ORIGINAL ARTICLE

MSEA: metabolite set enrichment analysis in the MeltDB metabolomics software platform: metabolic profiling of *Corynebacterium glutamicum* as an example

Marcus Persicke · Christian Rückert · Jens Plassmeier · Leonhardt Jonathan Stutz · Nikolas Kessler · Jörn Kalinowski · Alexander Goesmann · Heiko Neuweger

Abstract Gene set enrichment analysis (GSEA) has been successfully employed in transcriptomics and proteomics research for over 5 years. We have applied a modified GSEA approach to metabolomics, called metabolite set enrichment analysis (MSEA), and have integrated this method into the MeltDB platform. With the novel integrated functionality, we evaluate the applicability of the approach for metabolomics research by analyzing the metabolic profiles of industrial amino acid producer strains of *Corynebacterium glutamicum*. Metabolite sets are obtained from metabolic pathways defined in the KEGG and the newly created CglCyc databases. In the first experiment using MSEA, the metabolic profiling analyses compared glucose- and acetate-grown strains, and, in the second, a production strain series in which the carbon flow is shifted step-wise from the lysine pathway to the branching threonine, isoleucine, and methionine pathways. By identifying changes of the metabolic profile MSEA was able to identify the metabolic pathways activated while utilizing different carbon sources or those that have been genetically modified. The presented experimental results are publicly available at http://meltdb.cebitec.uni-bielefeld.de.

M. Persicke (✉) · C. Rückert · J. Plassmeier · J. Kalinowski
Institut für Genomforschung und Systembiologie, Centrum für Biotechnologie (CeBiTec), University of Bielefeld, Universitätsstrasse 27, 33615 Bielefeld, Germany
e-mail: MarcusP@CeBiTec.Uni-Bielefeld.DE

Present Address:
J. Plassmeier
Department of Biology, Massachusetts Institute of Technology, 31 Ames Street, Bldg. 68-346, Cambridge, MA 02142, USA

L. J. Stutz · N. Kessler · A. Goesmann · H. Neuweger
Computational Genomics, Centrum für Biotechnologie (CeBiTec), University of Bielefeld, Universitätsstrasse 27, 33615 Bielefeld, Germany

N. Kessler
Biodata Mining & Applied Neuroinformatics, Centrum für Biotechnologie (CeBiTec), University of Bielefeld, Universitätsstrasse 27, 33615 Bielefeld, Germany

1 Introduction

Metabolomics can be regarded as the most recent contribution to system-wide studies. It involves the analysis of the qualitative and quantitative collection of virtually all metabolites in the cell (the metabolome). As metabolomics is closely connected to the actual phenotype of an organism, it adds to the understanding of cellular systems. The results from metabolomics experiments can be connected to the genotype through biochemical pathways and gene regulatory networks (Fiehn 2002). A representative biochemical phenotype of an organism can be achieved through large-scale quantitative and qualitative measurements of sizeable numbers of cellular metabolites. This phenotypic information can be used to monitor and to assess the response of the biological system and the function of specific genes. In addition, metabolomics studies of mutants lacking or overexpressing an enzyme of unknown function can provide information on the specific biochemical pathway the enzyme is involved in. Thus, profiling the metabolome can often provide the most

conclusive and functional information of the "omics"-technologies. Coupled mass spectrometry instruments are currently one of the most widely applied technologies in metabolomics, as they provide rapid, sensitive, and selective qualitative and quantitative analyses (Dunn and Ellis 2005). Estimations for the number of metabolites present in living organisms range from hundreds to many thousands; for the model organism *Escherichia coli* = 750 metabolites have been ascertained (Nobeli et al. 2003). With the application of modern coupled technologies, a significant part of the metabolome (Hall 2006) can be analyzed. Yet, the extraction and analysis of entire metabolomes in a single step remains undoable due in part to the chemical complexity and heterogeneity of compounds (Goodacre et al. 2004). The number of estimated compounds for eukaryotic cells ranges from 4,000 to 20,000 (Fernie et al. 2004). In the entire plant kingdom, up to 200,000 metabolites are expected to exist. Therefore, truly comprehensive metabolomics studies will require new technical inventions (Hall 2006). Nonetheless, with current individual analytical techniques or the combination of various platforms, hundreds of metabolites can be effectively identified and relatively quantified.

While far from complete, the amount of data produced for the identifiable metabolites already presents interpretation challenges for identifying interesting changes. For other high-throughput techniques like transcriptomics and proteomics, gene set enrichment analysis (GSEA) has become a standard tool. A wide variety of applications implement the functionality that was first described by Mootha et al. (2003) and was improved upon by Subramanian et al. (2005). GSEA represents a robust technique for analyzing molecular profiling data and is used for analyzing gene expression by using pathway or ontology information. Gene sets are defined based on prior biological knowledge such as published information about biochemical pathways or co-expression in previous microarray experiments.

The central concept of GSEA has been the determination if the genes in a given set are enriched among the genes that are most differentially transcribed between two classes. Within GSEA, the genes are ordered based on a difference metric. Typically the SNR difference metric is used, which is simply the difference of the averages of the two classes divided by the sum of the standard deviations of the two diagnostic classes. In general, other difference metrics can be used such as m-values or probability values computed by simple *t*-tests. GSEA has been applied to a wide range of research questions with various software tools such as GSEA-P, J-Express, or Meta-GP that support the GSEA analysis for transcriptomics and proteomics datasets (Stavrum et al. 2008; Subramanian et al. 2007). A comprehensive overview of the large list of existing gene set enrichment tools has been detailed by Huang et al. (2009).

Nonetheless, and as noted earlier, it could be beneficial to evaluate how the GSEA approach performs when applied to the analysis of metabolomics data (Rubin 2006). Slight differences between the gene-centric and the metabolite-based profiling exist. For example, the range of concentration changes are magnitudes higher for metabolites compared to transcript levels or protein concentrations (van den Berg et al. 2006). One major challenge, however, is the fact that the number of identifiable metabolites that can be associated with metabolic pathways is much lower than the number of mRNAs of genes that can be measured via comprehensive microarray or cDNA sequencing methods, raising the need to evaluate the general applicability of the GSEA approach in the metabolomics context. A first implementation of the MSEA approach has been provided recently (Xia and Wishart 2010) focusing on metabolic pathways relevant to human metabolism. The analysis of metabolic profiles in the context of metabolic pathways has recently been detailed in the PAPi approach by Aggio et al. (2010). The algorithm generates Activity Scores for pathways present in the KEGG database based on metabolic profiles from different experimental conditions. The approach was evaluated on published experimental data from *Saccharomyces cerevisiae*. In our study we test the applicability of metabolite set enrichment analysis for metabolic profiles obtained from the amino acid producer *Corynebacterium glutamicum* and present adaptations of the original GSEA approach.

1.1 The MeltDB software platform

The automation of sample acquisition and subsequent high-throughput analysis is posing a challenge for the necessary systematic storage and computational processing of the experimental datasets. Whereas a multitude of specialized software systems for individual instruments and pre-processing methods exist, there is clearly a need for a free and platform-independent system that allows standardized integrated storage and analysis of data obtained from metabolomics experiments. We have implemented the platform independent MeltDB system (Neuweger et al. 2008) to address this need. MeltDB provides the storage, organization, and annotation of datasets generated in metabolomics experiments. The system offers functionality for the pre-processing of mass spectrometry datasets in netCDF, mzXML, and mzData file formats. The results of the pre-processing are able to be visualized and integrated within a functional genomics context and access to higher level statistical analysis is provided via the MeltDB web interface. For the evaluation and analysis of MSEA, we test the integration of the approach within the MeltDB software platform.

1.2 The amino acid producer *Corynebacterium glutamicum*

The Gram-positive soil bacterium *Corynebacterium glutamicum* is widely used for the production of industrially interesting amino acids. L-glutamate (1.5 million tons) and L-lysine (850,000 tons) are the major products produced and the amino acid market is growing at an annual rate of 7% (Hermann 2003; Leuchtenberger et al. 2005). Since the determination of the complete genome sequence of the *C. glutamicum* wild-type strain ATCC 13032 (Kalinowski et al. 2003) rational strain improvement is replacing the classical mutational approaches. In addition, knowledge of the genome sequence led to the development of genome-wide high-throughput technologies, e.g., transcriptomics (Hüser et al. 2003; Wendisch et al. 2006), proteomics (Burkovski 2006; Hansmeier et al. 2006), metabolomics (Krömer et al. 2005; Plassmeier et al. 2007), and fluxomics (Drysch et al. 2003; Hoon Yang et al. 2006). All these techniques have been applied to study and optimize several production processes, chief among them those for aspartate-derived amino acids like lysine or methionine.

