

The metabolome 18 years on: a concept comes of age

Douglas B. Kell^{1,2,3} · Stephen G. Oliver^{4,5}

Received: 17 August 2016 / Accepted: 17 August 2016 / Published online: 2 September 2016
© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract

Background The term ‘metabolome’ was introduced to the scientific literature in September 1998.

Aim and key scientific concepts of the review To mark its 18-year-old ‘coming of age’, two of the co-authors of that paper review the genesis of metabolomics, whence it has come and where it may be going.

Keywords Metabolome · Functional genomics · Systems biology · Precision medicine

1 Introduction

The great advances in biology leading up to the discovery of the structure of DNA and the definition of the genetic code (Cobb 2015; Judson 1979), and the tremendous strides made

since then, have been mainly pioneered by molecular genetic studies on model organisms such as *Escherichia coli* and yeasts (*Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*) (Castrillo and Oliver 2004). The genius of molecular genetics lay in the design of experiments whereby fundamental theories of the workings of living cells at the molecular level could be rigorously tested by performing experiments that had a qualitative read-out (either the cells grew or they did not; either colonies were blue or they were not). This was set to change when the first chromosome sequence to be completed (that of *S. cerevisiae* chromosome III; Oliver et al. 1992) revealed that only about 20 % of the protein-encoding genes had previously been discovered by classical genetics augmented by recombinant DNA technology. It was immediately evident that the normal course of genetic research, which proceeds from mutant phenotypes to the definition of the corresponding genotype, had to be reversed. Since DNA sequencing would define all the genes, in the future we would need to move from gene to function, rather than from function to gene (Kell and Oliver 2004) (Fig. 1). This functional analysis would need to be conducted using techniques that were every bit as comprehensive as genome sequencing, and so the different levels of ‘omic analysis were conceived (Oliver 1996).

Transcriptomics (the analysis of the complete complement of (m)RNA molecules in a cell, tissue, or organ) had the twin advantages of being most closely related to genomics and that it could be pursued using similar techniques—either by hybridisation of complementary nucleic acid strands or cDNA sequencing. Like the other functional ‘omes the transcriptome is context-dependent—it changes with the changing physiological, pathological, or developmental state of the cell. For yeast cells, the relationship between the genome and transcriptome is approximately one-to-one; introns and, therefore, differential splicing of mRNAs are

✉ Douglas B. Kell
dbk@manchester.ac.uk

✉ Stephen G. Oliver
sgo24@cam.ac.uk

¹ School of Chemistry, The University of Manchester, 131 Princess St, Manchester M1 7DN, UK

² Manchester Institute of Biotechnology, The University of Manchester, 131 Princess St, Manchester M1 7DN, UK

³ Centre for Synthetic Biology of Fine and Speciality Chemicals (SYNBIOCHEM), The University of Manchester, 131, Princess St, Manchester M1 7DN, UK

⁴ Cambridge Systems Biology Centre, University of Cambridge, Sanger Building, 80 Tennis Court Road, Cambridge CB2 1GA, UK

⁵ Department of Biochemistry, University of Cambridge, Sanger Building, 80 Tennis Court Road, Cambridge CB2 1GA, UK

Forward and reverse genomics

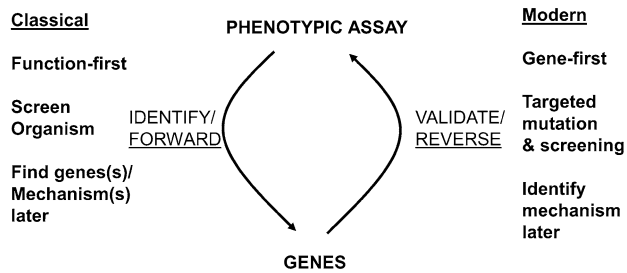


Fig. 1 The ‘forward’ and ‘reverse’ strategies that have been used to link genes and phenotypes. Classically, one would start with a function and seek gene(s) responsible. As it became clearer that most genes were phenotypically silent, it emerged from the systematic genome sequencing programs that only a small fraction of genes had been discovered in this way. The systematic genome sequencing programs also served to change this completely, as once one ‘had’ the genes it was necessary to discover their function. A similar story can be written for drug discovery (Kell 2013)

rare in yeast (Hirschman et al. 2006; Stajich et al. 2007). Proteomics (Wilkins et al. 1996) (the analysis of the complete complement of protein molecules) was also context-dependent, but the relationship (even in yeast) was one-to-many due to post-translational processing and modification of the primary polypeptides generated by protein synthesis. These were the “natural” ’omes that followed from the maxim that “DNA makes RNA makes protein” (and then apparently stops), a maxim that signalled still that ‘molecular biology’ for most people meant ‘macromolecular biology’. Despite its obvious importance in biotechnology (e.g. Bu’lock 1961; Dikicioglu et al. 2013; Nielsen and Keasling 2016), metabolism was seen at that time as something of a Cinderella subject (Griffin 2006), and only a few had pioneered such analyses.

2 A little pre-history

Although it was not called metabolomics, a few early workers had developed interests in using more or less comprehensive metabolic profiling systems to understand complex biological systems. Thus Williams, an early advocate of what we would now call ‘precision medicine’ (Williams 1956), recognised the potential utility of such methods, and the Hornings and their colleagues were at the forefront of instrumental implementations (Dalgliesh et al. 1966; Horning and Horning 1971). DBK carried out his D. Phil (1975–1978) in the laboratory of F. R. (‘Bob’) Whatley, whose colleague Bill Greenaway was explicitly developing GC–MS methods for the analysis of pathogenic fungi and the mode of action of fungicides. Partly because

of the help of an anonymous donor with an interest in the health-giving properties of propolis (Greenaway et al. 1991), the pressure to publish then was not so intense, and this kind of work only appeared rather subsequently (Grant et al. 1988). (It was also based on a naïve interpretation of the ‘crossover theorem’ (Chance and Williams 1955), and lacked the theoretical foundations that metabolic control analysis and systems biology—see below—could provide.) At the time, much of it involved improving the reproducibility, and the production (on a 5Mbyte “Winchester” hard disk the size of a bicycle wheel) of a database of mass spectra. Plus ça change, one might say!

