Genetic loci associated with coronary artery disease harbor evidence of selection and antagonistic pleiotropy

Sean G. Byars[1,2,*], Qinqin Huang[1], Lesley-Ann Gray[1,2], Samuli Ripatti[3,4,5], Gad Abraham[1,2], Stephen C. Stearns[6], Michael Inouye[1,2,3,*]

[1] Centre for Systems Genomics, School of BioSciences, The University of Melbourne, Parkville 3010, Victoria, Australia
[2] Department of Pathology, The University of Melbourne, Parkville, Victoria 3010, Australia
[3] National Institute for Health and Welfare, Helsinki, Finland
[4] Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, United Kingdom
[5] Department of Public Health, University of Helsinki, Helsinki, Finland
[6] Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT, USA

* Corresponding authors: Sean Byars (sean.byars@unimelb.edu.au), Michael Inouye (minouye@unimelb.edu.au)

Summary

Traditional genome-wide scans for positive selection have mainly uncovered selective sweeps associated with (near) monogenic traits. While selection on quantitative traits is much more common, very few such signals have been detected because of their polygenic nature. We searched for positive selection signals underlying coronary artery disease (CAD), which in both past and present has caused considerable disease in humans. In worldwide populations, we used novel approaches to quantify the relationship between polygenic selection signals and genetic risk of CAD in a genome-wide meta-analysis. We identified candidate loci that appear to have been directly modified by CAD selective pressures. Several loci showed significant positive relationships between selection and CAD genetic risk, and several showed negative relationships, depending on population. Top adaptive candidates were also associated with traits that could modify reproductive variation (fitness) suggesting antagonistic tradeoffs with genetically correlated phenotypes expressed earlier in life. Variants under selection at CAD loci also showed more evidence of gene regulatory effects in HapMap3 lymphoblastoid cell lines than those with no evidence of selection. Our study provides a novel approach for detecting selection on polygenic traits and evidence that modern human genomes have evolved in response to selection induced by CAD and other early-life traits sharing pleiotropic links with CAD.

Keywords: coronary artery disease, positive selection, antagonistic pleiotropy, integrated haplotype score, eQTL

Highlights:

- Widespread genomic signals of positive selection are present underlying coronary artery disease (CAD) loci
- Selection peaks that significantly associated with genetic risk suggest loci modified (in)directly by CAD
- Selection was more often associated with variants important for regulating gene expression
- CAD loci share many pleiotropic links with early-life traits suggesting antagonistic effects

Introduction

It is well established that modern human traits are a product of past evolutionary forces that have shaped heritable phenotypic and molecular variation, but we are far from understanding what diseases have driven natural selection and how this process has left its imprint across the genome. Although many recent genome-wide multi-population scans have searched for signatures of positive selection [1-9], these studies have detected few signals of selection on candidate loci associated with traits or diseases [10-12]. This suggests that classic 'selective sweeps' have been relatively rare in recent human history [13, 14] and that the tools currently used miss most of the smaller selection signals caused by diseases associated with polygenic traits [12]. This limits our understanding of how natural selection has acted on variation underlying complex diseases. In this study, we aimed to comprehensively identify positive selection signals underlying coronary artery disease (CAD) loci with methods designed to detect signals of recent positive selection. We also compared quantitative selection signals in 12 worldwide populations (HapMap3) with patterns of disease risk to identify signals of selection linked to CAD pressure.

Classic population genetics theory describes positive selection with the selective-sweep (or hard-sweep) model, in which a strongly advantageous mutation increases rapidly in frequency (often to fixation) resulting in reduced heterozygosity of nearby neutral polymorphisms due to genetic hitch-hiking [15, 16] and a longer haplotype with higher frequency. Many methods have been developed to detect these signatures [17, 18], including traditional tests that detect differentiation in allele frequencies among population (i.e. Wright's fixation index, Fst [19]) and more recently developed within population tests for extended haplotype homozygosity (i.e. integrated haplotype score, iHS [9]). Some of the most convincing examples of human adaptive evolution have been uncovered for traits influenced by single loci with large effects. For example, the lactase persistence (*LCT*) and Duffy-null (*DARC*) mutations affecting expression of key proteins in milk digestion [10] and malarial resistance [20] both display hallmarks of selective sweeps. Other loci that are not clearly monogenic but also show selective sweeps are associated with high-altitude tolerance (*EPAS1* [21]) and skin pigmentation (*SLC24A5* and *KITLG* [22]). These previous studies showed that rapid selective sweeps occurred around loci where alleles that were

previously rare or absent in populations had large effects on phenotypes.

Motivated by these initial successes and the increasing availability of global population data genotyped on higher resolution arrays (i.e. HapMap Project, 1000 Genomes Project), many genome-wide scans for candidate adaptive loci have recently been performed [11]. These studies suggest that selection may have operated on a variety of biological processes [10] in ways that differ among populations (i.e. local adaptation) [23], has been prevalent in genetic variation linked to metabolic processes [24], and may have often targeted intergenic regions and gene regulatory variants rather than protein-coding regions [12]. However, only the larger signals underlying monogenic traits are typically captured due to the lack of statistical power imposed by the need to correct for genome-wide multiple testing [18]. Most of these candidates also are not yet convincing due to inconsistencies between studies that utilized the same data [14], cannot be validated due to the absence of biological or functional information [25, 26], and perhaps because selective sweeps have actually been rare in human populations [27, 28].

In contrast to population genetics, in quantitative genetics rapid adaptation typically involves selection acting on quantitative traits that are highly polygenic [29, 30]. Under the 'infinitesimal (polygenic) model', such traits are likely to respond quickly to changing selective pressures through smaller frequency shifts in many polymorphisms already present in the population [13, 31]. Such alleles would not necessarily sweep to fixation, would produce smaller changes in surrounding heterozygosity, and would thus be hard to detect with most current population genetic methods [14, 26, 32]. Note that polygenic and classic sweep models are not mutually exclusive [13, 33], for alleles with small- and large-effects may both underlie a polygenic trait. Thus the degree to which candidate alleles will be detectable after a selective event will vary. Given that most common diseases are highly polygenic [34], this suggests a need to improve how we detect and understand adaptive signatures in the loci associated with polygenic traits.

Recent selection studies investigating polygenic traits have taken two approaches. The first scans for significant selection signals within genome-wide significant disease effect SNPs. For example, Ding and Kullo [35] found significant population differentiation (Fst) for 8 of 158 index SNPs underlying 36 cardiovascular disease phenotypes, and Raj et al. [36] observed elevated positive selection scores (Fst, iHS) for 37 of 416 index susceptibility SNPs underlying 10 inflammatory-diseases. The second approach tests if aggregated shifts in genome-wide significant allele frequencies are associated with phenotypic differences by population, latitudinal, or environmental gradients, which might indicate local adaptation. For example, Castro and Feldman [37] used 1300 index SNPs underlying many polygenic traits and found elevated adaptive signals (Fst and iHS) above background variation, and Turchin et al. [38] demonstrated moderately higher frequency of 139 height-increasing alleles in a Northern (taller) compared to Southern (shorter) European populations. These approaches all assume that the variants with the most significant p values are the most probable selection targets, but many if not most such variants are tagging tested or untested causal variants, which may themselves be of lower frequencies. This suggests an approach sensitive to more subtle signals of selection and disease risk is needed for polygenic selection.

We chose CAD as a model for examining polygenic selection signals underlying complex disease because it has (and continues to) impose considerable disease burden (selection pressure) in humans [39], its underlying genetic architecture has been extensively studied [40, 41] and many of its risk factors (cholesterol, blood pressure) have been under recent natural selection [42] related to potential pleiotropic effects or tradeoffs with CAD. Antagonistic pleiotropy describes gene effect on multiple linked traits where selection on one may cause fitness tradeoffs (i.e. disease, survival) in the other due to their negative genetic association [43]. Two common

misconceptions are that CAD is exclusively late age of onset and only occurs at appreciable frequency in contemporary humans. If that were true, selection might not have had either the opportunity or sufficient time to affect genetic variation associated with CAD. However, CAD manifests early in life [45, 46] and can be detected even in adolescence through degree of atherosclerosis [46, 47] and myocardial infarction events [48]. CAD is also a product of many heritable risk factors (cholesterol, weight, blood pressure) whose variation is expressed during the reproductive period, when CAD could drive selection directly or indirectly. Furthermore, CAD has impacted human populations since at least the ancient Middle Kingdom period, with studies finding the presence of atherosclerosis in Egyptian mummies [49]. This suggests that there has been enough time for genomic signatures of selection related to CAD to develop and be detectable in modern humans.

By combining several 1000 Genomes-imputed datasets including HapMap3 and Finnish SNP data, a large genetic meta-analysis of CAD, and HapMap3 gene expression data, we sought to address the reason(s) why CAD exists in humans by answering the following questions: 1) Has selection recently operated on CAD loci 2) How do selection signals underlying CAD loci vary among populations and are they enriched for gene regulatory effects? 3) Do candidate adaptive signatures overlap directly with CAD genetic risk and is this useful for highlighting disease-linked selection signals? 4) Do CAD-linked selection signals display functional effects and evidence of antagonistic pleiotropy, in that they are also linked to biological processes or traits influencing reproduction?

Results

To test for selection signals for variants directly linked with CAD, we utilized SNP summary statistics from 56 genome-wide significant CAD loci in Nikpay et al. [41], the most recent and largest CAD case-control GWAS meta-analysis to date, to identify 76 candidate genes for CAD (**Experimental Procedures**). Nikpay et al. used 60,801 CAD cases and 123,504 controls from a mix of individuals of mainly European (77%), south (13% India and Pakistan) and east (6% China and Korea) Asian, Hispanic and African American (~4%) descent with genetic variation imputed to a high-density using the 1000 Genomes reference panel. By investigating all SNPs in candidate CAD genes, we aimed to improve detection of smaller polygenic selection signals for the range of functional genic variants and short-range intergenic regulatory variants that would be missed with approaches that only consider genome-wide significant SNPs.

Signals of positive selection within coronary artery disease loci

We utilised the integrated Haplotype Score (iHS) to estimate positive selection for each SNP underlying candidate CAD genes within each population separately. Because iHS is typically

used to detect candidate adaptive SNPs where the selected alleles may not have reached fixation [9], this estimate is well suited for detecting recent signals of selection as opposed to other measures [18]. iHS is also better suited for detecting selection acting on standing variation in polygenic traits [18, 50].

Candidate selection signals were found for many of the 76 CAD genes within each of the 12 worldwide populations (11 HapMap3 populations and Finns; Fig. 1A for top 40 based on their association with CAD log odds genetic risk, Fig. S1 for all 76). These were defined as 'peaks' of significantly elevated iHS scores across SNPs within each gene-population combination, with the apex approximating the likely positional target of positive selection.

In the sample of all populations (Fig. 1A, largest iHS scores), most candidate selection signals were relatively small, but a few larger signals were detected. For example, out of the 912 gene-by-population combinations (Fig. S1), 354 (38%) contained weak-moderate candidate selection signals (significant iHS between 2-3), 84 (9%) contained moderate-strong signals (significant iHS between 3-4), and 6 (0.6%) had very strong signals (significant iHS > 4). The 6 largest selection signals were found in the following gene-population combinations: *BCAS3* in GIH (iHS=4.45), MEX (iHS=4.23) and CEU (iHS=4.86), *PEMT* in MKK (iHS=4.24), *ANKS1A* in LWK (iHS=4.03), and *CXCL12* in JPT (iHS=4.10), with all iHS p values <0.0001. Six genes (*BCAS3*, *SMG6*, *PDGFD*, *KSR2*, *SMAD3*, *HDAC9*) exhibited candidate selection signals consistently within all populations (Fig. 1A), and many genes also contained consistent selection signals for all populations within similar ancestral groups (e.g. African, European etc, Fig. 1A).

Within CAD genes, multiple candidate selection signals were sometimes present (particularly within larger genes, within separate linkage disequilibrium (LD)-blocks); these varied between and sometimes within a population. For example, in *PHACTR1* (~0.57mb in size, 14 introns) there are three main candidate selection signals in introns 4, 7 and 11 (see Fig. S2, comparing cross-population selection signals in *PHACTR1*) that were in separate LD-blocks (see Fig. 3C, LD plots). Within most populations, there was a broad and relatively weak set of candidate selection signals in intron 4 (the largest *PHACTR1* intron, ~300kb in length). Intron 4 is also the location of the published CAD index SNP (rs9369640) for *PHACTR1*. Three of the African populations had the highest iHS score for the same SNP in intron 4 (rs8180558) including ASW (iHS=2.4, P<0.05), LWK (iHS=2.8, P<0.01) and YRI (iHS=2.2, P<0.05), which is ~18kb upstream from the index CAD SNP (r^2 between rs8180558 and rs9369640 in *PHACTR1*: ASW=0.12; LWK=0.03; YRI=0.04). Peaks of *PHACTR1* selection signals within the three Asian populations were at rs4715043 in CHB (iHS=2.3, P<0.05) and rs6924689 in both CHD (iHS=2.9, P<0.01) and JPT (iHS=3.0,P<0.01). The GIH population contained the largest selection signal, also in intron 4, with an apex at rs4142300 (iHS=3.7, P<0.001, 75kb downstream of/r^2=0.07 with index CAD SNP rs9369640). This corresponded with the same apex SNP in intron 4 for TSI, though the TSI signal was weaker and non-significant (rs4142300, iHS=1.84); rs4142300 was also close to the apex SNP in CEU (rs9349350, iHS=2.0, P<0.05, r^2=0.92) and MEX (rs2015764, iHS=2.1, P<0.05, r^2=0.30). Other significant candidate selection signals were also present in intron 7 for three of the African populations (ASW, LWK, MKK), the CHD and GIH populations, with the largest intron 7 signal within MKK (SNP rs13191209, iHS=3.0, P<0.001). The last significant candidate selection signal within *PHACTR1* was found within intron 11 with the largest signal at rs9349549 (MKK iHS=2.9, P<0.01; CEU iHS=2.7, P<0.01; TSI iHS=3.0, P<0.01). Other interesting candidate selection signals present in other CAD genes (Fig. S1) are not discussed here. Such patterns suggest that candidate selection signals are complex and often do not correspond to the alleles with largest effect on CAD.

Relationship between CAD genetic risk and selection across populations

For each CAD gene within each population, we used a mixed effects linear model to regress SNP-based estimates of CAD log odds genetic risk (ln(OR), obtained from cardiogramplusc4d.org) against iHS selection scores (Experimental Procedures). We accounted for LD structure by including the first eigenvector from an LD matrix of correlations (r^2) between SNPs within each gene as a random effect.

For a subset of CAD loci, we found significant quantitative associations between disease risk and selection signals and for each of these the direction of this association was often consistent between populations (Fig. 1B). Furthermore, when compared to a null distribution of genes selected randomly from the genome, the strength of the CAD log odds versus selection signal at most loci was statistically significant (Fig. 1C). Fig. 1B shows 40 genes ranked based on those that showed the most consistent number of significant associations across the 12 populations, with those that showed fewer than four significant associations excluded. Positive and negative associations indicate elevated selection signals present in regions with higher or lower CAD log odds genetic risk, respectively.

In the comparison across populations, directionality of significant selection-risk associations tended to be most consistent for populations within the same ancestral group (Fig. 1B). For example, in *PHACTR1*, negative associations were present within all European populations (CEU, TSI, FIN), and in *NT5C2* strong positive associations were present in all East Asian populations (CHB, CHD, JPT). Other negative associations that were consistent across all populations within an ancestry group included five genes in Europeans (*COG5*, *ABO*, *ANKS1A*, *KSR2*, *FLT1*) and four genes (*LDLR*, *PEMT*, *KIAA1462*, *PDGFD*) in East Asians.

Additional consistent positive associations included four genes (*CNNM2*, *TEX41*, *NT5C2*, *MIA3*) in East Asians, three (*BCAS3*, *RAI1*, *KCNK5*) in Europeans, and one (*PPAP2B*) in Africans. In comparison to other ancestral groups, African populations showed fewer significant selection-risk associations (27.9% of all 76-gene x 12-population combinations) than Asians (31.5%) or Europeans (32.8%). Some associations were consistent in all but one population (e.g. *CNNM2*, *ABCG8* in Europeans; *BCAS3*, *KCNK5* in Asians; *CNNM2*, *TEX41* in Africans) or unique to one population within an ancestral group (e.g. *TEX41* in FIN, *COG5* in ASW).

Below we focus on *BCAS3* (Fig. 2) and *PHACTR1* (Fig. 3), two of the strongest selection-risk associations which, when adjusting for LD (**Experimental Procedures**), displayed varying directionality between at least two populations.

Genetic risk of CAD vs positive selection in BCAS3

The genetic risk of CAD for variants in *BCAS3* were positively correlated with an extremely large candidate adaptive signal in all European and two of three East Asian populations (Fig. 1B). For example in CEU, the largest iHS score was 4.85 and highly significant, and was elevated across most of *BCAS3* (Fig. 2B CEU, spanning introns 1-18 and various LD-blocks, Fig. 2C), which matched the approximate trends in CAD log odds giving rise to a highly significant positive correlation (Fig. 2A CEU). In contrast, in YRI there was no detectable selection signal close to the index SNP (Fig. 2B YRI), but weak-moderate signals were present towards the end of *BCAS3* (Fig. 2B YRI, introns 18-19, smaller LD-blocks Fig. 2C), which also corresponded with lower CAD log odds (Fig. 2B, YRI) thus giving rise to a significant negative correlation in Fig. 2A.

