Runs of homozygosity – association with coronary artery disease and gene expression in monocytes and macrophages

Paraskevi Christofidou,[1,17] Christopher P Nelson,[1,10,17] Majid Nikpay,[2,3] Liming Qu,[4] Mingyao Li,[4] Christina Loley,[5] Radoslaw Debiec,[1] Peter S Braund,[1] Matthew Denniff,[1] Fadi J Charchar,[6] Ares Rocanin Arjo,[7,8,9] David-Alexandre Trégouët,[7,8,9] Alison H Goodall,[1,10] Francois Cambien,[7,8,9] Willem H Ouwehand,[11] Robert Roberts,[2,3] Heribert Schunkert,[12,13] Christian Hengstenberg,[12,13] Muredach P Reilly,[14] Jeanette Erdmann,[15] Ruth McPherson,[2,3] Inke R König,[5] John R Thompson,[16] Nilesh J Samani,[1,10] Maciej Tomaszewski[1,10,*]

[1]*Department of Cardiovascular Sciences, University of Leicester, Leicester, LE3 9QP, UK*

[2]*Ruddy Cardiovascular Genetics Centre, University of Ottawa Heart Institute, Ottawa, Ontario, K1Y 4W7, Canada*

[3]*Atherogenomics Laboratory, University of Ottawa Heart Institute, Ottawa, Ontario, K1Y 3V5, Canada*

[4]*Department of Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, Pennsylvania, PA19104, USA*

[5]*Institute of Medical Biometry and Statistics, University of Lübeck, Lübeck, 23562, Germany*

[6] *Faculty of Science and Technology, School of Applied and Biomedical Sciences, Federation University Australia, Victoria, 3350, Australia*

[7]*ICAN Institute for Cardiometabolism and Nutrition, Paris, 75013, France*

[8]*INSERM, UMR_S 1166, Team Genomics & Pathophysiology of Cardiovascular Diseases, Paris, 75013, France*

[9]*Sorbonne Universités, UPMC University, Paris 06, UMR_S 1166, 75013, France*

[10]*NIHR Biomedical Research Unit in Cardiovascular Disease, Leicester, LE3 9QP, UK*

[11] *Department of Haematology, Cambridge Biomedical Campus, University of Cambridge and NHS Blood and Transplant, Cambridge, CB2 0PT, UK and Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1HH, UK*

[12]*Deutsches Herzzentrum München, Klinik für Herz- und Kreislauferkrankungen, Technische Universität München, Munich, 80636, Germany*

[13]*Deutsches Zentrum für Herz- und Kreislauf-Forschung (DZHK) e.V.(German Center for Cardiovascular Research), partner site Munich Heart Alliance, Munich, 80636, Germany*

[14]*Cardiovascular Institute, University of Pennsylvania, Philadelphia, Pennsylvania, PA 19148, USA*

[15]*Institute for Integrative and Experimental Genomics, University of Lübeck, Lübeck, 23562,Germany*

[16]*Department of Health Sciences, University of Leicester, Leicester, LE1 7RH, UK*

[17] These authors have contributed equally to this work

[*]Correspondence:mt142@le.ac.uk

Abstract

Runs of homozygosity (ROHs) are recognised signature of recessive inheritance. Contributions of ROHs to the genetic architecture of coronary artery disease and regulation of gene expression in cells relevant to atherosclerosis are not known.

Our combined analysis of 24,320 individuals from 11 populations of white European ethnicity showed an association between coronary artery disease and both the count and the size of ROHs. Individuals with coronary artery disease had approximately 0.63 (95% CI: 0.4-0.8) excess of ROHs when compared to coronary artery disease-free controls ($P=1.49x10^{-9}$). The average total length of ROHs was approximately 1046.92 (95% CI: 634.4-1459.5) kb greater in individuals with coronary artery disease than controls ($P=6.61x10^{-7}$). None of the identified individual ROHs was associated with coronary artery disease after correction for multiple testing. However, in aggregate burden analysis, ROHs favouring increased risk of coronary artery disease were much more common than those showing the opposite direction of association with coronary artery disease ($P=2.69x10^{-33}$). Individual ROHs showed significant associations with monocyte and macrophage expression of genes in their close proximity – subjects with several individual ROHs showed significant differences in the expression of 44 mRNAs in monocytes and 17 mRNAs in macrophages when compared to subjects without those ROHs.

This study provides evidence for an excess of homozygosity in coronary artery disease in outbred populations and suggest the potential biological relevance of ROHs in cells of importance to the pathogenesis of atherosclerosis.

Introduction

Coronary artery disease (CAD) is a complex, heterogeneous polygenic disorder. The largest genome-wide association (GWA) meta-analysis conducted to date reported 46 variants associated with risk of CAD (1). All these variants were identified assuming an additive mode of inheritance for CAD. It is increasingly recognised that the genetic architecture of complex disorders, including CAD, is not a simple composite of variants that operate exclusively under an additive mode of inheritance and that both dominant and recessive component may make important contributions overseen by conventional GWA studies.

The potential importance of recessively inherited variants to cardiovascular disease was suggested by several previous investigations. Firstly, history of parental consanguinity (and thus increased homozygosity) was associated with increased risk of premature myocardial infarction, independent of conventional CAD risk factors in a population of young adults of South Asian ethnicity (2). Secondly, studies in isolated populations with increased parental relatedness suggested possible associations between homozygosity or its proxies and the risk of both CAD (3-5) and other cardiovascular phenotypes (5-6). These associations were attributed to the increased levels of homozygosity and inbreeding depression (reduced biological fitness) in these populations (7).

Homozygosity mapping is a strategy with a potential to identify and quantify the recessive component of inheritance – long stretches (usually >1Mb) of consecutive homozygous genotypes are known as runs of homozygosity (ROHs) (8). In the human genome ROHs represent "re-union" of pieces of DNA from common ancestors in their descendants (8). Identical by descent, ROHs arise from background relatedness promoted by demographic processes that increase homozygosity and reduce population size, cultural/social factors that favour consanguinity and natural selection (9). A number of studies clearly demonstrated the presence of relatively frequent ROHs in outbred populations (8, 10-12). However, only a few studies successfully used the GWA derived homozygosity measures to examine their role in the genetic architecture of complex disorders in outbred populations (13-15). To the best of our knowledge, there is no study that examined whether homozygosity is associated with CAD in such populations and whether individual ROHs may play a role in regulation of gene expression within cells of key importance to atherosclerosis.

The primary goal of this project was a comprehensive analysis of association between genome-wide homozygosity measures and CAD in individuals of white European ancestry. A secondary analysis was undertaken to identify and quantify consensus ROHs overlapping across studies and explore their potential relevance to CAD individually and at the aggregate level. Finally, we explored the association of consensus ROHs and gene expression in human monocytes and macrophages. The overview of the strategy used is included in Figure 1.

Methods

Study cohorts

Genetic information on 24,320 biologically unrelated individuals (all of white European ancestry) was collected from 11 previous GWA studies (16). A total of 12,123 individuals with CAD and 12,197 CAD-free controls from the Wellcome Trust Case Control Consortium (WTCCC) (17), the German Myocardial Infarction Family Studies (GerMIFS-I, II and III) (17-19), InterHeart Study (ITH) (20), the Ottawa Heart Genomics Studies (OHGS-A, B and C) (21), PennCATH (22), Cleveland Clinic Gene Bank (CCGB) and Duke Cathgen Study (DUKE) were included in the analysis. In brief, six studies (GERMIFSI, GERMIFSII, GERMIFSIII, OHGS-A, PennCATH and WTCCC) were derived from the original CARDIoGRAM Consortium (16). Five additional studies (CCGB, DUKE, ITH, OHGS-B and OHGS-C), not originally a part of the Consortium (16) agreed to participate bringing a total number of examined populations to 11. Further information on the recruitment and phenotyping of the studies is available in Supplementary material including Table S1.

Genotyping and imputation

DNA was extracted from peripheral blood in all studies but British 1958 Birth Cohort (58BC) that contributed controls to WTCCC (23). Indeed, in these subjects DNA was extracted from cell lines (23). Information on single nucleotide polymorphisms (SNPs) included in the analysis is shown in Table S2. Only imputed genotypes that could be called with a posterior probability of ≥90% were

included in the analysis. SNPs were removed from further analysis if their minor allele frequency (MAF) was <1%, their genotype distribution deviated from Hardy-Weinberg equilibrium with $p<0.001$ or the genotype post-imputation call rate was ≤95%. Individual samples were removed from further analysis if the sample call rate was ≤95%. All quality control filters were applied individually at the cohort level.