3 The metabolome

Meanwhile, and while proteomics appeared daunting, performing functional analysis at the level of the metabolites appeared far more tractable since we calculated (wrongly, as it turned out: Jewison et al. 2012) that there were only 600–700 metabolites in the yeast cell—about an order of magnitude less than the number of protein-encoding genes (Goffeau et al. 1996). The complete complement of metabolites was also context-dependent, but there was no direct link to the genome since many genes may determine the synthesis and turnover of a single metabolite. Another major difficulty compared with the transcriptome and proteome was the recognition that the physical properties of metabolites were much more widely varied making the metabolites much more differentially extractable, and also that many were quite labile. On the other hand, the metabolic profile was directly and immediately linked to function, and potentially comprehensive methods of analysis (especially mass spectrometry and nuclear magnetic resonance) were available. Metabolic control analysis (MCA) (Fell 1992, 1996; Heinrich and Rapoport 1974; Heinrich and Schuster 1996; Kacser and Burns 1973; Kell et al. 1989; Kell and Westerhoff 1986), a precursor of modern metabolic network biology (Palsson 2006), had long explained why changes in the levels of individual genes or transcripts had relatively little effect on metabolic fluxes, but that they could necessarily—and for precisely the same reasons—have potentially very large effects on metabolite concentrations. Thus, we reasoned, also given that microbes tend to favour growth rate over growth yield (Westerhoff et al. 1983), that in order to maintain the fluxes through the metabolic networks at a relatively constant level, microbial cells would have to vary the concentrations of their constituent metabolites over a wide range—thus the concept, and the term, ‘metabolome’ was born (Oliver et al. 1998).

The initial test of the concept was pioneered in a collaborative effort between our laboratories (then in Manchester and Aberystwyth) and those of Kevin Brindle (in Cambridge) and

Hans Westerhoff/Karel van Dam (in Amsterdam). The idea was that we should be able to elucidate the role of genes of unknown function by comparing the metabolomes of their deletion mutants with those of the deletion mutants of genes of known function. This concept, often called “guilt by association” (Oliver 2000), and a standard strategy in the older ‘operational fingerprinting’ (Meuzelaar et al. 1982) and the newer machine learning (Goodacre et al. 1998), was to become a prevalent one in functional genomics. In this specific example, the use of metabolomics to reveal similarities between yeast mutants was termed FANCY, for Functional ANalysis by Co-responses in Yeast, by Bas Teusink (Teusink et al. 1998)—an acronym which, for better or worse, never caught on. For all that, the concept was robustly validated by the association of the metabolomes of *pfk26* and *pfk27* deletants, and also those of a number of nuclear *petite* mutants (Raamsdonk et al. 2001; Cornish-Bowden and Cárdenas 2001). What was remarkable about this proof-of-principle study was that it worked at all, given the small number of metabolites identified in the NMR analyses. The notion that it was only necessary to monitor the most connected metabolites was tested in Kevin Brindle’s lab, using classical biochemical analyses, but this only served to emphasise the importance of using just one analytical technique to quantify all metabolites. The discriminatory power of just a limited metabolome inspired DBK to suggest monitoring the metabolites excreted into the growth medium—the metabolic ‘footprint’ or exometabolome (Allen et al. 2003, 2004; Kaderbhai et al. 2003; Kell et al. 2005), of which more later. We also recognised that Direct Injection Mass Spectrometry (DIMS) could be used to speciate intact bacterial cells (Vaidyanathan et al. 2001) and other substances (Goodacre et al. 2002), and this DIMS approach has recently been exploited to great effect by Uwe Sauer and colleagues (Link et al. 2015) to analyse the endometabolome by directly injecting living cells into a high-resolution mass spectrometer.

However, the most important outcome of this study was that metabolomics was rapidly embraced across the biological research community, and especially by plant biologists (Fiehn 2002; Fiehn et al. 2000; Jenkins et al. 2004) despite (or perhaps because of; Quanbeck et al. 2012) the fact that higher plants are considered to have the largest and most complex metabolomes in the living world. (However, we note as a caveat that most microbes have still not been brought into laboratory culture and their many secondary metabolites decrypled (Kell et al. 2015a; Lewis et al. 2010)).

4 The previous 18 years

In a 2004 review (Kell 2004), one of us used the methods of text mining to analyse the areas in which metabolomics research was then most focused, identifying three main clusters: technological developments, the integration of

metabolomics with other ‘omics’ (Castrillo et al. 2007), and its use in predicting higher order properties such as disease. Shortly afterwards the Metabolomics Society and this journal were founded, with the annual meetings now attracting almost 1000 participants. The annual numbers of papers with the term metabolome* in their title or abstract continue to rise, and in 2015 amounted, at Web of Knowledge, to 3130 (in a total exceeding 18,000).

Consequently, the space available does not permit us to be even faintly comprehensive about the development of metabolomics—the papers in this journal provide an excellent starting point—but the massive improvement in mass spectrometric and chromatographic methods is clearly a huge driver (Dettmer et al. 2007; Makarov et al. 2006)) and has been so for us (e.g. (Begley et al. 2009; Dunn et al. 2011, 2015; Goodacre et al. 2004; O’Hagan et al. 2005; Zelena et al. 2009), as are improvements in mass precision and metabolite identifiability (Brown et al. 2009; Dunn et al. 2013; Kind and Fiehn 2007; Weber et al. 2011). We have also found the development of metabolic footprinting (Allen et al. 2003, 2004; Kell et al. 2005) (‘exometabolomics’) to be of value, and like many others have used both untargeted metabolomics and the related metabolic profiling (Goodacre et al. 2004) to discover new disease biomarkers (e.g. (Dunn et al. 2007; Kenny et al. 2005, 2010)).

