INVITED REVIEW

Genetics of human metabolism: an update

Gabi Kastenmüller¹⁴, Johannes Rafpler¹, Christian Gieger²³ and Karsten Suhre¹⁵,*

¹Institute of Bioinformatics and Systems Biology, ²Research Unit of Molecular Epidemiology and ³Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany, ⁴German Center for Diabetes Research, Neuherberg, Germany and ⁵Department of Physiology and Biophysics, Weill Cornell Medical College—Qatar, Doha, Qatar

*To whom correspondence should be addressed at: Weill Cornell Medical College—Qatar, Qatar Foundation—Education City, PO Box 24144, Doha, Qatar. Tel: +974 4492 8482; Email: karsten@suhre.fr

Abstract

Genome-wide association studies with metabolomics (mGWAS) identify genetically influenced metabotypes (GIMs), their ensemble defining the heritable part of every human’s metabolic individuality. Knowledge of genetic variation in metabolism has many applications of biomedical and pharmaceutical interests, including the functional understanding of genetic associations with clinical end points, design of strategies to correct dysregulations in metabolic disorders and the identification of genetic effect modifiers of metabolic disease biomarkers. Furthermore, it has been shown that GIMs provide testable hypotheses for functional genomics and metabolomics and for the identification of novel gene functions and metabolite identities. mGWAS with growing sample sizes and increasingly complex metabolic trait panels are being conducted, allowing for more comprehensive and systems-based downstream analyses. The generated large datasets of genetic associations can now be mined by the biomedical research community and provide valuable resources for hypothesis-driven studies. In this review, we provide a brief summary of the key aspects of mGWAS, followed by an update of recently published mGWAS. We then discuss new approaches of integrating and exploring mGWAS results and finish by presenting selected applications of GIMs in recent studies.

Introduction

Homeostasis of blood metabolites and excretion of metabolites through urine are crucial for maintaining human health. Therefore, dysregulations of metabolite levels often indicate critical physiological states, a fact that is widely used in clinical chemistry for the diagnosis of common metabolic diseases, such as diabetes and chronic kidney disease, but also of rare metabolic disorders, including inborn errors of metabolism, such as phenylketonuria and medium-chain acyl-CoA dehydrogenase deficiency. The number and variety of metabolites that are detectable and quantifiable in biological samples in a single experiment largely increased through the advent of modern metabolomics techniques, which are mainly based on nuclear magnetic resonance spectroscopy or mass spectrometry coupled to gas or liquid chromatography (1). As a consequence, blood and urine levels of a broad range of metabolites can be determined on an epidemiological scale from bio samples of thousands of individuals, opening new avenues to study human metabolism and its variation in health and disease based on large population studies (2).

Metabolite levels and their variation in the human body are influenced by various factors, specifically environmental conditions (day light, exposure to xenobiotics, etc.), lifestyle (nutrition...
habits, smoking and alcohol consumption, physical activity, etc.) and genetics. The genetic control of metabolite levels and its impact on human health are apparent in inborn errors of metabolism, where rare genetic variants disrupt a single gene that then leads to extreme, eventually toxic levels of related metabolites. Availability of population scale metabolomics and genotype data now allows to systematically investigate also the less apparent influences of more common and less deleterious genetic variants on human metabolism by conducting genome-wide association studies with metabolomics (mGWAS). This has been demonstrated in the first mGWAS by Gieger et al. (3), in which the authors performed a genome-wide scan for associations of common single-nucleotide polymorphisms (SNPs) with the levels of >350 metabolites and all pairwise ratios thereof. Despite the relatively small sample size of this first mGWAS (n = 284), they identified four common SNPs that significantly alter functionally related metabolic traits. Moreover, they found that ratios between metabolites can serve as proxies for the biochemical conversion of metabolites and thereby significantly strengthen the associations, a concept that has been helpful in many subsequent studies (4).

SNP-metabolite trait associations [metabolite quantitative trait loci (mQTLs)] identified in mGWAS typically cluster in groups of SNPs in high linkage disequilibrium (LD) that are associated with the same and/or biochemically related metabolites. Each of these groups of mQTLs defines a distinct genetically influenza metabolic (GIM). Many GIMs identified so far share most of the following properties (3): (i) The variance explained by the common genetic variants in the observed metabolic traits is often large, exceeding 10%, and the effect sizes are strong, with minor allele homozygotes sometimes displaying 50% differences in their estimated enzymatic throughput compared with major allele homozygotes. (ii) For the majority of GIMs, the metabolic traits can be functionally linked to an enzyme, a transporter or a regulator of metabolism that is encoded at the genetic locus (e.g. the associated metabolites being substrates or products of the encoded enzyme). (iii) Associations found in GWAS with clinical end points are enriched in GIMs, and the related metabolotypes are candidate intermediate traits on the pathway to the disease [for details on the concept of the intermediate phenotype, see (5)]. For instance, Ried et al. (6) demonstrated the potential of mGWAS results to infer asthma-related metabolic markers by the identification of potentially deregulated phospholipids that associate with asthma and asthma risk alleles.

Recent GWAS with Metabolic Traits

Recent mGWAS have extended previous studies in samples size, variety of metabolic traits and depth of genetic analysis (Table 1). As a result, the number of known GIMs largely increased. At the same time, most previously reported associations could be replicated in various cohorts, often using different metabolomics platforms and sometimes also different biofluids. Moreover, the functional interpretation of GIMs remarkably improved as recent studies started to systematically combine GIMs with additional data, such as expression quantitative trait loci (eQTLs) and metabolic pathway information. Thus, mGWAS published in recent years did not only lead to a more comprehensive and refined view on the inherited part of human metabolic individuality, but it also facilitated establishing potential links of this individuality to disease.

