Title: Regional polygenic covariance reveals heterogeneity in the shared heritability between complex traits

Guillaume Pare[1,2,3*], Shihong Mao[1], Wei Q. Deng[4]

1 Population Health Research Institute, Hamilton Health Sciences and McMaster University, Hamilton, Canada, 2 Population Genomics Program, Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Canada, 3 Department of Pathology and Molecular Medicine, McMaster University, Hamilton, Canada, 4 Department of Statistical Sciences, University of Toronto, Toronto, Canada

*Corresponding author: pareg@mcmaster.ca

Abstract

Complex traits can share a substantial proportion of their polygenic heritability. However, genome-wide polygenic correlations between a pair of traits can mask heterogeneity in their shared polygenic effects across loci. We propose a novel method (WML-RCP) to evaluate polygenic covariance between two complex traits in small genomic regions using summary association statistics. Our method makes no assumption about the causality of one trait on the other, but rather tests for evidence that the polygenic effect at a given region affects two traits concurrently. We show through simulations that our method is well calibrated and more powerful than other co-localisation methods under a polygenic model. As small genomic regions are more likely to harbour specific genetic effects, our method is ideal to identify heterogeneity

in shared polygenic covariance across regions. We illustrate the usefulness of our method by addressing three questions related to cardio-metabolic traits. First, we explore how regional polygenic covariance can inform on the strong epidemiological association between HDL cholesterol and coronary artery disease (CAD), suggesting a key role for triglycerides metabolism. Second, we identify a ~4Mb region including *PPP1R3B* on chromosome 8p23.1 with paradoxical polygenic covariance between triglycerides and BMI, as well as evidence of polygenic inheritance and pairwise covariance with multiple other metabolic traits. Finally, we investigate the potential role of PPARγ activators in the prevention of CAD. Our results provide a compelling argument that shared heritability between complex traits is highly heterogeneous across loci.

Introduction

Most complex traits follow a polygenic model of inheritance, whereby thousands of common genetic variants contribute to phenotypic variance. Furthermore, genetic variance is not spread evenly throughout the genome but rather tends to concentrate in specific regions[1-3]. While shared polygenic heritability between pairs of complex traits has been shown at a genome-wide level[4], there are currently no methods to evaluate regional polygenic covariance between two traits using summary association statistics. Existing methods for regional covariance either use individual-level data[5,6] or test for co-localisation of single variant associations without considering polygenic inheritance[7-10]. Nonetheless, the observation that a majority of polygenic heritability lies in variants associated below genome-wide significance coupled with the concentration of such associations at specific loci dictates the need for a method that can

estimate polygenic covariance within small (~1 Mb) regions. As each genetic region includes a different set of genes, genome-wide correlations will miss heterogeneity in the contribution of individual genes to shared heritability. Also, current co-localisation methods have been designed to identify linked eQTLs[7,11], which typically harbour very strong associations unlike most other complex traits.

We now propose a novel method (WML-RPC) to estimate the regional polygenic covariance between two traits, retaining all variants in a given region irrespective of linkage disequilibrium (LD) and using summary association statistics. Our method adopts a weighted maximum likelihood approach to estimate the regional polygenic variance of each trait and their polygenic covariance. It assumes random polygenic effects, or in other words, that multiple genetic variants are associated with a trait in each region. Our framework builds on previous work[1,12] and has the distinct advantage of being robust to misspecification of either LD or genetic effect sizes. Unlike other approaches, our method makes no assumption about the causal relationship of one trait over the other but rather is intended to test whether a single polygenic effect affects two traits concurrently at a given locus. In addition, as WML-RPC provides estimates for the strength of the polygenic covariance, it can be used to test for the presence of covariance or alternatively for deviation from a set level of covariance. We illustrate the usefulness of our method by applying it to three questions related to cardio-metabolic traits, bringing novel insights into the inverse association of HDL cholesterol with coronary artery disease (CAD), identifying a region on chromosome 8p23.1 with multiple strong signals of pairwise genetic covariance, and exploring the role of PPARγ activators in the prevention of CAD.

Results

Simulations with 1000 Genomes Project Haplotypes

We simulated two traits using phased 1000 Genomes (1000G) Project[13] haplotypes. The simulated regions comprised 700 SNPs, corresponding to a physical distance of ~1Mb, and summary association statistics were generated in two distinct populations of 100,000 participants each. Assuming realistic levels of genetic association, there was no type I error inflation when either or both traits were truly genetically associated in the absence of any genetic covariance. Power to detect genetic covariance was dependent on genetic effect sizes and strength of true underlying genetic covariance (Figure 1). We evaluated the influence of correlated error terms as could occur if summary association statistics were derived from overlapping sets of participants. With 25% of participants overlapping and the non-genetic correlation between traits set at $\rho = 0.2$, the impact on both type I error rate and power was minimal (Figures S1 and S2). We also tested a more extreme scenario that assumed a complete overlap in participants (i.e. correlation of non-genetic error terms of 0.2). Again, minimal type I error inflation was observed under the null hypothesis of no covariance (Figures S1 and S2). Finally, we sought to benchmark our method against two recently described co-localisation methods[8,10]. Both methods assess the possibility that a single causal variant underlies a genetic association with two traits. As expected given their model assumptions, neither method performed well in the presence of polygenic inheritance, with both showing inflated type I error rates and decreased power as compared to WML-RPC (Figure S3).

