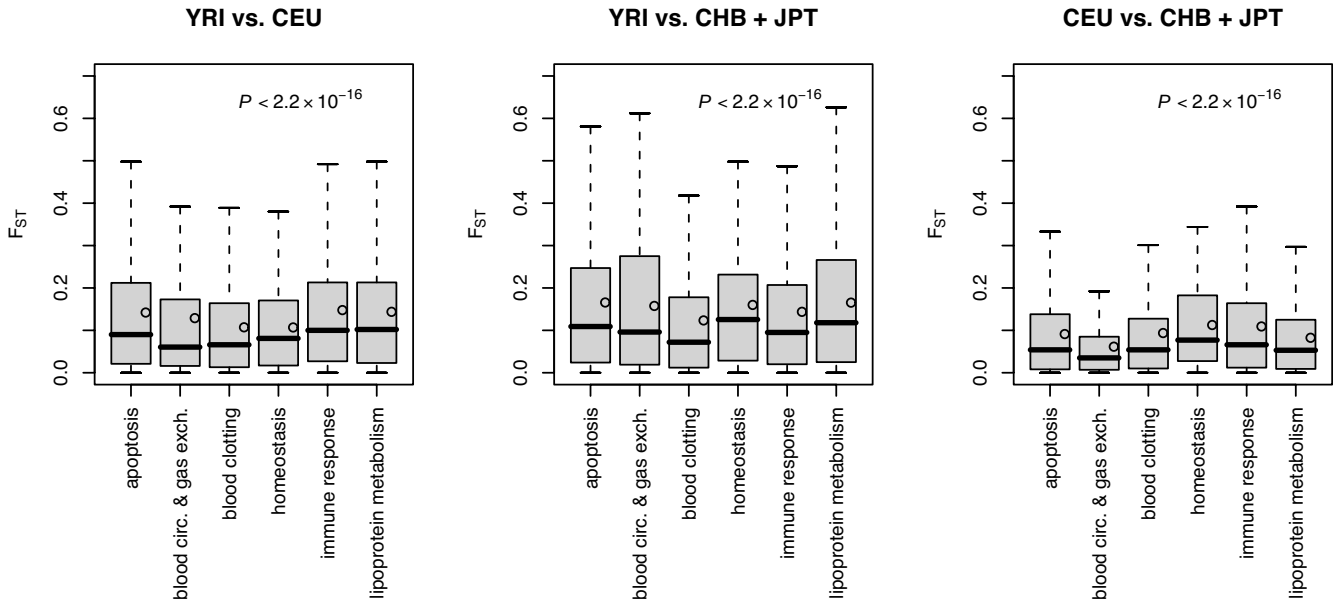

Figure 3
Boxplots of F_{ST} according to biological processes. Points in the box are the mean F_{ST} values. Outliers are not shown. *P* values were calculated by Kruskal-Wallis test. YRI, Yoruba in Ibadan, Nigeria; CEU, Utah residents with ancestry from northern and western Europe, CHB, Han Chinese in Beijing, China, and JPT, Japanese in Tokyo, Japan

(*ARHGEF1*, *RIPK1*, *BCL2L1*, *IL4*, *IL6*), as well as one gene each in 'blood clotting' (*F2*) and 'lipoprotein metabolism' (*PMVK*). The distribution of F_{ST} along the sequence for these genes is shown in Figure 6, indicating multiple SNP loci with a significantly high F_{ST}.

Discussion

The most direct way to study whether genetic risk factors vary among ethnic groups is to determine whether disease susceptibility variants differ in frequency and/or effect among groups [36]. Several studies have demonstrated that the genotype frequencies of SNPs in candidate genes for cardiovascular diseases (CVD) differ among populations [2,9,37]. Lanfear *et al*. [9] found higher frequencies of disease-associated genotypes in African-Americans than in European-Americans for polymorphisms in *GJA4*, *SERPINE1* (*PAI-1*) and *MMP3*. Two nonsense mutations in *PCSK9* that lead to lower plasma levels of low-density lipoprotein cholesterol are relatively common in African-Americans (2%) but rare in European-Americans (< 0.1%) [37]. Significant differences in allele frequencies were noted in the polymorphisms of *IL2*, *IL6*, and *IL10* among Blacks, Whites, and Asians [38]. In a meta-analysis, Ioannidis *et al*. [2] assessed 43 validated gene-disease associations across 697 study populations of various ethnicities and found that frequencies of polymorphisms in seven cardiovascular disease genes – *APOE*, *ACE*, *ITGB3*, *MTHFR*, *F2*, *PON1*, and *PON2* – varied significantly between ethnicities ($I^2 \geq 75\%$; I^2 being a measure of the extent to which the heterogeneity is not due to chance) [2]. These loci (except *APOE*) showed a large heterogeneity of 'race'-specific frequency of polymorphisms. In addition, a disease-associated mutation may be present at high frequency in one population but virtually absent in another, an example being a variant in the *SCN5A* gene (associated with cardiac arrhythmia) [39], which is present in African-Americans at an allele frequency of 0.132 and is not found in Europeans and Asians. However, the F_{ST} value may not be a 'sensitive' test and genes implicated in CVD susceptibility may not lead to exceptionally elevated F_{ST} values. For example, the F_{ST} value was not significantly higher for the nonsense mutation in *PCSK9* when comparing African-Americans (allele frequency: ~2%) and European-Americans (frequency < 0.1%).

