Available online at www.sciencedirect.com

ScienceDirect

Journal of Electrocardiology xx (2016) xxx – xxx

JOURNAL OF Electrocardiology

www.jecgonline.com

Genome-wide association studies and resting heart rate

Tuomas O. Kilpeläinen, PhD*

The Novo Nordisk Foundation center for Basic Metabolic Research, Section of Metabolic Genetics, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

Abstract Genome-wide association studies (GWASs) have revolutionized the search for genetic variants regulating resting heart rate. In the last 10 years, GWASs have led to the identification of at least 21 novel heart rate loci. These discoveries have provided valuable insights into the mechanisms and pathways that regulate heart rate and link heart rate to cardiovascular morbidity and mortality. GWASs capture majority of genetic variation in a population sample by utilizing high-throughput genotyping chips measuring genotypes for up to several millions of SNPs across the genome in thousands of individuals. This allows the identification of the strongest heart rate associated signals at genome-wide level. While GWASs provide robust statistical evidence of the association of a given genetic locus with heart rate, they are only the starting point for detailed follow-up studies to locate the causal variants and genes and gain further insights into the biological mechanisms underlying the observed associations.

Keywords: Gene; Genetic variant; Single nucleotide polymorphism; Genomewide association study; Heart rate; Cardiovascular disease

Introduction

High resting heart rate is a well-recognized but poorly understood risk factor for cardiovascular morbidity and mortality [1]. While higher heart rate is associated with increased incidence of cardiovascular events [1], it remains unclear whether the link is causal or whether heart rate is only secondary to other factors explaining the increased risk.

Resting heart rate is determined by both genetic factors and the environment. The overall contribution of genetic factors to heart rate is strong; heritability estimates range between 55% and 70% in studies comparing monozygotic and dizygotic twins [2–4]. Identifying the genetic determinants of resting heart rate would be important in order to increase our understanding of how heart rate regulation is linked to cardiovascular morbidity and mortality. Genetic discoveries could also unlock doors for novel risk markers and therapeutics for cardiovascular disorders.

In the early 2000's, genetic studies aiming to identify loci associated with heart rate applied a candidate gene approach, querying associations for variants in one or few genes hypothesized to be involved in heart rate regulation. While many positive findings were initially published, the results were inconsistent and difficult to replicate (e.g [5,6]). In the last 10 years, genome-wide association studies (GWASs) have revolutionized the study of the genetic basis of resting heart rate by identifying at least 21 novel heart rate associated loci [7–10]. In contrast to candidate gene studies, GWASs do not rely on previous biological hypotheses but utilize an agnostic approach, aiming to identify new genes previously unknown to play a role in a human trait or disease by interrogating hundreds of thousands to few millions of genetic variants across the genome in large population samples.

The present paper will describe how GWASs are performed, review the recent discoveries in GWASs of human resting heart rate, and elucidate the scientific and clinical impact of these discoveries.

Review

What is a genome-wide association study?

The aim of GWASs is to capture the majority of genetic variation in a population sample and relate these variants to heart rate or another human complex trait. To achieve this goal, GWASs examine associations of hundreds of thousands to millions of single nucleotide polymorphisms (SNPs), the most common form of human genetic variation, across the genome. The variation at SNPs is determined using high-throughput genotyping chips that allow the identification of heart rate loci at genome-wide level. Alternatively, DNA sequencing can be applied to achieve a

* The Novo Nordisk Foundation Center for Basic Metabolic Research, Section of Metabolic Genetics, Universitetsparken 1, DIKU building, DK-2100, Copenhagen, Denmark.

E-mail address: tuomas.kilpelainen@sund.ku.dk

http://dx.doi.org/10.1016/j.jelectrocard.2016.07.022
0022-0736/

more extensive coverage of genetic variants, but the high costs of sequencing have thus far limited its use in large-scale studies.

Typically, the SNP genotypes are coded as 0, 1, or 2, representing the number of "effect" (*e.g.* heart rate increasing) alleles an individual carries. The association between the number of effect alleles and heart rate is examined for all SNPs in the genome-wide dataset. Given the very large number of statistical tests performed in GWASs, the chance of a false-positive finding is high. Therefore, the P value required to consider an association significant must be very stringent. Hence the 'genome-wide significant' P value of 5×10^{-8} has been conventionally applied as the minimum threshold to be reached, corresponding to P = 0.05 corrected for a million independent statistical tests ($0.05/1{,}000{,}000 = 5 \times 10^{-8}$).