ROHs identification

A number of ROH definitions and methods of their detection have been proposed (12, 24-26). In this project ROHs were identified via the "Runs of Homozygosity" program – implemented in PLINK (v 1.07) (27). The adopted PLINK parameters are similar to the ones used in previous publications (28-29). A sliding window of 50 SNPs in 5,000 kb length region was used to scan the genome. To prevent underestimating the number and size of ROHs, 1 heterozygote and 2 missing calls in each window were permitted to allow for possible genotyping errors within a stretch of truly homozygous SNPs or other sources of artificial heterozygosity. A SNP was counted as a part of a ROH if >5% of windows spanning it were homozygous. These parameters were selected to minimize the probability of a window being called homozygous by chance. The existence of linkage disequilibrium (LD) blocks in DNA means that relatively short ROHs (those spanning from tens to hundreds kb) are very prevalent across the genome (30-33). In order to exclude these very common short tracts of homozygous SNPs, the minimum length for a ROH was set at 1Mb as used by several other studies (13, 28). Two additional parameters were added to ensure that estimates were not artificially inflated by apparently homozygous tracts in sparsely covered genomic regions. Firstly, the required minimum SNP density was set to 50, meaning at least 1 SNP had to be present per 50kb of DNA, and secondly, the maximum distance between two consecutive homozygous SNPs was set to 100kb. To ensure that the analysis captures only regions that are entirely homozygous between the first and the last SNP a threshold for the minimum number of SNPs constituting a ROH was selected. In line with previous studies on homozygosity of complex disorders, the minimum number of homozygous SNPs to qualify as a ROH in this project was set to 100 (13). There were no major differences in the results of ROH

analyses conducted under different thresholds of ROH calculation parameters in our pilot sensitivity analyses in WTCCC (data not shown).

Calculation of homozygosity measures

The following measures of homozygosity were calculated in each study (a) the average number of ROHs, (b) the total and average length of ROHs and (c) the proportion of the autosomal genome covered by ROHs (FROH) (34). The total number of ROHs was defined as the sum of all ROHs per individual. The average ROH number was calculated as the total number of runs divided by the total number of subjects. The average total ROH length is the sum of the length of each individual ROH per participant and was calculated by dividing the total ROH length by the number of individuals. The average ROH length was calculated by dividing the total genomic length of the ROHs by the total number of ROHs per individual. To calculate FROH, a percentage of homozygosity was calculated by summing ROHs>1Mb across the covered autosomal genome and dividing by the total autosomal base pairs represented in the SNP data (8). Specifically, the summed length of identified ROHs were divided by a factor of 2772.7 and subsequently converted to a percent by multiplying the dividend by 100. A factor of 2772.7 is the number of megabases covered by SNPs after imputation, which was calculated by summing the distance between the first and the last available consecutive SNP on each chromosomal arm for each of the 22 autosomes. Examination of the FROH distribution revealed that it was highly right-skewed. To facilitate the analysis, data were transformed using a rank-based inverse normal transformation.

Assuming the dispersion of general homozygosity indices similar to the observed measures, a study with 24,320 subjects including 12,123 individuals with CAD and 12,197 CAD-free controls has ≈80% power to detect a case-control difference of ≈0.25 in ROHs number, ≈4.7 kb in average ROH length and ≈520 kb in average total ROH length.

Definition of consensus ROH

It is not expected that a ROH will start and end at the same base pair positions for each individual, so we needed to define the size of the overlapping ROHs. For each study the size of a ROH was

determined as the consensus region of a homozygous run of SNPs (>1MB in length) overlapping between at least 5 individuals within a cohort. Consensus ROHs were then separated into groups and the numbers of cases and controls with the ROH were counted using PLINK in each cohort. This ROH definition was applied to each study and provided a set of consensus ROHs. Information on the size, location (start/end of the consensus ROH) and the number of SNPs in each consensus ROH were calculated for each study.

Identification of overlapping consensus ROHs

We combined these consensus ROHs from individual studies to identify regions overlapping between studies. This was achieved by comparing the positions of the consensus ROHs from each study against those from all other cohorts; this was repeated until we had identified how many studies overlapped for each individual consensus ROH. We allowed each individual study consensus ROH to overlap with more than one consensus ROHs from another individual study to allow longer ROHs to be included in multiple overlapping consensus ROHs. We then combined the numbers of cases and controls across studies with the overlapping consensus ROH.

Statistical analysis

Genetic architecture of homozygosity measures

We generated QQ plots to assess the quality of the data in each individual study. A regression analysis using individual level participant data adjusting for cohort, sex and age where possible was performed for the homozygosity measures. Age and sex were the only additional variables available in all studies. Additional sensitivity analyses using a rank-based inverse normal transformation of general homozygosity measures was performed to assess the influence of outliers. Further sensitivity analyses were conducted to exclude the effect of a cohort with different cell sources of DNA in cases and controls. Finally, to account for potential heterogeneity between individual studies in the combined analysis of data we first fitted a mixed model with cohort as a random effect in the analysis of association between CAD and four homozygosity measures. Secondly, we introduced a term of interaction between cohort and the outcome (CADxcohort or FROHxcohort) into the regression

models that examined association between CAD and each of four homozygosity measures. Further sensitivity association analyses were conducted for those measures of homozygosity that showed some evidence of heterogeneity between populations (defined as significance of interaction term at $P<0.05$). These sensitivity analyses examined the magnitude and the significance of association between homozygosity measures and CAD after exclusion of cohorts with statistically significant interaction terms. Each of the sensitivity analyses was conducted using regression models fitted with cohort as both fixed and random effect.

Association analysis between overlapping consensus ROHs and CAD

Overlapping consensus ROHs across populations were analysed using logistic regression with CAD as the response with adjustment for sex, age where possible, study and ROH (presence or absence) status (as the predictor of interest). To account for multiple testing we used a Bonferroni correction in this analysis ($P=2.94x10^{-6}$). A binomial test was used to examine if there was a deviation from the expected distribution (50/50) of overlapping consensus ROHs showing increased (OR>1) and decreased (OR<1) risk of CAD.

Gene expression analysis and ROHs

Approximately 600,000 directly genotyped SNPs were available in the Cardiogenics Study. For the sake of consistency with the other examined studies, imputation was performed based on HapMapII CEU build 36. The ROH definitions and the principles of genome-wide homozygosity analysis were similar to the one conducted in Stage A of this project (Figure 1).

Information on gene probes underneath consensus ROHs in monocytes (758 individuals) and macrophages (614 individuals) was extracted *in silico* from the microarray-based experiment conducted in the Cardiogenics Study reported previously (35-36). The comparative analysis was conducted at the probe level in *cis*, only assessing genes within 500kb of a ROH, by using linear regression adjusted for age, sex, recruitment center and the status (presence or absence) for consensus ROHs. After quality control filters, information on 11,336 gene probes was available to examine differences in expression of genes in monocytes and macrophages between individuals with ROHs

versus those without ROHs. Results were first adjusted using genomic control. We then calculated false discovery rate (FDR) q-values for all identified ROH-mRNA associations using a method by Storey and Tibshirani appropriate for correlated gene expression data (37). Associations with a FDR q-value of <0.01 were considered statistically significant.

All statistical analyses were undertaken in STATA v12.

Results

General characteristics of the study cohorts

A total of 24,320 individuals of European ancestry from 11 populations were included. The key characteristics of the cohorts (including definition of CAD in each study) used in the homozygosity analysis are summarised in Table S1.

General characteristics of homozygosity measures

The summary and distribution of homozygosity measures in each population are shown in Table S3 and Figure S1. A joint analysis of all subjects revealed on average 31.84±8.44 ROHs in autosomal DNA. The stretches of homozygous SNPs had an average length of 1360.58±127.19 kb and they covered on average a total length of 43.67±15.59 Mb. The number and length of ROHs per individual ranged from 4-276 and 1-29.4 Mb (respectively), in the overall sample.