The largest leap in increasing the number of GIMs in the last 2 years has been achieved in the recent study by Shin et al. (23). The authors reported 145 GIMs (Fig. 1), 84 of which were new. This mGWAS was based on relative quantification from liquid chromatography-mass spectrometry and gas chromatography-mass spectrometry (GC-MS) for 486 metabolites in blood samples of 7824 participants from two European cohorts. The metabolites and their associated genetic loci broadly cover representatives of all major metabolic pathways, providing the most comprehensive picture of how genetic variation affects homeostasis in blood metabolism to date. To facilitate further exploration and functional interpretation of the findings by the biomedical research community, all GIMs have been embedded into the metabolic network, which was reconstructed from the metabolomics data (Fig. 2). This network as well as a rich set of additional information (including mapped eQTLs, GWAS hits to clinical phenotypes, and drug-target information) is available at http://gwas.eu/si.

Recent mGWAS with similar or even larger sample sizes include the study by Draisma et al. (26), who analyzed 129 mostly lipid-related metabolites in 7478 individuals from seven European cohorts. They identified GIMs at 31 genomic loci, although only 4 of them were new. Nonetheless, the study refined patterns of associations of related lipid traits at already known loci and provided new insights into the complexity of lipid metabolism. Another example is the recent mGWAS by Rhee et al. (20), who identified specific patterns in the genetic associations of 46 triglycerides with different lengths of the fatty acid chain and degree of desaturation. Some of their loci have been previously linked to total triglyceride levels in blood, namely variants in glucokinase (hexokinase 4) Regulator (GCKR), FADS1-3 and APOA1/C3/A4/A5. While variants in GCKR showed stronger associations to triglycerides with shorter fatty acid chains, associations of variants in FADS1-3 were stronger with longer fatty acid chain lengths and a higher degree of desaturation, which is in line with the fatty acid desaturase function of the proteins encoded at the FADS1-3 locus.

While most early mGWAS were based on samples from individuals of European ancestry, Rueedi et al. (21) recently analyzed genetic associations of nuclear magnetic resonance (NMR)-derived metabolic traits in urine samples from 835 Europeans and replicated their findings in 601 samples from a Brazilian population with a diverse ethnic background. Another recent study by Yu et al. (22) performed an mGWAS in serum samples from almost 2000 individuals of African-American descent and thereby confirmed the robustness of various GIMs across different ethnicities, but also found novel independent variants at the same loci and even new GIMs, which have not been observed in Europeans before.

In addition to fully metabolome-wide mGWAS, many GWAS with smaller and more specialized sets of metabolic traits have been conducted. For instance, Stiles et al. (27) report genetic, anatomic and clinical determinants of human serum sterol and vitamin D levels, Ng et al. (28) analyzed genetic influences on blood levels of polychlorinated biphenyls and Xie et al. (29) conducted a GWAS focused on metabolites related to insulin sensitivity. The CHARGE Consortium (30) reported novel loci associated with plasma concentrations of four fatty acids in the de novo lipogenesis pathway. Further studies identified and refined associations with single metabolic traits, including disease-relevant markers, such as bilirubin (31–34), uric acid (35–38), dimethylarginine (39), homoaarginine (40) and creatinine (41). These studies could identify new loci due to their much larger sample sizes, in some cases over 100,000 individuals, while using less expensive biochemistry-based methods.

Interestingly, most conclusions from mGWAS in humans also apply to animals and plants. Ghazalpour et al. (42) studied the genetic regulation of mouse liver metabolite levels. By analyzing 283 metabolites in 104 inbred and recombinant inbred mouse...
strains, they identified 240 loci, the majority of which accounted for 20–40% of total metabolite variation. Remarkably, more than one-third of the loci that regulate liver metabolites in mice also correspond to human GIMs, supporting the similarity in genetic regulation of metabolites between mice and humans. Chen et al. (43) conducted an mGWAS in rice, covering 840 metabolites and 6.4 million SNPs obtained from 529 diverse accessions of Oryza sativa. This study identified hundreds of common variants influencing numerous secondary metabolites with large effect sizes, and reported 36 candidate genes that modulate levels of metabolites of potential physiological and nutritional importance. The authors concluded that mGWAS provide a powerful tool for large-scale interactive gene–metabolite annotation and identification, pathway elucidation and knowledge about crop improvement.

New Concepts
Identification of non-targeted metabolic traits

Many of the more recent mGWAS were based on non-targeted metabolomics approaches, which record all metabolite signals detectable by the specific NMR or MS method, including signals that could not be assigned to a specific biochemical molecule.

