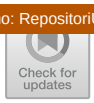

# Towards the Reconstruction of Integrated Genome-Scale Models of Metabolism and Gene Expression

Fernando Cruz[1(✉)], Diogo Lima[1], José P. Faria[2], Miguel Rocha[1], and Oscar Dias[1(✉)]

[1] Centre of Biological Engineering, University of Minho, 4710-057 Braga, Portugal
{fernando.cruz, odias}@ceb.uminho.pt, diogolima95@hotmail.com, mrocha@di.uminho.pt
[2] Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL, USA
jplfaria@anl.gov

**Abstract.** The reconstruction of integrated genome-scale models of metabolism and gene expression has been a challenge for a while now. In fact, various methods that allow integrating reconstructions of Transcriptional Regulatory Networks, gene expression data or both into Genome-Scale Metabolic Models have been proposed. Several of these methods are surveyed in this article, which allowed identifying their strengths and weaknesses concerning the reconstruction of integrated models for multiple prokaryotic organisms. Additionally, the main resources of regulatory information were also surveyed, as the existence of novel sources of regulatory information and gene expression data may contribute for the improvement of methodologies referred herein.

**Keywords:** Genome-Scale Metabolic Models · Genome-scale models of metabolism and gene expression · Regulation of gene expression · Databases of regulatory information

## 1 Control of Gene Expression in Prokaryotes

The optimal composition of the proteome in prokaryotes and eukaryotes changes considerably over time. In prokaryotic organisms, these changes often reflect the cell response to an ever-changing environment. Hence, the regulation of the gene expression is pivotal for controlling the optimal cellular composition of the proteome as a function of the consecutive environmental conditions.

The control of gene expression can occur at several potential stages of regulation [1]. Nevertheless, this review focuses primarily on the regulation of transcription initiation, as it likely is the main control stage of gene expression, in prokaryotic cells [2]. In addition, this study emphasizes the control of gene expression associated with the regulation of the cell metabolism.

Transcription is initiated when the holoenzyme RNA polymerase binds to a specific region of DNA known as the promoter [2]. Although there are consensus sequences for


F. Fdez-Riverola et al. (Eds.): PACBB 2019, AISC 1005, pp. 173–181, 2020.
https://doi.org/10.1007/978-3-030-23873-5_21

many promoters, these may vary considerably within the genome. As a result, the binding affinity of RNA polymerase is affected, and consequently the rate at which transcription is initiated. Due to this control over the initiation of transcription, these DNA sequences are often classified as strong or weak promoters.

Many of the principles assumed in the regulation of prokaryotic gene expression are based on the fact that genes are clustered into operons, which are regulated together [2]. In prokaryotes, genes are placed linearly and sequentially. A single functional mRNA molecule contains the information for the synthesis of multiple related proteins. A well-known example is the *lac* operon [3].

Considering the operons, these primary units of regulation of gene expression often comprise additional regulatory DNA sequences. The so-called cis-acting elements, also referred to Transcription Factor Binding Sites (TFBS), which are specific sites where gene transcription regulatory proteins bind to, directly or indirectly affect the initiation of transcription [2]. Also known as regulators or transcription factors (TFs), regulatory proteins are trans-acting elements that either induce or repress the expression of a given gene. A given regulator might coordinate the regulation of many operons. A network of operons with a common regulator is so called a regulon [1].

Besides the fundamental biological machinery described above, there are many other regulatory mechanisms for controlling gene expression in prokaryotes, such as transcriptional attenuation and gene regulation by recombination [1]. Yet, the scope of this review only encompasses biological processes that are quantitatively described in literature, databases and methodological approaches.

## 2 Regulatory Information Resources

Considering their type, there are two main resources of regulatory information. A considerable number of databases collect transcriptional data regarding elements of biological machinery that control gene expression. Valuable information regarding promoters, TFs, TFBS and operons among other, is often kept in these databases, once data is retrieved from literature or inferred with comparative genomics tools. On the other hand, a restricted number of databases centralize most gene expression data currently available. The development of high-throughput technologies, such as next-generation sequencing, contributed to the increase in the amount of gene expression data found in these public repositories.

