

REVIEW

# Genetic determinants of metabolism in health and disease: from biochemical genetics to genome-wide associations

Steven L Robinette<sup>1,2</sup>, Elaine Holmes<sup>1</sup>, Jeremy K Nicholson<sup>1,\*</sup> and Marc E Dumas<sup>1,\*</sup>

## Abstract

Increasingly sophisticated measurement technologies have allowed the fields of metabolomics and genomics to identify, in parallel, risk factors of disease; predict drug metabolism; and study metabolic and genetic diversity in large human populations. Yet the complementarity of these fields and the utility of studying genes and metabolites together is belied by the frequent separate, parallel applications of genomic and metabolomic analysis. Early attempts at identifying co-variation and interaction between genetic variants and downstream metabolic changes, including metabolic profiling of human Mendelian diseases and quantitative trait locus mapping of individual metabolite concentrations, have recently been extended by new experimental designs that search for a large number of gene-metabolite associations. These approaches, including metabolomic quantitative trait locus mapping and metabolomic genome-wide association studies, involve the concurrent collection of both genomic and metabolomic data and a subsequent search for statistical associations between genetic polymorphisms and metabolite concentrations across a broad range of genes and metabolites. These new data-fusion techniques will have important consequences in functional genomics, microbial metagenomics and disease modeling, the early results and implications of which are reviewed.

**Keywords** Metabonomics/metabolomics, Quantitative Trait Locus Mapping, Biochemical Genetics, NMR, MS

## Introduction

The last few decades have witnessed a radical change in the biological sciences, as the advent of the 'omics' era has brought large-scale measurement of genes [1,2], transcripts [3], proteins [4,5] and metabolites [6-8]. Genetics has undergone a high-throughput revolution, with genomic sciences enabling the rapid acquisition of genome-wide gene-expression profiles, polymorphisms and, more recently, whole genome sequences [9]. These advances have been matched by advances in the measurement of small-molecule metabolites in the associated fields of metabonomics [10,11] and metabolomics [8,12]. Like genomics, these fields aim for comprehensive measurement and analysis of variations, but in the complement of low-molecular-weight compounds within a cell, tissue or biofluid.

For the past 20 years, developments in genomics and metabonomics have progressed on parallel tracks, exchanging experimental designs, data-analysis techniques and applications in basic biology and medicine [13-16]. Yet genes and metabolites are intrinsically co-informational, each shedding light on complementary biological processes. The genome encodes the metabolic capacities of the cell (the microbiome also influences mammalian metabolism), and changes in the activity and function of enzymes, transporters and transcription factors resulting from genetic variations have a direct impact on the identities and quantities of both intracellular and extracellular metabolites. Metabolite concentrations are ultimately quantitative, phenotypic traits, the genetics of which are described by the quantitative trait locus (QTL) - a DNA sequence controlling the phenotypic outcome of the quantitative trait, such as a metabolite concentration. Since the origins of biochemical genetics over a century ago, the integrated study of genetics and metabolism has produced significant advances in the understanding of basic biological processes and in the diagnosis and treatment of human disease [17].

Metabolic profiling of single gene mutations [18] and QTL mapping of single metabolic traits [19] both

\*Correspondence: j.nicholson@imperial.ac.uk; m.dumas@imperial.ac.uk

<sup>1</sup>Biomolecular Medicine, Department of Surgery and Cancer, Faculty of Medicine, Imperial College London, Sir Alexander Fleming Building, Exhibition Road, South Kensington, London SW7 2AZ, UK

Full list of author information is available at the end of the article

represent early attempts at identifying gene-metabolite associations through omic sciences, by regressing one gene against many metabolites or one metabolite against many genes. In more recent studies, metabolome-wide profiling of biofluids by nuclear magnetic resonance (NMR) spectroscopy or mass spectrometry (MS) is combined with whole-genome profiling of single nucleotide polymorphisms (SNPs) to identify many gene-metabolite associations simultaneously, by regressing many metabolite levels against many polymorphisms [20]. While these studies, termed metabolomic quantitative trait locus (mQTL) mapping, or metabolomic genome-wide association studies (mGWAS), are often applied to human genes and human biofluid profiles, these techniques hold the additional promise of investigating interactions between gut microflora genes and biofluid metabolites. By regressing metagenomic sequences against metabolic profiles, metagenome-wide metabolome-wide associations provide insight into the metabolic cross-talk of bacterial and human genomes in the larger human superorganism [21,22]. This review discusses the highly analogous methods and applications of genomics and metabolomics, as well as recent attempts at integrating the two fields towards a more comprehensive and holistic understanding of gene function and the control of metabolic processes.

### **Developments in instrumentation and experimental design in metabolomics parallel those in genomics**

Increases in the speed, accuracy and coverage of genomic analysis have been mirrored by technological developments in the large-scale measurement of low-molecular-weight metabolites, using the two major analytical platforms of NMR spectroscopy and MS [23]. While these techniques feature varying strengths in coverage, sensitivity, selectivity for various chemical classes, reproducibility, provision of structural information, and sample-preparation requirements, both stand out in their capacity to measure a large number of small-molecule analytes in an untargeted fashion from complex biological mixtures, such as human biofluids [24].

One of the most popular analytical chemistry techniques, NMR spectroscopy has a long history of application in organic chemistry for structural identification and is used extensively in metabolomics. NMR spectroscopy is characterized by the following key properties, which are fit for purpose: (i) high dynamic range, with several biological nuclei, such as  $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$  or  $^{31}\text{P}$ , being accurately measured over a large range of concentrations; (ii) high linearity of signal intensity with concentration; and (iii) high reproducibility. In particular,  $^1\text{H}$  NMR spectroscopy is robust, provides a high degree of structural information for both one-dimensional  $^1\text{H}$  and

two-dimensional  $^1\text{H}$ - $^1\text{H}$  NMR, and is flexibly applied to extracts, biofluids and solid tissues using high-resolution magic angle spinning, and *in vivo* using magnetic resonance spectroscopy. Technological advancements in magnetic field strength with the introduction of 600, 800 and, recently, 1,000 MHz NMR spectrometers, pulse sequence experiments, and cryogenically cooled probes have increased the sensitivity and coverage for small-molecular-weight metabolites and lipid components from urine, plasma, serum and tissue samples [11].

MS, coupled to either liquid (LC) or gas (GC) chromatography, is also frequently applied to profile the metabolome [25-27]. LC- and GC-MS both boast high sensitivity to low-concentration analytes, as well as high-resolution chromatographic separation. While these techniques have faced challenges in compound identification, reproducibility and bias towards certain compound functional groups, the rapid pace of technological development in MS has sought to address many of these challenges. Advances in chromatography, including ultra-high-performance liquid chromatography (UPLC) and multidimensional gas chromatography (GC $\times$ GC, 3D-GC), have increased the speed and reproducibility of chromatography-coupled MS [28]. Additionally, advances in MS, including high-resolution time-of-flight (ToF) and quadrupole time-of-flight (qTOF) instruments, along with serial ion fragmentation (MS-MS, MS $^n$ ) have improved resolution, coverage and identification of low-molecular-weight species [29].

While many comparisons have been made between the utility of these techniques for metabolite analysis [30], advances in both have paralleled advances in speed, accuracy and coverage in genomics. As a result, broad-coverage metabolite profiling using NMR spectroscopy and MS has been increasingly used to answer the same questions as genomics, especially in the search for risk factors for disease at the population level and predictors of drug metabolism in personalized medicine [13-16]. These are mirrored in parallel developments in experimental design and data analysis.

Just as the introduction of the genome-wide association study (GWAS) began the search for associations between genome-wide polymorphisms with disease phenotypes in large population cohorts [15,31], the metabolome-wide association study (MWAS) has used large numbers of biofluid spectra and statistical regression to search for associations between metabolites present in human biofluids and both quantitative and binary physiological and pathological traits [16]. Two-class experimental designs, in which metabolite associations with disease are identified by statistical regression of metabolic profiles against binary variables for affected and control individuals, are common, and associations between metabolites and disease have been reported for obesity

and insulin resistance [32], prostate cancer [33], autism [34], ulcerative colitis [35] and more [36]. However, the recent application of metabolic profiling to large population cohorts, quantitative traits and population differences (termed MWAS) has revealed metabolite associations with diet and blood pressure [37], region and cardiovascular risk [38] and ethnicity [39].

