

## Machine-learning from *Pseudomonas putida* KT2440 transcriptomes reveals its transcriptional regulatory network

Hyun Gyu Lim<sup>a,b</sup>, Kevin Rychel<sup>a</sup>, Anand V. Sastry<sup>a</sup>, Gayle J. Bentley<sup>c,d</sup>, Joshua Mueller<sup>e</sup>, Heidi S. Schindel<sup>f</sup>, Peter E. Larsen<sup>g</sup>, Philip D. Laible<sup>g</sup>, Adam M. Guss<sup>d,f</sup>, Wei Niu<sup>e</sup>, Christopher W. Johnson<sup>c,d</sup>, Gregg T. Beckham<sup>c,d</sup>, Adam M. Feist<sup>a,b,h</sup>, Bernhard O. Palsson<sup>a,b,h,i,\*</sup>

<sup>a</sup> Department of Bioengineering, University of California San Diego, 9500 Gilman Dr., La Jolla, CA, 92093, USA

<sup>b</sup> Joint BioEnergy Institute, 5885 Hollis Street, 4th Floor, Emeryville, CA, 94608, USA

<sup>c</sup> Renewable Resources and Enabling Sciences Center, National Renewable Energy Laboratory, 15013 Denver West Parkway, Golden, CO, 80401, USA

<sup>d</sup> Agile BioFoundry, Lawrence Berkeley National Laboratory, 1 Cyclotron Rd, Emeryville, CA, 94720, USA

<sup>e</sup> Department of Chemical & Biomolecular Engineering, University of Nebraska-Lincoln, 1400 R St, Lincoln, NE, 68588, USA

<sup>f</sup> Biosciences Division, Oak Ridge National Laboratory, 5200 Bethel Valley Rd, Oak Ridge, TN, 37830, USA

<sup>g</sup> Biosciences Division, Argonne National Laboratory, 9700 South Cass Avenue, Lemont, IL, 60539, USA

<sup>h</sup> Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, 2800, Kgs, Lyngby, Denmark

<sup>i</sup> Department of Pediatrics, University of California, San Diego, CA, 92093, USA

### ARTICLE INFO

#### Keywords:

Machine learning  
Independent component analysis  
Transcriptome  
Systems biology  
*Pseudomonas putida*

### ABSTRACT

Bacterial gene expression is orchestrated by numerous transcription factors (TFs). Elucidating how gene expression is regulated is fundamental to understanding bacterial physiology and engineering it for practical use. In this study, a machine-learning approach was applied to uncover the genome-scale transcriptional regulatory network (TRN) in *Pseudomonas putida* KT2440, an important organism for bioproduction. We performed independent component analysis of a compendium of 321 high-quality gene expression profiles, which were previously published or newly generated in this study. We identified 84 groups of independently modulated genes (iModulons) that explain 75.7% of the total variance in the compendium. With these iModulons, we (i) expand our understanding of the regulatory functions of 39 iModulon associated TFs (e.g., HexR, Zur) by systematic comparison with 1993 previously reported TF-gene interactions; (ii) outline transcriptional changes after the transition from the exponential growth to stationary phases; (iii) capture group of genes required for utilizing diverse carbon sources and increased stationary response with slower growth rates; (iv) unveil multiple evolutionary strategies of transcriptome reallocation to achieve fast growth rates; and (v) define an osmotic stimulon, which includes the Type VI secretion system, as coordination of multiple iModulon activity changes. Taken together, this study provides the first quantitative genome-scale TRN for *P. putida* KT2440 and a basis for a comprehensive understanding of its complex transcriptome changes in a variety of physiological states.

### 1. Introduction

*Pseudomonas putida* KT2440 has received much attention as a workhorse for numerous biotechnological applications due to its versatile metabolism that allows it to utilize a broad spectrum of substrates and express stress-tolerant phenotypes (Nikel and de Lorenzo, 2018; Wackett, 2003). Despite its industrial importance, understanding of its genome-scale transcriptional regulatory network (TRN) is to date

limited compared to other widely used microorganisms in biotechnology, such as *Escherichia coli* and *Bacillus subtilis*. Therefore, unveiling its TRN is imperative to improving our mechanistic understanding of gene expression changes to maximize its biotechnological potential (Fang et al., 2020; Herrgård et al., 2004).