Table 1. List of published mGWAS in humans and selected GWAS with multiple metabolic traits (in the order of publication date)

| Biofluid       | Metabolic trait(s)                      | Traits (N) | Platform | Study population       | Cohort size | Loci (n) | References               |
|----------------|-----------------------------------------|------------|----------|------------------------|-------------|----------|--------------------------|
| Serum          | Targeted MS                             | 363 + all ratios | MS       | German                 | 284         | 4        | Gieger et al. (3)       |
| Serum          | Sphingolipids                           | 33 + 43 ratios | MS       | European               | 4400        | 5        | Hicks et al. (7)        |
| Serum          | Mainly phospholipids                    | 163 + 26 406 ratios | MS       | German, British        | 1809 + 422  | 9        | Illig et al. (8)        |
| Serum          | NMR-derived metabolites                 | 59 + 1661 ratios | NMR      | German                 | 862 + 992  | 5        | Suhre et al. (9)        |
| Serum          | Non-targeted MS, knowns                 | 276 + 37 179 ratios | MS       | German, British        | 1768 + 1052 | 37       | Suhre et al. (10)       |
| Urine and Plasma | Urine: NMR peaks, plasma: mainly phospholipids | 163 + ratios | NMR + MS | British                | 211         | 3        | Nicholson et al. (11)   |
| Serum          | Mainly lipid traits                     | 117 + 99 ratios | NMR      | Finnish                | 8330        | 31       | Kettunen et al. (12)    |
| Plasma         | Phospholipids + sphingolipids           | 153         | MS       | European               | 4034        | 35       | Demirkan et al. (13)    |
| Serum          | Mainly lipid traits                     | 117 + 99 ratios | NMR      | Finnish                | 8330        | 30       | Tukiainen et al. (14)   |
| Serum          | Mainly lipid traits and low-weight      | 130         | NMR      | Finnish, British       | 1905 + 4703 | 34       | Inouye et al. (15)      |
| Serum          | Non-targeted MS, unknowns               | 517         | MS       | German                 | 1768        | 34       | Krumsierek et al. (16)  |
| Urine          | NMR peaks                               | 2425        | NMR      | Brazil                 | 265         | 2        | Montoliu et al. (17)    |
| Plasma         | NMR peaks                               | 8600 + 124 750 ratios | NMR    | German                 | 1757        | 7        | Raffler et al. (18)     |
| Serum          | MS peaks                                | 6138        | MS       | Swedish                | 402 + 489   | 7        | Hong et al. (19)        |
| Plasma         | Amino acids, amines, polar metabolites, lipids | 217         | MS       | USA (European ancestry)| 2076        | 31       | Rhee et al. (20)        |
| Urine          | NMR peaks                               | 1276        | NMR      | European, Brazil       | 835 + 601   | 11       | Rueedi et al. (21)      |
| Serum          | Non-targeted MS, knowns and unknowns    | 308         | MS       | African                | 1260        | 19       | Yu et al. (22)          |
| Serum          | Non-targeted MS, knowns and unknowns    | 486 + 98 346 ratios | NMR    | European               | 7824        | 145      | Shin et al. (23)        |
| Serum          | Targeted MS (mainly phospholipids) + non-targeted MS (knowns) | 344 (151 + 193) | MS       | European               | 1809 + 843  | 12 new | Ried et al. (24)        |
| Serum          | NMR-derived metabolites                 | 42          | NMR      | European               | 2118        | 8        | Demirkan et al. (25)    |
| Serum          | Mainly phospholipids                    | 129         | MS       | European               | 7478 + 1182 | 31       | Draisma et al. (26)     |

Figure 1. Manhattan plot of the mGWAS by Shin et al. (23). Upward pointing P-values: TwinsUK cohort, downward pointing P-values: KORA population study. Only SNPs with association to raw metabolites (P < 10^{-6}) are displayed (no ratios). The green line indicates the genome-wide significance cutoff for the P-value (P < 10^{-8}). Loci that reach genome-wide significance in either cohort are indicated by a short vertical black line. Loci with P-values <10^{-30} are indicated with a red symbol. Loci that are further discussed in Figure 3 are highlighted and annotated [figure adapted from the Supplementary Material, Fig. S2 by Shin et al., Nature Genetics, 2014 (23)].
(unknown metabolite). In contrast to targeted approaches that use optimized methods for the quantification of a set of predefined metabolites, non-targeted approaches usually provide less precise (relative) quantifications, but this for a biochemically broader range of metabolites. While less biological information can be gained from GIMs with signals of unknown identity, it has been shown that genetic association of data for unknown metabolites may allow metabolite identification in MS (16) and NMR (18,21). Often both known and unknown signals were considered in mGWAS to explore the entire breadth of the metabolite data from these non-targeted approaches. In 2013, Raffler et al. (18) reported an mGWAS with non-targeted NMR traits in blood plasma. While the concept of testing NMR signal intensities has already been introduced in previous mGWAS (11,17), this study was the first where the concept of testing ratios between NMR signals at different spectral positions was applied, resulting in the identification of four additional loci that displayed genome-wide significant association signals. To elucidate the chemical identity of the metabolites that underlay the genetically associated NMR signals, the authors used pseudo-spectra to visualize either the strength of genetic associations of NMR signals (‘association spectra’) or the correlation between NMR signals and traits determined on complementary metabolomics platforms, such as MS and clinical biochemistry measurements (‘correlation spectra’). In 2014, Rueedi et al. (21) presented an NMR-based mGWAS in urine where such pseudo-spectra were automatically annotated. There, the authors introduced the ‘metabomatching’ approach that compares association spectra with pure compound spectra derived from the Human Metabolome Database (HMDB) (44). Using this metabomatching approach, Rueedi et al. (21) could identify the metabolic nature of the signals in 6 out of 11 loci reported in their study. Hence, several methods are now available that allow to deduce the biochemical identity of non-targeted metabolic traits using genetic association data.
**Exome sequencing and imputation**