Genetic risk of CAD vs positive selection in PHACTR1

For all European populations, *PHACTR1* (see CEU example, Fig. 3A) selection peaks were typically located within regions of consistently lower CAD log odds (Fig. 3B). This contrasted with most other non-European populations where the highest candidate selection peaks were located within regions with elevated CAD log odds (including the index CAD SNP rs9369640, intron 4). The largest selection peak in GIH (Fig. 3B) overlapped the CAD log odds peak in *PHACTR1* giving rise to the strong positive association seen in Fig. 3A. The two distinctive selection peaks in both CEU and GIH were separated by different LD-blocks (Fig 3C), suggesting that these may have developed independently within *PHACTR1*. Interestingly, the negative association found for the MKK population was due to the location of the selection peaks more closely matching those of the European populations in intron 11 (Fig. S2).

Enrichment of gene regulatory variants under selection at CAD loci

To establish whether variants with evidence of selection in CAD genes also showed evidence of function, we performed an eQTL scan in 8 HapMap3 populations with matched LCL gene expression. We compared all SNPs in each CAD locus against expression for each focal gene within each population. We found that SNPs with significant integrated Haplotype Scores (iHS) were often also involved in gene regulation, compared to SNPs with non-significant selection scores (Fig. 4, Kolmogorov-Smirnov test p value <0.001). To assess which biological pathways were enriched for the highest-ranked genes according to Fig. 1B, i.e. those where selection scores were most closely associated with CAD log odds genetic risk, we included the top 10 genes into the Enrichr analysis tool [51] and found that these genes are especially enriched in pathways related to metabolism, focal adhesion and transport of glucose and other sugars. More interestingly, we found connections to reproductive phenotypes in the associations of these genes with pathways, ontologies, cell types and transcription factors. For example, we found links to ovarian steroidogenesis and genes expressed in specific cell types and tissues including the ovary, endometrium and uterus (see Table S4 for Enrichr outputs).

Discussion

This study has identified many candidate adaptive signals which suggests that selection on CAD loci is much more widespread than previously appreciated (also see Supplementary Discussion). It has previously been suggested [12] and demonstrated [52] that selection on gene expression levels has been an important element of human adaptation in general. We confirm this result for CAD associated loci. Positive selection signals within CAD loci were more likely than random SNPs to be associated with gene expression levels in *cis* (Fig. 4).

We found evidence that some of these signals may be a result of selection pressures induced directly by CAD itself. This finding is important for highlighting genes that may have been

modified directly by selection on disease phenotypes and also for our general understanding of how quickly human genomes can respond to selection induced by changing environments. Subsequent biological process analyses and a thorough literature assessment (below) demonstrated that the loci most consistently associated with CAD genetic risk are also often linked to human reproduction, which suggests both their potential to respond to natural section and their possible role via antagonistic pleiotropy in the reproductive tradeoffs that would help to explain why CAD exists in human populations.

Coronary artery disease-induced changes to human genomes

One of our most interesting findings was the significant association between selection signals and CAD log odds genetic risk. This approach of integrating genome scans of positive selection with genome-wide genotype-phenotype data has been promoted previously as a tool to uncover biologically meaningful selection signals of recent human adaptation [12, 52] but has rarely been applied. Among the exceptions, Jarvis et al. [55] found a cluster of selection and association signals coinciding on chromosome 3 that included genes *DOCK3* and *CISH*, which are known to affect height in Europeans.

For highly-ranked genes (according to the number of significant associations present within the 12 populations) in Fig. 1B such as *BCAS3*, *CNNM2*, *TEX41*, *SMG6* and *PHACTR1*, the consistent overlap between selection and genetic risk of CAD suggests that many of these may have been modified by CAD-linked selective pressures. If so, then two conditions must have been met. Firstly, CAD was present for long enough to be involved in these genetic alterations, an evolutionary process which generally takes thousands of years. Indeed, precursors of CAD (i.e. atherosclerosis) are detectable in very early civilizations [49]. Secondly, the effects of CAD were directly or indirectly expressed during the reproductive period and trait variation was under natural selection due to its effects on reproductive success.

It is only possible for natural selection to directly act on CAD if those outcomes modify individual fitness relative to others in the same population. As outlined in the introduction, this is possible as CAD outcomes (i.e. myocardial infarction) do occur in young adults. However, early-life CAD outcomes are relatively rare, suggesting selection is more likely to operate indirectly on CAD via its risk factors (or other pleiotropically linked traits, discussed below), which provides a more likely explanation for the close associations we found between positive selection and genetic risk. Supporting this, phenotypic selection has been found operating on CAD risk factors [42], suggesting that these selection pressures are still present in modern humans.

Some genes had large signals of selection but showed weak or no consistent overlap with CAD genetic risk. For example *HDAC9* (Histone Deacetylase 9) shows extensive evidence for having undergone recent selection within most populations, especially those of European or Mexican decent, but little or no overlap with CAD risk was evident in most populations. This suggests positive selection has operated on this gene due to its effects on a trait unrelated to CAD, which may not be surprising given *HDAC9*'s broad biological roles (as a transcriptional regulator, cell-cycle progression) and association with other very different phenotypes including ulcerative colitis [57] and psychiatric disorders [58]. This further demonstrates that this approach is useful for separating candidate selection signals important for the disease or phenotype of interest from those that aren't.

Pleiotropic effects that establish the genetic foundations of tradeoffs

To further investigate whether top candidate adaptive loci for CAD modify fitness or share pleiotropic links with other traits that may modify fitness, we performed an extensive systematic literature search on the 40 top-ranked genes in Fig. 1 and a random set of 20 genes. If they have been under selection recently, they might still be associated with reproductive variation (i.e. fitness) in modern environments. We found that all 40 CAD genes shared at least one (often more) connection with fitness (Table S1-S2). Some appear to directly influence fitness (offspring number, age at menarche, menopause, survival), while many were associated with early-life reproductive traits that are likely to indirectly correlate with fitness including variation in ability to fertilize/conceive or fetal growth, development and survival. To test the novelty of this, we randomly chose 20 genes that were approximately the same size as the top 20 genes in Fig. 1. We only found three (out of 20) random genes with at least one potential link with fitness (Table S3). This suggests there are unique pleiotropic links between CAD and traits that have likely been under selection earlier in life.

Evidence for direct links between CAD genes and fitness (Table S1-S2) included genes associated with reproductive (*PPAP2B*, [59]) or twinning (*SMAD3*, [60]) capacity and number of offspring produced (e.g. *KIAA1462*, [61], *SLC22A5*, [62]). *PHACTR1*, *LPL*, *SMAD3*, *ABO* and *SLC22A5* may contribute to reproductive timing (menarche, menopause) in women [63-65] and animals [66]. Expression of *PHACTR1* [67], *KCNK5* [68], *MRAS* and *ADAMST7* [69] appear to regulate lactation capacity. Some gene deficiencies also cause pregnancy loss (e.g. *LDLR*, [70], *COL4A2*, [71]). Evidence for antagonistic links were much more common and included these: 25 genes shared links with traits expressed during pregnancy (Table S1-S2), i.e. variation that can negatively influence the health and survival outcomes of both the fetus and mother [72]. For example, a variant of *CDKN2B-AS1* significantly contributes to risk of fetal growth restriction [73], both *FLT1* [74] and *LPL* [75] are significantly differentially expressed in placental tissues from pregnancies with intrauterine growth restriction (IUGR), and preeclampsia and *LDLR*-deficient mice had litters with significant IUGR [76]. A further 29 and 19 genes were linked to traits that can directly influence female and male fertility, respectively (13 influence both) (Table S1-S2). For example, *BCAS3* and *PHACTR1* are highly expressed during human embryogenesis [77, 78], *SWAP70* is intensely expressed at the site of implantation [79], and *PHACTR1* may play a role in receptivity to implantation [80]. For *ABCG8* and *KSR2,* animal models provide further support as gene expression deficiency can cause infertility in females (*ABCG8*, [81]) and males (*KSR2*, [82]).

Pleiotropic connections were also apparent in the classification of specific disorders or from studies investigating single-gene effects. For example, women with polycystic ovarian syndrome (PCOS) have higher rates of infertility due to ovulation failure and modified cardiovascular disease risk factors (i.e. diabetes, obesity, hypertension [83]). A number of CAD genes in this study (e.g. *PHACTR1*, *LPL*, *PDGFD*, *IL6R*, *CNNM2*) are found differentially expressed in PCOS women [84-88], suggesting possible links between perturbed embryogenesis and angiogenesis. In males, this can be demonstrated with a mutation in *SLC22A5* that causes both cardiomyopathy and male infertility due to altered ability to break down lipids [89, 90]. More generally, many recent studies link altered cholesterol homeostasis with fertility, which is most apparent in patients suffering from hyperlipidemia or metabolic syndrome [91, 92].

To facilitate interpretation of selection occurring on early-life traits or CAD phenotypic risk factors that share pleiotropic connections and possible evolutionary tradeoffs with coronary artery disease, we present a conceptual figure (Fig. 5). These pleiotropic effects are important because many of them affect traits expressed early in life, some extremely early in life. Any allele that increases reproductive performance enough early in life to more than compensate for a loss of

associated fitness late in life will be selected [43]. Such a mechanism has been recently suggested to help explain the maintenance of polymorphic disease alleles in modern human populations [93]. Some previous studies have tested for such tradeoffs in humans using direct fitness-related phenotypes (e.g. [44]) although evidence for such a mechanism influencing human disease is currently lacking. Our approach examining antagonistic fitness effects for disease genes that displayed consistent selection-genetic risk associations in diverse worldwide populations provides support for such a mechanism influencing CAD. Here we have presented multiple cases in which such antagonistic pleiotropy appears to be present for genes associated with CAD, which may help to explain our vulnerability to the disease.

Study limitations

There are also some limitations to our approach. We utilized CAD genetic risk estimated from a meta-analysis based on predominantly European (77%) with smaller contributions from south/east Asian (19%), Hispanic and African American (~4%) ancestry [41]. Genetic risk variation for CAD might be different in the un-represented (i.e. Mexican) or less-represented (i.e. African) populations in this meta-analysis. If that were the case, it would reduce the usefulness of comparing selection and risk estimates in those populations. We also saw fewer significant selection-risk associations in the African populations (Fig. 1B), however this may be due to selection signals in the African populations being less obvious than those in East Asian and European populations, perhaps due to lesser linkage disequilibrium, as is consistent with results from previous studies [94]. Calculating disease risk and selection variation from populations within the same ancestral group might help resolve this, however it only represents a potential shortcoming for our cross-population analyses and not observations of antagonistic pleiotropy.

Summary

In this study, we found evidence that natural selection has recently operated on CAD associated variation. By comparing positive selection variation with genetic risk variation at known loci underlying CAD, we were able to identify and prioritize genes that have been the most likely targets of selection related to this disease across diverse human populations. That selection signals and the direction of selection-risk relationships varied among some populations suggests that CAD-driven selection has operated differently in these populations and thus that these populations might respond differently to similar heart disease prevention strategies. The pleiotropic effects that genes associated with CAD have on traits associated with reproduction that are expressed early in life strongly suggests some of the evolutionary reasons for the existence of human vulnerability to CAD.

Experimental Procedures

Defining loci linked to coronary artery disease

We started with the 56 lead index SNPs from Supplementary Table 5 in Nikpay et al. [41] corresponding to 56 CAD loci. When the index SNP was genic, all SNPs within that gene were extracted (using NCBI's dbSNP) including directly adjacent intergenic SNPs ±5000bp from untranslated regions (UTR) in LD>0.7 (with any respective genic SNP). When the index SNP was intergenic, that SNP and other directly adjacent SNPs ±5000bp and in LD>0.7 (with the index SNP) were extracted and combined with SNPs from the respective linked gene listed in Nikpay et al. [41] including SNPs ±5000bp from UTR regions in LD>0.7 with that gene. This resulted in SNP lists for 56 genes. To further explore other genes not directly connected with lead index SNPs, but that were found within the CAD loci identified by Nikpay et al. [41], we extracted SNPs within each of those genes (plus SNPs ±5000bp from UTR regions in LD>0.7 with that gene). This resulted in SNP lists for a further 20 genes, bringing the total number of candidate genes for CAD to 76.

The per-SNP log odds (ln(OR)) values for the 76 genes were obtained from Nikpay et al. [41] available at http://www.cardiogramplusc4d.org/downloads and used in the analysis described below.

Preparation of HapMap3 samples

Genotype data (1,457,897 SNPs, 1,478 individuals) were downloaded for 11 HapMap Phase 3 (release 3) populations (http://www.hapmap.org [95]) including: Yoruba from Ibadan, Nigeria (YRI), Maasai in Kinyawa, Kenya (MKK), Luhya in Webuye, Kenya (LWK), African ancestry in Southwest USA (ASW), Utah residents with ancestry from northern and western Europe from the CEPH collection (CEU), Tuscans in Italy (TSI), Japanese from Tokyo (JPT), Han Chinese from Beijing (CHB), Chinese in Metropolitan Denver, Colorado (CHD), Gujarati Indians in Houston, TX, USA (GIH), and Mexican ancestry in Los Angeles, CA, USA (MEX). We also included another HapMap3 population, the Finnish in Finland (FIN) sample (ftp://ftp.fimm.fi/pub/FIN_HAPMAP3 [96]). These data had already been pre-filtered, i.e. SNPs were excluded that were monomorphic, call rate < 95%, MAF<0.01, Hardy-Weinberg equilibrium P $<1x10^{-6}$ etc.

Before phasing and imputation, we performed a divergent ancestry check with flashpca [97] to check accuracy of population assignments, converted SNP data from build 36 to 37 with UCSC LiftOver (https://genome.ucsc.edu/cgi-bin/hgLiftOver), checked strand alignment in Plink v1.9 [98] to ensure all genotypes were reported on the forward strand, and kept only autosomal SNPs. To speed up imputation, data were first pre-phased with Shapeit v2 [99] using the duoHMM option that combines pedigree information to improve phasing and default values for window size (2Mb), per-SNP conditioning sates (100), effective population size (n=15000) and genetic maps from the 1000 Genomes Phase 3 b37 reference panel (ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/).

Phased data were imputed in 5 Mb chunks across each chromosome with Impute v2 [100]. We then removed any multiallelic SNPs (insertions, deletions etc) from the imputed data and excluded SNPs with call rate < 95%, HWE P $<1x10^{-6}$ and MAF<1%. The final dataset was then phased with Shapeit v2, and alleles were converted to ancestral and derived states using python script. Ancestral allele states came from 1000 Genomes Project FASTA files and derived 6-primate (human, gorilla, orangutan, chimp, macaque, marmoset) Enredo-Pecan-Ortheus alignment [101] from the Ensembl Compara 59 database [102].

Estimating signatures of recent selection

Integrated Haplotype Score (iHS): Using the package rehh [103] in R version 3.1.3, per SNP iHS scores were calculated within each population (after excluding non-founders) using methods described previously [9]. iHS could not be calculated for SNPs without an ancestral state, or whose population minor allele frequency is <5%, or for some SNPs that are close to chromosome ends or large regions without SNPs [9]. Rehh was also used to standardize (mean 0, variance 1) iHS values empirically to the distribution of available genome-wide SNPs with similar derived allele frequencies. For analyses in the main text, we considered a SNP to have a candidate selection signal if it had an absolute iHS score > 2, a permuted p value <0.05, and was within a 'cluster' of SNPs that also had elevated iHS scores. Although permuting p values is computationally more intensive, it provides more flexibility to detect smaller selection signals that may be incorrectly classified with the more stringent Bonferroni correction that is often applied to these estimates. For the analyses described below, even though we only used iHS estimates for the SNPs defined in the CAD genes (and additional SNPs for permutation purposes), we calculated per-SNP iHS scores genome-wide (rather than locally, i.e. within 1MB regions around focal SNPs), for this provides more accurate estimates because final adjustments are made relative to other genome-wide SNPs of similar sized derived allele frequency classes. P values for iHS scores were permuted based on comparison of nominal p values against 10000 randomly selected estimates from within the same derived allele frequency classes.

Comparing CAD genetic risk and quantitative selection signals

We first tested the null hypothesis that there is no association between CAD genetic risk and signals of positive selection for CAD genes. For each gene within each population, we used a mixed effects linear model to regress SNP-based estimates of CAD log odds (ln(OR)) genetic risk against selection scores (iHS) resulting in 912 separate regressions. To account for LD structure (and potential confounding of highly correlated SNPs) within each gene, we also included the first eigenvector derived from an LD matrix of correlations (r^2) between SNPs within each gene as a random effect. We chose to model LD structure with mixed-effects models rather than LD-prune because for many genes, the sample would have been too small for regression analyses. Also, it would be very difficult to properly capture both selection and the CAD log odds peaks needed to compare these variables. We accounted for multiple testing by permuting p values for each regression based on comparing each nominal p value against 10000 permuted p values derived from shuffling iHS scores.

Genes were then ranked based on the number of significant associations summed across the 12 populations. The 40 genes with at least four or more significant associations are shown in Fig. 1B. To illustrate the positional architecture of these selection-risk associations, plots for selected highly-ranked genes are shown in Fig. 2-3. By demonstrating how CAD genetic risk peaks and valleys correspond to variation in the magnitude of selection scores (iHS), this allowed visual assessment of potential modifications made to the phenotype-genotype map by selective pressures imposed directly or indirectly by CAD. It also helped us localize selection peaks within genes and compare them between populations. Similar peaks suggested similar selection and different peaks suggested local adaptation. This way of presenting the results also allowed us to detect the smaller adaptive shifts in allele frequencies typically expected to underlie selection on polygenic traits.

We then tested a second null hypothesis: that the selection-risk associations using the CAD genes are not unique compared to non-CAD associated loci. For each of the 76 CAD genes, we randomly (without replacement) chose 100 genes of similar length across the genome and

performed the same mixed effects regression procedure described above for each gene by population combination using both CAD log odds values from Nikpay et al. [41], iHS scores estimated from the SNP data, and the first LD eigenvector from SNPs within a gene. Permuted p values were derived by comparing the nominal p value for each CAD gene against the 100 null distribution p values from the non-CAD associated genes. Results are shown in Fig. 1C.