The production of lysine with *C. glutamicum* has become a field of extensive research in the last decades. In 2002, Onishi and co-workers presented a study where the exchange of a single amino acid in three genes of *C. glutamicum* ATCC 13032, namely the aspartokinase (*lysC*), the pyruvate decarboxylase (*pyc*), and the homoserine dehydrogenase (*hom*), led to a tremendous increase in lysine production (Ohnishi et al. 2002). The *pyc* mutation led to an enhanced flux from pyruvate to oxaloacetate, mutation of *lysC* generated a feedback deregulation, and the mutation of *hom* caused a leaky mutation, redirecting the flux of the lysine precursor aspartate-ß-semialdehyde into the lysine pathway. Strains carrying these mutations have become the "root" strains for further rational design of lysine producers. Recent studies in metabolomics and fluxomics revealed new insights in identifying "bottlenecks" in the lysine production, e.g., decreased TCA cycle activity lead to increased lysine formation or low biomass formation enhances diaminopimelate supply as a precursor (Becker et al. 2008, 2009; Drysch et al. 2003, 2004; El Massaoudi et al. 2003; Hoon Yang et al. 2006; Krömer et al. 2004; Neuweger et al. 2009).

The amino acid methionine is used in large amounts for animal nutrition, since their occurrence in plant material is very low. Until now methionine has been produced by a chemical process, producing a racemic mixture of D,L-methionine and employing rather hazardous chemicals (Leuchtenberger 1996). Since methionine, like lysine, is a member of the aspartic amino acid group and lysine is produced in large amounts in *C. glutamicum*, this strain is considered to be a good candidate for fermentative production. Toward this end, several studies elucidating the biosynthetic pathway of methionine in *C. glutamicum* have been performed (Hwang et al. 1999, 2002; Park et al. 1998; Rey et al. 2003, 2005; Rückert et al. 2003, 2005). In *C. glutamicum*, methionine synthesis begins at the branching point to lysine with homoserine dehydrogenase (*hom*) converting aspartate-ß-semialdehyde to homoserine. Homoserine dehydrogenase is feedback-inhibited by threonine (Morbach et al. 1996) and the *hom* gene is repressed by the master regulator of sulfur metabolism (Rey et al. 2003). Park and co-workers engineered a methionine producing strain derived from the lysine producer MH20-22B, yielding 2.9 g l^{-1} methionine (Park et al. 2007). The strain carries a deregulated homoserine dehydrogenase (hom^{FBR}) to abolish threonine-mediated feedback inhibition and a deletion of the homoserine kinase (*thrB*), eliminating flux of the precursor homoserine into the threonine pathway.

Besides the genetic modifications of *C. glutamicum* for the development of productions strains, the influence of different carbon sources on the metabolic behaviour was also of interest for strain optimisation and media design (Blombach and Seibold 2010; Kawaguchi et al. 2009; Kiefer et al. 2002; Muffler et al. 2002; Seibold et al. 2009; Wittmann et al. 2004). The analysis of the utilisation of glucose compared to acetate revealed clearly distinguishable metabolic profiles and different fluxes. The main differences were the higher fluxes in glycolysis and pentose phosphate-cycle (PPP) when glucose is used as carbon source, compared to the high fluxes in TCA cycle and glyoxylate shunt in an acetate-grown culture (Wendisch et al. 2000).

In this study we present the metabolic set enrichment analysis as a feature of the MeltDB software platform for identifying key pathways from metabolomics data sets in order to facilitate metabolome analysis. We evaluate the tool by analyzing the data published by Plassmeier and co-workers (2007), dealing with the identification of the 2-methylcitrate cycle in *C. glutamicum* ATCC 13032. For this experiment, *C. glutamicum* was fed with glucose, acetate, and a mixture of acetate and propionate. Furthermore, we present the metabolic profilng analysis of a series of *C. glutamicum* strains engineered towards a flux redirection from the lysine pathway to the branching threonine, isoleucine, and methionine pathway. This was done by analyzing the impact of three different alleles of the homoserine dehydrogenase with the purpose of gradually enhancing the flux into the methionine pathway. MSEA was used to identify the pathways impacted the most by these allele exchanges thus demonstrating its usefulness.

2 Materials and methods

2.1 Bacterial strains

The bacteria used in this study include several strains of *Corynebacterium glutamicum* ATCC 13032 (WT), the derivative lysine production strain *C. glutamicum* DM1730 carrying the mutations hom^{V59A} pyc^{P458S} $lysC^{T311I}$ Δpck (Seibold et al. 2006), and two derivatives of DM1730. The first, MP001, carries the wild type *hom* gene. The second strain, DM1795, has the hom^{V59A} gene replaced by the feedback deregulated hom^{FBR} allele. This strain was kindly provided by Evonik-Degussa AG (Künsebeck, Germany).

2.2 Media and growth conditions

In the glucose-acetate experiment (Plassmeier et al. 2007), Minimal medium MM1 (Tauch et al. 2002) was used for growth of *C. glutamicum* at a temperature of 30°C, with 4 g l^{-1} glucose or 4 g l^{-1} of sodium acetate as carbon source. All cultures for metabolite analysis were cultivated as triplicates in shaking flasks and harvested at OD_{600} of 3.

Allelic exchange of the *hom* gene: CGXII minimal medium was prepared according to Keilhauer et al. (1993), additionally containing 30 mg l^{-1} protocatechuic acid. *C. glutamicum* was cultivated in shaking flasks at a temperature of 30°C. For each strain, eight parallel cultivations were performed and harvested for metabolome analysis at OD_{600} of 10.

2.3 Chemicals

All chemicals and standard compounds were purchased from either Sigma–Aldrich–Fluka (Taufkirchen, Germany), Merck (Darmstadt, Germany), Roth (Karlsruhe, Germany), or Macherey–Nagel (Düren, Germany).

2.4 Harvesting and sample preparation of *C. glutamicum* cells for metabolome analysis

Cell harvesting by centrifugation was performed using the protocol of Plassmeier et al. (2007). Two milliliters of an exponentially growing bacterial culture were harvested by centrifugation in 2 ml screw cap vials at 16,100×*g* for 15 s at RT. After pelleting, the supernatant was removed quantitatively and the vials were immediately frozen in liquid nitrogen. The time until inactivation of the metabolism takes place was around 30 s. After freeze-drying, the biomass was disrupted using a Precellys homogenisator (Peqlab, Erlangen, Germany) and the hydrophilic metabolites were extracted with 80% methanol containing 10 μM ribitol (internal standard). Derivatization and GC–MS measurements were conducted as described previously (Plassmeier et al. 2007).

2.5 Metabolite analysis

The measurement of the metabolites was performed by injecting 1 μl of the derivatized metabolite sample into the GC–MS system consisting of a TraceGC gas chromatograph, a PolarisQ ion trap, and an AS1000 autosampler (Thermo Finnigan, Dreieich, Germany). For separation of the metabolites, a 30 m × 0.25 mm × 0.25 μm Equity-5 column (Suppelco) was used. The resulting chromatograms were converted to netCDF format and imported into the MeltDB system. Annotation of the chromatograms was done in accordance with the recommendations of the MSI and the chromatograms were organized in replicate groups according to their strain or the fed carbon source in Melt-DB. Peak detection, identification, and quantification were performed using the Xcalibur^{TM} software (Version 1.4, Thermo Finnigan, Dreieich, Germany). Subsequently, a processing setup defined and routinely employed at Bielefeld University, containing retention time windows of representative ions of 80 compounds, was applied (Plassmeier et al. 2007). The Xcalibur importer tool of MeltDB was used to transfer these results into the MeltDB data model and link the identified metabolites to the KEGG compound database. Therefore, the generic importer has been configured by a user defined mapping from library identifiers to KEGG compound IDs.

