



OPEN

DATA DESCRIPTOR

A manually curated compendium of expression profiles for the microbial cell factory *Corynebacterium glutamicum*

Angela Kranz^{1,2}✉, Tino Polen¹, Christian Kotulla¹, Annette Arndt³, Graziella Bosco¹, Michael Bussmann¹, Ava Chattopadhyay¹, Annette Cramer³, Cedric-Farhad Davoudi¹, Ursula Degner¹, Ramon Diesveld¹, Raphael Freiherr von Boeselager¹, Kim Gärtner¹, Cornelia Gätgens¹, Tobias Georgi¹, Christian Geraths¹, Sabine Haas¹, Antonia Heyer¹, Max Hünnefeld¹, Takeru Ishige¹, Armin Kabus¹, Nicolai Kallscheuer¹, Larissa Keuer¹, Simon Klaffl¹, Britta Kleine¹, Martina Kočan¹, Abigail Koch-Koerfges¹, Kim J. Kraxner¹, Andreas Krug¹, Aileen Krüger¹, Andreas Küberl¹, Mohamed Labib¹, Christian Lange¹, Christina Mack¹, Tomoya Maeda¹, Regina Mahr¹, Stephan Majda¹, Andrea Michel¹, Xenia Morosov¹, Olga Müller¹, Arun M. Nanda¹, Jens Nickel¹, Jennifer Pahlke¹, Eugen Pfeifer¹, Laura Plätzen¹, Paul Ramp¹, Doris Rittmann¹, Steffen Schaffer¹, Sandra Scheele¹, Stephanie Spelberg¹, Julia Schulte¹, Jens-Eric Schweitzer¹, Georg Sindelar¹, Ulrike Sorger-Herrmann¹, Markus Spelberg¹, Corinna Stansen¹, Apilaasha Tharmasothirajan¹, Jan van Ooyen¹, Philana van Summeren-Wesenhagen⁴, Michael Vogt¹, Sabrina Witthoff¹, Lingfeng Zhu¹, Bernhard J. Eikmanns³, Marco Oldiges¹, Georg Schaumann⁴, Meike Baumgart¹, Melanie Brocker¹, Lothar Eggeling¹, Roland Freudl¹, Julia Frunzke¹, Jan Marienhagen¹, Volker F. Wendisch^{1,5} & Michael Bott¹✉

Corynebacterium glutamicum is the major host for the industrial production of amino acids and has become one of the best studied model organisms in microbial biotechnology. Rational strain construction has led to an improvement of producer strains and to a variety of novel producer strains with a broad substrate and product spectrum. A key factor for the success of these approaches is detailed knowledge of transcriptional regulation in *C. glutamicum*. Here, we present a large compendium of 927 manually curated microarray-based transcriptional profiles for wild-type and engineered strains detecting genome-wide expression changes of the 3,047 annotated genes in response to various environmental conditions or in response to genetic modifications. The replicates within the 927 experiments were combined to 304 microarray sets ordered into six categories that were used for differential gene expression analysis. Hierarchical clustering confirmed that no outliers were present in the sets. The compendium provides a valuable resource for future fundamental and applied research with *C. glutamicum* and contributes to a systemic understanding of this microbial cell factory.

¹IBG-1: Biotechnology, Institute of Bio- and Geosciences, Forschungszentrum Jülich, D-52425, Jülich, Germany.

²IBG-4: Bioinformatics, Institute of Bio- and Geosciences, Forschungszentrum Jülich, D-52425, Jülich, Germany.

³Institute of Microbiology and Biotechnology, University of Ulm, D-89069, Ulm, Germany. ⁴SenseUp GmbH, c/o Campus Forschungszentrum, Wilhelm-Johnen-Strasse, D-52425, Jülich, Germany. ⁵Genetics of Prokaryotes, Biology & CeBITec, Bielefeld University, Universitaetsstr. 25, D-33615, Bielefeld, Germany. ✉e-mail: a.kranz@fz-juelich.de; m.bott@fz-juelich.de

Background & Summary

Corynebacterium glutamicum is a Gram-positive, facultatively anaerobic soil bacterium, which was discovered in the 1950s as a natural L-glutamate producer¹. Nowadays, *C. glutamicum* is established as an important industrial microorganism used for the large-scale production of L-glutamate (3.2 million tons/year) and L-lysine (2.6 million tons/year)^{2–6}. Efficient strains for the synthesis of several other amino acids such as L-leucine⁷, L-isoleucine⁸, L-valine^{9–12}, L-arginine¹³, or L-histidine^{14,15} have also been constructed. After the availability of the genome sequence^{16,17}, the product spectrum accessible with *C. glutamicum* was continuously extended and now includes for example various organic acids^{18–20}, biofuels such as ethanol²¹, isobutanol^{22,23} or 2-methyl-1-butanol and 3-methyl-1-butanol²⁴, carotenoids²⁵, plant secondary metabolites such as plant polyphenols^{26–28}, and heterologous proteins²⁹ such as antibodies. An essential prerequisite for the development of production strains of *C. glutamicum* by re-routing cellular metabolism was the development of efficient and reliable genetic tools enabling for example the deletion and overexpression of genes, genomic integration, or the introduction of modifications in promoter and operator (i.e. regulator-binding) sequences^{30–34}.

Rational development of microbial production strains does not only require knowledge about the metabolic network³⁵, but also of the regulatory networks that control metabolic fluxes. To analyse transcriptional changes of gene expression in response to environmental changes and metabolic shifts in the wild-type as well as engineered strains on a genome-wide scale, transcriptomics methods such as DNA microarrays or RNA-Seq are used. For *C. glutamicum*, many studies on transcriptional regulators (TRs)^{36–38}, including two-component signal transduction systems³⁹ and σ factors⁴⁰, have been performed to elucidate the complex network of transcriptional regulation in this bacterium. To understand the physiological functions of individual TRs, their target genes (also called regulon) need to be identified, e.g. by comparing changes in gene expression in strains lacking or overexpressing individual TR genes. In order to distinguish direct and indirect influences on target gene expression, the binding of the TRs to promoters has to be analyzed. For example, *in vivo* DNA-binding sites of individual TRs can be identified by chromatin immunoprecipitation (ChIP) in combination with microarrays (ChIP-Chip) or high-throughput sequencing (ChIP-Seq)⁴¹, or by using chromatin affinity purification followed by sequencing (ChAP-Seq)⁴².

Of the 159 genes encoding either DNA-binding TRs (139), response regulators of two-component systems (13), or σ factor subunits of the RNA polymerase (seven) in the genome of *C. glutamicum*^{17,38}, 93 TRs^{36,37,43}, seven two-component systems^{39,44–47}, and all seven sigma factors^{40,48–50} have been characterized so far. Information about the respective regulatory functions can be found in the web-based analysis platform CoryneRegNet^{43,51}. Despite this huge progress since the availability of the *C. glutamicum* genome sequences in 2003^{16,17}, the understanding of the transcriptional regulatory network in this bacterium is still far from being complete. In general, the possible number of regulatory interactions is equal to the number of transcription factors multiplied by the total number of genes in the genome multiplied by the number of environmental contexts in which the cell might find itself⁵². Therefore, comprehensive analysis of gene expression changes measured in a broad variety of different strains and growth conditions is necessary by using, for example, methods such as data mining or strategies such as knowledge discovery in databases^{53–55}.

