

## Review

Correspondence  
Weiwen Zhang  
Weiwen.Zhang@asu.edu

# Integrating multiple ‘omics’ analysis for microbial biology: application and methodologies

Weiwen Zhang,<sup>1</sup> Feng Li<sup>2</sup> and Lei Nie<sup>3</sup>

<sup>1</sup>Center for Ecogenomics, Biodesign Institute, Arizona State University, Tempe, AZ 85287-6501, USA

<sup>2</sup>Division of Biometrics II, Office of Biometrics/OTS/CDER/FDA, Silver Spring, MD 20993-0002, USA

<sup>3</sup>Division of Biometrics IV, Office of Biometrics/OTS/CDER/FDA, Silver Spring, MD 20993-0002, USA

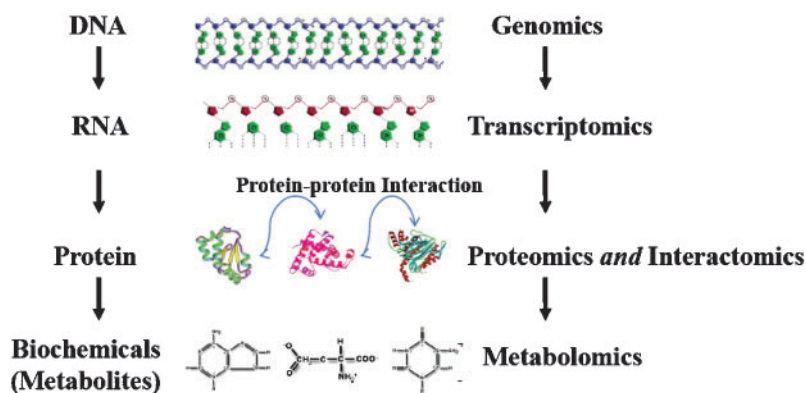
Recent advances in various ‘omics’ technologies enable quantitative monitoring of the abundance of various biological molecules in a high-throughput manner, and thus allow determination of their variation between different biological states on a genomic scale. Several popular ‘omics’ platforms that have been used in microbial systems biology include transcriptomics, which measures mRNA transcript levels; proteomics, which quantifies protein abundance; metabolomics, which determines abundance of small cellular metabolites; interactomics, which resolves the whole set of molecular interactions in cells; and fluxomics, which establishes dynamic changes of molecules within a cell over time. However, no single ‘omics’ analysis can fully unravel the complexities of fundamental microbial biology. Therefore, integration of multiple layers of information, the multi-‘omics’ approach, is required to acquire a precise picture of living micro-organisms. In spite of this being a challenging task, some attempts have been made recently to integrate heterogeneous ‘omics’ datasets in various microbial systems and the results have demonstrated that the multi-‘omics’ approach is a powerful tool for understanding the functional principles and dynamics of total cellular systems. This article reviews some basic concepts of various experimental ‘omics’ approaches, recent application of the integrated ‘omics’ for exploring metabolic and regulatory mechanisms in microbes, and advances in computational and statistical methodologies associated with integrated ‘omics’ analyses. Online databases and bioinformatic infrastructure available for integrated ‘omics’ analyses are also briefly discussed.

## Introduction

Due to revolutionary improvements in high-throughput DNA sequencing technologies, close to 1000 microbial genomes from almost all known major phylogenetic lineages have been fully sequenced, and many more are nearing completion (Medini *et al.*, 2008; Kyrpides, 2009). The computational-based annotation and comparative genomic analyses of DNA sequences have provided biologists with information regarding gene function, genome structures, biological pathways, metabolic and regulatory networks, and evolution of microbial genomes, which has greatly enhanced our understanding of microbial metabolism (Schoolnik, 2001; Ward & Fraser, 2005; Sharan & Ideker, 2006; Cardenas & Tiedje, 2008; Rocha, 2008). However, to fully elucidate microbial metabolism and its responses to environmental factors, it is necessary to include functional characterization and accurate quantification of all levels of gene products, mRNA, proteins and metabolites, as well as their interaction. In the past decade, significant efforts in improving analytical technologies pertaining to measuring mRNA, proteins and metabolites have been made. These efforts have led to the

generation of several new ‘omics’ research fields: transcriptomics, proteomics, metabolomics, interactomics and so on (Fiehn, 2001; Singh & Nagaraj, 2006; Lin & Qian, 2007; Kandpal *et al.*, 2009; Ishii & Tomita, 2009; Tang *et al.*, 2009a) (Fig. 1). In general, all experimental ‘omics’ approaches can be considered to share three major features in contrast to traditional procedures. The first feature is that, unlike traditional methods, ‘omics’ approaches are high-throughput, data-driven, holistic and top-down methodologies; the second feature is the attempt to understand the cell metabolism as one ‘integrated system’ rather than as mere collections of different parts by using information of the relationships between many measured molecular species (Ishii & Tomita, 2009); and the last feature is that these high-throughput ‘omics’ approaches generate large amounts of data and the analysis of these data often requires significant statistical and computational efforts. To date, a great deal of information regarding cellular metabolism has been acquired through application of individual ‘omics’ approaches (Park *et al.*, 2005).

While high-throughput ‘omics’ approaches to analyse molecules at different cellular levels are rapidly becoming



**Fig. 1.** Schematic diagram of various 'omics' technologies targeting different layers of cellular information.

available, it is also becoming clear that any single 'omics' approach may not be sufficient to characterize the complexity of biological systems (Gygi *et al.*, 1999). For example, the expression level of a given gene does not indicate the amount of protein produced, nor its location, biological activity or functional relationship with metabolites. Moreover, in cells many levels of regulation occur after genes have been transcribed, such as post-transcriptional, translational and post-translational regulation, and all forms of biochemical control such as allosteric or feedback regulation. For example, in a study by Ter Kuile & Westerhoff (2001), control of glycolysis was shown to be shared between metabolic, proteomic and genomic levels. Taking this view into account, it is hard to believe that functional genomics can stop at the mRNA level or any single level of information. Integrated multi-'omics' approaches have been applied recently and the studies have enabled researchers to unravel global regulatory mechanisms and complex metabolic networks in various eukaryotic organisms (Hegde *et al.*, 2003; Mootha *et al.*, 2003a, b; Alter & Golub, 2004). While more work still needs to be done in order to improve experimental protocols and computational methodologies so that integrative analysis of multiple large-scale 'omics' datasets can be used to generate new knowledge not accessible by analysis of a single data type alone, these early studies have clearly demonstrated that integrated 'omics' analysis may be a key to decipher complex biological systems. Some of the progress and challenges associated with integrated 'omics' studies have been reviewed previously in several excellent articles (Joyce & Palsson, 2006; De Keersmaecker *et al.*, 2006; Steinfath *et al.*, 2007). In this article we review recent progress, with a focus on the application of integrated 'omics' approaches in various prokaryotic microbial systems and on the advance in various computational methods in dealing with integrated 'omics' data. A brief introduction to various experimental 'omics' platforms is also presented.

### Experimental 'omics' approaches

Cells are living systems full of various functional molecules, which eventually determine the phenotype of the cells.

Such molecules include mRNA transcribed from DNA, proteins translated from mRNA, and various metabolites of small molecular mass generated by various enzymic activities (Fig. 1). Obviously, only analysing the DNA sequences of microbial genomes is not enough to obtain crucial information regarding the functionality of these molecules and the regulatory mechanisms involved in generating these molecules (Nierman *et al.*, 2000). To address the limitation in genome analysis, the last decade has witnessed significant growth in technologies pertaining to molecular biological assays to measure various cellular molecules, and these efforts have led to the establishment of various experimental 'omics' strategies. Several well-established 'omics' platforms that have been used in various integrated 'omics' studies are briefly described below.

### Transcriptomics

Transcriptomics, also called global analysis of gene expression or genome-wide expression profiling, is one of the tools to measure the whole set of all mRNA molecules, or 'transcripts', produced in one cell or a population of cells. Unlike genome sequencing and comparative genomics technologies that focus on DNA, which is static information for any given microbial species and normally does not change significantly in response to short-term external environmental changes, transcriptomics has enabled quantitative measurements of the dynamic expression of mRNA molecules and their variation between different states at the genome scale, thus reflecting the genes that are being actively expressed at any given time, with the exception of mRNA degradation phenomena (Ye *et al.*, 2001; Horak & Snyder, 2002). Several popular high-throughput transcriptomics strategies involve: (i) first identifying mRNAs that differ in their expression status under different experimental conditions and later defining the identity of the respective genes, for example by differential display or serial analysis of gene expression (SAGE); (ii) alternatively assessing changes in the expression of previously defined genes, for example by cDNA or oligonucleotide microarrays (Kagnoff & Eckmann, 2001), or chip-based nanolitre-volume reverse-transcript (RT)-

PCR, which measures gene expression for several thousands of genes simultaneously at higher sensitivity and accuracy than microarrays (Stedtfeld *et al.*, 2008); (iii) a recent approach employing next-generation sequencers (i.e. Roche 454 Sequencer, Illumina Genome Analyser and Applied Biosystems SOLiD System) for direct sequencing of the cDNA converted from whole transcriptomes (Frias-Lopez *et al.*, 2008; Gilbert *et al.*, 2008); and (iv) a more recent direct RNA sequencing (DRS) technology, which allows massively parallel sequencing of RNA molecules directly without prior synthesis of cDNA or the need for ligation/amplification steps (i.e. Helicos Genetic Analysis System) (Ozsolak *et al.*, 2009). In DRS both the abundance and the identity of mRNA molecules can be determined in one analytical process. In addition to a several thousand times higher sequencing throughput, a major advantage of next-generation sequencing over the traditional sequencing method is the dramatically increased degree of parallelism, which can be represented by the number of reads (i.e. the number of DNA templates that can be sequenced simultaneously) in a single sequencing run and the number of sequenced bases per day. Transcriptomics technologies have been used in various microbial systems to explore genome-wide transcriptional activity, and to define regulons, delineate operon structure and perform comparative genotyping (Ye *et al.*, 2001).

Combined with the chromatin immune-precipitation (ChIP) procedure, transcriptomics technology can also be used to investigate the genome-wide location and function of DNA-binding proteins. This so-called ChIP-Chip technology was initially established in yeast (Ren *et al.*, 2000), but has recently been applied to various bacterial systems (Grainger *et al.*, 2004; Herring *et al.*, 2005; Bruscella *et al.*, 2008; Engels *et al.*, 2008; Uyar *et al.*, 2009). Considering the smaller size of bacterial genomes, the technology could be even more powerful in studying genome-wide binding and gene expression regulatory systems in bacteria.