2.6 Metabolite set enrichment analysis

The MSEA extensions were integrated into the software framework of the MeltDB platform. MSEA can be applied to experiments with metabolic profiles originating from two different classes represented by replicate measurement groups in the MeltDB database. The metabolites present in both classes were ranked using a distance metric that compares the metabolic pool sizes that have been normalized in MeltDB towards the dry weight of the metabolite sample and an internal standard (Ribitol). As one possible metric is the difference of the log of the means of the metabolite pool sizes (m-values) which we have previously integrated in MeltDB for the ProMeTra application (Neuweger et al. 2009). In addition, we implemented the Signal-To-Noise criterion and the *t*-test significance metric. All metrics generate an ordered list *L* of all present metabolites. Given an a priori defined set of metabolites *MS* (e.g., metabolites present in a metabolic pathway), the goal of MSEA is to determine whether the members of *MS* are randomly distributed throughout *L* or predominantly found at the top or bottom. For the computation of the enrichment score (ES), we employed the method described

by Subramanian et al. (2005). An interesting extension in the context of metabolite research is the ordering of the m-values by the absolute values. When knock-out mutations of enzymes are analyzed, metabolite pool sizes will often increase and decrease in the same metabolic pathway. The rationale is that direct products of the enzyme decrease while substrates increase. To identify these effects, the ES function used is based on the absolute fold change. In addition to the computation of an ES, the computation of the p-value of that score for the experimental data was performed by the MeltDB implementation.

2.7 Extracting metabolite sets from the KEGG pathway database

The use of the KEGG compounds database provides a controlled vocabulary for the identification of metabolites (Kanehisa et al. 2006). Through the association of compounds detected in the GC– and LC–MS measurements to this controlled vocabulary, an unambiguous mapping to metabolic pathways present in the KEGG database is possible. The existing integration of the KEGG compounds database in the MeltDB system simplifies the generation of metabolite sets from the KEGG pathway database. Every metabolic pathway in KEGG is represented by the set of its associated metabolites and each pathway corresponds to a metabolite set (MS). Through the use of the KEGG pathway database, more than 200 predefined metabolite sets are readily available for the MSEA. It is notable that metabolites occurring more than once in a metabolic pathway, such as co-factors, are represented in a metabolite set only once.

2.8 CglCyc: reconstruction of metabolic pathways for *C. glutamicum*

To complement the metabolite sets obtained from the KEGG pathway database, we reconstructed the metabolic pathways of *C. glutamicum* ATCC 13032 using Pathway-Tools (Karp et al. 2010). We manually curated and extended the results in a locally installed CglCyc instance. For example, we adapted the metabolites of the pathway "methionine biosynthesis I". This pathway provided a metabolite set beginning with homoserine and ending with methionine. The original pathway used the homoserine *O*-succinyltransferase to convert homoserine to *O*-succinyl-homoserine. This enzyme is present in *Escherichia coli* (Born and Blanchard 1999), but *C. glutamicum* uses the homoserine *O*-acetyltransferase converting homoserine to *O*-acetyl-homoserine (Rückert et al. 2003). Therefore, the pathway "methionine biosynthesis I" was corrected by

adding the correct metabolite identifiers. As a second example, we renamed the pathway "respiration (anaerobic)", which is not present in *C. glutamicum*, to "TCA cycle and gluconeogenesis (upper part)/glycolysis (lower part)". By renaming this metabolic set, we fixed the occurrence of a non existing pathway in the CglCyc database and extended the metabolic pathways to a specific smaller part representing conditions under gluconeogenesis (Wendisch et al. 2000). Based on the reconstructed and curated metabolic pathways, 110 metabolite sets were defined that we integrated into the MSEA of MeltDB, harboring at minimal one of the possible detectable metabolites via GC–MS (Electronic Supplementary Material).

3 Results and discussion

We implemented the MSEA approach in MeltDB via a web based user interface. The interface for the MSEA functionality in MeltDB allows us to directly query the list of all metabolite sets obtained from the integrated KEGG database and the manually created CglCyc instance (Electronic Supplementary Material) initially obtained from the PathwayTools software. Users may specify which experimental conditions are to be compared and if the enrichment score for individual pathways or the complete list of metabolite pathways should be computed (Fig. 1). In the latter case, the best scoring metabolic pathway is presented to the user. In addition, the ordered list of metabolites identified in the experiment is given. Interactive access to the visualizations of both the *ES* score and the metabolite list via the user interface simplifies the analysis of the MSEA results. Further information such as the name and identifier of identifiable metabolites, the computed m-value, and a *t*-test probability is presented to the researcher.

The applicability of the method is presented in more detail with the analysis of metabolic profiling experiments conducted using amino acid production strains of *C. glutamicum*. For comparison of the MSEA results, we chose the presentation of the best five pathway hits, enabling a clear discrimination between the analyzed strains. Since metabolites are often present in more than one metabolite set, MSEA results with less than four matches to the identified metabolites could lead to false positive pathway identifications. Therefore, we excluded these hits from the analysis. MSEA was run with the m-value score computed from the replicate measurements of the wild type and mutant strains. The significance of the metabolites set analysis was tested by permuting the scores over the metabolites using 1,000 iterations each.

Fig. 1 The user interface for the MeltDB metabolite set enrichment analysis. After reference and query replicate groups are selected, the enrichment score for individual metabolite sets obtained from the KEGG pathway and/or CglCyc database can be computed. Alternatively, the best matching metabolite sets from all available pathways can be computed or the user may specify own metabolite sets using KEGG compound identifiers. In the default case, the method described by Subramanian et al. (2005) is used for ES score computation, alternatively the simple Kolmogorov–Smirnov statistic may be selected

3.1 Evaluation of MSEA by comparison of the metabolic profiles of *C. glutamicum* cultivated in minimal medium containing glucose or acetate as carbon source

The first evaluation of the metabolic set enrichment analysis was done on the published data of Plassmeier et al. (2007), dealing with cultivation of *C. glutamicum* in minimal medium with glucose or acetate as carbon source. The cultivations were performed in triplicates for each strain and metabolome analyses were performed using GC–MS. Cell harvesting was done by fast centrifugation, resulting in an inactivation time of the metabolism within 30 s. It is known that metabolites of the central metabolism have high turn over rates (~ 1 mmol s^{-1}; de Koning and van Dam 1992; Nöh et al. 2007; Schäfer et al. 1999) leading to a potentially wrong estimation of the metabolic state of the cells during growth if the inactivation time of the metabolism is long. Persicke and co-workers showed that the comparison of pool size ratios without total quantification of the metabolites was in the same range using harvesting methods that were fast or slower inactivating the metabolism (Persicke et al. 2010). MSEA were done using the m-values, calculated from the comparison of the acetate grown culture against the glucose grown culture. We chose this comparison, since a plethora of information is available on the influence of these carbon sources on the metabolism of this bacterium (Auchter et al. 2009; Gerstmeir et al. 2003; Han et al. 2008; Muffler et al. 2002; Plassmeier et al. 2007; Veit et al. 2009; Wendisch et al. 2000). The main difference between these two carbon sources is their entry point into metabolism. Glucose is taken up via the PEP-dependent sugar:phosphotransferase (PTS) system (Parche et al. 2001) and then metabolized via glycolysis and the tricarboxylic acid (TCA) cycle. Acetate mainly diffuses into the cells and is converted there to acetyl-CoA which is then used to fill the TCA cycle (Gerstmeir et al. 2003). During acetate consumption, gluconeogenesis must be active in order to fulfill the cellular requirements for sugar nucleotides. These principal differences are reflected by the results obtained via MSEA: the best five hits are the CglCyc pathways of the *pantothenate and coenzyme A biosynthesis*, *pentose phosphate pathway*, *pentose phosphate pathway* (*non-oxidative branch*), *pentose phosphate pathway* (*partial*) and *TCA cycle and gluconeogenesis* (*upper part*)/*glycolysis* (*lower part*) (Fig. 2a). The best pathway hit reflects the different utilization of acetate in comparison to glucose. As mentioned above, intracellular acetate is converted into acetyl-CoA to fulfill the TCA cycle. Coenzyme A is produced in *C. glutamicum* by the formation of pantothenate as precursor (Hüser et al. 2005). When utilizing acetate as sole carbon source, gluconeogenesis is performed, resulting in reduced fluxes in the pentose phosphate pathway (Wendisch et al. 2000) which is in accordance with the smaller pool sizes of the metabolites of the upper part of the glycolysis and the pentose phosphate pathway detected via MSEA. Conversely, as acetate consumption activates the glyoxylate shunt and leads to a higher flux in the TCA cycle (Gerstmeir et al. 2003; Hayashi et al. 2002; Muffler et al. 2002; Wendisch et al. 2000), elevated pool sizes of TCA cycle intermediates are found corresponding to the pathway hit *TCA cycle and gluconeogenesis* (*upper part*).