Over the past 20 years, researchers of the Institute of Bio- and Geosciences 1: Biotechnology (IBG-1, until 2010 named Institute of Biotechnology 1) at Forschungszentrum Jülich (Germany) have performed and collected the results of 1,146 microarray experiments in an in-house database⁵⁶. The majority of these microarray data (671 experiments) is not yet available publicly. Thus, in this study, we re-evaluated all experiments stored in our in-house database according to the Minimum Information About Microarray Experiments (MIAME) standards⁵⁷, curated them manually and uploaded them as a superseries of 927 experiments to Gene Expression Omnibus (GEO). Furthermore, we categorized the experiments and combined the replicates to 304 microarray sets that were used for differential gene expression analysis and hierarchical clustering analysis. This approach follows the FAIR principles^{58,59} and allows other researchers to use our data for future studies and contribute to a systemic understanding of the transcriptional regulatory network and the biology of *C. glutamicum* and its biotechnological application. Furthermore, the resulting knowledge gained for *C. glutamicum* can also benefit the understanding of phylogenetically related Actinobacteria, for example mycobacteria^{60,61}.

The complete workflow including processing of microarray experiments, re-evaluation of the data, and categorization is described in detail in the methods as well as validation section and visualized in Figs. 1 and 2.

Methods

Bacterial strains, plasmids, and cultivation conditions. All bacterial strains and plasmids that were used for microarray comparisons are listed in Supplementary Table S1. Strains bearing deletions, genomic integrations and/or promoter exchanges were constructed using a two-step homologous recombination protocol based on the suicide plasmid pK19mobsacB⁶². Construction of plasmids for (over-)expression of homologous or heterologous genes was performed by PCR amplification of the target sequences and cloning into suitable vectors by restriction and ligation or Gibson assembly⁶³. The deletion of genes was confirmed by PCR. Aerobic cultivation of *C. glutamicum* strains under standard conditions was performed in 50 mL of lysogeny broth (LB)⁶⁴, 50 mL brain-heart-infusion (BHI) medium, or 50 mL CGXII minimal medium⁶⁵ in 500 mL shaking flasks at 30 °C. In many cases, 2% (w/v) or 4% (w/v) glucose was used as carbon source. Changes in the composition of the used CGXII medium, such as the use of different carbon sources or the addition of special chemicals, as well as changes in the cultivation temperature are indicated. Plasmids were introduced into the indicated strains usually by electroporation⁶⁶. When required, gene expression was induced, e.g. by addition of isopropyl- β -D-thiogalactopyranoside (IPTG) to a final concentration between 10 μ M and 1 mM as indicated or by addition of anhydrotetracycline (typically 250 ng/ml). Antibiotics were used at the following concentrations: kanamycin, 25 μ g ml⁻¹; chloramphenicol, 10 μ g ml⁻¹; spectinomycin, 100 μ g ml⁻¹. Cells were usually grown until mid-exponential phase (optical density at 600 nm (OD₆₀₀) of ca. 4–5), harvested on ice, frozen in liquid nitrogen, and stored at –80 °C

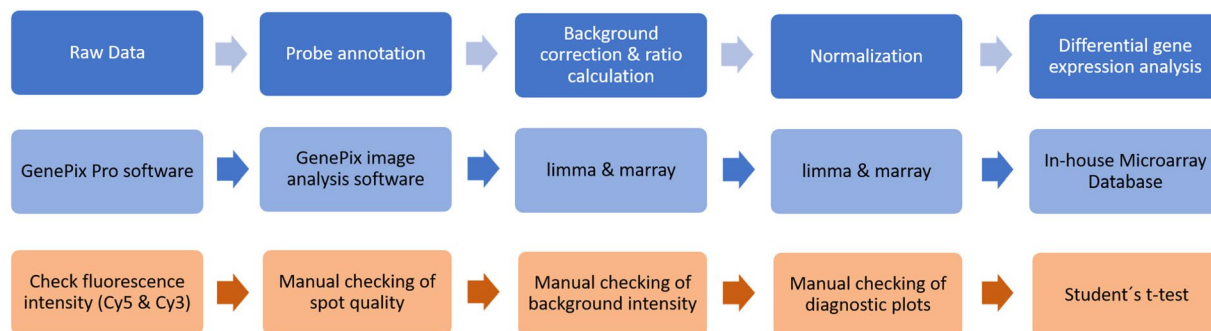


Fig. 1 Processing workflow applied to Agilent, Operon and PCR product-based microarrays. Boxes in blue represent the performed steps, boxes in light blue show the programs that were used, and the boxes in orange contain the quality checking procedures that were applied during the respective steps.

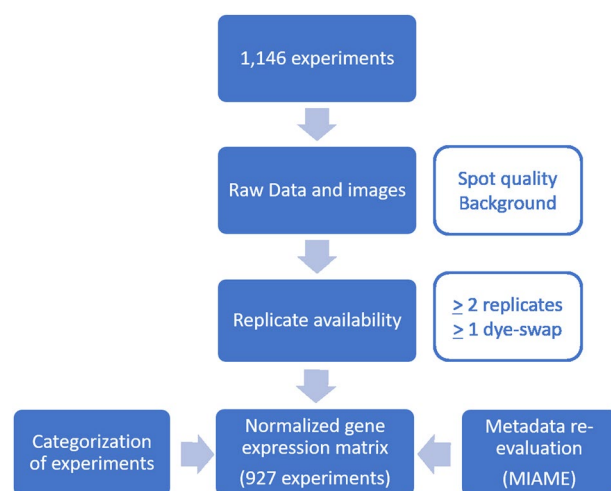


Fig. 2 DNA microarray re-evaluation pipeline. Overall, 1,146 experiments were re-evaluated during this study. After the re-evaluation and manual quality check, 927 high-quality experiments remained and were uploaded to GEO.

until further use for RNA isolation. In some experiments, cells were grown to an OD_{600} of ca. 5, washed in PBS buffer (137 mM NaCl, 2.7 mM KCl, 1.5 mM KH_2PO_4 , 8 mM Na_2HPO_4 , pH 7.4) or 0.9% (w/v) saline, transferred into new medium and harvested after cultivation for the indicated time. Experiments that required longer cultivation of the cells are designated accordingly. Aerobic and anaerobic cultivations in bioreactors were performed as described^{67–69}. For every gene expression comparison, two to nine biological replicates were performed.

RNA isolation and microarray hybridization. Total RNA was isolated using the RNeasy Kit (Qiagen, Hilden, Germany) or as described⁷⁰. Equal amounts of RNA (15–30 μ g) were used for random hexamer-primed synthesis of fluorescently labelled cDNA using the nucleotide analogues Cy3-dUTP or Cy5-dUTP (GE Healthcare, Eindhoven, Netherlands). Gene expression changes of two samples were compared using either self-made PCR product-based DNA microarrays⁷¹, custom-made DNA microarrays with 70-mer oligonucleotides obtained from Operon Biotechnologies (Cologne, Germany), or custom-made 4x44K 60mer DNA microarrays from Agilent Technologies (Waldbronn, Germany). Hybridization of mixtures of Cy3- and Cy5-labelled cDNA on the arrays and washing of the arrays were performed as described^{7,72,73}. Routinely, at least one dye-swap experiment was performed per gene expression comparison to avoid batch effects by consistently labelling identical experimental groups with the same dye. All oligonucleotides spotted on the arrays are based on the genome sequence entry NC_006958¹⁷.