Identifying functional targets of selection

To examine whether candidate adaptive signals within each gene corresponded to a gene's regulatory variation, we regressed SNPs within focal genes and gender against that gene's probe expression levels, which had previously been quantified in lymphoblastoid cell lines using Illumina's Human-6 v2 Expression BeadChip for eight of the 12 populations [104]. While selection related to CAD may have targeted regulatory variants important for other tissues/cell-types, gene expression data was only available for this cell-type. Given the central importance of circulating lymphoblastoid cells in CAD and its risk factors, we might expect this cell type a good candidate to search for association between selection signals and regulatory variants important for these genes. The raw gene microarray expression data had previously been normalized on a $\log_2$ scale using quantile normalization for replicates of a single individual then median normalization for each population [104]. P values for each SNP-probe association were permuted using 10000 permutations by randomly shuffling gene probes expression. P values were then extracted for the most significant iHS score for each gene-population combination and compared to the same number of p values randomly drawn from different LD blocks underlying SNPs with non-significant iHS scores across each gene-population combination. A Kolmogorov-Smirnov test was used to compare the distribution of p values from each. To examine what biological processes were associated with the top ranked genes from Fig. 1, we uploaded the top 10 genes into Enrichr (http://amp.pharm.mssm.edu/Enrichr/) to define associated pathways (i.e. KEGG 2015, kegg.jp/kegg), ontologies (MGI Mammalian phenotypes, informatics.jax.org), cell types (Cancer cell line Encyclopedia, broadinstitute.org/ccle) and transcription factors (ChEA 2015, amp.pharm.mssm.edu/lib/chea.jsp).

Author Contributions

Conceptualization, S.G.B. and M.I.; Methodology, S.G.B. and M.I.; Formal analysis, S.G.B. and Q.H.; Literature review, S.G.B.; Writing – original draft, S.G.B. and M.I.; Writing – review & editing, S.G.B., Q.H., L.G., S.R., G.A., S.C.S and M.I.; Visualization, S.G.B.; Funding acquisition, M.I.; Supervision, M.I.

Acknowledgements

This study was supported by the National Health and Medical Research Council (NHMRC) of Australia (grant no. 1062227) and the National Heart Foundation of Australia. MI was supported by a Career Development Fellowship co-funded by the NHMRC and the National Heart Foundation of Australia (no. 1061435). GA was supported by an NHMRC Peter Doherty Early Career Fellowship (no. 1090462). We are grateful to the CARDIoGRAMplusC4D consortium for making their large-scale genetic data available. A list of members of the consortium and the

contributing studies is available at www.cardiogramplusc4d.org.

References

1. Akey, J.M., et al., *Interrogating a high-density SNP map for signatures of natural selection.* Genome Res, 2002. **12**(12): p. 1805-14.
2. Bustamante, C.D., et al., *Natural selection on protein-coding genes in the human genome.* Nature, 2005. **437**(7062): p. 1153-7.
3. Carlson, C.S., et al., *Genomic regions exhibiting positive selection identified from dense genotype data.* Genome Res, 2005. **15**(11): p. 1553-65.
4. Kelley, J.L., et al., *Genomic signatures of positive selection in humans and the limits of outlier approaches.* Genome Res, 2006. **16**(8): p. 980-9.
5. Lao, O., et al., *Signatures of positive selection in genes associated with human skin pigmentation as revealed from analyses of single nucleotide polymorphisms.* Annals of Human Genetics, 2007. **71**: p. 354-369.
6. Sabeti, P.C., et al., *Detecting recent positive selection in the human genome from haplotype structure.* Nature, 2002. **419**(6909): p. 832-7.
7. Shriver, M.D., et al., *The genomic distribution of population substructure in four populations using 8,525 autosomal SNPs.* Hum Genomics, 2004. **1**(4): p. 274-86.
8. Tang, K., K.R. Thornton, and M. Stoneking, *A new approach for using genome scans to detect recent positive selection in the human genome.* Plos Biology, 2007. **5**(7): p. 1587-1602.
9. Voight, B.F., et al., *A map of recent positive selection in the human genome.* PLoS Biol, 2006. **4**(3): p. e72.
10. Grossman, S.R., et al., *Identifying Recent Adaptations in Large-Scale Genomic Data.* Cell, 2013. **152**(4): p. 703-713.
11. Haasl, R.J. and B.A. Payseur, *Fifteen years of genomewide scans for selection: trends, lessons and unaddressed genetic sources of complication.* Mol Ecol, 2015.
12. Scheinfeldt, L.B. and S.A. Tishkoff, *Recent human adaptation: genomic approaches, interpretation and insights.* Nat Rev Genet, 2013. **14**(10): p. 692-702.
13. Pritchard, J.K. and A. Di Rienzo, *Adaptation - not by sweeps alone.* Nature Reviews Genetics, 2010. **11**(10): p. 665-667.
14. Pritchard, J.K., J.K. Pickrell, and G. Coop, *The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation.* Curr Biol, 2010. **20**(4): p. R208-15.
15. Kaplan, N.L., R.R. Hudson, and C.H. Langley, *The Hitchhiking Effect Revisited.* Genetics, 1989. **123**(4): p. 887-899.
16. Smith, J.M. and J. Haigh, *The hitch-hiking effect of a favourable gene.* Genetics Research, 2007. **89**(5-6): p. 391-403.
17. Oleksyk, T.K., M.W. Smith, and S.J. O'Brien, *Genome-wide scans for footprints of natural selection.* Philos Trans R Soc Lond B Biol Sci, 2010. **365**(1537): p. 185-205.
18. Sabeti, P.C., et al., *Positive natural selection in the human lineage.* Science, 2006. **312**(5780): p. 1614-20.
19. Wright, S., *Genetical structure of populations.* Nature, 1950. **166**(4215): p. 247-9.

20. Hamblin, M.T., E.E. Thompson, and A. Di Rienzo, *Complex signatures of natural selection at the Duffy blood group locus.* American Journal of Human Genetics, 2002. **70**(2): p. 369-383.
21. Beall, C.M., et al., *Natural selection on EPAS1 (HIF2 alpha) associated with low hemoglobin concentration in Tibetan highlanders.* Proceedings of the National Academy of Sciences of the United States of America, 2010. **107**(25): p. 11459-11464.
22. Lamason, R.L., et al., *SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans.* Science, 2005. **310**(5755): p. 1782-6.
23. Fraser, H.B., *Gene expression drives local adaptation in humans.* Genome Research, 2013. **23**(7): p. 1089-1096.
24. Akey, J.M., *Constructing genomic maps of positive selection in humans: where do we go from here?* Genome Res, 2009. **19**(5): p. 711-22.
25. Sabeti, P.C., et al., *Genome-wide detection and characterization of positive selection in human populations.* Nature, 2007. **449**(7164): p. 913-U12.
26. Teshima, K.M., G. Coop, and M. Przeworski, *How reliable are empirical genomic scans for selective sweeps?* Genome Research, 2006. **16**(6): p. 702-712.
27. Fu, W. and J.M. Akey, *Selection and adaptation in the human genome.* Annu Rev Genomics Hum Genet, 2013. **14**: p. 467-89.
28. Hernandez, R.D., et al., *Classic Selective Sweeps Were Rare in Recent Human Evolution.* Science, 2011. **331**(6019): p. 920-924.
29. Falconer, D.S. and T.F.C. Mackay, *Introduction to quantitative genetics*. 4th ed. 1996, Harlow, England ; New York: Prentice Hall. xv, 464 p.
30. Grant, P.R. and B.R. Grant, *Predicting Microevolutionary Responses to Directional Selection on Heritable Variation.* Evolution, 1995. **49**(2): p. 241-251.
31. Hermisson, J. and P.S. Pennings, *Soft sweeps: Molecular population genetics of adaptation from standing genetic variation.* Genetics, 2005. **169**(4): p. 2335-2352.
32. Messer, P.W. and D.A. Petrov, *Population genomics of rapid adaptation by soft selective sweeps.* Trends in Ecology & Evolution, 2013. **28**(11): p. 659-669.
33. Chevin, L.M. and F. Hospital, *Selective Sweep at a Quantitative Trait Locus in the Presence of Background Genetic Variation.* Genetics, 2008. **180**(3): p. 1645-1660.
34. Welter, D., et al., *The NHGRI GWAS Catalog, a curated resource of SNP-trait associations.* Nucleic Acids Research, 2014. **42**(D1): p. D1001-D1006.
35. Ding, K.Y. and I.J. Kullo, *Geographic differences in allele frequencies of susceptibility SNPs for cardiovascular disease.* Bmc Medical Genetics, 2011. **12**.
36. Raj, T., et al., *Common Risk Alleles for Inflammatory Diseases Are Targets of Recent Positive Selection.* American Journal of Human Genetics, 2013. **92**(4): p. 517-529.
37. Casto, A.M. and M.W. Feldman, *Genome-Wide Association Study SNPs in the Human Genome Diversity Project Populations: Does Selection Affect Unlinked SNPs with Shared Trait Associations?* Plos Genetics, 2011. **7**(1).
38. Turchin, M.C., et al., *Evidence of widespread selection on standing variation in Europe at height-associated SNPs.* Nature Genetics, 2012. **44**(9): p. 1015-+.
39. Go, A.S., et al., *Heart disease and stroke statistics--2014 update: a report from the American Heart Association.* Circulation, 2014. **129**(3): p. e28-e292.

40. Deloukas, P., et al., *Large-scale association analysis identifies new risk loci for coronary artery disease.* Nature Genetics, 2013. **45**(1): p. 25-U52.
41. Nikpay, M., et al., *A comprehensive 1000 Genomes-based genome-wide association meta-analysis of coronary artery disease.* Nature Genetics, 2015. **47**(10): p. 1121-+.
42. Byars, S.G., et al., *Colloquium papers: Natural selection in a contemporary human population.* Proc Natl Acad Sci U S A, 2010. **107 Suppl 1**: p. 1787-92.
43. Williams, G.C., *Pleiotropy, Natural Selection, and the Evolution of Senescence.* Evolution, 1957. **11**(4): p. 398-411.
44. Wang, X., S.G. Byars, and S.C. Stearns, *Genetic links between post-reproductive lifespan and family size in Framingham.* Evol Med Public Health, 2013. **2013**(1): p. 241-53.
45. Jalowiec, D.A. and J.A. Hill, *Myocardial infarction in the young and in women.* Cardiovasc Clin, 1989. **20**(1): p. 197-206.
46. Rubin, J.B. and W.B. Borden, *Coronary Heart Disease in Young Adults.* Current Atherosclerosis Reports, 2012. **14**(2): p. 140-149.
47. Tuzcu, E.M., et al., *High prevalence of coronary atherosclerosis in asymptomatic teenagers and young adults: evidence from intravascular ultrasound.* Circulation, 2001. **103**(22): p. 2705-10.
48. Morillas, P., et al., *Characteristics and outcome of acute myocardial infarction in young patients - The PRIAMHO II study.* Cardiology, 2007. **107**(4): p. 217-225.
49. Allam, A.H., et al., *Atherosclerosis in ancient Egyptian mummies: the Horus study.* JACC Cardiovasc Imaging, 2011. **4**(4): p. 315-27.
50. Wollstein, A. and W. Stephan, *Inferring positive selection in humans from genomic data.* Investig Genet, 2015. **6**: p. 5.
51. Kuleshov, M.V., et al., *Enrichr: a comprehensive gene set enrichment analysis web server 2016 update.* Nucleic Acids Res, 2016. **44**(W1): p. W90-7.
52. Kudaravalli, S., et al., *Gene Expression Levels Are a Target of Recent Natural Selection in the Human Genome.* Molecular Biology and Evolution, 2009. **26**(3): p. 649-658.
53. Williamson, S.H., et al., *Localizing recent adaptive evolution in the human genome.* Plos Genetics, 2007. **3**(6): p. 901-915.
54. Kullo, I.J. and K.Y. Ding, *Patterns of population differentiation of candidate genes for cardiovascular disease.* Bmc Genetics, 2007. **8**.
55. Jarvis, J.P., et al., *Patterns of Ancestry, Signatures of Natural Selection, and Genetic Association with Stature in Western African Pygmies.* Plos Genetics, 2012. **8**(4): p. 299-313.
56. Comuzzie, A.G., et al., *Novel Genetic Loci Identified for the Pathophysiology of Childhood Obesity in the Hispanic Population.* Plos One, 2012. **7**(12).
57. Haritunians, T., et al., *Genetic Predictors of Medically Refractory Ulcerative Colitis.* Inflammatory Bowel Diseases, 2010. **16**(11): p. 1830-1840.
58. Lang, B., et al., *HDAC9 is implicated in schizophrenia and expressed specifically in post-mitotic neurons but not in adult neural stem cells.* Am J Stem Cells, 2012. **1**(1): p. 31-41.
59. Pokharel, K., et al., *Transcriptome profiling of Finnsheep ovaries during out-of-season breeding period.* Agricultural and Food Science, 2015. **24**: p. 1-9.

60. Mbarek, H., et al., *Identification of Common Genetic Variants Influencing Spontaneous Dizygotic Twinning and Female Fertility.* Am J Hum Genet, 2016. **98**(5): p. 898-908.
61. Huang, C., et al., *Efficient SNP Discovery by Combining Microarray and Lab-on-a-Chip Data for Animal Breeding and Selection.* Microarrays, 2015. **4**(4): p. 570-595.
62. Mote, B.E., et al., *Identification of genetic markers for productive life in commercial sows.* J Anim Sci, 2009. **87**(7): p. 2187-95.
63. Balgir, R.S., *Menarcheal age in relation to ABO blood group phenotypes and haemoglobin-E genotypes.* J Assoc Physicians India, 1993. **41**(4): p. 210-1.
64. Pyun, J.A., et al., *Genome-wide association studies and epistasis analyses of candidate genes related to age at menarche and age at natural menopause in a Korean population.* Menopause, 2014. **21**(5): p. 522-529.
65. Spencer, K.L., et al., *Genetic Variation and Reproductive Timing: African American Women from the Population Architecture Using Genomics and Epidemiology (PAGE) Study.* Plos One, 2013. **8**(2).
66. Rempel, L.A., et al., *Association analyses of candidate single nucleotide polymorphisms on reproductive traits in swine.* J Anim Sci, 2010. **88**(1): p. 1-15.
67. Patel, O.V., et al., *Homeorhetic adaptation to lactation: comparative transcriptome analysis of mammary, liver, and adipose tissue during the transition from pregnancy to lactation in rats.* Functional & Integrative Genomics, 2011. **11**(1): p. 193-202.
68. Wang, M., et al., *MicroRNA expression patterns in the bovine mammary gland are affected by stage of lactation.* J Dairy Sci, 2012. **95**(11): p. 6529-35.
69. Colodro-Conde, L., et al., *A twin study of breastfeeding with a preliminary genome-wide association scan.* Twin Res Hum Genet, 2015. **18**(1): p. 61-72.
70. McLean, M.P., Z. Zhao, and G.C. Ness, *Reduced hepatic LDL-receptor, 3-hydroxy-3-methylglutaryl coenzyme A reductase and sterol carrier protein-2 expression is associated with pregnancy loss in the diabetic rat.* Endocrine, 1995. **3**(10): p. 695-703.
71. Kuo, D.S., C. Labelle-Dumais, and D.B. Gould, *COL4A1 and COL4A2 mutations and disease: insights into pathogenic mechanisms and potential therapeutic targets.* Hum Mol Genet, 2012. **21**(R1): p. R97-110.
72. Lin, S., et al., *Pre-eclampsia has an adverse impact on maternal and fetal health.* Transl Res, 2015. **165**(4): p. 449-63.
73. Sayed, A.A.A., *Molecular genetic studies in pregnancies affected by preeclampsia and intrauterine growth restriction*. 2011, University of Nottingham.
74. Fritz, R.B., *Trophoblast Retrieval And Isolation From e Cervix (tric) For Non-Invasive Prenatal Genetic Diagnosis And Prediction Of Abnormal Pregnancy Outcome*. 2015, Wayne State University.
75. Tabano, S., et al., *Placental LPL gene expression is increased in severe intrauterine growth-restricted pregnancies.* Pediatr Res, 2006. **59**(2): p. 250-3.
76. Bhasin, K.K., et al., *Maternal low-protein diet or hypercholesterolemia reduces circulating essential amino acids and leads to intrauterine growth restriction.* Diabetes, 2009. **58**(3): p. 559-66.