When comparing the MSEA results performed with the CglCyc database and the KEGG database, two pathway hits were found in both databases, *pentose phosphate pathway* at first position and the *TCA cycle* at fourth position (Fig. 2a/c). Using the KEGG database, no
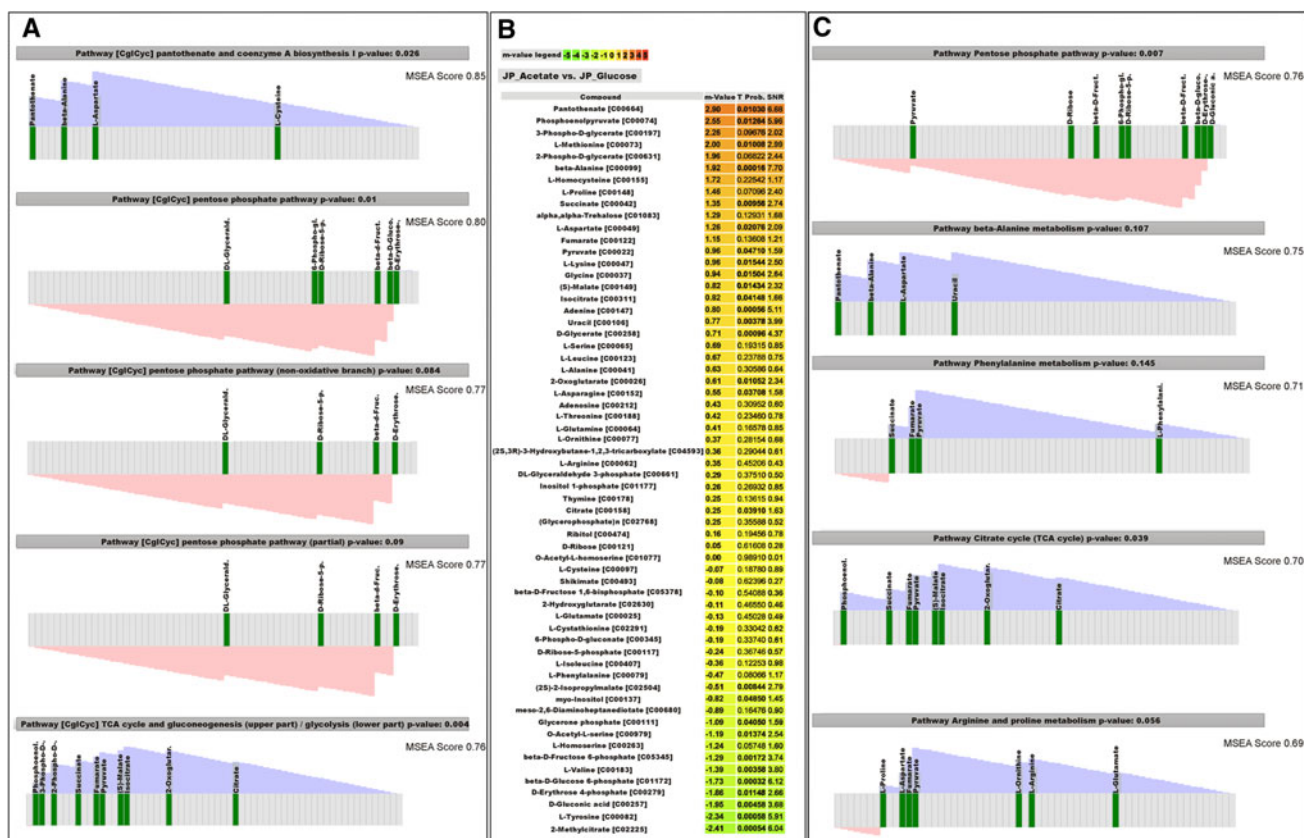
Fig. 2 The visualization of the enrichment analysis results generated out of SVG graphics from the web interface. The best five results for the CglCyc (**a**) and the KEGG (**c**) metabolic pathways exhibiting the highest enrichment scores were presented. The running ES score is plotted using *blue* color if the metabolites belonging to MS are predominantly found in the top part of the ordered List *L*. The *red* color indicates that the metabolites are found at the lower part of *L*. All metabolites present in the experiment are represented as *gray*

boxes and the ones belonging to MS are labeled and highlighted in *green*. In the middle, the computed m-values (**b**), the *t*-test probabilities, and the signal to noise ratios (SNR) are listed for all compounds identified in the analyzed metabolomics experiment of *C. glutamicum* ATCC 13032 grown on the carbon sources glucose and acetate. For the background of the table cells, a color mapping function of the m-values from *red* to *green* was chosen

pathway hit for gluconeogenesis was found under the best five hits. Due to the different behavior of the metabolites corresponding to the gluconeogenesis pathway, higher pools in the upper part and lower pools in the upper part, the MSEA score for this static KEGG pathway was low. A main drawback using the KEGG database for MSEA can be observed here. The pathways in the KEGG database are designed to present a comprehensive view on the cellular metabolism, resulting in a low granularity of many pathways. As MSEA works better with compartmentalized pathways and KEGG pathways cannot be redefined by the user, use of these pathways may give poor results for MSEA.

The other three pathway hits were different in comparison to the MSEA results using the CglCyc database. The pathway corresponding to the second best hit was the *ß-alanine metabolism*. This pathway represents a part of the *pantothenate and coenzyme A biosynthesis* (Hüser et al. 2005), being the best hit using the CglCyc database. Here,

four metabolites were identified, aspartate, ß-alanine, pantothenate and uracil. This highlights the second major drawback of using the KEGG database for MSEA, the presence of metabolic routes in the pathway that are not present in the organism analyzed with MSEA. For example, the uracil degradation pathway to ß-alanine is most probably not present in *C. glutamicum*. False positive metabolite identifications in pathways will cause over/under representation of a pathway. Here, ß-alanine biosynthesis would not be identified when excluding pathways with less than four metabolite identifiers. Therefore, it might be necessary to check whether changing the minimal number of matched metabolites in the pathways delivers more meaningful results.

The last two pathways identified were the *phenylalanine metabolism* and the *arginine and proline metabolism*. It is hard to decide whether these hits should be considered a real true positive as several of the metabolites responsible for the high MSEA scores are also found in two of the best

five scoring pathways, the pentose phosphate pathway and the TCA cycle. This is due to the interconnectedness of the different pathways which cannot be disentangled properly. Still, the presence of other, unique metabolites provides circumstantial evidence that these hits are indeed correct.

Since the results of the MSEA for the glucose and acetate experiment in *C. glutamicum* delivered better results with the CglCyc database, we used this for the following analyses.

3.2 Analysis of the metabolic shift caused by allelic exchange of the homoserine dehydrogenase gene (*hom*) in *C. glutamicum* strains by metabolome analysis

As a second test case, the metabolic profiles of a strain development line originating from a genetically well-defined lysine producer were analyzed. In these strains, the branching point between lysine and threonine biosynthesis, homoserine dehydrogenase was engineered in a step-wise manner, ranging from a leaky mutation to a wild-type gene and to a feedback-deregulated *hom* allele. The strains used were the *C. glutamicum* wild-type ATCC 13032 as common reference, the lysine production strain *C. glutamicum* DM1730 (relevant mutation: *hom*[V59A], leaky *hom* allele) and the derivatives *C. glutamicum* MP001 (*hom*[WT], wild-type allele) and *C. glutamicum* DM1795 (*hom*[FBR], encoding a feedback-resistant enzyme). Thus the applicability of MSEA to analyze production strain development lines with gradual changes in a single enzyme was tested.