Fluorescence determination and probe annotation. The fluorescence of DNA microarrays was determined at 532 nm (Cy3-dUTP) and 635 nm (Cy5-dUTP) at 5 or 10 μ m resolution with a GenePix 4000B laser scanner and GenePix Pro Software (Molecular Devices, Sunnyvale, USA). Raw data files of fluorescence images were saved in TIFF format and further processed using the GenePix image analysis software. Probe annotation was used to map each spot on the scanned microarray chip to genes. For this purpose, gene array list (GAL) files designed specifically for the platform used were loaded into the analysis software and aligned to the probes to

perform spot detection. Multiple probes that mapped onto the same gene ID were summarized by their median values. All GAL files were uploaded to the GEO database (GPL29897).

Background correction, ratio calculation and normalization. The raw results were saved as GPR files and further processed using the BioConductor R packages *limma* and *marray* (<http://www.bioconductor.org>). To achieve background correction the respective *limma* function was used to subtract the background intensity from the foreground intensity for each spot. Afterwards, the Cy5/Cy3 or Cy3/Cy5 (for dye-swap experiments) ratios for each background-subtracted spot were calculated and normalized. The purpose of normalization is to remove sources of systemic variation in the measured fluorescence intensities that might influence the differential expression analysis (e.g. different labelling efficiencies, different scanning parameters or scanning properties). Here, we used the R package *marray* to perform loess normalization followed by diagnostic-plot generation (volcano plots, boxplots, MA-plots). Normalized ratios of medians reflecting the relative mRNA level were filtered for a signal-to-noise ratio ((F635Median/B635Median) or (F532Median/B532Median)) higher than 3.

Differential gene expression analysis. For further analysis, loess-normalized data were stored in the in-house DNA microarray database⁵⁶ including metadata about each experiment. The in-house DNA microarray database is a custom-made database based on MySQL running on a Linux host combined with a graphical user interface (self-made Java applications) allowing display and analysis of stored microarray data⁵⁶. Routinely, replicates were compared with each other and obvious outlier experiments were removed. Afterwards, replicates belonging to the same experimental setup were combined to sets and then used to perform differential gene expression analysis. For this purpose, normalized log-transformed RNA levels from two or more biological replicate experiments (in case of dye-swaps the reciprocal values were used) were averaged and *p*-value calculation based on Student's *t*-test was performed. For each set, only genes showing statistically significant expression ratios (*p*-value ≤ 0.05) were considered for further analysis steps. Overall, 927 individual microarray experiments were combined to 304 sets.

Quality check. Experimental setup, sample preparation, microarray platform, and equipment used for scanning and analysis of the DNA microarrays can influence technical variation. This so-called batch effect can lead to incorrect results and subsequently to misinterpretation of results. To evaluate possible batch effects, we performed hierarchical clustering analysis using R. All results were visualized using the *ggplot2* package in R.

Data Records

The raw data (GPR files) for all 927 experiments alongside relevant metadata and unfiltered differential expression results (normalized ratio data) as a matrix were uploaded to the GEO database⁷⁴.

In addition, the following files are freely available on Zenodo (<https://doi.org/10.5281/zenodo.6842664>)⁷⁵:

- Normalized expression matrix: Normalized signal intensities (for channel Cy5 and Cy3) resulting from the pre-processing of microarrays with the R packages *limma* and *marray* for every experiment.
- Filtered differential expression results: Matrix with significantly expressed genes (≥ 2 -fold up- and ≥ 2 -fold-down-regulated) with an adjusted *p*-value ≤ 0.05 for each set.

Technical Validation

Re-evaluation and categorization of microarray experiments. All downstream analysis steps of expression data rely on the quality of the data. Therefore, we manually re-evaluated and curated all microarray experiments (1,146) that were pre-processed as described in the Methods section and saved in our in-house microarray database before they were uploaded to the GEO database. The following criteria were applied for this purpose (Fig. 2):

- Two or more replicates were available for each experimental setup.
- At least one replicate for each experimental setup was a dye-swap experiment.
- Non-normalized raw data and images were available.
- Every image obtained (see Microarray data processing) was checked manually for overall quality regarding spot quantity and background fluorescence. Only those chips showing a spot coverage of at least 80% and only minor background were included in the next steps.
- Only experiments with metadata that fulfilled the MIAME criteria⁵⁷ were considered for upload.

After re-evaluation, 927 experiments were uploaded to GEO.

Alongside the raw data uploaded to the GEO repository, the following detailed metadata was added: (i) Strain background (for further information about strains and plasmids used in this study see Supplementary Table S1). (ii) Media and other supplements used for cultivation of strains (see Methods for media composition). Usually, cells were harvested during the exponential growth phase (OD₆₀₀ ca. 5) and then used for RNA isolation. Exceptions are stated. (iii) Number of replicates. (iv) The date (month and year) the experiment was performed. (v) The platform that was used for the specific microarray experiment (see Methods for further information). (vi) Additional information on the strains or genes that were deleted/over-expressed with respect to their function. (vii) If the strain/gene was already characterized and published, the associated PubMed ID (PMID) was added. (viii) If the microarray data were already part of a publication, the PMID and (if available) the GEO accession number were added. In order to ensure that the collected metadata are consistent and

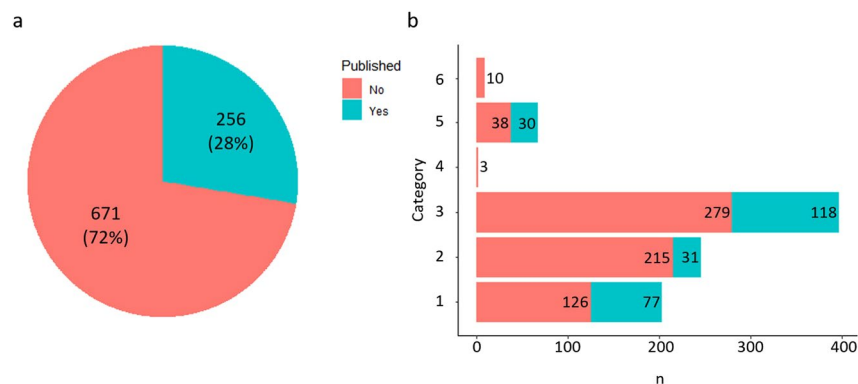


Fig. 3 Overview about publication status and categorization of experiments. **(a)** Of the 927 experiments that are described in this publication, 256 had already been uploaded to GEO and/or are part of a publication in a peer-reviewed scientific journal. **(b)** Number of experiments in the six experimental categories (see section Re-evaluation and categorization of microarray experiments for further information) and publication status. Categories: (1) WT condition A vs. WT condition B. (2) Plasmid-based gene overexpression in parental strain vs. parental strain with empty vector control. (3) Deletion mutant vs. parental strain. (4) Gene silencing strain vs. parental strain. (5) Promoter exchange strain vs. parental strain. (6) Producer strain vs. WT or other producer strain.

standardized, we used unified vocabularies to describe the experimental data. Of the 927 experiments 256 (28%) were already part of a publication (120) and/or uploaded to GEO (136) (Fig. 3a).