Comparison of homozygosity measures between individuals with CAD and CAD-free controls

The distribution of average ROH number, average ROH length and average total ROH length for cases and controls in each population are shown in Figure S2. As a general trend, the distributions of homozygosity measures were comparable between cases and controls from the same populations.

The combined analysis of 11 populations revealed statistically significant differences in homozygosity levels between individuals with CAD and controls after adjustment for study, sex and age (Table 1). On average, individuals with CAD had 0.63 ROHs more than controls (β=0.63, 95% CI: 0.42-0.83,

P=1.49x10^{-9}). The length of ROHs in individuals with CAD was on average 4.50 kb longer compared to controls (β=4.50, 95% CI: 0.85-8.15, P=0.016). The average total length of ROHs in the autosomal genome was 1046.92kb greater in individuals with CAD than controls (β=1046.92, 95% CI: 634.37-1459.48, P=6.61x10^{-7}). Logistic regression analysis revealed that every 1 standard deviation increase in FROH was associated with a 13% increase in CAD risk (OR=1.13, 95% CI: 1.09-1.17, P=1.57x10^{-11}). The association results were similar using cohort as both fixed (Table 1) and random effect (Table S4).

We have conducted a number of further sensitivity analyses to confirm robustness of our findings. Firstly, using a rank based inverse normal transformation for ROH number, average and total length of ROHs we confirmed that outliers had no noticeable influence on the estimates of association between CAD and these homozygosity measures (data not shown). Secondly, exclusion of a proportion of WTCCC controls whose DNA was extracted from cell lines (the British 1958 Birth Cohort) rather than peripheral blood (like the rest of cases and controls) had no major impact on the significance of our data (Table S5). Thirdly, accounting for some heterogeneity between populations in the analysis of association between CAD and average ROH number and FROH did not affect the significance of our findings. Indeed; exclusion of cohorts with significant interaction terms did not reduce the magnitude or statistical significance of association between CAD and average ROH number or FROH (Table S6). In fact, correcting for the residual heterogeneity has increased the statistical significance of the findings.

Identification and characterization of the overlapping consensus ROHs

A joint analysis of all consensus ROHs identified across 11 examined populations revealed a total of 16,989 overlapping consensus ROHs shared by at least 2 populations. An example of an overlapping consensus ROH segment is illustrated in Figure S3. These overlapping consensus ROHs were further classified into four groups based on their SNP enrichment (Table S7).

Association between individual overlapping consensus ROHs and CAD

After correction for multiple testing calculated at $P=2.94x10^{-6}$, none of the 16,989 identified overlapping consensus ROHs were associated with CAD. The most significant association between an overlapping consensus ROH on chromosome 3 (present in 138 subjects from 7 populations) was assigned a nominal level of statistical significance of $P=1.99x10^{-4}$. Details of the 10 most statistically significant overlapping consensus ROHs are listed in Table S8.

Overlapping consensus ROHs and CAD – aggregate burden analysis

Based on the magnitude of crude OR from the analysis of association with CAD, each of 16,989 overlapping consensus ROHs was classified as potentially increasing (OR>1.0) or decreasing (OR<1.0) risk of CAD. Under a null hypothesis of no association between overlapping consensus ROHs and CAD, the distribution of overlapping consensus ROHs between both categories should be even (50/50). However, a clear deviation from this expected distribution was apparent – there was an excess of overlapping consensus ROHs increasing risk of CAD over those classified as potentially protective (55/45), a difference inconsistent with chance ($P=2.69x10^{-33}$) (Table 2). The same over-representation of ROHs increasing the risk of CAD was apparent in each category of overlapping consensus ROHs after stratification based on number of SNPs in each consensus segment (Table 2).

Association between consensus ROHs and gene expression measures in human monocytes and macrophages

To explore if the presence of individual consensus ROHs is related to expression of genes that map onto or near specific ROHs, we combined genome-wide consensus ROHs with data from monocyte and macrophage transcriptome profiling in the Cardiogenics Study. In brief 11,336 gene probes were expressed in human monocytes and macrophages. A total of 3,223 consensus ROHs were identified in the genome-wide homozygosity analysis of Cardiogenics. After FDR-based correction for multiple testing (q-value <0.01), 44 consensus ROH-mRNA associations retained their statistical significance in monocytes (Figure 2, Table 3). The most significant association was identified between ROH on chromosome 16 and expression of dihydrouridine synthase 2 (*DUS2L* [MIM 609707]) – subjects with this consensus ROH had on average 0.23 (95%CI: 0.19-0.26) lower expression of *DUSL2* in

monocytes when compared to those without this ROH after adjustment for age, sex and centre of recruitment ($P=5.74x10^{-30}$, $q=6.51x10^{-26}$). After the correction for multiple testing, 17 consensus ROH-mRNA associations retained their statistical significance in macrophages (Figure 2, Table 4). The most significant association was identified on chromosome 3 with the expression of WD repeat domain 6 (*WDR6* [MIM 606031]) – subjects with this consensus ROH had on average 0.19 (95%CI: 0.14-0.25) lower expression of *WDR6* in macrophages when compared to those without this ROH after adjustment for age, sex and centre of recruitment ($P=4.57x10^{-10}$, $q=5.18x10^{-6}$).

Discussion

Our analysis of homozygosity produced several important findings. Firstly, we provided evidence for enrichment of homozygosity in individuals with CAD when compared to CAD-free controls. Secondly, while none of the identified overlapping consensus ROHs was associated with CAD individually, their aggregate burden analysis showed over-representation of overlapping consensus ROHs favouring increased risk of CAD when compared to those showing the opposite direction of association with CAD. Taken together, these data indicate that homozygosity may be an important risk factor for CAD and suggest that the accumulation of multiple recessive variants may be an important component of the genetic architecture of CAD. Finally, the analysis of human monocyte and macrophage transcriptomes suggested that many individual consensus ROHs may carry biologically active variants with a potential to affect expression of genes located either within these ROHs or their close proximity.

Our analysis of homozygosity measures provided further evidence for presence of common ROHs in outbred populations (8, 11-12). Detection of ROHs in homozygosity mapping tends to vary along the genome due to differences in informativeness of haplotypes and differences in haplotype genealogies (38). The relative performance of homozygosity mapping is influenced by population demographic processes and the extent of selection against causal variants (38). ROHs were widely distributed across the entire genome and were present on each human autosome. The number of ROHs was a function of chromosomal length and the average ROH number and average total length of ROHs were

dependent on chromosomal size, showing good agreement with previous studies in populations of European ancestry (29).

Inbred individuals tend to have higher rates of congenital disorders and lower survival and fertility rates (inbreeding depression) (7, 39). FROH is a measure of inbreeding effects and correlates most highly with the homozygous mutational load, the putative causal mechanism underlying inbreeding depression (34). This homozygosity measure has low prediction error variance, especially when SNP density is high (14). However, given the small variation in genome-wide FROH in unrelated individuals, large sample sizes are necessary to detect inbreeding depression for likely effect sizes (14). Previous studies investigating the effects of FROH on human complex traits with relatively small sample sizes have failed to find significant inbreeding effects (28, 34, 40-43); most likely because they were underpowered. Furthermore, very small studies (n<1,000) that did find significant inbreeding depression effects using FROH (13) may greatly overestimated the size of effects. The sample size of this study provides a well-powered tool for finding signatures of homozygosity on CAD risk. As a result, we were able to quantify a measure of global homozygosity in relation to CAD – an estimated 13% increase in CAD risk per 1 standard deviation increase in FROH.

We appreciate that our analysis did not reveal statistically significant associations between individual overlapping consensus ROHs and CAD, possibly because many of them have low population frequency and this may affect the power of the analysis. Indeed, whilst some overlapping consensus homozygous sequences are common and may coincide with previously identified ROH islands (29) or long haplotypes, a majority of the consensus regions overlapping across populations occur in low frequencies (<5%) or indeed are very rare (<1%).