## Proteomics

Proteins are vital parts of living organisms, as they are the major components for building the cellular structure and they also serve as catalytic enzymes in metabolic pathways, and as signal transduction proteins in regulatory pathways of cells (Graham *et al.*, 2007). The ability to identify and measure protein molecules on a large scale depends on recent advances in high-throughput proteomics methodologies, which aim at the simultaneous analysis of all proteins expressed by a cell, tissue or organism in a specific physiological condition. The term proteomics was coined to make an analogy with genomics and transcriptomics as a tool for the large-scale study of proteins, particularly their functions and structures (Wilkins *et al.*, 1996; James, 1997). Proteomics has heavily relied on two major strategies for the separation and visualization of proteins: (i) two-dimensional PAGE (2D-PAGE), in which proteins are

separated according to their isoelectric point and mass, followed by mass spectrometric identification, and (ii) gel-free profiling procedures such as multidimensional separations coupling micro-scale separations [most commonly micro-capillary liquid chromatographic (mLC) fractionation of protein tryptic digests] with automated tandem mass spectrometry (LC-MS/MS) (Baggerman *et al.*, 2005; Nie *et al.*, 2008). 2D-PAGE was first introduced in 1975 (O'Farrell, 1975), and typically allows the detection of only a few hundred proteins after their separation, thereby covering only 'low-complexity' proteomes. With recent application of various fluorescent dyes in this approach, such as 2D fluorescence difference gel electrophoresis (2D DIGE), the coverage and sensitivity of gel-based protein analysis have been significantly improved (Sonck *et al.*, 2009); however, it is still not compatible with gel-free proteomics techniques that can detect several thousand proteins of large dynamic ranges in a single run and that can also provide better detection for low-abundance proteins and insoluble membrane-spanning proteins (Wolff *et al.*, 2006). Even with some improvement, currently neither the 2D-PAGE nor the LC-MS/MS approach is close to saturating identifications in even small bacterial proteomes: typically only 20–40 % of the proteome can be detected, and these are identification approaches and do not provide quantitative information. In order to obtain more accurate proteome measurements, quantitative proteomic methods, such as stable isotope labelling-based isotope-coded affinity tags (ICAT), isobaric tag for relative and absolute quantification (iTRAQ) (Yan *et al.*, 2008) or label-free comparative quantitative proteomics, need to be employed (Haqqani *et al.*, 2008). Proteomics has been extensively used to explore microbial metabolism, differentiation and relationship with environments (for recent reviews see Cash, 2000; Fraser & Rappuoli, 2005; Norbeck *et al.*, 2006; Graham *et al.*, 2007; Zhao & Poh, 2008; Lacerda & Reardon, 2009).

## Interactomics

It has been suggested that cellular life is organized through a complex protein interaction network, in which many proteins participating in the same functional pathways tend to aggregate together in multi-component protein complexes. The large-scale detection of these aggregated proteins, i.e. 'interactomics', thus represents one of the important directions of functional proteomics and it could provide novel insight into microbial cellular metabolism (Singh & Nagaraj, 2006). In addition, the interaction between proteins and DNA is crucial for regulating gene expression and thus key to cellular regulatory networks (Missiuro *et al.*, 2009). The interactome of cells can generally be obtained by three approaches. The first one is to apply computational methods for the systematic identification of protein interactions in bacteria (Parrish *et al.*, 2007; Kim *et al.*, 2008; Rodriguez-Llorente *et al.*, 2009). As one example of this approach, Rodriguez-Llorente *et al.* (2009) recently performed a large-scale

reconstruction of interactomes involved in establishing symbiosis in *Sinorhizobium meliloti*. The study identified 263 novel proteins potentially associated with the Symbiosis Interactome, and the topology of the Symbiosis Interactome was used to guide experimental techniques attempting to validate novel proteins involved in different stages of symbiosis. The second approach is proteomic analysis of the protein complexes, typically performed using affinity tag/pull-down of large protein complexes and then MS/MS characterization at a proteome level (Butland *et al.*, 2005; Gingras *et al.*, 2005; Krogan *et al.*, 2006; Díaz-Mejía *et al.*, 2009). Using this approach, some subsets of the interactomes, such as cell-envelope-associated proteome and interactome in *Escherichia coli*, have been studied (Díaz-Mejía *et al.*, 2009). However, only recently the first bacterial genome-scale interactome has been characterized in *E. coli* K-12. In this study, Hu *et al.* (2009) performed an extensive proteomic survey using affinity-tagged *E. coli* K-12 strains and generated comprehensive genomic context inferences to derive a high-confidence compendium for virtually the entire proteome, consisting of 5993 putative physical interactions and 74 776 putative functional associations. Clustering of the respective probabilistic networks revealed putative orphan membership in discrete multi-protein complexes and functional modules together with annotated gene products, whereas a machine-learning strategy based on network integration implicated the orphans in specific biological processes. The third approach is large-scale genetic interaction screening by phenotype analysis of gene deletion mutants (Tong *et al.*, 2004; Boone *et al.*, 2007; Butland *et al.*, 2008; Typas *et al.*, 2008). Using the genetic interaction screening approach, two research groups independently performed large-scale quantitative analyses of genetic interactions using double mutant strains in *E. coli* (Butland *et al.*, 2008; Typas *et al.*, 2008). These studies have revealed the presence of new reactions and novel pathways that involve not only functionally uncharacterized genes, but also well-characterized genes. Together, they have improved our understanding of gene function and network connections in bacteria.

### Metabolomics and fluxomics

In cells, the rate of enzymic reactions is also regulated by the concentrations of substrates and products (Ter Kuile & Westerhoff, 2001). Moreover, for most organisms, there is no direct relationship between cellular metabolites and genes in the same way as for mRNA and proteins. For example, *Saccharomyces cerevisiae* has fewer than 600 low-molecular-mass cellular intermediates (Oliver *et al.*, 1998), whereas its genome contains ~6200 protein-encoding genes (Raamsdonk *et al.*, 2001). Metabolomics, as a method to define the small-molecule diversity in the cell and to display differences in small molecule abundance, shows many advantages in terms of metabolic analyses because metabolites are the functional entities within the cells and their concentration levels vary as a consequence of genetic

or physiological changes (Raamsdonk *et al.*, 2001). Metabolomics analysis is typically performed by employing gas chromatography time-of-flight mass spectrometry (GC-TOF), high-performance liquid chromatography-mass spectrometry (LC-MS) or capillary electrophoresis-mass spectrometry (CE-MS) instruments, nuclear magnetic resonance (NMR) spectroscopy, and more recently vibrational spectroscopy (of which the resolution and sensitivity are considered as being lower than mass spectrometry). Metabolomics analysis can also be performed through a combined application of several technologies together in order to achieve high coverage and better identification (Kell, 2004; Dunn *et al.*, 2005). Several metabolomic studies aimed at the non-biased comprehensive study of metabolites have been reported in recent years for various microbes including *E. coli* and *S. cerevisiae* (Tweeddale *et al.*, 1998, 1999; Raamsdonk *et al.*, 2001; Soga *et al.*, 2003; Castrillo *et al.*, 2003; van der Werf *et al.*, 2005; Garcia *et al.*, 2008). Compared with transcriptomics and proteomics, technologies used to profile end products of gene expression (e.g. metabolites) are less mature, and most of the metabolomics studies done so far are not yet sufficiently comprehensive (on average only a few dozen metabolites identified in most of these studies) and their measurement accuracy also needs further improvement (Fiehn, 2001; Kell, 2004; Cascante & Marin, 2008). However, these studies have demonstrated that microbial metabolomics could be a powerful tool in deciphering microbial metabolism and bridging the phenotype-genotype gap since it amplifies changes in the proteome and provides a better representation of the phenotype of an organism than any other methods (Cascante & Marin, 2008).

Moreover, in some recent applications, microbial metabolomics studies using isotope-labelled intermediate metabolites combined with various types of dynamic metabolic flux modelling have been suggested as a useful tool in investigations of complex and large-scale metabolic systems, and the term 'fluxomics' has been coined to describe this study area (Forster *et al.*, 2002; Toya *et al.*, 2007; Ishii & Tomita, 2009; Tang *et al.*, 2009a, b). Improvements are still needed in both instruments and data-analysing software to improve identification coverage and accuracy in order to make metabolomics compatible with other 'omics' technologies (Dunn, 2008).

### Recent application of integrated 'omics' in microbial systems

#### Integrated transcriptomics and proteomics

Although one would hypothesize that the correlation between mRNA expression levels and protein abundance will be strong based on the central dogma of molecular genetics, support for this hypothesis from early experimental data is not immediately apparent. Most recent studies have either failed to find a significant correlation between protein and mRNA abundances (Gygi *et al.*, 1999)



or have observed only a weak correlation (Ideker *et al.*, 2001; Greenbaum *et al.*, 2002; Washburn *et al.*, 2003; Zhang *et al.*, 2006b; Nie *et al.*, 2007). It has been suggested that the discrepancy arises from several factors, including protein regulation by post-translational modification, post-transcriptional regulation of protein synthesis, differences in the half-lives of mRNA and proteins, possible functional requirement for protein binding, and significant levels of experimental error (Greenbaum *et al.*, 2002; Beyer *et al.*, 2004; Park *et al.*, 2005). However, this discrepancy also highlights the importance of applying transcriptomic analysis in combination with proteomic analysis, since the biological systems may not be faithfully represented by either of the single individual approaches (Park *et al.*, 2005). Integrated transcriptomics and proteomics is the most commonly used method among various multi-'omics' approaches. Several recent combined transcriptomic and proteomic analyses in microbial systems are summarized in Table 1. The studies can be classified into three major categories based on how the data were analysed.

The first type of study is to use transcriptomics and proteomics data to complement each other, so that the detection bias from each of the technologies can be avoided and a better coverage of the metabolic changes can be achieved. For example, bacterial *c*-type cytochromes undergo a complex post-translational maturation process involving covalent attachment of haem groups. This modification can change the charge state in the gas phase and cause atypical fragmentation of peptides, resulting in loss of detection by current proteomics technology (Aubert *et al.*, 1998). By using the integrated 'omics' approach, the expression of *c*-type cytochromes involved in electron-transport processes was demonstrated in *Desulfovibrio vulgaris* (Nie *et al.*, 2006c; Zhang *et al.*, 2006a, b). This is especially important since most of these studies are aimed at obtaining global metabolic responses after certain treatments (e.g. stress or drug) (Mader *et al.*, 2002a; Budde *et al.*, 2006; Scherl *et al.*, 2006; Brown *et al.*, 2006) (Table 1).

The second type of study is to use transcriptomic and proteomic data for cross-validation purposes. When a good correlation or consensus at both RNA and protein level is found, conclusions can then be made with high confidence (Corbin *et al.*, 2003; Lee *et al.*, 2003; Yoon *et al.*, 2003; Nunez *et al.*, 2006).