In addition to studies of disease risk, both metabolites and genes have been queried to predict drug metabolism. Paralleling previous work in pharmacogenomics, the introduction of pharmaco-metabonomics has demonstrated that drug metabolism can be predicted from the metabolite composition of urinary biofluids before drug administration [13,40]. Recent applications of pharmaco-metabonomics have highlighted metabolic predictors of acetaminophen toxicity in animals [13] and humans [41,42], capecitabine toxicity [43] and microbial influences on drug detoxification [41].

### **Early integration: metabolic profiling of Mendelian traits and QTL mapping of single metabolites**

Since the discovery of alkaptonuria by Archibald Garrod in the early 20th century, the measurement of metabolites has been used as a proxy to identify human genetic diseases, especially inborn errors of metabolism. Uniform newborn screening for multiple inborn errors of metabolism, including urea-cycle disorders and amino- and organic-acidurias, using heel-prick testing with GC-MS-MS exemplifies the power of targeted metabolomic analysis to diagnose these diseases and enable the interruption of pathological processes resulting from genetic mutations [44]. More recently, untargeted NMR spectroscopy and MS have been used to diagnose known inborn errors [45-48], identify novel inborn errors using biofluid profiling [49,50] and identify complex downstream metabolic consequences [51-53] and biomarkers of organ pathology resulting from genetic mutations [54-56].

Untargeted metabolic profiling of biofluids, especially urine and serum, is a powerful technique for diagnosing inborn errors of metabolism with often non-specific clinical presentation, as metabolic intermediates accumulated in biofluid compartments can be easily identified [18,57,58]. As a result, diagnosis of suspected inborn errors has been reported for many Mendelian diseases [45-50], especially using NMR spectroscopy. In some cases, metabolic profiling of biofluids from patients with suspected inborn errors has led to the discovery of previously undescribed diseases, with the identification of causal genes following the description of metabolic perturbations, as occurred with aminoacylase 1 deficiency and beta-ureidopropionase deficiency [49,50].

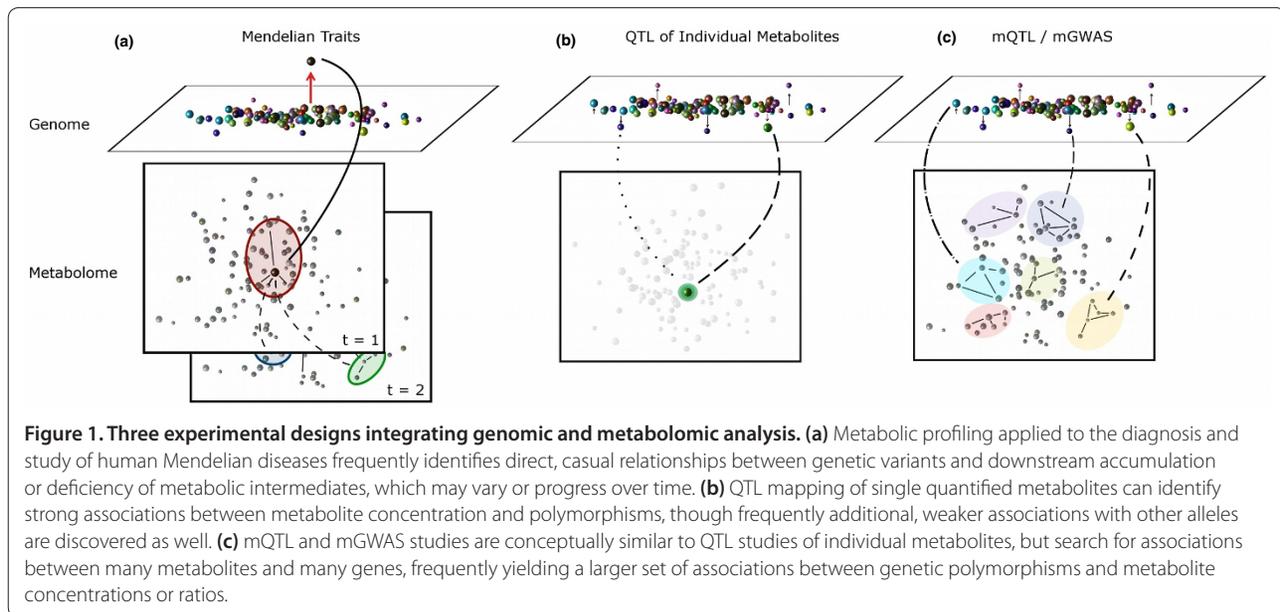
While mutations in enzymes and transporters can often be readily diagnosed by biofluid profiling, and a strong mechanistic link is easily inferred between the

disruption of a metabolic pathway and resulting accumulation or depletion of metabolic intermediates, many Mendelian diseases result in more complex, progressive organ-specific or multi-organ pathology [59]. In these cases, metabolic profiling has been applied to identify sites of lesions, describe progression and attempt to identify proxy small-molecular biomarkers of the disease. Examples of this include the identification of markers of autosomal dominant polycystic kidney disease [55], comparison of the urinary profiles of several genetic forms of renal Fanconi syndrome [56] and description of abnormal brain metabolism in Smith-Lemli-Optiz syndrome [54]. Additionally, metabolite flux analysis using isotopically labeled metabolites suggests an additional way to apply metabolic profiling techniques to study the impact of genetic mutations [59].

The genomic corollary of using metabolic profiling to study Mendelian genes is QTL mapping of single metabolic traits. The advent of whole-genome SNP analysis and the use of QTL mapping for quantitative traits, such as height [60], led to interest in identifying genetic loci associated with quantitative variation in individual metabolite levels [19]. Examples of this include mapping of serum leptin levels to genes on human chromosome 2 in multiple human populations [61,62], associations with plasma triglyceride levels [63,64] and identification of genetic variants associated with high-density lipoprotein (HDL) levels [65,66]. Recent studies have investigated associations between serum lipid fractions and polymorphisms, constituting an intermediate between traditional metabolite QTL mapping and mQTL/mGWAS [67]. Like many QTL mapping studies, attempts to identify loci significantly associated with biofluid levels of single metabolites frequently indicate a large number of genetic associations. This is almost certainly an indication of complex, multigenic control processes regulating energy metabolism and homeostasis, and the identification of large numbers of multiple loci provides an important documentation of genes involved in complex metabolic pathways. However, a large number of loci each contributing to a potentially small percentage of observed variance in metabolite levels complicates direct interpretation of genotype-phenotype relationships in these cases. Figure 1 shows a schematic illustration of gene-metabolite correlations in biochemical genetics, traditional QTL mapping, and mQTL/mGWAS.

### **Identifying the genetic determinants of the metabolome: mQTL and mGWAS**

GWAS currently requires increasingly large cohorts to ensure discovery of new genes associated with disease phenotypes [68]. Although this approach is very efficient, the biological relevance of these associations can be difficult to assess. The identification of phenotypes



related to disease mechanism, onset and progression represents a promising research avenue.

The systematic search for molecular endophenotypes (that is, internal phenotypes) that can be mapped onto the genome began with the quantitative genetic analysis of gene-expression profiles, referred to as genetical genomics [69] or expression QTL (eQTL) mapping [70]. Treating genome-wide gene-expression profiles as quantitative traits was originally developed in model organisms and applied to humans [70,71]. In eQTL mapping, *cis*-regulatory associations between genomic variations and gene-expression levels are discovered by integrated analysis of quantitative gene-expression profiles and SNPs. The identification of a SNP at a gene locus affecting its own expression represents a powerful self-validation. However, eQTL mapping presents a series of drawbacks: (i) frequently analyzed cell lines often have altered gene expression, and access to biopsy samples from organs directly relevant to pathology is often impossible; and (ii) due to the gene-centric nature of eQTL mapping, this approach bypasses the biological consequences of the endophenotypes generating the association.