Identification of causative variants in GWAS is always a major challenge. Intermediate phenotypes with large effect sizes, as obtained from metabolomics, may be useful to figure out the causative variants of associations with complex clinical end points when the genetic structure of the association with the intermediate trait and the disease end point are the same. For the first time, exome-sequencing-based metabolomics associations were reported by Demirkol et al. (25). The authors first performed an NMR-based mGWAS in serum using microarray genotype data of 2118 individuals. Following the mGWAS, they selected a subset of the GWAS cohort (n = 921) and performed exome sequencing on candidate genes within eight candidate loci. They identified seven variants in or near four genes that modulate metabolite levels independently of the GWAS hits. For instance, the common SNP rs1047891 is a missense variant in CPS1 that was tagged by the genotyped SNP rs715. Interestingly, the association of rs1047891 to glycine was also previously reported by Shin et al. (23) using a fine-mapping approach of candidate loci based on 1000 Genomes Project data imputed genetic variants [see the Supplementary Material, Table S8 in Shin et al. (23)]. However, except for one variant, this fine-mapping approach did not significantly strengthen any association signal that was not already apparent using the genotyped SNPs. Fine mapping of association signals using imputed or exome-sequenced variants may thus help single out potentially causative SNPs; but for the moment, there is little evidence that it will help find truly novel association signals at the mGWAS level. This may change with larger sample numbers.

**Non-additive genetic models**

In most mGWAS, association models between genotype and metabolotype assume additive effects, which is intuitively supported by the observation that many GIMs are related to rate limiting steps in enzymatic reactions or transport processes, which suggests a dose-dependent response to genetic variability. Tsepilov et al. (45) investigated systematically non-additive effects on a large panel of serum metabolites and all possible ratios (n = 22 801) in a population-based study (n = 1785). They found that most genetic effects on metabolite concentrations and ratios were indeed additive, with a few notable exceptions that may allow to understand the genetic control of these loci more deeply. However, with larger sample sizes, rare variants with potentially larger effect sizes and recessive modes of inheritance may be detected.

**Epistasis and Mendelian randomization**

Due to their large number and strong effect sizes, GIMs provide a test bed to identify and investigate more complex genetic and metabolic interactions. Shin et al. (23) provided an example of genetic interaction between two gene variants (epistasis), rs10469966 (NAT8) and rs4488133 (PYROXD2), and blood metabolite levels. The same study also reported an example of a Mendelian randomization analysis to establish causation, where expression of THEM4 was shown to mediate the association between rs6693388 and the ratio of linoleate (18:2n6) to 5,8-tetradecadien. However, it seems that as for now there were only few cases that display association signals that were strong enough to single out statistically significant cases of epistasis and Mendelian randomization.

**Multi-phenotype mGWAS**

Strategies that combine several phenotypes in an mGWAS may potentially detect additional genetic loci and further their functional characterization. Inouye et al. (15) showed that association testing of multiple correlated phenotypes offers better power.
than univariate analysis of single traits. Ried et al. (24) applied phenotype set enrichment analysis (PSEA), a method that tests sets of metabolites for association enrichment at genetic loci. In addition to confirming previously reported GIMs, the authors identified and validated 12 new loci.

**Augmenting GIMs with functional information**

A crucial step in the interpretation of mGWAS results is to put the identified associations into the context of results from other association studies, including associations of genetic variants to traits at different phenotype layers. Most recently published mGWAS studies, therefore, go beyond the mere reporting of genetic associations by combining their results with additional -omics datasets (i.e. eQTLs) (23,25), by linking individual associations in a systems level approach (16,46), and by adding clinical association data to establish complex gene-to-disease networks (for an example, see Fig. 5). However, collecting and integrating publicly available association data still present a major bottleneck in the evaluation of mGWAS results, in large part due to the fact that data from different sources are reported on different, frequently non-overlapping scales. For example, Lin et al. (26) recently established the SNiPA web service (http://www.snipa.org). SNiPA contains LD information as well as functional annotations for almost all genetic variants of the latest 1000 Genomes Project dataset. The functional annotations include regulatory elements, eQTL associations to clinical traits and also associations to metabolic traits as provided by the NHGRI-EBI GWAS catalog (http://www.ebi.ac.uk/gwas). SNiPA provides several user-friendly, interactive tools that are especially useful in the context of mGWAS, such as regional association plots and LD plots that are enriched with functional annotations. In particular, links to the Metabolomics GWAS server (http://www.gwas.eu) allow direct access to association data from the Suhe et al. study (10) and Shin et al. study (23) mGWAS, including many unpublished associations below the conservative level of genome-wide significance, that may prove to represent true positive associations when combined with additional evidence from user-provided studies (49).