To facilitate the finding of specific data sets within this collection (Supplementary Table S2) by the user, they were categorized according to specific criteria: (1) WT condition A vs. WT condition B: experiments in which the transcriptome of *C. glutamicum* wild type (WT) was compared under two different growth conditions A and B. (2) Plasmid-based gene overexpression in parental strain vs. parental strain with empty vector control: experiments in which the influence of plasmid-based overexpression of homologous or heterologous genes in *C. glutamicum* parental strains (WT, deletion mutants or strains with genomically integrated promoter-fusions to genes encoding fluorescence proteins) on the transcriptome was studied by comparison with the parental strain carrying the plasmid vector only. (3) Deletion mutant vs. parental strain: experiments in which the transcriptome of a deletion mutant, i.e. a strain lacking one or several genes in the genome, that can also bear additional changes such as promoter exchanges, was compared with the transcriptome of the parental strain. (4) Gene silencing strain vs. parental strain: experiments in which the transcriptome of a strain with a silenced gene was compared with the transcriptome of the parental strain. (5) Promoter exchange strain vs. parental strain: experiments in which the transcriptome of strains bearing one or more genomically integrated promoter exchanges was compared with the transcriptome of the parental strain. (6) Producer strain vs. WT or other producer strain: experiments in which the transcriptome of a genetically engineered producer strain was compared with the transcriptome of the WT or another producer strain. An overview about the number of experiments that were associated with the described categories as well as their publication status can be found in Fig. 3b.

Overview about microarray sets and experiments with respect to regulated genes and functional categorization. Differentially regulated genes that are at least 2-fold up- or down-regulated (p -value ≤ 0.05) within microarray sets are of particular interest for scientific analysis to identify, for example, the regulon of a transcription factor or regulatory networks. Among the 304 sets of microarray comparisons described here, differentially regulated genes (≥ 2 -fold) were found for 291 sets. The large majority of these sets (280) comprised less than 500 regulated genes out of 3,047 annotated genes, whereas 10 sets comprised between 500 and 1,000 regulated genes and one set showed an even larger number ($>1,000$) of regulated genes (Fig. 4). The latter two groups of sets and those without more than 2-fold regulated genes are listed in Supplementary Table S3. The trend that more than 90% of the sets contained less than 500 regulated genes is also reflected within five of the six categories described above.

In a previous study, the *C. glutamicum* genes were sorted into 22 functional groups⁷⁶. An analysis of the percentage of regulated genes in these groups based on our microarray data revealed that between 12% and 50% were differentially regulated (Fig. 5). When comparing ≥ 2 -fold up-regulated genes (5%–21%) with ≥ 2 -fold down-regulated genes (7%–29%), it becomes obvious that in all 22 functional groups the number of both kinds of differentially regulated genes is very similar. With 50% the functional group “Carbon source transport and metabolism”⁷⁷ has by far the highest proportion of differentially regulated genes. The smallest fraction of differentially regulated genes (12%) was found in the functional group “Cell division, chromosome partitioning”.

When analyzing the allocation to the 22 functional groups of those genes that were at least 30-times among the top 20 of regulated genes within all sets, genes involved in transport of carbon, inorganic ions and further metabolites were comparatively often differentially regulated (Supplementary Table S4, Supplementary Table S5).

To analyse the overall correlation between the sets, we performed hierarchical clustering of differentially regulated genes across all sets (Fig. 6). Overall, this analysis showed that several clusters could be identified and

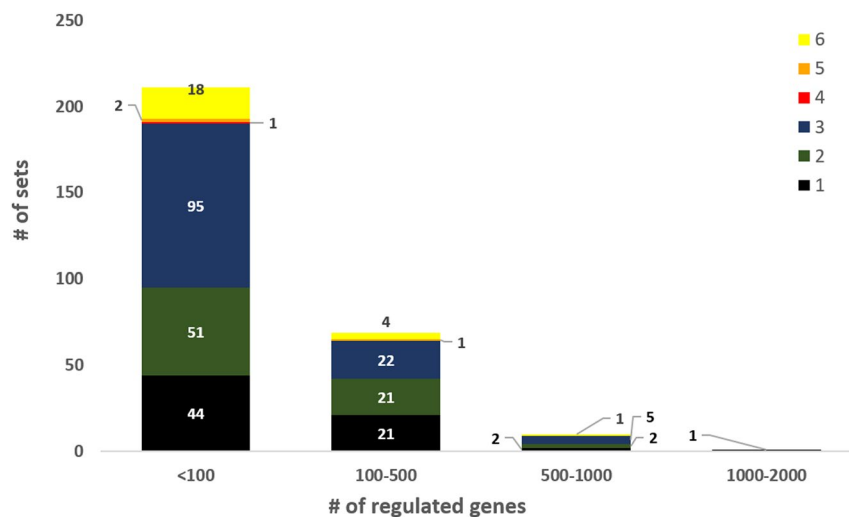


Fig. 4 Number of sets with <100, 100–500, 500–1000, 1000–2000 and >2000 regulated genes (>2-fold up- and >2-fold-down-regulated, p -value ≤ 0.05) with respect to their categorization (see section Re-evaluation and categorization of microarray experiments for further information). Categories: (1) WT condition A vs. WT condition B. (2) Plasmid-based gene overexpression in parental strain vs. parental strain with empty vector control. (3) Deletion mutant vs. parental strain. (4) Gene silencing strain vs. parental strain. (5) Promoter exchange strain vs. parental strain. (6) Producer strain vs. WT or other producer strain.



Fig. 5 Proportion of genes associated to the functional categories⁷⁶ that are ≥ 2 -fold up-regulated (blue), ≥ 2 -fold down-regulated (orange) or unchanged (grey) among the microarray sets described. The numbers at the right side indicate the genes allocated to the corresponding functional category.

no obvious outlier sets were observed. In order to check the reliability of the analysis, we took a closer look at selected clusters with respect to their biological context. One cluster (purple bar) comprised experiments analysing the effects of protein secretion with different signal peptides and target proteins, which revealed differential expression of genes related to cell envelope stress⁷⁸. A second cluster (green bar) involved experiments studying the influence of the DNA-intercalating agent mitomycin C, causing differential expression of a large set of stress response genes⁷⁷. A third cluster (orange) comprised experiments analysing the effects of varying iron concentrations, in which the genes of the DtxR and RipA regulons were identified⁷⁸. In summary, the correlation analysis showed no obvious batch effects, because many clusters comprised sets involving the usage of microarray chips from different platforms for the hybridization, the respective experiments were performed at different time points during the data collection period and/or the biological context favoured this cluster (Supplementary Table S6).

Usage Notes

Here, we provide the first manually curated compendium of expression profiles for the microbial cell factory *C. glutamicum* comprising 927 high-quality microarray experiments that were summarized in 304 sets comprising the mean values for at least two replicates per experimental setup.