Depending on the methodology used for reconstructing integrated genome-scale models of metabolism and gene expression, one may resort to either transcriptional regulatory data, raw gene expression data or a combination of both.

### 2.1 Databases of Transcriptional Information

Databases of prokaryotic transcriptional information store valuable information on regulatory interactions that take place inside the cell. Information contained in these databases often describes the biological machinery responsible for controlling the gene expression as a function of changes in the environmental conditions.

According to their representativity, databases of regulatory information can be categorized into two groups: organism-specific and non-organism specific. Table S1 of supplementary file 1 (https://bit.ly/2WonbG0) provides the most relevant databases of regulatory information grouped by the corresponding scope.

Comprehensive organism-specific databases are only available for model organisms such as *Escherichia. coli* and *Bacillus subtilis*, or well-known bacteria such as *Mycobacterium tuberculosis*, Gamma-proteobacteria, Mycobacteria and Cyanobacteria. These databases are the result of collaborative task forces aimed at collecting regulatory information on a single organism, which are spread all over literature and other resources, like databases of gene expression data.

For instance, the Database of Transcriptional Regulation in *B. subtilis* (DBTBS) comprises a collection of experimentally validated gene regulatory relations and the corresponding TFBS of the bacterium genes [4]. Recently, the reconstruction of the Transcriptional Regulatory Network (TRN) for *B. subtilis* combined information available in this database with data of less comprehensive databases and a gene expression dataset [5]. This work, by Faria *et al.* [5], also proposed a novel representation of fundamental units of function within a cell called Atomic Regulons (ARs) [6].

Another set of curated regulatory interactions can be obtained for *E. coli* in the RegulonDB [7]. The authors present this database as a unified resource for transcriptional regulation in *E. coli* K-12. In the latest version, an additional effort for incorporating high-throughput-generated binding data was made, extending the understanding of gene expression in the model organism. However, non-model organisms' databases are less comprehensive. Some of these resources comprise regulatory DNA motifs, respective TFs and regulatory networks of less described bacteria.

These databases are gold standards of regulatory information for a single prokaryotic organism. Hence, these should be assessed to infer high-quality TRNs or integrate regulatory information with Genome-Scale Metabolic (GSM) models. Nevertheless, they may lack regulatory interactions found in recent data.

Non-organism specific databases offer limited information. These resources contain information for vast phylogenetic clades, including specific elements of the biological machinery of regulation of gene expression. Comprehensive information, such as the regulatory interactions between TFs and target genes, can, nevertheless, be obtained with comparative genomics approaches.

For example, RegPrecise [8] represents at least 14 taxonomic groups of bacteria, with a collection of transcriptional regulons, determined with comparative genomics approaches, inferred from high-quality manually-curated transcriptional regulatory interactions, namely called regulogs [9]. PRODORIC2 is another database that includes manually curated and unique collection of TFBS for a considerable range of bacteria [10]. Other databases, shown in Table S1, also provide relevant information regarding the regulation of gene expression in prokaryotes, such as putative operons, promoters, TFs and TFBS for multiple species of bacteria. These databases are useful for comparative genomics-based approaches towards the reconstruction of TRN.

### 2.2 Databases of Gene Expression Data

Up to now, the main sources of gene expression data were based on high-throughput transcriptomics technologies, namely microarrays [11] and RNA-seq [12, 13]. Whereas the former contributed for the initial steps of the research in this area, the latter is responsible for a paradigm shift (see below). Other techniques for measuring gene expression level, such as ChIP-chip [14], SAGE [15], or ChIP-seq [16] are worth mentioning as well.

Expression profiling-based techniques such as microarrays [11] and SAGE [15] allow measuring the level of gene expression and quantifying the amount of mRNA, respectively. Besides the nature of these possible outputs, genome binding experiments also provide insights over DNA-protein binding targets [14]. NGS-based technologies have the advantage of being sensitive while providing whole-genome direct measurements of mRNA without previous knowledge of the genome sequence [12, 13, 16]. These techniques are also able to detect transcription starting sites [17].