Gene variants that interact with geographic- or population-specific environmental factors may be under strong positive or diversifying selection pressure [40]. A measure of population differentiation (i.e., F_{ST}) has been used to quantify the degree to which populations are subdivided for particular genetic variants. While other measures, such as the nearest-neighbor statistic (S_{nn}) [41], and *c* parameter (i.e., measuring how isolated a population has been) [42] are used to identify the level of population differentiation, the unbiased estimator of F_{ST} is simple and easy to calculate. We calculated F_{ST} for SNPs in a set of genes in causal pathways for CVD, to identify the patterns of differ-

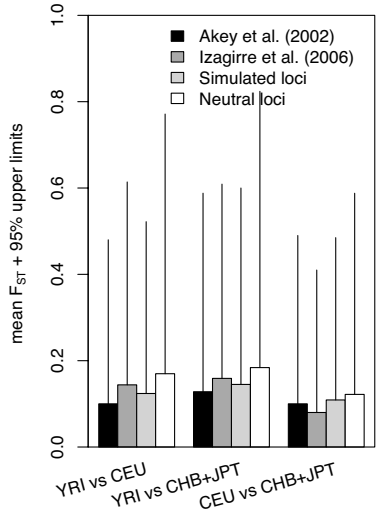

Figure 4
Mean and 95% upper limits of F_{ST} distributions. Black, dark grey, and light grey bars represent the mean of F_{ST} distribution found for the autosomal SNPs analyzed in Akey *et al.* [18], Izagirre *et al.* [47], and coalescent simulations, respectively. White bars represent the mean of F_{ST} in the 'neutral' autosomal loci (14,792 SNPs) from the intergenic regions in this study. The 95% upper limits are placed on top of the mean value of F_{ST}.

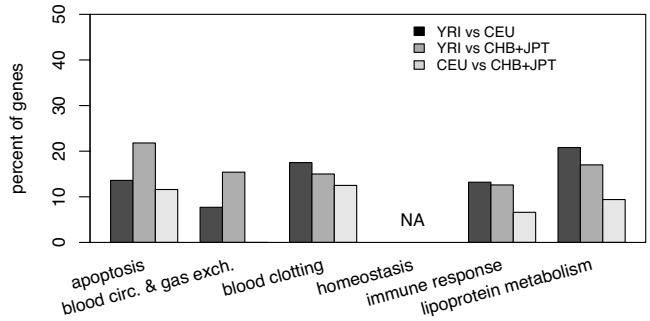

Figure 5
Percentage of genes in different biological processes with significantly high F_{ST} (empirical $P \leq 0.05$) in at least one SNP.

ences for allele frequencies from one ethnic group to another.

To avoid false positive results due to genotyping error and ascertainment bias [18,25,43], we studied SNPs that were in Hardy-Weinberg equilibrium, had a minor allele frequency (MAF) of ≥ 0.10 in at least one population (i.e., common SNPs), and cosegregated in all three populations [25]. The relatively low genotype error rate (0.3%) in HapMap data [33] has likely had a limited impact on estimates of F_{ST} [18]. Consistent with previous reports [44], most common SNPs were shared between populations in the present study and 10–20% of the common variants present in one population are not necessarily common in the other population (i.e., private SNPs), especially when comparing Africans and non-Africans. However, whether the 'private' or shared common SNPs contribute to ethnic differences of CVD risk needs further investigation. The approximate chi-square distribution of F_{ST} (Figure 1b) is similar to the distribution for the entire HapMap data [25]. The distribution of pairwise F_{ST} is similar for the YRI vs. CEU and YRI vs. CHB + JPT, consistent with previous reports [28,33,44]. The lower level of population differentiation in CEU vs. CHB + JPT supports a recent split between these two populations.

Given that candidate loci with large F_{ST} values might have undergone local adaptation [18,45], we hypothesized that F_{ST} values would be higher in putatively functional variants than in putatively nonfunctional variants. Common functional SNPs with large F_{ST} values may influence variation in disease susceptibility among different populations. A high divergence of allele frequency was noted among putatively functional (e.g., nonsynonymous sites and SNPs in 5', and 3' UTRs) and non-functional SNPs (Figure 2). In addition, the pattern varied in different pairwise population comparisons and was most obvious for the Africans vs. non-Africans comparison. An extreme

Table 2: Number of genes in different biological processes with significantly higher F_{ST} (empirical $P \leq 0.05$) in at least one SNP

Biological process (gene number)	Pairwise population comparisons			**Total**
	YRI vs. CEU	YRI vs. CHB+JPT	CEU vs. CHB+JPT	
Apoptosis (147)	20 (13.6%)	32 (21.8%)	17 (11.6%)	**46 (31.3%)**
Blood circulation and gas exchange (13)	1 (7.7%)	2 (15.4%)	0 (0%)	**2 (15.4%)**
Blood clotting (40)	7 (17.5%)	6 (15.0%)	5 (12.5%)	**9 (22.5%)**
Homeostasis (9)	-	-	-	**-**
Immune response (151)	20 (13.2%)	19 (12.6%)	10 (6.6%)	**35 (23.2%)**
Lipoprotein metabolism (53)	11 (20.8%)	9 (17.0%)	5 (9.4%)	**18 (34.0%)**
Total (364)	**51 (14.0%)**	**63 (17.3%)**	**32 (8.8%)**	**110 (30.2%)**

example is a non-synonymous SNP in *F2* (rs5896, Met165 → Thr) with higher F_{ST} in YRI vs. CHB + JPT (F_{ST} = 1.000) and CEU vs. CHB + JPT (F_{ST} = 0.950), at a MAF of 0.00, 0.05, and 1.00 in YRI, CEU and CHB + JPT, respectively. This finding has also been reported in a prior analysis of the Phase I HapMap data [33].