The third type of study is to utilize transcriptomic and proteomic data to reveal novel biological insights that are not accessible through any single type of 'omics' dataset. For example, the discrepancy between transcriptomics and proteomics data has been used to suggest possible post-transcription regulation involved in chill-adaptation in *Bacillus subtilis* (Budde *et al.*, 2006) and physiological adjustment of *Halobacterium salinarum* in response to changes in oxygen availability (Schmid *et al.*, 2007). In one recent study using proteomic and transcriptomic data, Nie

*et al.* (2006a) used multiple regression analysis of whole-genome mRNA expression and LC-MS/MS proteome abundance data collected from *D. vulgaris* grown in three conditions, to gain insights into how the mRNA–protein correlation may be affected by various sequence features related to translation efficiency. The analysis suggested that the mRNA–protein correlation is affected primarily by the factors important during the elongation stage, i.e. codon usage and amino acid composition. In contrast, factors related to translation initiation and termination, such as stop codon context and the Shine–Dalgarno sequence, appear to be less important in *D. vulgaris* (Nie *et al.*, 2006a). In another recent study of growth and stationary phase adaptation in *Streptomyces coelicolor*, independent principal components analyses of isobaric stable isotope labelled peptides (iTRAQ)-derived shotgun proteomic data and DNA microarray-derived transcriptome data revealed that the prominent patterns in both protein and mRNA domains are surprisingly well correlated (Jayapal *et al.*, 2008). By employing a systematic concordance analysis, the researchers estimated that over 30 % of the analysed genes likely exhibited significantly divergent patterns, of which nearly one-third even displayed opposing trends. Integrating the data with biological information, they discovered that certain groups of functionally related genes exhibit mRNA–protein discordance in a similar fashion. These observations suggest that differences between mRNA and protein synthesis/degradation mechanisms are prominent in microbes while reaffirming the plausibility of such mechanisms acting in a concerted fashion at a protein complex or subpathway level (Jayapal *et al.*, 2008). Overall, these studies constitute excellent examples of how large-scale transcriptomic and proteomic data can be integrated to gain novel insights into metabolic and regulatory mechanisms of microbial cellular processes.

### Integrated transcriptomics and metabolomics

Integrated transcriptomics and metabolomics analysis is a powerful tool to build the relationship between information elements – genes/transcripts – and functional elements – metabolites – in cells. It was first shown by fungal and plant biologists that transcript and metabolic profiles can be analysed in parallel, and from pairwise transcript–metabolite correlation analysis functionally important metabolites (Askenazi *et al.*, 2003) and genes (Urbanczyk-Wochniak *et al.*, 2003) could be identified (Hoefgen & Nikiforova, 2008). In recent years, the integrated transcriptomics and metabolomics approach has also been applied in various microbial systems. In one study, an integrated transcriptomics, metabolomics and fluxomics analysis of lysine-producing *Corynebacterium glutamicum* ATCC 13287 was performed at different stages of batch culture (Kromer *et al.*, 2004). Good correlation was found between fluxomics and transcriptomics data for certain pathways. For example, the phase shift from growth to lysine production was accompanied by a decrease in glucose uptake flux and the redirection of flux from the

**Table 1.** Recent application of integrated transcriptomics and proteomics studies in microbial systems

Species	Study goals	Key results	Reference
<i>Bacillus subtilis</i>	To study gene/protein expression patterns in response to amino acid availability (with and without 0.2 % Casamino acids)	Genes downregulated by adding Casamino acids were mainly involved in amino acid biosynthesis, such as <i>argC</i> , <i>comK</i> and <i>cotE</i> . Most significantly regulated genes (at least twofold changes) were confirmed by the proteome analysis.	Mader <i>et al.</i> (2002a)
<i>B. subtilis</i>	To characterize the pleiotropic DegS-DegU regulon by combining proteomic and transcriptomic approaches	Besides genes already known to be under the control of DegU-P, novel putative members of this regulon were identified.	Mader <i>et al.</i> (2002b)
<i>B. subtilis</i>	To study the response to the overproduction of an insoluble protein PorA from <i>Neisseria meningitidis</i>	Good correlation between transcriptome and proteome data was observed. Expression of the heat-shock genes <i>dnaK</i> , <i>groEL</i> , <i>grpE</i> , <i>clpP</i> , <i>clpC</i> and two ribosomal protein genes, <i>rpsB</i> and <i>rplJ</i> , was increased.	Jurgen <i>et al.</i> (2001)
<i>B. subtilis</i>	To analyse the adaptational responses of <i>B. subtilis</i> to low temperature	Proteome analysis of chill-adapted cells indicated a major contribution of post-transcriptional regulation phenomena in adaptation to low temperature.	Budde <i>et al.</i> (2006)
<i>Burkholderia pseudomallei</i>	To compare genomes, transcriptomes and proteomes of two natural isolates of <i>B. pseudomallei</i> , the causative agent of the human disease melioidosis	Proteins expressed in a strain-specific manner were similarly correlated at the gene expression level, but up to 38 % of the global proteomic variation between strains comprised proteins expressed in both strains but associated with strain-specific protein isoforms.	Ou <i>et al.</i> (2005)
<i>Desulfovibrio vulgaris</i>	To study the effects of multiple sequence features related to translational efficiency on mRNA-protein correlation	The mRNA-protein correlation was affected the most by the features at elongation stages, i.e. codon usage and amino acid composition (5.3–15.7 % and 5.8–11.9 % of the total variation of mRNA-protein correlation, respectively). Taken together, all sequence features contributed to 15.2–26.2 % of the total variation of mRNA-protein correlation.	Nie <i>et al.</i> (2006a)
<i>D. vulgaris</i>	To study the effects of excess NaCl on <i>D. vulgaris</i>	It was shown that import of osmoprotectants, such as glycine betaine and ectoine, is the primary mechanism used by <i>D. vulgaris</i> to counter hyperionic stress. Several efflux systems were also highly upregulated, as was the ATP synthesis pathway. Integration of data from multiple methods allowed development of a conceptual model for the salt stress response in <i>D. vulgaris</i> that can be compared to those in other micro-organisms.	Mukhopadhyay <i>et al.</i> (2006)
<i>D. vulgaris</i>	To study the responses of the anaerobic, sulfate-reducing organism <i>D. vulgaris</i> to low-oxygen exposure (0.1 % O <sub>2</sub> )	In contrast to 0.1 % O <sub>2</sub> exposure, air exposure was detrimental to both growth rate and viability and caused dramatic changes at both the transcriptome and proteome levels. The results highlight the differences in the cell-wide responses to low and high O <sub>2</sub> levels in <i>D. vulgaris</i> and suggest that while exposure to air is highly detrimental to <i>D. vulgaris</i> , this bacterium can successfully cope with periodic exposure to low O <sub>2</sub> levels in its environment.	Mukhopadhyay <i>et al.</i> (2007)
<i>D. vulgaris</i>	To determine expressivity of genes based on proteomics and transcriptomics data and provide a more functionally based annotation	Expression profiles of 1234 hypothetical proteins and conserved genes were used from transcriptomic datasets, complemented with shotgun LC-MS/MS and AMT tag proteomic data. 1212 of these genes were transcribed, with 786 producing detectable proteins. There was no evidence for expression of 17 predicted genes.	Elias <i>et al.</i> (2009)
<i>Escherichia coli</i>	To study the metabolic and physiological changes of <i>E. coli</i> during the high cell density cultivation	The patterns of gene expression found by proteome and transcriptome analysis were mostly similar. Genes/proteins up- and downregulated were identified. The metabolic responses of <i>E. coli</i> in different growth phases during high cell density culture were described.	Yoon <i>et al.</i> (2003)
<i>E. coli</i>	To compare the parent strain <i>E. coli</i> W3110 and the L-threonine-overproducing mutant <i>E. coli</i> TF5015	Genes displaying distinctly different expression levels are mostly related to the biosynthesis and/or metabolism of aspartate family amino acids and central intermediary metabolism. The differently expressed protein levels are in good agreement with relative mRNA levels of corresponding genes.	Lee <i>et al.</i> (2003)

**Table 1.** cont.

Species	Study goals	Key results	Reference
<i>E. coli</i>	To compare transcriptome and proteome of <i>E. coli</i> MG1655 grown in minimal medium with glycerol as the carbon source	1147 different <i>E. coli</i> proteins were identified. Among them, 1113 (97 %) were expressed at high level (3-fold higher than whole genome average) on Affymetrix GeneChips.	Corbin <i>et al.</i> (2003)
<i>Geobacter sulfurreducens</i>	To define the regulon of the sigma factor RpoS in <i>G. sulfurreducens</i> by using DNA microarray and proteomics	Analysis of a subset of the <i>rpoS</i> mutant proteome indicated that 15 major protein species showed reproducible differences in abundance relative to those of the wild-type strain. Protein identification by MS indicated that the expression of seven of these proteins correlated with the microarray data.	Nunez <i>et al.</i> (2006)
<i>Halobacterium salinarum</i>	To study the physiological adjustment of <i>H. salinarum</i> in response to changes in oxygen availability	Dynamic temporal analysis of relationships between transcription and translation of key genes suggested several important mechanisms for cellular sustenance under anoxia as well as specific instances of post-transcriptional regulation.	Schmid <i>et al.</i> (2007)
<i>Mycobacterium tuberculosis</i>	To decipher the virulence networks regulated by the transcriptional regulator PhoP	A combined transcriptomic and proteomic analysis revealed that PhoP orchestrates a variety of functions implicated in <i>M. tuberculosis</i> virulence and persistence, making <i>phoP</i> mutants promising vaccine candidates.	Gonzalo-Asensio <i>et al.</i> (2008)
<i>Polaromonas</i> sp. JS666	To identify genes/proteins and pathways involved in metabolizing <i>cis</i> -dichloroethene (cDCE)	2D-gel analysis revealed upregulation of glutathione S-transferase, cyclohexanone monooxygenase and haloacid dehalogenase. DNA microarray experiments confirmed the proteomics findings that these genes were among the most highly upregulated by cDCE. The upregulation of genes with antioxidant functions and the inhibition of cDCE degradation by elevated oxygen levels suggest that cDCE induces an oxidative stress response.	Jennings <i>et al.</i> (2009)
<i>Staphylococcus aureus</i>	To study the correlation of proteomic and transcriptomic profiles of <i>S. aureus</i> during the post-exponential phase of growth	Correlation of the two datasets revealed that 42 % of the identified proteins (248 proteins) were amongst the top 25 % of genes with highest mRNA signal intensities, and 69 % of the identified proteins (406 proteins) were amongst the top 50 % with the highest mRNA signal intensities. The fact that the remaining 31 % of proteins were not strongly expressed at the RNA level indicates either that some low-abundance proteins were identified or that some transcripts or proteins showed extended half-lives.	Scherl <i>et al.</i> (2005)
<i>S. aureus</i>	To study the response of three isogenic strains of <i>S. aureus</i> to vancomycin or teicoplanin treatments	Only partial correlation was obtained between transcriptomic and proteomic results during stationary phase. Several overexpressed genes represent previously reported targets, while series of genes and proteins possibly involved in the glycopeptide resistance mechanism were discovered, including regulators, global regulator attenuator, hyper-mutability factor or hypothetical proteins.	Scherl <i>et al.</i> (2006)
<i>Shewanella oneidensis</i>	To characterize the ferric uptake regulator (Fur) modulon in <i>S. oneidensis</i>	Using the integrated transcriptomics and proteomics approach, the study identified nine probable operons (containing 24 genes) and 15 individual ORFs, either with unknown functions or encoding products annotated as transport or binding proteins, that are predicted to be direct targets of Fur-mediated repression.	Wan <i>et al.</i> (2004)
<i>S. oneidensis</i>	To examine knockout effects of the gene encoding SO1377 (gi24372955), a member of the conserved, hypothetical, bacterial protein family COG2268	Whole-genome expression (RNA and protein) profiles revealed numerous gene and protein expression changes relative to the wild-type control, including some involved in iron metabolism, oxidative damage protection and respiratory electron transfer, e.g. complex IV of the respiratory chain.	Gao <i>et al.</i> (2006)

