Evaluating the Impact of Physiological Variability in Genome-Wide Association Studies of Resting Heart Rate

Stefan van Duijvenboden[1,2], Julia Ramírez[1,2], William J Young[1,3], Andrew Tinker[1], Patricia B Munroe[1], Pier D Lambiase[2,3], Michele Orini[2,3]

[1]Queen Mary University of London, London, UK
[2]University College London, London, UK
[3]Barts Heart Centre, St Bartholomew's Hospital, London, UK

Abstract

Genome-wide association studies (GWAS) have discovered hundreds of genetic loci for resting heart rate (RHR). However, the impact of intra-individual variation in RHR on GWAS results is unclear. We evaluated this impact by analyzing two RHR recordings from N ~61,000 subjects from UK Biobank. In addition, we modelled variations in RHR as independent white zero-mean Gaussian noise with a standard deviation of 0.5x, 1x, and 2x the standard deviation of the difference between the original RHR values (4, 8, and 16 bpm, respectively). The two original RHR recordings were highly correlated ($\rho=0.77$), but results from the genetic analyses were slightly different: the number of genome-wide significant ($p < 5x10^{-8}$) variants at the locus with the strongest reported association (MYH6): n=39 vs. n=34; the p-value of the corresponding lead-variant, $3.6x10^{-24}$ vs. $2.1x10^{-19}$; and the estimated heritability 20.0% vs. 16.7%. Simulated data showed an inverse relationship between RHR variation and genetic association strength and heritability. Results formally demonstrate the impact of intra-individual RHR variability on the discovery of genetic variants in single-measurement studies.

1. Introduction

Elevated resting heart rate (RHR) is a heritable marker of cardiovascular mortality with an estimated heritability between 50-70% in twin studies[1, 2]. Consequently, several genome-wide association studies (GWAS) have been undertaken to elucidate potential mechanisms underlying the genetic regulation of RHR[3-5].

Importantly, the reliability and interpretation of GWAS results relies on the reproducibility of the phenotypic data. From this perspective, heart rate is not constant but varies as a result of physiological (autonomic and hormonal control) and environmental factors (time of day, mental stress, etc)[6]. As a result, RHR measurements taken from the same individuals can differ significantly depending on when, where and how the measurements were taken. Studies investigating the genetic associations of blood pressure have estimated a 20% gain in statistical power with long-term average measurements as compared to single-measurement associations. However, it is unclear how the level of variation in RHR affects the discovery of genetic variants. Increased may introduce false-negative GWAS results by increasing the p-values of associations to values above significance threshold. Alternatively, it is also possible that increased variation may promote overestimation of certain associations leading to false-positive results (e.g. spurious associations). Replication of genetic association results in an independent cohort is important to lessen the likelihood of false positives. However, if replication data is based on data with increased variation in intra-individual RHR, it may impair replication of genetic associations.

The objective of this work was to evaluate the effect of variability in RHR measurements on genetic association results by conducting a physiological and a simulation study based on a large dataset of two repeated recordings of RHR from the UK Biobank study[7].

2. Methods

Materials

Two automated recordings of RHR were obtained from N=74,725 participants from the UK Biobank study. RHR was both recorded during a blood pressure assessment (referred to as RHR recording 1), and during the pre-exercise resting phase of an ECG exercise stress test (referred to as RHR recording 2). We applied genetic quality control to exclude individuals with poor genotype quality including high heterozygosity / missingness and sex discordance, as supplied by UK Biobank (N=3126) and restricted our analysis to individuals with European ancestry (N=65,042). In line with previous studies on RHR, we further excluded individuals with existing cardiovascular conditions known to affect RHR (including atrial fibrillation, history of myocardial infarction or heart failure, (supra)-ventricular tachycardia) as well as individuals on RHR altering medications (non-dihydropyridine calcium antagonists). Individuals with
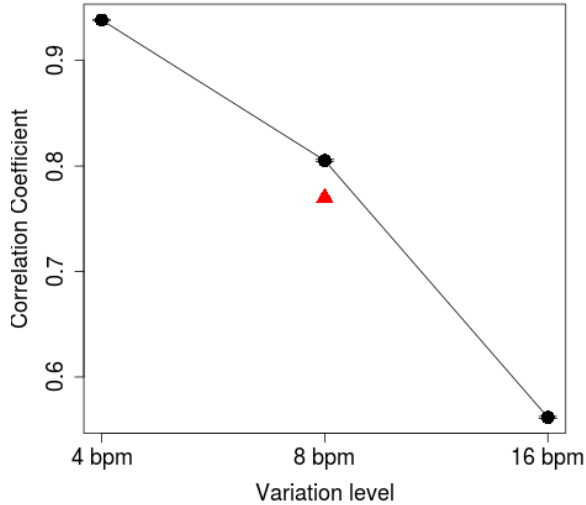
Fig. 1: Correlation between original and simulated RHR measurements. Black circles and error bars show the median correlation coefficient and interquartile range for each simulated level of variation. For comparison, the red triangle shows the variation and correlation measured between the two original recordings.

extreme RHR measurements (<40 or >120 bpm) were also excluded. The remaining 60,913 individuals were selected for analysis.