To provide a reasonable biological explanation for the F_{ST} values, we considered the sampling distribution of the F_{ST} estimates. One of the methods is based on numerical sampling or permutation procedures so that F_{ST} is estimated and the proportion of values larger than or equal to the one estimated from the observed data set will yield the unbiased *P*-value of the test [46]. Yet another method involves the use of variances in actual values of F_{ST} to detect regions of exceptional F_{ST} values, defined as population-average values more than three standard deviations from the chromosomal average [25]. An alternative strategy is based on coalescent theory that simulates the histories of the samples or the populations. An expected distribution of F_{ST} for 15,559 simulated SNPs was generated under the calibrated demographic model [32] (Figure 1c~d). Recently, an empirical distribution has been used to test the significance level of F_{ST} [18,47,48]. Instead of whole-genome empirical distribution of F_{ST} [18,34], we used an empirical distribution of F_{ST} from 'neutral' loci in 289 and 17 intergenic regions from autosomal chromosomes and X chromosome, respectively. The number of SNPs for empirical neutral distribution (14,792 from autosomal chromosomes and 372 from the X chromosome) takes into account the multiple testing incurred in our evaluation of 15,559 SNPs. This 'neutral' empirical distribution of F_{ST} is most likely shaped by only demography and therefore the *P* values of F_{ST} estimated from the empirical distribution may represent a more reliable indicator of selection. The mean and 95% upper limits of this 'neutral' distribution are slightly higher than the previously used 'neutral' empirical [18,47] or the simulated distribution (Figure 4), indicating the statistical test using the empirical distribution is more conservative. Although none of the SNPs or the genes remained significant after correction for multiple testing using false discovery rate [49], the method of population differentiation can used as an exploratory tool for detecting local adaptation [47].

Genes are subjected to different evolutionary constraints depending on their biological functions and genes with a higher population differentiation are likely to have been more readily influenced by the environment [40]. For instance, Grossman *et al*. [50] found that F_{ST} values for 'apoptosis' genes among Ashkenazi, Sephardic and Arab Israelis are low. In the present study, most striking was the high mean F_{ST} in the biological process of 'apoptosis' (F_{ST} = 0.166) and 'lipoprotein metabolism' (F_{ST} = 0.165) between YRI vs. CHB + JPT, but lower F_{ST} in CEU vs. CHB + JPT (0.091 and 0.082, respectively) (Figure 3). The patterns suggested that differential environments pressures may have accounted for the varying F_{ST} among biological processes for different populations.

The biological processes of 'apoptosis' and 'lipoprotein metabolism' showed an excess of high F_{ST} values, which may be the result of local adaptation or genetic drift (Table 2, Figure 5). The combined information of F_{ST} over loci provides a means to quantify the level of population differentiation in a given gene (Figure 6) [35]. A summary of functions for these genes relevant to CVD is shown in Table S3 (see additional file 1). For three gene loci – *F2*, *IL4*, and *IL6* – significantly higher F_{ST} values in CEU vs. CHB + JPT were noted (Table 3). Previous studies also demonstrated a higher population differentiation in *IL4* between Europeans and East Asians [51]. These findings suggest that different selective factors might be exerting locus-specific effects in populations with different geographic origin. Most of the genes under local adaptation were in the 'immune response' and 'apoptosis' pathways (Table 3). How these biological processes may promote differential susceptibility to CVD among Africans and non-Africans need further study.

Given that CVD susceptibility varies among populations, genes that are responsible for such variations should also differ among populations. Hence, regardless of whether drift or selection is responsible, the approach of looking among a set of candidate genes for those with highest F_{ST} values should help identify candidate genes to explain differences in CVD susceptibility. The present study identifies genes in etiological pathways of CVD with a high F_{ST} among populations, and should be considered as a means of generating new hypotheses to test (i.e., reprioritize candidate genes rather than identify new ones). Putative functional SNPs with a high F_{ST} should be investigated further for confirmation of functional effects and should be included in CVD association studies among populations. In addition, the differential patterns of F_{ST} among biological processes and functional pathways may provide insight into the mechanisms contributing to varying CVD susceptibility among different populations.