**Table 1.** cont.

Species	Study goals	Key results	Reference
<i>S. oneidensis</i>	To characterize the dynamic molecular response of <i>S. oneidensis</i> to an acute chromate shock	Differential proteomics was used to complement the transcriptome data, resulting in comparable induction and repression patterns for a subset of corresponding proteins. The response of <i>S. oneidensis</i> to chromate shock appears to require a combination of different regulatory networks that involve genes with annotated functions in oxidative stress protection, detoxification, protein stress protection, iron and sulfur acquisition, and SOS-controlled DNA repair mechanisms.	Brown <i>et al.</i> (2006)
<i>Streptomyces coelicolor</i>	To analyse growth and stationary-phase adaptation in <i>S. coelicolor</i>	By employing a systematic concordance analysis, the study estimated that over 30 % of the genes analysed likely exhibited significantly divergent patterns, of which nearly one-third displayed opposing trends. Integrating these data with biological information, certain groups of functionally related genes were shown to exhibit mRNA–protein discordance in a similar fashion.	Jayapal <i>et al.</i> (2008)
<i>Synechocystis</i> sp. PCC 6803	To study the heat-shock response of <i>Synechocystis</i> sp. PCC 6803	The findings indicated that levels of the mRNAs and proteins of chaperonins were well correlated. However, the level of elongation factors is mainly regulated at the protein level.	Suzuki <i>et al.</i> (2006)
<i>Thermobifida fusca</i>	To compare the extracellular proteome of <i>T. fusca</i> grown on cellobiose with that of cells grown on glucose	Transcriptional analysis of genes encoding extracellular proteins suggested that their expression is controlled at the transcriptional level and that their expression also is induced by cellulose.	Chen & Wilson (2007)

tricarboxylic acid (TCA) cycle towards anaplerotic carboxylation and lysine biosynthesis; a significant correlation between flux and expression was also observed for glucose-6-phosphate dehydrogenase, transaldolase and transketolase and for most TCA cycle genes. However, cytoplasmic malate dehydrogenase expression increased despite a reduction of the TCA cycle flux, probably related to its contribution to NADH regeneration under conditions of reduced growth. In addition, most genes for lysine biosynthesis showed a constant expression level, despite a marked change of the metabolic flux, indicating that they are strongly regulated at the metabolic level. Glyoxylate cycle genes were continuously expressed, but the pathway exhibited *in vivo* activity only in the later stage. The most pronounced changes in gene expression during cultivation were found for enzymes at entry points into glycolysis, the pentose phosphate pathway, the TCA cycle and lysine biosynthesis, indicating that these might be of special importance for transcriptional control in *C. glutamicum*. The results demonstrated that the metabolic changes at different layers are not always the same, and that an integrated approach is valuable for fully understanding the microbial systems (Kromer *et al.*, 2004).

In another recent integrated transcriptomics and metabolomics study targeting the infection mechanism of *Rhodococcus fascians* into *Arabidopsis thaliana*, the transcript results showed that *R. fascians* caused a pronounced cytokinin response, illustrated by the activation of cytokinin perception, signal transduction and homeostasis.

Gene Ontology categorization of the differentially expressed genes hinted at a significant impact of infection on the primary metabolism of the host. This was confirmed by subsequent metabolite profiling; for example, invertase transcripts and activities were strongly enhanced upon infection, which probably accounted for the observed increase in the hexose:sucrose ratio and illustrating the establishment of a sink (Depuydt *et al.*, 2009). In a study to compare the aerobic and anaerobic fermentations of *Zymomonas mobilis*, Yang *et al.* (2009) found that greater amounts of end products such as acetate, lactate and acetoin were detected under aerobic conditions and there was only 1.7 % of the amount of ethanol present aerobically as there was anaerobically. However, they did not observe any significant differences in gene expression between aerobic and anaerobic conditions in the early exponential growth phase. They also found that differences in energy-spilling pathways due to oxygen, and uncoupled growth, later led to large differences in ethanol yield and transcriptomic profiles between anaerobic and aerobic stationary-phase cultures (Yang *et al.*, 2009).

#### Other integrated 'omics' studies

To obtain a complete view of microbial metabolism, researchers have recently made efforts to utilize more than two layers of information to seek insights into biological systems. Ishii *et al.* (2007) studied the response of *E. coli* cells to genetic and environmental perturbations. In this study, multiple 'omics' measurements were performed,



including DNA microarray for 4213 *E. coli* genes and qRT-PCR for 85 genes involved in central carbon metabolism, 2D DIGE for approximately 2000 proteins and quantitative methods using liquid chromatography-mass spectrometry/mass spectrometry (LCMS/MS) for 57 proteins involved in central carbon metabolism, CE-TOFMS for 579 metabolites and  $^{13}\text{C}$ -labelled glucose for fluxome analysis. Analysis of metabolic enzyme gene disruptants revealed unexpectedly small changes in mRNA and proteins for most disruptants. Overall, metabolite levels were also stable, reflecting the rerouting of fluxes in the metabolic network. In contrast, the cells actively regulated enzyme levels to maintain a stable metabolic state in response to changes in growth rate. *E. coli* thus seems to use complementary strategies that result in a metabolic network robust against perturbations (Ishii *et al.*, 2007). In another study, Trauger *et al.* (2008) studied the adaptation response of *Pyrococcus furiosus* when shifting from its optimal 95 °C to 72 °C using integrated transcriptomic, proteomic and metabolomic analysis. The complementary information obtained by the various 'omics' techniques was used to catalogue and correlate the overall molecular changes.

## Computational methodologies related to integrated 'omics' analyses

### Methodologies for integrated transcriptomics and proteomics

Several recent studies have shown that only 20–40 % of the variations in protein concentration can be attributed to variable mRNA levels (Nie *et al.*, 2006b; Brockmann *et al.*, 2007). In addition to various biological factors and technical limitations, it has been suggested that the poor correlation between transcriptomic and proteomic data may stem from the inadequacy of available statistical tools to compensate for biases in the data collection methodologies. As summarized in one review (Nie *et al.*, 2007), chances of capturing correlation patterns between transcriptomic and proteomic data could be improved if some of the statistical challenges can be properly addressed, such as proper data transformation and normalization, better statistical tools to deal with experimental measurement errors and missing proteomic data, and tools for handling the nonlinearity property of correlation between transcriptomic and proteomic data. Some progress has been made in recent years; for example, Nie *et al.* (2006a) proposed a zero-inflated Poisson regression model to address issues with missing proteomics data. The key improvement of this model is that the undetected proteins are also taken into consideration, thus allowing an estimation of protein expression even when the proteins are experimentally undetectable due to technical limitations. In addition, the predicted values can be used to correct the measured protein abundance level for experimentally detected proteins by considering their mRNA levels. Fagan *et al.* (2007) used multivariate statistical methods to study gene expression and protein abundance data from studies of the human malarial parasite life cycle and the NCI-60 cancer

cell line. They used the co-inertia analysis (CIA) method to visualize gene and proteomic expression data stemming from the same biological samples and to explore the relationships between two or more datasets, and used principal components analysis or correspondence analysis for data exploration on single datasets. After projecting gene ontology (GO) information onto these plots, the approach can visualize gene expression, protein abundance and GO classes in the same low-dimensional projections and identify GO classes that are likely to be of biological importance (Fagan *et al.*, 2007). Recently, Torres-García *et al.* (2009) used a stochastic Gradient Boosted Trees approach to uncover possible nonlinear relationships between transcriptomic and proteomic data, and to predict protein abundance for the proteins not experimentally detected based on relevant predictors such as mRNA abundance, cellular role, molecular mass, sequence length, protein length, G + C content and triple codon counts.

In addition to the data-driven exploratory approaches mentioned above, some studies have taken biologically based approaches which aim to describe the relationship between proteomic data and mRNA through a mechanistic and dynamic model based on biological relationship. For example, Mehra *et al.* (2003) used a mechanistic genome-wide model for translation that provides mapping between changes in mRNA levels and changes in protein levels. In the model, the results present the correlation between mRNA and protein levels as a function of both the kinetic parameters and concentration of ribosomes at the reference state. In another study, Mogilevskaya *et al.* (2009) presented an approach to developing mathematical description of metabolic pathways, and then to integrate reaction kinetics and different types of experimental data including proteomic, mRNA and metabolite data.

### Methodologies for integrated transcriptomics and metabolomics

The multivariate analysis methods applied to integrating transcriptomics and proteomics (introduced in the previous section) can be well adapted here. In attempts to encompass the profiling of RNA and metabolites of potato, Urbanczyk-Wochniak *et al.* (2003, 2007) used the principal components analysis method to compare the discriminatory power of metabolic and RNA profiling to distinguish different potato tuber systems, and used Spearman's rank-based correlation to determine the relationship between genes and their paired metabolite. In another study, a batch-learning self-organizing network was used to classify the metabolites and the transcripts of *Arabidopsis* according to their time-dependent pattern of changes; the results showed that the metabolites and genes regulated by the same mechanism tended to be clustered together (Hirai *et al.*, 2005). Pir *et al.* (2006) used partial least squares (PLS) to model the metabolic variables as a function of the transcriptome profile. The experimental design allowed the analyses to discriminate between the effects that the growth

medium, dilution rate and deletion of specific genes had on the transcriptome and metabolite profiles (Pir *et al.*, 2006).