Traditionally, genome-scale TRNs have been reconstructed in a bottom-up fashion by revealing regulons, one at a time, which are defined as the set genes regulated by the same transcription factor (TF)

\* Corresponding author. Department of Bioengineering, University of California San Diego, 9500 Gilman Drive La Jolla, CA, 92093-0412, USA.  
E-mail address: [palsson@ucsd.edu](mailto:palsson@ucsd.edu) (B.O. Palsson).

<https://doi.org/10.1016/j.ymben.2022.04.004>

Received 20 January 2022; Received in revised form 23 March 2022; Accepted 23 April 2022

Available online 27 April 2022

1096-7176/© 2022 The Authors. Published by Elsevier Inc. on behalf of International Metabolic Engineering Society. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

(Santos-Zavaleta et al., 2019). Specifically, the binding of TFs to putative promoter sites was historically investigated *in vitro* using low-throughput experimental techniques such as an electrophoretic mobility shift assay (Hellman and Fried, 2007) and a fluorescence polarization assay (Checovich et al., 1995; Lundblad et al., 1996). Although recent DNA affinity purification (DAP) sequencing method (Bartlett et al., 2017) enabled a high throughput studies, this method is still limited to *in vitro* interactions. Genome-wide *in vivo* binding studies are still relying on chromatin immunoprecipitation (ChIP) (Gao et al., 2021; Huang et al., 2019) sequencing technologies. These genome-wide binding results are typically compared with differentially expressed genes from TF knock-out and overexpression studies to distinguish functional binding sites from nonfunctional ones. Although these methods are essential to identify and validate regulons, elucidation of the genome-scale TRN, focusing on one known TF at a time is very laborious as there are approximately 500 proteins involved in transcriptional regulation in *P. putida* KT2440. Thus, a broader strategy is needed to elucidate the global TRN.

Recently, independent component analysis (ICA) has been shown to efficiently uncover bacterial genome-scale TRNs from gene expression profiles, obtained by using methods such as microarray hybridization and RNA sequencing (Poudel et al., 2020; Rychel et al., 2020; Sastry et al., 2019). ICA is an unsupervised blind source separation algorithm, which can identify statistically independent components (ICs) from mixed signals (Comon, 1994; Hyvärinen, 1999; Liebermeister, 2002); it was found to perform the best in a comparative analysis of signal extraction algorithms for TRN inference (Saelens et al., 2018). ICs from gene expression profiles are independently modulated groups of genes (iModulons), which are likely under the regulation of the same underlying biological signal (e.g. TF or genetic perturbation). iModulons are similar to regulons (Sastry et al., 2021), but not identical since they were obtained in a top-down data-driven manner. While Principal Component Analysis (PCA) is often used to identify variance within datasets, principal components (PCs) are statistical measures that usually lack clear biological interpretations. However, we have demonstrated that iModulons recapitulate known regulatory mechanisms and can thus be considered as knowledge-based representations of the composition of the transcriptome. Indeed, recent studies successfully applied ICA for model bacteria (e.g., *E. coli* (Sastry et al., 2019), *Bacillus subtilis* (Rychel et al., 2020)) and even less-studied bacteria (e.g., *Staphylococcus aureus* (Poudel et al., 2020), *Mycobacterium tuberculosis* (Yoo et al., 2022), *Sulfolobus acidocaldarius* (Chauhan et al., 2021), *P. aeruginosa* (Rajput et al., 2021)); these efforts allowed rapid identification of strain-specific genome-scale TRNs and their systematic comparison across species (Sastry et al., 2021).