Simulation study

In addition to physiological variability observed by comparing the two RHR recordings, we also simulated physiological variability at different levels by creating additive zero-mean Gaussian noise models with standard deviation 0.5x, 1x, and 2x σHR (4 ,8, and 16bpm, respectively), where σHR represents the standard deviation of the difference between the two original RHR measurements. These values were added in a random order to the first of the two original RHR recordings, as the corresponding heritability was higher compared to the second recording. The simulation was repeated 10 times for each variability level.

Genetic studies

We performed a genetic association study of chromosome 14, which harbours the locus with the strongest reported association for RHR (*MYH6*)[3]. Each RHR trait (two original traits and 30 simulated traits) were analysed using linear mixed model method (BOLT-LMM)[8] under the additive genetic model including ~328,000 imputed single nucleotide variants (SNVs) with minor allele frequency ≥ 1% and imputed quality > 0.3. The model included sex, age, and age^2 as covariates, in line with previous studies[3]. Three methods were then used to compare the genetic results between traits and we used the results from the first original RHR recording as reference value:

1) Counting the number of genome-wide significant (*P* < 5x10^{-8}) variants at *MYH6*, and measuring the *P*-value of the lead variant (rs365990, discovered in RHR recording 1).

2) Counting the number of genome-wide significant loci and the overlap with loci discovered for RHR recording 1.

3) Comparing the estimated heritability.

Individual loci were defined based on genomic distance of >500 Kb to each side of the lead variants. The heritability was estimated using BOLT-REML[8] with the same covariates included as in the GWAS.

3.   Results

Physiological study

The correlation between the two original RHR measurements was 0.77. From both recordings, the strongest genetic associations were obtained from the first recording: lead variant rs365990, *P*-value 3.6 x 10^{-24}, compared to rs422068 with *P*-value 2.1 x 10^{-19} for the second recording. Both variants were located within only 3kb of each other. The number of genome-wide significant SNVs at *MYH6* was also higher for the first recording: N=39 compared to N=34 (all included in the N=39) for RHR recording 2. Finally, the estimated heritability was higher for recording 1 compared to recording 2: 20.0% versus 16.9%, respectively.