3.3 MSEA of the lysine production strain *C. glutamicum* DM1730

The five top scoring pathways delivered by MSEA were selected for a more detailed inspection. Here, the minimal pathway matches were five metabolites. Surprisingly, the pathways for *formyl-tetrahydrofolate* (*formyl-THF*) *biosynthesis I/II* are found as the top scoring hits in the strain DM1730 (Table 1, Electronic Supplementary Fig. 1). These two pathway hits were performed on the same identified metabolites, serine, homocysteine, glutamate, glycine and methionine. Formyl-THF and its derivatives are not measurable using the applied GC–MS method, as the substances have high molecular weights and are therefore not volatile at the temperatures used. MSEA for these pathways showed a higher abundance of the matched metabolites than the wild type. Conversely, the end products of the formyl-THF biosynthesis are methionine and glycine (Rückert et al. 2003), both metabolites are not affected in comparison to the wild type (Fig. 3). In fact, the precursors of the formyl-THF biosynthesis, homocysteine and serine, showed a higher abundance in comparison to the wild type, indicating of a lower formation of formyl-THF. This may be due to the fact that the pool of methionine, acting as precursor for the C^1-metabolism via *S*-adenosyl-methionine (Lu 2000), shows the same pool size as the wild type, indicating an optimal supply for growth and leading to accumulation of precursors.

The third top pathway hit, *lysine biosynthesis I* is found to be on target. As this strains carry mutations redirecting

Table 1 MSEA results of the comparison of the strains *C. glutamicum* DM1730, MP001 and DM1730 with the wild type

Strain	Pathway	Higher/lower abundance	MSEA-score	*p*-value
DM1730	1. *Formyl-THF biosynthesis I*	Higher	0.86	0.021
	2. *Formyl-THF biosynthesis II*	Higher	0.86	0.018
	3. *Lysine biosynthesis I*	Higher	0.83	0.013
	4. *TCA cycle and gluconeogenesis* (*upper part*)/*glycolysis* (*lower part*)	Lower	0.80	0.010
	5. *Superpathway of lysine, threonine and methionine biosynthesis I*	Higher	0.66	0.058
MP001	1. *Threonine biosynthesis*	Higher	0.86	0.008
	2. *TCA cycle and gluconeogenesis* (*upper part*)/*glycolysis* (*lower part*)	Lower	0.82	0.004
	3. *Superpathway of lysine, threonine and methionine biosynthesis I*	Higher	0.74	0.004
	4. *Superpathway of methionine biosynthesis* (*by transsulfuration*)	Higher	0.74	0.026
	5. *Aspartate superpathway*	Higher	0.72	0.003
DM1795	1. *Threonine biosynthesis*	Higher	0.91	0.009
	2. *Isoleucine biosynthesis I* (*from threonine*)	Higher	0.85	0.034
	3. *Superpathway of methionine biosynthesis* (*by transsulfuration*)	Higher	0.83	0.021
	4. *Methionine biosynthesis I*	Higher	0.77	0.044
	5. *TCA cycle and gluconeogenesis* (*upper part*)/*glycolysis* (*lower part*)	Lower	0.76	0.015

Presented are the MSEA calculations and the abundance scoring for the best five pathways involved in these comparisons
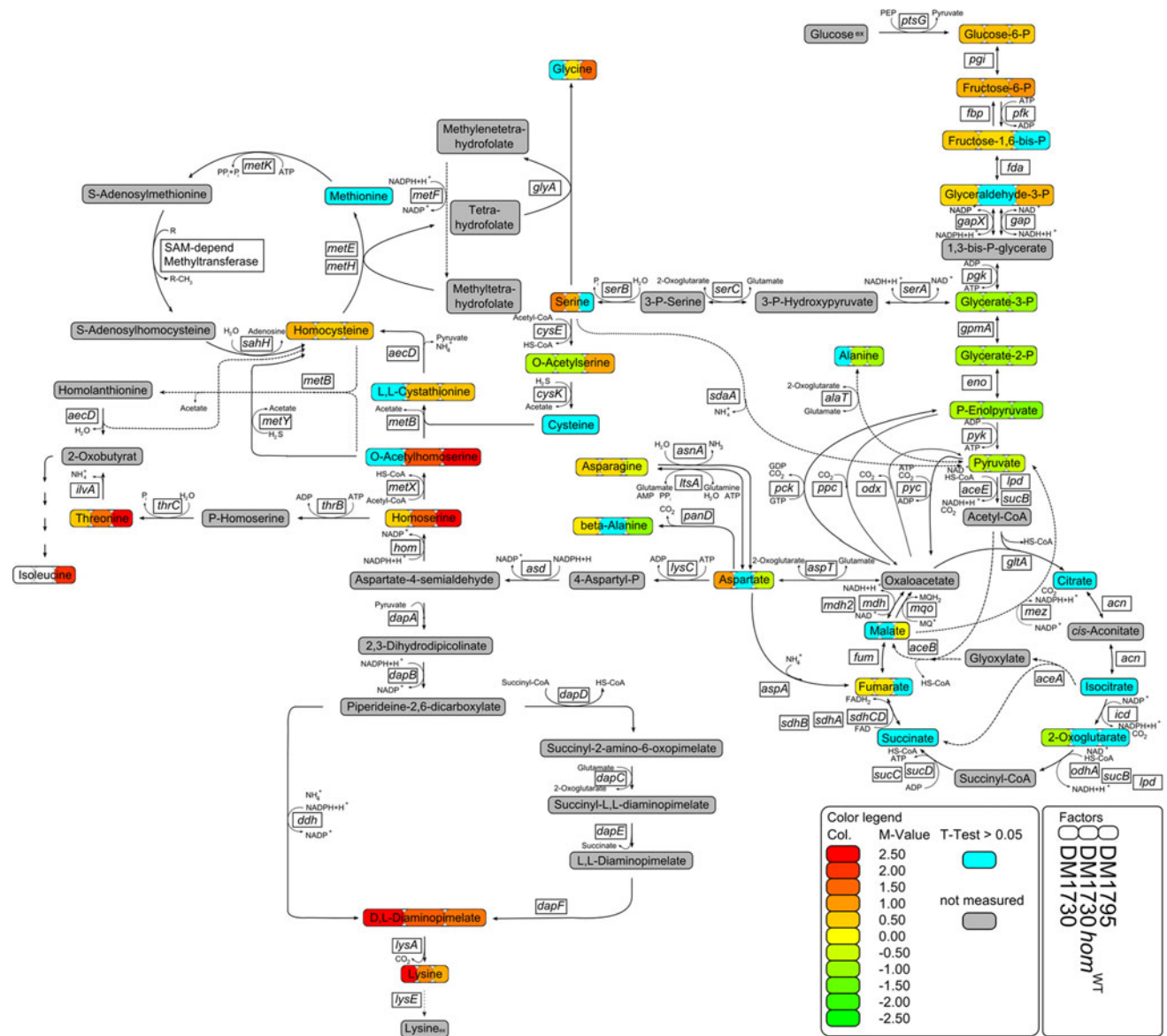
Fig. 3 Visualization of the m-values obtained from the metabolite profiles of *C. glutamicum* DM1730, MP001 and DM1795 in comparison to the wild type by the software tool ProMeTra. The m-values of the metabolites are presented in ellipses in color code from *green* to *red*. Not measured metabolites are shown in *grey* and metabolites, showing no significant difference (*t*-test > 0.05) between the analyzed strains and the wild type, are given in *blue*

carbon flow towards lysine biosynthesis (pyc^{P458S}, $lysC^{T311I}$, Δpck) (Ohnishi et al. 2002), this result delivers another proof of principle.

In the fourth position, the CglCyc pathway hit belonging to the carbohydrate- and energy metabolism group was found in the strain DM1730, namely *TCA cycle and gluconeogenesis (upper part)/glycolysis (lower part)*. Since the strains were cultivated in media containing glucose as sole carbon source and harvested in the logarithmic growth phase, it can be assumed that this hit is due to differences in the lower part of glycolysis as no gluconeogenesis is needed for the cell growth at these conditions. The matched

metabolites belong to the TCA cycle and the lower part of the glycolysis and showed lower pool sizes than the *C. glutamicum* wild type (Fig. 3). Due to the *pyc* (pyruvate carboxylase) and *lysC* (aspartate kinase) mutations in the strain DM1730, resulting in a higher flux from pyruvate to oxaloacetate and its efflux into the lysine pathway, it was already expected that these metabolite pools are affected (Ikeda et al. 2006).