77. Kakourou, G., et al., *Investigation of gene expression profiles before and after embryonic genome activation and assessment of functional pathways at the human metaphase II oocyte and blastocyst stage.* Fertility and Sterility, 2013. **99**(3): p. 803-+.
78. Siva, K., et al., *Human BCAS3 Expression in Embryonic Stem Cells and Vascular Precursors Suggests a Role in Human Embryogenesis and Tumor Angiogenesis.* Plos One, 2007. **2**(11).
79. Liu, J., et al., *Expression of SWAP-70 in the uterus and feto-maternal interface during embryonic implantation and pregnancy in the rhesus monkey (Macaca mulatta).* Histochem Cell Biol, 2006. **126**(6): p. 695-704.
80. Zhou, L., et al., *Local injury to the endometrium in controlled ovarian hyperstimulation cycles improves implantation rates.* Fertil Steril, 2008. **89**(5): p. 1166-76.
81. Solca, C., G.S. Tint, and S.B. Patel, *Dietary xenosterols lead to infertility and loss of abdominal adipose tissue in sterolin-deficient mice.* J Lipid Res, 2013. **54**(2): p. 397-409.
82. Moretti, E., et al., *Ultrastructural study of spermatogenesis in KSR2 deficient mice.* Transgenic Res, 2015. **24**(4): p. 741-51.
83. Dokras, A., *Cardiovascular disease risk in women with PCOS.* Steroids, 2013. **78**(8): p. 773-6.
84. Kenigsberg, S., et al., *Gene expression microarray profiles of cumulus cells in lean and overweight-obese polycystic ovary syndrome patients.* Mol Hum Reprod, 2009. **15**(2): p. 89-103.
85. Manneras-Holm, L., A. Benrick, and E. Stener-Victorin, *Gene expression in subcutaneous adipose tissue differs in women with polycystic ovary syndrome and controls matched pair-wise for age, body weight, and body mass index.* Adipocyte, 2014. **3**(3): p. 190-6.
86. Salilew-Wondim, D., et al., *Polycystic ovarian syndrome is accompanied by repression of gene signatures associated with biosynthesis and metabolism of steroids, cholesterol and lipids.* Journal of Ovarian Research, 2015. **8**.
87. Scotti, L., et al., *Platelet-derived growth factor BB and DD and angiopoietin1 are altered in follicular fluid from polycystic ovary syndrome patients.* Mol Reprod Dev, 2014. **81**(8): p. 748-56.
88. Yan, L., et al., *Expression of apoptosis-related genes in the endometrium of polycystic ovary syndrome patients during the window of implantation.* Gene, 2012. **506**(2): p. 350-4.
89. Kilic, M., et al., *Identification of Mutations and Evaluation of Cardiomyopathy in Turkish Patients with Primary Carnitine Deficiency.* Jimd Reports - Case and Research Reports, 2011/3, 2012. **3**: p. 17-23.
90. Tamai, I., *Pharmacological and pathophysiological roles of carnitine/organic cation transporters (OCTNs: SLC22A4, SLC22A5 and Slc22a21).* Biopharmaceutics & Drug Disposition, 2013. **34**(1): p. 29-44.
91. Maqdasy, S., et al., *Cholesterol and male fertility: What about orphans and adopted?* Molecular and Cellular Endocrinology, 2013. **368**(1-2): p. 30-46.
92. Schisterman, E.F., et al., *Lipid concentrations and semen quality: the LIFE study.* Andrology, 2014. **2**(3): p. 408-15.

93. Carter, A.J. and A.Q. Nguyen, *Antagonistic pleiotropy as a widespread mechanism for the maintenance of polymorphic disease alleles.* BMC Med Genet, 2011. **12**: p. 160.
94. Granka, J.M., et al., *Limited evidence for classic selective sweeps in African populations.* Genetics, 2012. **192**(3): p. 1049-64.
95. International HapMap, C., et al., *A second generation human haplotype map of over 3.1 million SNPs.* Nature, 2007. **449**(7164): p. 851-61.
96. Surakka, I., et al., *Founder population-specific HapMap panel increases power in GWA studies through improved imputation accuracy and CNV tagging.* Genome Research, 2010. **20**(10): p. 1344-1351.
97. Abraham, G. and M. Inouye, *Fast principal component analysis of large-scale genome-wide data.* PLoS One, 2014. **9**(4): p. e93766.
98. Chang, C.C., et al., *Second-generation PLINK: rising to the challenge of larger and richer datasets.* Gigascience, 2015. **4**: p. 7.
99. O'Connell, J., et al., *A general approach for haplotype phasing across the full spectrum of relatedness.* PLoS Genet, 2014. **10**(4): p. e1004234.
100. Howie, B.N., P. Donnelly, and J. Marchini, *A flexible and accurate genotype imputation method for the next generation of genome-wide association studies.* PLoS Genet, 2009. **5**(6): p. e1000529.
101. Paten, B., et al., *Genome-wide nucleotide-level mammalian ancestor reconstruction.* Genome Res, 2008. **18**(11): p. 1829-43.
102. Flicek, P., et al., *Ensembl 2012.* Nucleic Acids Res, 2012. **40**(Database issue): p. D84-90.
103. Gautier, M. and R. Vitalis, *rehh: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure.* Bioinformatics, 2012. **28**(8): p. 1176-7.
104. Stranger, B.E., et al., *Patterns of Cis Regulatory Variation in Diverse Human Populations.* Plos Genetics, 2012. **8**(4): p. 272-284.
105. Stearns, S.C., *The evolution of life histories*. 1992, Oxford ; New York: Oxford University Press. xii, 249 p.
106. Abraham, G., et al., *Genomic prediction of coronary heart disease.* bioRxiv, 2016.

Figure legends

Figure 1. Association of coronary artery disease (CAD) genetic risk and positive signatures of selection in 12 worldwide populations. The 40 of 76 CAD genes investigated are shown that have at least four significant selection-risk associations in Panel B across all 12 populations. **Panel A.** Magnitude and significance of largest positive selection signal (integrated haplotype score, iHS) within each gene-population combination. P values (circles within squares) were obtained from 10000 permutations. Bonferroni corrected p-value limit also shown (α=0.05/76=0.000657) with closed circles. **Panel B.** Null hypothesis: no association between CAD genetic risk and positive selection, tested using mixed effects model with SNP estimates of CAD log odds genetic risk and iHS while accounting for gene LD structure as a random effect (first eigenvector from LD matrix per gene). Scaled regression coefficients were obtained directly from regressions, each p value from 10000 permutations. **Panel C.** Null hypothesis: association between genetic risk and positive selection for SNPs within CAD genes no different than non-CAD associated genes. Permuted p values were estimated by comparing each p value in Panel B against 100 nominal p values obtained by randomly choosing (without replacement) 100 non-CAD associated genes of similar size across the genome and using the same mixed effects model setup as described above. **Populations**. Grouped by ancestry, African (ASW, African ancestry in Southwest USA; MKK, Maasai in Kinyawa, Kenya; YRI, Yoruba from Ibadan, Nigeria; LWK, Luhya in Webuye, Kenya), East-Asian (CHB, Han Chinese subjects from Beijing; CHD, Chinese in Metropolitan Denver, Colorado; JPT, Japanese subjects from Tokyo), European (CEU, Utah residents with ancestry from northern and western Europe from the CEPH collection; TSI, Tuscans in Italy; FIN, Finnish in Finland), GIH (Gujarati Indians in Houston, TX, USA), MEX (Mexican ancestry in Los Angeles, CA, USA).

Figure 2. Quantitative links between coronary artery disease risk and selection signals in *BCAS3*. A. Correlation between selection signals (iHS) and coronary artery disease (CAD) log odds genetic risk (log odds, ln(OR)), both represented as absolute values. Red line/upper right value, β from mixed effects regression. **B.** Base pair positional comparison of selection signals and CAD genetic risk across *BCAS3*. Blue points, CAD log odds values; grey-orange or non-significant-significant points, iHS scores. Horizontal bar shows *BCAS3* gene (and intron) span and location of lead index SNP. Blue/orange lines are smoothed lines estimated with loess function in R. **C.** LD plots, r^2. Populations: CEU, Utah residents with ancestry from northern and western Europe from the CEPH collection; YRI, Yoruba from Ibadan, Nigeria.

Figure 3. Quantitative links between coronary artery disease risk and selection signals in *PHACTR1*. A. Correlation between selection signals (iHS) and coronary artery disease (CAD) log odds genetic risk (ln(OR)), both represented as absolute values. Red line/upper right value, β from mixed effects regression. **B.** Base pair positional comparison of selection signals and CAD genetic risk across *PHACTR1*. Blue points, CAD log odds values; grey-orange or non-significant-significant points, iHS scores. Horizontal bar shows *PHACTR1* gene (and intron) spans and location of index SNP if present. **C.** LD plots, r^2. Populations: CEU, Utah residents with ancestry from northern and western Europe from the CEPH collection; GIH, Gujarati Indians in Houston, TX, USA.

Figure 4: Comparing positive selection with gene regulation. Summary distribution of permuted eQTL p values for SNPs with (left) or without (right) a significant selection signal. SNPs with a significant selection signal (iHS) were chosen by taking the largest significant positive selection signal (if one was present) within each gene-population combination. The same number of SNPs without a significant selection signal were also randomly drawn across all gene-

population combinations for comparison. These SNPs were used in an eQTL analysis where they were regressed (including gender as a covariate) against their associated gene probe's expression.

Figure 5. Conceptual figure of potential evolutionary tradeoffs between coronary artery disease (CAD) burden and other phenotypes as a consequence of antagonistic pleiotropy (AP) [43]. As a simple example, AP describes gene effect on two traits (pleiotropy) that oppositely (antagonistic) affect individual fitness at different ages. Selection on that gene conferring a fitness advantage and disadvantage at different ages depends on the size and timing of the effects. An advantage during the ages with the highest probability of reproduction (between~20-45 years of age in humans) would increase fitness (lifetime reproductive success) more than a similarly sized disadvantage at later ages would decrease it. This concept is part of the well-known evolutionary theory of ageing, which describes tradeoffs in energy invested into growth, reproduction and survival [105]. In the figure above, intense natural selection occurring on CAD loci as a result of fitness advantages (+ signs, red text callout box 1.) conferred by genetically correlated risk factors ('CAD risk factors' box) or early-life traits ('early-life traits' box) trades off with the deleterious effects of these genes on fitness (i.e. CAD burden) later in life (- sign, red text callout box 2.) where the intensity of selection is weak. This occurs because of the negative relationship between genetic effects on early vs late-life traits (- sign, red text callout box 3.), which could help explain the high prevalence and maintenance of CAD in modern human populations. Over shorter timescales, lifetime probability of CAD is modified by a combination of genetic and environmental risk factors (e.g. [106]). There is a good evidence that such antagonistic effects have operated on CAD loci given: significant associations between CAD genetic risk and selection we found (Fig 1-2); CAD genes also underlie many early-life traits known to modify fitness (Table S2); phenotypic selection has been found operating on CAD phenotypic risk factors [42].