However, overlapping consensus ROH–based analysis provided several important insights into the genetic architecture of CAD. Firstly, it revealed that none of the identified overlapping consensus regions were exclusive to CAD cases. Secondly, it showed that none of the identified overlapping consensus regions have been implicated in previous CAD GWA studies. This is not perhaps surprising given that GWA studies have analysed data under an additive model of inheritance, whilst ROH analysis assumes recessive mode of inheritance. Thirdly, the aggregate burden analysis of overlapping consensus ROHs showed an excess of the regions favouring increased risk of CAD over

those that tend to protect against it supporting the hypothesis that individuals with CAD may have accumulated more recessive variants than controls. This cumulative excess of ROHs with dispersed distribution across the genome rather than effects of specific variants recessively inherited by a subset of affected individuals (44) appears a more likely driver for the increased homozygosity in CAD. Additional work is needed to unravel the exact synergistic effects of multiple recessive variants on global homozygosity levels and their relevance to CAD. Finally, within the constraints of interpretational limitations discussed above, the overlapping consensus ROH-based analysis provides a useful example of comparing individual ROHs across studies.

Our study also examined the effects of consensus ROHs on the expression of genes in human cells relevant to atherosclerosis. One of the most obvious explanations for the biological meaning of the associations between consensus ROHs and the expression of genes underneath them or in close proximity to them is that they are signatures of functionally active recessive variants with a potential to affect transcription in monocytes and macrophages. Neither of the identified genes showed the immediately obvious biological relevance to atherosclerosis or CAD. Indeed, the most significant association in monocytes was identified with the expression of *DUS2L*. This gene encodes a cytoplasmic protein that catalyzes the conversion of uridine residues to dihydrouridine in the D-loop of tRNA. The encoded protein may affect the rate of translation by inhibiting an interferon-induced protein kinase (45). The most significant association in macrophages was with the expression of *WDR6*. This gene encodes a member of the WD repeat protein family implicated in regulation of cell growth arrest (46). Future studies will be necessary to elucidate the molecular and clinical mechanisms of these associations.

We should mention that definition of ROHs varies across studies in the published literature. The discrepancy between definitions of ROHs across studies makes their comparisons difficult (47). In order to combine the ROHs information across several studies we made use of both directly genotyped and imputed SNPs to increase the coverage of the genome and extract the maximum possible information. The use of imputed SNPs helped to increase similarity of genomic coverage across studies and made the data more comparable as ROHs were defined based on similar SNP sets. We applied a requirement that a genotype could only be called if the posterior probability was >90%

and then applied call rate filters to remove SNPs which were called in <95% of individuals. This has an effect of thinning the imputed data, but is not equivalent to having the same effect as LD pruning. We chose not to use LD-driven elimination of SNPs from our datasets as such pruning may act as a potential confounder in comparative ROH analyses of different populations – the local level of LD determines the effective number of SNPs used for ROH definition (29). Although LD pruning brings the benefit of selecting independent SNPs, previous studies have shown that LD-based pruning may reduce informativeness of datasets and lead to a loss in power to detect ROHs (48). Many recently published studies defined a ROH to include a minimum number of 50 – 65 SNPs for pruned and unpruned data (44, 49-53). Therefore, we have accounted for imputation-driven increase in SNP density by extension of the minimal number of SNPs that define a ROH to 100.

We acknowledge the potential confounding of undetected clonal mosaicism manifesting as chromosomal abnormalities (i.e. uniparental disomy) that may mirror runs of homozygosity in GWA based on DNA extracted from peripheral blood. Indeed, both Laurie et al. (54) and Jacobs et al. (55) revealed that clones of cells with such chromosomal anomalies are present in free of cancer (apparently healthy) individuals included in GWA studies and that increasing age is associated with augmented risk of clonal mosaicism in the human genome (54-56). This in essence means that comparative analysis of homozygosity in DNA extracted from peripheral blood between cases and controls that differ in age may be potentially affected by increased rates of such clonal chromosomal abnormalities (and thus inflated homozygosity measures) in the older group. Whilst we were not able to directly quantify the prevalence of such abnormalities in the populations included in this project, we reason that a correction for such a potential confounding would further increase rather than reduce the significance of our association findings. Indeed, CAD-free controls were generally much older than individuals with CAD in a majority of studies included in this project. Thus, a correction for age-driven clonal mosaicism-related chromosomal abnormalities interpreted as ROHs in genome-wide analysis would reduce the rates of homozygosity in the older groups (mostly - CAD-free controls) further increasing the case-control difference. To this end, we also acknowledge the different cell types as a source of DNA for experiments in one of 11 cohorts included in this project. Indeed, a part of control group in WTCCC (58BC cohort) was genotyped using DNA extracted from cell lines (23).

DNA extracted from cell lines may be associated with higher rates of chromosomal aberrations that may resemble ROHs in GWAs. However, we confirmed that exclusion of those subjects had no major effect on findings from our analysis of association between CAD and homozygosity measures.

Summary

Genome-wide homozygosity analysis revealed statistically significant differences in the genome-wide homozygosity levels between individuals with CAD and CAD-free controls. The aggregate burden analysis of overlapping consensus ROHs showed their over-representation amongst subjects with CAD, suggesting that accumulation of recessive variants may increase the risk of CAD. Finally, the presence of associations between consensus ROHs and gene expression in human monocytes and macrophages suggest that many individual ROHs may be signatures of biologically active recessive variants with a potential to regulate transcription.

Supplemental data

Supplemental data include three figures and eight tables.

Acknowledgments

This study was supported by the Alumni Association of University of Leicester PhD studentship and International Mentoring Travel Award by American Heart Association to PC. MT is supported by the British Heart Foundation. NJS holds a personal chair supported by the British Heart Foundation and is a UK NIHR senior investigator. CPN is funded by the British Heart Foundation. The Cardiogenics project was supported by the European Union 6th Framework Program (LSHM-CT-2006-037593). The UKBS collection of Common Controls has been funded by the Wellcome Trust grant 084183/Z/07/Z and by NIHR programme grant to NHSBT (RP-PG-0310-1002). The collection was established as part of the WTCCC. We would like to thank the participants and members of the

CARDIoGRAM and the Cardiogenics consortia. The remaining authors report no conflicts. Members of the Cardiogenics Consortium are listed in the Supplement.

References

1. CARDIoGRAMplusC4D Consortium, Deloukas, P., Kanoni, S., Willenborg, C., Farrall, M., Assimes., T.L., Thompson, J.R., Ingelsson, E.,Saleheen, D., Erdmann, J., et al. (2013). Large-scale association analysis identifies new risk loci for coronary artery disease. Nat Genet. *45*,25-33.

2. Ismail, J., Jafar, T.H., Jafary, F.H., White, F., Faruqui, A.M., and Chaturvedi, N. (2004). Risk factors for non-fatal myocardial infarction in young South Asian adults. Heart *90*,259-263.

3. Shami, S.A., Qaisar, R., and Bittles, A.H. (1991). Consanguinity and adult morbidity in Pakistan. Lancet *338*,954.

4. Puzyrev, V.P., Lemza, S.V., Nazarenko, L.P., and Panphilov, V.I. (1992). Influence of genetic and demographic factors on etiology and pathogenesis of chronic disease in north Siberian aborigines. Arctic Med Res. *51*,136-42.

5. Rudan, I., Rudan, D., Campbell, H., Carothers, A., Wright, A., Smolej-Narancic, N., Janicijevic, B., Jin, L., Chakraborty, R., Deka, R., et al. (2003). Inbreeding and risk of late onset complex disease. J Med Genet. *40*,925-32.

6. Campbell, H., Carothers, A.D., Rudan, I., Hayward, C., Biloglav, Z., Barac, L., Pericic, M., Janicijevic, B., Smolej-Narancic, N., Polasek, O., et al. (2007). Effects of genome-wide heterozygosity on a range of biomedically relevant human quantitative traits. Hum Mol Genet. *16*,233-41.

7. Charlesworth, D., and Willis, J.H. (2009). The genetics of inbreeding depression. Nat Rev Genet. *10*,783-96.

8. McQuillan, R., Leutenegger, A.L., Abdel-Rahman, R., Franklin, C.S., Pericic, M., Barac-Lauc, L., Smolej-Narancic, N., Janicijevic, B., Polasek, O., Tenesa, A., et al. (2008). Runs of homozygosity in European populations. Am J Hum Genet. *83*,359-72.

9. Szpiech, Z.A., Xu, J., Pemberton, T.J., Peng, W., Zöllner, S., Rosenberg, N.A., Li, J.Z. (2013). Long runs of homozygosity are enriched for deleterious variation. Am J Hum Genet. *93*,90-102.