**Application of GIMS for Hypothesis Generation**

mGWAS results are more and more used as starting hypothesis for deeper functional research. The great potential of mGWAS as hypothesis generating tool is demonstrated when mGWAS ‘rediscover’—sometimes decades-old—findings from biochemical experiments in the setting of modern genomic and metabolomic studies (50). One interesting example is SNP rs37369, which is a non-synonymous variant (Val140Ile) in the coding region of alanine-glyoxylate aminotransferase 2 (AGXT2). Several mGWAS linked this variant to changes in homeostasis of plasma beta-aminoisobutyrate (BAIB) (20), serum symmetric/asymmetric di-methylarginine (39) and serum homoarginine (40). Furthermore, SNP rs37369 was associated with elevated urinary excretion of BAIB (9,11,21) with BAIB concentrations >10 times higher in urine of homozygotes for the minor SNP allele. Since BAIB is one of AGXT2’s substrates (51), it was hypothesized that rs37369 is causative for hyper-beta-aminoisobutyric aciduria (9,11), a inheritable trait first described in the early 1950s (52,53). This mGWAS-generated hypothesis was recently validated by Kittel et al. (54) by in vitro studies where the authors demonstrated that the rs37369 polymorphism results in a significantly lower AGXT2 enzyme activity when compared with the wild type. These studies, inspired by a single mGWAS result, rejuvenated research interest in AGXT2, indicating that altered AGXT2 activity may contribute to the pathogenesis of cardiovascular, renal, neurological and hematological diseases, and highlighted the unique role of AGXT2 at the intersection of key mitochondrial pathways, and even as a potential drug target (50). Inspired by an observed co-association of BAIB and certain triglyceride levels at the AGXT2 locus in their mGWAS, Rhee et al. (20) further set out to investigate the relationship between BAIB metabolism and lipid homeostasis. Using a morpholino knockdown of agxt2 in zebrafish, the authors could indeed establish a functional link between BAIB, cholesterol ester and TAG metabolism, concluding that their data ‘provides an example how the breadth of gene, metabolite, and phenotype data [. . .] can provide a springboard for research in metabolism’.

Another example where mGWAS data could have helped to formulate a hypothesis on a solute transporter’s specificity has just been published by Nguyen et al. (55). Among the 12 novel loci reported by Ried et al. (24) was MFSD2A, a member of the major facilitator superfamily. The mouse homologue (Mfsd2a) was identified by Nguyen et al. (55) as a transporter for the essential omega-3 fatty acid docosahexaenoic acid. The authors showed that this previously orphan transporter transports lysolipids, writing: ‘Unexpectedly, cell-based studies indicate that Mfsd2a transports DHA in the form of lysophosphatidyldicholine (LPC), but not unesterified fatty acid, in a sodium-dependent manner. Notably, Mfsd2a transports common plasma LPCs carrying long-chain fatty acids such LPC oleate and LPC palmitate, but not LPCs with less than a 14-carbon acyl chain’. This finding was reflected in Ried et al.’s study data, which associated this locus with acyl-bound LPCs with side chain lengths C16:0, C17:0, C18:1, C20:4 and C18:0. This finding was further replicated by Draisma et al. (26), who found an association of SNP rs75259794 with these same LPCs, the top association at a P-value of 2.8 × 10\(^{-13}\). Thus, had this information been available to the Nguyen et al., they might actually have expected that Mfsd2a transports LPCs and have designed their experiments accordingly.

Pharmaceutical companies are exploring innovative ways of drug development from candidate selection to clinical proof of concept (56). One compelling application of GWAS for the validation of therapeutic targets through human genetics has been suggested by Plenge et al. (57). The authors describe the concept of using dose-response curves based on multiple functional variants in genes of pharmaceutical interest in order to prioritize molecular targets in drug development. However, for such an approach to be applicable, the target gene is required to harbor a causal variant that is unequivocally associated with a medical trait of interest, the biological function of the causal gene and causal variant needs to be known, and the gene has to harbor multiple causal variants of known biological function, thereby enabling the generation of genotype–phenotype dose–response curves. These criteria are unfortunately rarely met by genetic associations with clinical end points, especially in the case of complex disorders. However, when a GIM is overlapping such a locus, then, due to its generally much stronger effect size and its more direct link to the genetic variant, this GIM can provide intermediate and functionally relevant readouts (5) that allow for the identification of multiple causal variants in order to satisfy the criteria laid out by Plenge et al. (57).

**Conclusion and Outlook**

With the era of GWAS with single clinical end points reaching maturation in meta-analyses that basically include all cohorts available worldwide for any given disease, what will the future bring
for GWAS with metabolic traits? Meta-analyses of mGWAS are certainly the next big step. However, merging data from different metabolomics platforms may require additional efforts in developing statistical methods that allow to combine phenotype information that match only partially (e.g. some platforms can differentiate between isoleucine and leucine, others only report their sum; some platforms resolve lipid traits to a high degree of detail, while other methods provide access to aggregated traits; MS or NMR peaks may be machine specific and hard to match between studies). Larger cohorts shall be phenotyped by multiple platforms in samples from the same donors. Moreover, samples from multiple biofluids are being collected in current population studies, allowing to study metabolic correlations between them, including possible geno-
type-dependent interactions with the human gut or saliva microbiome. A number of GWAS with other disease-relevant -omics phenotypes have also been conducted, including genome-wide gene expression (58), proteomics (59) and protein glycosylation (60), and further epigenome-wide association studies between DNA methylation and metabolomics (61) and studies on transcriptome–metabolome associations (62). We thus expect to see an exciting new series of GWAS in deeply omics-phenotyped cohorts, where the combination of multiple disease-relevant quantitative traits may allow the identification of functionally relevant genotype–pathway associations. In order to leverage the full potential of the multi-omics GWAS, new computational approaches for systems biology and network-based association tests need to be developed. We hope that the resulting growing number of GIMs will lead to a significant increase in our functional understanding of genetic variation in human metabolism and its pathologies, knowing that every single newly discovered GIM holds to potential to reveal an exciting and potentially health-relevant story.

Conflict of Interest statement. None declared.

