Biochimica et Biophysica Acta 1842 (2014) 1923–1931

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbadis

Review

Genetics of the human metabolome, what is next?☆

Harish Dharuri [a], Ayşe Demirkan [a,b], Jan Bert van Klinken [a], Dennis Owen Mook-Kanamori [c,d,e], Cornelia M. van Duijn [b], Peter A.C. 't Hoen [a], Ko Willems van Dijk [a,c,*]

[a] Department of Human Genetics, Leiden University Medical Center, Leiden, Netherlands
[b] Genetic Epidemiology Unit, Departments of Epidemiology and Clinical Genetics, Erasmus University Medical Center, Rotterdam, Netherlands
[c] Department of Endocrinology, Leiden University Medical Center, Leiden, Netherlands
[d] Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, Netherlands
[e] Department of Physiology and Biophysics, Weill Cornell Medical College—Qatar, Doha, Qatar

ARTICLE INFO

Article history:
Received 29 November 2013
Received in revised form 4 May 2014
Accepted 28 May 2014
Available online 4 June 2014

Keywords:
Metabolomics
Genome-wide association studies (GWAS)
Pathway
Analysis
Systems biology

ABSTRACT

Increases in throughput and decreases in costs have facilitated large scale metabolomics studies, the simultaneous measurement of large numbers of biochemical components in biological samples. Initial large scale studies focused on biomarker discovery for disease or disease progression and helped to understand biochemical pathways underlying disease. The first population-based studies that combined metabolomics and genome wide association studies (mGWAS) have increased our understanding of the (genetic) regulation of biochemical conversions. Measurements of metabolites as intermediate phenotypes are a potentially very powerful approach to uncover how genetic variation affects disease susceptibility and progression. However, we still face many hurdles in the interpretation of mGWAS data. Due to the composite nature of many metabolites, single enzymes may affect the levels of multiple metabolites and, conversely, levels of single metabolites may be affected by multiple enzymes. Here, we will provide a global review of the current status of mGWAS. We will specifically discuss the application of prior biological knowledge present in databases to the interpretation of mGWAS results and discuss the potential of mathematical models. As the technology continuously improves to detect metabolites and to measure genetic variation, it is clear that comprehensive systems biology based approaches are required to further our insight in the association between genes, metabolites and disease. This article is part of a Special Issue entitled: From Genome to Function.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

The "inborn errors of metabolism" as defined by Garrod at the beginning of the twentieth century depict the first clearly recognized examples of specific genetic defects leading to the accumulation of metabolites in body fluids [1]. For example, in alkaptonuria, a genetic defect in the enzyme homogentisate 1,2-dioxygenase leads to the accumulation of homogentisic acid and its oxide alkapton in plasma and urine. Detection of alkapton in urine is relatively simple in that exposure of urine from affected patients to air results in black discoloration that is readily detected by eye. Alkaptonuria is transmitted as a recessive Mendelian trait with near complete penetrance and is an example of a rare metabolic disease caused by rare genetic variants [2].

Changes in plasma metabolites are also pathogenic hallmarks of common metabolic diseases such as type-2 diabetes. The defining metabolic marker for type 2 diabetes is glucose, but hyperglycemia co-occurs with changes in a variety of additional metabolites including amino acids, lipids and lipoproteins. The high heritability of type 2 diabetes is not explained by rare genetic variants segregating in families, but is thought to be caused by a variety of, and presumably combination of common genetic variants. This paradigm is referred to as "common disease–common variant" hypothesis and is pursued in so-called genome wide association studies (GWAS). In GWAS, genome wide genotyping platforms measure genotypes for hundred thousand to millions of single nucleotide polymorphisms (SNPs) with minor allele frequencies (MAF) generally larger than 0.05 and test each of those SNPs for association with a specific trait [3]. A large number of GWAS have been performed with a variety of both binary traits (e.g. type 2 diabetes) and quantitative traits (e.g. fasting glucose levels). These studies have successfully uncovered genetic variants that contribute to disease risk and also to the variation in quantitative phenotypes [4]. For