10. Broman, K.W., and Weber, J.L. (1999). Long homozygous chromosomal segments in reference families from the centre d'Etude du polymorphisme humain. Am J Hum Genet. *65*,1493-500.

11. Li, L.H., Ho, S.F., Chen, C.H., Wei, C.Y., Wong, W.C., Li, L.Y., Hung, S.I., Chung, W.H., Pan, W.H., Lee, M.T., et al. (2006). Long contiguous stretches of homozygosity in the human genome. Hum Mutat. *27*,1115-21.

12. Gibson, J., Morton, N.E., and Collins, A. (2006). Extended tracts of homozygosity in outbred human populations. Hum Mol Genet. *15*,789-95.

13. Lencz, T., Lambert, C., DeRosse, P., Burdick, K.E., Morgan, T.V., Kane, J.M., Kucherlapati, R., and Malhotra, A.K. (2007). Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia. Proc Natl Acad Sci. *104*,19942-7.

14. Keller, M.C., Simonson, M.A., Ripke, S., Neale, B.M., Gejman, P.V., Howrigan, D.P., Lee, S.H., Lencz, T., Levinson, D.F., Sullivan, P.F., et al. (2012). Runs of homozygosity implicate autozygosity as a schizophrenia risk factor. PLoS Genet. *8*,e1002656.

15. Power, R.A., Keller, M.C., Ripke, S., Abdellaoui, A., Wray, N.R., Sullivan, P.F., MDD PGC Working Group, and Breen, G. (2014). A recessive genetic model and runs of homozygosity in major depressive disorder. Am J Med Genet B Neuropsychiatr Genet. *165B.*157-66.

16. Schunkert, H., König, I.R., Kathiresan, S., Reilly, M.P., Assimes, T.L., Holm, H., Preuss, M., Stewart, A.F., Barbalic, M., Gieger, C., et al. (2011). Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. Nat Genet. *43,*333-8.

17. Samani, N.J., Erdmann, J., Hall, A.S., Hengstenberg, C., Mangino, M., Mayer, B., Dixon, R.J., Meitinger, T., Braund, P., Wichmann, H.E., et al. (2007). Genomewide association analysis of coronary artery disease. N Engl J Med. *357,*443-53.

18. Erdmann, J., Grosshennig, A., Braund, P.S., König, I.R., Hengstenberg, C., Hall, A.S., Linsel-Nitschke, P., Kathiresan, S., Wright, B., Trégouët, D.A., et al. (2009). New susceptibility locus for coronary artery disease on chromosome 3q22.3. Nat Genet. *41,*280-2.

19. Erdmann, J., Willenborg, C., Nahrstaedt, J., Preuss, M., König, I.R., Baumert, J., Linsel-Nitschke, P., Gieger, C., Tennstedt, S., Belcredi, P.,et al. (2011). Genome-wide association study identifies a new locus for coronary artery disease on chromosome 10p11.23. Eur Heart J. *32,*158-68.

20. Yusuf, S., Hawken, S., Ounpuu, S., Dans, T., Avezum, A., Lanas, F., McQueen, M., Budaj, A., Pais, P., Varigos, J., et al. (2004). Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. Lancet. *364,*937-52.

21. McPherson, R., Pertsemlidis, A., Kavaslar, N., Stewart, A., Roberts, R., Cox, D.R., Hinds, D.A., Pennacchio, L.A., Tybjaerg-Hansen, A., Folsom, A.R.,et al. (2007). A common allele on chromosome 9 associated with coronary heart disease. Science *316*,1488-1491.

22. Lehrke, M., Millington, S.C., Lefterova, M., Cumaranatunge, R.G., Szapary, P., Wilensky, R., Rader, D.J., Lazar, M.A., and Reilly, M.P. (2007). CXCL16 is a marker of inflammation, atherosclerosis, and acute coronary syndromes in humans. J Am Coll Cardiol. *49*,442-9.

23. Burton, P.R., Clayton, D.G., Cardon, L.R., Craddock, N., Deloukas, P., Duncanson, A., Kwiatkowski, D.P., McCarthy, M.I., Ouwehand, W.H., Samani, N.J., et al. (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. *447*,661-78.

24. Auton, A., Bryc, K., Boyko, A.R., Lohmueller, K.E., Novembre, J., Reynolds, A., Indap, A., Wright, M.H., Degenhardt, J.D., Gutenkunst, R.N., et al. (2009). Global distribution of genomic diversity underscores rich complex history of continental human populations. Genome Res. *19*,795-803.

25. Curtis, D., Vine, E., and Knight, J. (2008). Study of regions of extended homozygosity provides a powerful method to explore haplotype structure of human populations. Ann Hum Genet. *72*,261-78.

26. MacLeod, I.M., Meuwissen, T.H., Hayes, B.J., and Goddard, M.E. (2009). A novel predictor of multilocus haplotype homozygosity: comparison with existing predictors. Genet Res *91*,413-26.

27. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. *81*,559-75.

28. Nalls, M.A., Simon-Sanchez, J., Gibbs, J.R., Paisan-Ruiz, C., Bras, J.T., Tanaka, T., Matarin, M., Scholz, S., Weitz, C., Harris, T.B.,et al. (2009). Measures of autozygosity in decline: globalization, urbanization, and its implications for medical genetics. PLoS Genet. *5*,e1000415.

29. Nothnagel, M., Lu, T.T., Kayser, M., and Krawczak, M. (2010). Genomic and geographic distribution of SNP-defined runs of homozygosity in Europeans. Hum Mol Genet. *19*, 2927-35.

30. Abecasis, G.R., Ghosh, D., and Nichols, T.E. (2005). Linkage disequilibrium: ancient history drives the new genetics. Hum Hered. *59*,118-24.

31. Wall, J.D., and Pritchard, J.K. (2003). Haplotype blocks and linkage disequilibrium in the human genome. Nat Rev Genet. *4*,587-97.

32. Abecasis, G.R., Noguchi, E., Heinzmann, A., Traherne, J.A., Bhattacharyya, S., Leaves, N.I., Anderson, G.G., Zhang, Y., Lench, N.J., Carey, A., et al. (2001). Extent and distribution of linkage disequilibrium in three genomic regions. Am J Hum Genet. *68*,191-197.

33. Reich, D.E., Cargill, M., Bolk, S., Ireland, J., Sabeti, P.C., Richter, D.J., Lavery, T., Kouyoumjian, R., Farhadian, S.F., Ward, R., et al. (2001). Linkage disequilibrium in the human genome. Nature *411*,199-204.

34. McQuillan, R., Eklund, N., Pirastu, N., Kuningas, M., McEvoy, B.P., Esko, T., Corre, T., Davies, G., Kaakinen, M., Lyytikäinen, L.P., et al. (2012). Evidence of inbreeding depression on human height. PLoS Genet. *8*, e1002655.

35. Heinig, M., Petretto, E., Wallace, C., Bottolo, L., Rotival, M., Lu, H., Li, Y., Sarwar, R., Langley, S.R., Bauerfeind, A., et al. (2010). A trans-acting locus regulates an anti-viral expression network and type 1 diabetes risk. Nature *467*,460-464.

36. Rotival, M., Zeller, T., Wild, P., Maouche, S., Szymczak, S., Schillert, A., Castagné, R., Deiseroth, A., Proust, C., Brocheton, J., et al. (2011). Integrating Genome-Wide Genetic Variations and Monocyte Expression Data Reveals Trans-Regulated Gene Modules in Humans. PLoS Genet. *7*,e1002367.

37. Storey, J.D., and Tibshirani, R. (2003). Statistical significance for genome wide studies. PNAS. *100*, 9440–9445

38. Browning, S.R., and Thompson, E.A. (2012). Detecting rare variant associations by identity-by-descent mapping in case-control studies. Genetics *190*,1521-31.

39. Keller, M.C., Visscher, P.M., and Goddard, M.E. (2011). Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. Genetics *189*, 237-249.

40. Spain, S.L., Cazier, J.B., CORGI Consortium, Houlston, R., Carvajal-Carmona, L., and Tomlinson I. (2009). Colorectal cancer risk is not associated with increased levels of homozygosity in a population from the United Kingdom. Cancer Res. *69*, 7422-7429.