Funding

K.S. is supported by ‘Biomedical Research Program’ funds at Weill Cornell Medical College in Qatar, a program funded by the Qatar Foundation. The statements made herein are solely the responsibility of the author. C.G. is supported by a grant from the Helmholtz Association (Helmholtz-Russia Joint Research Group 310), the European Union’s Seventh Framework Programme (FP7-Health-FS-2012) under Grant Agreement No. 305280 (MiMOmics) and BMBF eMed project: eAtheroSysMed—Systems medicine of myocardial infarction and stroke. Funding to pay the Open Access publication charges for this article was provided by Weill Cornell Medical College in Qatar.

References

1. Nicholson, J.K. and Lindon, J.C. (2008) Systems biology: metabonomics. Nature, 455, 1054–1056.
2. Suhre, K. and Gieger, C. (2012) Genetic variation in metabolic phenotypes: study designs and applications. Nat. Rev. Genet., 13, 759–769.
3. Gieger, C., Geistlinger, L., Altmaier, E., Hrabe de Angelis, M., Kronenberg, F., Meitingen, T., Mewes, H.W., Wichmann, H.E., Weinberger, K.M., Adamski, J. et al. (2008) Genetics meets metabolomics: a genome-wide association study of metabolite profiles in human serum. PLoS Genet., 4, e1000282.
4. Petersen, A.K., Krumsieck, J., Wagele, B., Theis, F.J., Wichmann, H.E., Gieger, C. and Suhre, K. (2012) On the hypothesis-free testing of metabolite ratios in genome-wide and metabolome-wide association studies. BMC Bioinformatics, 13, 120.
5. Kronenberg, F. (2012) Metabolic traits as intermediate phenotypes. In Suhre, K. (ed) Genetics Meets Metabolomics. Springer, New York, pp. 255–264.
6. Ried, J.S., Baurecht, H., Stuckler, F., Krumsieck, J., Gieger, C., Heinrich, J., Kabesch, M., Prehn, C., Peters, A., Rodriguez, E. et al. (2013) Integrative genetic and metabolite profiling analysis suggests altered phosphatidylcholine metabolism in asthma. Allergy, 68, 629–636.
7. Hicks, A.A., Pramstaller, P.P., Johansson, A., Vitart, V., Rudan, I., Ugocsai, P., Aulchenko, Y., Franklin, C.S., Liebisch, G., Erdmann, J. et al. (2009) Genetic determinants of circulating sphingolipid concentrations in European populations. PLoS Genet., 5, e1000672.
8. Illig, T., Gieger, C., Zhai, G., Romisch-Margl, W., Wang-Sattler, R., Prehn, C., Altmaier, E., Kastenmüller, G., Kato, B.S., Mewes, H.W. et al. (2010) A genome-wide perspective of genetic vari-
ation in human metabolism. Nat. Genet., 42, 137–141.
9. Suhre, K., Wallaschofski, H., Raf J. Kl., Friedlich, N., Haring, R., Michael, K., Wasner, C., Krebs, A., Kronenberg, F., Chang, D. et al. (2011) A genome-wide association study of metabolic traits in human urine. Nat. Genet., 43, 565–569.
10. Suhre, K., Shin, S.-Y., Petersen, A.K., Mohney, R.P., Meredith, D., Wägele, B., Altmaier, E., CARDioGRAM, Deloukas, P., Erdmann, J. et al. (2011) Human metabolic individuality in biomedical and pharmaceutical research. Nature, 477, 54–60.
11. Nicholson, G., Rantalaainen, M., Li, J.V., Maher, A.D., Malmodin, D., Ahmadi, K.R., Faber, J.H., Barrett, A., Min, J.L., Rayner, N.W. et al. (2011) A genome-wide metabolic QTL analysis in Europeans implicates two loci shaped by recent positive selection. PLoS Genet., 7, e1002270.
12. Kettunen, J., Tukiainen, T., Sarin, A.P., Ortega-Alonso, A., Tikkanen, E., Lyytikainen, L.P., Kangas, A.J., Soininen, P., Wurtz, P., Silander, K. et al. (2012) Genome-wide association study identifies multiple loci influencing human serum metabolite levels. Nat. Genet., 44, 269–276.
13. Demirkiran, A., van Duijn, C.M., Ugocsai, P., Isaacs, A., Pramstaller, P.P., Liebisch, G., Wilson, J.F., Johansson, A., Rudan, I., Aulchenko, Y.S. et al. (2012) Genome-wide association study identifies novel loci associated with circulating phospho-
and sphingolipid concentrations. PLoS Genet., 8, e1002490.
14. Tukiainen, T., Kettunen, J., Kangas, A.J., Lyytikainen, L.P., Soi-
ninen, P., Sarin, A.P., Tikkanen, E., O’Reilly, P.F., Savolainen, M.J., Kaski, K. et al. (2012) Detailed metabolic and genetic characterization reveals new associations for 30 known lipid loci. Hum. Mol. Genet., 21, 1444–1455.
15. Inouye, M., Ripatti, S., Kettunen, J., Lyytikainen, L.P., Oksala, N., Laurila, P.P., Kangas, A.J., Soininen, P., Savolainen, M.J., Viikari, J. et al. (2012) Novel loci for metabolic networks and multi-tissue expression studies reveal genes for atheroscle-
rosis. PLoS Genet., 8, e1002907.
16. Krumsieck, J., Suhre, K., Evans, A.M., Mitchell, M.W., Mohney, R.P., Milburn, M.V., Wagele, B., Romisch-Margl, W., Illig, T., Adamski, J. et al. (2012) Mining the unknown: a systems approach to metabolite identification combining genetic and metabolic information. PLoS Genet., 8, e1003005.