Insights into the relationship between HDL cholesterol and coronary artery disease

To illustrate how regional polygenic covariance can provide novel epidemiological insights, we first explored the genetic relationship between HDL cholesterol (HDLc) and coronary artery disease (CAD) using summary association statistics from large genetic meta-analyses. Blood HDLc concentration is one of the strongest predictor of decreased CAD risk in epidemiological studies[14], yet the causality of this association remains controversial. Several Mendelian randomization studies have been conducted to address this question, supporting a lack of causal relationship[15-19]. Furthermore, pharmacological interventions to raise HDLc have thus far been disappointing[20-22], further strengthening the hypothesis of a non-causal relationship. If the relationship truly is non-causal, then one or more upstream biological pathways can be expected to jointly affect HDLc concentration and risk of CAD, thus explaining the strong epidemiological association. In other words, there must exist some underlying causal risk factor(s) that leads to decreased HDL cholesterol while increasing the risk of CAD, even if HDLc itself is an epiphenomenon. Regional polygenic covariance can help address this question by identifying regions whose effects on both HDLc and CAD are consistent with epidemiological studies, providing insights into the identity of those biological pathways responsible for the strong epidemiological association.

We divided the genome into 2,687 regions of ~ 1 Mb and determined which regions showed evidence of polygenic covariance between HDLc and CAD. Keeping only the 673 regions with at least nominal evidence ($p < 0.05$) of polygenic association with either HDLc or CAD, we tested for regional polygenic covariance and applied a conservative Bonferroni

correction (p <0.05/673). Consistent with a non-causal role of HDLc in CAD, none of the 4 regions identified are directly involved in HDL production (e.g. the *APOA1* locus) and heterogeneity in polygenic covariance was present, with one region having positive polygenic covariance while others had negative covariance (Table 1). The three regions with significant negative polygenic covariance between HDLc and CAD are of particular interest since they could potentially underpin the epidemiological association. Tellingly, all three of them were located at loci known to be directly related to triglycerides metabolism. Both the LPL and TRIB1 loci are strongly associated with fasting triglycerides[23] while APOE is linked to the postprandial regulation of triglyceride-rich lipoproteins[24]. A single region had significant (p = 6.2 x10^{-5}) positive covariance, encompassing the gene encoding for hepatic lipase (*LIPC*). LIPC deficiency is known to lead to increased HDLc[25] and triglycerides-rich intermediate-density lipoproteins (IDL)[26], but its role in CAD remains controversial. Consistently, genome-wide significant positive polygenic covariance between HDLc and triglycerides was also observed at the locus, the only locus across the genome with positive covariance (data not shown). Overall, our results support the hypothesis the role of HDLc as a marker of triglycerides levels can help explain the strong epidemiological association with CAD. Triglycerides are known causal mediators of CAD[19]. Their levels are notoriously variable and can increase dramatically in the post-prandial state. As HDLc concentrations are both more stable and inversely correlated to triglycerides concentrations, they can provide a surrogate for long-term exposure to triglycerides. Indeed, non-fasting triglycerides, although seldom measured, have been shown to better predict CAD risk than fasting measurements[27]. Our results also suggest that high HDLc caused by decreased LIPC activity increases risk of CAD.

Triglycerides and Body Mass Index (BMI)

Increased BMI is strongly associated with increased blood triglycerides (TG) levels and evidence suggests this relationship reflects a causal effect[28,29]. We sought to determine if, even in the context of this well-established relationship, there were loci affecting TG and BMI in opposite directions. Similar to the previous analysis, we only kept the 841 regions with at least nominal evidence ($p < 0.05$) of polygenic association with either TG or BMI, and then tested for regional polygenic covariance, applying a conservative Bonferroni correction ($p < 0.05/841$). A single region on chromosome 8p23.1 had significant negative polygenic covariance (Figure 2 and Table S1). A closer inspection revealed that this region was flanked by three other regions showing a similar pattern of association. Polygenic associations were stronger than univariate ones, as evidenced by comparing polygenic regional association *p*-values to minimum univariate *p*-value in the corresponding regions. Merging all four regions together resulted in a much stronger polygenic association signal, as expected. The merged region itself is flanked by poorly characterized and difficult to sequence regions. Remarkably, polygenic covariance was also observed with multiple other metabolic traits (Table S2). The extended region encompasses the *PPP1R3B* gene, among others[30,31]. Product of the *PPP1R3B* gene is the regulatory subunit of PP1, a phosphatase involved in the modulation of glycogen synthesis in the liver and whole-body glucose homeostasis[32]. This locus has been implicated in genetic associations with several traits, including fasting glucose, fasting insulin[33], lipids[34], C-reactive protein[35] and plasma lactate[36].