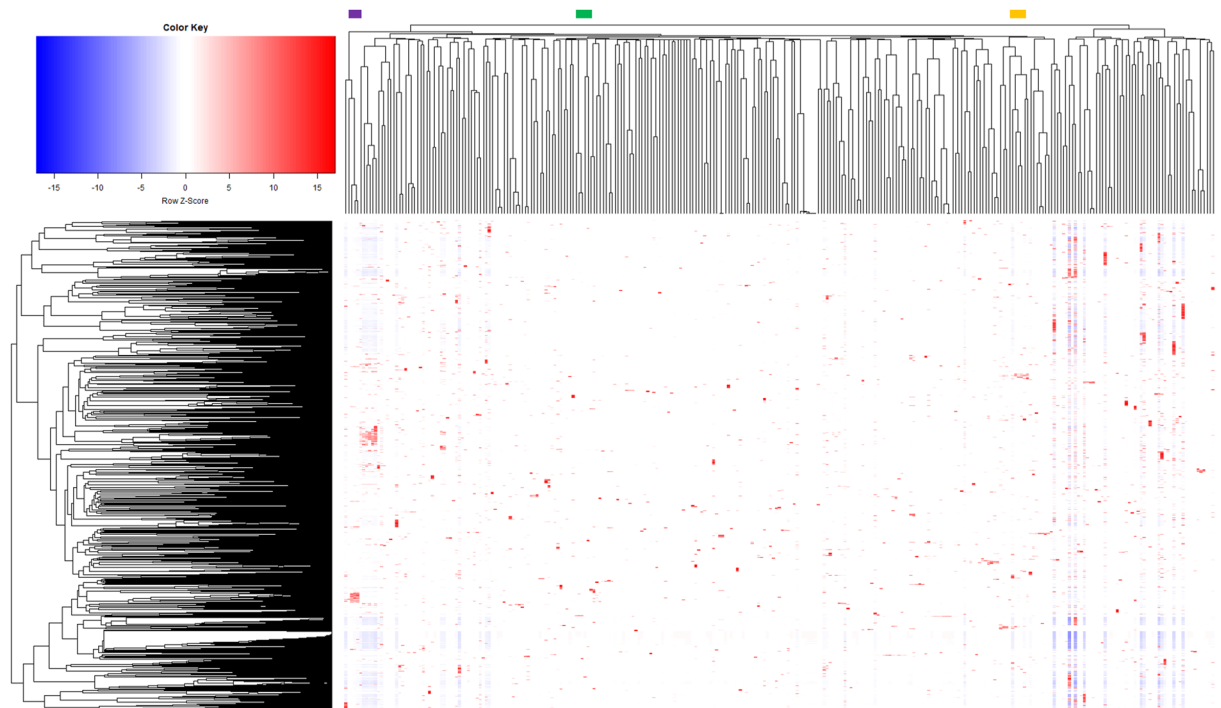


Fig. 6 Hierarchical clustering of differentially regulated genes across sets. Unsupervised hierarchical clustering was performed using median log₂ values of the 3021 genes that were differentially expressed (p -value ≤ 0.05) at least 2-fold within strain comparison, with uncentered Pearson correlations and complete clustering. Example clusters are highlighted with different coloured bars on top and are explained in the text.

The availability of data for different strains as well as conditions and the resulting high variability among the amount of differentially regulated genes offers a valuable and rich resource for further analysis. For instance, it can be used to perform gene co-expression analysis to predict gene functions that are so far unknown⁷⁹. Furthermore, our data collection can easily be integrated with other transcriptomic data (e.g. RNA-Seq or ChAP-Seq) to systematically identify transcriptional regulatory interactions by applying for example machine learning methods^{53–55,80}. In addition, the data can also be used to study the perturbome of *C. glutamicum* to identify a set of core genes that are responsible for a common stress response towards different kind of perturbations⁸¹.

In summary, our manually curated compendium of expression profiles will have a high value for the community by contributing to the further systemic understanding and biotechnological exploitation of *C. glutamicum*.

Code availability

Raw data files of fluorescence images were analyzed by quantitative Image analysis using the GenePix image analysis software (GenePix Pro 6.0; <https://axon-genepix-pro.software.informer.com/6.0/>). In order to assign fluorescence signals to annotated genes gene array list (GAL) files were used (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE169361>)⁷⁴. Processing of the raw data obtained (GPR-files) was performed with the BioConductor R-packages limma version 3.14 (<https://bioconductor.org/packages/release/bioc/html/limma.html>) and marray version 3.14 (<https://bioconductor.org/packages/release/bioc/html/marray.html>) to achieve background correction of spot intensities, ratio calculation/normalization, and diagnostic-plot generation for array quality control. Statistical analysis steps were performed within our in-house microarray database as well as with Excel (Microsoft).

Received: 22 February 2022; Accepted: 18 August 2022;

Published online: 01 October 2022

References

- Kinoshita, S., Udaka, S. & Shimono, M. Studies on the amino acid fermentation. Part 1. Production of L-glutamic acid by various microorganisms. *The Journal of General and Applied Microbiology* **3**, 193–205, <https://doi.org/10.2323/jgam.3.193> (1957).
- Eggeling, L. & Bott, M. A giant market and a powerful metabolism: L-lysine provided by *Corynebacterium glutamicum*. *Appl Microbiol Biotechnol* **99**, 3387–3394, <https://doi.org/10.1007/s00253-015-6508-2> (2015).
- van Ooyen, J., Noack, S., Bott, M., Reth, A. & Eggeling, L. Improved L-lysine production with *Corynebacterium glutamicum* and systemic insight into citrate synthase flux and activity. *Biotechnol Bioeng* **109**, 2070–2081, <https://doi.org/10.1002/bit.24486> (2012).
- Becker, J., Zelder, O., Häfner, S., Schröder, H. & Wittmann, C. From zero to hero – design-based systems metabolic engineering of *Corynebacterium glutamicum* for L-lysine production. *Metab Eng* **13**, 159–168, <https://doi.org/10.1016/j.ymben.2011.01.003> (2011).
- Jin, C. & Bao, J. Lysine Production by Dry Biorefining of Wheat Straw and Cofermentation of *Corynebacterium glutamicum*. *J Agric Food Chem* **69**, 1900–1906, <https://doi.org/10.1021/acs.jafc.0c07902> (2021).
- Wendisch, V. F. Metabolic engineering advances and prospects for amino acid production. *Metab Eng* **58**, 17–34, <https://doi.org/10.1016/j.ymben.2019.03.008> (2020).