Figure 1

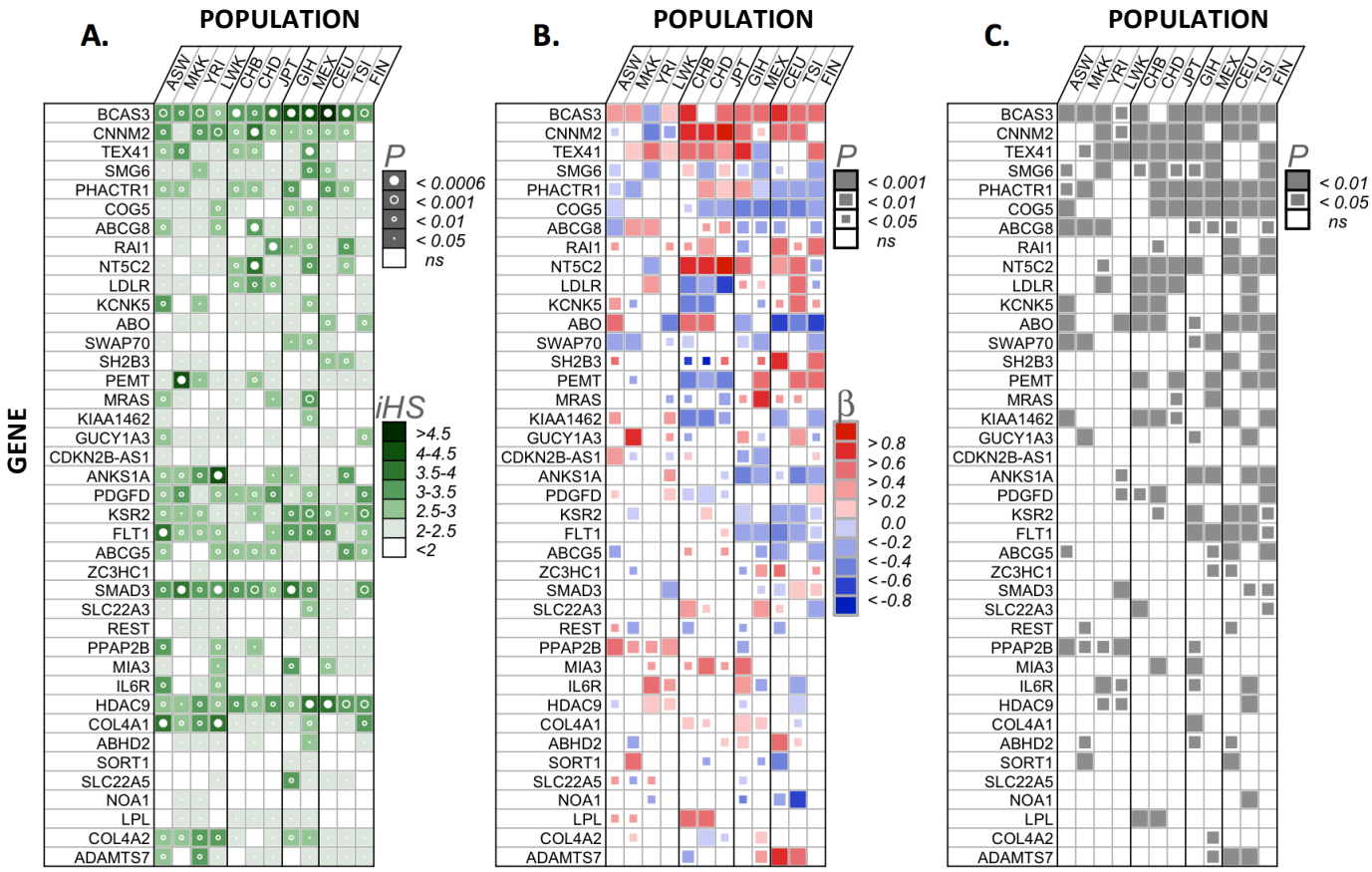


Figure 2

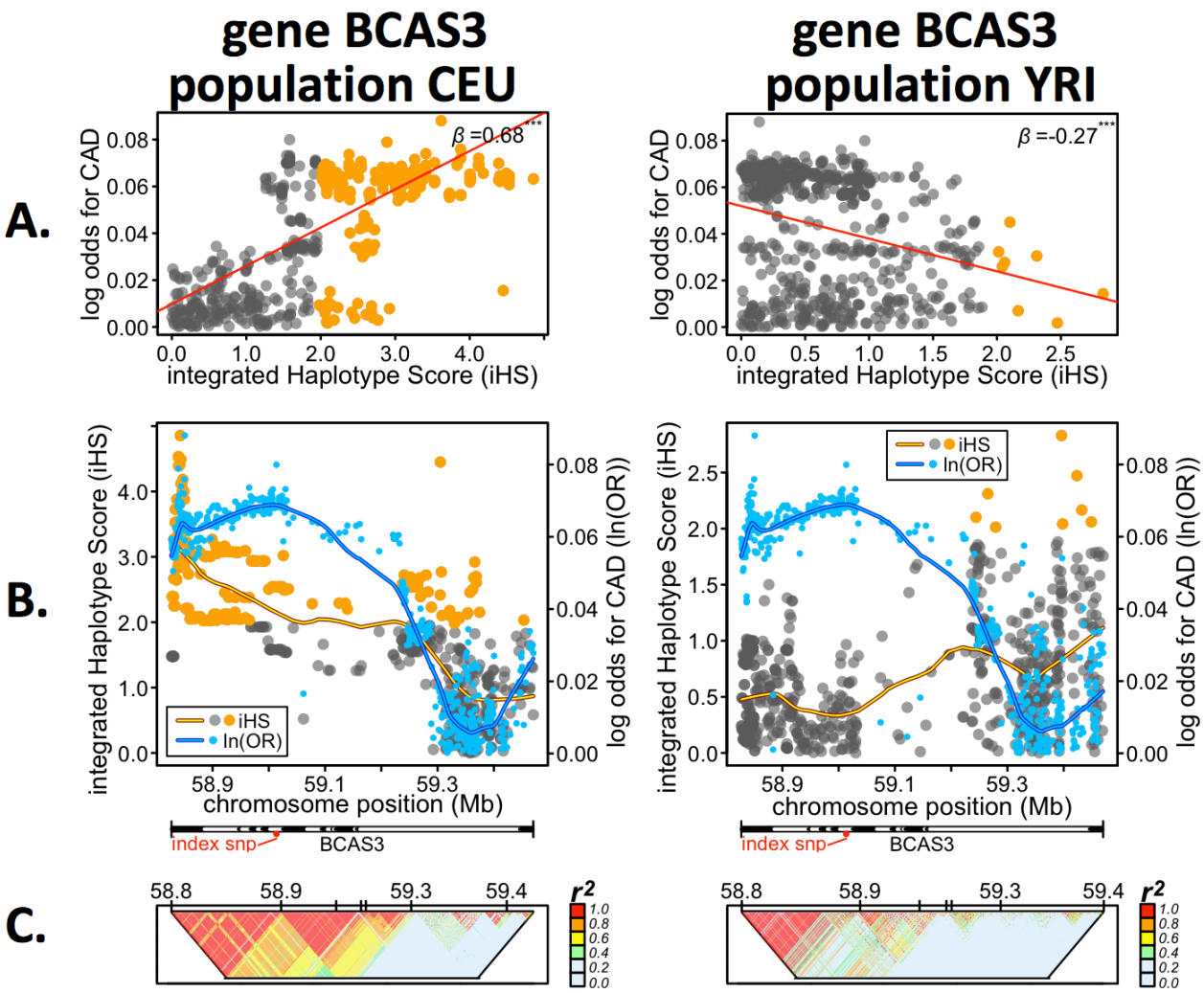


Figure 3

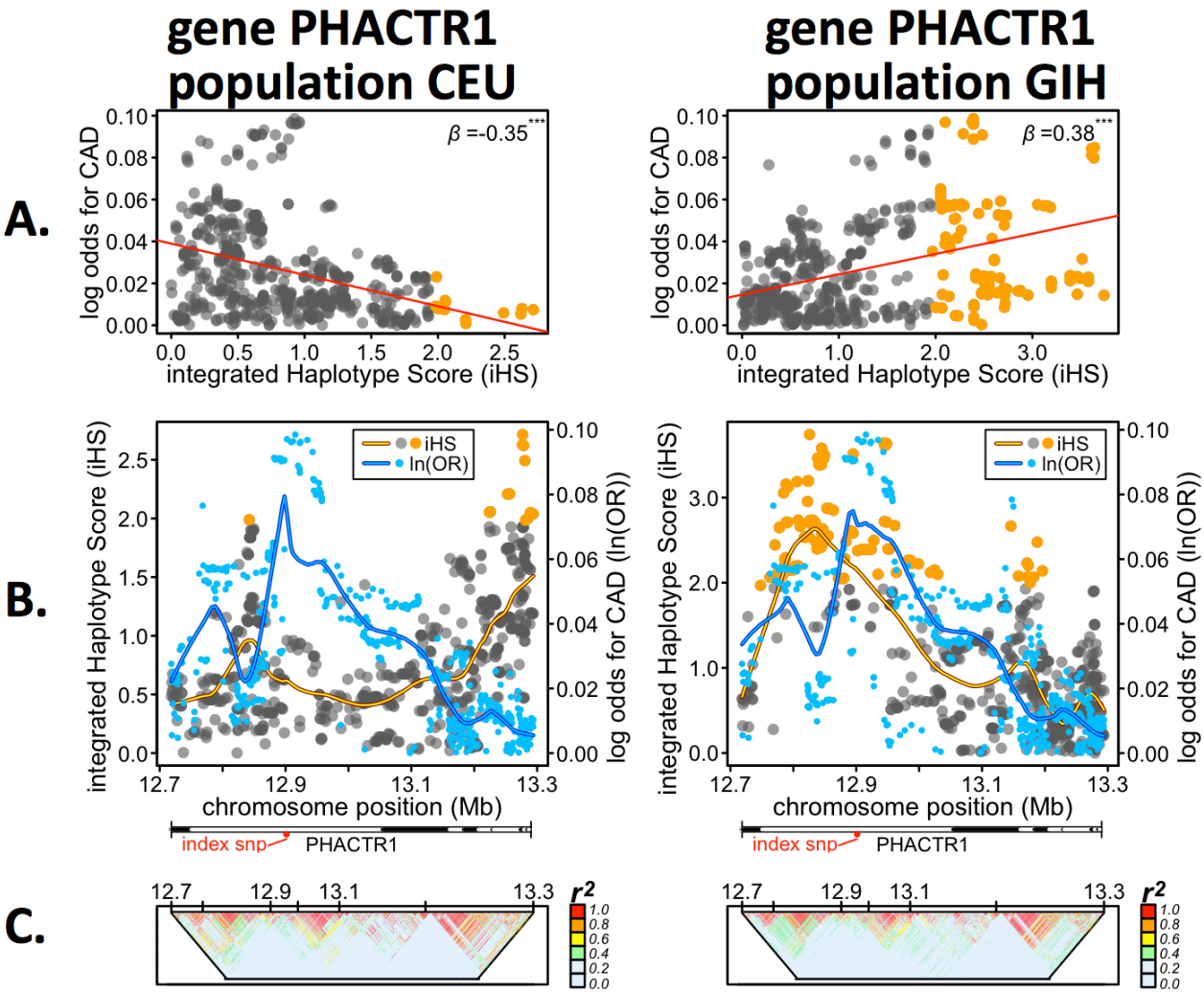


Figure 4

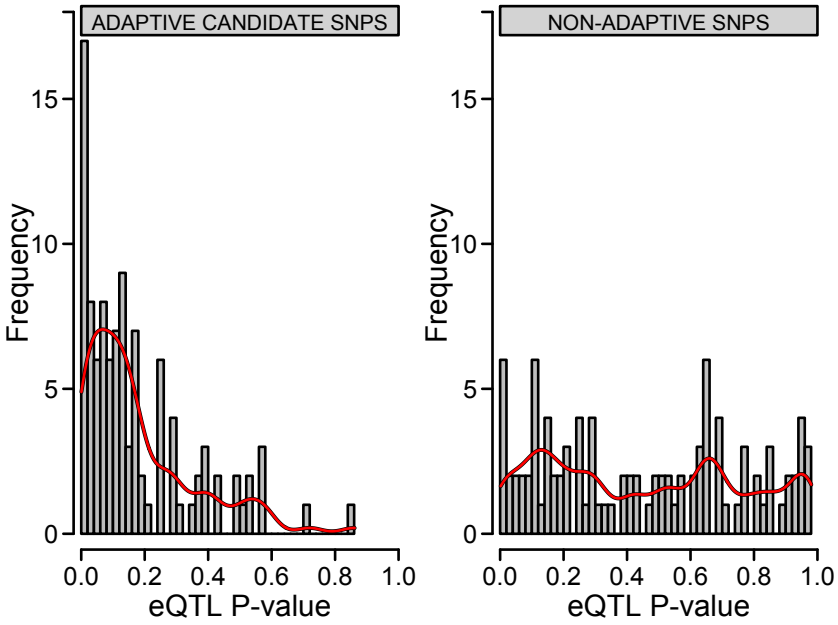
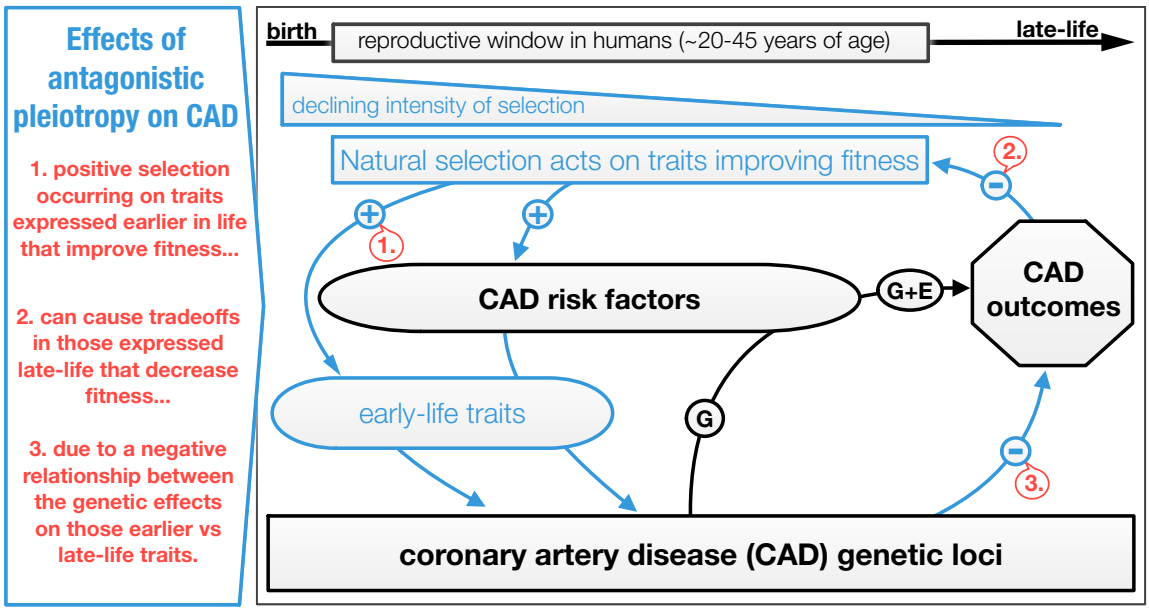


Figure 5



Supplementary Discussion

Widespread signals of positive selection on CAD loci

Evidence of candidate positive selection signals for CAD loci were widespread, with many genes having significant iHS scores of small-medium size (i.e. iHS score range: 2-3) with four genes (*BCAS3*, *ANKS1A*, *CXCL12*, *PEMT*) harboring large selection signals (iHS >4), two of which had previously been identified as having strong selection signals including *BCAS3* (Breast Carcinoma Amplified Sequence 3) in the HapMap3 CEU population [1] and *PHACTR1* (phosphatase and actin regulator 1) across the ASW, CEU and CHB/CHD HapMap3 populations [2]. Twelve genes contained SNPs with selection scores that remained significant after correction for multiple testing (Fig. 1A). The consistency of smaller, less significant selection signals for several genes within most populations (i.e. *CNNM2*, *PHACTR1*, *PDGFD*) strongly suggest that these may be smaller and possibly valid incomplete selective sweeps that are typically missed due to stringency of multiple-correction thresholds and lack of validation across multiple populations.

These patterns match expectations from the polygenic model of selection that predicts that selection on complex traits mostly involves smaller shifts in many underlying loci; it is the likely reason why so few large selection signals have been found underlying complex traits in general [3, 4] and those underlying cardiovascular disease phenotypes in particular [5, 6]. For example, Kullo & Ding 2007 [6] found that 110 out of 364 genes in pathways associated with cardiovascular disease (i.e. inflammation, insulin, p53, Ras, cholesterol biosynthesis etc) had significantly higher Fst (empirical $P<0.05$) in at least one SNP between 4 populations, but none remained significant after correction for multiple testing. In a later study, Ding & Kullo 2011 [5] found that 8 out of 158 genome-wide significant SNPs in genes for 36 cardiovascular disease phenotypes and related traits (CHD, hypertension, stroke, BMI, lipids etc) had significantly elevated Fst between 52 populations in the Human Genome Diversity Project.

It is difficult to compare selection candidates we found in the 76 CAD associated genes with results from these two previous studies as full sets of gene lists and Fst estimates were not available for either, and they used loci underlying much broader cardiovascular disease phenotypes than our more current list of specific CAD loci [7]. Nevertheless, due to fine-scale imputation with the 1000 Genomes Panel, our study suggests that many more loci related to cardiovascular disease have been recently modified by natural selection than previously identified. The larger sample of SNPs also likely improved reliability of iHS p values, with many more estimates available per MAF bin used to standardize iHS measures [8].

The Fst measures used in the Ding and Kullo studies also differ qualitatively from the iHS scores we used. Fst captures allele frequency differences between populations and is less sensitive to detecting alleles that have undergone recent selection [9], while the iHS statistic detects whether common alleles are carried on unusually long haplotypes within populations and should be better at capturing more recent smaller selection signals [8]. Lastly, by considering not just genome-wide significant index SNPs, we were able to detect smaller selection signals within CAD loci that were consistent across populations and would have otherwise been missed. *PHACTR1* is a good example of this – several smaller candidate selection signals were found (iHS ranging from 2-3.8) where peak selection signals did not span the index SNP location - sometimes signals were in different introns within the same locus (Fig. S2).

Supplementary Discussion

References

1. Sabeti, P.C., et al., *Genome-wide detection and characterization of positive selection in human populations.* Nature, 2007. **449**(7164): p. 913-U12.
2. Williamson, S.H., et al., *Localizing recent adaptive evolution in the human genome.* Plos Genetics, 2007. **3**(6): p. 901-915.
3. Fu, W. and J.M. Akey, *Selection and adaptation in the human genome.* Annu Rev Genomics Hum Genet, 2013. **14**: p. 467-89.
4. Hernandez, R.D., et al., *Classic Selective Sweeps Were Rare in Recent Human Evolution.* Science, 2011. **331**(6019): p. 920-924.
5. Ding, K.Y. and I.J. Kullo, *Geographic differences in allele frequencies of susceptibility SNPs for cardiovascular disease.* Bmc Medical Genetics, 2011. **12**.
6. Kullo, I.J. and K.Y. Ding, *Patterns of population differentiation of candidate genes for cardiovascular disease.* Bmc Genetics, 2007. **8**.
7. Nikpay, M., et al., *A comprehensive 1000 Genomes-based genome-wide association meta-analysis of coronary artery disease.* Nature Genetics, 2015. **47**(10): p. 1121-+.
8. Voight, B.F., et al., *A map of recent positive selection in the human genome.* PLoS Biol, 2006. **4**(3): p. e72.
9. Sabeti, P.C., et al., *Positive natural selection in the human lineage.* Science, 2006. **312**(5780): p. 1614-20.

Supplemental Information

Table S1. Literature search results - potential pleiotropic links between CAD risk genes and fitness-related phenotypes