The importance of metabolomics databases (Haug et al. 2013; Skogerson et al. 2011; Wishart et al. 2013; Zhu et al. 2013) and the need to make metabolomics data publically available (Rocca-Serra et al. 2016; Salek et al. 2015) cannot be stressed too highly.

An important trend is the use of ^{13}C labelling for measuring fluxes (Zamboni et al. 2009), as well as the integration of experimental metabolomics with the genome-wide metabolic networks that are becoming available (Herrgård et al. 2008; Swainston et al. 2016; Thiele et al. 2013). Equivalently, and sadly, an important non-trend is any major improvement in the proper use of statistical and related (machine learning) methods in biological (Ioannidis 2005) and especially metabolomics (Broadhurst and Kell 2006) studies.

5 Quo vadis? How will the full potential of metabolomics be revealed?

“It has been said that we always overestimate what we can do in two years and underestimate what we can do in twenty.”

P. Ball & L. Garwin (Ball and Garwin 1992)

Given the above caveat, we do not seek to be overly predictive, but some trends are obvious. The improvement in sample scale (with (Dunn et al. 2011, 2015) or

potentially without (Lewis et al. 2016) the need for drift correction) is clearly one, and this will be aided by the continuing development of inter-laboratory comparisons (Abate-Pella et al. 2015) and standards for data, data analysis, and interoperability and data integration (Goodacre et al. 2007; Grapov et al. 2015; Salek et al. 2013, 2015; Sansone et al. 2007). Such things will assist greatly in the development of personalised medicine and its integration with wearable technologies. As well as the anticipated trends in sensitivity, moving towards the necessary single-cell analyses, it is clear that many more metabolites remain to be discovered, even in simple hosts (Carbonell et al. 2013, 2014) (probably as a result of enzyme promiscuity Currin et al. 2015; Jeffryes et al. 2015). Such analyses are greatly aided by the use of proper descriptors of small molecule structures, such as SMILES (Weininger 1988) and InChI (Coles et al. 2005; Heller et al. 2013; Spjuth et al. 2013), that allow cheminformatic reasoning about properties such as drug-metabolite similarities (Dobson et al. 2009b; O'Hagan and Kell 2015b, O'Hagan and Kell 2016; O'Hagan et al. 2015).

Another trend will be further automation of instrument tuning (Bradbury et al. 2015), non-invasive methods (Ratnay et al. 2014), and an increased portability of instrumentation such that it may even be used in the field (as is now the case for genomics (Ashton et al. 2015; Kilianski et al. 2015) and biometrics). This is clearly assisted by 'ambient mass spectrometry' (Cooks et al. 2006), and the impressive 'iKnife' (Alexander et al. 2016; Balog et al. 2013) pioneering of such measurements in the operating theatre. This kind of development will be especially important in terms of environmental metabolomics (Bundy et al. 2009) and the 'exposome' (the integrated load of xenobiotics that an individual has accumulated in his/her lifetime) (Athersuch and Keun 2015; Rappaport et al. 2014). The extensive data that will be generated will be harvested via the 'Internet of Things' (Ellis et al. 2015), scientific reasoning will be further automated (King et al. 2004, 2009; Williams et al. 2015), and in an era where the methods of 'artificial intelligence' are starting to show human-level abilities, at least in restricted domains (Koza 2010; Mnih et al. 2015; Silver et al. 2016), we shall be wise to exploit such methods.

At least as judged by their appearance in the literature, some enzymes in a given organism are much more greatly studied than are others, a phenomenon referred to as 'publication asymmetry' (César-Razquin et al. 2015). As assessed in that paper (César-Razquin et al. 2015), solute carriers (SLCs (Hediger et al. 2004)) or transporters are the most neglected group of genes in the human genome. Our own analyses also point up their major importance in flux control (Walter et al. 1987), drug transport (Dobson et al. 2009a; Dobson and Kell 2008; Kell 2013, 2015a, 2015b, 2016; Kell et al. 2013, 2011; Kell and Oliver 2014; Lanthaler et al. 2011;

Mendes et al. 2015; O'Hagan and Kell 2015a) and biotechnology (Kell et al. 2015b). Thus we consider that, although challenging, compartment-based metabolomics, where such transporters are necessarily involved, is likely to become a substantial field of itself. Indeed, our improved understanding of a special compartment called the microbiome shows that not all of the genes and metabolites involved in supposedly non-communicable diseases even arise from the host (Honda and Littman 2016; Potgieter et al. 2015; Wang et al. 2011; Wikoff et al. 2009).

Biological studies will be much aided by the ability to manipulate genomes at will. Henrik Kacser, as a major part of his motivation for developing MCA in the first place, had long ago explained why much more sensitive analyses are possible with haploids than with diploids (Kacser and Burns 1981). Thus, a particularly nice example was given by the work of Superti-Furga and colleagues (Winter et al. 2014) on a near-haploid cell line showing that at least 99.5 % of the uptake of the drug sepantronium bromide proceeded through a specific transporter, and thus that any transbilayer flux was negligible.

The original paper (Oliver et al. 1998) concluded "many of these techniques are sufficiently general that, once they have been tried and tested in the experimentally tractable yeast system, they should be directly applicable to the study of the functional genomics of higher organisms". Certainly this has been borne out, and overall, then, metabolomics has had a very healthy childhood and adolescence. Perhaps now the exposome, and even more comprehensive studies, will usher in the (for us much-vaunted (Kell 2004, 2006; Kell et al. 2005) but largely awaited) integration of metabolomics and systems biology. If it does, it will have been well worth the wait.