In this study, to elucidate the genome-scale TRN of *P. putida* KT2440, we perform ICA for an RNA-seq compendium, named *putida*PRECISE321, consisting of 321 high-quality gene expression profiles, of which 305 were previously published and 16 were newly generated in this study (Supplementary Data). We obtained 84 iModulons, which explain 75.7% of the variation in the compendium. With the iModulons, we show that the genome-scale TRN in *P. putida* KT2440 can be effectively uncovered by identifying regulatory target genes of 38 TFs. In cases where the iModulon did not match previously described regulons well, we propose improved gene-regulator interactions that are corroborated by motif analysis. In addition, we demonstrate the ability of iModulons to comprehensively interpret transcriptional responses to environmental or evolutionary changes. We demonstrate the iModulon clustering, effectively representing a stimulon, to suggest an activating condition of a less-characterized functional cluster (i.e., the Type VI Secretion System). Collectively, the 84 iModulons delineate the genome-scale TRN in *P. putida* KT2440 by deconvolution of complex transcriptomic responses orchestrated by multiple TFs. Detailed, searchable, interactive dashboards for all 84 iModulons are publicly available on iModulonDB (<https://imodulondb.org>) (Rychel et al., 2021).

## 2. Results

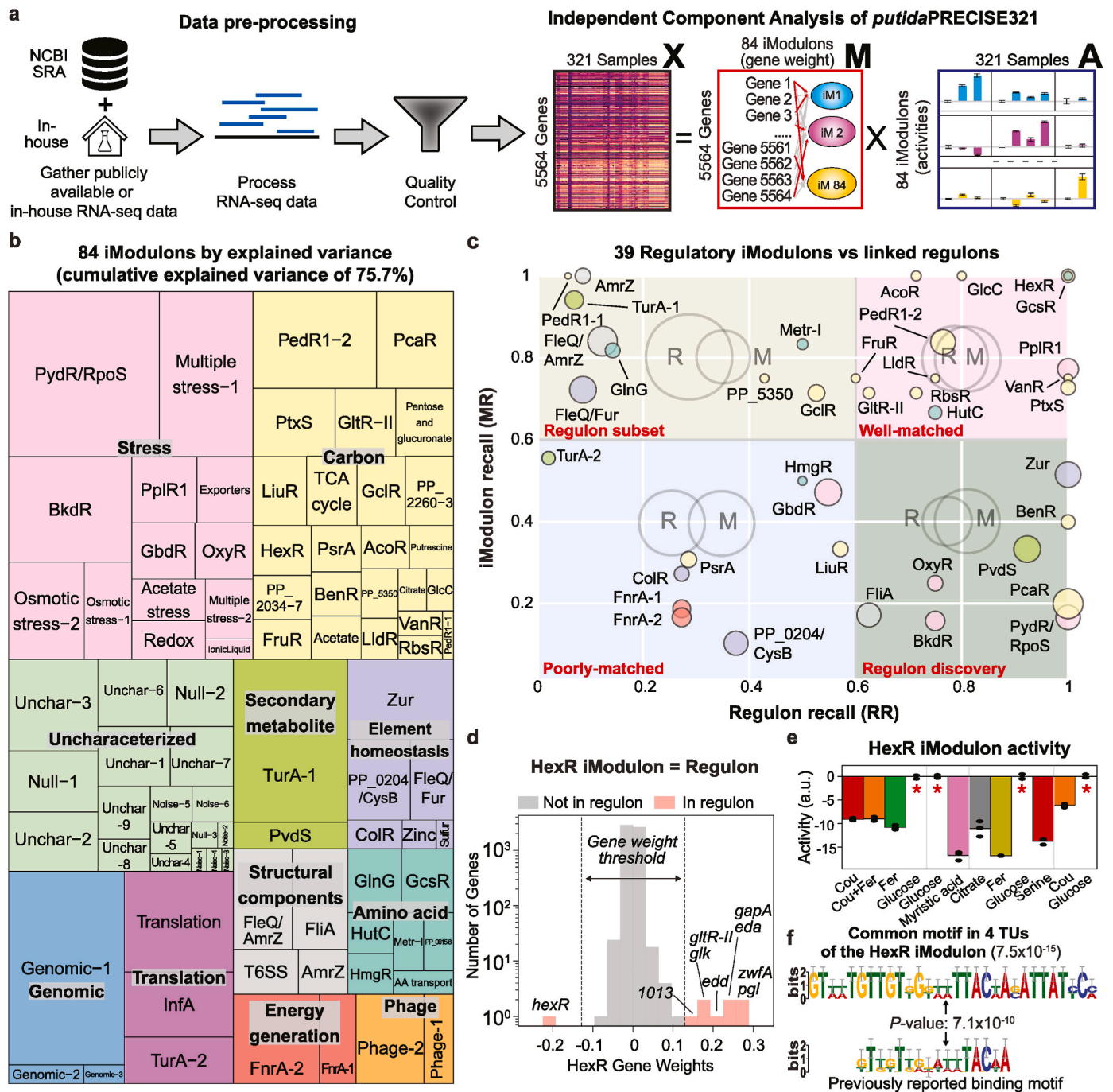
### 2.1. ICA reveals 84 independently modulated groups of genes in *P. putida* KT2440