Immediately following the success of the eQTL mapping approach [70], in which *cis*-regulatory associations between genomic variations and gene-expression levels are discovered by integrated analysis of quantitative gene-expression profiles and SNPs, metabolic profiles were included as endophenotypic quantitative traits. This led to mapping of multiple quantitative metabolic traits directly onto the genome to identify mQTLs in plants [72,73], then in animal models [74,75]. In mQTL mapping, individuals are genotyped and phenotyped in

parallel and the resulting genome-wide and metabolome-wide profiles are then quantitatively correlated (Box 1). mQTL mapping presents a significant advantage over gene-expression products such as transcripts [70] or proteins [76]: the ever-increasing coverage of the metabolome allows a glimpse at the real molecular endpoints, which are closer to the disease phenotypes of interest. Following the success of mQTL mapping in plants [72,73] and then in mammalian models [75], this approach was quickly followed by the development of mGWAS in humans cohorts ([77-83], see also the review by J Adamski [84]).

One of the distinctive features of mGWAS is the intrinsically parallel identification of associations between monogenetically determined metabolic traits and their causative gene variants (see Table 1 for a list of human mQTL-metabolite associations).

The mechanistic explanation of gene/metabolite associations identified by mQTL mapping can be difficult. The simplest case corresponds to associations between genes encoding enzymes and metabolites, which are either substrates or products of the enzyme they are associated with [74,75] (Figure 2). This corresponds to a direct *cis*-acting mechanism. Also, one of the interesting discoveries from results obtained by Suhre *et al.* is that a number of gene variants causing metabolic variation correspond to solute transporter genes, as the majority of the genes in this category belong to the solute carrier (SLC) family [78,81,82]. Again, this corresponds to a direct mechanistic link. In other cases, the link between gene variants and their associated metabolites can demonstrate pathway, rather than direct, connectivity, such as polymorphisms in enzymes associated with

### Box 1. Mathematical modeling for mQTL identification

The statistical analysis involved in mQTL mapping and mGWAS does not currently differ substantially from the statistical methods used to identify genetic loci associated with single quantitative traits. mQTL and mGWAS involve independent QTL mapping of each metabolite identified by metabolic profiling, though accurate analysis is dependent upon proper preprocessing of both genomic and metabolomic data. Associations are identified using techniques such as Haley-Knott regression implemented in the R/QTL package, which uses local information about surrounding markers [103], or typical univariate association tests such as  $\chi^2$  or Cochran-Armitage trend tests implemented in PLINK [104]. The results of mQTL and association mapping are typically displayed using a logarithm of odds (LOD,  $-\log_{10}(P \text{ value})$ ) score, which allows establishment of genome/metabolome LOD score maps [74,75], or more classical Manhattan plots [77,78,81,82] (Figure 2).

The main challenge in mQTL data modeling is multiple correlation testing. Assuming the use of high-resolution metabolic profiles (1,000 to 10,000 features) and genome-wide SNP coverage (600,000 SNPs), a typical metabolome-wide GWAS can apply between 600,000,000 and 6,000,000,000 univariate tests. Given the number of tests involved, there are numerous opportunities for false discoveries and multiple testing corrections are required to account for this. Genome-wide significance levels can be estimated using Bonferroni correction [77], but also using Benjamini and Hochberg or Benjamini and Yakutieli corrections [105]. Finally, permutation and resampling methods also provide empirical estimates for false discovery thresholds [74,79].

metabolites several reactions downstream of the compound directly acted upon by the enzyme itself (as observed with *NTSE* polymorphisms and inosine).

More opaque associations may be *trans*-acting in a broader sense: the causative gene variant can be a molecular switch, and the metabolite it is associated with is in fact regulated indirectly by this molecular switch (further down in the regulation events). This is particularly the case when the causative gene variant encodes a transcription factor, inducing the medium- to long-term expression of entire gene networks, or when the gene variant encodes a kinase or a phosphatase regulating entire pathways on much shorter time-scales. Unlike *cis*-acting mQTL/metabolite associations, which can be seen as self-validation of the causative gene at the locus, *trans*-acting mQTL associations present the challenge of identification of the most relevant causative gene at the locus. If a SNP is associated with a metabolite, the closest gene at the locus is not necessarily the most relevant candidate, and further investigation of a larger biological network, such as protein-protein interactions [85], may be necessary to identify mechanistic relationships between genetic variants and downstream metabolism. Despite these challenges, which are familiar to

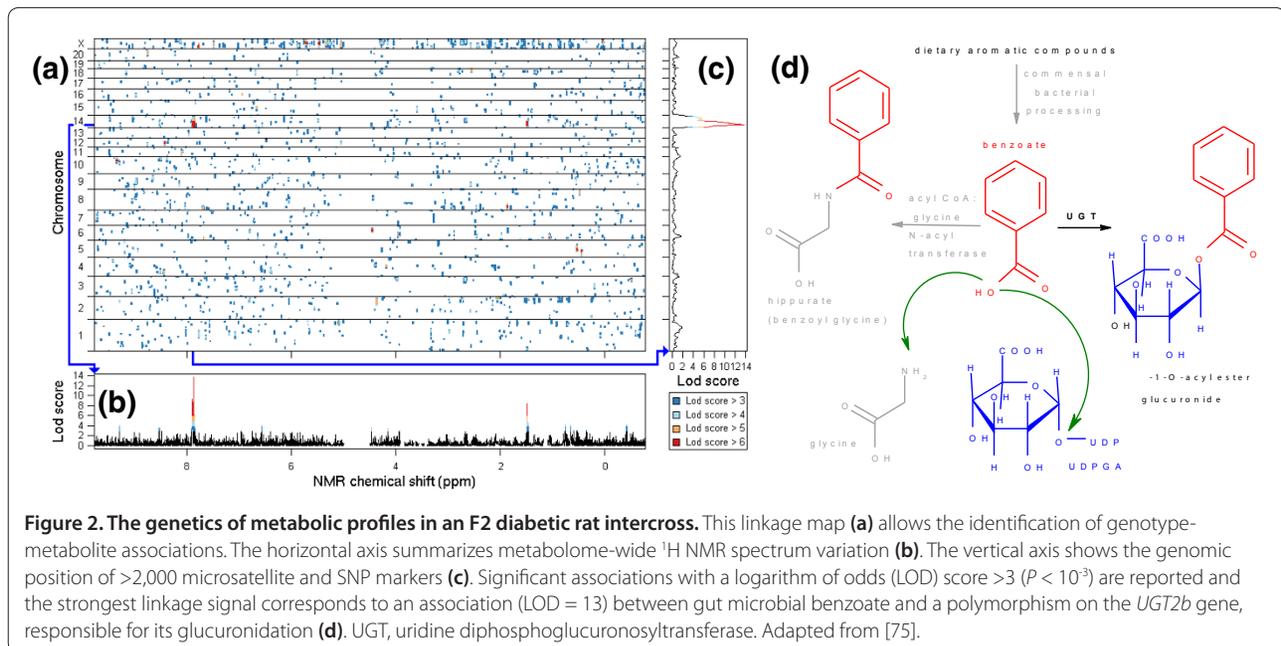
practitioners of biochemical genetics, statistical identification of gene-metabolite associations by mQTL and mGWAS promises to significantly advance current understandings of gene function, metabolic regulation and mechanisms of pathology.

### A glimpse of our extended genome with microbiome-metabolome associations

The functional genomic association studies and data-integration approaches described above rely predominantly on mammalian genome sequences and their annotation (excluding MWAS, which makes use only of metabolite profiling data and does not, as such, require genomic data). However, human phenotypes result from the interaction of several sets of genes: the karyome, the chondriome and the microbiome, respectively corresponding to eukaryotic chromosomes, mitochondrial chromosomes and, finally, gut bacterial prokaryotic chromosomes. The latest human gut microbiome gene catalogue identified 3.3 million non-redundant genes [86], which was dubbed 'our other genome', and the bacterial species composition of the gut microbiome varies from one individual to another, but this variation is stratified, not continuous, and suggests the existence of stable bacterial communities, or 'enterotypes' [87].