Our study has several limitations. First, since population differentiation detects local adaptation in geographically separate populations within the last ~75,000 years [52], the present study cannot identify genes subject to natural selection before this time scale. Phylogenetic analyses [52] detect evolutionary changes preceding this time period whereas nucleotide diversity and LD-based tests [22,53] may be helpful in further investigating genomic regions with significantly high F_{ST} values. Second, the fluctuation of allele frequencies due to a relatively small sample size could affect the robustness of our inferences. Thus, esti-

Table 3: Genes with a significantly high weighted-average F_{ST} ($P \leq 0.05$)

Symbol	Gene name	YRI vs. CEU		YRI vs. CHB + JPT		CEU vs. CHB + JPT		Biological processes
		F_{ST}	*P*	F_{ST}	*P*	F_{ST}	*P*	
GRB2	growth factor receptor-bound protein 2	0.720	0.035*	0.828	0.023*	0.048	0.556	Immune response
IKBKB	inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase beta	0.658	0.043*	0.555	0.071	0.034	0.605	Immune response
IL4	interleukin 4	0.262	0.220	0.205	0.329	0.512	0.043*	Immune response/Apoptosis
IL6	interleukin 6	0.108	0.478	0.431	0.119	0.570	0.037*	Immune response/Apoptosis
ARHGEF1	Rho guanine nucleotide exchange factor 1	0.769	0.030*	0.532	0.078	0.149	0.263	Apoptosis
RIPK1	receptor (*TNFRSF*)-interacting serine-threonine kinase 1	0.642	0.045*	0.544	0.074	0.070	0.456	Apoptosis
BCL2L1	BCL2-like 1	0.278	0.204	0.679	0.043*	0.265	0.127	Apoptosis
PMVK	phosphomevalonate kinase	0.563	0.059*	0.652	0.048*	0.097	0.380	Lipoprotein metabolism
F2	coagulation factor II	0.413	0.109	0.536	0.077	0.473	0.049*	Blood clotting

* Genes with a significantly higher weighted-average F_{ST}

mation of allele frequencies in a larger sample would be needed to confirm our results. Even though common SNPs (MAF > 0.1) that conformed to Hardy-Weinberg equilibrium (HWE) were included, we cannot completely address the issues of ascertainment bias of SNPs [43]. However, SNPs with MAF < 0.1 and not in HWE could have biological relevance since natural selection, not genotyping error also could lead to deviation from HWE. Third, we did not adjust for recombination rate in our analyses. We found a negative correlation between the weighted-average F_{ST} and recombination rate (based on the recombination map of Kong et al. [54]), although the correlation was not statistically significant (analyses not shown). The recombination rate in genes with a high F_{ST} listed in Figure 6 (except GRB2 and PRIK1) was below the average recombination rate (1cM/MB) across the human genome. Finally, a complete catalogue of etiologic pathways implicated in CVD is yet to be established.

Conclusion

In summary, the present study of genes in etiologic pathways for CVD revealed greater population differentiation in putative functional SNPs in these genes, as well as significant variation in F_{ST} based on different biological processes relevant to CVD. The biological processes of 'apoptosis' and 'lipoprotein metabolism' showed an excess of genes with high F_{ST}. In addition, the pattern varied in different pairwise population comparisons. SNP loci (especially putatively functional SNPs) and genes with a significantly higher population differentiation should be considered high priority for investigating genetic factors influencing differences in CVD risk among populations.

Methods

Genes in the etiologic pathways for cardiovascular disease (CVD)

Based on a search of the literature in PUBMED [55], we identified 37 functional pathways implicated in CVD (a summary of the functional pathways, the number of genes in each pathway and corresponding references is presented in Table S1, see additional file 1). We explored 416 genes from these pathways using the Panther classification system [56,57]. These genes were classified into the following biological processes relevant to CVD: 1) apoptosis; 2) blood circulation & gas exchange; 3) blood clotting; 4) homeostasis; 5) immunity and defense; and 6) lipid fatty acid & steroid metabolism. Examples of candidate genes in various pathways were: caspase and TNF/TNF receptor gene family in the 'apoptosis' process; endothelin and nitric oxide synthase 3 genes in the 'blood circulation & gas exchange' process; those genes involved in the intrinsic and extrinsic coagulation pathway in the 'blood clotting' process; the gene family of insulin receptor substrate in the 'homeostasis' process; those genes participating in the inflammation response, such as the interleukin gene family, in the 'immunity and defense' process; and the arachidonate-lipoxygenase, and phospholipase gene family in the 'lipid fatty acid and steroid metabolism' process.

Genotype data

Using the National Center for Biotechnology Information (NCBI) reference sequence [58], we aligned the sequence of messenger RNA of each gene with the human chromosome sequence (NCBI build 35). Based on the alignment, the genotype data for single nucleotide polymorphisms (SNPs) in each gene were obtained from HapMap database (Phase II) [24,33]. The HapMap data includes 90 individuals (30 trios) from the Yoruba in Ibadan, Nigeria (YRI), 90 individuals (30 trios) in Utah residents with