Genotype reference panels and imputation

The implementation of GWASs was enabled by increasing knowledge of the human genome over the early 2000's. The Human Genome Project, completed in 2003, provided the backbone for GWASs by sequencing all 3 billion base pairs of the human genome [11]. The International HapMap Project, launched in 2002, extended the work by cataloguing the locations of 3.1 million common SNPs in diverse human populations and by making the information publically available [12]. Importantly, HapMap also provided information on correlation patterns between common SNPs across the genome. Correlation patterns exist as nearby genetic variants are often inherited together, making it possible to cover the majority of common genetic variation by genotyping only a subset of variants that are informative of many other SNPs. Indeed, the design of genotyping microchips has strongly relied on SNP correlations catalogued in the HapMap database to capture the majority of common genetic variation by carefully selecting informative tag SNPs.

To make as efficient use of chip genotyping data as possible, the genotypes of SNPs that have not been genotyped are often imputed based on the HapMap reference panel [13]. Recently, an extended reference panel from the 1000 Genomes Project has largely replaced the HapMap panel as it provides a more dense coverage of both common and less frequent genetic variants across the human genome [14]. Imputing SNPs from the same reference panel across many population samples makes it possible to generate consistent SNP data across multiple studies even if the studies have used different chips for genotyping, thus maximizing SNP sample sizes in meta-analyses of GWASs. While imputation allows coherent meta-analyses of GWASs, it also increases the resolution of GWASs by providing a more dense coverage of SNPs, helping to fine-map the association signal [15].

Follow-up of GWAS discoveries

The discoveries in GWASs are statistical associations that do not tell us directly about the mechanisms of how the association takes place. The identification of a novel heart rate increasing signal indicates that one or more genetic variants in the genetic region have biological functions that lead to the observed association. A critical aim after the initial observation is to identify the causal variants and their target genes. Only few of the GWAS-identified variants are predicted to impact protein-coding regions with the vast majority being located in non-coding, intronic or intergenic, regions and likely to modify the function of one or more nearby genes [16]. Most GWAS-identified heart rate loci contain several genes that could explain the observed associations and the causal genes have not been unequivocally established. For some identified variants, the correlation pattern with other variants is complex and spans very large regions in the genome, making the identification of the causal genes and variants difficult (Fig. 1).

The follow-up of GWAS discoveries involves various strategies to identify the causal variants and genes and gain insight into their role and function. Examples of follow-up analyses commonly applied in GWASs are studies of correlation of the identified variants with nearby functional variants or with the expression of nearby genes, analyses of overlap with functional regulatory elements of the genome, and studies in transgenic cell culture or animal models [16]. *In silico* follow-up of the identified variants in other GWASs for related traits may give a more refined phenotypic understanding of the observed associations. For example, examining associations of GWAS-identified heart rate loci in GWASs of ECG-traits and cardiovascular morbidity and mortality has given important insights into the physiological roles of the identified loci [8].

GWAS discoveries for heart rate

Implementation of the GWAS approach has led to a rapid pace of discovery of novel genetic loci unequivocally associated with resting heart rate. Since 2009, seven GWASs of resting heart rate have been published, identifying at least 21 novel heart rate associated loci (Table 1).

The first published GWAS for resting heart rate included 8842 Korean individuals, identifying a locus near the *GJA1* (*gap junction protein, alpha 1*) gene, well known for its role in the synchronized contraction of the heart and a major component of cardiac gap junctions. In addition, a second heart rate associated locus was identified near the *CD34* (*hematopoietic progenitor cell antigen CD34*) gene. Soon after the first GWAS, another GWAS of heart rate was published, including up to 23,112 Icelanders [10]. The study identified one novel heart rate locus in the *MYH6* gene encoding alpha-myosin cardiac heavy chain protein.

Larger sample sizes were required for the identification of additional genetic loci with smaller effects, leading to an extensive collaboration to pool results in a meta-analysis of 15 GWASs within the Cohorts for Heart and Aging Research in Genome Epidemiology (CHARGE) Consortium [9]. In this meta-analysis, including a total of 38,991 individuals of European ancestry, four novel heart rate loci were identified in or near the *SOX5*, *SLC35F1*, *SLC12A9,* and *FADS1* genes. In addition, the consortium confirmed the three previously published loci in or near the *GJA1*, *CD34*, and *MYH6* genes (Table 1).