41. Vine, A.E., McQuillin, A., Bass, N.J., Pereira, A., Kandaswamy, R., Robinson, M., Lawrence, J., Anjorin, A., Sklar, P., Gurling, H.M.,et al. (2009). No evidence for excess runs of homozygosity in bipolar disorder. Psychiatr Genet. *19*,165-70.

42. Enciso-Mora, V., Hosking, F.J., and Houlston, R.S. (2010). Risk of breast and prostate cancer is not associated with increased homozygosity in outbred populations. Eur J Hum Genet. *18*, 909-14.

43. Hosking, F.J., Papaemmanuil, E., Sheridan, E., Kinsey, S.E., Lightfoot, T., Roman, E., Irving, J.A., Allan, J.M., Taylor, M., Tomlinson, I.P., et al. (2010). Genome-wide homozygosity signatures and childhood acute lymphoblastic leukemia risk. Blood *115*, 4472–4477.

44. Ghani, M., Sato, C., Lee, J.H., Reitz, C., Moreno, D., Mayeux, R., St George-Hyslop, P., and Rogaeva, E. (2013). Evidence of recessive Alzheimer disease loci in a Caribbean Hispanic data set: genome-wide survey of runs of homozygosity. JAMA Neurol. 70, 1261-7.

45. Mittelstadt, M., Frump, A., Khuu, T., Fowlkes, V., Handy, I., Patel, C.V., Patel, R.C. (2008). Interaction of human tRNA-dihydrouridine synthase-2 with interferon-induced protein kinase PKR. Nucleic Acids Res. *36*, 998-1008.

46. Xie, X., Wang, Z., Chen, Y. (2007). Association of LKB1 with a WD-repeat protein WDR6 is implicated in cell growth arrest and p27 (Kip1) induction. Mol Cell Biochem. *301*,115-22.

47. Howrigan, D.P., Simonson, M.A., and Keller, M.C. (2011). Detecting autozygosity through runs of homozygosity: a comparison of three autozygosity detection algorithms. BMC Genomics *12*,460.

48. Ku, C.S., Naidoo, N., Teo, S.M., and Pawitan, Y. (2011). Regions of homozygosity and their impact on complex diseases and traits. Hum Genet *129*,1–15.

49. Di Gaetano, C., Fiorito, G., Ortu, M.F., Rosa, F., Guarrera, S., Pardini, B., Cusi, D., Frau, F., Barlassina, C., Troffa, C., et al. (2014). Sardinians genetic background explained by runs of homozygosity and genomic regions under positive selection. PLoS One *9*,e91237.

50. Heron, E.A., Cormican, P., Donohoe, G., O'Neill, F.A., Kendler, K.S., Riley, B.P., Wellcome Trust Case Control Consortium 2, Gill, M., Corvin, A.P., Morris, D.W. (2014). No evidence that runs of homozygosity are associated with schizophrenia in an Irish genome-wide association dataset. Schizophr Res. *154*,79-82.

51. Lin, P.I., Kuo, P.H., Chen, C.H., Wu, J.Y., Gau, S.S., Wu, Y.Y., and Liu, S.K. (2013). Runs of homozygosity associated with speech delay in autism in a taiwanese han population: evidence for the recessive model. PLoS One. *16*,e72056.

52. McWhirter, R.E., Thomson, R.J., Marthick, J.R., Rumbold, A.R., Brown, M.A., Taylor-Thomson, D., Maypilama, E.L., Condon, J.R., and Dickinson, J.L. (2014). Runs of homozygosity and a cluster of vulvar cancer in young Australian Aboriginal women. GynecolOncol. *133*,421-6.

53. Wang, C., Xu, Z., Jin, G., Hu, Z., Dai, J., Ma, H., Jiang, Y., Hu, L., Chu, M., Cao, S., et al. (2013). Genome-wide analysis of runs of homozygosity identifies new susceptibility regions of lung cancer in Han Chinese. J Biomed Res. *27*,208-14.

54. Laurie, C.C., Laurie, C.A., Rice, K., Doheny, K.F., Zelnick, L.R., McHugh, C.P., Ling, H., Hetrick, K.N., Pugh, E.W., Amos, C., et al. (2012). Detectable clonal mosaicism from birth to old age and its relationship to cancer. Nat Genet. *44*, 642-50.

55. Jacobs, K.B., Yeager, M., Zhou, W., Wacholder, S., Wang, Z., Rodriguez-Santiago, B., Hutchinson, A., Deng, X., Liu, C., Horner, M.J., et al. (2012). Detectable clonal mosaicism and its relationship to aging and cancer. Nat Genet. *44*, 651-8.

56. Machiela, M.J., Chanock, S.J. (2013). Detectable clonal mosaicism in the human genome. Semin Hematol. *50*, 348-59.

Figure titles and legends

Figure 1: Overview of the project strategy.

Figure 2: Analysis of association between the presence of consensus ROHs and gene expression in the Cardiogenics Study – genome-wide signal-intensity plot. A – human monocytes, B – human macrophages, the Y axis shows the logarithmic level of statistical significance (P-value) for association of each consensus ROH with expression, X-axis – 22 autosomal chromosomes in numerical order

Tables

Table 1: Differences in homozygosity measures between individuals with coronary artery disease and controls – combined analysis.

Measure	Un-adjusted analysis			Age-adjusted analysis			Age- and sex-adjusted analysis		
	β-coefficient/ OR	95% CI	P-value	β-coefficient/ OR	95% CI	P-value	β-coefficient/ OR	95% CI	P-value
Average ROH number	0.39	0.20, 0.57	3.92×10^{-5}	0.67	0.47, 0.86	4.12×10^{-11}	0.63	0.42, 0.83	1.49×10^{-9}
Average ROH length (kb)	2.51	-0.80, 5.82	0.14	5.19	1.64, 8.75	0.004	4.50	0.85, 8.15	0.016
Average total length of ROHs (kb)	688.11	314.44, 1061.78	3.08×10^{-4}	1145.15	743.95, 1546.35	2.24×10^{-8}	1046.92	634.37, 1459.48	6.61×10^{-7}
FROH	1.07	1.03, 1.10	8.63×10^{-5}	1.14	1.10, 1.18	2.21×10^{-13}	1.13	1.09, 1.17	1.57×10^{-11}

Data for number, average length and total length of homozygosity runs (ROHs) are β-coefficients [with respective confidence intervals and level of statistical significance (P-value) from regressing homozygosity measures on case-control status with adjustment for cohort, age and sex (where appropriate)]; data on FROH (proportion of autosomal genome in ROHs) are expressed as odds ratios (OR) of coronary artery disease risk (with confidence intervals and level of statistical significance) with adjustment for cohort, age and sex (where appropriate); 95% CI – 95% confidence intervals, ROHs – runs of homozygosity.

Table 2: Frequency of overlapping consensus ROHs with a potentially increasing or decreasing risk of coronary artery disease – analysis stratified on number of single nucleotide polymorphisms in overlapping consensus ROHs.

	Overlapping consensus ROHs	Expected	Observed	P-value
Overall	↑ CAD risk	50% - 8494.5	54.6% - 9278	2.69×10^{-33}
	↓ CAD risk	50% - 8494.5	45.4% - 7711	
Group 0 2-9 SNPs	↑ CAD risk	50% - 926	55.1% - 1020	1.37×10^{-5}
	↓ CAD risk	50% - 926	44.9% - 832	
Group 1 10-49 SNPs	↑ CAD risk	50% - 3009	53.4% - 3214	1.33×10^{-7}
	↓ CAD risk	50% - 3009	46.6% - 2804	
Group 2 50-99 SNPs	↑ CAD risk	50% - 1704	55.0% - 1876	4.09×10^{-9}
	↓ CAD risk	50% - 1704	45.0% - 1532	
Group 3 100+ SNPs	↑ CAD risk	50% - 2855.5	55.5% - 3168	1.39×10^{-16}
	↓ CAD risk	50% - 2855.5	44.6% - 2543	

Data are counts and percentages, CAD – coronary artery disease, SNPs – single nucleotide polymorphism, overlapping consensus ROHs were classified as increasing (↑) and decreasing (↓) risk of CAD based on the magnitude of their odds ratio (OR) for CAD; OR>1.0 – increasing risk of CAD, OR<1.0 – decreasing risk of CAD, P-value – level of statistical significance from binomial test, the size (and thus the number of SNPs) in each overlapping consensus ROH depends on the length of the consensus sequence common for the studies – from 2 to 100+ SNPs were identified in these regions.