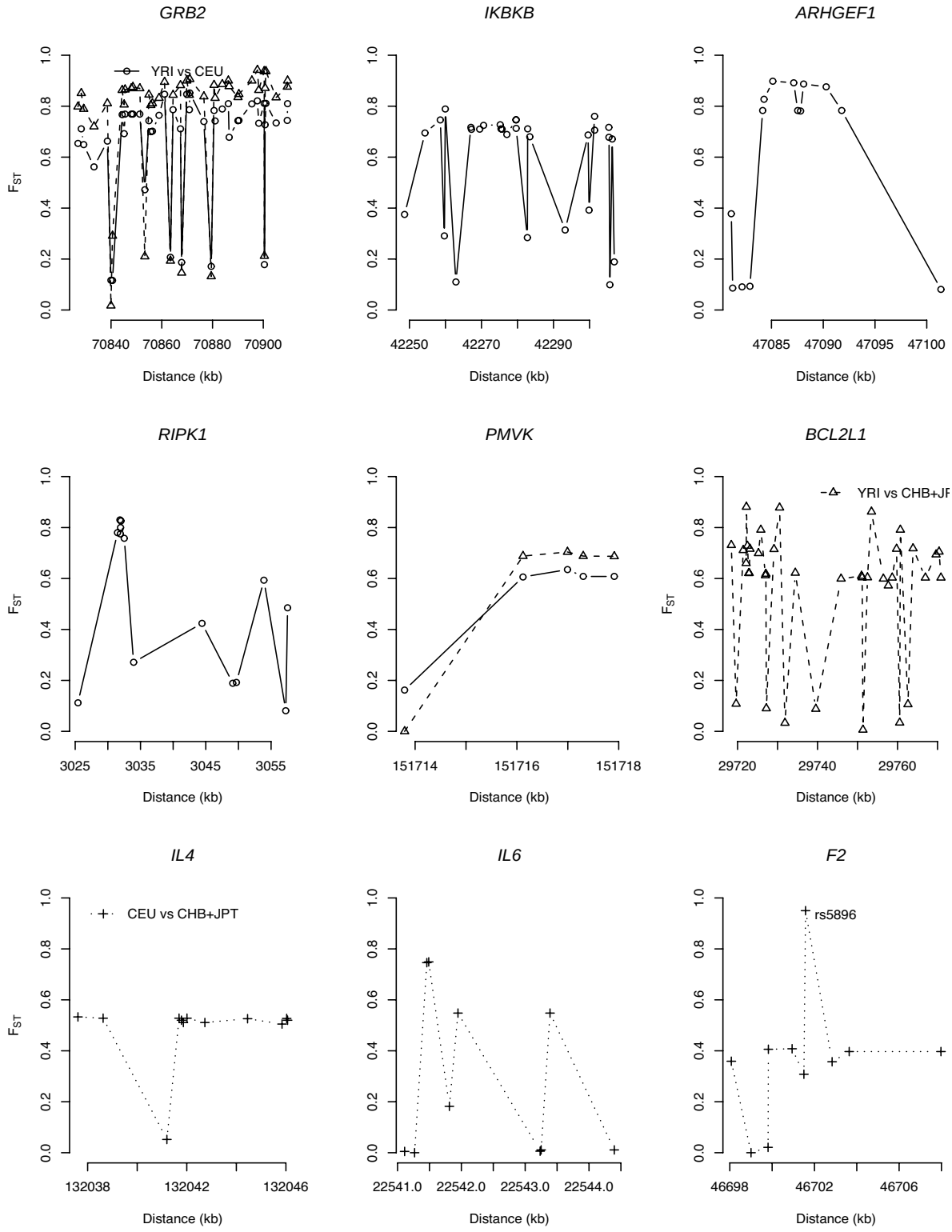

Figure 6

F_{ST} profile for nine genes with a significantly higher weighted-average F_{ST}. The X-axis indicates the chromosomal position (kb). See Table 3 for gene names. The average recombination rate (cM/MB) for the genes is: *GRB2*, 1.78; *IKBKB*, 0.76; *ARHGEF1*, 0.81; *RIPK1*, 2.06; *PMVK*, 1.07; *BCL2L1*, 0.86; *IL4*, 0.94; IL6, 1.10; *F2*, 0.66.

ancestry from Northern and Western Europe (CEU), 45 unrelated Han Chinese in Beijing, China (CHB) and 45 unrelated Japanese in Tokyo, Japan (JPT). For each gene, map information for SNPs was obtained from NCBI and the genome annotation database at University of California, Santa Cruz (UCSC) [27]. The reference mRNA sequence was annotated as 5' untranslated regions (5' UTR), coding (synonymous and non-synonymous), intronic, and 3' untranslated regions (3' UTR).

Calculation of F_{ST}

The estimate of population differentiation, F_{ST}, measures relatedness of pairs of alleles within a population relative to the total populations [16,35]. We calculated an unbiased small-sample estimator of F_{ST} as described by Weir [16,59]. If there are n_i alleles sampled from the i^{th} of r populations, the sample frequency of the SNP allele u in the i^{th} subpopulation is $\tilde{p}_{iu}$, and a weighted average of p_u across population is $\bar{p}_u = \frac{1}{\sum_i n_i}\sum_{i=1}^{r} n_i \tilde{p}_{iu}$. Two mean squares were defined as,

$$MSG_u = \frac{1}{\sum_{i=1}^{r}(n_i - 1)}\sum_{i=1}^{r} n_i \tilde{p}_{iu}(1-\tilde{p}_{iu}),$$ and

$$MSP_u = \frac{1}{r-1}\sum_{i=1}^{r} n_i (\tilde{p}_{iu} - \bar{p}_u)^2,$$ where MSG_u and MSP_u denote the observed mean square errors for loci with populations and between populations, respectively. The moment estimator of F_{ST} was defined as, $F_{ST} = \frac{MSP_u - MSG_u}{MSP_u + (n_c - 1)MSG_u}$, where, n_c is the average sample size across samples that also incorporates and corrects for the variance in sample size over subpopulations, $n_c = \frac{1}{r-1}\left(\sum_{i=1}^{r} n_i - \frac{\sum_{i=1}^{r} n_i^2}{\sum_{i=1}^{r} n_i}\right)$.