Thiazolidinediones, PPARγ and risk of CAD

Pharmacological activation of PPARγ with thiazolidinediones is used to treat and prevent diabetes. However, the role of thiazolidinediones in prevention of CAD is controversial. *Post hoc* analyses of randomized trials identified a potential deleterious effect of thiazolidinediones on CAD risk[37], which led to the removal of all thiazolidinediones from clinical use except one, pioglitazone. Based on these observations, a large clinical trial addressing the issue of CAD risk reduction by rosiglitazone was stopped early[38], thus leaving this important clinical question unanswered. The controversy was further fueled by the recent publication of the IRIS trial showing a significant *reduction* in cardiovascular events in individuals randomized to pioglitazone[39]. Regional polygenic covariance can provide insights into the issue. We tested the region surrounding *PPARG* (+/- 500 Kb) for evidence of association with cardiometabolic traits. As expected from the known pharmacological effects of thiazolidinediones[40], significant ($p < 0.05$) regional associations were observed with diabetes, triglycerides, HDLc, LDLc and BMI. We then tested this set of traits for polygenic covariance with diabetes and CAD (Table 2). Significant and positive polygenic covariance was observed between diabetes and triglycerides, triglycerides and CAD, and LDLc and CAD. Significant negative polygenic variance was observed between diabetes and BMI. However, polygenic covariance was not significant between diabetes and LDLc, or diabetes and CAD. Polygenic covariance between LDLc and CAD is of particular interest since pioglitazone has recently been shown to reduce LDL particle number and size [41,42]. This observation and the polygenic covariance with triglycerides support the hypothesis that the protective effect of pioglitazone (and perhaps other thiazolidinediones) on CAD risk is the consequence of its beneficial effect on atherogenic lipoproteins.

Discussion

We herein propose a novel method to estimate regional polygenic covariance between two traits. Our method is distinct from other co-localisation tests as it is based on a polygenic model of inheritance and it makes no assumptions about the causal relationship between traits. This approach is particularly attractive when studying complex traits with strong polygenic inheritance, where any single genetic association is unlikely to fully capture a large proportion of genetic effects. Our method has several other advantages, including the ability to adjust for LD, the possibility to test specific hypotheses regarding polygenic covariance, and the use of summary association statistics as inputs. WML-RPC has wide ranging applications, as we have illustrated. It can help discover biological pathways explaining epidemiological associations such as for HDLc and CAD, identify regions with complex patterns of polygenic covariance, or help gain insights into the role of single genes or drug targets.

Our examples make a compelling argument that shared heritability is highly dependent on regional genetic effects. Even for established relationships such as the one between BMI and triglycerides, some regions could be found where increased BMI was linked to decreased triglycerides. Unless a locus has a direct effect on a risk factor (e.g. the *APOB* or *LDLR* loci on LDLc), it thus cannot be assumed covariance implies a causal effect of the risk factor on outcome. For instance, genetic covariance between HDLc and CAD at the *LIPC* locus, combined with prior knowledge of the effect of LIPC on intermediate density lipoproteins, suggests that decreased LIPC activity leads to both increased HDLc and CAD risk. Inclusion of that locus in Mendelian randomization studies may thus result in biased inferences about the causal role of HDLc in CAD. These considerations stress the importance of taking the biological

effects of each genetic region into account before concluding on the relationship between a risk factor and outcome. Knowledge of biological effects can also provide insights into epidemiological relationships, such as regions with negative covariance between HDLc and CAD pointing to triglycerides metabolism as a key factor to explain the epidemiological association.

WML-RPC can also be used to explore candidate gene regions. We found that regional polygenic associations recapitulate the effects of PPARγ agonist thiazolidinediones on cardio-metabolic traits. Our results support the hypothesis thiazolidinediones can reduce CAD risk through their effect on lipids, particularly LDLc and triglycerides. In line with this hypothesis, recent data have shown pioglitazone decreased the concentration of atherogenic lipoproteins [41,42]. However, genetic covariance with CAD was only significant with LDLc and triglycerides but not diabetes itself, as might have been expected given triglycerides had significant covariance with both diabetes and CAD. While this could have been the results of the play of chance, it is also possible that genetic variants regulating PPARG function vary from one tissue to the other, such that genetic regulation of LDLc at the *PPARG* locus (and thus risk of CAD) only partially overlaps with its effect on diabetes. Indeed, tissue-specific effects of PPARG have been described [40], with adipocytes mainly responsible for glycemic effects and hepatocytes regulating atherogenic lipoprotein metabolism [43]. It is likewise possible thiazolidinediones have varying affinities for different tissues. This illustrates a further advantage of our method as it is agnostic to gene regulation mechanisms and thus not dependent on known eQTL associations, which may vary according to tissue and cellular context.

There are some limitations. First, WML-RPC assumes that a single genetic effect (i.e. gene) underlies covariance at each region, which might not always be the case even within small regions. Second, some loci might not fit a polygenic model, for example when there is a single very strong association at a locus, and other methods might be better suited. However, as we have shown at the *PPP1R3B* locus there clearly are situations where the polygenic model best captures genetic variance and covariance. Third, statistical power to detect genetic covariance depends on sample size and genetic variance. While confident in regions identified using stringent statistical criteria, many other truly covariant regions have likely been missed. Fourth, many regions have no known candidate genes, in which case our method can point to regions of interest but not necessarily biological interpretations. Nonetheless, improving knowledge of gene function and regulation, combined with the expanding repertoire of genome-wide association studies, should provide increasing opportunities for WML-RPC to lead to novel insights into complex traits.