Before integrating multi-‘omics’ data, each ‘omics’ dataset must be preprocessed, involving procedures such as normalization or scaling, missing value imputation, and dimension reduction or feature selection. Lê Cao *et al.* (2008, 2009) proposed a sparse PLS approach to combine integration and simultaneous variable (e.g. gene) selection in one step. In this approach, the PLS was penalized by the sum of the absolute values of the coefficients through least absolute shrinkage and selection operator (LASSO) (Tibshirani, 1996), therefore automatically eliminating variables (e.g. genes) with negligible effects. The model selection approach, together with the smoothly clipped absolute deviation approach (Fan & Li, 2001) is effective in analysing data with sparsity (e.g. only a few genes have significant effects).

### Methodologies for integrated proteomics and metabolomics

Some general statistical tools, such as unsupervised multivariate exploratory data mining, correlation network topology analysis, pattern recognition, principal components analysis and sparse partial least squares regression have been used for integrative proteomic and metabolomic analysis (Weckwerth *et al.*, 2004; Morgenthal *et al.*, 2006; Weckwerth & Morgenthal, 2005; Morgenthal *et al.*, 2007; Wienkoop *et al.*, 2008; Lê Cao *et al.*, 2008, 2009). The results demonstrated that good pattern recognition and association identification through proper statistical tools are important for the biological interpretation. In a recent effort to define complex signatures of disease that involve signals from multiple types of biomolecules, Webb-Robertson *et al.* (2009) presented a Bayesian approach to integration that uses posterior probabilities to assign class memberships to samples using individual and multiple data sources; these probabilities are based on lower-level likelihood functions derived from standard statistical learning algorithms. The approach was demonstrated by integrating two proteomic datasets and one metabolic dataset from microbial infections of mice; the results showed that integration of the different datasets improved classification accuracy to ~89 % from the best individual dataset at ~83 %.

### Methodologies for integrating interactomics with other ‘omics’

The interactomics data reveal the physical or functional interaction between components of different layers in a cell. Integrating interactomics with other ‘omics’ data and combining biological mechanistic models leads to more complete insight into the reconstruction of the biological regulation network and thus helps more accurate prediction of the cellular dynamics. While so far no study integrating interactomics with other ‘omics’ data has been published in microbial systems, it is expected that such

integrative studies will appear in the near future. To address the computational capacity needed for this type of integration, some established methodologies, such as dynamic Bayesian networks, artificial neural networks and dynamic control system theory, can play a role, although further exploration is still needed.

### Databases and bioinformatic infrastructures for integrated ‘omics’ analysis

With more cognate experimental ‘omics’ data available, the need to develop public data repositories and to improve data sharing becomes urgent. Ideally, such databases will also serve as the object for the intensive bioinformatics effort for generating new hypotheses and eventually deriving useful information from all the cognate ‘omics’ data collected. To achieve this goal, the databases will not only have a data storage function to keep all the ‘omics’ data and related experimental information for single querying, but should also be further jointly analysed by sophisticated computational and visualization tools available for data interpretation, such as identification of global and local correlation pattern between heterogeneous ‘omics’ datasets, and description of the concordance between mRNA and protein levels for specific functional classes of proteins (Waters *et al.*, 2006). In the past few years, several such online databases have been established and made available for microbiology studies (Martínez-Cruz *et al.*, 2003; Jones *et al.*, 2004; Faria-Campos *et al.*, 2007; Adler *et al.*, 2008). Among them, several are dedicated to specific microbial species, such as ChlamyCys for *Chlamydomonas reinhardtii* (May *et al.*, 2009) and EchoBASE for *E. coli* (Misra *et al.*, 2005), while others are more general and can be used for cognate ‘omics’ data from any species. One good example of these databases is PARE (Protein Abundance and mRNA Expression), which can be used to determine the global correlation between transcriptomic and proteomic data and identify correlation outliers, and also allows users to select subsets of proteins to carry out the analysis (Yu *et al.*, 2007). Currently, almost all of these databases are for integrated transcriptomics and proteomics studies, although some allow incorporation of genomic data (Watson, 2005) or protein–protein interactome data (Maraziotis *et al.*, 2007). Most of the current databases are not equipped with enough statistical tools for integrative analysis. Immediate effort is needed to develop databases which can integrate metabolomics or other ‘omics’ data, and to incorporate various novel computational methods or software specifically designed for integrated ‘omics’ study.

### Concluding remarks

Due to financial constraints and availability issues, very few studies employed more than one ‘omics’ technology in any single investigation in the past. However, our survey of the literature suggests that more and more multi-‘omics’

studies have been performed in recent years. It shows that the use of these integrated approaches to analyse a complex process at different levels can provide new insights into microbial biology. Although more efforts are still needed to improve the 'omics' technologies in terms of their detection coverage, sensitivity and specificity, so far the results obtained from these microbe-based studies strongly suggest that integration of knowledge at different levels in the cascade from genes to proteins and further to metabolic fluxes at a genomic scale is a powerful tool, and will be pivotal for understanding how the individual components in the system interact and influence overall cell metabolism. It is also clear from these studies that the ability to conduct multi-'omics' analyses would represent an additional and novel means to generate discrete and testable biological hypotheses from large-scale high-throughput datasets. For example, a strong correlation between transcriptomic or proteomic data can serve as confirmation for the discovery of an induced response to a treatment, and the lack of a strong correlation can help detect experimental errors or suggest the possibility of a biological uncoupling between the corresponding levels of the respective mRNA and protein species (Nie *et al.*, 2007). In the future, such integrated 'omics' investigation of various cellular molecules (i.e. transcriptome, proteome and metabolome) and their interaction in cells (i.e. interactomes) could lead to a quantified description/model of cellular metabolism at a genome scale that can serve as a foundation for further hypothesis-driven investigation (Schilling *et al.*, 1999; Ishii & Tomita, 2009). Eventually those efforts will lead to fundamental new insights into microbial cellular metabolism.

## Disclaimer

Views expressed in this paper are the authors' professional opinions and do not necessarily represent the official position of the US Food and Drug Administration.

## References

- Adler, P., Reimand, J., Jänes, J., Kolde, R., Peterson, H. & Vilo, J. (2008). EGGanin: pathway animations for high-throughput data. *Bioinformatics* **24**, 588–590.
- Alter, O. & Golub, G. H. (2004). Integrative analysis of genomescale data by using pseudoinverse projection predicts novel correlation between DNA replication and RNA transcription. *Proc Natl Acad Sci U S A* **101**, 16577–16582.
- Askenazi, M., Driggers, E. M., Holtzman, D. A., Norman, T. C., Iverson, S., Zimmer, D. P., Boers, M. E., Blomquist, P. R., Martinez, E. J. & other authors (2003). Integrating transcriptional and metabolite profiles to direct the engineering of lovastatin-producing fungal strains. *Nat Biotechnol* **21**, 150–156.
- Aubert, C., Leroy, G., Bianco, P., Forest, E., Bruschi, M. & Dolla, A. (1998). Characterization of the cytochromes *c* from *Desulfovibrio desulfuricans* G201. *Biochem Biophys Res Commun* **242**, 213–218.
- Baggerman, G., Vierstraete, E., De Loof, A. & Schoofs, L. (2005). Gel-based versus gel-free proteomics: a review. *Comb Chem High Throughput Screen* **8**, 669–677.
- Beyer, A., Hollunder, J., Nasheuer, H. P. & Wilhelm, T. (2004). Post-transcriptional expression regulation in the yeast *Saccharomyces cerevisiae* on a genomic scale. *Mol Cell Proteomics* **3**, 1083–1092.
- Boone, C., Bussey, H. & Andrews, B. J. (2007). Exploring genetic interactions and networks with yeast. *Nat Rev Genet* **8**, 437–449.
- Brockmann, R., Beyer, A., Heinisch, J. J. & Wilhelm, T. (2007). Posttranscriptional expression regulation: what determines translation rates? *PLOS Comput Biol* **3**, e57.
- Brown, S. D., Thompson, M. R., Verberkmoes, N. C., Chourey, K., Shah, M., Zhou, J., Hettich, R. L. & Thompson, D. K. (2006). Molecular dynamics of the *Shewanella oneidensis* response to chromate stress. *Mol Cell Proteomics* **5**, 1054–1071.
- Bruscella, P., Eraso, J. M., Roh, J. H. & Kaplan, S. (2008). The use of chromatin immunoprecipitation to define PpsR binding activity in *Rhodobacter sphaeroides* 2.4.1. *J Bacteriol* **190**, 6817–6828.
- Budde, I., Steil, L., Scharf, C., Völker, U. & Bremer, E. (2006). Adaptation of *Bacillus subtilis* to growth at low temperature: a combined transcriptomic and proteomic appraisal. *Microbiology* **152**, 831–853.
- Butland, G., Peregrín-Alvarez, J. M., Li, J., Yang, W., Yang, X., Canadien, V., Starostine, A., Richards, D., Beattie, B. & other authors (2005). Interaction network containing conserved and essential protein complexes in *Escherichia coli*. *Nature* **433**, 531–537.
- Butland, G., Babu, M., Díaz-Mejía, J. J., Bohdana, F., Phanse, S., Gold, B., Yang, W., Li, J., Gagarinova, A. G. & other authors (2008). eSGA: *E. coli* synthetic genetic array analysis. *Nat Methods* **5**, 789–795.
- Cardenas, E. & Tiedje, J. M. (2008). New tools for discovering and characterizing microbial diversity. *Curr Opin Biotechnol* **19**, 544–549.
- Cascante, M. & Marin, S. (2008). Metabolomics and fluxomics approaches. *Essays Biochem* **45**, 67–82.
- Cash, P. (2000). Proteomics in medical microbiology. *Electrophoresis* **21**, 1187–1201.
- Castrillo, J. I., Hayes, A., Mohammed, S., Gaskell, S. J. & Oliver, S. G. (2003). An optimized protocol for metabolome analysis in yeast using direct infusion electrospray mass spectrometry. *Phytochemistry* **62**, 929–937.
- Chen, S. & Wilson, D. B. (2007). Proteomic and transcriptomic analysis of extracellular proteins and mRNA levels in *Thermobifida fusca* grown on cellobiose and glucose. *J Bacteriol* **189**, 6260–6265.
- Corbin, R. W., Paliy, O., Yang, F., Shabanowitz, J., Platt, M., Lyons, C. E., Jr, Root, K., McAuliffe, J., Jordan, M. I. & other authors (2003). Toward a protein profile of *Escherichia coli*: comparison to its transcription profile. *Proc Natl Acad Sci U S A* **100**, 9232–9237.
- De Keersmaecker, S. C., Thijs, I. M., Vanderleyden, J. & Marchal, K. (2006). Integration of omics data: how well does it work for bacteria? *Mol Microbiol* **62**, 1239–1250.
- Depuydt, S., Trenkamp, S., Fernie, A. R., Eiftieh, S., Renou, J. P., Vuylsteke, M., Holsters, M. & Vereecke, D. (2009). An integrated genomics approach to define niche establishment by *Rhodococcus fascians*. *Plant Physiol* **149**, 1366–1386.
- Díaz-Mejía, J. J., Babu, M. & Emili, A. (2009). Computational and experimental approaches to chart the *Escherichia coli* cell-envelope-associated proteome and interactome. *FEMS Microbiol Rev* **33**, 66–97.
- Dunn, W. B. (2008). Current trends and future requirements for the mass spectrometric investigation of microbial, mammalian and plant metabolomes. *Phys Biol* **5**, 11001.
- Dunn, W. B., Bailey, N. J. & Johnson, H. E. (2005). Measuring the metabolome: current analytical technologies. *Analyst* **130**, 606–625.
- Elias, D. A., Mukhopadhyay, A., Joachimiak, M. P., Drury, E. C., Redding, A. M., Yen, H. C., Fields, M. W., Hazen, T. C., Arkin, A. P. & other authors (2009). Expression profiling of hypothetical genes in