F_{ST} was estimated for each SNP locus and a weighted-average F_{ST} was estimated for each gene [35,60]. F_{ST} can be negative when levels of differentiation are close to zero and/or sample sizes are small, indicating no population differentiation at these loci [35]. In our analysis, we assigned a value of zero to negative F_{ST} values. The program for calculating F_{ST} for each SNP locus was written in Perl and is available from the authors upon request. The weighted-average F_{ST} value combining information over loci [35] was calculated using the 'Genepop' software (version 3.4) [61].

Expected distribution of F_{ST} *under a calibrated demographic model*

We used coalescent theory to obtain the expected distribution of F_{ST} under a calibrated demographic model for Africans, Europeans, and Asians [32]. Using the program 'cosi' [62], we simulated a one megabase (MB) region 1,000 times under the 'best-fitting' population parameters for the three populations. The 'best-fitting' set of parameters yielded good agreement with all aspects (including allele frequency spectrum, fraction of alleles that are ancestral, linkage disequilibrium, and F_{ST}) of the observed data in the human genome [32]. Pairwise F_{ST} among populations was calculated for each simulated SNP.

Significance of F_{ST}

Population demographic history, such as migration among sub-populations, can also influence F_{ST} [63]. By comparing the F_{ST} of an individual locus to the empirical distribution, it is possible to distinguish between genetic drift and natural selection without having to take population demographic history into account [64]. To assess the statistical significance of F_{ST} values for SNPs, we selected 289 intergenic regions across the autosomal genome and 17 intergenic regions across chromosome X to obtain a neutral distribution of F_{ST}, based on the annotation tables of human chromosomes from UCSC database. These regions were separated by at least one MB from the closest exon and did not include centromeric regions. Each region spanned an average of 1.52 MB and in total composed 466.10 MB. Following the method of Izagirre *et al*. [47], we defined 'neutral' SNPs as follows: 1) separated by at least 25 kb from each other, 2) genotyped in all three populations, and 3) MAF ≥ 0.10 in at least one of the three populations. In all, 14,792 SNPs for autosomal chromosomes and 372 SNPs for the X chromosome satisfied these criteria. We used the observed 'neutral' distribution to assess the significance level of the F_{ST} for each SNP (P value, one-sided) in the autosomal and X chromosome regions separately. We focused on the significantly higher values of F_{ST} for local adaptation, although significantly lower F_{ST} might result from balancing selection.

All statistical analyses were performed using *R*. Analysis of variance (ANOVA) was performed to compare linear regressions of F_{ST} against logarithm of MAF [ln(MAF)] with and without the terms for the SNP category and a ln(MAF) × category interaction.

Authors' contributions

IJK and KD designed the project, analyzed the data, and wrote the paper. Both authors read and approved the final manuscript.

Additional material

Additional file 1

Patterns of population differentiation of candidate genes for cardiovascular disease. Table S1 describes the summary of functional pathways implicated in atherosclerosis; Table S2 shows mean F_{ST} values for SNPs of different categories, conditioned on MAF; and Table S3 is the summary of function for genes with a significant weighted-average F_{ST}.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2156-8-48-S1.doc]

Acknowledgements

We acknowledge the Research Computing Facility of Mayo Clinic Rochester and the Supercomputing Institute of University of Minnesota, Minneapolis, for technical support.