In conclusion, we present a novel method to estimate regional polygenic covariance using summary association statistics. WML-RPC can estimate polygenic covariance within relatively small genetic regions, enabling a more detailed characterization of genetic covariance than genome-wide genetic correlations. Our method can be used to identify pathways shared between two traits, pinpoint regions of interest, or test specific hypotheses for a given gene. Our examples illustrate the heterogeneity in pairwise genetic covariance across loci. They support the notion that genetic effects are specific to each region and that unless a locus directly affects a risk factor, caution must be exercised when making causal inferences.

Methods

Estimation of regional genetic variance and covariance

We recently described a simple procedure to estimate regional genetic variance using summary association statistics, adjusting for linkage disequilibrium (LD)[12]. We now propose to adapt this procedure to estimate polygenic covariance between a pair of traits using a weighted maximum likelihood (WML) approach. Suppose the genotype matrix is fixed while the true genetic effect is a random vector β, whose individual components, i.e. the effect size of SNPs, $I = 1,2,\ldots,m$, have mean zero and variance σ^2. The true, unobserved, genetic model can be expressed as:

$$Y = X_{n\times m}\beta + \epsilon \text{ (Equation 1)}$$

where ε is a vector of standard normal error with identity variance covariance matrix and the genetic variance is given by $m\sigma^2$. Without loss of generalizability, we assume the observed quantitative trait (y) and the n x m genotype matrix $\boldsymbol{X}$ standardized to have zero mean and unit variance throughout. The pairwise LD (r^2) between two SNPs k and l is denoted by $r^2_{k,l}$. For a SNP d, the following LD adjustment (η_d) can be defined as the summation of LD between the d^{th} SNP and 200 SNPs upstream and downstream:

$$\eta_d = \sum_{j=d-200}^{d+200} r^2_{d,j} \text{ (Equation 2)}$$

with a distance of 200 SNPs assumed sufficient to ensure the extend of LD (other values might be used). Only including SNPs with summary GWAS statistics in the sum, variance explained by each SNP *d* is given by:

$$R_d^2 = \frac{b_d^2}{\eta_d} \text{ (Equation 3)}$$

where b_d denotes the univariate regression coefficient commonly reported in GWAS results with sample size *N* (assuming genotypes from external GWASs have also been standardized to have zero mean and unit variance). Assuming a strictly additive genetic model where each SNP contributes additively to a trait without any interaction or haplotype effects, we have previously shown[12] that $\sum_{d=1}^{m} R_d^2$ is an estimator of the regional variance explained $m\sigma^2$ by demonstrating the approximated equivalence between the expected total genetic variance over a region

$$E[\sum_{d=1}^{m} R_d^2] = m\sigma^2 + \frac{1}{N}\sum_{d=1}^{m}\frac{1}{\eta_d} \quad \text{(Equation 4)}$$

and the multiple linear regression variance explained

$$E[R^2] = m\sigma^2 + \frac{m}{N} \quad \text{(Equation 5)}$$

when the sample size is sufficiently large.

Since the true genetic effects are given by a random vector β, this implies:

$$\boldsymbol{b} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \frac{(X'X)(X'X)}{N^2} + \frac{(X'X)}{N^2})$$

or marginally:

$$b_d \sim \mathcal{N}(0, \eta_d \sigma^2 + \frac{1}{N}) \quad \text{(Equation 6)}$$

As we are interested in estimation of σ^2 via the surrogate $\sum_{d=1}^{m} \frac{b_d^2}{\eta_d}$, the following weighted likelihood function is maximized to find $\hat{\sigma}^2$:

$$log\mathcal{L}(\sigma^2|b) = -\sum_{d=1}^{m} \frac{1}{2\eta_d} (\log\left(\eta_d \sigma^2 + \frac{1}{N}\right) + \frac{b_d^2}{\eta_d \sigma^2 + \frac{1}{N}}) \quad \text{(Equation 7)}$$

where the log-likelihood of each observed $\widehat{b_d}$ is weighted by the inverse of the LD adjustment such that if two SNPs were in complete LD, then effectively only one SNP contributes to the log-likelihood for the genetic variance.

This framework can be extended to study the genetic covariance between a pair of traits. In this scenario we have $b_d^{(1)}$ and $b_d^{(2)}$ the summary association statistics for trait 1 and 2, respectively, following a bivariate normal distribution:

$$\begin{pmatrix} \boldsymbol{b_d^{(1)}} \\ \boldsymbol{b_d^{(2)}} \end{pmatrix} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma_d}) \text{ (Equation 8)}$$

with covariance matrix:

$$\boldsymbol{\Sigma_d} = \begin{bmatrix} \eta_d \sigma_1^2 + \dfrac{1}{N_1} & \mathrm{Cov}_{gen} + \mathrm{Cov}_{error} \\ \mathrm{Cov}_{gen} + \mathrm{Cov}_{error} & \eta_d \sigma_2^2 + \dfrac{1}{N_2} \end{bmatrix}$$

Where σ_1^2 and σ_2^2 are the genetic variance of trait 1 and 2, respectively, and N_1 and N_2 the corresponding sample sizes. Cov_{gen} represents the genetic covariance between both traits whereas Cov_{error} is the error term covariance and can be assumed to be zero.