- Desulfovibrio vulgaris* leads to improved functional annotation. *Nucleic Acids Res* 37, 2926–2939.
- Engels, V., Lindner, S. N. & Wendisch, V. F. (2008). The global repressor SugR controls expression of genes of glycolysis and of the L-lactate dehydrogenase LdhA in *Corynebacterium glutamicum*. *J Bacteriol* 190, 8033–8044.
- Fagan, A., Culhane, A. C. & Higgins, D. G. (2007). A multivariate analysis approach to the integration of proteomic and gene expression data. *Proteomics* 7, 2162–2171.
- Fan, J. & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J Am Stat Assoc* 96, 1348–1360.
- Faria-Campos, A. C., Gomes, R. R., Moratelli, F. S., Rausch-Fernandes, H., Franco, G. R. & Campos, S. V. (2007). BNDb – Biomolecules Nucleus Database: an integrated proteomics and transcriptomics database. *Genet Mol Res* 6, 937–945.
- Fiehn, O. (2001). Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks. *Comp Funct Genomics* 2, 155–168.
- Forster, J., Gombert, A. K. & Nielsen, J. (2002). A functional genomics approach using metabolomics and *in silico* pathway analysis. *Biotechnol Bioeng* 79, 703–712.
- Fraser, C. M. & Rappuoli, R. (2005). Application of microbial genomic science to advanced therapeutics. *Annu Rev Med* 56, 459–474.
- Frias-Lopez, J., Shi, Y., Tyson, G. W., Coleman, M. L., Schuster, S. C., Chisholm, S. W. & DeLong, E. F. (2008). Microbial community gene expression in ocean surface waters. *Proc Natl Acad Sci U S A* 105, 3805–3810.
- Gao, W., Liu, Y., Giometti, C. S., Tollaksen, S. L., Khare, T., Wu, L., Klingeman, D. M., Fields, M. W. & Zhou, J. (2006). Knock-out of SO1377 gene, which encodes the member of a conserved hypothetical bacterial protein family COG2268, results in alteration of iron metabolism, increased spontaneous mutation and hydrogen peroxide sensitivity in *Shewanella oneidensis* MR-1. *BMC Genomics* 7, 76.
- Garcia, D. E., Baidoo, E. E., Benke, P. I., Pingitore, F., Tang, Y. J., Villa, S. & Keasling, J. D. (2008). Separation and mass spectrometry in microbial metabolomics. *Curr Opin Microbiol* 11, 233–239.
- Gilbert, J. A., Field, D., Huang, Y., Edwards, R., Li, W., Gilna, P. & Joint, I. (2008). Detection of large numbers of novel sequences in the metatranscriptomes of complex marine microbial communities. *PLoS One* 3, e3042.
- Gingras, A. C., Aebersold, R. & Raught, B. (2005). Advances in protein complex analysis using mass spectrometry. *J Physiol* 563, 11–21.
- Gonzalo-Asensio, J., Mostowy, S., Harders-Westervreen, J., Huygen, K., Hernández-Pando, R., Thole, J., Behr, M., Gicquel, B. & Martín, C. (2008). PhoP: a missing piece in the intricate puzzle of *Mycobacterium tuberculosis* virulence. *PLoS One* 3, e3496.
- Graham, R. L. J., Graham, C. & McMullan, G. (2007). Microbial proteomics: a mass spectrometry primer for biologists. *Microb Cell Fact* 6, 26.
- Grainger, D. C., Overton, T. W., Reppas, N., Wade, J. T., Tamai, E., Hobman, J. L., Constantinidou, C., Struhl, K., Church, G. & Busby, S. J. (2004). Genomic studies with *Escherichia coli* MelR protein: applications of chromatin immunoprecipitation and microarrays. *J Bacteriol* 186, 6938–6943.
- Greenbaum, D., Jansen, R. & Gerstein, M. (2002). Analysis of mRNA expression and protein abundance data: an approach for the comparison of the enrichment of features in the cellular population of proteins and transcripts. *Bioinformatics* 18, 585–596.
- Gygi, S. P., Rochon, Y., Franza, B. R. & Aebersold, R. (1999). Correlation between protein and mRNA abundance in yeast. *Mol Cell Biol* 19, 1720–1730.
- Haqqani, A. S., Kelly, J. F. & Stanimirovic, D. B. (2008). Quantitative protein profiling by mass spectrometry using label-free proteomics. *Methods Mol Biol* 439, 241–256.
- Hegde, P. S., White, I. R. & Debouck, C. (2003). Interplay of transcriptomics and proteomics. *Curr Opin Biotechnol* 14, 647–651.
- Herring, C. D., Raffaele, M., Allen, T. E., Kanin, E. I., Landick, R., Ansari, A. Z. & Palsson, B. Ø. (2005). Immobilization of *Escherichia coli* RNA polymerase and location of binding sites by use of chromatin immunoprecipitation and microarrays. *J Bacteriol* 187, 6166–6174.
- Hirai, M. Y., Klein, M., Fujikawa, Y., Yano, M., Goodenowe, B., Yamazaki, Y., Kanaya, S., Nakamura, Y., Kitayama, M. & other authors (2005). Elucidation of gene-to-gene and metabolite-to-gene networks in *Arabidopsis* by integration of metabolomics and transcriptomics. *J Biol Chem* 280, 25590–25595.
- Hoefgen, R. & Nikiforova, V. J. (2008). Metabolomics integrated with transcriptomics: assessing systems response to sulfur-deficiency stress. *Physiol Plant* 132, 190–198.
- Horak, C. E. & Snyder, M. (2002). Global analysis of gene expression in yeast. *Funct Integr Genomics* 2, 171–180.
- Hu, P., Janga, S. C., Babu, M., Díaz-Mejía, J. J., Butland, G., Yang, W., Pogoutse, O., Guo, X., Phanse, S. & other authors (2009). Global functional atlas of *Escherichia coli* encompassing previously uncharacterized proteins. *PLoS Biol* 7, e96.
- Ideker, T., Thorsson, V., Ranish, J. A., Christmas, R., Buhler, J., Eng, J. K., Bumgarner, R., Goodlett, D. R., Aebersold, R. & Hood, L. (2001). Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* 292, 929–934.
- Ishii, N. & Tomita, M. (2009). Multi-omics data-driven systems biology of *E. coli*. In *Systems Biology and Biotechnology of Escherichia coli*, pp. 41–57. Edited by S. Y. Lee, Dordrecht, Netherlands: Springer.
- Ishii, N., Nakahigashi, K., Baba, T., Robert, M., Soga, T., Kanai, A., Hirasawa, T., Naba, M., Hirai, K. & other authors (2007). Multiple high-throughput analyses monitor the response of *E. coli* to perturbations. *Science* 316, 593–597.
- James, P. (1997). Protein identification in the post-genome era: the rapid rise of proteomics. *Q Rev Biophys* 30, 279–333.
- Jayapal, K. P., Philp, R. J., Kok, Y. J., Yap, M. G., Sherman, D. H., Griffin, T. J. & Hu, W. S. (2008). Uncovering genes with divergent mRNA-protein dynamics in *Streptomyces coelicolor*. *PLoS One* 3, e2097.
- Jennings, L. K., Chartrand, M. M., Lacrampe-Couloume, G., Lollar, B. S., Spain, J. C. & Gossett, J. M. (2009). Proteomic and transcriptomic analyses reveal genes upregulated by *cis*-dichloroethene in *Polaromonas* sp. strain JS666. *Appl Environ Microbiol* 75, 3733–3744.
- Jones, A., Hunt, E., Wastling, J. M., Pizarro, A. & Stoeckert, C. J., Jr (2004). An object model and database for functional genomics. *Bioinformatics* 20, 1583–1590.
- Joyce, A. R. & Palsson, B. O. (2006). The model organism as a system: integrating ‘omics’ data sets. *Nat Rev Mol Cell Biol* 7, 198–210.
- Jurgen, B., Hanschke, R., Sarvas, M., Hecker, M. & Schweder, T. (2001). Proteome and transcriptome based analysis of *Bacillus subtilis* cells overproducing an insoluble heterologous protein. *Appl Microbiol Biotechnol* 55, 326–332.
- Kagnoff, M. F. & Eckmann, L. (2001). Analysis of host responses to microbial infection using gene expression profiling. *Curr Opin Microbiol* 4, 246–250.
- Kandpal, R., Saviola, B. & Felton, J. (2009). The era of omics unlimited. *Biotechniques* 46, 351–355.
- Kell, D. B. (2004). Metabolomics and systems biology: making sense of the soup. *Curr Opin Microbiol* 7, 296–307.