Gene Rank*	CAD Gene	Full name	Species	Potential fitness effects/links	Potential fitness class**	Causative factor/model	Reference
1	BCAS3	Breast Carcinoma Amplified Seque	Human/M	embryogenesis	female potential fertility	highly expressed in developing oocytes	Siva 2007 - Human BCAS3 Expression in Embryonic Stem Cells and Vascular Precursors Suggests a Role in Human Embryogenesis and Tumor Angiogenesis
	BCAS3	Breast Carcinoma Amplified Seque	Mouse	embryogenesis	female potential fertility	BCAS3 upregulated in developmentally incompetent mouse c	Zucotti 2008 - Maternal Oct-4 is a potential key regulator of the developmental competence of mouse oocytes
2	CNNM2	Cyclin And CBS Domain Divalent M	Human	pregnancy-related blood pressure	pregnancy outcomes	CNNM2 expression	Nestler 2014 - CNNM2 Cyclin And CBS Domain Divalent Metal Cation Transport Mediator 2
	CNNM2	Cyclin And CBS Domain Divalent M	Mouse	pregnancy complications, hypoxia	pregnancy outcomes	CNNM2 down-regulated (-2.5 fold change) during pregnancy	Gheorghe 2007 - Gene expression patterns in the hypoxic murine placenta: A role in epigenesis?
3	TEX41	Testis Expressed 41 (Non- Protein	Human	fetal IUGR, developmental delays	pregnancy outcomes	triplication involving TEX41	Yuan 2015 - A de novo triplication on 2q22.3 including the entire ZEB2 gene associated with global developmental delay, multiple congenital anomalies and behavioral abnormalities), Molecular Cytogenetics
4	SMG6	Nonsense Mediated MRNA Decay	Mouse	altered embryogenesis	female potential fertility	knockout or siRNA-mediated knockdown studies	Bao 2015 - UPF2, a nonsense-mediated mRNA decay factor, is required for prepubertal Sertoli cell development and male fertility by ensuring fidelity of the transcriptome
5	PHACTR1	Phosphatase And Actin Regulator :	Human	reproductive timing	reproductive outcomes	PHACTR1 variation	Spencer 2013 - Genetic Variation and Reproductive Timing: African American Women from the Population Architecture Using Genomics and Epidemiology (PAGE) Study
	PHACTR1	Phosphatase And Actin Regulator :	Human	oocyte function	female potential fertility	PHACTR1 highly expressed	Kakourou 2013 - Investigation of gene expression profiles before and after embryonic genome activation and assessment of functional pathways at the human metaphase IIoocyte and blastocyst stage
	PHACTR1	Phosphatase And Actin Regulator :	Human	placental inflammatory responses	pregnancy outcomes	PHACTR1 down-regulated	Saben 2013 - Early growth response protein-1 mediates lipotoxicity-associated placental inflammation: role in maternal obesity
	PHACTR1	Phosphatase And Actin Regulator :	Human	endomtrium implantation receptivity	female potential fertility	PHACTR1 8-fold upregulated	Zhou 2008 - Local injury to the endometrium in controlled ovarian hyperstimulation cycles improves implantation rates
	PHACTR1	Phosphatase And Actin Regulator :	Mouse	uterus functioning	female potential fertility	PHACTR1 1.4-1.9 fold change	Newbold 2007 - Developmental exposure to diethylstilbestrol alters uterine gene expression that may be associated with uterine neoplasia later in life
	PHACTR1	Phosphatase And Actin Regulator :	Rat	lactation	reproductive outcomes	PHACTR1 expression (4.7 fold change) in mammary tissues	Patel 2011 - Homeorhetic adaptation to lactation: comparative transcriptome analysis of mammary, liver, and adipose tissue during the transition from pregnancy to lactation in rats
6	COG5	Component of Oligomeric Golgi Co	Drosophila	spermatogenesis	male potential fertility	COG5 expression	Farkas 2003 - The Drosophila Cog5 homologue is required for cytokinesis, cell elongation, and assembly of specialized golgi architecture during spermatogenesis
	COG5	Component of Oligomeric Golgi Co	Human	intrauterine growth	pregnancy outcomes	COG5 expression	Foulquier 2009 - COG defects, birth and rise!
7	ABCG8	ATP-Binding Cassette, Sub-Family	Mouse	infertility	female potential fertility	Knockout mice deficient Abcg8	Solca 2013 - Dietary xenosterols lead to infertility and loss of abdominal adipose tissue in sterolin-deficient mice
	ABCG8	ATP-Binding Cassette, Sub-Family	Human	fetal distress, asphyxial events, intrauterine death	pregnancy outcomes	intrahepatic cholestasis of pregnancy (ICP), enterohepatic cir	Dixon 2016 - The pathophysiology of intrahepatic cholestasis of pregnancy
8	RAI1	Retinoic Acid Induced 1	Human/M	Growth retardation, embryonic-postnatal development	pregnancy outcomes	knock-out mouse model for Smith-Magenis syndrome	Bi 2005 - Inactivation of Rai1 in mice recapitulates phenotypes observed in chromosome engineered mouse models for Smith–Magenis syndrome
	RAI1	Retinoic Acid Induced 1	Mouse	Growth retardation, impaired motor and sensory coordin	pregnancy outcomes	Transgenic mice over-expressing RAI1	Girirajan 2008 in Girirajan 2009 - Abnormal maternal behaviour, altered sociability, and impaired serotonin metabolism in Rai1-transgenic mice
9	NTSC2	Nucleotidase, Cytosolic II	Human	Female reproduction	female potential fertility	NTSC2 is over expressed in fallopian tube, uterine endometri	Shen 2013 - Estradiol Regulation of Nucleotidases in Female Reproductive Tract Epithelial Cells and Fibroblasts
	NTSC2	Nucleotidase, Cytosolic II	Human	fetal growth, birthweight, postnatal growth & metabolism	pregnancy outcomes	fetal genotype	Horikoshi 2013 - New loci associated with birth weight identify genetic links between intrauterine growth and adult height and metabolism
10	LDLR	Low Density Lipoprotein Receptor	Human/M	IUGR in offspring of LDLR-/- mice. Childhood obesity.	pregnancy outcomes	LDLR mouse model for IUGR/modification	Bhasin 2009 - Maternal Low-Protein Diet or Hypercholesterolemia Reduces Circulating Essential Amino Acids and Leads to Intrauterine Growth Restriction
	LDLR	Low Density Lipoprotein Receptor	Human	Placental regulation of cholesterol	pregnancy outcomes	Maternal lipid profile affecting placental protein expression c	Chiasson 2007 - Influence of maternal lipid profile on placental protein expression of LDLr and SR-BI
	LDLR	Low Density Lipoprotein Receptor	Mouse	Placental regulation of cholesterol	pregnancy outcomes	Maternal-fetal transfer of lipids	Overbrgh 1995 - Expression of mouse alpha-macroglobulins, lipoprotein rece r-related protein, LDL receptor, apolipoprotein E, and lipoprotein lipase in pregnancy
	LDLR	Low Density Lipoprotein Receptor	Rat	Pregnancy loss	fetal/offspring mortality	LDLR rat model for diabetes	McLean 1995 - Reduced hepatic LDL-receptor, 3-hydroxy-3-methylglutaryl coenzyme A reductase and sterol carrier protein-2 expression is associated with pregnancy loss in the diabetic rat
11	KCNK5	Potassium Channel, Two Pore Dor	Human	Fertility - sperm volume	male potential fertility	KCNK5 protein and mRNA levels	Yeung 2008 - Potassium channels involved in human sperm volume regulation—quantitative studies at the protein and mRNA levels
	KCNK5	Potassium Channel, Two Pore Dor	Human	Male infertility	male potential fertility	Sperm inability to fertilize egg	Cooper 2007 - Involvement of Potassium and Chloride Channels and Other Transporters in Volume Regulation by Spermatozoa
	KCNK5	Potassium Channel, Two Pore Dor	Mouse	Male infertility	male potential fertility	Sperm volume	Barfield 2005 - Characterization of potassium channels involved in volume regulation of human spermatozoa
	KCNK5	Potassium Channel, Two Pore Dor	Primate	Male fertility	male potential fertility	Sperm function	Chow 2006 - Expression of two-pore domain potassium channels in nonhuman primate sperm
	KCNK5	Potassium Channel, Two Pore Dor	Mouse	Female fertility	female potential fertility	Oocyte survival/viability	Kang 2015 - TASK-2 Expression Levels are Increased in Mouse Cryopreserved Ovarie
	KCNK5	Potassium Channel, Two Pore Dor	Cow	Lactation	reproductive outcomes	KCNK5 expression	Wang 2012 - MicroRNA expression patterns in the bovine mammary gland are affected by stage of lactation
12	ABO	ABO Blood Group (Transferase A, /	Human	Birth weight, maternal age at child-bearing	pregnancy outcomes	ABO variation	Gloria-Bottini 2011 - Effect of smoking and ABO blood groups on maternal age at child bearing and on birth weight.
	ABO	ABO Blood Group (Transferase A, /	Human	Fetal growth restriction	pregnancy outcomes	ABO variation	Clark 2011 - The influence of maternal Lewis, Secretor and ABO(H) blood groups on fetal growth restriction
	ABO	ABO Blood Group (Transferase A, /	Human	fetal hypoxia, pregnancy complications, hemolytic disease	pregnancy outcomes	ABO incompatibility	Dean 2005 - Chapter 4Hemolytic disease of the newborn - In Blood Groups and Red Cell Antigens
	ABO	ABO Blood Group (Transferase A, /	Human	protection against malaria	pregnancy outcomes	ABO varaition	Pathirana 2005 - ABO-blood-group types and protection against severe, Plasmodium falciparum malaria
	ABO	ABO Blood Group (Transferase A, /	Human	age at menarche	reproductive outcomes	ABO blood group phenotypes	Balgir 1993 - Menarcheal age in relation to ABO blood group phenotypes and haemoglobin-E genotypes
	ABO	ABO Blood Group (Transferase A, /	Human	male infertility	male potential fertility	sperm concentration/function	Abdollahi 2014 - Association of ABO Blood Group System and Anti-Sperm Antibody with Male Infertility
	ABO	ABO Blood Group (Transferase A, /	Human	pregnancy complications	pregnancy outcomes	preeclampsia	2016 Franchini - Relationship between ABO blood group and pregnancy complications: a systematic literature analysis
	ABO	ABO Blood Group (Transferase A, /	Human	female fertility, embryo implantation	female potential fertility	ABO variation	Mengoli 2015 - ABO blood group and fertility: a single-centre study
13	SWAP70	SWAP Switching B-Cell Complex 7(	Monkey	female fertility, implantation, placentation	female potential fertility	SWAP-70 expression	Liu 2006 - Expression of SWAP-70 in the uterus and feto-maternal interface during embryonic implantation and pregnancy in the rhesus monkey (Macaca mulatta)
	SWAP70	SWAP Switching B-Cell Complex 7(	Human	fetal growth restriction	pregnancy outcomes	preeclampsia	Sitras 2014 - Gene expression profile in cardiovascular disease and preeclampsia: A meta-analysis of the transcriptome based on raw data from human studies deposited in Gene Expression Omnibus
14	SH2B3	SH2B Adaptor Protein 3	Human	intrauterine/postnatal growth	pregnancy outcomes	SH2B3 variation	Padidela 2013 - Effects of polymorphisms in the growth hormone and insulin-like growth factor axis on intrauterine and postnatal growth
	SH2B3	SH2B Adaptor Protein 3	Human	male testicular function	male potential fertility	SH2B3 variation	Slack 2010 - Regulation of Lifespan, Metabolism, and Stress Responses by the Drosophila SH2B Protein, Lnk
15	PEMT	Phosphatidylethanolamine N-Meth	Human	fetal growth, placental function	pregnancy outcomes	Choline metabolism/PEMT expression	Hogeveen 2012 - Umbilical choline and related methylamines betaine and dimethylglycine in relation to birth weight
	PEMT	Phosphatidylethanolamine N-Meth	Human	premature birth	pregnancy outcomes	PEMT variation [744CC genotype]	Zhu 2016 - Choline Intake During Pregnancy and Genetic Polymorphisms Influence Choline Metabolism in Chinese Preterms Receiving Total Parenteral Nutrition Therapy
	PEMT	Phosphatidylethanolamine N-Meth	Human	sperm quality	male potential fertility	PEMT variation [27774G.C]	Lazaros 2012 - Phosphatidylethanolamine N-methyltransferase and choline dehydrogenase gene polymorphisms are associated with human sperm concentration
	PEMT	Phosphatidylethanolamine N-Meth	Human	Fetal growth, placental function	pregnancy outcomes	mRNA levels of PEMT	Khot 2014 - Expression of Genes Encoding Enzymes Involved in the One Carbon Cycle in Rat Placenta is Determined by Maternal Micronutrients (Folic Acid, Vitamin B12) and Omega-3 Fatty Acids
	PEMT	Phosphatidylethanolamine N-Meth	Mouse	embryo survival/viability during pre-implantation	female potential fertility	PEMT expression	Cheng 2016 - Requirement of Leukemia Inhibitory Factor or Epidermal Growth Factor for Pre-Implantation Embryogenesis via JAK/STAT3 Signaling Pathways
	PEMT	Phosphatidylethanolamine N-Meth	Human	fetal development	pregnancy outcomes	PEMT expression	Yan 2013 - Pregnancy alters choline dynamics: results of a randomized trial using stable isotope methodology in pregnant and nonpregnant women
16	MRAS	Muscle RAS Oncogene Homolog	Mouse	male testicular function	male potential fertility	MRAS expression	Wei 2012 - Wei 2013 - Transcriptome profiling of the developing postnatal mouse testis using next-generation sequencing
	MRAS	Muscle RAS Oncogene Homolog	Human	embryo implantation	female potential fertility	MRAS regulation by androgen and progesterone receptors	Cloke 2008 - "The Androgen and Progesterone Receptors Regulate Distinct Gene Networks and Cellular Functions in Decidualizing Endometrium
	MRAS	Muscle RAS Oncogene Homolog	Mouse	embryo pluripotency	female potential fertility	MRAS expression	Palmqvist 2005 - "Correlation of Murine Embryonic Stem Cell Gene Expression Profiles with Functional Measures of Pluripotency
	MRAS	Muscle RAS Oncogene Homolog	Human	breastfeeding capacity	reproductive outcomes	MRAS expression	Colodro-Conde 2014 - A Twin Study of Breastfeeding With a Preliminary Genome-Wide Association Scan
17	KIAA1462		Geese	offspring number	reproductive outcomes	KIAA1462 expression	Huang 2015 - "Efficient SNP Discovery by Combining Microarray and Lab-on-a-Chip Data for Animal Breeding and Selection
	KIAA1462		Human	birth-related myometrial gene expression	pregnancy outcomes	KIAA1462 expression	Chan 2014 - Assessment of myometrial transcriptome changes associated with spontaneous human labour by high-throughput RNA-seq
	KIAA1462		Mouse	female reproduction	female potential fertility	KIAA1462 highly expressed in oocytes & ovaries	Fledel-Alon 2011 - Variation in Human Recombination Rates and Its Genetic Determinants
	KIAA1462		Pig	fetal growth	pregnancy outcomes	KIAA1462 expression	Pérez-Montarelo 2014 - Identification of genes regulating growth and fatness traits in pig through hypothalamic transcriptome analysis
	KIAA1462		Human	embryo implantation	female potential fertility	KIAA1462 differential expression	2011 Aghajanova - Comparative Transcriptome Analysis of Human Trophectoderm and Embryonic Stem Cell-Derived Trophoblasts Reveal Key Participants in Early Implantation
18	GUCY1A3	Guanylate Cyclase 1, Soluble, Alph	Cattle	embryo implantation	female potential fertility	GUCY1A3 expression	Ponsuksili 2012 - Gene Expression and DNA-Methylation of Bovine Pretransfer Endometrium Depending on Its Receptivity after In Vitro-Produced Embryo Transfer
	GUCY1A3	Guanylate Cyclase 1, Soluble, Alpha 3		embryo implantation	female potential fertility	GUCY1A3 expression	Duzyj 2014 - PreImplantation factor (PIF*) promotes embryotrophic and neuroprotective decidual genes: effect negated by epidermal growth factor
	GUCY1A3	Guanylate Cyclase 1, Soluble, Alph	Human	placental functioning	pregnancy outcomes	GUCY1A3 expression	Sedlmeier 2014 - Human placental transcriptome shows sexually dimorphic gene expression and responsiveness to maternal dietary n-3 long-chain polyunsaturated fatty acid intervention during pregnancy
	GUCY1A3	Guanylate Cyclase 1, Soluble, Alph	Human	birth weight	pregnancy outcomes	GUCY1A3 expression	Taal 2013 - Early Growth, Cardiovascular and Renal Development The Generation R Study
	GUCY1A3	Guanylate Cyclase 1, Soluble, Alph	Human	fetal growth, birthweight, postnatal growth & metabolism	pregnancy outcomes	fetal genotype	Horikoshi 2013 - New loci associated with birth weight identify genetic links between intrauterine growth and adult height and metabolism
19	CDKN2B-AS1	CDKN2B Antisense RNA 1	Human	fertility	female potential fertility	CDKN2B-AS1 linked endometriosis	Pagliardini 2015 - Replication and meta-analysis of previous genome-wide association studies confirm vezatin as the locus with the strongest evidence for association with endometriosis
	CDKN2B-AS1	CDKN2B Antisense RNA 1	Human	fetal growth restriction	pregnancy outcomes	CDKN2B-AS1 variant	Abd-Rabou 2014 - Molecular genetic studies in pregnancies affected by preeclampsia and intrauterine growth restriction
20	ANKS1A	Ankyrin Repeat And Sterile Alpha I	Cattle	fertility	female potential fertility	ANKS1A expression in endometrium and corpus luteum	Moore 2015 - Differentially expressed genes in endometrium and corpus luteum of Holstein cows selected for high and low fertility are enriched for sequence variants associated with fertility
	ANKS1A	Ankyrin Repeat And Sterile Alpha I	Cattle	fertility	female potential fertility	ANKS1A 6.7 fold upregulated in blastocysts	Gad 2012 - Transcriptome profiling of bovine blastocysts developed under alternative culture conditions during specific stages of development
	ANKS1A	Ankyrin Repeat And Sterile Alpha I	Human/M	male fertility	male potential fertility	ANKS1A expression	Tüttelmann 2011 - Copy Number Variants in Patients with Severe Oligozoospermia and Sertoli-Cell-Only Syndrome
21	PDGFD	Platelet Derived Growth Factor D	Human	female fertility	female potential fertility	ovarian hyperstimulation	Scotti 2014 - Platelet-derived growth factor BB and DD and angiopoietin1 are altered in follicular fluid from polycystic ovary syndrome patients
	PDGFD	Platelet Derived Growth Factor D	Human	female reproduction	female potential fertility	PDGFD expressed in oocytes	Bonnet 2015 - Spatio-Temporal Gene Expression Profiling during In Vivo Early Ovarian Folliculogenesis: Integrated Transcriptomic Study and Molecular Signature of Early Follicular Growth
	PDGFD	Platelet Derived Growth Factor D	Mouse	male/female reproduction	female potential fertility	PDGFD expression	Sharov 2008 - Effects of aging and calorie restriction on the global gene expression profiles of mouse testis and ovary
	PDGFD	Platelet Derived Growth Factor D		female reproductive function	female potential fertility	PDGFD down-regulated in endometrium	Popovici 2006 - Gene Expression Profiling of Human Endometrial- Trophoblast Interaction in a Coculture Model
	PDGFD	Platelet Derived Growth Factor D	Rat	female reproductive function	female potential fertility	PDGFD expression	Sleer 2007 - Cell-Type Localization of Platelet-Derived Growth Factors and Receptors in the Postnatal Rat Ovary and Follicle1
	PDGFD	Platelet Derived Growth Factor D	Human	pregnancy complication, preeclampsia	pregnancy outcomes	PDGFD down-regulated in placenta	Sitrasa, 2009 - Differential Placental Gene Expression in Severe Preeclampsia
22	KSR2	Kinase Suppressor Of Ras 2	Mouse	male fertility	male potential fertility	KSR2 -/- knockout mouse model for spermatogenesis	Moretti, 2015 - Ultrastructural study of spermatogenesis in KSR2 deficient mice
	KSR2	Kinase Suppressor Of Ras 2	Cattle	female reproductive function	female potential fertility	KSR2 up-regulated in epithelial cells	Cerny 2015 - A transcriptomal analysis of bovine oviductal epithelial cells collected during the follicular phase versus the luteal phase of the estrous cycle
	KSR2	Kinase Suppressor Of Ras 2	Mouse	offspring growth	fetal/offspring mortality	KSR2 -/- knockout mouse model	Henry 2014 - Obesity-dependent dysregulation of glucose homeostasis in kinase suppressor of ras 2–/– mice
23	FLT1	Fms-Related Tyrosine Kinase 1	Human	fetal development	pregnancy outcomes	FLT1 expression	Kaipainen 1993 - The related FLT4, FLT1, and KDR receptor tyrosine kinases show distinct expression patterns in human fetal endothelial cells
	FLT1	Fms-Related Tyrosine Kinase 1	Mouse	offspring viability, fetal growth	pregnancy outcomes	FLT1 knockdown	Khankin 2012 - Hemodynamic, Vascular, and Reproductive Impact of FMS-Like Tyrosine Kinase 1 (FLT1) Blockade on the Uteroplacental Circulation During Normal Mouse Pregnancy
	FLT1	Fms-Related Tyrosine Kinase 1	Human	pregnancy loss	fetal/offspring mortality	immune responses to placental malaria	Atis Muehlenbachs 2008 - Natural selection of FLT1 alleles and their association with malaria resistance in utero
	FLT1	Fms-Related Tyrosine Kinase 1		female reproduction	female potential fertility	FLT1 expression in oocytes	Bonnet 2013 - An overview of gene expression dynamics during early ovarian folliculogenesis: specificity of follicular compartments and bi-directional dialog
	FLT1	Fms-Related Tyrosine Kinase 1	Human	intrauterine growth restriction	pregnancy outcomes	FLT1 upregulated	Fritz 2015 - Trophoblast Retrieval And Isolation From The Cervix (tric) For Non-Invasive Prenatal Genetic Diagnosis And Prediction Of Abnormal Pregnancy Outcome
	FLT1	Fms-Related Tyrosine Kinase 1	Human	fetal growth	pregnancy outcomes	FLT1 expression in placenta	Korevaar 2014 - Soluble Flt1 and Placental Growth Factor Are Novel Determinants of Newborn Thyroid (Dys)Function: The Generation R Study
	FLT1	Fms-Related Tyrosine Kinase 1	Human	female reproduction	female potential fertility	FLT1 expression in oocytes	Dan 2014 - Pathogenesis and stem cell therapy for premature ovarian failure
	FLT1	Fms-Related Tyrosine Kinase 1	Human	female reproduction, implantation	female potential fertility	FLT1 expression in uterus	Dey 2004 - Molecular Cues to Implantation
	FLT1	Fms-Related Tyrosine Kinase 1	Human	female reproduction	pregnancy outcomes	FLT1 expression in placenta, fetal tissues	Sood 2006 - Gene expression patterns in human placenta
	FLT1	Fms-Related Tyrosine Kinase 1	Human	intrauterine growth restriction	pregnancy outcomes	FLT1 expression during pregnancy	Tsao 2005 - Excess Soluble fms-Like Tyrosine Kinase 1 and Low Platelet Counts in Premature Neonates of Preeclamptic Mothers
24	ABCG5	ATP-Binding Cassette, Sub-Family	Rat	intrauterine growth restriction	pregnancy outcomes	Rat model of IUGR	Chen 2015 - Effects of intrauterine growth restriction and high-fat diet on serum lipid and transcriptional levels of related hepatic genes in rats
	ABCG5	ATP-Binding Cassette, Sub-Family G, Member		trophoblast, blastocyst development	female potential fertility	ABCG5 gene expression	Ozawa 2012 - Global gene expression of the inner cell mass and trophectoderm of the bovine blastocyst
25	ZC3HC1	Zinc Finger, C3HC-Type Containing 1		male fertility	male potential fertility	meiosis disruptors	Archambeault 2014 - Disrupting the male germ line to find infertility and contraception targets
	ZC3HC1	Zinc Finger, C3HC-Type Containing	Mouse	pregnancy establishment, maintenance, conceptus surviv	female potential fertility	ZC3HC1 expression, 1.57 fold change	Niklaus 2006 - Mining the Mouse Transcriptome of Receptive Endometrium Reveals Distinct Molecular Signatures for the Luminal and Glandular Epithelium
26	SMAD3	SMAD Family Member 3		folliculogenesis	female potential fertility	SMAD3 expression	Xu 2002 - Stage-Specific Expression of Smad2 and Smad3 During Folliculogenesis
	SMAD3	SMAD Family Member 3	Mouse/Ra	oocyte function	female potential fertility	SMAD3 expression	Ethier 2001 - Roles of activin and its signal transduction mechanisms in reproductive tissues
	SMAD3	SMAD Family Member 3		estrogen receptor interactions	female potential fertility	SMAD3 expression	Matsuda 2001 - Cross-talk between Transforming Growth Factor- and Estrogen Receptor Signaling through Smad3
	SMAD3	SMAD Family Member 3	Rat	testis function	male potential fertility	SMAD3 expression	Xu 2003 - Developmental and Stage-Specific Expression of Smad2 and Smad3 in Rat Testis
	SMAD3	SMAD Family Member 3	Human	age at natural menopause	reproductive outcomes	SMAD3 interaction	Pyun 2014 - Genome-wide association studies and epistasis analyses of candidate genes related to age at menarche and age at natural menopause in a Korean population
	SMAD3	SMAD Family Member 3	Human	twinning capcity	reproductive outcomes	SMAD3 genotype (rs17293443-C)	Mbarek 2016 - Identification of Common Genetic Variants Influencing Spontaneous Dizygotic Twinning and Female Fertility
	SMAD3	SMAD Family Member 3	Human	female fertility and fecundity	female potential fertility	SMAD3 promotes proliferation and steroidogenesis of human	Liu 2014 - Effects of Smad3 on the proliferation and steroidogenesis in human ovarian luteinized granulosa cells
	SMAD3	SMAD Family Member 3	Mouse	embryo viability	female potential fertility	SMAD3 signaling	Dunn 2004 - Combinatorial activities of Smad2 and Smad3 regulate mesoderm formation and patterning in the mouse embryo
	SMAD3	SMAD Family Member 3	Human	spermatogenesis, male reproduction	male potential fertility	SMAD3 expression	Itman 2011 - Smad3 Dosage Determines Androgen Responsiveness and Sets the Pace of Postnatal Testis Development
27	SLC22A3	Solute carrier family 22, extra neu	Human	placental functioning	pregnancy outcomes	SLC22A3 expression	Jacob 2005 - Gametes and embryo epigenetic reprogramming affect developmental outcome: Implication for assisted reproductive technologies
	SLC22A3	Solute carrier family 22, extra neu	Human	fetal development, fetal-placental resource provisioning	pregnancy outcomes	SLC22A3 expression	Nelissen 2011 - Epigenetics and the placenta
	SLC22A3	Solute carrier family 22, extra neu	Human	fetal-placental functioning	pregnancy outcomes	SLC22A3 expression changes during pregnancy	Monk 2006 - Limited evolutionary conservation of imprinting in the human placenta
	SLC22A3	Solute carrier family 22, extra neu	Human	fetal-placental functioning	pregnancy outcomes	SLC22A3 expression by trimester	Berveiller 2015 - Limited evolutionary conservation of imprinting in the human placenta
28	REST	RE1-Silencing Transcription Factor	Mouse	embryo functioning	female potential fertility	REST regulatory networks	Johnson 2008 - REST regulates distinct transcriptional networks in embryonic and neural stem cells.
29	PPAP2B	Phospholipid Phosphatase 3	Human	endometriosis, female fertility	female potential fertility	PPAP2B -1.69 fold change	Burney 2007 - Gene Expression Analysis of Endometrium Reveals Progesterone Resistance and Candidate Susceptibility Genes in Women with Endometriosis
	PPAP2B	Phospholipid Phosphatase 3	Human/ro	gametogenesis	male potential fertility	PPAP2B expression	Chalmel 2007 - The conserved transcriptome in human and rodent male gametogenesis
	PPAP2B	Phospholipid Phosphatase 3	Sheep	breeding capacity	reproductive outcomes	PPAP2B association	Pokharel1 2015 - Transcriptome profiling of Finnsheep ovaries during out-of-season breeding period
	PPAP2B	Phospholipid Phosphatase 3	Human	pregnancy complications	pregnancy outcomes	PPAP2B 1.36 fold upregulated in placental tissues of preeclar	Yan 2013 - Screening for preeclampsia pathogenesis related genes
	PPAP2B	Phospholipid Phosphatase 3	Human	embryo viability	female potential fertility	spontaneous abortion due to parthenogenesis	Liu 2010 - Genome-wide Gene Expression Profiling Reveals Aberrant MAPK and Wnt Signaling Pathways Associated with Early Parthenogenesis
	PPAP2B	Phospholipid Phosphatase 3	Human	embryo implantation	female potential fertility	PPAP2B differential expression	Aghajanova 2011 - Comparative Transcriptome Analysis of Human Trophectoderm and Embryonic Stem Cell-Derived Trophoblasts Reveal Key Participants in Early Implantation
	PPAP2B	Phospholipid Phosphatase 3		female reproductive function	female potential fertility	PPAP2B up-regulated in endometrium	Popovici 2006 - Gene Expression Profiling of Human Endometrial- Trophoblast Interaction in a Coculture Model
30	MIA3	Melanoma Inhibitory Activity Fami	Cattle	ovarian functioning	female potential fertility	MIA3 4.6 fold upregulated	Hatzirodos 2014 - Transcriptome profiling of granulosa cells from bovine ovarian follicles during atresia
	MIA3	Melanoma Inhibitory Activity Fami	Mouse	placental (dys)function	female potential fertility	MIA3 expression in early trophoblast differentiation	Chaynel Moore 2011 - Characterizing the Nature of the Xst199 Mutation and the Normal Function of the TANGO1 Protein in Trophoblast Cells
31	IL6R	Interleukin 6 Receptor	Pig	endometrium functioning	pregnancy outcomes	IL6R differentially expressed in endometrium	Østrup 2010 - Differential Endometrial Gene Expression in Pregnant and Nonpregnant Sows
	IL6R	Interleukin 6 Receptor	Cattle	endometrium functioning	pregnancy outcomes	IL6R 3.38 fold up-regulated during pregnancy	Mansouri-Attia 2009 - Endometrium as an early sensor of in vitro embryo manipulation technologies
	IL6R	Interleukin 6 Receptor	Human	endometrium functioning in PCOS women	female potential fertility	IL6R up-regulated	Yan 2012 - Expression of apoptosis-related genes in the endometrium of polycystic ovary syndrome patients during the window of implantation
	IL6R	Interleukin 6 Receptor	Human	pre-term birth snps	pregnancy outcomes	IL6R significantly associated	Anum 2009 - Genetic Contributions to Disparities in Preterm Birth (review)
32	HDAC9	Histone Deacetylase 9	Human	oocyte function	female potential fertility	HDAC9 expression	Zhang 2007 - Distinct sets of developmentally regulated genes that are expressed by human oocytes and human embryonic stem cells
	HDAC9	Histone Deacetylase 9	Cattle	male fertility	male potential fertility	germ cell production	Hering 2014 - Genome-Wide Association Study for Sperm Concentration in Holstein-Friesian Bulls
	HDAC9	Histone Deacetylase 9	Pig	birth weight	pregnancy outcomes	HDAC9 expression	Zhang 2014 - Genome Wide Screening of Candidate Genes for Improving Piglet Birth Weight Using High and Low Estimated Breeding Value Populations