Acknowledgments SGO thanks both the BBSRC and the UK Technology Strategy Board (Grants BB/C5051140/2 and BB/L004437/1: '13TSB_SynBio'), as well the European Commission (7th Framework Programme BIOEDGE Contract No: 289126), for research funds. DBK thanks the Biotechnology and Biological Sciences Research Council (BBSRC) for financial support (Grant BB/M017702/1). This is a contribution from the Manchester Centre for Synthetic Biology of Fine and Speciality Chemicals (SYNBIO-CHEM). We apologise to the many readers whose work was not cited.

Compliance with ethical standards

Conflict of interest DBK and SGO have no conflict of interest to declare.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Abate-Pella, D., Freund, D. M., Ma, Y., et al. (2015). Retention projection enables accurate calculation of liquid chromatographic retention times across labs and methods. *Journal of Chromatography A*, *1412*, 43–51.
- Alexander, J., Gildea, L., Balog, J., et al. (2016). A novel methodology for in vivo endoscopic phenotyping of colorectal cancer based on real-time analysis of the mucosal lipidome: A prospective observational study of the iKnife. *Surgical Endoscopy* 1–10.
- Allen, J. K., Davey, H. M., Broadhurst, D., et al. (2003). High-throughput characterisation of yeast mutants for functional genomics using metabolic footprinting. *Nature Biotechnology*, *21*, 692–696.
- Allen, J., Davey, H. M., Broadhurst, D., et al. (2004). Discrimination of the modes of action of antifungal substances by use of metabolic footprinting. *Applied and Environmental Microbiology*, *70*, 6157–6165.
- Ashton, P. M., Nair, S., Dallman, T., et al. (2015). MinION nanopore sequencing identifies the position and structure of a bacterial antibiotic resistance island. *Nature Biotechnology*, *33*, 296–300.
- Athersuche, T. J., & Keun, H. C. (2015). Metabolic profiling in human exposome studies. *Mutagenesis*, *30*(6), 755–762.
- Ball, P., & Garwin, L. (1992). Science at the atomic scale. *Nature*, *355*, 761–766.
- Balog, J., Sasi-Szabó, L., Kinross, J., et al. (2013). Intraoperative tissue identification using rapid evaporative ionization mass spectrometry. *Science Translational Medicine*, *5*, 194ra93.
- Begley, P., Francis-McIntyre, S., Dunn, W. B., et al. (2009). Development and performance of a gas chromatography-time-of-flight mass spectrometry analysis for large-scale non-targeted metabolomic studies of human serum. *Analytical Chemistry*, *81*, 7038–7046.
- Bradbury, J., Genta-Jouve, G., Allwood, J. W., et al. (2015). MUSCLE: Automated multi-objective evolutionary optimization of targeted LC-MS/MS analysis. *Bioinformatics*, *31*, 975–977.
- Broadhurst, D., & Kell, D. B. (2006). Statistical strategies for avoiding false discoveries in metabolomics and related experiments. *Metabolomics*, *2*, 171–196.
- Brown, M., Dunn, W. B., Dobson, P., et al. (2009). Mass spectrometry tools and metabolite-specific databases for molecular identification in metabolomics. *Analyst*, *134*, 1322–1332.
- Bu'lock, J. D. (1961). Intermediary metabolism and antibiotic synthesis. *Advances in Applied Microbiology Physiology*, *3*, 293–333.
- Bundy, J. G., Davey, M. P., & Viant, M. R. (2009). Environmental metabolomics: A critical review and future perspectives. *Metabolomics*, *5*, 3–21.
- Carbonell, P., Parutto, P., Herisson, J., et al. (2014). XTMS: Pathway design in an eXTended metabolic space. *Nucleic Acids Research*, *42*, W389–W394.
- Carbonell, P., Planson, A. G., & Faulon, J. L. (2013). Retrosynthetic design of heterogeneous pathways. *Systems Metabolic Engineering*, *985*, 149–173.
- Castrillo, J. I., & Oliver, S. G. (2004). Yeast as a touchstone in post-genomic research: Strategies for integrative analysis in functional genomics. *Journal of Biochemistry and Molecular Biology*, *37*, 93–106.
- Castrillo, J. I., Zeef, L. A., Hoyle, D. C., et al. (2007). Growth control of the eukaryote cell: a systems biology study in yeast. *Journal of Biology*, *6*, 4.