The weighted likelihood function can be adapted using genetic variance estimates from the previous weighted likelihood:

$$\log\mathcal{L}\left(\mathrm{Cov}_{gen} \middle| \boldsymbol{b_d}, \hat{\sigma}_1^2, \hat{\sigma}_2^2, \mathrm{Cov}_{error}\right) = -\sum_{d=1}^{m} \frac{1}{2\eta_d} \left(\log|\boldsymbol{\Sigma_d}| + \boldsymbol{b_d}^T \boldsymbol{\Sigma_d}^{-1} \boldsymbol{b_d}\right) \text{ (Equation 9)}$$

The maximum likelihood estimates of Cov_{gen} enables the use of a likelihood ratio test for hypothesis testing. While Cov_{error} could be estimated, we found that under realistic scenarios its effect is negligible and has therefore been set to zero for current analyses. This might not be ideal when the correlation of error terms is very strong, in which case a non-zero Cov_{error} could be used in the likelihood estimation. As a note, Cov_{gen} estimated from empirical data can cause numerical estimates of the genetic correlation to be higher than 1 or lower than -1, and will correspondingly be set to 1 or -1. Also note that stable and meaningful estimates of Cov_{gen} can only be obtained when both trait 1 and 2 have non-zero genetic variance. Finally, although the aim of the method is to estimate regional polygenic covariance, it is important to confirm there is

no bias in regional polygenic variance under the null hypothesis of no genetic association, which we checked using simulations (Figure S4).

Simulations using 1000 Genomes Project data and study of cardio-metabolic traits

We tested our method using summary association statistics from large genetic meta-analyses of cardiometabolic traits, including coronary artery disease[44], LDL cholesterol, HDL cholesterol, triglycerides[45], type 2 diabetes[46], body mass index[47], and blood pressure[48]. We identified a common set of SNPs among all corresponding meta-analyses and subsequently divided the genome into blocks of ~1Mb minimizing inter-block LD, as described in[12]. We used 1000G[13] participants of European descent as the reference panel for LD as it is the dominant ancestry in the studies included.

References

1. Pare, G., Asma, S. & Deng, W.Q. Contribution of large region joint associations to complex traits genetics. *PLoS Genet* **11**, e1005103 (2015).
2. Loh, P.R. *et al.* Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. *Nat Genet* **47**, 1385-92 (2015).
3. Shi, H., Kichaev, G. & Pasaniuc, B. Contrasting the Genetic Architecture of 30 Complex Traits from Summary Association Data. *Am J Hum Genet* **99**, 139-53 (2016).
4. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat Genet* **47**, 1236-41 (2015).
5. Visscher, P.M. *et al.* Statistical power to detect genetic (co)variance of complex traits using SNP data in unrelated samples. *PLoS Genet* **10**, e1004269 (2014).
6. Lee, S.H., Yang, J., Goddard, M.E., Visscher, P.M. & Wray, N.R. Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. *Bioinformatics* **28**, 2540-2 (2012).
7. Zhu, Z. *et al.* Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet* **48**, 481-7 (2016).
8. Chun, S. *et al.* Limited statistical evidence for shared genetic effects of eQTLs and autoimmune-disease-associated loci in three major immune-cell types. *Nat Genet* (2017).
9. Hormozdiari, F. *et al.* Colocalization of GWAS and eQTL Signals Detects Target Genes. *Am J Hum Genet* **99**, 1245-1260 (2016).
10. Pickrell, J.K. *et al.* Detection and interpretation of shared genetic influences on 42 human traits. *Nat Genet* **48**, 709-17 (2016).
11. Mancuso, N. *et al.* Integrating Gene Expression with Summary Association Statistics to Identify Genes Associated with 30 Complex Traits. *Am J Hum Genet* **100**, 473-487 (2017).
12. Pare, G., Mao, S. & Deng, W.Q. A method to estimate the contribution of regional genetic associations to complex traits from summary association statistics. *Sci Rep* **6**, 27644 (2016).
13. Genomes Project, C. *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74 (2015).
14. Emerging Risk Factors, C. *et al.* Major lipids, apolipoproteins, and risk of vascular disease. *JAMA* **302**, 1993-2000 (2009).
15. Frikke-Schmidt, R. *et al.* Association of loss-of-function mutations in the ABCA1 gene with high-density lipoprotein cholesterol levels and risk of ischemic heart disease. *JAMA* **299**, 2524-32 (2008).
16. Johannsen, T.H. *et al.* Hepatic lipase, genetically elevated high-density lipoprotein, and risk of ischemic cardiovascular disease. *J Clin Endocrinol Metab* **94**, 1264-73 (2009).
17. Haase, C.L., Tybjaerg-Hansen, A., Grande, P. & Frikke-Schmidt, R. Genetically elevated apolipoprotein A-I, high-density lipoprotein cholesterol levels, and risk of ischemic heart disease. *J Clin Endocrinol Metab* **95**, E500-10 (2010).