References

1. Risch N, Burchard E, Ziv E, Tang H: **Categorization of humans in biomedical research: genes, race and disease.** *Genome Biol* 2002, **3(7):**comment2007.
2. Ioannidis JP, Ntzani EE, Trikalinos TA: **'Racial' differences in genetic effects for complex diseases.** *Nat Genet* 2004, **36(12):**1312-1318.
3. Keys A: **Coronary heart disease in seven countries.** *Circulation* 1970, **41(4 Suppl):**1-211.
4. Gupta V, Nanda NC, Yesilbursa D, Huang WY, Gupta V, Li Q, Gomez CR: **Racial differences in thoracic aorta atherosclerosis among ischemic stroke patients.** *Stroke* 2003, **34(2):**408-412.
5. Davey Smith G, Neaton JD, Wentworth D, Stamler R, Stamler J: **Mortality differences between black and white men in the USA: contribution of income and other risk factors among men screened for the MRFIT. MRFIT Research Group. Multiple Risk Factor Intervention Trial.** *Lancet* 1998, **351(9107):**934-939.
6. Yano K, Reed DM, McGee DL: **Ten-year incidence of coronary heart disease in the Honolulu Heart Program. Relationship to biologic and lifestyle characteristics.** *Am J Epidemiol* 1984, **119(5):**653-666.
7. Anand SS, Razak F, Yi Q, Davis B, Jacobs R, Vuksan V, Lonn E, Teo K, McQueen M, Yusuf S: **C-reactive protein as a screening test for cardiovascular risk in a multiethnic population.** *Arterioscler Thromb Vasc Biol* 2004, **24(8):**1509-1515.
8. Yusuf S, Reddy S, Ounpuu S, Anand S: **Global burden of cardiovascular diseases: part I: general considerations, the epidemiologic transition, risk factors, and impact of urbanization.** *Circulation* 2001, **104(22):**2746-2753.
9. Lanfear DE, Marsh S, Cresci S, Shannon WD, Spertus JA, McLeod HL: **Genotypes associated with myocardial infarction risk are more common in African Americans than in European Americans.** *J Am Coll Cardiol* 2004, **44(1):**165-167.
10. Bolli P: **The question of the role of ethnicity on cardiovascular risk: does it matter where we come from?** *J Hypertens* 2005, **23(7):**1331-1333.
11. Cavalli-Sforza LL, Feldman MW: **The application of molecular genetic approaches to the study of human evolution.** *Nat Genet* 2003, **33 Suppl:**266-275.
12. Cavalli-Sforza LL Piazza A, Menozzi P.: **History and Geography of Human Genes.** Princeton , Princeton University Press; 1994.
13. Chikhi L, Destro-Bisol G, Bertorelle G, Pascali V, Barbujani G: **Clines of nuclear DNA markers suggest a largely neolithic ancestry of the European gene pool.** *Proc Natl Acad Sci U S A* 1998, **95(15):**9053-9058.
14. Young JH, Chang YP, Kim JD, Chretien JP, Klag MJ, Levine MA, Ruff CB, Wang NY, Chakravarti A: **Differential susceptibility to hypertension is due to selection during the out-of-Africa expansion.** *PLoS Genet* 2005, **1(6):**e82.
15. Excoffier L: **Analysis of population subdivision.** In *Handbook of statistical genetics* Edited by: Balding DJ, Bishop M, Cannings C. Chichester, West Sussex , John Wiley & Sons, LTD; 2001.
16. Weir BS, Hill WG: **Estimating F-statistics.** *Annu Rev Genet* 2002, **36:**721-750.
17. Lewontin RC, Krakauer J: **Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms.** *Genetics* 1973, **74(1):**175-195.
18. Akey JM, Zhang G, Zhang K, Jin L, Shriver MD: **Interrogating a high-density SNP map for signatures of natural selection.** *Genome Res* 2002, **12(12):**1805-1814.
19. Kayser M, Brauer S, Stoneking M: **A genome scan to detect candidate regions influenced by local natural selection in human populations.** *Mol Biol Evol* 2003, **20(6):**893-900.
20. Schlotterer C: **Towards a molecular characterization of adaptation in local populations.** *Curr Opin Genet Dev* 2002, **12(6):**683-687.
21. Ryan AW, Mapp J, Moyna S, Mattiangeli V, Kelleher D, Bradley DG, McManus R: **Levels of interpopulation differentiation among different functional classes of immunologically important genes.** *Genes Immun* 2006, **7(2):**179-183.
22. Ding K, Kullo IJ: **Molecular evolution of 5' flanking regions of 87 candidate genes for atherosclerotic cardiovascular disease.** *Genet Epidemiol* 2006, **30(7):**557-569.
23. Rockman MV, Hahn MW, Soranzo N, Loisel DA, Goldstein DB, Wray GA: **Positive selection on MMP3 regulation has shaped heart disease risk.** *Curr Biol* 2004, **14(17):**1531-1539.
24. **The International HapMap Project** [http://www.hapmap.org]
25. Weir BS, Cardon LR, Anderson AD, Nielsen DM, Hill WG: **Measures of human population structure show heterogeneity among genomic regions.** *Genome Res* 2005, **15(11):**1468-1476.
26. **NCBI SNP database** [http://www.ncbi.nlm.nih.gov/projects/SNP/]
27. **USCS genome browser [genome.ucsc.edu].** .
28. Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, Feldman MW: **Genetic structure of human populations.** *Science* 2002, **298(5602):**2381-2385.
29. Bamshad MJ, Wooding S, Watkins WS, Ostler CT, Batzer MA, Jorde LB: **Human population genetic structure and inference of group membership.** *Am J Hum Genet* 2003, **72(3):**578-589.
30. Shriver MD, Kennedy GC, Parra EJ, Lawson HA, Sonpar V, Huang J, Akey JM, Jones KW: **The genomic distribution of population substructure in four populations using 8,525 autosomal SNPs.** *Hum Genomics* 2004, **1(4):**274-286.
31. Bamshad M, Wooding S, Salisbury BA, Stephens JC: **Deconstructing the relationship between genetics and race.** *Nat Rev Genet* 2004, **5(8):**598-609.
32. Schaffner SF, Foo C, Gabriel S, Reich D, Daly MJ, Altshuler D: **Calibrating a coalescent simulation of human genome sequence variation.** *Genome Res* 2005, **15(11):**1576-1583.
33. Altshuler D, Brooks LD, Chakravarti A, Collins FS, Daly MJ, Donnelly P: **A haplotype map of the human genome.** *Nature* 2005, **437(7063):**1299-1320.
34. Rockman MV, Hahn MW, Soranzo N, Zimprich F, Goldstein DB, Wray GA: **Ancient and recent positive selection transformed opioid cis-regulation in humans.** *PLoS Biol* 2005, **3(12):**e387.
35. Weir BS, Chockerham CC: **Estimating F-statistics for the analysis of population structure.** *Evolution* 1984, **38:**1358-1370.
36. Bamshad M: **Genetic influences on health: does race matter?** *JAMA* 2005, **294(8):**937-946.
37. Cohen J, Pertsemlidis A, Kotowski IK, Graham R, Garcia CK, Hobbs HH: **Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in PCSK9.** *Nat Genet* 2005, **37(2):**161-165.
38. Hoffmann SC, Stanley EM, Cox ED, DiMercurio BS, Koziol DE, Harlan DM, Kirk AD, Blair PJ: **Ethnicity greatly influences cytokine gene polymorphism distribution.** *Am J Transplant* 2002, **2(6):**560-567.
39. Splawski I, Timothy KW, Tateyama M, Clancy CE, Malhotra A, Beggs AH, Cappuccio FP, Sagnella GA, Kass RS, Keating MT: **Variant of SCN5A sodium channel implicated in risk of cardiac arrhythmia.** *Science* 2002, **297(5585):**1333-1336.
40. Garte S: **Locus-specific genetic diversity between human populations: an analysis of the literature.** *Am J Hum Biol* 2003, **15(6):**814-823.
41. Hudson RR: **A new statistic for detecting genetic differentiation.** *Genetics* 2000, **155(4):**2011-2014.