- César-Razquin, A., Snijder, B., Frappier-Brinton, T., et al. (2015). A call for systematic research on solute carriers. *Cell*, *162*, 478–487.
- Chance, B., & Williams, G. R. (1955). Respiratory enzymes in oxidative phosphorylation. III The steady state. *Journal of Biological Chemistry*, *217*, 409–427.
- Cobb, M. (2015). *Life's greatest secret: The race to crack the genetic code*. London: Profile Books.
- Coles, S. J., Day, N. E., Murray-Rust, P., Rzepa, H. S., & Zhang, Y. (2005). Enhancement of the chemical semantic web through the use of InChI identifiers. *Organic & Biomolecular Chemistry*, *3*, 1832–1834.
- Cooks, R. G., Ouyang, Z., Takats, Z., & Wiseman, J. M. (2006). Ambient mass spectrometry. *Science*, *311*, 1566–1570.
- Cornish-Bowden, A., & Cárdenas, M. L. (2001). Silent genes given voice. *Nature*, *409*, 571–572.
- Currin, A., Swainston, N., Day, P. J., & Kell, D. B. (2015). Synthetic biology for the directed evolution of protein biocatalysts: navigating sequence space intelligently. *Chemical Society Reviews*, *44*, 1172–1239.
- Dalgliesh, C. E., Horning, E. C., Horning, M. G., Knox, K. L., & Yarger, K. (1966). A gas-liquid-chromatographic procedure for separating a wide range of metabolites occurring in urine or tissue extracts. *Biochemical Journal*, *101*, 792–810.
- Dettmer, K., Aronov, P. A., & Hammock, B. D. (2007). Mass spectrometry-based metabolomics. *Mass Spectrometry Reviews*, *26*, 51–78.
- Dikiçioğlu, D., Pir, P., & Oliver, S. G. (2013). Predicting complex phenotype-genotype interactions to enable yeast engineering: *Saccharomyces cerevisiae* as a model organism and a cell factory. *Biotechnology Journal*, *8*, 1017–1034.
- Dobson, P. D., & Kell, D. B. (2008). Carrier-mediated cellular uptake of pharmaceutical drugs: An exception or the rule? *Nature Reviews Drug Discovery*, *7*, 205–220.
- Dobson, P., Lanthaler, K., Oliver, S. G., & Kell, D. B. (2009a). Implications of the dominant role of cellular transporters in drug uptake. *Current Topics in Medicinal Chemistry*, *9*, 163–184.
- Dobson, P. D., Patel, Y., & Kell, D. B. (2009b). “Metabolite-likeness” as a criterion in the design and selection of pharmaceutical drug libraries. *Drug Discovery Today*, *14*, 31–40.
- Dunn, W. B., Broadhurst, D., Begley, P., et al. (2011). TheHusermet consortium, procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nature Protocols*, *6*, 1060–1083.
- Dunn, W. B., Broadhurst, D. I., Sasalu, D., et al. (2007). Serum metabolomics reveals many novel metabolic markers of heart failure, including pseudouridine and 2-oxoglutarate. *Metabolomics*, *3*, 413–426.
- Dunn, W. B., Erban, A., Weber, R. J. M., et al. (2013). Mass Appeal: Metabolite identification in mass spectrometry-focused untargeted metabolomics. *Metabolomics*, *9*, S44–S66.
- Dunn, W. B., Lin, W., Broadhurst, D., et al. (2015). Molecular phenotyping of a UK population: Defining the human serum metabolome. *Metabolomics*, *11*, 9–26.
- Ellis, D. I., Muhamadali, H., Haughey, S. A., et al. (2015). Point-and-shoot: rapid quantitative detection methods for on-site food fraud analysis—moving out of the laboratory and into the food supply chain. *Anal Meth*, *7*, 9401–9414.
- Fell, D. A. (1992). Metabolic Control Analysis—a survey of its theoretical and experimental development. *Biochemical Journal*, *286*, 313–330.
- Fell, D. A. (1996). *Understanding the control of metabolism*. London: Portland Press.
- Fiehn, O. (2002). Metabolomics: The link between genotypes and phenotypes. *Plant Molecular Biology*, *48*, 155–171.
- Fiehn, O., Kopka, J., Dormann, P., Altmann, T., Trethewey, R. N., & Willmitzer, L. (2000). Metabolite profiling for plant functional genomics. *Nature Biotechnology*, *18*, 1157–1161.
- Goffeau, A., Barrell, B. G., Bussey, H., et al. (1996). Life with 6000 genes. *Science*, *274*, 546–567.