17. Montoliu, I., Genick, U., Ledda, M., Collino, S., Martin, F.P., le Coutre, J. and Rezzi, S. (2013) Current status on genome-
metabolome-wide associations: an opportunity in nutrition research. Genes Nutr., 8, 19–27.
18. Raf J. Kl., Romisch-Margl, W., Petersen, A.K., Pagel, P., Blochl, F., Hengstenberg, C., Illig, T., Meisinger, C., Stark, K., Wichmann, H.E. et al. (2013) Identification and MS-assisted inter-
pretation of genetically influenced NMR signals in human plasma. Genome. Med., 5, 13.
26. Draisma, H.H., Pool, R., Kobl, M., Jansen, R., Petersen, A.K.,
28. Ng, E., Salihovic, S., Monica Lind, P., Mahajan, A., Syvanen,
25. Demirkan, A., Henneman, P., Verhoeven, A., Dharuri, H.,
23. Shin, S.-Y., Fauman, E.B., Petersen, A.-K., Krumsiek, J., Santos,
30. Wu, J.H., Lemaitre, R.N., Sebastiani, P., Solovieff, N., Hartley, S.W.,
32. Chen, G., Ramos, E., Adeyemo, A., Shriner, D., Zhou, J., Dou-
27. Donnell, C.J., Barron-Casella, E., Bean, C.J.
31. ilton, J.N., Sebastieni, P., Solovieff, N., Spector, T.D., Prehn, C.,
35. Johnson, A.D., Kavousi, M., Smith, A.V., Chen, M.H., Dehghan,
34. Johnson, A.D., Kavousi, M., Smith, A.V., Chen, M.H., Dehghan,
33. Kang, T.W., Kim, H.J., Ju, H., Kim, J.H., Jeon, Y.J., Lee, H.C., Kim,
30. Wu, J.H., Lemaitre, R.N., Manichaikul, A., Guan, W., Tanaka,
29. Rhee, E.P., Ho,J.E., Chen, M.H., Shen, D., Cheng, S., Larson, M.G.,
21. Ruedi, R., Ledda, M., Nicholls, A.W., Salek, R.M., Marques-
Vidal, P., Morya, E., Sameshima, K., Montoliv, I., Di, Siva, L.,
Collino, S. et al. (2014) Genome-wide association study of metabolic traits reveals novel gene-metabolite-disease links. PLoS Genet., 10, e1004132.
24. Yu, B., Zheng, Y., Alexander, D., Morrison, A.C., Coresh, J. and
31. ilton, J.N., Sebastiani, P., Solovieff, N., Hartley, S.W.,
22. Yu, B., Zheng, Y., Alexander, D., Morrison, A.C., Coresh, J. and
Boerwinkle, E. (2014) Genetic determinants in human serum metabolism. Eur. J. Hum. Genet., 20, 463–468.
23. Kang, T.W., Kim, H.J., Ju, H., Kim, J.H., Jeon, Y.J., Lee, H.C., Kim,
K.K., Kim, J.W., Lee, S., Kim, J.Y. et al. (2010) Genome-wide asso-
ciation of serum bilirubin levels in Korean population. Hum. Mol. Genet., 19, 3672–3678.
25. Demirkan, A., Henneman, P., Verhoeven, A., Dharruri, H.,
Amin, N., van Klinken, J.B., Karssen, L.C., de Vries, B., Meiss-
ner, A., Goraller, S. et al. (2015) Insight in genome-wide associ-
34. Johnson, A.D., Kavousi, M., Smith, A.V., Chen, M.H., Dehghan,
33. Kang, T.W., Kim, H.J., Ju, H., Kim, J.H., Jeon, Y.J., Lee, H.C., Kim,
K.K., Kim, J.W., Lee, S., Kim, J.Y. et al. (2010) Genome-wide asso-
ciation of serum bilirubin levels in Korean population. Hum. Mol. Genet., 19, 3672–3678.
26. Draisma, H.H., Pool, R., Kobl, M., Jansen, R., Petersen, A.K.,
30. Wu, J.H., Lemaitre, R.N., Manichaikul, A., Guan, W., Tanaka,
29. Rhee, E.P., Ho,J.E., Chen, M.H., Shen, D., Cheng, S., Larson, M.G.,
21. Ruedi, R., Ledda, M., Nicholls, A.W., Salek, R.M., Marques-
Vidal, P., Morya, E., Sameshima, K., Montoliv, I., Di, Siva, L.,
Collino, S. et al. (2014) Genome-wide association study of metabolic traits reveals novel gene-metabolite-disease links. PLoS Genet., 10, e1004132.
24. Yu, B., Zheng, Y., Alexander, D., Morrison, A.C., Coresh, J. and
Boerwinkle, E. (2014) Genetic determinants influencing human serum metabolism among African Americans. PLoS Genet., 10, e1004212.
23. Shin, S.-Y., Fauman, E.B., Petersen, A.-K., Krumsiek, J., Santos,
R., Huang, J., Arnold, M., Erte, I., Forgetta, V., Yang, T.-P. et al. (2014) An atlas of genetic influences on human blood metabolites. Nat. Genet., 46, 543–550.
21. Ruedi, R., Ledda, M., Nicholls, A.W., Salek, R.M., Marques-
Vidal, P., Morya, E., Sameshima, K., Montoliv, I., Di, Siva, L.,
Collino, S. et al. (2014) Genome-wide association study of metabolic traits reveals novel gene-metabolite-disease links. PLoS Genet., 10, e1004132.
24. Yu, B., Zheng, Y., Alexander, D., Morrison, A.C., Coresh, J. and
Boerwinkle, E. (2014) Genetic determinants influencing human serum metabolism among African Americans. PLoS Genet., 10, e1004212.
23. Shin, S.-Y., Fauman, E.B., Petersen, A.-K., Krumsiek, J., Santos,
R., Huang, J., Arnold, M., Erte, I., Forgetta, V., Yang, T.-P. et al. (2014) An atlas of genetic influences on human blood metabolites. Nat. Genet., 46, 543–550.
46. Krumsiek, J., Suhre, K., Illig, T., Adamski, J. and Theis, F.J. (2011) Gaussian graphical modeling reconstructs pathway reactions from high-throughput metabolomics data. BMC Syst. Biol., 5, 21.

