Archives of Biochemistry and Biophysics xxx (2015) 1–9

Contents lists available at ScienceDirect

Archives of Biochemistry and Biophysics

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

Review article

Biochemical insights from population studies with genetics and metabolomics

Karsten Suhre [a, b, *], Johannes Raffler [b], Gabi Kastenmüller [b, c]

[a] Department of Physiology and Biophysics, Weill Cornell Medical College – Qatar, Doha, Qatar
[b] Institute of Bioinformatics and Systems Biology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany
[c] German Center for Diabetes Research, Neuherberg, Germany

ARTICLE INFO

Article history:
Received 31 July 2015
Received in revised form
28 September 2015
Accepted 28 September 2015
Available online xxx

Keywords:
Metabolomics
Genome-wide association study
Genetic variation
Metabolic individuality
Partial correlation

ABSTRACT

Genome-wide association studies with concentrations of hundreds of small molecules in samples collected from thousands of individuals (mGWAS) access otherwise inaccessible natural genetic experiments and their influence on the metabolic capacities of the human body. By sampling the natural metabolic and genetic variability that is present in the general population, mGWAS identified over 150 associations between genetic variants and variation in the metabolic composition of human body fluids. Many of these genetic variants were found to be located in enzyme or transporter coding genes, whose functions match the biochemical nature of the associated metabolites. Associations identified by mGWAS can reveal novel biochemical knowledge, such as the function of uncharacterized genes, the biochemical identity of small molecules, and the structure of entire biochemical pathways. Here we review findings of recent mGWAS and discuss concrete examples of how their results can be interpreted in a biochemical context. We describe online resources that are available for mining mGWAS results. In this context, we present two concepts that also find more general applications in the field of metabolomics: strengthening of associations by looking at ratios between metabolite pairs and reconstruction of metabolic pathways by Gaussian graphical modeling.

Contents

* Corresponding author. Weill Cornell Medical College – Qatar, Qatar Foundation, Education City, PO Box 24144, Doha, Qatar.
E-mail address: karsten@suhre.fr (K. Suhre).

http://dx.doi.org/10.1016/j.abb.2015.09.023
0003-9861/

1. Introduction

Over one hundred fifty years of biochemical and molecular biology experimentation have helped to uncover the function and substrate specificities of many enzymes in organisms ranging from microbes to humans. This knowledge, generated by generations of biochemists and molecular biologists, fills today's reference volumes on biochemistry. Most early discoveries regarding enzyme and transporter function and their place on these biochemical pathway maps were the result of biomolecular experiments, initiated, for example, by the discovery of a microbial strain that lacked a specific metabolic capacity or the characterization of a human patient with a rare inborn error of metabolism. These investigations were generally hypothesis-driven and therefore used targeted biochemical assays that determined the concentrations of only a few biochemicals at a time.

This situation changed drastically with the recent advent of modern, highly sensitive and mass-resolving mass-spectrometers (MS) and strong nuclear magnetic resonance spectroscopes (NMR) [1]. Using MS- and NMR-based metabolomics approaches, hundreds and even thousands of small molecules (metabolites) can now be detected, identified and quantified in any biological sample, using only micro-liter amounts of fluid or tissue extract [2,3]. Equipped with high-throughput robotic processing capabilities, modern metabolomics platforms can analyze blood and urine samples from large population-based epidemiological studies with thousands of participants. Combined with the availability of high-throughput genotyping arrays, epidemiological studies are now generating biochemical hypotheses at a large scale by analyzing influences of genetic variation on metabolic phenotypes in genetic association studies [4].

However, biochemistry and epidemiology research rarely overlap: Molecular and cell-based assays and *in vivo* models are generally implemented to understand the molecular basis of a disease, while population-based cohort studies, combined with deep biomedical phenotyping, are applied to identify the predisposing factors of the same disease. In our experience, there is little natural overlap and flow of ideas between these communities. We therefore review here recent findings from population-based genome-wide association studies with metabolomics, with the objective to make these results more readily accessible to biochemists and molecular biologists.