- Goodacre, R., Broadhurst, D., Smilde, A., et al. (2007). Proposed minimum reporting standards for data analysis in metabolomics. *Metabolomics*, 3, 231–241.
- Goodacre, R., Timmins, É. M., Burton, R., et al. (1998). Rapid identification of urinary tract infection bacteria using hyperspectral whole-organism fingerprinting and artificial neural networks. *Microbiology*, 144, 1157–1170.
- Goodacre, R., Vaidyanathan, S., Bianchi, G., & Kell, D. B. (2002). Metabolic profiling using direct infusion electrospray ionisation mass spectrometry for the characterisation of olive oils. *Analyst*, 127, 1457–1462.
- Goodacre, R., Vaidyanathan, S., Dunn, W. B. (2004). Metabolomics by numbers: acquiring and understanding global metabolite data. *Trends in Biotechnology*, 22, 245–252.
- Grant, B. R., Greenaway, W., & Whatley, F. R. (1988). Metabolic changes during development of *Phytophthora palmivora* examined by Gas-Chromatography Mass-Spectrometry. *Journal of General Microbiology*, 134, 1901–1911.
- Grapov, D., Wanichthanarak, K., & Fiehn, O. (2015). MetaMapR: Pathway independent metabolomic network analysis incorporating unknowns. *Bioinformatics*, 31, 2757–2760.
- Greenaway, W., May, J., Scaysbrook, T., & Whatley, F. R. (1991). Identification by gas chromatography-mass spectrometry of 150 compounds in propolis. *Zeitschrift für Naturforschung C*, 46, 111–121.
- Griffin, J. L. (2006). The Cinderella story of metabolic profiling: does metabolomics get to go to the functional genomics ball? *Philosophical Transactions of the Royal Society of London B*, 361, 147–161.
- Haug, K., Salek, R. M., Conesa, P., et al. (2013). MetaboLights—an open-access general-purpose repository for metabolomics studies and associated meta-data. *Nucleic Acids Research*, 41, D781–D786.
- Hediger, M. A., Romero, M. F., Peng, J. B., et al. (2004). The ABCs of solute carriers: Physiological, pathological and therapeutic implications of human membrane transport proteins. *Pflügers Archiv*, 447, 465–468.
- Heinrich, R., & Rapoport, T. A. (1974). A linear steady-state treatment of enzymatic chains. General properties, control and effector strength. *European Journal of Biochemistry*, 42, 89–95.
- Heinrich, R., & Schuster, S. (1996). *The regulation of cellular systems*. New York: Chapman & Hall.
- Heller, S., McNaught, A., Stein, S., et al. (2013). InChI—the worldwide chemical structure identifier standard. *Journal of Cheminformatics*, 5, 7.
- Herrgård, M. J., Swainston, N., Dobson, P., et al. (2008). A consensus yeast metabolic network obtained from a community approach to systems biology. *Nature Biotechnology*, 26, 1155–1160.
- Hirschman, J. E., Balakrishnan, R., Christie, K. R., et al. (2006). Genome Snapshot: A new resource at the *Saccharomyces Genome Database* (SGD) presenting an overview of the *Saccharomyces cerevisiae* genome. *Nucleic Acids Research*, 34, D442–D445.
- Honda, K., & Littman, D. R. (2016). The microbiota in adaptive immune homeostasis and disease. *Nature*, 535, 75–84.
- Horning, E. C., & Horning, M. G. (1971). Metabolic profiles: Gas-phase methods for analysis of metabolites. *Clinical Chemistry*, 17, 802–809.
- Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, 2, e124.
- Jeffreyes, J. G., Colastani, R. L., Elbadawi-Sidhu, M., et al. (2015). MINEs: Open access databases of computationally predicted enzyme promiscuity products for untargeted metabolomics. *Journal of Cheminformatics*, 7, 44.
- Jenkins, H., Hardy, N., Beckmann, M., et al. (2004). A proposed framework for the description of plant metabolomics experiments and their results. *Nature Biotechnology*, 22, 1601–1606.
- Jewison, T., Knox, C., Neveu, V., et al. (2012). YMDB: The yeast metabolome database. *Nucleic Acids Research*, 40, D815–D820.
- Judson, H. F. (1979). *The eighth day of creation: Makers of the revolution in biology*. New York: Touchstone Books.
- Kacser, H., & Burns, J. A. (1973). The control of flux in Davies. In D. D. Davies (Ed.), *Rate control of biological processes. symposium of the society for experimental biology* (pp. 65–104). Cambridge: Cambridge University Press.
- Kacser, H., & Burns, J. A. (1981). The molecular basis of dominance. *Genetics*, 97, 639–666.
- Kaderbhai, N. N., Broadhurst, D. I., Ellis, D. I., et al. (2003). Functional genomics via metabolic footprinting: Monitoring metabolite secretion by *Escherichia coli* tryptophan metabolism mutants using FT-IR and direct injection electrospray mass spectrometry. *Comparative Functional Genomics*, 4, 376–391.
- Kell, D. B. (2004). Metabolomics and systems biology: making sense of the soup. *Current Opinion in Microbiology*, 7, 296–307.
- Kell, D. B. (2006). Systems biology, metabolic modelling and metabolomics in drug discovery and development. *Drug Disc Today*, 11, 1085–1092.
- Kell, D. B. (2013). Finding novel pharmaceuticals in the systems biology era using multiple effective drug targets, phenotypic screening, and knowledge of transporters: Where drug discovery went wrong and how to fix it. *FEBS Journal*, 280, 5957–5980.
- Kell, D. B. (2015a). The transporter-mediated cellular uptake of pharmaceutical drugs is based on their metabolite-likeness and not on their bulk biophysical properties: Towards a systems pharmacology. *Perspectives on Science*, 6, 66–83.
- Kell, D. B. (2015b). What would be the observable consequences if phospholipid bilayer diffusion of drugs into cells is negligible? *Trends in Pharmacological Sciences*, 36, 15–21.
- Kell, D.B. (2016) How drugs pass through biological cell membranes—a paradigm shift in our understanding? *Beilstein Magazine* 2, http://www.beilstein-institut.de/download/628/09_kell.pdf.
- Kell, D. B., Brown, M., Davey, H. M., et al. (2005). Metabolic footprinting and Systems Biology: The medium is the message. *Nature Reviews Microbiology*, 3, 557–565.
- Kell, D. B., Dobson, P. D., Bilsland, E., & Oliver, S. G. (2013). The promiscuous binding of pharmaceutical drugs and their transporter-mediated uptake into cells: What we (need to) know and how we can do so. *Drug Disc Today*, 18, 218–239.
- Kell, D. B., Dobson, P. D., & Oliver, S. G. (2011). Pharmaceutical drug transport: The issues and the implications that it is essentially carrier-mediated only. *Drug Disc Today*, 16, 704–714.
- Kell, D. B., & Oliver, S. G. (2004). Here is the evidence, now what is the hypothesis? The complementary roles of inductive and hypothesis-driven science in the post-genomic era. *Bioessays*, 26, 99–105.
- Kell, D. B., & Oliver, S. G. (2014). How drugs get into cells: tested and testable predictions to help discriminate between transporter-mediated uptake and lipoidal bilayer diffusion. *Frontiers in Pharmacology*, 5, 231.
- Kell, D. B., Potgieter, M., & Pretorius, E. (2015a). Individuality, phenotypic differentiation, dormancy and ‘persistence’ in culturable bacterial systems: Commonalities shared by environmental, laboratory, and clinical microbiology. *F1000Research*, 4, 179.
- Kell, D. B., Swainston, N., Pir, P., & Oliver, S. G. (2015b). Membrane transporter engineering in industrial biotechnology and whole-cell biocatalysis. *Trends in Biotechnology*, 33, 237–246.
- Kell, D. B., van Dam, K., & Westerhoff, H. V. (1989). Control analysis of microbial growth and productivity. *Symp. Soc. Gen. Microbiol.*, 44, 61–93.