Abbreviations: mGWAS, metabolite genome wide association study; MAF, minor allele frequency; SNP, single nucleotide polymorphism; LD, linkage disequilibrium; eQTL, expression quantitative trait locus; GIM, genetically influenced metabotype; MS, mass spectrometry; GC, gas chromatography; LC, liquid chromatography; NMR, nuclear magnetic resonance; TCF7L1, transcription factor 7-like 2; FADS, fatty acid desaturase; GLS2, glutamine synthase; SLC16A9, solute carrier family 16 member 9; GGM, Gaussian graphical modeling; PSEA, phenotype set enrichment analysis; GSMM, genome scale metabolic model; CBA, constraint-based analysis

☆ This article is part of a Special Issue entitled: From Genome to Function.

* Corresponding author at: Department of Human Genetics, Leiden University Medical Center, Leiden, Netherlands.
E-mail address: KoWvD@lumc.nl (K. Willems van Dijk).

http://dx.doi.org/10.1016/j.bbadis.2014.05.030
0925-4439/© 2014 Elsevier B.V. All rights reserved.

example, for type-2 diabetes, thus far, more than 60 risk loci have been identified, giving novel insights into the complex pathophysiology of the disease. However, the risk attributed to individual SNPs in the vicinity of even the strongest candidate gene, transcription factor 7-like 2 (*TCF7L2*), is relatively modest (odds ratios of 1.5–1.7) [5]. Moreover, the combined genetic loci discovered to date explain only a small proportion (less than 5%) of the observed heritability of type 2 diabetes. Thus, a significant proportion of the observed heritability remains to be uncovered [6].

Since a large proportion of the SNPs discovered through GWAS are intergenic or lie within the intronic regions of genes, rather than in the protein coding sequences, the genetic basis for the association is often not obvious. It is possible that the SNPs discovered through GWAS are in linkage disequilibrium (LD) with the real causal variant that is not captured by the platform. This hypothesis to uncover "missing heritability" is currently being tested by many labs using next generation deep sequencing approaches to screen the whole genome or whole exome to locate the functional variants. Unfortunately, thus far, these approaches have met with relatively limited success. This lack of success may be associated with our inability to recognize the causative variants among the many detected variants. Alternatively, GWAS hits may constitute expression quantitative trait loci (eQTLs) influencing the expression level of one or more genes nearby (*cis*-eQTLs), or at a distant physical location (*trans*-eQTLs) [7,8]. Recently, a combination of RNA and genome sequencing has provided in-depth insight into the relation between genetic variation and transcriptome variation and their association with functional variation [9].

Whereas it is often difficult to determine the effect of GWAS-discovered SNPs on nearby or distant genes, it is clear that many different genes and loci are involved in the pathogenesis of complex diseases such as type 2 diabetes. In addition, it is also clear that environmental factors including lifestyle (i.e. diet and physical activity) affect the development of diabetes. Therefore, it may be more appropriate to consider common metabolic disorders such as diabetes as the outcome of a variety and often combination of mild "inborn errors of metabolism" in conjunction with the environment. These mild "inborn errors of metabolism" would be reflected by differences in the concentrations of metabolites in cells and/or body fluids and could provide insight into the "missing heritability". The terms "genetically determined metabotype" (GDM) [10] and "genetically influenced metabotype" (GIM) have been coined for this [11]. GIM has been defined as relatively prevalent genetic variants that lead to substantial modification in the efficiency of metabolic conversions [12]. The combination of GIMs in any given individual determines his metabolic individuality and thus, in combination with environment and lifestyle, the risk for metabolic disorders such as type 2 diabetes.