Table 3: Average mRNA expression differences in human monocytes between individuals with and without consensus ROHs – Cardiogenics Study.

Chr	Location (Mb)	SNPs	Individuals with consensus ROHs	Gene	β (95% CI)	P-value
16	663.38-663.69	6	454	*DUS2L*	-0.23 (-0.26, -0.19)	$5.74x10^{-30}$
16	669.09 669.17	4	228	*DUS2L*	-0.16 (-0.20, -0.12)	$4.55x10^{-13}$
21	293.68-296.11	91	48	*C21ORF7*	0.70 (0.51, 0.89)	$1.33x10^{-11}$
3	491.91-492.85	28	225	*WDR6*	-0.15 (-0.20, -0.11)	$4.67x10^{-11}$
15	414.96-415.58	23	331	*ZSCAN29*	0.07 (0.05, 0.09)	$1.81x10^{-10}$
15	414.27-414.76	18	325	*ZSCAN29*	0.07 (0.05, 0.09)	$1.93x10^{-10}$
15	413.87-413.97	2	322	*ZSCAN29*	0.07 (0.05, 0.09)	$4.22x10^{-10}$
3	491.75-491.76	2	233	*WDR6*	-0.14 (-0.18, -0.10)	$5.27x10^{-10}$
15	416.22-416.83	6	352	*ZSCAN29*	0.07 (0.05, 0.09)	$7.88x10^{-10}$
15	413.33-413.38	2	303	*ZSCAN29*	0.07 (0.05, 0.09)	$1.34x10^{-9}$
3	491.14-491.38	10	227	*WDR6*	-0.14 (-0.18, -0.10)	$1.63x10^{-9}$
15	417.06-417.07	2	353	*ZSCAN29*	0.06 (0.04, 0.08)	$3.62x10^{-9}$
3	487.46-490.15	50	227	*WDR6*	-0.14 (-0.18, -0.09)	$4.08x10^{-9}$
3	493.65 - 495.29	43	198	*WDR6*	-0.14 (-0.18, -0.09)	$1.53x10^{-8}$
16	660.34 - 660.47	3	467	*ATP6V0D1*	0.10 (0.06, 0.13)	$6.41x10^{-8}$

16	663.38 - 663.69	6	454	*ATP6V0D1*	0.09 (0.06, 0.13)	$8.29x10^{-8}$
3	486.37 - 486.47	8	242	*WDR6*	-0.12 (-0.16, -0.08)	$1.21x10^{-7}$
6	286.55 - 287.10	10	200	*ZNF193*	0.09 (0.06, 0.12)	$1.87x10^{-7}$
15	411.65 - 411.82	4	271	*ZSCAN29*	0.06 (0.04, 0.08)	$1.91x10^{-7}$
6	286.02 - 286.23	11	200	*ZNF193*	0.09 (0.06, 0.12)	$2.10x10^{-7}$
3	486.10 - 486.19	3	248	*WDR6*	-0.12 (-0.16, -0.07)	$3.37x10^{-7}$
16	669.09 - 669.17	4	228	*SLC7A6*	0.10 (0.06, 0.14)	$6.87x10^{-7}$
6	288.03 - 288.04	2	200	*ZNF193*	0.09 (0.06, 0.12)	$7.78x10^{-7}$
4	1037.94 - 1042.37	96	62	*MANBA*	-0.15 (-0.21, -0.10)	$8.14x10^{-7}$
16	658.30 - 658.86	10	452	*ATP6V0D1*	0.09 (0.05, 0.12)	$1.25x10^{-6}$
10	756.07 - 757.01	17	177	*VCL*	0.10 (0.06, 0.13)	$2.22x10^{-6}$
13	956.09 - 958.69	101	71	*UGCGL2*	0.11 (0.07, 0.15)	$2.49x10^{-6}$
6	284.18 - 284.55	14	212	*ZNF193*	0.08 (0.05, 0.11)	$3.00x10^{-6}$
6	284.56 - 285.08	35	212	*ZNF193*	0.08 (0.05, 0.11)	$3.96x10^{-6}$
16	670.10 - 670.32	10	138	*SLC7A6*	0.12 (0.07, 0.17)	$3.98x10^{-6}$
6	308.72 - 308.75	6	45	*HLA-C*	0.26 (0.16, 0.37)	$4.35x10^{-6}$
16	670.10 – 670.32	10	138	*DUS2L*	-0.12 (-0.17, -0.07)	$5.55\ x10^{-6}$
6	282.81 – 283.17	18	197	*ZNF193*	0.08 (0.05, 0.11)	$7.49\ x10^{-6}$
10	759.90 – 760.13	3	168	*VCL*	0.09 (0.05, 0.13)	$7.62\ x10^{-6}$

10	755.26 – 755.30	2	177	*VCL*	0.09 (0.05, 0.13)	$1.20x10^{-5}$
3	1592.98 – 1599.09	187	19	*RARRES1*	0.14 (0.08, 0.20)	$1.24x10^{-5}$
2	1354.79 – 1362.11	139	378	*MCM6*	-0.08 (-0.12, -0.05)	$1.67x10^{-5}$
20	332.89 – 333.03	3	204	*UQCC*	0.08 (0.05, 0.12)	$2.05x10^{-5}$
6	279.79 – 280.41	21	194	*ZNF193*	0.07 (0.04, 0.11)	$2.41x10^{-5}$
16	669.09 – 669.17	4	228	*SLC7A6*	0.09 (0.05, 0.13)	$2.68x10^{-5}$
16	663.38 – 663.69	6	454	*DPEP3*	0.07 (0.04, 0.10)	$2.70x10^{-5}$
10	761.02 – 761.36	8	144	*ADK*	-0.15 (-0.22, -0.09)	$3.19x10^{-5}$
16	670.10 – 670.32	10	138	*SLC7A6*	0.10 (0.06, 0.15)	$3.20x10^{-5}$
15	419.40 – 419.56	3	400	*ZSCAN29*	0.04 (0.02, 0.06)	$3.47x10^{-5}$

Chr – chromosome, SNPs – number of SNPs in overlapping consensus region, β – regression coefficient, 95% CI – 95% confidence intervals, P-value – level of statistical significance for difference in gene expression between individuals with and without consensus ROHs adjusted for demographic parameters and genomic control. All identified associations have a false discovery rate q-value <0.01. ***DUS2L*** – dihydrouridine synthase 2, ***C21orf7*** [MIM 611110] – MAP3K7 C-Terminal like, ***WDR6*** – WD repeat domain 6, ***ZSCAN29*** – zinc finger and SCAN domain containing 29, ***ATP6V0D1*** [MIM 607028] – ATPase, H+ transporting, lysosomal 38kDa V0 subunit d1, ***ZNF193*** [MIM 602246] – zinc finger and SCAN domain containing 9, ***SLC7A6*** [MIM 605641] – solute carrier family 7 member 6, ***MANBA*** [MIM 609489] – mannosidase beta a, lysosomal, ***VCL*** [MIM 193065] – vinculin, ***UGCGL2*** [MIM 605898] **–** UDP-Glucose ceramide glucosyltransferase-like 2, ***HLA-C*** [MIM 142840] – major histocompatibility complex, class I, C, ***RARRES1*** [MIM 605090] – retinoic acid receptor responder (tazarotene induced) 1, ***MCM6*** [MIM 601806] – minichromosome maintenance complex component 6, ***UQCC*** [MIM 611797] **–** ubiquinol-cytochrome c reductase complex assembly factor 3, ***DPEP3*** [MIM 609926] **–** dipeptidase 3, ***ADK*** [MIM 102750] **–** adenosine

kinase. Gene symbols appear more than once in the table because different consensus ROHs can be mapped onto the same gene array probe if there is more than one ROH within 100kb of the probe.