Functional genomics data repositories, like GEO [18] and ArrayExpress [19], store gene expression data for a wide diversity of organisms, including bacteria. Furthermore, both databases respect the Minimum Information About a Microarray Experiment (MIAME) [20] and provide query and browsing tools for analyzing and retrieving gene expression data. Other databases of gene expression data derived from microarray and RNA-seq experiments are COLOMBOS [21] and M3D [22]. Both databases provide comprehensive compendia of bacterial gene expression normalized and downstream processed data. These databases are of extreme importance for reconstructing novel TRNs or determining sets of co-expressed genes using *de novo* reverse engineering-based approaches. Besides, the mentioned resources of gene expression data have already been used for reconciling gene expression data with GSM models.

GEO and ArrayExpress were surveyed as these are the major sources of gene expression data to date. The type and amount of available expression studies, as well as availability of NGS-based techniques, are summarized in Fig. 1A and B. The distribution of gene expression bacterial data was also retrieved. As shown in the Figure S1-C and S1-D of the supplementary file 1 (https://bit.ly/2WonbG0), GEO was further analyzed by collecting the availability of experimental series throughout the years and determining the most-represented bacterial species, respectively.

This survey shows that most data available in both databases is from expression profiling and transcription profiling studies, with 82094 (70%) and 8150 (66%) experimental series for GEO and ArrayExpress, respectively (Fig. 1A and B). In 2012, NGS-based studies represented approximately 2% of the data available in GEO [23].

Although most expression series are derived from microarray-based studies, 64% and 77% in GEO and ArrayExpress, respectively, as of February 2019, the proportion of NGS studies (36% and 23%) has risen significantly. These numbers are aligned with predictions for high-throughput sequencing techniques [23]. GEO has seen a consistent increase in publicly available gene expression experimental series since 2012, at a rate of approximately 11000 series a year (Figure S1 – C of the supplementary file 1 available at https://bit.ly/2WonbG0).

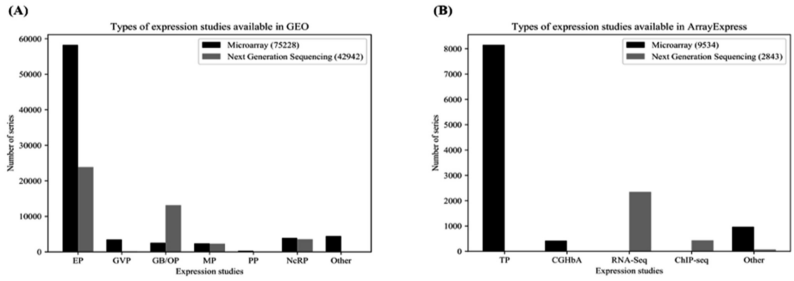


**Fig. 1.** Survey of the GEO and ArrayExpress databases. Types of available expression studies in GEO (**A**) and ArrayExpress (**B**) for a total of 118170 and 12375 series, respectively. EP (Expression profiling); GVP (Genome Variation Profiling); GB/OP (Genome binding/occupancy profiling); MP (Methylation profiling); PP (Protein profiling); NcRNAP (Non-coding RNA profiling); TP (Transcription profiling); CGHbA (Comparative genomic hybridization by array).

Analyzing the amount of experimental series for each species revealed that the proportion of bacterial-associated data is about 8% and 11% in GEO and ArrayExpress, respectively. As depicted in the Figure S1 – D of the supplementary file 1 (https://bit.ly/2WonbG0), the most-represented bacterial species in GEO as well as in ArrayExpress (data not shown) is *E. coli*. Additionally, we found that Proteobacteria, Firmicutes and Actinobacteria are extensively represented phyla, including 18 of the 20 most-represented bacterial species.