In the next CglCyc pathway hit of the MSEA, the *superpathway of lysine, threonine and methionine biosynthesis I* was found for strain DM1730. The pool sizes of the metabolites belonging to this pathway also showed higher

levels than in the wild type strain (Fig. 3). Here, smaller pool sizes for threonine and homoserine were expected, since the strain DM1730 harbors the leaky *hom* gene. The higher pools of threonine and homoserine might be due to a higher abundance of the precursor aspartate-ß-semialdehyde in the strain DM1730, resulting form the efflux of oxaloacetate into the pathway of the aspartate-derived amino acids.

3.4 MSEA of the strain *C. glutamicum* MP001

The MSEA for the strain MP001, containing the *hom*^{WT} allele, delivers the *threonine biosynthesis* as best hit (Table 1, Electronic Supplementary Fig. 2). Using the ProMeTra pathway visualization for a closer inspection, it was found that the exchange of the leaky *hom* gene against the wild type *hom* gene resulted in a higher enzyme activity of the *hom* gene and therefore resulting in higher pool sizes of metabolites of the threonine and the branching methionine pathway (Fig. 3), thus increasing the MSEA score for *threonine biosynthesis*. We were aware of the fact metabolite pools do not directly reflect fluxes, but in case of the *hom* gene, the increased enzyme activity might potentially result in an increased metabolic flow at this point. The equilibrium of the enzyme reaction is shifted towards the product side (homoserine).

The next two CglCyc pathway hits of the MSEA results of the strain MP001 were found likewise for the strain DM1730, namely *TCA cycle and gluconeogenesis (upper part)/glycolysis (lower part)* and *superpathway of lysine, threonine and methionine biosynthesis I*. Here it becomes obvious that if the strain harbors the wild type *hom* gene, the pools of threonine and methionine are higher than in the strain DM1730, leading to higher MSEA scores for this pathway hit. Furthermore, the increased pool sizes of the metabolites of the threonine and methionine pathway caused by the wild type *hom* gene generate a higher demand of the precursor aspartate-ß-semialdehyde. This leads to lower pool sizes of metabolites in the lysine biosynthesis and the pool of aspartate as well as those of its precursors in the TCA cycle and the lower glycolysis (Fig. 3) thus increasing the MSEA score for the CglCyc hit *TCA cycle and gluconeogenesis (upper part)/glycolysis (lower part)* in comparison to the strain DM1730.

The *superpathway of methionine biosynthesis (by sulfhydrylation)* is the pathway hit at fourth position. As mentioned above, the incorporation of the wild type *hom* gene shifts the carbon flow in the strain MP001, resulting in higher pool sizes of the metabolites of the threonine and methionine pathways. In particular, homoserine and *O*-acetyl-homoserine showed higher pool sizes in the strain MP001 compared to the strain DM1730.

The fifth pathway hit for the strain MP001 was the *aspartate superpathway*, showing the same metabolite identifiers as the *superpathway of lysine, threonine and methionine biosynthesis I* until dihydroxy acetone phosphate. This metabolite belongs to the nicotinamide adenine dinucleotide (NAD) biosynthesis present in this superpathway. NAD and its precursors are not measurable with GC–MS, as formyl-THF their molecular weights are too high. Since the *superpathway of lysine, threonine and methionine biosynthesis I* had a high MSEA score, it was to be expected that this pathway occurs in the best five pathway hits.

3.5 MSEA of the strain *C. glutamicum* DM1795

When comparing the MSEA calculations for the strain *C. glutamicum* DM1795 to the other two strains, we observed that other pathways were involved here. The two best pathway hits were *threonine biosynthesis* and *isoleucine biosynthesis (from threonine)* (Table 1, Electronic Supplementary Fig. 3). Since the feedback-deregulated *hom* gene leads to an enhanced production of threonine, higher pool sizes of isoleucine can be observed due to the fact that threonine is a precursor for this metabolite (Guillouet et al. 2001) (Fig. 3).

The next two MSEA pathway hits belong to sulfur containing metabolites, *superpathway of methionine biosynthesis (by transsulfuration)*, and *methionine biosynthesis I*. The matched metabolites in these CglCyc pathways were mostly the same, only the metabolites aspartate, pyruvate and cystathionine differ in the pathways. In the case of aspartate, the *pathway of methionine biosynthesis I* starts at homoserine according to PathwayTools (Karp et al. 2010). The absence or presence of cystathionine and pyruvate in methionine pathways is due to the fact that *C. glutamicum* utilizes *O*-acetyl-homoserine by a branched pathway, the transsulfuration- and the sulfhydrylation-pathway (Hwang et al. 2002; Lee and Hwang 2003). The feedback-deregulated *hom* gene shifts the metabolic profile of the strain DM1975 to small metabolite pools corresponding to the lysine pathway and to high pool sizes of metabolites of the threonine and methionine pathway. In particular, the pool sizes of threonine, homoserine and *O*-acetyl-homoserine were higher compared to the strains *C. glutamicum* DM1730 and MP001 (Fig. 3).

At the fifth position, the pathway *TCA cycle and gluconeogenesis (upper part)/glycolysis (lower part)* can be found. Since the strains DM1730, MP001 and DM1795 share the same mutations except the homoserine dehydrogenase, it is a consistency check for MSEA that this pathway was identified in all three strains.

Closer inspection of the pathway for methionine metabolism revealed that conversion of *O*-acetyl-homoserine into

homocysteine or cystathionine is apparently limiting in this strain as *O*-acetyl-homoserine was found to accumulate. Also, the pools of serine and glycine in the three strains harboring the three different *hom* genes were found to be altered significantly and in an opposing way. While the pool of serine is lowering with rising metabolite pools corresponding to the methionine pathway, the glycine pool increases. This might be a sign for a higher conversion of homocysteine into methionine by MetE and MetH, the two homocysteine methyltransferases (Rückert et al. 2003). The methyl group derives from methyl-THF, which is regenerated by GlyA and MetF, leading to the conversion of serine into glycine.

As we showed in this paper, MSEA supports generating knowledge from metabolomics data. As metabolites often occur in more than one metabolic pathway, it might be difficult to interpret the results of metabolic profiling analyses. MSEA is able to identify metabolic pathways where the metabolites show the same behavior, thus facilitating data analysis.

4 Concluding remarks

MeltDB was designed as platform-independent software for the analysis and integration of metabolomics experiments. We extended the MeltDB platform by the metabolite set enrichment, which can directly be applied for the identification of metabolic pathways that exhibit the main changes in metabolite pool levels under different experimental conditions. As presented in our evaluation with metabolic profiles obtained from *C. glutamicum* wild type and mutant strains with known genetic modifications, the MSEA approach is able to identify the metabolic pathways associated to the modifications. Users have to keep in mind that MSEA is dependent on the quality and the comprehensiveness of the metabolic data. In addition, success is constrained to the composition and complexity of metabolic databases.

We curated manually a CglCyc database that up to now contains 119 biosynthetic pathways which are directly accessible to the implemented MSEA approach. In the future, we will extend the CglCyc pathway database to improve the analysis of larger metabolite profiles generated by GC–MS or LC–MS. Apart from the analysis of bacterial production and wild-type strains, the general MSEA approach enables identification of metabolic differences on the level of pathways for all types of metabolomics and metabolic profiling experiments. The direct integration of the approach in MeltDB ensures that this novel analysis approach can be readily applied to existing and future metabolomics experiments.

References

Aggio, R. B., Ruggiero, K., & Villas-Boas, S. G. (2010). Pathway activity profiling (PAPi): From the metabolite profile to the metabolic pathway activity. *Bioinformatics, 26*(23), 2969–2976.

Auchter, M., Arndt, A., & Eikmanns, B. J. (2009). Dual transcriptional control of the acetaldehyde dehydrogenase gene *ald* of *Corynebacterium glutamicum* by RamA and RamB. *Journal of Biotechnology, 140*(1–2), 84–91.

Becker, J., Klopprogge, C., Schroder, H., & Wittmann, C. (2009). Metabolic engineering of the tricarboxylic acid cycle for improved lysine production by *Corynebacterium glutamicum*. *Applied and Environmental Microbiology, 75*(24), 7866–7869.

Becker, J., Klopprogge, C., & Wittmann, C. (2008). Metabolic responses to pyruvate kinase deletion in lysine producing *Corynebacterium glutamicum*. *Microbial Cell Factories, 7*, 8.