2. Genome-wide association studies with metabolomics

Naturally occurring genetic variation as in inborn errors of metabolism has the potential to uncover the function, regulation and substrate specificity of biochemically active proteins, such as enzymes, solute transporters, and regulators of metabolism. Thus, the study of rare inborn errors of metabolism with large effects on metabolic homeostasis led to the identification and characterization of many enzymes. Prominent examples are the enzymes *phenylalanine hydroxylase* and *homogentisate 1,2-dioxygenase*, which were functionally characterized by studying phenylketonuria [5] and alkaptonuria [6], respectively. Over 100 years ago, Archibald Garrod assumed that “inborn errors of metabolism” are “merely extreme examples of variations of chemical behavior which are probably everywhere present in minor degrees” and that this “chemical individuality [confers] predisposition to and immunities from the various mishaps which are spoken of as diseases” [6,7]. Epidemiological studies, including the Framingham Heart Study [8], KORA [9], TwinsUK [10], SHIP [11], the Rotterdam Study [12], the African Americans in the Atherosclerosis Risk in Communities (ARIC) Study [13], and many others, collected demographic, health and life-style related information from thousands of individuals from the general population, and biobanked samples of blood, urine, and other body fluids that were then analyzed using genomics, transcriptomics, proteomics, metabolomics and other large scale -omics technologies. Genome-wide genotyping arrays allow for a broad characterization of the common, naturally occurring genetic variance in these study populations. Combined with a comprehensive quantification of the biochemical composition of matching samples using metabolomics, single nucleotide polymorphisms (SNPs) that associate with even small differences in metabolic phenotypes (metabotypes) can then be detected (Fig. 1).

A SNP-metabolite association at a genetic locus is referred to as a “metabolic quantitative trait locus” (mQTL). Due to the inherent correlation between metabolites that are on a same biochemical pathway, there are in general multiple metabolites that associate with a single genetic variant. Likewise, due to the high correlation between neighboring genetic variants (also known as linkage disequilibrium, see [14]), there are in general many SNPs that associate with the same metabolites at a given locus. The strength of these associations may vary between studies, dependent on the population size, the metabolomics platform that is used and other factors that determine the non-genetic part of the variation of the metabolic trait.

We call the ensemble of genotype-dependent differences in the concentrations of one or several metabolites at a genetic locus a “genetically influenced metabotype” (GIM). It should be born in mind that SNPs reported in GWAS are generally genotyping-array based tagging SNPs. These SNPs do not necessarily correspond to the causative variant that affects a gene function, possibly by modifying the protein structure through an amino acid change or a modification of a splice site, or by changing the gene's expression by altering a transcription factor binding site or a DNA methylation site. A GIM is thus used as a loosely defined term that refers to the association of a set of correlated SNPs with a set of correlated metabolites, possibly detected in multiple studies, where the GIM might even be labeled with a different putative causative gene.

Today, genome-wide association studies with metabolic traits (mGWAS) have identified over 150 GIMs [15–27] in general populations, confirming Garrod's conjecture in many instances [4]. These “variations of chemical behavior” (GIMs/mQTLs) are clearly “everywhere present in minor degrees” and represent what Archibald Garrod referred to as our “chemical individuality”.

3. Properties of genetically influenced metabotypes

A prototype of a GIM is the association of SNP rs174548 (and correlated SNPs) with arachidonic acid, several biochemically related omega-3 and omega-6 polyunsaturated fatty acids (PUFAs), and a large number of PUFA-containing glycerophospholipids, including phosphatidylcholines, phosphatidylethanolamines, and phosphatidylinositols [15,17,27,28]. SNP rs174548 is located near the *fatty acid desaturase 1* gene (FADS1). The enzymatic function of FADS1 is that of a delta-5 desaturase. It inserts a fourth double bond into dihomolinolenate (C20:3) to form arachidonic acid (C20:4) (Fig. 1). In the case of the FADS1 GIM, its metabotype consists of PUFAs that are upstream and downstream of the FADS1 reaction, and of metabolites that incorporate these PUFAs, i.e. through the de-novo synthesis of glycerophospholipds in the Kennedy pathway. Interestingly, one of the strongest associations of SNP rs174548 were observed with the ratio of the product and substrate pair (C20:4/C20:3) of the FADS1 enzymatic reaction. It can be shown that, under a number of simplifications, this ratio is proportional to the conversion rate of the FASD1 reaction [15]. In this case there is, thus, a perfect match between the function of the gene and the associated metabolic phenotype. Had the function of FADS1 not

Please cite this article in press as: K. Suhre, et al., Biochemical insights from population studies with genetics and metabolomics, Archives of Biochemistry and Biophysics (2015), http://dx.doi.org/10.1016/j.abb.2015.09.023