The classical identification of associations between gut bacteria and metabolites has been performed on a case-by-case basis for decades. However, the correlation of metabolic profiles with multiple gut bacterial abundance profiles was initiated a few years ago with the introduction of bacteria/metabolite association networks [21]. Semi-quantitative characterizations of microbial populations using denaturing gradient gel electrophoresis (DGGE) and fluorescent *in situ* hybridization (FISH) have yielded associations with obesity and related metabolites [88]. Recently, the introduction of high-throughput sequencing of bacterial 16S rDNA profiles and correlation with metabolic profiles has greatly increased the coverage and quantification of microbial species [89]. The correlation of metabolic profiles with 16S rDNA microbiome profiles provides a strategy for the identification of co-variation between metabolites and bacterial taxa, and such associations point to the production or regulation of metabolic biosynthesis by these microbes.

Given these early successes, the integration of metabolome-wide experimental profiles with metagenome-wide metabolic reconstruction models obtained from full microbiome sequencing should provide a clear insight into the functional role of the gut microbiome, especially the synthesis of metabolites and resultant impacts on human metabolism. This critical need for a marriage between metabolomics/metabonomics and metagenomics has been clearly identified for several years [90]. How



new experimental data change our understanding of our commensal microflora remains to be seen.

### Future directions - the rise of sequencing and consequences for genome-metabolome data fusion

Genomics is currently undergoing yet another revolution, as next-generation sequencing technologies increase the accuracy, coverage and read-length, and drastically decrease the cost of whole-exome sequencing (WES) and whole-genome sequencing (WGS). The introduction of third-generation sequencing technologies in the near future promises to continue this trend [91]. Consequently, the near term promises a dramatic expansion in the availability of sequence data both in the laboratory and in the clinic. The relevance of the explosion of sequence data to the continued integration of metabonomic and genomic data is twofold: first, an opportunity for metabonomics to contribute to the increased clinical presence of omics sciences led by genome sequencing; and second, a challenge to develop methods of integrating metabolic profiles with sequences rather than polymorphisms.

The introduction of WES and WGS into the clinic is already well underway, with success stories including discoveries of new Mendelian disorders [92,93] and successful therapy designed on the basis of mutation discovery [94]. Of known and suspected human Mendelian diseases, molecular bases have been identified for over 3,000, with another approximately 3,700 phenotypes suspected of having a Mendelian basis [95,96]. As sequencing identifies an increasing number of variants with associations to disease, the rate-limiting step in

genomic medicine will move from discovery to functional annotation of sequence variants. Metabolite profiling, along with other high-throughput measurement and data-analysis technologies, may find increasing acceptance in medicine, as investigators rush to keep up with a deluge of sequence data. Increasingly, routine genome sequencing will create a significant resource for large-scale population studies like those currently used to identify gene-metabolite associations, and will critically include rare variants not captured by polymorphism data.

Recent results demonstrate the potential power of integrating WES/WGS with metabolic profiling and mQTL/mGWAS. In 2011, a publication in the *New England Journal of Medicine* reported the discovery by WES of a novel human Mendelian disease in which rare mutations in the *NT5E* gene result in arterial calcifications due to loss of CD73 (encoded by *NT5E*) function, which converts AMP to adenosine in the vasculature [97]. Within the year, an independent metabolome-wide GWAS study published in *Nature* reported a statistically significant association between a SNP near the *NT5E* locus (rs494562) and inosine concentration in human serum, as part of a much larger set of gene-metabolite associations (Table 1) [81]. While in this case the publication of the statistical association followed the description of the human phenotype, future genetic studies will be greatly aided by gene-metabolite co-variation discovered by association studies.

Despite the significant opportunity represented by low-cost sequencing for the integration of genomic and metabonomic data and the identification of gene-metabolite

**Table 1. Human gene-metabolite associations identified by mQTL/mGWAS**

Metabolite	Biofluid	SNP ID	Local gene	P-value	Reference(s)
Trimethylamine	Urine	rs7072216	<i>PYROXD2 (C10orf33)</i>	7.90E-15	[79]
N-acetylated compound(s)	Urine	rs9309473	<i>ALMS1, NAT8, TPRKB, DUSP11</i>	1.40E-11	[79]
3-Amino-isobutyrate	Urine	rs37369	<i>AGXT2</i>	1.1E-06	[79]
				3.17E-75	[82]
2-Hydroxyisobutyrate	Urine	rs830124	<i>WDR66, HPD</i>	1.59E-15	[82]
Dimethylamine	Plasma	rs6584194	<i>PYROXD2 (C10orf33)</i>	8.10E-03	[79]
Sphingomyelin SM C14:10	Serum	rs9309413	<i>PLEK</i>	1.95E-09	[77]
Lysine	Serum	rs992037	<i>PARK2</i>	1.20E-07	[77]
Sphingomyelin SM(OH,COOH) C18:2	Serum	rs1148259 (rs1200826)	<i>ANKRD30A</i>	3.04E-09	[77]
Phosphatidylcholine PC aa C36:4	Serum	rs174548	<i>FADS1</i>	4.52E-08	[77]
Phosphatidylethanolamine PE aa C38:6	Serum	rs4775041	<i>LIPC</i>	9.66E-08	[77]
C0	Serum	rs7094971	<i>SLC16A9</i>	3.80E-20	[78]
N-Acetylmethionine	Serum	rs13391552	<i>NAT8</i>	5.40E-252	[81]
5-Oxoproline	Serum	rs6558295	<i>OPLAH</i>	1.50E-59	[81]
Androsterone sulfate	Serum	rs17277546	<i>CYP3A4</i>	8.70E-40	[81]
Urate	Serum	rs4481233	<i>SLC2A9</i>	5.50E-34	[81]
Glycine	Serum	rs2216405	<i>CPS1</i>	1.60E-27	[81]
		rs7422339	<i>CPS1</i>	2.12E-24	[83]
Succinylcarnitine	Serum	rs2652822	<i>LACTB</i>	7.20E-27	[81]
Isobutyrylcarnitine	Serum	rs662138	<i>SLC22A1</i>	7.30E-25	[81]
Aspartylphenylalanine	Serum	rs4329	<i>ACE</i>	8.20E-20	[81]
Serine	Serum	rs477992	<i>PHGDH</i>	2.60E-14	[81]
Inosine	Serum	rs494562	<i>NT5E</i>	7.40E-13	[81]
Proline	Serum	rs2023634	<i>PRODH</i>	2.00E-22	[81]
α-Hydroxyisovalerate	Serum	rs2403254	<i>HPS5</i>	1.00E-20	[81]
Bradykinin, des-arg(9)	Serum	rs4253252	<i>KLKB1</i>	6.60E-18	[81]
Glutamine	Serum	rs2657879	<i>GLS2</i>	3.10E-17	[81]
Isovalerylcarnitine	Serum	rs272889	<i>SLC22A4</i>	7.40E-16	[81]
Decanoylcarnitine	Serum	rs8396	<i>ETFDH</i>	5.50E-15	[81]
Carnitine	Serum	rs7094971	<i>SLC16A9</i>	3.40E-14	[81]

Shown here are the SNP-metabolite associations with the highest statistical significance, as in [77,79,81-83]. Associations with metabolite concentration were reported for a total of 28 unique SNPs, as shown above. Associations with ratios of multiple metabolites were reported for an additional 30 unique SNPs, but are not included in this table.

associations, several challenges stand in the way of routine paired analysis of sequences and metabolic profiles. The first of these is the challenge of discovering significant associations with low sample numbers. Many of the successes reported in clinical sequencing have made use of data from a small number of patients, and sometimes only a sole patient. In these cases, potentially disease-causative variants are often identified using filtering strategies rather than statistical analysis [98,99]. While the diagnosis of human inborn errors of metabolism from single-patient biofluid NMR spectra demonstrates the potential of metabolic profiling to work with low sample numbers, lack of statistical validation

means that the 'biological signal' in these cases must be quite marked. A second challenge is a dearth of tools for statistical analysis of sequence data. While QTL mapping using SNPs is well established, statistical techniques for QTL mapping with both rare and common variants are just beginning to be introduced [100]. It is likely that increased availability of large-scale population sequence data from initiatives such as the 1000 Genomes Project [101,102] and ClinSeq [103] will spur the development of statistical methods that can be deployed to identify gene-metabolite associations.