18. Voight, B.F. *et al.* Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study. *Lancet* **380**, 572-80 (2012).
19. Holmes, M.V. *et al.* Mendelian randomization of blood lipids for coronary heart disease. *Eur Heart J* **36**, 539-50 (2015).
20. Barter, P.J. *et al.* Effects of torcetrapib in patients at high risk for coronary events. *N Engl J Med* **357**, 2109-22 (2007).
21. Keech, A. *et al.* Effects of long-term fenofibrate therapy on cardiovascular events in 9795 people with type 2 diabetes mellitus (the FIELD study): randomised controlled trial. *Lancet* **366**, 1849-61 (2005).
22. Schwartz, G.G. *et al.* Effects of dalcetrapib in patients with a recent acute coronary syndrome. *N Engl J Med* **367**, 2089-99 (2012).
23. Global Lipids Genetics, C. *et al.* Discovery and refinement of loci associated with lipid levels. *Nat Genet* **45**, 1274-83 (2013).
24. Bergeron, N. & Havel, R.J. Prolonged postprandial responses of lipids and apolipoproteins in triglyceride-rich lipoproteins of individuals expressing an apolipoprotein epsilon 4 allele. *J Clin Invest* **97**, 65-72 (1996).
25. Connelly, P.W. & Hegele, R.A. Hepatic lipase deficiency. *Crit Rev Clin Lab Sci* **35**, 547-72 (1998).
26. Ruel, I.L., Couture, P., Cohn, J.S. & Lamarche, B. Plasma metabolism of apoB-containing lipoproteins in patients with hepatic lipase deficiency. *Atherosclerosis* **180**, 355-66 (2005).
27. Bansal, S. *et al.* Fasting compared with nonfasting triglycerides and risk of cardiovascular events in women. *JAMA* **298**, 309-16 (2007).
28. Fall, T. *et al.* The role of adiposity in cardiometabolic traits: a Mendelian randomization analysis. *PLoS Med* **10**, e1001474 (2013).
29. Freathy, R.M. *et al.* Common variation in the FTO gene alters diabetes-related metabolic traits to the extent expected given its effect on BMI. *Diabetes* **57**, 1419-26 (2008).
30. Smith, S., Giriat, I., Schmitt, A. & de Lange, T. Tankyrase, a poly(ADP-ribose) polymerase at human telomeres. *Science* **282**, 1484-7 (1998).
31. Borowiec, M. *et al.* Mutations at the BLK locus linked to maturity onset diabetes of the young and beta-cell dysfunction. *Proc Natl Acad Sci U S A* **106**, 14460-5 (2009).
32. Mehta, M.B. *et al.* Hepatic protein phosphatase 1 regulatory subunit 3B (Ppp1r3b) promotes hepatic glycogen synthesis and thereby regulates fasting energy homeostasis. *J Biol Chem* (2017).
33. Manning, A.K. *et al.* A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nat Genet* **44**, 659-69 (2012).
34. Teslovich, T.M. *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707-13 (2010).
35. Dehghan, A. *et al.* Meta-analysis of genome-wide association studies in >80 000 subjects identifies multiple loci for C-reactive protein levels. *Circulation* **123**, 731-8 (2011).
36. Tin, A. *et al.* GCKR and PPP1R3B identified as genome-wide significant loci for plasma lactate: the Atherosclerosis Risk in Communities (ARIC) study. *Diabet Med* **33**, 968-75 (2016).

37. Nissen, S.E. & Wolski, K. Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. *N Engl J Med* **356**, 2457-71 (2007).
38. Punthakee, Z. *et al.* Design, history and results of the Thiazolidinedione Intervention with vitamin D Evaluation (TIDE) randomised controlled trial. *Diabetologia* **55**, 36-45 (2012).
39. Kernan, W.N. *et al.* Pioglitazone after Ischemic Stroke or Transient Ischemic Attack. *N Engl J Med* **374**, 1321-31 (2016).
40. Ahmadian, M. *et al.* PPARgamma signaling and metabolism: the good, the bad and the future. *Nat Med* **19**, 557-66 (2013).
41. Mani, P. *et al.* Favorable Impact on LDL Particle Size in Response to Treatment With Pioglitazone is Associated With Less Progression of Coronary Atherosclerosis in Patients With Type 2 Diabetes. *J Am Coll Cardiol* **66**, 328-9 (2015).
42. Nakano, K. *et al.* Effect of pioglitazone on various parameters of insulin resistance including lipoprotein subclass according to particle size by a gel-permeation high-performance liquid chromatography in newly diagnosed patients with type 2 diabetes. *Endocr J* **57**, 423-30 (2010).
43. Duan, Y. *et al.* Peroxisome Proliferator-activated receptor gamma activation by ligands and dephosphorylation induces proprotein convertase subtilisin kexin type 9 and low density lipoprotein receptor expression. *J Biol Chem* **287**, 23667-77 (2012).
44. Nikpay, M. *et al.* A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet* **47**, 1121-30 (2015).
45. Willer, C.J. *et al.* Discovery and refinement of loci associated with lipid levels. *Nat Genet* **45**, 1274-83 (2013).
46. Mahajan, A. *et al.* Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nat Genet* **46**, 234-44 (2014).
47. Locke, A.E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197-206 (2015).
48. Ehret, G.B. *et al.* Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* **478**, 103-9 (2011).