## 3 Methods for Integrating TRN Reconstruction or Gene Expression Data in GSM Models

As of 2001, several methods have been developed for assisting in the reconstruction and analysis of integrated genome-scale models of metabolism and gene expression [23–25]. The main theory, type of implementation and major drawbacks associated with these methods were addressed, to understand how they comply regulatory information with metabolism. Table S2 of supplementary file 1 (https://bit.ly/2WonbG0) summarizes their main requirements, implementations and drawbacks.

Contrasting with previous reviews [23–25], this review was extended to include more methods. Blazier and Papin [24] reviewed MADE [26], E-FLUX [27] and PROM [28] by highlighting the methods advantages and limitations. Afterwards, Machado and Herrgård [25] revised and evaluated those methods plus tFBA [29] and the method by Lee *et al.* [30], using two gene expression datasets of *E. coli* and one of *Saccharomyces cerevisiae*. Additionally, rFBA [31], SR-FBA [32], PROM [28] and tFBA [29] were already classified and categorized according to the deviations from traditional phenotype simulation with FBA [23]. Besides of these previous reports, TIGER toolbox [33], GIM$^3$E [34], FlexFlux [35], TRFBA [36], CoRegFlux [37] and ME-models [38] were never surveyed before.

Methods were grouped by the type of implementation, namely whether they integrate TRN reconstructions, gene expression data or both. Regardless of classifying

these methodologies in toolboxes, simple algorithms, computational tools, advantages or drawbacks, the main deviations to the standard constraint-based modeling approach and Flux Balance Analysis (FBA) framework [39] are also presented.

Unlike ME-models [38] and GIM$^3$E [34], all methods addressed in this study allow the simulation of phenotypes for multiple environmental and genetic conditions using integrated genome-scale models of metabolism and transcriptional information. Note that, GIMM$^3$E and ME-models require the utilization of additional omics data such as exometabolomics and proteomics, respectively.

## 4 Discussion

The present study is aimed at highlighting the required resources, features and limitations of the latest efforts towards the reconstruction of integrated genome-scale models of metabolism and gene expression. The existence of new sources of regulatory information and gene expression data (e.g. RNA-seq and ChIP-seq) opens previously closed doors for introducing new methods. Although the main share of gene expression data is still from microarray expression studies, the number of datasets and series obtained by NGS-based technologies is on the rise. Nevertheless, most of the methods surveyed in this article, except the method by Lee *et al.* [30], only used microarray expression data.

None of the methods for integrating gene expression with metabolic models, previously evaluated by Machado and coworkers [25], outperforms each other in phenotype predictions. Furthermore, simple growth maximization with parsimony FBA (pFBA) [40], performed as well as the evaluated methods, namely MADE [26], E-FLUX [27], PROM [28], tFBA [29] and the method by Lee *et al.* [30]. This indicates that the promising results reported by these methods are just mere artifacts.

In fact, most results presented by such tools might be related with rigid constraints created around the nature of the gene expression dataset. Whereas some methods require large-scale gene expression datasets to be robust, others resort to mapping levels of gene expression directly with the reactions bounds which might not be the case for all organisms and datasets. The first methods to ever be developed (RFBA [31] and SR-FBA [32]) limit the solution space by removing possible solutions with Boolean logic. Complex formulations, requirements for large-scale or specific gene expression datasets (that are scarce for some bacteria groups) and incongruences obtained in recent benchmarking tests, pose a hard challenge for using these methods out of the scope they were developed for.

Some of the major drawbacks are not a repercussion of the methodologies itself but rather due to the difficulty in propagating their implementations to other organisms, especially those poorly documented. Hence, the perspective of reconstructing integrated genome-scale models of metabolism and gene expression for diverse prokaryotes rather than well-known organisms is still a complex endeavor. Nevertheless, this gap can be overcome by novel approaches, such as using comparative genomics tools for determining *in silico* regulatory mechanisms that affect metabolism [23].