Blombach, B., & Seibold, G. M. (2010). Carbohydrate metabolism in *Corynebacterium glutamicum* and applications for the metabolic engineering of L-lysine production strains. *Applied Microbiology and Biotechnology, 86*(5), 1313–1322.

Born, T. L., & Blanchard, J. S. (1999). Enzyme-catalyzed acylation of homoserine: Mechanistic characterization of the *Escherichia coli metA*-encoded homoserine transsuccinylase. *Biochemistry, 38*(43), 14416–14423.

Burkovski, A. (2006). Proteomics of *Corynebacterium glutamicum*: Essential industrial bacterium. *Methods of Biochemical Analysis, 49*, 137–147.

de Koning, W., & van Dam, K. (1992). A method for the determination of changes of glycolytic metabolites in yeast on a subsecond time scale using extraction at neutral pH. *Analytical Biochemistry, 204*(1), 118–123.

Drysch, A., El Massaoudi, M., Mack, C., et al. (2003). Production process monitoring by serial mapping of microbial carbon flux distributions using a novel sensor reactor approach: II–(13)C-labeling-based metabolic flux analysis and L-lysine production. *Metabolic Engineering, 5*(2), 96–107.

Drysch, A., El Massaoudi, M., Wiechert, W., de Graaf, A. A., & Takors, R. (2004). Serial flux mapping of *Corynebacterium glutamicum* during fed-batch L-lysine production using the sensor reactor approach. *Biotechnology and Bioengineering, 85*(5), 497–505.

Dunn, W. B., & Ellis, D. I. (2005). Metabolomics: Current analytical platforms and methodologies. *Trends in Analytical Chemistry, 4*, 285–294.

El Massaoudi, M., Spelthahn, J., Drysch, A., de Graaf, A., & Takors, R. (2003). Production process monitoring by serial mapping of microbial carbon flux distributions using a novel sensor reactor approach: I–Sensor reactor system. *Metabolic Engineering, 5*(2), 86–95.

Fernie, A. R., Trethewey, R. N., Krotzky, A. J., & Willmitzer, L. (2004). Metabolite profiling: From diagnostics to systems

biology. *Nature Reviews. Molecular Cell Biology, 5*(9), 763–769.

Fiehn, O. (2002). Metabolomics—the link between genotypes and phenotypes. *Plant Molecular Biology, 48*(1–2), 155–171.

Gerstmeir, R., Wendisch, V. F., Schnicke, S., et al. (2003). Acetate metabolism and its regulation in *Corynebacterium glutamicum*. *Journal of Biotechnology, 104*(1–3), 99–122.

Goodacre, R., Vaidyanathan, S., Dunn, W. B., Harrigan, G. G., & Kell, D. B. (2004). Metabolomics by numbers: Acquiring and understanding global metabolite data. *Trends in Biotechnology, 22*(5), 245–252.

Guillouet, S., Rodal, A. A., An, G. H., et al. (2001). Metabolic redirection of carbon flow toward isoleucine by expressing a catabolic threonine dehydratase in a threonine-overproducing *Corynebacterium glutamicum*. *Applied Microbiology and Biotechnology, 57*(5–6), 667–673.

Hall, R. D. (2006). Plant metabolomics: From holistic hope, to hype, to hot topic. *New Phytologist, 169*(3), 453–468.

Han, S. O., Inui, M., & Yukawa, H. (2008). Transcription of *Corynebacterium glutamicum* genes involved in tricarboxylic acid cycle and glyoxylate cycle. *Journal of Molecular Microbiology and Biotechnology, 15*(4), 264–276.

Hansmeier, N., Chao, T. C., Puhler, A., Tauch, A., & Kalinowski, J. (2006). The cytosolic, cell surface and extracellular proteomes of the biotechnologically important soil bacterium *Corynebacterium effic*iens YS-314 in comparison to those of *Corynebacterium glutamicum* ATCC 13032. *Proteomics, 6*(1), 233–250.

Hayashi, M., Mizoguchi, H., Shiraishi, N., et al. (2002). Transcriptome analysis of acetate metabolism in *Corynebacterium glutamicum* using a newly developed metabolic array. *Bioscience, Biotechnology, and Biochemistry, 66*(6), 1337–1344.

Hermann, T. (2003). Industrial production of amino acids by coryneform bacteria. *Journal of Biotechnology, 104*(1–3), 155–172.

Hoon Yang, T., Wittmann, C., & Heinzle, E. (2006). Respirometric 13C flux analysis–Part II: In vivo flux estimation of lysine-producing *Corynebacterium glutamicum*. *Metabolic Engineering, 8*(5), 432–446.

Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2009). Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research, 37*(1), 1–13.

Hüser, A. T., Becker, A., Brune, I., et al. (2003). Development of a *Corynebacterium glutamicum* DNA microarray and validation by genome-wide expression profiling during growth with propionate as carbon source. *Journal of Biotechnology, 106*(2–3), 269–286.

Hüser, A. T., Chassagnole, C., Lindley, N. D., et al. (2005). Rational design of a *Corynebacterium glutamicum* pantothenate production strain and its characterization by metabolic flux analysis and genome-wide transcriptional profiling. *Applied and Environmental Microbiology, 71*(6), 3255–3268.

Hwang, B. J., Kim, Y., Kim, H. B., et al. (1999). Analysis of *Corynebacterium glutamicum* methionine biosynthetic pathway: Isolation and analysis of *metB* encoding cystathionine gamma-synthase. *Molecules and Cells, 9*(3), 300–308.

Hwang, B. J., Yeom, H. J., Kim, Y., & Lee, H. S. (2002). *Corynebacterium glutamicum* utilizes both transsulfuration and direct sulfhydrylation pathways for methionine biosynthesis. *Journal of Bacteriology, 184*(5), 1277–1286.

Ikeda, M., Ohnishi, J., Hayashi, M., & Mitsuhashi, S. (2006). A genome-based approach to create a minimally mutated *Corynebacterium glutamicum* strain for efficient L-lysine production. *Journal of Industrial Microbiology & Biotechnology, 33*(7), 610–615.

Kalinowski, J., Bathe, B., Bartels, D., et al. (2003). The complete Corynebacterium glutamicum ATCC 13032 genome sequence and its impact on the production of L-aspartate-derived amino acids and vitamins. *Journal of Biotechnology, 104*(1–3), 5–25.

Kanehisa, M., Goto, S., Hattori, M. et al. (2006). From genomics to chemical genomics: New developments in KEGG. *Nucleic Acids Research, 34*(Database issue), D354–D357.

Karp, P. D., Paley, S. M., Krummenacker, M., et al. (2010). Pathway Tools version 13.0: Integrated software for pathway/genome informatics and systems biology. *Briefings in Bioinformatics, 11*(1), 40–79.

Kawaguchi, H., Sasaki, M., Vertes, A. A., Inui, M., & Yukawa, H. (2009). Identification and functional analysis of the gene cluster for L-arabinose utilization in *Corynebacterium glutamicum*. *Applied and Environmental Microbiology, 75*(11), 3419–3429.

Keilhauer, C., Eggeling, L., & Sahm, H. (1993). Isoleucine synthesis in *Corynebacterium glutamicum*: Molecular analysis of the ilvB-ilvN-ilvC operon. *Journal of Bacteriology, 175*(17), 5595–5603.

Kiefer, P., Heinzle, E., & Wittmann, C. (2002). Influence of glucose, fructose and sucrose as carbon sources on kinetics and stoichiometry of lysine production by *Corynebacterium glutamicum*. *Journal of Industrial Microbiology & Biotechnology, 28*(6), 338–343.

Krömer, J. O., Fritz, M., Heinzle, E., & Wittmann, C. (2005). In vivo quantification of intracellular amino acids and intermediates of the methionine pathway in *Corynebacterium glutamicum*. *Analytical Biochemistry, 340*(1), 171–173.

Krömer, J. O., Sorgenfrei, O., Klopprogge, K., Heinzle, E., & Wittmann, C. (2004). In-depth profiling of lysine-producing *Corynebacterium glutamicum* by combined analysis of the transcriptome, metabolome, and fluxome. *Journal of Bacteriology, 186*(6), 1769–1784.

Lee, H. S., & Hwang, B. J. (2003). Methionine biosynthesis and its regulation in *Corynebacterium glutamicum*: Parallel pathways of transsulfuration and direct sulfhydrylation. *Applied Microbiology and Biotechnology, 62*(5–6), 459–467.