	HDAC9	Histone Deacetylase 9	Human/M	oocyte function	female potential fertility	HDAC9 expression	Assou 2006 - The human cumulus–oocyte complex gene-expression profile
	HDAC9	Histone Deacetylase 9	Human	birth-related myometrial gene expression	pregnancy outcomes	HDAC9 expression	Chan 2014 - Assessment of myometrial transcriptome changes associated with spontaneous human labour by high-throughput RNA-seq
33	**COL4A1**	Collagen, Type IV, Alpha 1	Pig	neonate survival	pregnancy outcomes	COL4A1 expression	Jiang 2007 - Expression of X-linked genes in deceased neonates and surviving cloned female piglets
	COL4A1	Collagen, Type IV, Alpha 1	Human	testis function	male potential fertility	COL4A1 expression	Chen 2013 - The Wilms tumor gene, Wt1, maintains testicular cord integrity by regulating the expression of Col4a1 and Col4a2
	COL4A1	Collagen, Type IV, Alpha 1	Mouse	folliculogenesis	female potential fertility	COL4A1 expression	Borup 2016 - Competence Classification of Cumulus and Granulosa Cell Transcriptome in Embryos Matched by Morphology and Female Age
	COL4A1	Collagen, Type IV, Alpha 1	Human	fetal survival	fetal/offspring mortality	COL4A1 mutation	Garel 2013 - Fetal intracerebral hemorrhage and COL4A1 mutation: promise and uncertainty
	COL4A1	Collagen, Type IV, Alpha 1		fetal/placenta growth and development	pregnancy outcomes	COL4A1 expression	Rodriguez-Zas 2008 - Biological interpretations of transcriptomic profiles in mammalian oocytes and embryos
34	**ABHD2**	Abhydrolase Domain Containing 2		male fertility	male potential fertility	ABHD2 expression	Gerhardt 2016 - Progesterone and Endocannabinoid Interaction Alters Sperm Activation
35	**SORT1**	Sortilin 1	Human	endometrium functioning	pregnancy outcomes	SORT1 expression during labour	Chan 2014 - Assessment of myometrial transcriptome changes associated with spontaneous human labour by high-throughput RNA-seq
	SORT1	Sortilin 1		ovarian functioning	female potential fertility	SORT1 up-regulated	Zhao 2012 Effect of luteal-phase support on endometrial microRNA expression following controlled ovarian stimulation
	SORT1	Sortilin 1	Rat	ovarian functioning	female potential fertility	SORT1 expression	Wang 2012 - Effect of CdCl_2 on Expression of Sortilin(Sort1) in Rat Ovary
	SORT1	Sortilin 1	Human	embryo implantation	female potential fertility	SORT1 differential expression	Aghajanova 2011 - Comparative Transcriptome Analysis of Human Trophectoderm and Embryonic Stem Cell-Derived Trophoblasts Reveal Key Participants in Early Implantation
36	**SLC22A5**	Solute Carrier Family 22 (Organic C	Mouse	male infertility	male potential fertility	SLC22A5 mutation	Toshimori 1999 - Dysfunctions of the epididymis as a result of primary carnitine deCciency in juvenile visceral steatosis mice
	SLC22A5	Solute Carrier Family 22 (Organic C	Pig	reproductive variation, offspring born alive and total born	reproductive outcomes	SLC22A5 genotype	Mote 2014 - Identification of genetic markers for productive life in commercial sows.
	SLC22A5	Solute Carrier Family 22 (Organic C	Pig	age at puberty	reproductive outcomes	SLC22A5 genotype	Rempel 2014 - Association analyses of candidate single nucleotide polymorphisms on reproductive traits in swine.
37	**NOA1**	Nitric Oxide Associated 1	Human	Male fertility, testicular functioning	male potential fertility	NOA1 expression	Okada 2008 - Genome-wide expression of azoospermia testes demonstrates a specific profile and implicates ART3 in genetic susceptibility
	NOA1	Nitric Oxide Associated 1	Mouse	embryo/trophoblast viability	female potential fertility	NOA1-deficient mouse model	Omary 2011 - NOA1 is an essential GTPase required for mitochondrial protein synthesis
38	**LPL**	Lipoprotein Lipase	Human	pregnancy complications	pregnancy outcomes	LPL expression	Schmella 2015 - The -93T/G LPL Promoter Polymorphism Is Associated With Lower Third-Trimester Triglycerides in Pregnant African American Women
	LPL	Lipoprotein Lipase	Human	male infertility	male potential fertility	sperm DNA fragmentation related to LPL expression	Intasqui1 2013 - Sperm nuclear DNA fragmentation rate is associated with differential protein expression and enriched functions in human seminal plasma
	LPL	Lipoprotein Lipase	Human	reproductive timing	reproductive outcomes	LPL expression	Spencer 2013 - Genetic Variation and Reproductive Timing: African American Women from the Population Architecture Using Genomics and Epidemiology (PAGE) Study
	LPL	Lipoprotein Lipase	Human	intrauterine growth restriction	pregnancy outcomes	LPL-mediated fetal-placental nutrient transfer	Tabano 2006 - Placental LPL Gene Expression Is Increased in Severe Intrauterine Growth-Restricted Pregnancies
	LPL	Lipoprotein Lipase	Human/M	placental functioning	pregnancy outcomes	LPL expression	Lindegaard 2005 - Endothelial and lipoprotein lipases in human and mouse placenta
	LPL	Lipoprotein Lipase		fetal/placental resource transfer, pregnancy complication	pregnancy outcomes	LPL expression	Lager 2012 - Regulation of Nutrient Transport across the Placenta
	LPL	Lipoprotein Lipase	Human	testis/spermatogenesis	male potential fertility	LPL expression	Nielsen 2009 - Lipoprotein lipase and endothelial lipase in human testis and in germ cell neoplasms
	LPL	Lipoprotein Lipase	Mouse	Placental regulation of cholesterol	pregnancy outcomes	Maternal-fetal transfer of lipids	Overbrgh 1995 - Expression of mouse alpha-macroglobulins, lipoprotein rece r-related protein, LDL receptor, apolipoprotein E, and lipoprotein lipase in pregnancy
39	**COL4A2**	Collagen, Type IV, Alpha 2	Mouse	fetal viability	fetal/offspring mortality	Mouse knockout model for COL4A2	Kuo 2012 - COL4A1 and COL4A2 mutations and disease: insights into pathogenic mechanisms and potential therapeutic targets
	COL4A2	Collagen, Type IV, Alpha 2	Human	testis function	male potential fertility	COL4A2 expression	Chen 2013 - The Wilms tumor gene, Wt1, maintains testicular cord integrity by regulating the expression of Col4a1 and Col4a2
	COL4A2	Collagen, Type IV, Alpha 2	Human	offspring viability	fetal/offspring mortality	COL4A2 expression	Yoneda 2012 - De Novo and Inherited Mutations in COL4A2, Encoding the Type IV Collagen α2 Chain Cause Porencephaly
40	**ADAMST7**	ADAM Metallopeptidase With Thr	Mouse	embryogenesis	female potential fertility	COL4A2 expression	Hurskainen 1999 - ADAM-TS5, ADAM-TS6, and ADAM-TS7, novel members of a new family of zinc metalloproteases. General features and genomic distribution of the ADAM-TS family.
	ADAMST7	ADAM Metallopeptidase With Thr	Dog	mammary tissue functioning	reproductive outcomes	ADAMST7 up-regulated in mammary tissues	Rao 2009 - GENE EXPRESSION PROFILES OF PROGESTIN-INDUCED CANINE MAMMARY HYPERPLASIA AND SPONTANEOUS MAMMARY TUMORS
	ADAMST7	ADAM Metallopeptidase With Thr	Human	breastfeeding capacity	reproductive outcomes	ADAMST7 expression	Colodro-Conde 2014 - A Twin Study of Breastfeeding With a Preliminary Genome-Wide Association Scan

Notes:
*** 'gene rank' based on Fig. 1B (i.e. number of significant genetic risk * selection associations)**
****'classes' column colouring defined further**

male potential fertility	processes leading up to/during fertilization
female potential fertility	processes leading up to/during fertilization
fetal/offspring mortality	survival of fetus/offspring
reproductive outcomes	reproductive timing, actual number of offspring produced
pregnancy outcomes	processes occurring during pregnancy negatively affecting fetus/mother (not including death)

Search criteria:
- For each CAD gene, google scholar was used to search for studies using the 'Search terms' (below) and the gene name (BCAS3 is used as an example)
- For each search, only the first page of results was considered. Search results most consistent with all search terms are ranked by page, thus the most relevant results were always on the first page. This approach was also employed to keep this literature search tractable in terms of time (i.e. a search for each of the terms below for one gene usually took 1 hour +).
- We also used the GWAS Catalog (https://www.ebi.ac.uk/gwas/) using the gene name to search for further potential links to fitness related traits

Search terms:
"BCAS3" and "reproduction" and gene and -"noncommercial use, distribution, and reproduction in any"
"BCAS3" and "fitness" and gene
"BCAS3" and "fertility" and gene
"BCAS3" and "menarche" and gene
"BCAS3" and "menopause" and gene
"BCAS3" and "birth" or "birth weight"
"BCAS3" and "pregnancy" and gene
"BCAS3" and "placenta" and gene
"BCAS3" and "implantation" and gene
"BCAS3" and "oocyte" and gene
"BCAS3" and "sperm" and gene
"BCAS3" and "testis"

Table S2. Summary of types of pleiotropic effects CAD genes have on fitness (based on Table S1)