A user-friendly tool implementing different methods and approaches as a function of the available data would likely shed some light on the reconstruction of these methods. Lastly, reconstructing models that incorporate TRN reconstructions should be more advantageous when compared with only the reconciliation of gene expression data into a new FBA-based formulation. Firstly, the model would provide comprehensive knowledge regarding the metabolic and regulatory events occurring inside the cell. Secondly, the various approaches as well as the amount of data available for reconstructing and integrating TRN into GSM models would ease the diffusion of this approach to most bacteria having a sequenced genome. This hypothetical computational tool would therefore be able to combine different sources of regulatory information available in the resources discussed above, which are rarely combined.

**Acknowledgements.** This study was supported by the Portuguese Foundation for Science and Technology (FCT) under the scope of the strategic funding of UID/BIO/04469/2019 unit and BioTecNorte operation (NORTE-01-0145-FEDER-000004) funded by the European Regional Development Fund under the scope of Norte2020-Programa Operacional Regional do Norte. Fernando Cruz holds a doctoral fellowship (SFRH/BD/139198/2018) funded by the FCT. The authors thank project SHIKIFACTORY100 - Modular cell factories for the production of 100 compounds from the shikimate pathway (814408) funded by the European Commission.

## References

1. Nelson, D.L., Cox, M.M.: Lehninger Principles of Biochemistry. W.H. Freeman, New York (2008)
2. Lodish, H., Berk, A., Zipursky, S.L., Matsudaira, P., Baltimore, D., Darnell, J.: Molecular Cell Biology. W.H. Freeman, New York (2000)
3. Willson, C., Pebrin, D., Cohn, M., Jacob, F., Monod, J.: Non-inducible mutants of the regulator gene in the "lactose" system of Escherichia coli. J. Mol. Biol. **8**, 582–592 (1964)
4. Sierro, N., Makita, Y., de Hoon, M., Nakai, K.: DBTBS: a database of transcriptional regulation in Bacillus subtilis containing upstream intergenic conservation information. Nucleic Acids Res. **36**, D93–D96 (2008)
5. Faria, J.P., Overbeek, R., Taylor, R.C., Conrad, N., Vonstein, V., Goelzer, A., Fromion, V., Rocha, M., Rocha, I., Henry, C.S.: Reconstruction of the regulatory network for bacillus subtilis and reconciliation with gene expression data. Front. Microbiol. **7**, 275 (2016)
6. Faria, J.P., Davis, J.J., Edirisinghe, J.N., Taylor, R.C., Weisenhorn, P., Olson, R.D., Stevens, R.L., Rocha, M., Rocha, I., Best, A.A., DeJongh, M., Tintle, N.L., Parrello, B., Overbeek, R., Henry, C.S.: Computing and applying atomic regulons to understand gene expression and regulation. Front. Microbiol. **7**, 1819 (2016)
7. Santos-Zavaleta, A., Sánchez-Pérez, M., Salgado, H., Velázquez-Ramírez, D.A., Gama-Castro, S., Tierrafría, V.H., Busby, S.J.W., Aquino, P., Fang, X., Palsson, B.O., Galagan, J.E., Collado-Vides, J.: A unified resource for transcriptional regulation in Escherichia coli K-12 incorporating high-throughput-generated binding data into RegulonDB version 10.0. BMC Biol. **16**, 91 (2018)
8. Novichkov, P.S., Brettin, T.S., Novichkova, E.S., Dehal, P.S., Arkin, A.P., Dubchak, I., Rodionov, D.A.: RegPrecise web services interface: programmatic access to the transcriptional regulatory interactions in bacteria reconstructed by comparative genomics. Nucleic Acids Res. **40**, W604–W608 (2012)