Leuchtenberger, W. (1996) Amino acids: Technical production and use. In H. J. Rehm, G. Reed, A. Pühler, & P. Stadler (Eds.), Biotechnology (pp. 466–502). Weinheim: VCH.

Leuchtenberger, W., Huthmacher, K., & Drauz, K. (2005). Biotechnological production of amino acids and derivatives: Current status and prospects. *Applied Microbiology and Biotechnology, 69*(1), 1–8.

Lu, S. C. (2000). S-Adenosylmethionine. *International Journal of Biochemistry and Cell Biology, 32*(4), 391–395.

Mootha, V. K., Lindgren, C. M., Eriksson, K.-F., et al. (2003). PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature Genetics, 34*(3), 267–273.

Morbach, S., Sahm, H., & Eggeling, L. (1996). L-Isoleucine production with *Corynebacterium glutamicum*: Further flux increase and limitation of export. *Applied and Environmental Microbiology, 62*(12), 4345–4351.

Muffler, A., Bettermann, S., Haushalter, M., et al. (2002). Genome-wide transcription profiling of *Corynebacterium glutamicum* after heat shock and during growth on acetate and glucose. *Journal of Biotechnology, 98*(2–3), 255–268.

Neuweger, H., Albaum, S. P., Dondrup, M., et al. (2008). MeltDB: A software platform for the analysis and integration of metabolomics experiment data. *Bioinformatics, 24*(23), 2726–2732.

Neuweger, H., Persicke, M., Albaum, S. P., et al. (2009). Visualizing post genomics data-sets on customized pathway maps by ProMeTra-aeration-dependent gene expression and metabolism of *Corynebacterium glutamicum* as an example. *BMC Systems Biology, 3*, 82.

Nobeli, I., Ponstingl, H., Krissinel, E. B., & Thornton, J. M. (2003). A structure-based anatomy of the *E. coli* metabolome. *Journal of Molecular Biology, 334*(4), 697–719.

Nöh, K., Gronke, K., Luo, B., et al. (2007). Metabolic flux analysis at ultra short time scale: Isotopically non-stationary 13C labeling experiments. *Journal of Biotechnology, 129*(2), 249–267.

Ohnishi, J., Mitsuhashi, S., Hayashi, M., et al. (2002). A novel methodology employing *Corynebacterium glutamicum* genome information to generate a new L-lysine-producing mutant. *Applied Microbiology and Biotechnology, 58*(2), 217–223.

Parche, S., Burkovski, A., Sprenger, G. A., et al. (2001). *Corynebacterium glutamicum*: A dissection of the PTS. *Journal of Molecular Microbiology and Biotechnology, 3*(3), 423–428.

Park, S. D., Lee, J. Y., Kim, Y., Kim, J. H., & Lee, H. S. (1998). Isolation and analysis of *metA*, a methionine biosynthetic gene encoding homoserine acetyltransferase in *Corynebacterium glutamicum. Molecules and Cells, 8*(3), 286–294.

Park, S. D., Lee, J. Y., Sim, S. Y., Kim, Y., & Lee, H. S. (2007). Characteristics of methionine production by an engineered *Corynebacterium glutamicum* strain. *Metabolic Engineering, 9*(4), 327–336.

Persicke, M., Plassmeier, J., Neuweger, H. et al. (2010). Size exclusion chromatography-an improved method to harvest *Corynebacterium glutamicum* cells for the analysis of cytosolic metabolites. *Journal of Biotechnology*. doi:10.1016/j.jbiotec.2010.08.016.

Plassmeier, J., Barsch, A., Persicke, M., Niehaus, K., & Kalinowski, J. (2007). Investigation of central carbon metabolism and the 2-methylcitrate cycle in *Corynebacterium glutamicum* by metabolic profiling using gas chromatography-mass spectrometry. *Journal of Biotechnology, 130*(4), 354–363.

Rey, D. A., Nentwich, S. S., Koch, D. J., et al. (2005). The McbR repressor modulated by the effector substance *S*-adenosylhomocysteine controls directly the transcription of a regulon involved in sulphur metabolism of *Corynebacterium glutamicum* ATCC 13032. *Molecular Microbiology, 56*(4), 871–887.

Rey, D. A., Pühler, A., & Kalinowski, J. (2003). The putative transcriptional repressor McbR, member of the TetR-family, is involved in the regulation of the metabolic network directing the synthesis of sulfur containing amino acids in *Corynebacterium glutamicum. Journal of Biotechnology, 103*(1), 51–65.

Rubin, E. (2006). Circumventing the cut-off for enrichment analysis. *Briefings in Bioinformatics, 7*(2), 202–203.

Rückert, C., Koch, D. J., Rey, D. A., et al. (2005). Functional genomics and expression analysis of the *Corynebacterium glutamicum fpr2-cysIXHDNYZ* gene cluster involved in assimilatory sulphate reduction. *BMC Genomics, 6*, 121.

Rückert, C., Pühler, A., & Kalinowski, J. (2003). Genome-wide analysis of the L-methionine biosynthetic pathway in *Corynebacterium glutamicum* by targeted gene deletion and homologous complementation. *Journal of Biotechnology, 104*(1–3), 213–228.

Schäfer, U., Boos, W., Takors, R., & Weuster-Botz, D. (1999). Automated sampling device for monitoring intracellular metabolite dynamics. *Analytical Biochemistry, 270*(1), 88–96.

Seibold, G., Auchter, M., Berens, S., Kalinowski, J., & Eikmanns, B. J. (2006). Utilization of soluble starch by a recombinant *Corynebacterium glutamicum* strain: Growth and lysine production. *Journal of Biotechnology, 124*(2), 381–391.

Seibold, G. M., Wurst, M., & Eikmanns, B. J. (2009). Roles of maltodextrin and glycogen phosphorylases in maltose utilization and glycogen metabolism in *Corynebacterium glutamicum. Microbiology, 155*(Pt 2), 347–358.

Stavrum, A. K., Petersen, K., Jonassen, I., & Dysvik, B. (2008). Analysis of gene-expression data using J-Express. *Current Protocols in Bioinformatics*, Chap. 7, Unit 7.3.

Subramanian, A., Kuehn, H., Gould, J., Tamayo, P., & Mesirov, J. P. (2007). GSEA-P: A desktop application for gene set enrichment analysis. *Bioinformatics, 23*(23), 3251–3253.

Subramanian, A., Tamayo, P., Mootha, V. K., et al. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America, 102*(43), 15545–15550.

Tauch, A., Götker, S., Pühler, A., Kalinowski, J., & Thierbach, G. (2002). The alanine racemase gene alr is an alternative to antibiotic resistance genes in cloning systems for industrial *Corynebacterium glutamicum* strains. *Journal of Biotechnology, 99*(1), 79–91.

van den Berg, R. A., Hoefsloot, H. C., Westerhuis, J. A., Smilde, A. K., & van der Werf, M. J. (2006). Centering, scaling, and transformations: Improving the biological information content of metabolomics data. *BMC Genomics, 7*, 142.

Veit, A., Rittmann, D., Georgi, T., et al. (2009). Pathway identification combining metabolic flux and functional genomics analyses: acetate and propionate activation by *Corynebacterium glutamicum. Journal of Biotechnology, 140*(1–2), 75–83.

Wendisch, V. F., Bott, M., Kalinowski, J., Oldiges, M., & Wiechert, W. (2006). Emerging *Corynebacterium glutamicum* systems biology. *Journal of Biotechnology, 124*(1), 74–92.

Wendisch, V. F., de Graaf, A. A., Sahm, H., & Eikmanns, B. J. (2000). Quantitative determination of metabolic fluxes during coutilization of two carbon sources: comparative analyses with *Corynebacterium glutamicum* during growth on acetate and/or glucose. *Journal of Bacteriology, 182*(11), 3088–3096.

Wittmann, C., Kiefer, P., & Zelder, O. (2004). Metabolic fluxes in *Corynebacterium glutamicum* during lysine production with sucrose as carbon source. *Applied and Environmental Microbiology, 70*(12), 7277–7287.

Xia, J., & Wishart, D. S. (2010). MSEA: A web-based tool to identify biologically meaningful patterns in quantitative metabolomic data. *Nucleic Acids Research, 38*(Suppl), W71–W77.