- Kell, D. B., & Westerhoff, H. V. (1986). Metabolic control theory: Its role in microbiology and biotechnology. *FEMS Microbiology Reviews*, 39, 305–320.
- Kenny, L. C., Broadhurst, D. I., Dunn, W., et al. (2010). Robust early pregnancy prediction of later preeclampsia using metabolomic biomarkers. *Hypertension*, 56, 741–749.
- Kenny, L. C., Dunn, W. B., Ellis, D. I., et al. (2005). Novel biomarkers for pre-eclampsia detected using metabolomics and machine learning. *Metabolomics*, 1, 227–234. doi:10.1007/s11306-005-0003-1.
- Kilianski, A., Haas, J. L., Corriveau, E. J., et al. (2015). Bacterial and viral identification and differentiation by amplicon sequencing on the MinION nanopore sequencer. *Gigascience*, 4, 12.
- Kind, T., & Fiehn, O. (2007). Seven golden rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. *BMC Bioinformatics*, 8, 105.
- King, R. D., Rowland, J., Oliver, S. G., et al. (2009). The automation of science. *Science*, 324, 85–89.
- King, R. D., Whelan, K. E., Jones, F. M., et al. (2004). Functional genomic hypothesis generation and experimentation by a robot scientist. *Nature*, 427, 247–252.
- Koza, J. R. (2010). Human-competitive results produced by genetic programming. *Genetic Programming and Evolvable Machines*, 11, 251–284.
- Lanthaler, K., Bilsland, E., Dobson, P., et al. (2011). Genome-wide assessment of the carriers involved in the cellular uptake of drugs: A model system in yeast. *BMC Biology*, 9, 70.
- Lewis, K., Epstein, S., D’Onofrio, A., & Ling, L. L. (2010). Uncultured microorganisms as a source of secondary metabolites. *The Journal of Antibiotics (Tokyo)*, 63, 468–476.
- Lewis, M.R., Pearce, J.T., Spagou, K., et al. (2016) Development and Application of UPLC-ToF MS for Precision Large Scale Urinary Metabolic Phenotyping. *Analytical Chemistry*. doi:10.1021/acs.analchem.6b01481.
- Link, H., Fuhrer, T., Gerosa, L., Zamboni, N., & Sauer, U. (2015). Real-time metabolome profiling of the metabolic switch between starvation and growth. *Nature Methods*, 12, 1091–1097.
- Makarov, A., Denisov, E., Kholomeev, A., et al. (2006). Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. *Analytical Chemistry*, 78, 2113–2120.
- Mendes, P., Oliver, S. G., & Kell, D. B. (2015). Fitting transporter activities to cellular drug concentrations and fluxes: Why the bumblebee can fly. *Trends in Pharmacological Sciences*, 36, 710–723.
- Meuzelaar, H. L. C., Haverkamp, J., & Hileman, F. D. (1982). *Pyrolysis mass spectrometry of recent and fossil biomaterials*. Amsterdam: Elsevier.
- Mnih, V., Kavukcuoglu, K., Silver, D., et al. (2015). Human-level control through deep reinforcement learning. *Nature*, 518, 529–533.
- Nielsen, J., & Keasling, J. D. (2016). Engineering cellular metabolism. *Cell*, 164, 1185–1197.
- O’Hagan, S., Dunn, W. B., Brown, M., Knowles, J. D., & Kell, D. B. (2005). Closed-loop, multiobjective optimisation of analytical instrumentation: Gas-chromatography-time-of-flight mass spectrometry of the metabolomes of human serum and of yeast fermentations. *Analytical Chemistry*, 77, 290–303.
- O’Hagan, S., & Kell, D. B. (2015a). The apparent permeabilities of Caco-2 cells to marketed drugs: Magnitude, and independence from both biophysical properties and endogenous similarities. *PeerJ*, 3, e1405.
- O’Hagan, S., & Kell, D. B. (2015b). Understanding the foundations of the structural similarities between marketed drugs and endogenous human metabolites. *Frontiers in Pharmacology*, 6, 105.
- O’Hagan, S. and Kell, D.B. (2016) MetMaxStruct: A Tversky-similarity-based strategy for analysing the (sub)structural similarities of drugs and endogenous metabolites. *Frontiers in Pharmacology*, in press.
- O’Hagan, S., Swainston, N., Handl, J., & Kell, D. B. (2015). A ‘rule of 0.5’ for the metabolite-likeness of approved pharmaceutical drugs. *Metabolomics*, 11, 323–339.
- Oliver, S. G. (1996). From DNA sequence to biological function. *Nature*, 379, 597–600.
- Oliver, S. G. (2000). Guilt-by-association goes global. *Nature*, 403, 601–603.
- Oliver, S. G., Winson, M. K., Kell, D. B., & Baganz, F. (1998). Systematic functional analysis of the yeast genome. *Trends in Biotechnology*, 16, 373–378.
- Oliver, S. G., Vandraart, Q. J. M., Agostonicarbone, M. L., et al. (1992). The complete DNA sequence of yeast chromosome III. *Nature*, 357, 38–46.
- Palsson, B. Ø. (2006). *Systems biology: Properties of reconstructed networks*. Cambridge: Cambridge University Press.
- Potgieter, M., Bester, J., Kell, D. B., & Pretorius, E. (2015). The dormant blood microbiome in chronic, inflammatory diseases. *FEMS Microbiology Reviews*, 39, 567–591.
- Quanbeck, S. M., Brachova, L., Campbell, A. A., et al. (2012). Metabolomics as a hypothesis-generating functional genomics tool for the annotation of arabidopsis thaliana genes of “unknown function”. *Frontiers in Plant Science*, 3, 15.
- Raamsdonk, L. M., Teusink, B., Broadhurst, D., et al. (2001). A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. *Nature Biotechnology*, 19, 45–50.
- Rappaport, S. M., Barupal, D. K., Wishart, D., Vineis, P., & Scalbert, A. (2014). The blood exposome and its role in discovering causes of disease. *Environmental Health Perspectives*, 122, 769–774.
- Ratray, N. J. W., Hamrang, Z., Trivedi, D. K., et al. (2014). Taking your breath away: metabolomics breathes life into personalized medicine. *Trends in Biotechnology*, 32, 538–548.
- Rocca-Serra, P., Salek, R. M., Arita, M., et al. (2016). Data standards can boost metabolomics research, and if there is a will, there is a way. *Metabolomics*, 12, 14.
- Salek, R. M., Neumann, S., Schober, D., et al. (2015). COordination of Standards in Metabolomics (COSMOS): Facilitating integrated metabolomics data access. *Metabolomics*, 11, 1587–1597.
- Salek, R. M., Steinbeck, C., Viant, M. R., et al. (2013). The role of reporting standards for metabolite annotation and identification in metabolomic studies. *GigaScience*, 2, 13.
- Sansone, S. A., Fan, T., Goodacre, R., et al. (2007). The metabolomics standards initiative. *Nature Biotechnology*, 25, 846–848.
- Silver, D., Huang, A., Maddison, C. J., et al. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, 529, 484–489.
- Skogerson, K., Wohlgemuth, G., Barupal, D. K., & Fiehn, O. (2011). The volatile compound binbase mass spectral database. *BMC Bioinformatics*, 12, 321.
- Spjuth, O., Berg, A., Adams, S., & Willighagen, E. L. (2013). Applications of the InChI in cheminformatics with the CDK and Bioclipse. *Journal of Cheminformatics*, 5, 14.
- Stajich, J. E., Dietrich, F. S., & Roy, S. W. (2007). Comparative genomic analysis of fungal genomes reveals intron-rich ancestors. *Genome Biology*, 8, R223.
- Swainston, N., Smallbone, K., Hefzi, H., et al. (2016). Recon 2.2: From reconstruction to model of human metabolism. *Metabolomics*, 12, 109.
- Teusink, B., Baganz, F., Westerhoff, H. V., & Oliver, S. G. (1998). Metabolic Control Analysis as a tool in the elucidation of the function of novel genes. In M. F. Tuite & A. J. P. Brown (Eds.), *Methods in microbiology: Yeast gene analysis* (pp. 297–336). London: Academic Press.