42. Nicholson G, Smith AV, Jonsson F, Gustafsson O, Stefanssonand K, Donnelly P: **Assessing population differentiation and isolation from single-nucleotide polymorphism data.** *Journal of the Royal Statistical Society Series B* 2002, **64:**695-715.
43. Clark AG, Hubisz MJ, Bustamante CD, Williamson SH, Nielsen R: **Ascertainment bias in studies of human genome-wide polymorphism.** *Genome Res* 2005, **15(11):**1496-1502.
44. Carlson CS, Eberle MA, Rieder MJ, Smith JD, Kruglyak L, Nickerson DA: **Additional SNPs and linkage-disequilibrium analyses are necessary for whole-genome association studies in humans.** *Nat Genet* 2003, **33(4):**518-521.
45. Hinds DA, Stuve LL, Nilsen GB, Halperin E, Eskin E, Ballinger DG, Frazer KA, Cox DR: **Whole-genome patterns of common DNA variation in three human populations.** *Science* 2005, **307(5712):**1072-1079.
46. Balloux F, Lugon-Moulin N: **The estimation of population differentiation with microsatellite markers.** *Mol Ecol* 2002, **11(2):**155-165.
47. Izagirre N, Garcia I, Junquera C, de la Rua C, Alonso S: **A Scan for Signatures of Positive Selection in Candidate Loci for Skin Pigmentation in Humans.** *Mol Biol Evol* 2006, **23(9):**1697-1706.
48. Teshima KM, Coop G, Przeworski M: **How reliable are empirical genomic scans for selective sweeps?** *Genome Res* 2006, **16(6):**702-712.
49. Storey JD, Tibshirani R: **Statistical significance for genomewide studies.** *Proc Natl Acad Sci U S A* 2003, **100(16):**9440-9445.
50. Grossman I, Avidan N, Singer C, Paperna T, Lancet D, Beckmann JS, Miller A: **Genomic profiling of interpopulation diversity guides prioritization of candidate-genes for autoimmunity.** *Genes Immun* 2004, **5(6):**493-504.
51. Rockman MV, Hahn MW, Soranzo N, Goldstein DB, Wray GA: **Positive selection on a human-specific transcription factor binding site regulating IL4 expression.** *Curr Biol* 2003, **13(23):**2118-2123.
52. Sabeti PC, Schaffner SF, Fry B, Lohmueller J, Varilly P, Shamovsky O, Palma A, Mikkelsen TS, Altshuler D, Lander ES: **Positive natural selection in the human lineage.** *Science* 2006, **312(5780):**1614-1620.
53. Wang H, Ding K, Zhang Y, Jin L, Kullo IJ, He F: **Comparative and Evolutionary Pharmacogenetics of ABCB1: Complex Signatures of Positive Selection on Coding and Regulatory Regions.** *Pharmacogenet Genomics* 2007, **17(8):**667-78.
54. Kong A, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, Richardsson B, Sigurdardottir S, Barnard J, Hallbeck B, Masson G, Shlien A, Palsson ST, Frigge ML, Thorgeirsson TE, Gulcher JR, Stefansson K: **A high-resolution recombination map of the human genome.** *Nat Genet* 2002, **31(3):**241-247.
55. **PUBMED** [http://www.ncbi.nlm.nih.gov/pubmed]
56. **PANTHER classification system** [http://panther.appliedbiosystems.com]
57. Mi H, Lazareva-Ulitsky B, Loo R, Kejariwal A, Vandergriff J, Rabkin S, Guo N, Muruganujan A, Doremieux O, Campbell MJ, Kitano H, Thomas PD: **The PANTHER database of protein families, subfamilies, functions and pathways.** *Nucleic Acids Res* 2005, **33(Database issue):**D284-8.
58. **NCBI Reference sequence** [http://www.ncbi.nlm.nih.gov/RefSeq/]
59. Weir BS: **Genetic data analysis II.** Sunderland, MA , Sinauer Associated; 1986.
60. Hudson RR, Slatkin M, Maddison WP: **Estimation of levels of gene flow from DNA sequence data.** *Genetics* 1992, **132(2):**583-589.
61. **Genepop software** [http://wbiomed.curtin.edu.au/genepop/]
62. **cosi program** [http://www.broad.mit.edu/~sfs/cosi]
63. Nei M, Maruyama T: **Letters to the editors: Lewontin-Krakauer test for neutral genes.** *Genetics* 1975, **80(2):**395.
64. Hamblin MT, Thompson EE, Di Rienzo A: **Complex signatures of natural selection at the Duffy blood group locus.** *Am J Hum Genet* 2002, **70(2):**369-383.