7. Vogt, M. *et al.* Pushing product formation to its limit: metabolic engineering of *Corynebacterium glutamicum* for L-leucine overproduction. *Metab Eng* **22**, 40–52, <https://doi.org/10.1016/j.ymben.2013.12.001> (2014).
8. Vogt, M. *et al.* The contest for precursors: channelling L-isoleucine synthesis in *Corynebacterium glutamicum* without byproduct formation. *Appl Microbiol Biotechnol* **99**, 791–800, <https://doi.org/10.1007/s00253-014-6109-5> (2015).
9. Mahr, R. *et al.* Biosensor-driven adaptive laboratory evolution of L-valine production in *Corynebacterium glutamicum*. *Metab Eng* **32**, 184–194, <https://doi.org/10.1016/j.ymben.2015.09.017> (2015).
10. Blombach, B., Arndt, A., Aucther, M. & Eikmanns, B. J. L-valine production during growth of pyruvate dehydrogenase complex-deficient *Corynebacterium glutamicum* in the presence of ethanol or by inactivation of the transcriptional regulator SugR. *Appl Environ Microbiol* **75**, 1197–1200, <https://doi.org/10.1128/AEM.02351-08> (2009).
11. Han, G. *et al.* Improvement of L-Valine Production by Atmospheric and Room Temperature Plasma Mutagenesis and High-Throughput Screening in *Corynebacterium glutamicum*. *ACS Omega* **5**, 4751–4758, <https://doi.org/10.1021/acsomega.9b02747> (2020).
12. Wiechert, J., Gätgens, C., Wirtz, A. & Frunzke, J. Inducible Expression Systems Based on Xenogeneic Silencing and Counter-Silencing and Design of a Metabolic Toggle Switch. *ACS Synth Biol* **9**, 2023–2038, <https://doi.org/10.1021/acssynbio.0c00111> (2020).
13. Jiang, Y., Sheng, Q., Wu, X. Y., Ye, B. C. & Zhang, B. L-arginine production in *Corynebacterium glutamicum*: manipulation and optimization of the metabolic process. *Crit Rev Biotechnol* **41**, 172–185, <https://doi.org/10.1080/07388551.2020.1844625> (2021).
14. Schwentner, A. *et al.* Modular systems metabolic engineering enables balancing of relevant pathways for L-histidine production with *Corynebacterium glutamicum*. *Biotechnol Biofuels* **12**, 65, <https://doi.org/10.1186/s13068-019-1410-2> (2019).
15. Kulis-Horn, R. K., Persicke, M. & Kalinowski, J. *Corynebacterium glutamicum* ATP-phosphoribosyl transferases suitable for L-histidine production—Strategies for the elimination of feedback inhibition. *J Biotechnol* **206**, 26–37, <https://doi.org/10.1016/j.jbiotec.2015.04.001> (2015).
16. Ikeda, M. & Nakagawa, S. The *Corynebacterium glutamicum* genome: features and impacts on biotechnological processes. *Appl Microbiol Biotechnol* **62**, 99–109, <https://doi.org/10.1007/s00253-003-1328-1> (2003).
17. Kalinowski, J. *et al.* The complete *Corynebacterium glutamicum* ATCC 13032 genome sequence and its impact on the production of L-aspartate-derived amino acids and vitamins. *J Biotechnol* **104**, 5–25, [https://doi.org/10.1016/s0168-1656\(03\)00154-8](https://doi.org/10.1016/s0168-1656(03)00154-8) (2003).
18. Wieschalka, S., Blombach, B., Bott, M. & Eikmanns, B. J. Bio-based production of organic acids with *Corynebacterium glutamicum*. *Microb Biotechnol* **6**, 87–102, <https://doi.org/10.1111/1751-7915.12013> (2013).
19. Kallscheuer, N. & Marienhagen, J. *Corynebacterium glutamicum* as platform for the production of hydroxybenzoic acids. *Microb Cell Fact* **17**, 70, <https://doi.org/10.1186/s12934-018-0923-x> (2018).
20. Kogure, T., Suda, M., Hiraga, K. & Inui, M. Protocatechuate overproduction by *Corynebacterium glutamicum* via simultaneous engineering of native and heterologous biosynthetic pathways. *Metab Eng* **65**, 232–242, <https://doi.org/10.1016/j.ymben.2020.11.007> (2021).
21. Jojima, T. *et al.* Metabolic engineering for improved production of ethanol by *Corynebacterium glutamicum*. *Appl Microbiol Biotechnol* **99**, 1165–1172, <https://doi.org/10.1007/s00253-014-6223-4> (2015).
22. Hasegawa, S., Jojima, T., Suda, M. & Inui, M. Isobutanol production in *Corynebacterium glutamicum*: Suppressed succinate by-production by *pckA* inactivation and enhanced productivity via the Entner-Doudoroff pathway. *Metab Eng* **59**, 24–35, <https://doi.org/10.1016/j.ymben.2020.01.004> (2020).
23. Blombach, B. & Eikmanns, B. J. Current knowledge on isobutanol production with *Escherichia coli*, *Bacillus subtilis* and *Corynebacterium glutamicum*. *Bioeng Bugs* **2**, 346–350, <https://doi.org/10.4161/bbug.2.6.17845> (2011).
24. Vogt, M., Brüsseler, C., Ooyen, J. V., Bott, M. & Marienhagen, J. Production of 2-methyl-1-butanol and 3-methyl-1-butanol in engineered *Corynebacterium glutamicum*. *Metab Eng* **38**, 436–445, <https://doi.org/10.1016/j.ymben.2016.10.007> (2016).
25. Heider, S. A. & Wendisch, V. F. Engineering microbial cell factories: Metabolic engineering of *Corynebacterium glutamicum* with a focus on non-natural products. *Biotechnol J* **10**, 1170–1184, <https://doi.org/10.1002/biot.201400590> (2015).
26. Kallscheuer, N. *et al.* Construction of a *Corynebacterium glutamicum* platform strain for the production of stilbenes and (2S)-flavanones. *Metab Eng* **38**, 47–55, <https://doi.org/10.1016/j.ymben.2016.06.003> (2016).
27. Milke, L., Aschenbrenner, J., Marienhagen, J. & Kallscheuer, N. Production of plant-derived polyphenols in microorganisms: current state and perspectives. *Appl Microbiol Biotechnol* **102**, 1575–1585, <https://doi.org/10.1007/s00253-018-8747-5> (2018).
28. Wolf, S. *et al.* Advances in metabolic engineering of *Corynebacterium glutamicum* to produce high-value active ingredients for food, feed, human health, and well-being. *Essays Biochem* **65**, 197–212, <https://doi.org/10.1042/EBC20200134> (2021).
29. Freudl, R. Beyond amino acids: Use of the *Corynebacterium glutamicum* cell factory for the secretion of heterologous proteins. *J Biotechnol* **258**, 101–109, <https://doi.org/10.1016/j.jbiotec.2017.02.023> (2017).
30. Bakkes, P. J. *et al.* Improved pEKEx2-derived expression vectors for tightly controlled production of recombinant proteins in *Corynebacterium glutamicum*. *Plasmid* **112**, 102540, <https://doi.org/10.1016/j.plasmid.2020.102540> (2020).
31. Kirchner, O. & Tauch, A. Tools for genetic engineering in the amino acid-producing bacterium *Corynebacterium glutamicum*. *J Biotechnol* **104**, 287–299, [https://doi.org/10.1016/s0168-1656\(03\)00148-2](https://doi.org/10.1016/s0168-1656(03)00148-2) (2003).
32. Kortmann, M., Kuhl, V., Klaffl, S. & Bott, M. A chromosomally encoded T7 RNA polymerase-dependent gene expression system for *Corynebacterium glutamicum*: construction and comparative evaluation at the single-cell level. *Microb Biotechnol* **8**, 253–265, <https://doi.org/10.1111/1751-7915.12236> (2015).
33. Li, C., Swofford, C. A., Rückert, C. & Sinskey, A. J. Optimizing recombinering in *Corynebacterium glutamicum*. *Biotechnol Bioeng*, <https://doi.org/10.1002/bit.27737> (2021).
34. Tenhaef, N., Stella, R., Frunzke, J. & Noack, S. Automated Rational Strain Construction Based on High-Throughput Conjugation. *ACS Synth Biol* <https://doi.org/10.1021/acssynbio.0c00599> (2021).
35. Kjeldsen, K. R. & Nielsen, J. In silico genome-scale reconstruction and validation of the *Corynebacterium glutamicum* metabolic network. *Biotechnol Bioeng* **102**, 583–597, <https://doi.org/10.1002/bit.22067> (2009).
36. Brinkrolf, K., Brune, I. & Tauch, A. The transcriptional regulatory network of the amino acid producer *Corynebacterium glutamicum*. *J Biotechnol* **129**, 191–211, <https://doi.org/10.1016/j.jbiotec.2006.12.013> (2007).
37. Toyoda, K. & Inui, M. Regulons of global transcription factors in *Corynebacterium glutamicum*. *Appl Microbiol Biotechnol* **100**, 45–60, <https://doi.org/10.1007/s00253-015-7074-3> (2016).
38. Schröder, J. & Tauch, A. Transcriptional regulation of gene expression in *Corynebacterium glutamicum*: the role of global, master and local regulators in the modular and hierarchical gene regulatory network. *FEMS Microbiol Rev* **34**, 685–737, <https://doi.org/10.1111/j.1574-6976.2010.00228.x> (2010).
39. Bott, M. & Brocker, M. Two-component signal transduction in *Corynebacterium glutamicum* and other corynebacteria: on the way towards stimuli and targets. *Appl Microbiol Biotechnol* **94**, 1131–1150, <https://doi.org/10.1007/s00253-012-4060-x> (2012).
40. Pátek, M. & Nesvera, J. Sigma factors and promoters in *Corynebacterium glutamicum*. *J Biotechnol* **154**, 101–113, <https://doi.org/10.1016/j.jbiotec.2011.01.017> (2011).
41. Myers, K. S., Park, D. M., Beauchene, N. A. & Kiley, P. J. Defining bacterial regulons using ChIP-seq. *Methods* **86**, 80–88, <https://doi.org/10.1016/j.ymeth.2015.05.022> (2015).
42. Harada, R. & Nepveu, A. Chromatin affinity purification. *Methods Mol Biol* **809**, 237–253, https://doi.org/10.1007/978-1-61779-376-9_16 (2012).
43. Parise, M. T. D. *et al.* CoryneRegNet 7, the reference database and analysis platform for corynebacterial gene regulatory networks. *Sci Data* **7**, 142, <https://doi.org/10.1038/s41597-020-0484-9> (2020).