- Kim, J. G., Park, D., Kim, B. C., Cho, S. W., Kim, Y. T., Park, Y. J., Cho, H. J., Park, H., Kim, K. B. & other authors (2008). Predicting the interactome of *Xanthomonas oryzae* pathovar *oryzae* for target selection and DB service. *BMC Bioinformatics* 9, 41.
- Krogan, N. J., Cagney, G., Yu, H., Zhong, G., Guo, X., Ignatchenko, A., Li, J., Pu, S., Datta, N. & other authors (2006). Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* 440, 637–643.
- Kromer, J. O., Sorgenfrei, O., Klopprogge, K., Heinze, E. & Wittmann, C. (2004). In-depth profiling of lysine-producing *Corynebacterium glutamicum* by combined analysis of the transcriptome, metabolome, and fluxome. *J Bacteriol* 186, 1769–1784.
- Kyrpides, N. C. (2009). Fifteen years of microbial genomics: meeting the challenges and fulfilling the dream. *Nat Biotechnol* 27, 627–632.
- Lacerda, C. M. & Reardon, K. F. (2009). Environmental proteomics: applications of proteome profiling in environmental microbiology and biotechnology. *Brief Funct Genomic Proteomic* 8, 75–87.
- Lê Cao, K. A., Rossouw, D., Robert-Granié, C. & Besse, P. (2008). A sparse PLS for variable selection when integrating omics data. *Stat Appl Genet Mol Biol* 7, Article 35.
- Lê Cao, K. A., Martin, P. B., Robert-Granié, C. & Besse, P. (2009). Sparse canonical methods for biological data integration: application to a cross-platform study. *BMC Bioinformatics* 10, 34.
- Lee, J. H., Lee, D. E., Lee, B. U. & Kim, H. S. (2003). Global analyses of transcriptomes and proteomes of a parent strain and an L-threonine-overproducing mutant strain. *J Bacteriol* 185, 5442–5451.
- Lin, J. & Qian, J. (2007). Systems biology approach to integrative comparative genomics. *Expert Rev Proteomics* 4, 107–119.
- Mader, U., Homuth, G., Scharf, C., Buttner, K., Bode, R. & Hecker, M. (2002a). Transcriptome and proteome analysis of *Bacillus subtilis* gene expression modulated by amino acid availability. *J Bacteriol* 184, 4288–4295.
- Mader, U., Antelmann, H., Buder, T., Dahl, M. K., Hecker, M. & Homuth, G. (2002b). *Bacillus subtilis* functional genomics: genome-wide analysis of the DegS-DegU regulon by transcriptomics and proteomics. *Mol Genet Genomics* 268, 455–467.
- Maraziotis, I. A., Dimitrakopoulou, K. & Bezerianos, A. (2007). Growing functional modules from a seed protein via integration of protein interaction and gene expression data. *BMC Bioinformatics* 8, 408.
- Martínez-Cruz, L. A., Rubio, A., Martínez-Chantar, M. L., Labarga, A., Barrio, I., Podhorski, A., Segura, V., Sevilla Campo, J. L., Avila, M. A. & Mato, J. M. (2003). GARBAN: genomic analysis and rapid biological annotation of cDNA microarray and proteomic data. *Bioinformatics* 19, 2158–2160.
- May, P., Christian, J. O., Kempa, S. & Walther, D. (2009). ChlamyCyc: an integrative systems biology database and web-portal for *Chlamydomonas reinhardtii*. *BMC Genomics* 10, 209.
- Medini, D., Serruto, D., Parkhill, J., Relman, D. A., Donati, C., Moxon, R., Falkow, S. & Rappuoli, R. (2008). Microbiology in the post-genomic era. *Nat Rev Microbiol* 6, 419–430.
- Mehra, A., Lee, K. H. & Hatzimanikatis, V. H. (2003). Insights into the relation between mRNA and protein expression patterns. I. Theoretical considerations. *Biotechnol Bioeng* 84, 822–833.
- Misra, R. V., Horler, R. S., Reindl, W., Goryanin, I. I. & Thomas, G. H. (2005). EchoBASE: an integrated post-genomic database for *Escherichia coli*. *Nucleic Acids Res* 33, D329–D333.
- Missiuro, P. V., Liu, K., Zou, L., Ross, B. C., Zhao, G., Liu, J. S. & Ge, H. (2009). Information flow analysis of interactome networks. *PLOS Comput Biol* 5, e1000350.
- Mogilevskaya, E., Bagrova, N., Plyusnina, T., Gizzatkulov, N., Metelkin, E., Goryacheva, E., Smirnov, S., Kosinsky, Y., Dorodnov, A. & other authors (2009). Kinetic modeling as a tool to integrate multilevel dynamic experimental data. *Methods Mol Biol* 563, 197–218.
- Mootha, V. K., Bunkenborg, J., Olsen, J. V., Hjerrild, M., Wisniewski, J. R., Stahl, E., Bolouri, M. S., Ray, H. N., Sihag, S. & other authors (2003a). Integrated analysis of protein composition, tissue diversity, and gene regulation in mouse mitochondria. *Cell* 115, 629–640.
- Mootha, V. K., Lepage, P., Miller, K., Bunkenborg, J., Reich, M., Hjerrild, M., Del-monte, T., Villeneuve, A., Sladek, R. & other authors (2003b). Identification of a gene causing human cytochrome c oxidase deficiency by integrative genomics. *Proc Natl Acad Sci U S A* 100, 605–610.
- Morgenthal, K., Weckwerth, W. & Steuer, R. (2006). Metabolomic networks in plants: Transitions from pattern recognition to biological interpretation. *Biosystems* 82, 108–117.
- Morgenthal, K., Wienkoop, S., Wolschin, F. & Weckwerth, W. (2007). Integrative profiling of metabolites and proteins: improving pattern recognition and biomarker selection for systems level approaches. *Methods Mol Biol* 358, 57–75.
- Mukhopadhyay, A., He, Z., Alm, E. J., Arkin, A. P., Baidoo, E. E., Borglin, S. C., Chen, W., Hazen, T. C., He, Q. & other authors (2006). Salt stress in *Desulfovibrio vulgaris* Hildenborough: an integrated genomics approach. *J Bacteriol* 188, 4068–4078.
- Mukhopadhyay, A., Redding, A. M., Joachimiak, M. P., Arkin, A. P., Borglin, S. E., Dehal, P. S., Chakraborty, R., Geller, J. T., Hazen, T. C. & other authors (2007). Cell-wide responses to low-oxygen exposure in *Desulfovibrio vulgaris* Hildenborough. *J Bacteriol* 189, 5996–6010.
- Nie, L., Wu, G. & Zhang, W. (2006a). Correlation of mRNA expression and protein abundance affected by multiple sequence features related to translational efficiency in *Desulfovibrio vulgaris*: a quantitative analysis. *Genetics* 174, 2229–2243.
- Nie, L., Wu, G. & Zhang, W. (2006b). Correlation between mRNA and protein abundance in *Desulfovibrio vulgaris*: a multiple regression to identify sources of variations. *Biochem Biophys Res Commun* 339, 603–610.
- Nie, L., Wu, G., Brockman, F. J. & Zhang, W. (2006c). Integrated analysis of transcriptomic and proteomic data of *Desulfovibrio vulgaris*: zero-inflated Poisson regression models to predict abundance of undetected proteins. *Bioinformatics* 22, 1641–1647.
- Nie, L., Wu, G., Culley, D. E., Scholten, J. C. & Zhang, W. (2007). Integrative analysis of transcriptomic and proteomic data: challenges, solutions and applications. *Crit Rev Biotechnol* 27, 63–75.
- Nie, L., Wu, G. & Zhang, W. (2008). Statistical application and challenges in global gel-free proteomic analysis by mass spectrometry. *Crit Rev Biotechnol* 28, 297–307.
- Nierman, W., Eisen, J. A. & Fraser, C. M. (2000). Microbial genome sequencing 2000: new insights into physiology, evolution and expression analysis. *Res Microbiol* 151, 79–84.
- Norbeck, A. D., Callister, S. J., Monroe, M. E., Jaitly, N., Elias, D. A., Lipton, M. S. & Smith, R. D. (2006). Proteomic approaches to bacterial differentiation. *J Microbiol Methods* 67, 473–486.
- Nunez, C., Esteve-Núñez, A., Giometti, C., Tollaksen, S., Khare, T., Lin, W., Lovley, D. R. & Methé, B. A. (2006). DNA microarray and proteomic analyses of the RpoS regulon in *Geobacter sulfurreducens*. *J Bacteriol* 188, 2792–2800.
- O'Farrell, P. H. (1975). High resolution two-dimensional electrophoresis of proteins. *J Biol Chem* 250, 4007–4021.
- Oliver, S. G., Winson, M. K., Kell, D. B. & Baganz, F. (1998). Systematic functional analysis of the yeast genome. *Trends Biotechnol* 16, 373–378.
- Ou, K., Ong, C., Koh, S. Y., Rodrigues, F., Sim, S. H., Wong, D., Ooi, C. H., Ng, K. C., Jikuya, H. & other authors (2005). Integrative

genomic, transcriptional, and proteomic diversity in natural isolates of the human pathogen *Burkholderia pseudomallei*. *J Bacteriol* **187**, 4276–4285.

Ozsolak, F., Platt, A. R., Jones, D. R., Reifengerger, J. G., Sass, L. E., McInerney, P., Thompson, J. F., Bowers, J., Jarosz, M. & Milos, P. M. (2009). Direct RNA sequencing. *Nature* **461**, 814–818.

Park, S. J., Lee, S. Y., Cho, J., Kim, T. Y., Lee, J. W., Park, J. H. & Han, M. J. (2005). Global physiological understanding and metabolic engineering of microorganisms based on omics studies. *Appl Microbiol Biotechnol* **68**, 567–579.

Parrish, J. R., Yu, J., Liu, G., Hines, J. A., Chan, J. E., Mangiola, B. A., Zhang, H., Pacifico, S., Fotouhi, F. & other authors (2007). A proteome-wide protein interaction map for *Campylobacter jejuni*. *Genome Biol* **8**, R130.

Pir, P., Kirdar, B., Hayes, A., Önsan, Z., Ülgen, K. & Oliver, S. (2006). Integrative investigation of metabolic and transcriptomic data. *BMC Bioinformatics* **7**, 203.

Raamsdonk, L. M., Teusink, B., Broadhurst, D., Zhang, N., Hayes, A., Walsh, M. C., Berdon, J. A., Brindle, K. M., Kell, D. B. & other authors (2001). A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. *Nat Biotechnol* **19**, 45–50.

Ren, B., Robert, F., Wyrick, J. J., Aparicio, O., Jennings, E. G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N. & other authors (2000). Genome-wide location and function of DNA-binding proteins. *Science* **290**, 2306–2309.

Rocha, E. P. (2008). The organization of the bacterial genome. *Annu Rev Genet* **42**, 211–223.

Rodriguez-Llorente, I., Caviedes, M. A., Dary, M., Palomares, A. J., Cánovas, F. M. & Peregrín-Alvarez, J. M. (2009). The Symbiosis Interactome: a computational approach reveals novel components, functional interactions and modules in *Sinorhizobium meliloti*. *BMC Syst Biol* **3**, 63.

Scherl, A., François, P., Bento, M., Deshusses, J. M., Charbonnier, Y., Converset, V., Huyghe, A., Walter, N., Hoogland, C. & other authors (2005). Correlation of proteomic and transcriptomic profiles of *Staphylococcus aureus* during the post-exponential phase of growth. *J Microbiol Methods* **60**, 247–257.

Scherl, A., François, P., Charbonnier, Y., Deshusses, J. M., Koessler, T., Huyghe, A., Bento, M., Stahl-Zeng, J., Fischer, A. & other authors (2006). Exploring glycopeptide-resistance in *Staphylococcus aureus*: a combined proteomics and transcriptomics approach for the identification of resistance-related markers. *BMC Genomics* **7**, 296.

Schilling, C. H., Edwards, J. S. & Palsson, B. O. (1999). Toward metabolic phenomics: analysis of genomic data using flux balances. *Biotechnol Prog* **15**, 288–295.