Simulation study

The correlation between original and simulated data reduced with increasing levels of RHR variation, correlation coefficient (cc)=0.93, 0.81, and 0.56 for 0.5, 1, and 2x σHR (Fig. 1), respectively. The median correlation coefficient for variation level 1x σHR (8 bpm) was similar to the correlation between both original RHR recordings (0.77 vs. 0.81, respectively). The number of genome-wide significant SNVs at *MYH6* decreased with increasing levels of variation with median values of: N=34 (interquartile range (IQR): 32-35), 28 (IQR: 26-33), and 1 (IQR: 0-5) for 0.5x, 1x, and 2x σHR respectively, versus N=39 in the original data (Fig. 2A). A similar trend was observed for the change in p-value of the lead variant at this locus: median –log10(*P*-value): 21.3 (IQR: 20.7-21.7), 15.4 (IQR: 14.6-17.1), and 6.9 (IQR: 6.6-7.9), respectively compared to 23.4 for the original data and 7.3 (p=5x10^{-8}) for the generally accepted genome-wide significant threshold (Fig. 2B). The total number of discovered loci also decreased with increasing levels of variation in RHR: median number of discovered loci: N=3 (IQR: 3-3), 2 (IQR: 2-3), and 1 (IQR: 0-1) for 0.5x, 1x, and 2x σHR respectively. All loci discovered in the simulated data overlapped with the loci identified in the original data, and therefore no non-overlapping loci were detected. Finally, heritability estimates decreased with increasing levels of variation in RHR (Fig 2C): 17.4% (IQR: 17.4%-17.5%), 12.8% (IQR: 12.5%-13.1%), and
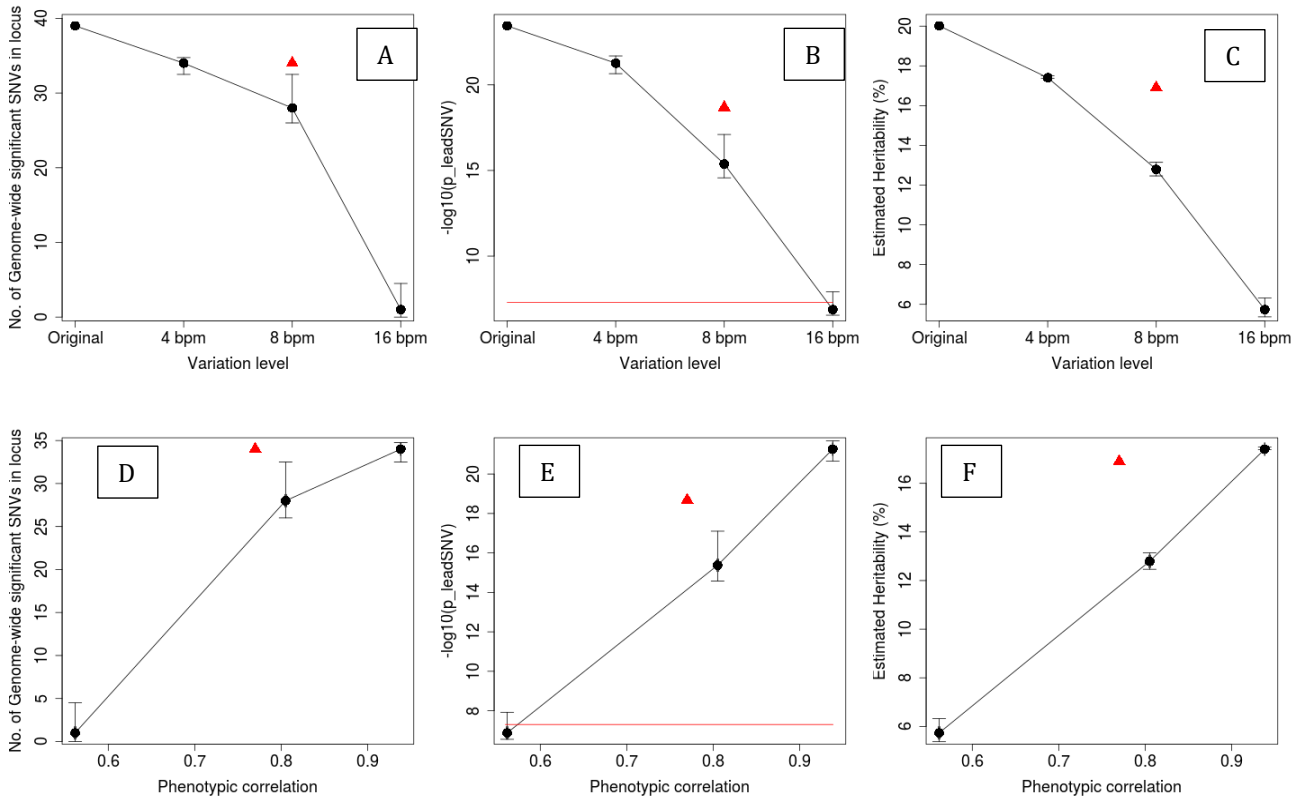
Fig. 2: Effect of simulated and real physiological variation in RHR on GWAS results and heritability estimates. (A) number of genome-wide significant (GWS) SNVs at MYH6 locus, (B) –log10 of the P-value of the lead variant and genome-wide significance threshold (p=5x10⁻8, red horizontal line) , and (C) estimated heritability, all plotted as function of the simulated variation level and with respect to the original results. Bottom row shows same parameters but plotted against phenotypic correlation between the simulated and original RHR data. Black circles represent median values from the simulation results and corresponding interquartile ranges are given by error bars. For comparison, red triangles present the results from comparing the two original RHR recordings.

5.7% (IQR: 5.4%-6.3%) for 0.5x, 1x, and 2x σHR, respectively, compared to 20.0% for the original data. Fig. 2D, 2E, and 2F show a direct relationship between genetic association and heritability results when plotted as function of the phenotypic correlation.

4.   Discussion

This work formally demonstrates the impact of physiological variability in RHR on the discovery of associated genetic variants. The main finding is that increased variation in RHR attenuates the strength of genetic associations and heritability estimates, but we did not observe evidence for spurious associations (e.g. false-positive associations).