47. Sainz, I.M., Pixley, R.A. and Colman, R.W. (2007) Fifty years of research on the plasma kallikrein-kinin system: from protein structure and function to cell biology and in-vivo pathophysiology. Thromb. Haemost., 98, 77–83.

48. Arnold, M., Rafler, J., Pfeuffer, A., Suhre, K. and Kastenmüller, G. (2015) SNiPA: an interactive, genetic variant-centered annotation browser. Bioinformatics, 31, 1334–1336.

49. Ho, Y.Y., Baechler, E.C., Ortmann, W., Behrens, T.W., Graham, R.R., Bhangale, T.R. and Pan, W. (2014) Using gene expression to improve the power of genome-wide association analysis. Hum. Hered., 78, 94–103.

50. Rodionov, R.N., Jarzebska, N., Weiss, N. and Lentz, S.R. (2014) AGXT2: a promiscuous aminotransferase. Trends Pharmacol. Sci., 35, 575–582.

51. Kontani, Y., Kaneko, M., Kikugawa, M., Fujimoto, S. and Tamaki, N. (1993) Identity of \( \beta \)-3-aminoisobutyrate-pyruvate aminotransferase with alanine-glyoxylate aminotransferase 2. Biochim. Biophys. Acta., 1156, 161–166.

52. Harris, H. (1953) Family studies on the urinary excretion of beta-aminoisobutyric acid. Ann. Eugen., 18, 43–49.

53. Crumpler, H.R., Dent, C.E., Harris, H. and Westall, R.G. (1951) beta-Aminoisobutyric acid (alpha-methyl-beta-alanine); a new amino-acid obtained from human urine. Nature, 167, 307–308.

54. Kittel, A., Muller, F., Konig, J., Mieth, M., Sticht, H., Zolk, O., Kralj, A., Heinrich, M.R., Fromm, M.F. and Maas, R. (2014) Alanine-glyoxylate aminotransferase 2 (AGXT2) polymorphisms have considerable impact on methylarginine and beta-aminoisobutyrate metabolism in healthy volunteers. PLoS One, 9, e88544.

55. Nguyen, L.N., Ma, D., Shui, G., Wong, P., Cazenave-Gassiot, A., Zhang, X., Wenk, M.R., Goh, E.L. and Silver, D.L. (2014) Mfsd2a is a transporter for the essential omega-3 fatty acid docosahexaenoic acid. Nature, 509, 503–506.

56. Owens, P.K., Raddad, E., Miller, J.W., Stille, J.R., Olovich, K.G., Smith, N.V., Jones, R.S. and Scherer, J.C. (2015) A decade of innovation in pharmaceutical R&D: the Chorus model. Nat. Rev. Drug Discov., 14, 17–28.

57. Plenge, R.M., Scollnick, E.M. and Altshuler, D. (2013) Validating therapeutic targets through human genetics. Nat. Rev. Drug Discov., 12, 581–594.

58. Westra, H.J., Peters, M.J., Esko, T., Yaghoobkar, H., Schurmann, C., Kettunen, J., Christiansen, M.W., Fairfax, B.P., Schramm, K., Powell, J.E. et al. (2013) Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat. Genet., 45, 1238–1243.

59. Meizer, D., Perry, J.R., Hernandez, D., Corsi, A.M., Stevens, K., Rafferty, I., Laurentani, F., Murray, A., Gibbs, J.R., Paolisso, G. et al. (2008) A genome-wide association study identifies protein quantitative trait loci (pQTLs). PLoS Genet., 4, e1000072.

60. Lauc, G., Essafi, A., Huffman, J.E., Hayward, C., Knezevic, A., Kattla, J.J., Polasek, O., Gornik, O., Vitart, V., Abrahams, J.L. et al. (2010) Genomics meets glycomics—the first GWAS study of human N-glycome identifies HNF1alpha as a master regulator of plasma protein fucosylation. PLoS Genet., 6, e1001256.

61. Petersen, A.K., Zeilinger, S., Kastenmüller, G., Romisch-Margl, W., Brugger, M., Peters, A., Meisinger, C., Strauch, K., Hengstenberg, C., Pagel, P. et al. (2014) Epigenetics meets metabolomics: an epigenome-wide association study with blood serum metabolic traits. Hum. Mol. Genet., 23, 534–545.

62. Bartel, J., Krumsiek, J., Schramm, K., Adamski, J., Gieger, C., Herder, C., Carstensen, M., Peters, A., Rathmann, W., Roden, M. et al. (2015) The human blood metabolome-transcriptome interface. PLoS Genet., 11, e1005274.