Schmid, A. K., Reiss, D. J., Kaur, A., Pan, M., King, N., Van, P. T., Hohmann, L., Martin, D. B. & Baliga, N. S. (2007). The anatomy of microbial cell state transitions in response to oxygen. *Genome Res* **17**, 1399–1413.

Schoolnik, G. K. (2001). The accelerating convergence of genomics and microbiology. *Genome Biol* **2**, REPORTS4009.

Sharan, R. & Ideker, T. (2006). Modeling cellular machinery through biological network comparison. *Nat Biotechnol* **24**, 427–433.

Singh, O. V. & Nagaraj, N. S. (2006). Transcriptomics, proteomics and interactomics: unique approaches to track the insights of bioremediation. *Brief Funct Genomic Proteomic* **4**, 355–362.

Soga, T., Ohashi, Y., Ueno, Y., Naraoka, H., Tomita, M. & Nishioka, T. (2003). Quantitative metabolome analysis using capillary electrophoresis mass spectrometry. *J Proteome Res* **2**, 488–494.

Sonck, K. A., Kint, G., Schoofs, G., Vander Wauven, C., Vanderleyden, J. & De Keersmaecker, S. C. (2009). The proteome

of *Salmonella typhimurium* grown under in vivo-mimicking conditions. *Proteomics* **9**, 565–579.

Stedtfeld, R. D., Baushke, S. W., Turlousse, D. M., Miller, S. M., Stedtfeld, T. M., Gulari, E., Tiedje, J. M. & Hashsham, S. A. (2008). Development and experimental validation of a predictive threshold cycle equation for quantification of virulence and marker genes by high-throughput nanoliter-volume PCR on the OpenArray platform. *Appl Environ Microbiol* **74**, 3831–3838.

Steinfath, M., Reipsilber, D., Scholz, M., Walther, D. & Selbig, J. (2007). Integrated data analysis for genome-wide research. *EXS* **97**, 309–329.

Suzuki, I., Simon, W. J. & Slabas, A. R. (2006). The heat shock response of *Synechocystis* sp. PCC 6803 analysed by transcriptomics and proteomics. *J Exp Bot* **57**, 1573–1578.

Tang, Y. J., Martin, H. G., Myers, S., Rodriguez, S., Baidoo, E. E. & Keasling, J. D. (2009a). Advances in analysis of microbial metabolic fluxes via <sup>13</sup>C isotopic labeling. *Mass Spectrom Rev* **28**, 362–375.

Tang, Y. J., Martin, H. G., Dehal, P. S., Deutschbauer, A., Llorca, X., Meadows, A., Arkin, A. & Keasling, J. D. (2009b). Metabolic flux analysis of *Shewanella* spp. reveals evolutionary robustness in central carbon metabolism. *Biotechnol Bioeng* **102**, 1161–1169.

Ter Kuile, B. H. & Westerhoff, H. V. (2001). Transcriptome meets metabolome: hierarchical and metabolic regulation of the glycolytic pathway. *FEBS Lett* **500**, 169–171.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J Roy Statist Soc Ser B Methodol* **58**, 267–288.

Tong, A. H., Lesage, G., Bader, G. D., Ding, H., Xu, H., Xin, X., Young, J., Berriz, G. F., Brost, R. L. & other authors (2004). Global mapping of the yeast genetic interaction network. *Science* **303**, 808–813.

Torres-Garcia, W., Zhang, W., Runger, G. C., Johnson, R. H. & Meldrum, D. R. (2009). Integrative analysis of transcriptomic and proteomic data of *Desulfovibrio vulgaris*: a non-linear model to predict abundance of undetected proteins. *Bioinformatics* **25**, 1905–1914.

Toya, Y., Ishii, N., Hirasawa, T., Naba, M., Hirai, K., Sugawara, K., Igarashi, S., Shimizu, K., Tomita, M. & Soga, T. (2007). Direct measurement of isotopomer of intracellular metabolites using capillary electrophoresis time-of-flight mass spectrometry for efficient metabolic flux analysis. *J Chromatogr A* **1159**, 134–141.

Trauger, S. A., Kalisak, E., Kalisiak, J., Morita, H., Weinberg, M. V., Menon, A. L., Poole, F. L., II, Adams, M. W. & Siuzdak, G. (2008). Correlating the transcriptome, proteome, and metabolome in the environmental adaptation of a hyperthermophile. *J Proteome Res* **7**, 1027–1035.

Tweeddale, H., Notley-McRobb, L. & Ferenci, T. (1998). Effect of slow growth on metabolism of *Escherichia coli*, as revealed by global metabolite pool (metabolome) analysis. *J Bacteriol* **180**, 5109–5116.

Tweeddale, H., Notley-McRobb, L. & Ferenci, T. (1999). Assessing the effect of reactive oxygen species on *Escherichia coli* using a metabolome approach. *Redox Rep* **4**, 237–241.

Typas, A., Nichols, R. J., Siegle, D. A., Shales, M., Collins, S. R., Lim, B., Braberg, H., Yamamoto, N., Takeuchi, R. & other authors (2008). High-throughput, quantitative analyses of genetic interactions in *E. coli*. *Nat Methods* **5**, 781–787.

Urbanczyk-Wochniak, E., Luedemann, A., Kopka, J., Selbig, J., Roessner-Tunali, U., Willmitzer, L. & Fernie, A. R. (2003). Parallel analysis of transcript and metabolic profiles: a new approach in systems biology. *EMBO Rep* **4**, 989–993.

Urbanczyk-Wochniak, E., Willmitzer, L. & Fernie, A. R. (2007). Integrating profiling data: using linear correlation to reveal coregulation of transcript and metabolites. *Methods Mol Biol* **358**, 77–85.

Uyar, E., Kurokawa, K., Yoshimura, M., Ishikawa, S., Ogasawara, N. & Oshima, T. (2009). Differential binding profiles of StpA in wild-

type and *hns* mutant cells: a comparative analysis of cooperative partners by chromatin immunoprecipitation-microarray analysis. *J Bacteriol* **191**, 2388–2391.

van der Werf, M. J., Jellema, R. H. & Hankemeier, T. (2005). Microbial metabolomics: replacing trial-and-error by the unbiased selection and ranking of targets. *J Ind Microbiol Biotechnol* **32**, 234–252.

Wan, X. F., Verberkmoes, N. C., McCue, L. A., Stanek, D., Connelly, H., Hauser, L. J., Wu, L., Liu, X., Yan, T. & other authors (2004). Transcriptomic and proteomic characterization of the Fur regulon in the metal-reducing bacterium *Shewanella oneidensis*. *J Bacteriol* **186**, 8385–8400.

Ward, N. & Fraser, C. M. (2005). How genomics has affected the concept of microbiology. *Curr Opin Microbiol* **8**, 564–571.

Washburn, M. P., Koller, A., Oshiro, G., Ulaszek, G., Plouffe, D., Deciu, C., Winzler, E. & Yates, J. R., III (2003). Protein pathway and complex clustering of correlated mRNA and protein expression analyses in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A* **100**, 3107–3112.

Waters, K. M., Pounds, J. G. & Thrall, B. D. (2006). Data merging for integrated microarray and proteomic analysis. *Brief Funct Genomic Proteomic* **5**, 261–272.

Watson, M. (2005). ProGenExpress: visualization of quantitative data on prokaryotic genomes. *BMC Bioinformatics* **6**, 98.

Webb-Robertson, B. J., McCue, L. A., Beagley, N., McDermott, J. E., Wunschel, D. S., Varnum, S. M., Hu, J. Z., Isern, N. G., Buchko, G. W. & other authors (2009). A Bayesian integration model of high-throughput proteomics and metabolomics data for improved early detection of microbial infections. *Pac Symp Biocomput* 451–463.

Weckwerth, W. & Morgenthal, K. (2005). Metabolomics: from pattern recognition to biological interpretation. *Drug Discov Today* **10**, 1551–1558.

Weckwerth, W., Wenzel, K. & Fiehn, O. (2004). Process for the integrated extraction, identification and quantification of metabolites, proteins and RNA to reveal their co-regulation in biochemical networks. *Proteomics* **4**, 78–83.

Wienkoop, S., Morgenthal, K., Wolschin, F., Scholz, M., Selbig, J. & Weckwerth, W. (2008). Integration of metabolomic and proteomic phenotypes. *Mol Cell Proteomics* **7**, 1725–1736.

Wilkins, M. R., Pasquali, C., Appel, R. D., Ou, K., Golaz, O., Sanchez, J., Yan, J. X., Gooley, A. A., Hughes, G. & other authors (1996). From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Biotechnology (N Y)* **14**, 61–65.

Wolff, S., Otto, A., Albrecht, D., Zeng, J. S., Büttner, K., Glückmann, M., Hecker, M. & Becher, D. (2006). Gel-free and gel-based proteomics in *Bacillus subtilis*: a comparative study. *Mol Cell Proteomics* **5**, 1183–1192.

Yan, W., Hwang, D. & Aebersold, R. (2008). Quantitative proteomic analysis to profile dynamic changes in the spatial distribution of cellular proteins. *Methods Mol Biol* **432**, 389–401.

Yang, S., Tschaplinski, T. J., Engle, N. L., Carroll, S. L., Martin, S. L., Davison, B. H., Palumbo, A. V., Rodriguez, M., Jr & Brown, S. D. (2009). Transcriptomic and metabolomic profiling of *Zymomonas mobilis* during aerobic and anaerobic fermentations. *BMC Genomics* **10**, 34.

Ye, R. W., Wang, T., Bedzyk, L. & Croker, K. M. (2001). Applications of DNA microarrays in microbial systems. *J Microbiol Methods* **47**, 257–272.

Yoon, S. H., Han, M. J., Lee, S. Y., Jeong, K. J. & Yoo, J. S. (2003). Combined transcriptome and proteome analysis of *Escherichia coli* during high cell density culture. *Biotechnol Bioeng* **81**, 753–767.

Yu, E. Z., Burba, A. E. & Gerstein, M. (2007). PARE: a tool for comparing protein abundance and mRNA expression data. *BMC Bioinformatics* **8**, 309.

Zhang, W., Gritsenko, M. A., Moore, R. J., Culley, D. E., Nie, L., Petritis, K., Strittmatter, E. F., Camp, D. G., II, Smith, R. D. & Brockman, F. J. (2006a). A proteomic view of *Desulfovibrio vulgaris* metabolism as determined by liquid chromatography coupled with tandem mass spectrometry. *Proteomics* **6**, 4286–4299.

Zhang, W., Culley, D. E., Scholten, J. C., Hogan, M., Vitiritti, L. & Brockman, F. J. (2006b). Global transcriptomic analysis of *Desulfovibrio vulgaris* on different electron donors. *Antonie Van Leeuwenhoek* **89**, 221–237.

Zhao, B. & Poh, C. L. (2008). Insights into environmental bioremediation by microorganisms through functional genomics and proteomics. *Proteomics* **8**, 874–881.