44. Heyer, A. *et al.* The two-component system ChrSA is crucial for haem tolerance and interferes with HrrSA in haem-dependent gene regulation in *Corynebacterium glutamicum*. *Microbiology (Reading)* **158**, 3020–3031, <https://doi.org/10.1099/mic.0.062638-0> (2012).
45. Koçan, M. *et al.* Two-component systems of *Corynebacterium glutamicum*: deletion analysis and involvement of the PhoS-PhoR system in the phosphate starvation response. *J Bacteriol* **188**, 724–732, <https://doi.org/10.1128/JB.188.2.724-732.2006> (2006).
46. Keppel, M. *et al.* HrrSA orchestrates a systemic response to heme and determines prioritization of terminal cytochrome oxidase expression. *Nucleic Acids Res* **48**, 6547–6562, <https://doi.org/10.1093/nar/gkaa415> (2020).
47. Kleine, B. *et al.* The three-component system EsrISR regulates a cell envelope stress response in *Corynebacterium glutamicum*. *Mol Microbiol* **106**, 719–741, <https://doi.org/10.1111/mmi.13839> (2017).
48. Busche, T., Šilar, R., Pičmanová, M., Pátek, M. & Kalinowski, J. Transcriptional regulation of the operon encoding stress-responsive ECF sigma factor SigH and its anti-sigma factor RshA, and control of its regulatory network in *Corynebacterium glutamicum*. *BMC Genomics* **13**, 445, <https://doi.org/10.1186/1471-2164-13-445> (2012).
49. Taniguchi, H. *et al.* Physiological roles of sigma factor SigD in *Corynebacterium glutamicum*. *BMC Microbiol* **17**, 158, <https://doi.org/10.1186/s12866-017-1067-6> (2017).
50. Dostálová, H. *et al.* Overlap of Promoter Recognition Specificity of Stress Response Sigma Factors SigD and SigH in *Corynebacterium glutamicum* ATCC 13032. *Front Microbiol* **9**, 3287, <https://doi.org/10.3389/fmicb.2018.03287> (2018).
51. Baumbach, J., Brinkrolf, K., Czaja, L. F., Rahmann, S. & Tauch, A. CoryneRegNet: an ontology-based data warehouse of corynebacterial transcription factors and regulatory networks. *BMC Genomics* **7**, 24, <https://doi.org/10.1186/1471-2164-7-24> (2006).
52. Faith, J. J. *et al.* Large-scale mapping and validation of *Escherichia coli* transcriptional regulation from a compendium of expression profiles. *PLoS Biol* **5**, e8, <https://doi.org/10.1371/journal.pbio.0050008> (2007).
53. Yousef, M., Kumar, A. & Bakir-Gungor, B. Application of Biological Domain Knowledge Based Feature Selection on Gene Expression Data. *Entropy (Basel)* **23**, <https://doi.org/10.3390/e23010002> (2020).
54. Toro-Dominguez, D. *et al.* A survey of gene expression meta-analysis: methods and applications. *Brief Bioinform* <https://doi.org/10.1093/bib/bba019> (2020).
55. Frawley, W. J., Piatetsky-Shapiro, G., & Matheus, C. J. Knowledge Discovery in Databases: An Overview. *AI Magazine* **13** (1992).
56. Polen, T. & Wendisch, V. F. Genomewide expression analysis in amino acid-producing bacteria using DNA microarrays. *Appl Biochem Biotechnol* **118**, 215–232, <https://doi.org/10.1385/abab:118:1-3:215> (2004).
57. Brazma, A. Minimum Information About a Microarray Experiment (MIAME)—successes, failures, challenges. *ScientificWorldJournal* **9**, 420–423, <https://doi.org/10.1100/tsw.2009.57> (2009).
58. Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**, 160018, <https://doi.org/10.1038/sdata.2016.18> (2016).
59. Wilkinson, M. D. *et al.* Addendum: The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **6**, 6, <https://doi.org/10.1038/s41597-019-0009-6> (2019).
60. Gao, B. & Gupta, R. S. Phylogenetic framework and molecular signatures for the main clades of the phylum *Actinobacteria*. *Microbiol Mol Biol Rev* **76**, 66–112, <https://doi.org/10.1128/MMBR.05011-11> (2012).
61. Turkarlan, S. *et al.* A comprehensive map of genome-wide gene regulation in *Mycobacterium tuberculosis*. *Sci Data* **2**, 150010, <https://doi.org/10.1038/sdata.2015.10> (2015).
62. Schäfer, A. *et al.* Small mobilizable multi-purpose cloning vectors derived from the *Escherichia coli* plasmids pK18 and pK19: selection of defined deletions in the chromosome of *Corynebacterium glutamicum*. *Gene* **145**, 69–73, [https://doi.org/10.1016/0378-1119\(94\)90324-7](https://doi.org/10.1016/0378-1119(94)90324-7) (1994).
63. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* **6**, 343–345, <https://doi.org/10.1038/nmeth.1318> (2009).
64. Bertani, G. Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*. *J Bacteriol* **62**, 293–300, <https://doi.org/10.1128/JB.62.3.293-300.1951> (1951).
65. Keilhauer, C., Eggeling, L. & Sahm, H. Isoleucine synthesis in *Corynebacterium glutamicum*: molecular analysis of the *ilvB-ilvN-ilvC* operon. *J Bacteriol* **175**, 5595–5603, <https://doi.org/10.1128/jb.175.17.5595-5603.1993> (1993).
66. van der Rest, M. E., Lange, C. & Molenaar, D. A heat shock following electroporation induces highly efficient transformation of *Corynebacterium glutamicum* with xenogeneic plasmid DNA. *Appl Microbiol Biotechnol* **52**, 541–545, <https://doi.org/10.1007/s002530051557> (1999).
67. Hemmerich, J. *et al.* Screening of a genome-reduced *Corynebacterium glutamicum* strain library for improved heterologous cutinase secretion. *Microb Biotechnol* **13**, 2020–2031, <https://doi.org/10.1111/1751-7915.13660> (2020).
68. Litsanov, B., Brocker, M. & Bott, M. Toward homosuccinate fermentation: metabolic engineering of *Corynebacterium glutamicum* for anaerobic production of succinate from glucose and formate. *Appl Environ Microbiol* **78**, 3325–3337, <https://doi.org/10.1128/AEM.07790-11> (2012).
69. Michel, A., Koch-Koerfges, A., Krumbach, K., Brocker, M. & Bott, M. Anaerobic growth of *Corynebacterium glutamicum* via mixed-acid fermentation. *Appl Environ Microbiol* **81**, 7496–7508, <https://doi.org/10.1128/AEM.02413-15> (2015).
70. Wendisch, V. F. *et al.* Isolation of *Escherichia coli* mRNA and comparison of expression using mRNA and total RNA on DNA microarrays. *Anal Biochem* **290**, 205–213, <https://doi.org/10.1006/abio.2000.4982> (2001).
71. Lange, C., Rittmann, D., Wendisch, V. F., Bott, M. & Sahm, H. Global expression profiling and physiological characterization of *Corynebacterium glutamicum* grown in the presence of L-valine. *Appl Environ Microbiol* **69**, 2521–2532, <https://doi.org/10.1128/aem.69.5.2521-2532.2003> (2003).
72. Frunzke, J., Engels, V., Hasenbein, S., Gätgens, C. & Bott, M. Co-ordinated regulation of gluconate catabolism and glucose uptake in *Corynebacterium glutamicum* by two functionally equivalent transcriptional regulators, GntR1 and GntR2. *Mol Microbiol* **67**, 305–322, <https://doi.org/10.1111/j.1365-2958.2007.06020.x> (2008).
73. Polen, T., Krämer, M., Bongaerts, J., Wubbolts, M. & Wendisch, V. F. The global gene expression response of *Escherichia coli* to L-phenylalanine. *J Biotechnol* **115**, 221–237, <https://doi.org/10.1016/j.jbiotec.2004.08.017> (2005).
74. Kranz, A., Polen, T. & Bott, M. A compendium of expression profiles for *Corynebacterium glutamicum* ATCC13032. *GEO* <https://identifiers.org/geo:GSE169361> (2022).
75. Kranz, A., Polen, T. & Bott, M. A compendium of expression profiles for *Corynebacterium glutamicum*. *Zenodo*. <https://doi.org/10.5281/zenodo.6842664> (2022).
76. Baumgart, M. *et al.* *Corynebacterium glutamicum* Chassis C1*: Building and Testing a Novel Platform Host for Synthetic Biology and Industrial Biotechnology. *ACS Synth Biol* **7**, 132–144, <https://doi.org/10.1021/acssynbio.7b00261> (2018).
77. Frunzke, J., Bramkamp, M., Schweitzer, J. E. & Bott, M. Population Heterogeneity in *Corynebacterium glutamicum* ATCC 13032 caused by prophage CGP3. *J Bacteriol* **190**, 5111–5119, <https://doi.org/10.1128/JB.00310-08> (2008).
78. Jurischka, S. *et al.* A secretion biosensor for monitoring Sec-dependent protein export in *Corynebacterium glutamicum*. *Microb Cell Fact* **19**, 11, <https://doi.org/10.1186/s12934-019-1273-z> (2020).
79. van Dam, S., Vösa, U., van der Graaf, A., Franke, L. & de Magalhães, J. P. Gene co-expression analysis for functional classification and gene-disease predictions. *Brief Bioinform* **19**, 575–592, <https://doi.org/10.1093/bib/bbw139> (2018).
80. Parise, D. *et al.* On the Consistency between Gene Expression and the Gene Regulatory Network of *Corynebacterium glutamicum*. *Netw Syst Med* **4**, 51–59, <https://doi.org/10.1089/nsm.2020.0014> (2021).
81. Molina Mora, J. A. *et al.* A first perturbome of *Pseudomonas aeruginosa*: Identification of core genes related to multiple perturbations by a machine learning approach. *Biosystems* **205**, 104411, <https://doi.org/10.1016/j.biosystems.2021.104411> (2021).