- Thiele, I., Swainston, N., Fleming, R. M. T., et al. (2013). A community-driven global reconstruction of human metabolism. *Nature Biotechnology*, *31*, 419–425.
- Vaidyanathan, S., Rowland, J. J., Kell, D. B., & Goodacre, R. (2001). Rapid discrimination of aerobic endospore-forming bacteria via electrospray ionization mass spectrometry of whole cell suspensions. *Analytical Chemistry*, *73*, 4134–4144.
- Walter, R. P., Morris, J. G., & Kell, D. B. (1987). The roles of osmotic stress and water activity in the inhibition of the growth, glycolysis and glucose phosphotransferase system of *Clostridium pasteurianum*. *Journal of General Microbiology*, *133*, 259–266.
- Wang, Z., Klipfell, E., Bennett, B. J., et al. (2011). Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature*, *472*, 57–63.
- Weber, R. J., Southam, A. D., Sommer, U., & Viant, M. R. (2011). Characterization of isotopic abundance measurements in high resolution FT-ICR and Orbitrap mass spectra for improved confidence of metabolite identification. *Analytical Chemistry*, *83*, 3737–3743.
- Weininger, D. (1988). SMILES, a chemical language and information system. I. Introduction to methodology and encoding rules. *Journal of Chemical Information and Computer Sciences*, *28*, 31–36.
- Westerhoff, H. V., Hellingwerf, K. J., & van Dam, K. (1983). Thermodynamic efficiency of microbial growth is low but optimal for maximal growth rate. *Proceedings of the National Academy of Sciences of the United States of America*, *80*, 305–309.
- Wikoff, W. R., Anfora, A. T., Liu, J., et al. (2009). Metabolomics analysis reveals large effects of gut microflora on mammalian blood metabolites. *Proceedings of the National Academy of Sciences of the United States of America*, *10*, 3698–3703.
- Wilkins, M. R., Pasquali, C., Appel, R. D., et al. (1996). From proteins to proteomes: Large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Biotechnology*, *14*, 61–65.
- Williams, R. J. (1956). *Biochemical Individuality*. New York: John Wiley.
- Williams, K., Bilsland, E., Sparkes, A., et al. (2015). Cheaper faster drug development validated by the repositioning of drugs against neglected tropical diseases. *Journal of the Royal Society, Interface*, *12*, 20141289.
- Winter, G. E., Radic, B., Mayor-Ruiz, C., et al. (2014). The solute carrier SLC35F2 enables YM155-mediated DNA damage toxicity. *Nature Chemical Biology*, *10*, 768–773.
- Wishart, D. S., Jewison, T., Guo, A. C., et al. (2013). HMDB 3.0—The Human Metabolome Database in 2013. *Nucleic Acids Research*, *41*, D801–D807.
- Zamboni, N., Fendt, S.-M., Ruhl, M., & Sauer, U. (2009). ¹³C-based metabolic flux analysis. *Nature Protocols*, *4*, 878–892.
- Zelena, E., Dunn, W. B., Broadhurst, D., et al. (2009). Development of a robust and repeatable UPLC-MS method for the long-term metabolomic study of human serum. *Analytical Chemistry*, *81*, 1357–1364.
- Zhu, Z. J., Schultz, A. W., Wang, J., et al. (2013). Liquid chromatography quadrupole time-of-flight mass spectrometry characterization of metabolites guided by the METLIN database. *Nature Protocols*, *8*, 451–460.