Of the omics sciences, genomics and metabolomics are uniquely complementary, the strengths of each addressing

weaknesses of the other. Genes are (mostly) static, an 'upstream' blueprint controlling dynamic biological processes. The identities and quantities of 'downstream' metabolites capture both genetic and environmental influences, and can be measured serially to assess variation through time. Genomic studies often struggle to establish a firm link between genetic variants and phenotypic observations, and while metabonomics provides a closer proxy to phenotype, it is often difficult to infer underlying causality from variations in metabolism. Together, the integrated application of genomics and metabonomics promises a bridging of the gap between genotype and phenotype through intermediate metabolism, to help annotate genes of unknown function, genetic controls of metabolism, and mechanisms of disease.

#### Abbreviations

DGGE, denaturing gradient gel electrophoresis; FISH, fluorescent *in situ* hybridization; GC, gas chromatography; GWAS, genome-wide association study; HDL, high density lipoprotein; LC, liquid chromatography; mGWAS, metabolomic genome-wide association study; mQTL, metabolomic quantitative trait locus; MS, mass spectrometry; MWAS, metabolome-wide association study; NMR, nuclear magnetic resonance; QTL, quantitative trait locus; qToF, quadrupole time-of-flight; SNP, single nucleotide polymorphism; ToF, time-of-flight; UPLC, ultra-performance liquid chromatography; WES, whole exome sequencing; WGS, whole genome sequencing.

#### Acknowledgements

SLR acknowledges support from the NSF Graduate Research Fellowship Program and from the Marshall Aid Commemoration Commission. M-ED is funded by Nestlé (RDLS015375), Agence Nationale de la Recherche (ANR-08-GENO-030-02), and EU-FP7 EURATRANS (HEALTH-F4-2010-241504).

#### Competing interests

The authors declare that they have no competing interests.

#### Author details

<sup>1</sup>Biomolecular Medicine, Department of Surgery and Cancer, Faculty of Medicine, Imperial College London, Sir Alexander Fleming Building, Exhibition Road, South Kensington, London SW7 2AZ, UK. <sup>2</sup>Medical Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA.

Published: 30 April 2012

#### References

- Hood L, Galas D: **The digital code of DNA.** *Nature* 2003, **421**:444-448.
- Collins FS, Green ED, Guttmacher AE, Guyer MS, Institute UNHGR: **A vision for the future of genomics research.** *Nature* 2003, **422**:835-847.
- Wang Z, Gerstein M, Snyder M: **RNA-Seq: a revolutionary tool for transcriptomics.** *Nat Rev Genet* 2009, **10**:57-63.
- Liotta LA, Ferrari M, Petricoin E: **Clinical proteomics: written in blood.** *Nature* 2003, **425**:905.
- Tyers M, Mann M: **From genomics to proteomics.** *Nature* 2003, **422**:193-197.
- Nicholson J, Connelly J, Lindon J, Holmes E: **Metabonomics: a platform for studying drug toxicity and gene function.** *Nat Rev Drug Discov* 2002, **1**:153-161.
- Nicholson JK, Wilson ID: **Understanding 'global' systems biology: metabonomics and the continuum of metabolism.** *Nat Rev Drug Discov* 2003, **2**:668-676.
- Fiehn O: **Metabonomics - the link between genotypes and phenotypes.** *Plant Mol Biol* 2002, **48**:155-171.
- Diamandis EP: **Next-generation sequencing: a new revolution in molecular diagnostics?** *Clin Chem* 2009, **55**:2088-2092.
- Nicholson J, Lindon J, Holmes E: **'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via**

- multivariate statistical analysis in biological NMR spectroscopic data. *Xenobiotica* 1999, **29**:1181-1189.
- Beckonert O, Keun H, Ebbels T, Bundy J, Holmes E, Lindon J, Nicholson J: **Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts.** *Nat Protocols* 2007, **2**:2692-2703.
  - Oliver SG, Winson MK, Kell DB, Baganz F: **Systematic functional analysis of the yeast genome.** *Trends Biotechnol* 1998, **16**:373-378.
  - Andrew Clayton T, Lindon JC, Cloarec O, Antti H, Charuel C, Hanton G, Provost J-P, Le Net J-L, Baker D, Walley RJ, Everett JR, Nicholson JK: **Pharmacometabonomic phenotyping and personalized drug treatment.** *Nature* 2006, **440**:1073-1077.
  - Evans WE, McLeod HL: **Pharmacogenomics - drug disposition, drug targets, and side effects.** *N Engl J Med* 2003, **348**:538-549.
  - McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JPA, Hirschhorn JN: **Genome-wide association studies for complex traits: consensus, uncertainty and challenges.** *Nat Rev Genet* 2008, **9**:356-369.
  - Nicholson JK, Holmes E, Elliott P: **The metabolome-wide association study: a new look at human disease risk factors.** *J Proteome Res* 2008, **7**:3637-3638.
  - Rinaldo P, Hahn S, Matern D: **Clinical biochemical genetics in the twenty-first century.** *Acta Paediatr Suppl* 2004, **93**:22-26; discussion 27.
  - Constantinou M, Papakonstantinou E, Spraul M, Sevastiadou S, Costalos C, Koupparis M, Shulpis K, Tsantili-Kakoulidou A, Mikros E: **H-1 NMR-based metabonomics for the diagnosis of inborn errors of metabolism in urine.** *Anal Chim Acta* 2005, **542**:169-177.
  - Almasy L, Blangero J: **Human QTL linkage mapping.** *Genetica* 2009, **136**:333-340.
  - Ala-Korpela M, Kangas AJ, Inouye M: **Genome-wide association studies and systems biology: together at last.** *Trends Genet* 2011, **27**:493-498.
  - Martin F-PJ, Dumas M-E, Wang Y, Legido-Quigley C, Yap IKS, Tang H, Zirah S, Murphy GM, Cloarec O, Lindon JC, Sprenger N, Fay LB, Kochhar S, van Bladeren P, Holmes E, Nicholson JK: **A top-down systems biology view of microbiome-mammalian metabolic interactions in a mouse model.** *Mol Syst Biol* 2007, **3**:112.
  - Nicholson JK, Holmes E, Wilson ID: **Gut microorganisms, mammalian metabolism and personalized health care.** *Nat Rev Microbiol* 2005, **3**:431-438.
  - Dunn W, Ellis D: **Metabonomics: current analytical platforms and methodologies.** *Trac-Trend Anal Chem* 2005, **24**:285-294.
  - Dunn W, Bailey N, Johnson H: **Measuring the metabolome: current analytical technologies.** *Analyst* 2005, **130**:606-625.
  - Chan ECY, Pasikanti KK, Nicholson JK: **Global urinary metabolic profiling procedures using gas chromatography-mass spectrometry.** *Nat Protoc* 2011, **6**:1483-1499.
  - Dettmer K, Aronov PA, Hammock BD: **Mass spectrometry-based metabolomics.** *Mass Spectrom Rev* 2007, **26**:51-78.
  - Want EJ, Wilson ID, Gika H, Theodoridis G, Plumb RS, Shockcor J, Holmes E, Nicholson JK: **Global metabolic profiling procedures for urine using UPLC-MS.** *Nat Protoc* 2010, **5**:1005-1018.
  - Lei Z, Huhman DV, Sumner LW: **Mass spectrometry strategies in metabolomics.** *J Biol Chem* 2011, **286**:25435-25442.
  - Benton HP, Wong DM, Trauger SA, Siuzdak G: **XCMS2: processing tandem mass spectrometry data for metabolite identification and structural characterization.** *Anal Chem* 2008, **80**:6382-6389.
  - Fernie A, Trethewey R, Krotzky A, Willmitzer L: **Innovation - metabolite profiling: from diagnostics to systems biology.** *Nat Rev Mol Cell Biol* 2004, **5**:763-769.
  - Consortium WTCC: **Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls.** *Nature* 2007, **447**:661-678.
  - Newgard CB, An J, Bain JR, Muehlbauer MJ, Stevens RD, Lien LF, Haqq AM, Shah SH, Arlotto M, Slentz CA, Rochon J, Gallup D, Ilkayeva O, Wenner BR, Yancy WS, Eisenstein H, Musante G, Surwit RS, Millington DS, Butler MD, Svetkey LP: **A branched-chain amino acid-related metabolic signature that differentiates obese and lean humans and contributes to insulin resistance.** *Cell Metab* 2009, **9**:311-326.
  - Sreekumar A, Poisson LM, Rajendiran TM, Khan AP, Cao Q, Yu J, Laxman B, Mehra R, Lonigro RJ, Li Y, Nyati MK, Ahsan A, Kalyana-Sundaram S, Han B, Cao X, Byun J, Omenn GS, Ghosh D, Pennathur S, Alexander DC, Berger A, Shuster JR, Wei JT, Varambally S, Beecher C, Chinnaiyan AM: **Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression.** *Nature* 2009, **457**:910-914.