Acknowledgements

The authors would like to thank Prof. Dr. Björn Usadel (IBG-4: Bioinformatics, Forschungszentrum Jülich) for supporting this work.

Author contributions

A.Kra. contributed to study design and concept, performed re-evaluation, manual curation, and categorization of the DNA microarray data, contributed to the upload of the data to GEO, and wrote the manuscript. T.P. contributed to manufacturing of DNA microarrays, designed, and established the in-house microarray database, supervised, and contributed to data acquisition and analysis, and performed upload of the data to GEO. C.K. improved the in-house microarray database. A.A., G.B., M.Bu., A.Ch., A.Cr., C.D., U.D., R.D., K.G., C.Gä., T.G., C.Ge., S.H., A.H., M.H., T.I., A.Ka., N.K., L.K., S.K., B.K., M.K., A.K.-K., K.K., E.K., A.Kru., A.Krü., A.Kü., M.L., C.L., C.M., T.M., R.M., S.M., A.M., X.M., O.M., A.N., J.N., J.P., E.P., L.P., P.R., S.S., S.Sche., S.Sp., J.S., J.E.S., G.Si., U.S., M.S., C.S., A.T., J.v.O., P.v.S.-W., A.V., M.V., R.F.v.B., S.W. and L.Z. contributed to data acquisition and analysis. D.R. contributed to manufacturing of DNA microarrays and contributed to data acquisition and analysis. L.E., B.J.E., R.F., J.M., M.O. and G.Sc. supervised data acquisition and analysis. M.Ba., M.Br., J.F. and V.F.W. supervised and contributed to data acquisition and analysis. M.Bo. designed and conceptualized the study, supervised data acquisition and analysis, and revised the manuscript. All authors read and approved the manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41597-022-01706-7>.

Correspondence and requests for materials should be addressed to Angela Kranz or Michael Bott.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022