9. Novichkov, P.S., Kazakov, A.E., Ravcheev, D.A., Leyn, S.A., Kovaleva, G.Y., Sutormin, R. A., Kazanov, M.D., Riehl, W., Arkin, A.P., Dubchak, I., Rodionov, D.A.: RegPrecise 3.0 – a resource for genome-scale exploration of transcriptional regulation in bacteria. BMC Genom. **14**, 745 (2013)
10. Eckweiler, D., Dudek, C.-A., Hartlich, J., Brötje, D., Jahn, D.: PRODORIC2: the bacterial gene regulation database in 2018. Nucleic Acids Res. **46**, D320–D326 (2018)
11. Young, R.A.: Biomedical discovery with DNA arrays. Cell **102**, 9–15 (2000)
12. Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., Wold, B.: Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods **5**, 621–628 (2008)
13. Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., Snyder, M.: The transcriptional landscape of the yeast genome defined by RNA sequencing. Science (80-.) **320**, 1344–1349 (2008)
14. Iyer, V.R., Horak, C.E., Scafe, C.S., Botstein, D., Snyder, M., Brown, P.O.: Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. Nature **409**, 533–538 (2001)
15. Velculescu, V.E., Zhang, L., Zhou, W., Vogelstein, J., Basrai, M.A., Bassett, D.E., Hieter, P., Vogelstein, B., Kinzler, K.W.: Characterization of the yeast transcriptome. Cell **88**, 243–251 (1997)
16. Johnson, D.S., Mortazavi, A., Myers, R.M., Wold, B.: Genome-wide mapping of in vivo protein-DNA interactions. Science (80-.) **316**, 1497–1502 (2007)
17. Price, M.N., Deutschbauer, A.M., Kuehl, J.V., Liu, H., Witkowska, H.E., Arkin, A.P.: Evidence-based annotation of transcripts and proteins in the sulfate-reducing bacterium Desulfovibrio vulgaris Hildenborough. J. Bacteriol. **193**, 5716–5727 (2011)
18. Barrett, T., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., Marshall, K.A., Phillippy, K.H., Sherman, P.M., Holko, M., Yefanov, A., Lee, H., Zhang, N., Robertson, C.L., Serova, N., Davis, S., Soboleva, A.: NCBI GEO: archive for functional genomics data sets—update. Nucleic Acids Res. **41**, D991–D995 (2012)
19. Kolesnikov, N., Hastings, E., Keays, M., Melnichuk, O., Tang, Y.A., Williams, E., Dylag, M., Kurbatova, N., Brandizi, M., Burdett, T., Megy, K., Pilicheva, E., Rustici, G., Tikhonov, A., Parkinson, H., Petryszak, R., Sarkans, U., Brazma, A.: ArrayExpress update—simplifying data submissions. Nucleic Acids Res. **43**, D1113–D1116 (2015)
20. Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., Stoeckert, C., Aach, J., Ansorge, W., Ball, C.A., Causton, H.C., Gaasterland, T., Glenisson, P., Holstege, F.C.P., Kim, I.F., Markowitz, V., Matese, J.C., Parkinson, H., Robinson, A., Sarkans, U., Schulze-Kremer, S., Stewart, J., Taylor, R., Vilo, J., Vingron, M.: Minimum information about a microarray experiment (MIAME)—toward standards for microarray data. Nat. Genet. **29**, 365–371 (2001)
21. Moretto, M., Sonego, P., Dierckxsens, N., Brilli, M., Bianco, L., Ledezma-Tejeida, D., Gama-Castro, S., Galardini, M., Romualdi, C., Laukens, K., Collado-Vides, J., Meysman, P., Engelen, K.: COLOMBOS v3.0: leveraging gene expression compendia for cross-species analyses. Nucleic Acids Res. **44**, D620–D623 (2016)
22. Faith, J.J., Driscoll, M.E., Fusaro, V.A., Cosgrove, E.J., Hayete, B., Juhn, F.S., Schneider, S. J., Gardner, T.S.: Many Microbe Microarrays Database: uniformly normalized Affymetrix compendia with structured experimental metadata. Nucleic Acids Res. **36**, D866–D870 (2007)
23. Faria, J.P., Overbeek, R., Xia, F., Rocha, M., Rocha, I., Henry, C.S.: Genome-scale bacterial transcriptional regulatory networks: reconstruction and integrated analysis with metabolic models. Brief. Bioinform. **15**, 592–611 (2014)
24. Blazier, A.S., Papin, J.A.: Integration of expression data in genome-scale metabolic network reconstructions. Front. Physiol. **3**, 299 (2012)