34. Yap IKS, Angley M, Veselkov KA, Holmes E, Lindon JC, Nicholson JK: **Urinary metabolic phenotyping differentiates children with autism from their unaffected siblings and age-matched controls.** *J Proteome Res* 2010, **9**:2996-3004.
35. Bjerrum JT, Nielsen OH, Hao F, Tang H, Nicholson JK, Wang Y, Olsen J: **Metabonomics in ulcerative colitis: diagnostics, biomarker identification, and insight into the pathophysiology.** *J Proteome Res* 2010, **9**:954-962.
36. Vinayavekhin N, Homan EA, Saghatelian A: **Exploring disease through metabolomics.** *Acc Chem Biol* 2010, **5**:91-103.
37. Holmes E, Loo RL, Stampler J, Bictash M, Yap IKS, Chan Q, Ebbels T, de Iorio M, Brown IJ, Veselkov KA, Daviglus ML, Kesteloot H, Ueshima H, Zhao L, Nicholson JK, Elliott P: **Human metabolic phenotype diversity and its association with diet and blood pressure.** *Nature* 2008, **453**:396-400.
38. Yap IKS, Brown IJ, Chan Q, Wijeyesekera A, Garcia-Perez I, Bictash M, Loo RL, Chadeau-Hyam M, Ebbels T, de Iorio M, Maibaum E, Zhao L, Kesteloot H, Daviglus ML, Stampler J, Nicholson JK, Elliott P, Holmes E: **Metabolome-wide association study identifies multiple biomarkers that discriminate north and south chinese populations at differing risks of cardiovascular disease INTERMAP Study.** *J Proteome Res* 2010, **9**:6647-6654.
39. Bictash M, Ebbels TM, Chan Q, Loo RL, Yap IKS, Brown IJ, de Iorio M, Daviglus ML, Holmes E, Stampler J, Nicholson JK, Elliott P: **Opening up the "black box": metabolic phenotyping and metabolome-wide association studies in epidemiology.** *J Clin Epidemiol* 2010, **63**:970-979.
40. Nicholson JK, Wilson ID, Lindon JC: **Pharmacometabonomics as an effector for personalized medicine.** *Pharmacogenomics* 2011, **12**:103-111.
41. Clayton TA, Baker D, Lindon JC, Everett JR, Nicholson JK: **Pharmacometabonomic identification of a significant host-microbiome metabolic interaction affecting human drug metabolism.** *Proc Natl Acad Sci U S A* 2009, **106**:14728-14733.
42. Winnike JH, Li Z, Wright FA, Macdonald JM, O'Connell TM, Watkins PB: **Use of pharmaco-metabonomics for early prediction of acetaminophen-induced hepatotoxicity in humans.** *Clin Pharmacol Ther* 2010, **88**:45-51.
43. Backshall A, Sharma R, Clarke SJ, Keun HC: **Pharmacometabonomic profiling as a predictor of toxicity in patients with inoperable colorectal cancer treated with capecitabine.** *Clin Cancer Res* 2011, **17**:3019-3028.
44. Wilcken B, Wiley V, Hammond J, Carpenter K: **Screening newborns for inborn errors of metabolism by tandem mass spectrometry.** *N Engl J Med* 2003, **348**:2304-2312.
45. Engelke UFH, Tassinari M, Hayek J, de Vries M, Bilos A, Vivi A, Valensin G, Buoni S, Zannolli R, Brussel W, Kremer B, Salomons GS, Veendrick-Meeke MJBM, Kluijtmans LAJ, Morava E, Wevers RA: **Guanidinoacetate methyltransferase (GAMT) deficiency diagnosed by proton NMR spectroscopy of body fluids.** *NMR Biomed* 2009, **22**:538-544.
46. Maschke S, Wahl A, Azaroual N, Boulet O, Crunelle V, Imbenotte M, Foulard M, Vermeersch G, Lhermitte M: **H-1-NMR analysis of trimethylamine in urine for the diagnosis of fish-odour syndrome.** *Clin Chim Acta* 1997, **263**:139-146.
47. Moolenaar S, Engelke U, Abeling N, Mandel H, Duran M, Wevers R: **Prolidase deficiency diagnosed by H-1 NMR spectroscopy of urine.** *J Inher Metab Dis* 2001, **24**:843-850.
48. Sewell AC, Murphy HC, Iles RA: **Proton nuclear magnetic resonance spectroscopic detection of sialic acid storage disease.** *Clin Chem* 2002, **48**:357-359.
49. Engelke UFH, Sass JO, Van Coster RN, Gerlo E, Olbrich H, Krywawych S, Calvin J, Hart C, Omrans H, Wevers RA: **NMR spectroscopy of aminoacylase 1 deficiency, a novel inborn error of metabolism.** *NMR Biomed* 2008, **21**:138-147.
50. Moolenaar SH, Göhlich-Ratmann G, Engelke UF, Spraul M, Humpfer E, Dvortsak P, Voit T, Hoffmann GF, Bräutigam C, van Kuilenburg AB, van Gennip A, Vreken P, Wevers RA: **beta-Ureidopropionase deficiency: a novel inborn error of metabolism discovered using NMR spectroscopy on urine.** *Magn Reson Med* 2001, **46**:1014-1017.
51. Engelke U, Kremer B, Kluijtmans L, van der Graaf M, Morava E, Loupatty F, Wanders R, Moskau D, Loss S, van den Bergh E, Wevers R: **NMR spectroscopic studies on the late onset form of 3-methylglutaconic aciduria type I and other defects in leucine metabolism.** *NMR Biomed* 2006, **19**:271-278.
52. Holmes E, Foxall P, Spraul M, Farrant R, Nicholson J, Lindon J: **750 MHz H-1 NMR spectroscopy characterisation of the complex metabolic pattern of urine from patients with inborn errors of metabolism: 2-hydroxyglutaric aciduria and maple syrup urine disease.** *J Pharmaceut Biomed* 1997, **15**:1647-1659.
53. Wikoff WR, Gangoi JA, Barshop BA, Siuzdak G: **Metabolomics identifies perturbations in human disorders of propionate metabolism.** *Clin Chem* 2007, **53**:2169-2176.
54. Caruso P, Poussaint T, Tzika A, Zurakowski D, Astrakas L, Elias E, Bay C, Irons M: **MRI and H-1 MRS findings in Smith-Lemli-Opitz syndrome.** *Neuroradiology* 2004, **46**:3-14.
55. Gronwald W, Klein MS, Zeltner R, Schulze B-D, Reinhold SW, Deutschmann M, Immervoll A-K, Böger CA, Banas B, Eckardt K-U, Oefner PJ: **Detection of autosomal dominant polycystic kidney disease by NMR spectroscopic fingerprinting of urine.** *Kidney Int* 2011, **79**:1244-1253.
56. Vilasi A, Cutillas PR, Maher AD, Zirah SFM, Capasso G, Norden AWG, Holmes E, Nicholson JK, Unwin RJ: **Combined proteomic and metabolomic studies in three genetic forms of the renal Fanconi syndrome.** *Am J Physiol Renal Physiol* 2007, **293**:F456-F467.
57. Moolenaar S, Engelke U, Wevers RA: **Proton nuclear magnetic resonance spectroscopy of body fluids in the field of inborn errors of metabolism.** *Ann Clin Biochem* 2003, **40**:16-24.
58. Oostendorp M, Engelke U, Willemsen M, Wevers R: **Diagnosing inborn errors of lipid metabolism with proton nuclear magnetic resonance spectroscopy.** *Clin Chem* 2006, **52**:1395-1405.
59. Lanpher B, Brunetti-Pierri N, Lee B: **Inborn errors of metabolism: the flux from Mendelian to complex diseases.** *Nat Rev Genet* 2006, **7**:449-460.
60. Willemsen G, Boomsma DI, Beem AL, Vink JM, Slagboom PE, Posthuma D: **QTLs for height: results of a full genome scan in Dutch sibling pairs.** *Eur J Hum Genet* 2004, **12**:820-828.
61. Comuzzie AG, Hixson JE, Almasy L, Mitchell BD, Mahaney MC, Dyer TD, Stern MP, MacCluer JW, Blangero J: **A major quantitative trait locus determining serum leptin levels and fat mass is located on human chromosome 2.** *Nat Genet* 1997, **15**:273-276.
62. Rotimi CN, Comuzzie AG, Lowe WL, Luke A, Blangero J, Cooper RS: **The quantitative trait locus on chromosome 2 for serum leptin levels is confirmed in African-Americans.** *Diabetes* 1999, **48**:643-644.
63. Arnett DK, Miller MB, Coon H, Ellison RC, North KE, Province M, Leppert M, Eckfeldt JH: **Genome-wide linkage analysis replicates susceptibility locus for fasting plasma triglycerides: NHLBI Family Heart Study.** *Hum Genet* 2004, **115**:468-474.
64. Duggirala R, Blangero J, Almasy L, Dyer TD, Williams KL, Leach RJ, O'Connell P, Stern MP: **A major susceptibility locus influencing plasma triglyceride concentrations is located on chromosome 15q in Mexican Americans.** *Am J Hum Genet* 2000, **66**:1237-1245.
65. Arya R, Duggirala R, Almasy L, Rainwater DL, Mahaney MC, Cole S, Dyer TD, Williams K, Leach RJ, Hixson JE, MacCluer JW, O'Connell P, Stern MP, Blangero J: **Linkage of high-density lipoprotein-cholesterol concentrations to a locus on chromosome 9p in Mexican Americans.** *Nat Genet* 2002, **30**:102-105.
66. Liija HE, Suviolahti E, Soro-Paavonen A, Hiekkalinna T, Day A, Lange K, Sobel E, Taskinen M-R, Peltonen L, Perola M, Pajukanta P: **Locus for quantitative HDL-cholesterol on chromosome 10q in Finnish families with dyslipidemia.** *J Lipid Res* 2004, **45**:1876-1884.
67. Petersen A-K, Stark K, Musameh MD, Nelson CP, Römisch-Margl W, Kremer W, Raffler J, Krug S, Skurk T, Rist MJ, Daniel H, Hauner H, Adamski J, Tomaszewski M, Döring A, Peters A, Wichmann H-E, Kaess BM, Kalbitzer HR, Huber F, Pfahler V, Samani NJ, Kronenberg F, Dieplinger H, Illig T, Hengstenberg C, Suhre K, Gieger C, Kastenmüller G: **Genetic associations with lipoprotein subfractions provide information on their biological nature.** *Hum Mol Genet* 2012, **21**:1433-1443.
68. Lango Allen H, Estrada K, Lettre G, Berndt SI, Weedon MN, Rivadeneira F, Willer CJ, Jackson AU, Vedantam S, Raychaudhuri S, Ferreira T, Wood AR, Weyant RJ, Segre AV, Speliotes EK, Wheeler E, Soranzo N, Park J-H, Yang J, Gudbjartsson D, Heard-Costa NL, Randall JC, Qi L, Vernon Smith A, Mägi R, Pastinen T, Liang L, Heid IM, Luan Ja, Thorleifsson G, et al.: **Hundreds of variants clustered in genomic loci and biological pathways affect human height.** *Nature* 2010, **467**:832-838.
69. Jansen RC, Nap JP: **Genetical genomics: the added value from segregation.** *Trends Genet* 2001, **17**:388-391.
70. Schadt EE, Monks SA, Drake TA, Lusk AJ, Che N, Colinayo V, Ruff TG, Milligan SB, Lamb JR, Cavet G, Linsley PS, Mao M, Stoughton RB, Friend SH: **Genetics of gene expression surveyed in maize, mouse and man.** *Nature* 2003, **422**:297-302.
71. Dixon AL, Liang L, Moffatt MF, Chen W, Heath S, Wong KCC, Taylor J, Burnett E, Gut I, Farrall M, Lathrop GM, Abecasis GR, Cookson WOC: **A genome-wide association study of global gene expression.** *Nat Genet* 2007, **39**:1202-1207.