Author Contributions:

G. P. designed the experiment; G.P and W.Q.D. wrote the manuscript; S. M. analyzed the data and prepared tables and figures; All authors reviewed the manuscript.

Competing Financial Interests:

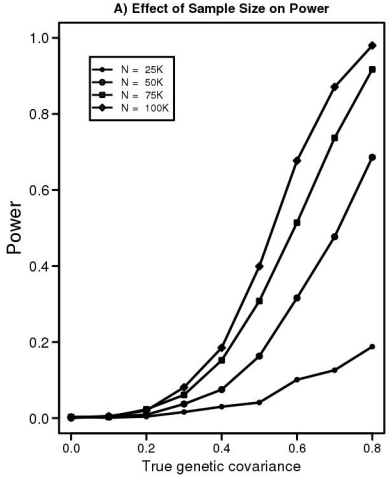
The authors declare no competing financial interests.

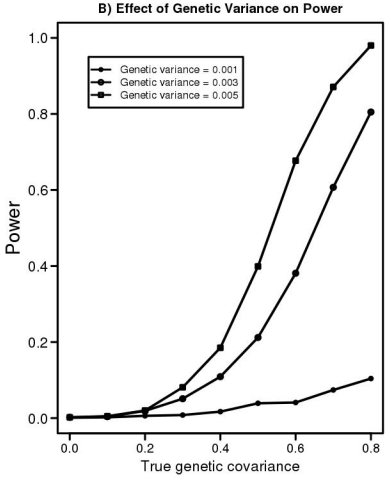
Figures Legends

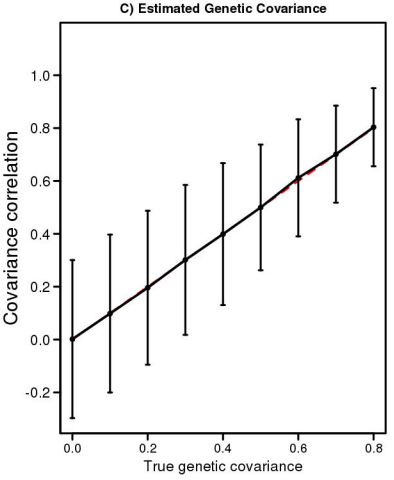
Figure 1: Performance of WML-RPC in simulated data using 1000 Genomes Project haplotypes. Power to detect polygenic covariance at an alpha level of 0.001 according to true polygenic covariance in 1,000 simulated replicates with 700 SNPs from the 1000G Project. In panel A, sample size ranged from 25K to 100K individuals while keeping the true regional genetic variance constant at 0.005 for each trait. In Panel B, sample size was set at 100K individuals, but the genetic variance varied from 0.001 to 0.005. In Panel C, the mean (SD) estimated genetic covariance is illustrated as a function of the true (red dashed line) genetic covariance, assuming a sample size of 100K and genetic variance of 0.005 for both traits.

Figure 2: Genetic associations and polygenic covariance of BMI and triglycerides at the 8p23.1 locus

Genetic associations with BMI at the 8p23.1 locus are illustrated with blue points (univariate SNP *p*-values), short horizontal lines (regional genetic variance *p*-values) and long horizontal line (regional genetic variance *p*-value of the merged region). Corresponding *p*-values for triglycerides are shown in red. BMI-triglycerides regional polygenic covariance *p*-values are similarly illustrated with cyan horizontal lines.


A) Effect of Sample Size on Power
N = 25K
N = 50K
N = 75K
N = 100K
Power
True genetic covariance
0.0
0.2
0.4
0.6
0.8
1.0


B) Effect of Genetic Variance on Power
Genetic variance = 0.001
Genetic variance = 0.003
Genetic variance = 0.005
Power
True genetic covariance
0.0
0.2
0.4
0.6
0.8
1.0


C) Estimated Genetic Covariance
Covariance correlation
True genetic covariance
-0.2
0.0
0.2
0.4
0.6
0.8
1.0