2. Metabolomics measurements

The detection of GIMs has been facilitated by technological developments in the field of metabolomics, where it is now possible to simultaneously measure hundreds of metabolites in large sets of biological samples using automated procedures, and at relatively low cost (10s of euros per sample). A variety of metabolomics platforms are available, all having their own characteristics. Generally speaking, the metabolomics techniques can be divided in two types of platforms and two types of approaches. Metabolomics platforms based on mass spectrometry (MS) in general require extensive sample preparation and are used in-line with gas or liquid chromatography (GC-MS and LC-MS). In contrast, nuclear magnetic resonance (NMR) based platforms require relatively limited sample preparation and the samples can be analyzed without prior separation procedures. MS and NMR based platforms can be employed for targeted and/or non-targeted approaches. In a targeted approach, the platform is optimized for detection of a set of predefined metabolites and absolute or relative concentrations are determined using internal standards. In contrast, in a non-targeted approach, the platform is optimized to capture global snapshots of the test and reference samples and reports the differences. To subsequently identify the metabolites underlying the differential signal in the untargeted approach, additional analyses are required that are frequently challenging. Therefore, metabolomics datasets from a non-targeted approach often contain a large number of 'unknown' compounds. The main characteristic of all metabolomics platforms is that a subset of compounds can be detected based on common chemical properties of these compounds rather than their biological relatedness. No single analytical technique exists that is suitable for the identification and quantification of all endogenous metabolites in a sample.

Excellent reviews on the possibilities and challenges of the different metabolomics platforms and approaches are available [13–15]. In general, NMR spectroscopy is highly reproducible and quantitative. However, NMR spectroscopy is relatively insensitive and metabolite identification relies on specialized and mostly proprietary spectral deconvolution algorithms. These algorithms may not always identify the same metabolites and may not always base the identification of a specific metabolite on the same spectral signal. In contrast, MS based platforms provide highly precise information on metabolite mass from which identity can often be inferred. However, metabolite quantification requires spiked internal standards. Thus, a common challenge in metabolomics on any platform is the reproducibility of reported metabolite levels across different laboratories. In addition to these platform-specific challenges, additional variability may be caused by differences in instrumentation and experimental setup conditions such as sample preparation and extraction method, collection protocols, source material (plasma, serum, urine, etc.), but also sample storage conditions and batch effects. These aspects all require careful consideration when replicating observations and pooling metabolomics data for meta-analyses.

3. Genome wide association studies of metabolomics data

Since metabolomics data are (semi)quantitative, they are suited for metabolomics GWAS (mGWAS), uncovering genetic variants that affect metabolite levels. One of the first studies employed an MS-based platform that could identify and quantify up to 363 metabolites in 284 individuals [10]. The study reported that common SNPs explained up to 12% of the observed variance in metabolite levels. Moreover, the study determined that the explained variance could be dramatically increased by considering ratios of metabolites. This is because analyzing ratios of metabolite concentrations potentially reduces the variation in the dataset when the pair of metabolites is related to the substrate and product of a given enzymatic reaction. Furthermore, where a SNP impacts such a metabolic reaction, consideration of ratios leads to a dramatic reduction in the *p*-value of association. For example, rs174548, a SNP in an intron of the fatty acid delta-5 desaturase 1 (*FADS1*) gene is associated with a phosphatidylcholine moiety, PC C36:4 (36 denotes the number of carbons in the side chains and 4 denotes the number of double bonds) levels with a *p*-value of 4.52×10^{-8}, slightly above the genome-wide threshold. However, association of the same SNP with the ratio of PC C36:4/PC C36:3 has a *p*-value of 2.4×10^{-22}, a reduction by 14 orders of magnitude. The FADS1 enzyme introduces a double bond in long chain polyunsaturated fatty acids and the moieties PC C36:3 and PC C36:4 are related to the substrate and product of this enzymatic reaction.

A consistent theme that has emerged from mGWAS is that significant SNP–metabolite associations point to the underlying biological mechanism. This is in contrast to GWAS of clinical endpoints where unravelling the underlying mechanism is often much more challenging. In addition to FADS1, several other associations have shown that the functional nature of the gene matches with the biochemical characteristics of the associated metabolite. For example, SNPs in the gene *GLS2* (glutamine synthase 2) have been found associated with glutamine [16,17].This is a biologically plausible association because the enzyme

Download English Version:

https://daneshyari.com/en/article/1904695

Download Persian Version:

https://daneshyari.com/article/1904695

Daneshyari.com