Gene Rank	Gene	male potential fertility	female potential fertility	sum(both male and female fertility)	pregnancy outcomes	reproductive outcomes	fetal/offspring mortality	sum
1	BCAS3		1					2
2	CNNM2				1			1
3	TEX41				1			1
4	SMG6		1					1
5	PHACTR1		1		1	1		3
6	COG5	1			1			2
7	ABCG8	1	1	1	1		1	5
8	RAI1				1			1
9	NT5C2		1		1			2
10	LDLR				1		1	2
11	KCNK5	1	1	1		1		4
12	ABO	1	1	1	1	1	1	6
13	SWAP70		1		1			2
14	SH2B3	1			1		1	3
15	PEMT	1	1	1	1			4
16	MRAS	1	1	1		1		4
17	KIAA1462		1		1	1		3
18	GUCY1A3		1		1		1	3
19	CDKN2B-AS1		1		1			2
20	ANKS1A	1	1	1				3
21	PDGFD		1		1			2
22	KSR2	1	1	1			1	4
23	FLT1		1		1		1	3
24	ABCG5		1		1			2
25	ZC3HC1	1	1	1				3
26	SMAD3	1	1	1		1		4
27	SLC22A3				1			1
28	REST		1					1
29	PPAP2B	1	1	1	1	1		5
30	MIA3		1					1
31	IL6R		1		1			2
32	HDAC9	1	1	1	1			4
33	COL4A1	1	1	1	1		1	5
34	ABHD2	1						1
35	SORT1		1		1			2
36	SLC22A5	1				1		2
37	NOA1	1	1	1				3
38	LPL	1			1	1		3
39	COL4A2	1					1	2
40	ADAMST7		1			1		2
	sum	19	29	13	25	10	9	
							sum	106
						average		2.7

Supplemental Information

Table S3. Literature search results* - assessing fitness links for 20 randomly selected non-CAD associated genes

CAD gene	Random Gene**	Full name	Species	Potential fitness effects/links	Potential fitness class	Causative factor/model	Reference
BCAS3	**STPG2**	Sperm Tail PG-Rich Repeat Containi	-	-	-	-	-
CNNM2	**CFAP44**	Cilia And Flagella Associated Proteir	-	-	-	-	-
TEX41	**SHISA9**	Shisa Family Member 9	-	-	-	-	-
SMG6	**TANGO6**	Transport And Golgi Organization 6	-	-	-	-	-
PHACTR1	**SUMF1**	Sulfatase Modifying Factor 1	Mouse	embryogenesis	female potential fertility	sumf1 upregulated in developmentally incompetent mouse oocytes	Zucotti 2008 - Maternal Oct-4 is a potential key regulator of the developmental competence of mouse oocytes
COG5	**FRMD5**	FERM Domain Containing 5	-	-	-	-	-
ABCG8	**ASIC5**	Acid Sensing Ion Channel Subunit Fa	-	-	-	-	-
RAI1	**ZNF516**	Zinc Finger Protein 516	Human	endometriosis	female potential fertility	ZNF516 appears to be involved in endometriosis	Sun 2014 - Genome-wide profiling of long noncoding ribonucleic acid expression patterns in ovarian endometriosis by microarray
NTSC2	**LANCL1-AS1**	LANCL1 Antisense RNA 1	-	-	-	-	-
LDLR	**SYT13**	Synaptotagmin 13	-	-	-	-	-
KCNK5	**FAM53A**	Family With Sequence Similarity 53	-	-	-	-	-
ABO	**TTC22**	Tetratricopeptide Repeat Domain 2	-	-	-	-	-
SWAP70	**RNF157**	Ring Finger Protein 157	Cattle	oocyte/follicle maturation (oocyte quality)	female potential fertility	In cattle model, RNF157 2.24 differentially upregulated between the I	Janowski 2012 - Incidence of apoptosis and transcript abundance in bovine follicular cells is associated with the quality of the enclosed oocyte
	RNF157	Ring Finger Protein 157	Human	early peripheral blood gene expression during pregnancy related to preeclampsi	pregnancy complications/outc	RNF157 is -1.65 fold significantly (P=0.01) downregulated in peripher	Enquobahrie 2011 - Maternal peripheral blood gene expression in early pregnancy and preeclampsia
SH2B3	**PLBD1-AS1**	PLBD1 Antisense RNA 1	-	-	-	-	-
PEMT	**WAC**	WW Domain Containing Adaptor W	-	-	-	-	-
MRAS	**TMEM178A**	Transmembrane Protein 178A	-	-	-	-	-
KIAA1462	**PLEKHD1**	Pleckstrin Homology And Coiled-Coi	-	-	-	-	-
GUCY1A3	**MACC1**	Metastasis Associated In Colon Can	-	-	-	-	-
CDKN2B-AS1	**CACNA2D4**	Calcium Voltage-Gated Channel Au	-	-	-	-	-
ANKS1A	**NWD2**	NACHT And WD Repeat Domain Coi	-	-	-	-	-

Notes:

*The same search terms/criteria were used for random genes as for the CAD genes (see 'Notes' under Table S1)

**The random gene was chosen randomly from the lists of random genes that were chosen for the analysis presented in Fig. 1C. I.e. For each CAD gene, 100 randomly chosen genes (without replacement) that were ~ the same size.

Table S4. Selected Enrichr analysis outputs for top 10-ranked CAD genes with highest genetic risk-selection associations from Fig. 1

PATHWAYS - KEGG 2016 Pathways

Index	Name	P-value	Adjusted p-value	Z-score	Combined score
1	Bile secretion_Homo sapiens_hsa04976	0.0006325	0.008222	-1.77	8.48
2	Ovarian steroidogenesis_Homo sapiens_hsa04913	0.02878	0.07482	-1.85	4.79
3	Fat digestion and absorption_Homo sapiens_hsa04975	0.02374	0.07482	-1.76	4.57
4	Nicotinate and nicotinamide metabolism_Homo sapiens_hsa00760	0.01700	0.07482	-1.75	4.53
5	Aldosterone synthesis and secretion_Homo sapiens_hsa04925	0.04596	0.09556	-1.87	4.39
6	ABC transporters_Homo sapiens_hsa02010	0.02542	0.07482	-1.65	4.28
7	Toxoplasmosis_Homo sapiens_hsa05145	0.06617	0.09558	-1.71	4.01
8	Hepatitis C_Homo sapiens_hsa05160	0.07427	0.09655	-1.61	3.76
9	Pyrimidine metabolism_Homo sapiens_hsa00240	0.05911	0.09558	-1.60	3.75
10	mRNA surveillance pathway_Homo sapiens_hsa03015	0.05145	0.09556	-1.56	3.66

ONTOLOGIES - MGI Mammalian Phenotype Level 3

Index	Name	P-value	Adjusted p-value	Z-score	Combined score
1	MP0000003_abnormal_adipose_tissue_	0.01005	0.1061	-2.31	5.19
2	MP0003718_maternal_effect_	0.01434	0.1061	-2.10	4.71
3	MP0005395_other_phenotype_	0.01434	0.1061	-1.88	4.21
4	MP0002139_abnormal_hepatobiliary_system_	0.005946	0.1061	-1.86	4.16
5	MP0009389_abnormal_extracutaneous_pigme_	0.03729	0.1769	-1.87	3.23
6	MP0002168_other_aberrant_phenotype_	0.03869	0.1769	-1.67	2.90
7	MP0001764_abnormal_homeostasis_	0.01658	0.1061	-1.22	2.74
8	MP0005501_abnormal_skin_physiology_	0.05987	0.2395	-1.60	2.29
9	MP0000358_abnormal_cell_content/_	0.07746	0.2680	-1.41	1.86
10	MP0005253_abnormal_eye_physiology_	0.09214	0.2680	-1.37	1.81

CELL TYPES - Cancer Cell Line Encyclopedia

Index	Name	P-value	Adjusted p-value	Z-score	Combined score
1	IGR1_SKIN	0.002319	0.1055	-1.95	4.38
2	HEYA8_OVARY	0.03735	0.1638	-2.31	4.18
3	OVK18_OVARY	0.003838	0.1055	-1.79	4.03
4	HTK_HAEMATOPOIETIC_AND_LYMPHOID_TISSUE	0.03195	0.1638	-2.13	3.85
5	HS944T_SKIN	0.04174	0.1638	-2.08	3.76
6	MFE296_ENDOMETRIUM	0.04564	0.1638	-2.05	3.71
7	HS746T_STOMACH	0.05146	0.1638	-2.03	3.67
8	WM983B_SKIN	0.05724	0.1638	-1.96	3.54
9	NCIH650_LUNG	0.08805	0.1638	-1.95	3.53
10	TE10_OESOPHAGUS	0.05387	0.1638	-1.94	3.51

TRANSCRIPTION - ChEA 2015

Index	Name	P-value	Adjusted p-value	Z-score	Combined score
1	CTNNB1_20460455_ChIP-Seq_HCT116_Human	4.305e-7	0.00005554	-2.06	20.15
2	TRIM28_17542650_ChIP-ChIP_NTERA2_Human	0.4994	0.5017	-24.98	17.23
3	ESR1_20079471_ChIP-ChIP_T-47D_Human	0.0007167	0.009462	-2.99	13.94
4	FOXP1_22492998_ChIP-Seq_STRATIUM_Mouse	0.005027	0.03393	-4.06	13.72
5	ESR1_22446102_ChIP-Seq_UTERI_Mouse	0.0002684	0.005797	-2.46	12.65
6	FOXA2_19822575_ChIP-Seq_HepG2_Human	0.000003903	0.0002810	-1.20	9.82
7	TFEB_21752829_ChIP-Seq_HELA_Human	0.0003687	0.007099	-1.92	9.50
8	BCL11B_21912641_ChIP-Seq_STHDH STRIUM _Mouse	0.008180	0.04530	-2.86	8.85
9	ARNT_22903824_ChIP-Seq_MCF7_Human	0.0007447	0.009462	-1.83	8.55
10	CIITA_18437201_ChIP-ChIP_Raji B and iDC_Human	0.01114	0.05347	-2.91	8.53

Figure S1

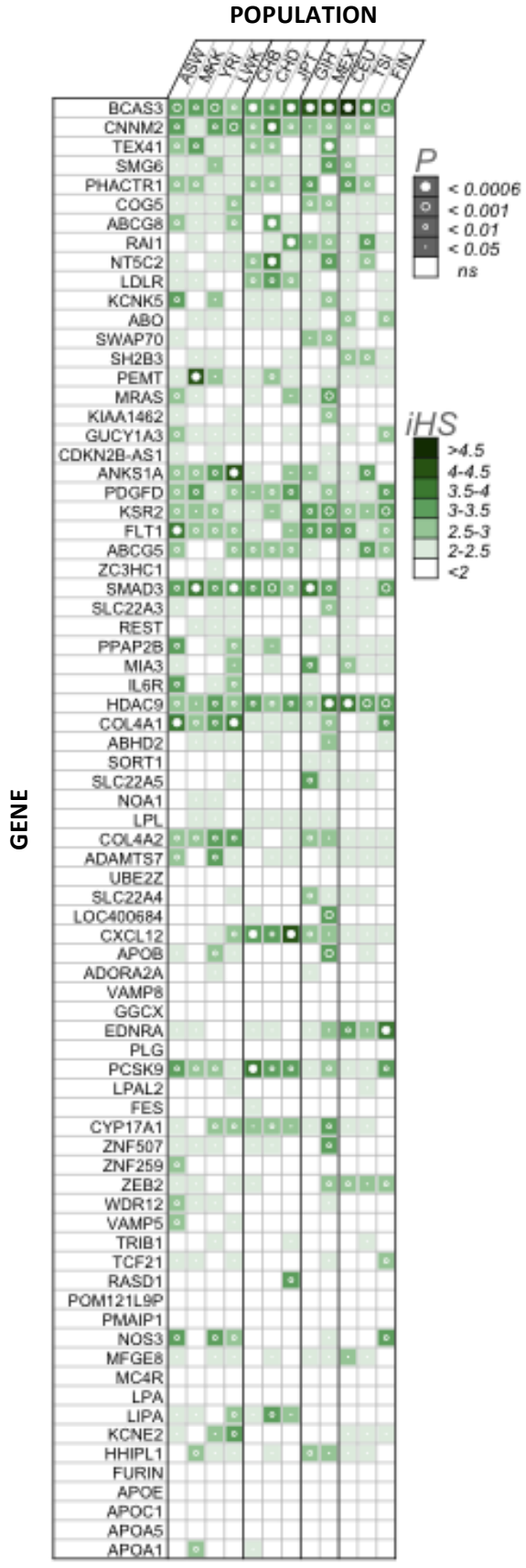


Figure S1: Association of coronary artery disease (CAD) risk and genomic signatures of selection in 12 worldwide populations. All 76 genes are shown ranked according to Fig. 1B. Boxes show magnitude and significance of largest positive selection signal (integrated haplotype score, iHS) within each gene-population combination. P values (circles within squares) were obtained from 10000 permutations. Bonferroni corrected p value limit also shown (α=0.05/76=0.000657) with closed circles. **Populations**. Grouped by common ancestry, African (ASW, African ancestry in Southwest USA; MKK, Maasai in Kinyawa, Kenya; YRI, Yoruba from Ibadan, Nigeria; LWK, Luhya in Webuye, Kenya), East-Asian (CHB, Han Chinese subjects from Beijing; CHD, Chinese in Metropolitan Denver, Colorado; JPT, Japanese subjects from Tokyo), European (CEU, Utah residents with ancestry from northern and western Europe from the CEPH collection; TSI, Tuscans in Italy; FIN, Finnish in Finland), GIH (Gujarati Indians in Houston, TX, USA), MEX (Mexican ancestry in Los Angeles, CA, USA).

Figure S2

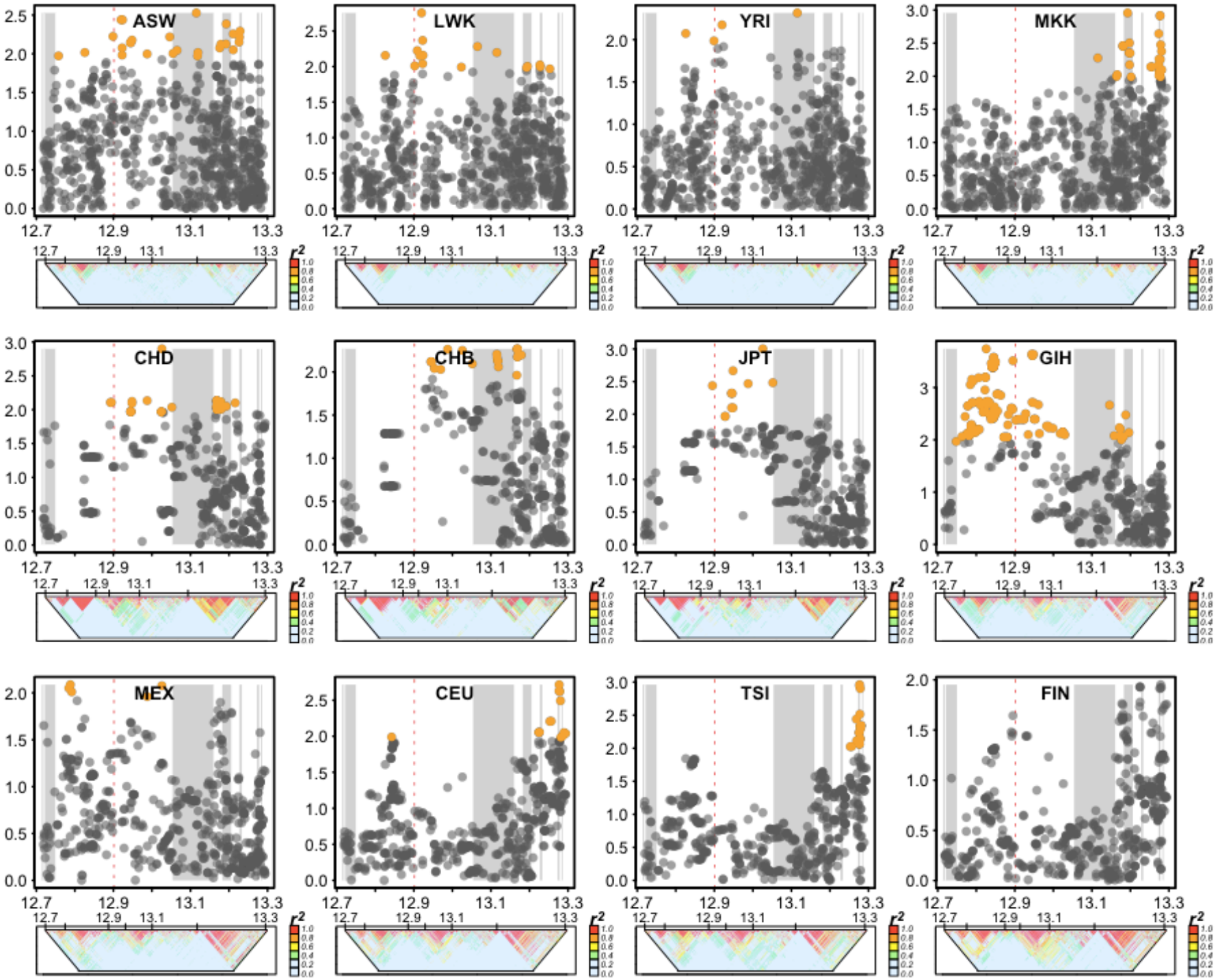


Figure S2: Comparing cross-population candidate selection signals in *PHACTR1*. Per-SNP integrated Haplotype Scores (iHS) plotted by chromosome position within *PHACTR1* (including LD plots below each) for 12 worldwide populations. Permuted p value significance for each score coded by color (grey, non-significant; orange, $p < 0.05$). Red dashed line indicates position of index SNP for *PHACTR1*. Grey columns in background represent intron spans. Populations are clustered by common ancestry, African (ASW, African ancestry in Southwest USA; MKK, Maasai in Kinyawa, Kenya; YRI, Yoruba from Ibadan, Nigeria; LWK, Luhya in Webuye, Kenya), East-Asian (CHB, Han Chinese subjects from Beijing; CHD, Chinese in Metropolitan Denver, Colorado; JPT, Japanese subjects from Tokyo), European (CEU, Utah residents with ancestry from northern and western Europe from the CEPH collection; TSI, Tuscans in Italy; FIN, Finnish in Finland), GIH (Gujarati Indians in Houston, TX, USA), MEX (Mexican ancestry in Los Angeles, CA, USA).