25. Machado, D., Herrgård, M.: Systematic evaluation of methods for integration of transcriptomic data into constraint-based models of metabolism. PLoS Comput. Biol. **10**, e1003580 (2014)
26. Jensen, P.A., Papin, J.A.: Functional integration of a metabolic network model and expression data without arbitrary thresholding. Bioinformatics **27**, 541–547 (2011)
27. Colijn, C., Brandes, A., Zucker, J., Lun, D.S., Weiner, B., Farhat, M.R., Cheng, T.-Y., Moody, D.B., Murray, M., Galagan, J.E.: Interpreting expression data with metabolic flux models: predicting mycobacterium tuberculosis mycolic acid production. PLoS Comput. Biol. **5**, e1000489 (2009)
28. Chandrasekaran, S., Price, N.D.: Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in Escherichia coli and Mycobacterium tuberculosis. Proc. Natl. Acad. Sci. U.S.A **107**, 17845–17850 (2010)
29. van Berlo, R.J.P., de Ridder, D., Daran, J.-M., Daran-Lapujade, P.A.S., Teusink, B., Reinders, M.J.T.: Predicting metabolic fluxes using gene expression differences as constraints. IEEE/ACM Trans. Comput. Biol. Bioinform. **8**, 206–216 (2011)
30. Lee, D., Smallbone, K., Dunn, W.B., Murabito, E., Winder, C.L., Kell, D.B., Mendes, P., Swainston, N.: Improving metabolic flux predictions using absolute gene expression data. BMC Syst. Biol. **6**, 73 (2012)
31. Covert, M.W., Schilling, C.H., Palsson, B.: Regulation of gene expression in flux balance models of metabolism. J. Theor. Biol. **213**, 73–88 (2001)
32. Shlomi, T., Eisenberg, Y., Sharan, R., Ruppin, E.: A genome-scale computational study of the interplay between transcriptional regulation and metabolism. Mol. Syst. Biol. **3**, 101 (2007)
33. Jensen, P.A., Lutz, K.A., Papin, J.A.: TIGER: Toolbox for integrating genome-scale metabolic models, expression data, and transcriptional regulatory networks. BMC Syst. Biol. **5**, 147 (2011)
34. Schmidt, B.J., Ebrahim, A., Metz, T.O., Adkins, J.N., Palsson, B.Ø., Hyduke, D.R.: GIM3E: condition-specific models of cellular metabolism developed from metabolomics and expression data. Bioinformatics **29**, 2900–2908 (2013)
35. Marmiesse, L., Peyraud, R., Cottret, L.: FlexFlux: combining metabolic flux and regulatory network analyses. BMC Syst. Biol. **9**, 93 (2015)
36. Motamedian, E., Mohammadi, M., Shojaosadati, S.A., Heydari, M.: TRFBA: an algorithm to integrate genome-scale metabolic and transcriptional regulatory networks with incorporation of expression data. Bioinformatics. **33**, btw772 (2017)
37. Banos, D.T., Trébulle, P., Elati, M.: Integrating transcriptional activity in genome-scale models of metabolism. BMC Syst. Biol. **11**, 134 (2017)
38. Lloyd, C.J., Ebrahim, A., Yang, L., King, Z.A., Catoiu, E., O'Brien, E.J., Liu, J.K., Palsson, B.O.: COBRAme: a computational framework for genome-scale models of metabolism and gene expression. PLoS Comput. Biol. **14**, e1006302 (2018)
39. Orth, J.D., Thiele, I., Palsson, B.O.: What is flux balance analysis? Nat. Publ. Gr. **28**, 245–248 (2010)
40. Lewis, N.E., Hixson, K.K., Conrad, T.M., Lerman, J.A., Charusanti, P., Polpitiya, A.D., Adkins, J.N., Schramm, G., Purvine, S.O., Lopez-Ferrer, D., Weitz, K.K., Eils, R., König, R., Smith, R.D., Palsson, B.Ø.: Omic data from evolved E. coli are consistent with computed optimal growth from genome-scale models. Mol. Syst. Biol. **6**, 390 (2010)