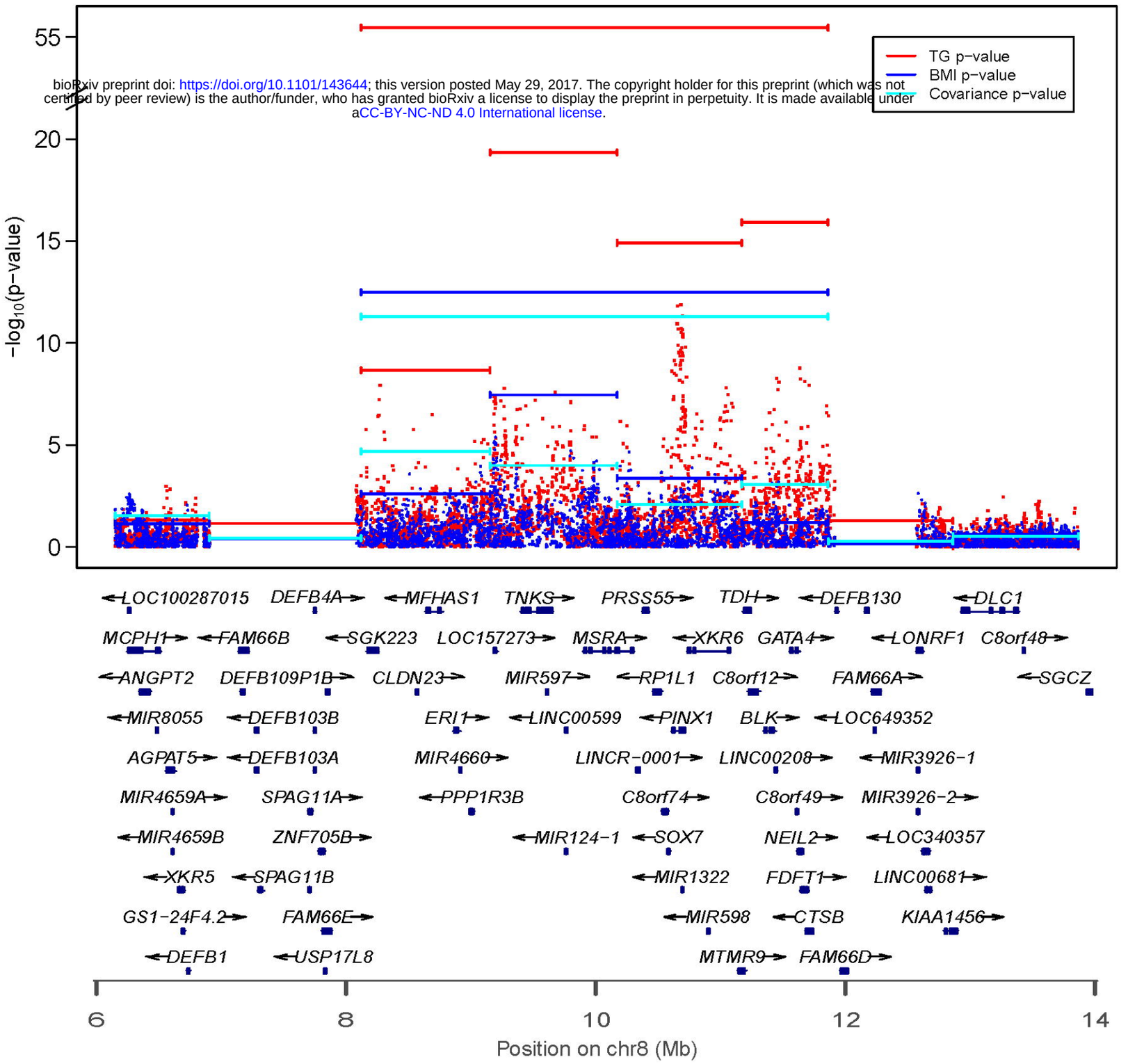


Table 1: Regions with significant polygenic covariance between HDLc and CAD

CHR	Position (Kb)	Candidate gene	Candidate Gene function	HDLc polygenic regional *p*-value	CAD polygenic regional *p*-value	Polygenic covariance (95% CI)	Polygenic covariance *p*-value
8	18,869-19,875	LPL	Lipoprotein triglyceride lipase	1.03E-266	7.90E-05	-0.93 (-1.00, -0.74)	1.31E-07
8	126,042-127,062	TRIB1	Regulation of hepatic lipogenesis	5.67E-39	4.45E-03	-1.00 (-1.00, -0.74)	4.00E-05
15	58,311-59,348	LIPC	Hepatic triglyceride lipase	<1.00E-323	2.12E-01	1.00 (0.85, 1.00)	6.17E-05
19	44,789-45,840	APOE	Catabolism of triglyceride-rich lipoproteins	1.50E-31	7.51E-07	-0.86 (-1.00, -0.66)	2.76E-06

Table 2: Polygenic covariance at the PPARG locus

Trait	Regional association *p*-value	Polygenic covariance with diabetes (95%CI)	Polygenic covariance with diabetes *p*-value	Polygenic covariance with CAD (95%CI)	Polygenic covariance with CAD *p*-value
BMI	0.036	-0.98 (-1.00,-0.25)	0.012	0.18 (-0.95,1.00)	0.76
Diabetes	1.62E-03	N/A	N/A	0.41 (-0.56,1.00)	0.42
HDL	4.19E-03	-0.51 (-1.00,0.12)	0.14	0.27 (-0.78,1.00)	0.61
LDL	2.99E-13	0.36 (-0.14,0.87)	0.18	0.99 (0.36,1.00)	1.71E-3
Triglycerides	5.20E-06	0.93 (0.54,1.00)	9.9E-4	0.90 (0.00,1.00)	0.039