Variability in RHR was observed between the two original RHR recordings. The fact that he second recording was taken just before an exercise stress test may have increased the variability as subjects may have experienced an anticipatory stress response to the actual stress test RHR[9]. We observed that the genetic associations were attenuated in the second recording demonstrating a potential effect of increased RHR variability. This finding supports previous studies that have shown that averaging blood pressure measurements

increased statistical power to detect genetic associations compared to a single-measurement GWAS[10].

The results from the physiological study were compatible with the results from the simulation study. The simulation study highlights important implications on GWAS results. For example, even for 8 bpm of variation, on average only 2 out the 3 loci were discovered and the estimated heritability dropped from 20% to approximately 15%. Although our results are based on the analysis of a single chromosome, our results suggest a potential loss of discoverable loci over the whole genome because of (increased) physiological variation in heart rate.

This work is a first explorative study on the effect of RHR variation on GWAS results and has some limitations. RHR variations were modelled as independent white Gaussian noise, whereas, physiologically, the magnitude of the variation may depend on the RHR itself. For example, Fig. 2 shows that simulation data slightly overestimates the decrease in genetic association strength compared to the original data. Furthermore, we did not observe evidence of spurious associations, but the number of simulations was limited and we only investigated a section of the genome (chromosome 14). Future studies may increase the number of genetic variants to further evaluate false-

positive associations. However, given the consistency of increasing *P*-values and decreasing heritability measures, we suspect false-positive results are unlikely to occur.

In summary, this work highlights the important implications of physiological variability of RHR on the genetic association results. The importance of averaging measurements to improve accuracy is well established, but our results provide a more detailed overview of the impact of different levels of variation on genetic results for a single-measurement RHR GWAS. Whilst results focus on RHR, they may have broader implications, as many cardiovascular traits are prone to physiological variation.

Acknowledgments

References

[1]   A. I. Aladin *et al.*, "Relation of resting heart rate to risk for all-cause mortality by gender after considering exercise capacity (the Henry Ford exercise testing project)," (in eng), *Am J Cardiol,* vol. 114, no. 11, pp. 1701-6, Dec 1 2014, doi: 10.1016/j.amjcard.2014.08.042.

[2]   M. W. Russell, I. Law, P. Sholinsky, and R. R. Fabsitz, "Heritability of ECG measurements in adult male twins," *Journal of electrocardiology,* vol. 30, pp. 64-68, 1998.

[3]   M. Den Hoed *et al.*, "Identification of heart rate–associated loci and their effects on cardiac conduction and rhythm disorders," *Nature genetics,* vol. 45, no. 6, p. 621, 2013.

[4]   M. E. van den Berg *et al.*, "Discovery of novel heart rate-associated loci using the Exome Chip," *Human molecular genetics,* vol. 26, no. 12, pp. 2346-2363, 2017, doi: 10.1093/hmg/ddx113.

[5]   R. N. Eppinga *et al.*, "Identification of genomic loci associated with resting heart rate and shared genetic predictors with all-cause mortality," vol. 48, no. 12, pp. 1557-1564, 2016, doi: 10.1038/ng.3708.

[6]   W. F. Boron and E. L. Boulpaep, *Medical Physiology*. Elsevier Health Sciences, 2008.

[7]   T. J. Littlejohns, C. Sudlow, N. E. Allen, and R. Collins, "UK Biobank: opportunities for cardiovascular research," *European heart journal,* vol. 40, no. 14, pp. 1158-1166, 2017.

[8]   P. R. Loh *et al.*, "Efficient Bayesian mixed-model analysis increases association power in large cohorts," (in eng), *Nat Genet,* vol. 47, no. 3, pp. 284-90, Mar 2015, doi: 10.1038/ng.3190.

[9]   D. B. Penzien, K. G. Hursey, H. Kotses, and H. A. Beazel, "The effects of anticipatory stress on heart rate and T-wave amplitude," (in eng), *Biol Psychol,* vol. 15, no. 3-4, pp. 241-8, Nov-Dec 1982, doi: 10.1016/0301-0511(82)90046-1.

[10]   S. K. Ganesh *et al.*, "Effects of long-term averaging of quantitative blood pressure traits on the detection of genetic associations," (in eng), *Am J Hum Genet,* vol. 95, no. 1, pp. 49-65, Jul 3 2014, doi: 10.1016/j.ajhg.2014.06.002.

Address for correspondence:

Name: Stefan van Duijvenboden
Full postal address: Institute of Cardiovascular Science, UCL, Gower Street, London, UK, WC1E 6BT.
E-mail address: ucemvan@ucl.ac.uk