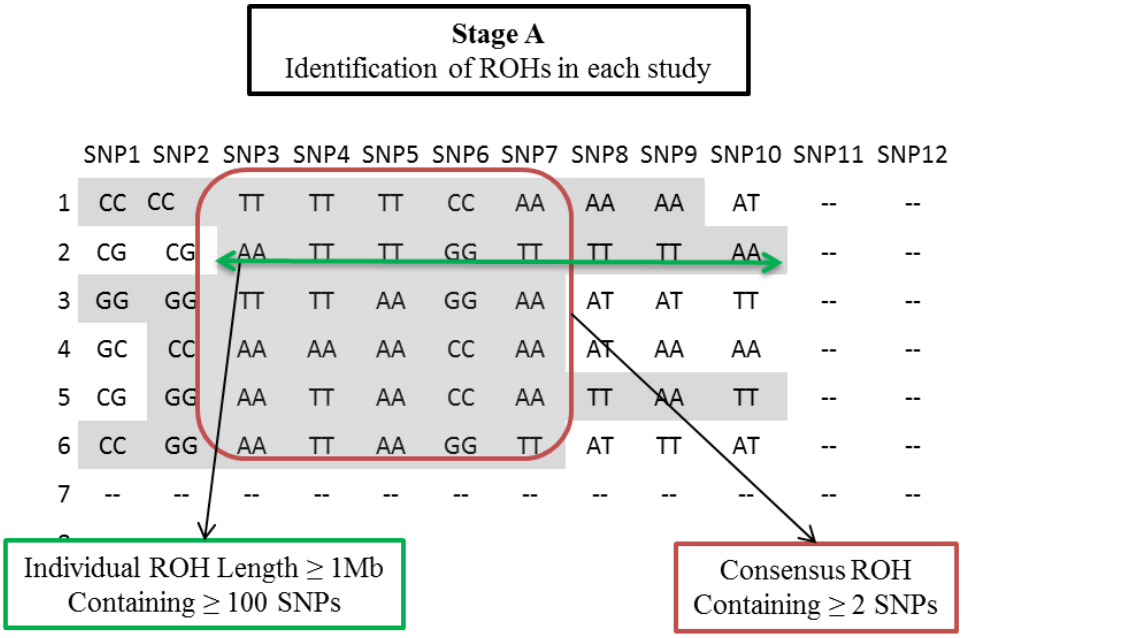
Table 4: Average mRNA expression differences in human macrophages between individuals with and without consensus ROHs – Cardiogenics Study.

Chr	Location (Mb)	SNPs	Individuals with consensus ROHs	Gene	β (95% CI)	P-value
3	491.91-492.85	28	225	*WDR6*	-0.19 (-0.25, -0.14)	$4.57x10^{-10}$
3	491.14-491.38	10	227	*WDR6*	-0.19 (-0.24, -0.13)	$1.61x10^{-9}$
3	491.75-491.76	2	233	*WDR6*	-0.17 (-0.23, -0.12)	$1.86x10^{-8}$
15	414.96-415.58	23	331	*ZSCAN29*	0.08 (0.05, 0.10)	$2.02x10^{-8}$
15	414.27-414.76	18	325	*ZSCAN29*	0.07 (0.05, 0.10)	$2.34x10^{-8}$
3	487.46-490.15	50	227	*WDR6*	-0.17 (-0.23, -0.11)	$3.90x10^{-8}$
15	413.87-413.97	2	322	*ZSCAN29*	0.07 (0.05, 0.10)	$4.85x10^{-8}$
15	413.33-413.38	2	303	*ZSCAN29*	0.07 (0.05, 0.09)	$3.02x10^{-7}$
15	417.06-417.07	2	353	*ZSCAN29*	0.07 (0.04, 0.09)	$3.46x10^{-7}$
15	416.22-416.83	6	352	*ZSCAN29*	0.07 (0.04, 0.09)	$3.71x10^{-7}$
3	486.37-486.47	8	242	*WDR6*	-0.15 (-0.21, -0.10)	$7.81x10^{-7}$
3	493.65-495.29	43	198	*WDR6*	-0.16 (-0.22, -0.10)	$7.91x10^{-7}$
3	486.10-486.19	3	248	*WDR6*	-0.15 (-0.20, -0.09)	$1.49x10^{-6}$
17	264.48 – 265.34	25	106	*EVI2A*	-0.24 (-0.33, -0.14)	$5.22x10^{-6}$
15	413.33 – 413.38	2	303	*ADAL*	0.04 (0.02, 0.06)	$5.75x10^{-6}$

15	413.87 – 413.97	2	322	*ADAL*	0.04 (0.02, 0.05)	$1.04x10^{-5}$
16	663.38 – 663.69	6	454	*DUS2L*	-0.10 (-0.15, -0.06)	$1.50x10^{-5}$

Chr – chromosome, SNPs – number of SNPs in overlapping consensus region, β – regression coefficient, 95% CI – 95% confidence intervals, P-value – level of statistical significance for difference in gene expression between individuals with and without consensus ROHs adjusted for demographic parameters and genomic control. All identified associations have a false discovery rate q-value <0.01. ***WDR6*** – WD repeat domain 6, ***ZSCAN29*** – zinc finger and SCAN domain containing 29, ***EVI2A*** [MIM 158380] – ecotropic viral integration site 2A, ***ADAL*** – adenosine deaminase-like, ***DUS2L*** – dihydrouridine synthase 2. Gene symbols appear more than once in the table because different consensus ROHs can be mapped onto the same gene array probe if there is more than one ROH within 100kb of the probe.

Figure 1.



Stage B (Primary objective)
Calculation of homozygosity measures (Average ROH number, Average ROH length, Average total ROH length, FROH)

- Genetic architecture of homozygosity
- Comparison of homozygosity measures between CAD patients and controls

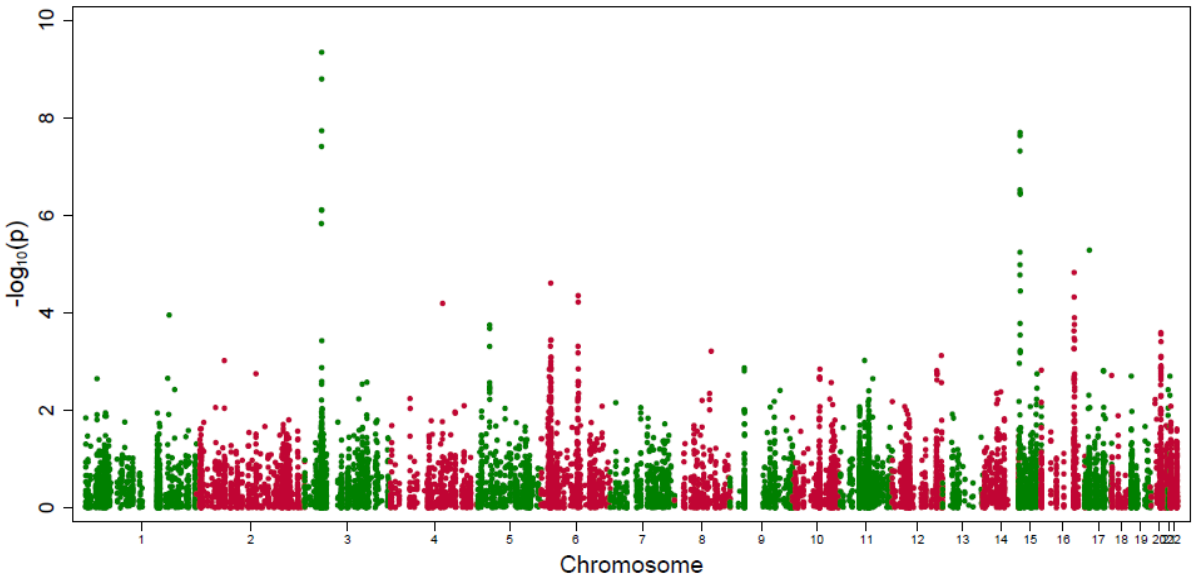
Stage C (Secondary objective)
Identification of consensus ROHs within studies (consensus region of a homozygous run of SNPs in at least 5 individuals within a study)

- Identification of overlapping consensus ROHs across cohorts
- Association between overlapping consensus ROHs and CAD
- Aggregate burden analysis

Stage D (Secondary objective)
Association between consensus ROHs and gene expression in human monocytes and macrophages (ROH carriers vs. non-carriers)

Figure 2.

A. Macrophages



B. Monocytes