72. Keurentjes JJB, Fu J, de Vos CHR, Lommen A, Hall RD, Bino RJ, van der Plas LHW, Jansen RC, Vreugdenhil D, Koornneef M: **The genetics of plant metabolism.** *Nat Genet* 2006, **38**:842-849.
73. Schauer N, Semel Y, Roessner U, Gur A, Balbo I, Carrari F, Pleban T, Perez-Melis A, Bruedigam C, Kopka J, Willmitzer L, Zamir D, Fernie AR: **Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement.** *Nat Biotechnol* 2006, **24**:447-454.
74. Cazier J-B, Kaisaki PJ, Argoud K, Blaise BJ, Veselkov K, Ebbels TMD, Tsang T, Wang Y, Bihoreau M-T, Mitchell SC, Holmes EC, Lindon JC, Scott J, Nicholson JK, Dumas M-E, Gauguier D: **Untargeted metabolome quantitative trait locus mapping associates variation in urine glycerate to mutant glycerate kinase.** *J Proteome Res* 2012, **11**:631-642.
75. Dumas M-E, Wilder SP, Bihoreau M-T, Barton RH, Fearnside JF, Argoud K, D'Amato L, Wallis RH, Blancher C, Keun HC, Baunsgaard D, Scott J, Sidelmann UG, Nicholson JK, Gauguier D: **Direct quantitative trait locus mapping of mammalian metabolic phenotypes in diabetic and normoglycemic rat models.** *Nat Genet* 2007, **39**:666-672.
76. Klose J, Nock C, Herrmann M, Stühler K, Marcus K, Blüggel M, Krause E, Schalkwyk LC, Rastan S, Brown SDM, Büssov K, Himmelbauer H, Lehrach H: **Genetic analysis of the mouse brain proteome.** *Nat Genet* 2002, **30**:385-393.
77. Gieger C, Geistlinger L, Altmajer E, Hrabé de Angelis M, Kronenberg F, Meitinger T, Mewes H-W, Wichmann H-E, Weinberger KM, Adamski J, Illig T, Suhre K: **Genetics meets metabolomics: a genome-wide association study of metabolite profiles in human serum.** *PLoS Genet* 2008, **4**:e1000282.
78. Illig T, Gieger C, Zhai G, Römisch-Margl W, Wang-Sattler R, Pohn C, Altmajer E, Kastenmüller G, Kato BS, Mewes H-W, Meitinger T, de Angelis MH, Kronenberg F, Soranzo N, Wichmann H-E, Spector TD, Adamski J, Suhre K: **A genome-wide perspective of genetic variation in human metabolism.** *Nat Genet* 2010, **42**:137-141.
79. Nicholson G, Rantalainen M, Li JV, Maher AD, Malmmodin D, Ahmadi KR, Faber JH, Barrett A, Min JL, Rayner NW, Toft H, Krestyaninova M, Viksna J, Neogi SG, Dumas M-E, Sarkans U, Consortium M, Donnelly P, Illig T, Adamski J, Suhre K, Allen M, Zondervan KT, Spector TD, Nicholson JK, Lindon JC, Baunsgaard D, Holmes E, McCarthy MI, Holmes CC: **A genome-wide metabolic QTL analysis in Europeans implicates two loci shaped by recent positive selection.** *PLoS Genet* 2011, **7**:e1002270.
80. Nicholson G, Rantalainen M, Maher AD, Li JV, Malmmodin D, Ahmadi KR, Faber JH, Hallgrimsdóttir IB, Barrett A, Toft H, Krestyaninova M, Viksna J, Neogi SG, Dumas M-E, Sarkans U, Consortium TM, Silverman BW, Donnelly P, Nicholson JK, Allen M, Zondervan KT, Lindon JC, Spector TD, McCarthy MI, Holmes E, Baunsgaard D, Holmes CC: **Human metabolic profiles are stably controlled by genetic and environmental variation.** *Mol Syst Biol* 2011, **7**:525.
81. Suhre K, Shin S-Y, Petersen A-K, Mohny RP, Meredith D, Wägele B, Altmajer E, CARDIoGRAM, Deloukas P, Erdmann J, Grundberg E, Hammond CJ, de Angelis MH, Kastenmüller G, Köttgen A, Kronenberg F, Mangino M, Meisinger C, Meitinger T, Mewes H-W, Milburn MV, Pohn C, Raffler J, Ried J, Ried JS, Römisch-Margl W, Samani NJ, Small KS, Wichmann H-E, Zhai G, Illig T, et al.: **Human metabolic individuality in biomedical and pharmaceutical research.** *Nature* 2011, **477**:54-60.
82. Suhre K, Wallaschofski H, Raffler J, Friedrich N, Haring R, Michael K, Wasner C, Krebs A, Kronenberg F, Chang D, Meisinger C, Wichmann H-E, Hoffmann W, Völzke H, Völker U, Teumer A, Biffar R, Kocher T, Felix SB, Illig T, Kroemer HK, Gieger C, Römisch-Margl W, Nauck M: **A genome-wide association study of metabolic traits in human urine.** *Nat Genet* 2011, **43**:565-569.
83. Mittelstrass K, Ried JS, Yu Z, Krumsiek J, Gieger C, Pohn C, Roemisch-Margl W, Polonikov A, Peters A, Theis FJ, Meitinger T, Kronenberg F, Weidinger S, Wichmann HE, Suhre K, Wang-Sattler R, Adamski J, Illig T: **Discovery of sexual dimorphisms in metabolic and genetic biomarkers.** *PLoS Genet* 2011, **7**:e1002215.
84. Adamski J: **Genome-wide association studies (GWAS) with metabolomics.** *Genome Med* 2012, **4**: 34
85. Davidovic L, Navratil V, Bonaccorso CM, Catania MV, Bardoni B, Dumas M-E: **A metabolomic and systems biology perspective on the brain of the fragile X syndrome mouse model.** *Genome Res* 2011, **21**:2190-2202.
86. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, Nielsen T, Pons N, Levenez F, Yamada T, Mende DR, Li J, Xu J, Li S, Li D, Cao J, Wang B, Liang H, Zheng H, Xie Y, Tap J, Lepage P, Bertalan M, Batto J-M, Hansen T, Le Paslier D, Linneberg A, Nielsen HB, Pelletier E, Renault P, et al.: **A human gut microbial gene catalogue established by metagenomic sequencing.** *Nature* 2010, **464**:59-65.
87. Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, Fernandes GR, Tap J, Bruls T, Batto J-M, Bertalan M, Borrrel N, Casellas F, Fernandez L, Gautier L, Hansen T, Hattori M, Hayashi T, Kleerebezem M, Kurokawa K, Leclerc M, Levenez F, Manichanh C, Nielsen HB, Nielsen T, Pons N, Poulain J, Qin J, Sicheritz-Ponten T, Tims S, et al.: **Enterotypes of the human gut microbiome.** *Nature* 2011, **473**:174-180.
88. Waldram A, Holmes E, Wang Y, Rantalainen M, Wilson ID, Tuohy KM, McCartney AL, Gibson GR, Nicholson JK: **Top-down systems biology modeling of host metabolite-microbiome associations in obese rodents.** *J Proteome Res* 2009, **8**:2361-2375.
89. Claus SP, Ellero SL, Berger B, Krause L, Bruttin A, Molina J, Paris A, Want EJ, de Waziers I, Cloarec O, Richards SE, Wang Y, Dumas M-E, Ross A, Rezzi S, Kochhar S, van Bladeren P, Lindon JC, Holmes E, Nicholson JK: **Colonization-induced host-gut microbial metabolic interaction.** *MBio* 2011, **2**:e00271-00210.
90. Turnbaugh PJ, Gordon JI: **An invitation to the marriage of metagenomics and metabolomics.** *Cell* 2008, **134**:708-713.
91. Blow N: **DNA sequencing: generation next-next.** *Nat Methods* 2008, **5**:267-274.
92. Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, Shendure J, Bamshad MJ: **Exome sequencing identifies the cause of a mendelian disorder.** *Nat Genet* 2010, **42**:30-35.
93. St Hilaire C, Ziegler SG, Markello TC, Brusco A, Groden C, Gill F, Carlson-Donohoe H, Lederman RJ, Chen MY, Yang D, Siegenthaler MP, Arduino C, Mancini C, Freudenthal B, Stanescu HC, Zdebek AA, Chaganti RK, Nussbaum RL, Kleta R, Gahl WA, Boehm M: **NT5E mutations and arterial calcifications.** *N Engl J Med* 2011, **364**:432-442.
94. Whorley EA, Mayer AN, Syverson GD, Helbling D, Bonacci BB, Decker B, Serpe JM, Dasu T, Tschannen MR, Veith RL, Basehore MJ, Broeckel U, Tomita-Mitchell A, Arca MJ, Casper JT, Margolis DA, Bick DP, Hessner MJ, Routes JM, Verbsky JW, Jacob HJ, Dimmock DP: **Making a definitive diagnosis: successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease.** *Genet Med* 2011, **13**:255-262.
95. **Online Mendelian Inheritance in Man** [http://omim.org/statistics/entry]
96. Green ED, Guyer MS, Institute NHGR: **Charting a course for genomic medicine from base pairs to bedside.** *Nature* 2011, **470**:204-213.
97. Bamshad MJ, Ng SB, Bigham AW, Tabor HK, Emond MJ, Nickerson DA, Shendure J: **Exome sequencing as a tool for Mendelian disease gene discovery.** *Nat Rev Genet* 2011, **12**:745-755.
98. Robinson PN, Krawitz P, Mundlos S: **Strategies for exome and genome sequence data analysis in disease-gene discovery projects.** *Clin Genet* 2011, **80**:127-132.
99. Yang H-C, Chen C-W: **Region-based and pathway-based QTL mapping using a p-value combination method.** *BMC Proc* 2011, **5** Suppl 9:S43.
100. Consortium GP: **A map of human genome variation from population-scale sequencing.** *Nature* 2010, **467**:1061-1073.
101. Mills RE, Walter K, Stewart C, Handsaker RE, Chen K, Alkan C, Abyzov A, Yoon SC, Ye K, Cheetham RK, Chinwalla A, Conrad DF, Fu Y, Grubert F, Hajirasouliha I, Hormozdiari F, Iakoucheva LM, Iqbal Z, Kang S, Kidd JM, Konkel MK, Korn J, Khurana E, Kural D, Lam HYK, Leng J, Li R, Li Y, Lin C-Y, Luo R, et al.: **Mapping copy number variation by population-scale genome sequencing.** *Nature* 2011, **470**:59-65.
102. Biesecker LG, Mullikin JC, Facio FM, Turner C, Cherukuri PF, Blakesley RW, Bouffard GG, Chines PS, Cruz P, Hansen NF, Teer JK, Maskeri B, Young AC, Program NCS, Manolio TA, Wilson AF, Finkel T, Hwang P, Arai A, Remaley AT, Sachdev V, Shamburek R, Cannon RO, Green ED: **The ClinSeq Project: piloting large-scale genome sequencing for research in genomic medicine.** *Genome Res* 2009, **19**:1665-1674.
103. Feenstra B, Skovgaard IM, Broman KW: **Mapping quantitative trait loci by an extension of the Haley-Knott regression method using estimating equations.** *Genetics* 2006, **173**:2269-2282.
104. Purcell S, Neale B, Todd-Brown K, Thomas M, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC: **PLINK: a tool set for whole-genome association and population-based linkage analyses.** *Am J Hum Genet* 2007, **81**:559-575.
105. Blaise BJ, Shintu L, Elena B, Emsley L, Dumas M-E, Toulhoat P: **Statistical recoupling prior to significance testing in nuclear magnetic resonance based metabolomics.** *Anal Chem* 2009, **81**:6242-6251.

doi:10.1186/gm329

**Cite this article as:** Robinette SL, et al.: Genetic determinants of metabolism in health and disease: from biochemical genetics to genome-wide associations. *Genome Medicine* 2012, **4**:30.