We generated a compendium of RNA-seq datasets for *P. putida* KT2440 by downloading, aligning, and performing rigorous quality control on all public datasets available from SRA as of Aug 31, 2020 and in-house datasets using our established pipeline (Sastry et al., 2021) (See Methods; Fig. 1a and Supplementary Fig. 1). We named the compendium *putida*PRECISE321, for *putida* Precision RNA-seq Expression Compendium for Independent Signal Exploration. It consists of 321 samples from 118 unique experimental conditions across 21 projects (Bentley et al., 2020; Bojanović et al., 2017; Jayakody et al., 2018; Lei et al., 2020; Lim et al., 2020, 2021; Miyazaki et al., 2018; Nakamura et al., 2020; Pobre et al., 2020; Xiao et al., 2018). We centered each project to a baseline condition to remove batch effects (Supplementary Fig. 2) and then used the ICA algorithm (Sastry et al., 2021) to extract all independent expression signals (iModulons) from the data. The results, therefore, represent co-regulated gene sets and regulator activities across all genes perturbed by any available high-quality transcriptome for *P. putida* KT2440.

The ICA of *putida*PRECISE321 revealed 84 iModulons. Each iModulon has member gene weights (stored in the columns of a matrix M), from which outliers are identified as “member genes”, and activities in each sample, or condition (stored in the rows of a matrix A). The 84 iModulons contain genes with closely related biological functions (Fig. 1b and Supplementary Fig. 3a). The median gene number in the 84 iModulons was 10 (Supplementary Fig. 3b). Together, these iModulons include 1265 genes (23% of the total annotated ORFs), whose expression levels actively changed under the conditions represented in the dataset. A total of 75.7% of the gene expression variance was mechanistically explained by these iModulons, which is comparable to the variance found by PCs (Supplementary Fig. 3c).

The iModulons were characterized and categorized into four groups, depending on their level of characterization and likely underlying mechanisms: regulatory, functional, genomic, and uncharacterized iModulons. “Regulatory” iModulons (46.4% of the 84 iModulon) have significant overlaps (false discovery rates, FDRs  $<10^{-4}$ ) with previously reported regulons mostly from bottom-up experiments. We compiled 1993 experimentally validated or predicted interactions, as part of this study, from multiple databases and literature (see Methods). Consequently, we found 39 iModulons associated with TFs (Supplementary Table 1). In addition, 24 “functional” iModulons (28.6%) were defined by enrichment of genes found in a given metabolic pathway from the KEGG database, or by inferring roles based on gene functional annotations or their active conditions. Additionally, three “genomic” iModulons (3.5%) contain differentially expressed genes due to genomic variations (e.g., gene deletion) in the host strains. The remaining 18 iModulons (21.4%) were classified as “uncharacterized”. They typically contained short hypothetical genes or genes with unknown functions or unclear active expression conditions. “Uncharacterized” iModulons represent potential discoveries of new regulons. Regulatory or functional iModulons were named based on their regulators (if associated) or related characteristics (see Methods). Genomic or uncharacterized iModulons were numbered in order based on the number of genes that they contain, with smaller iModulons appearing first.

To investigate the accuracy of TRN uncovered by iModulons, we systematically compared the 39 regulatory iModulons with associated TF regulons using two metrics (Fig. 1c): iModulon recall (MR) and regulon recall (RR), where MR is the fraction of shared genes in a given iModulon and RR is the fraction of shared genes in a linked TF regulon (Supplementary Fig. 4a). Depending on these two metrics, regulatory iModulons were further categorized into four groups (one for each quadrant of Fig. 1c) with a threshold value of 0.6 for both MR and RR: well-matched, regulon subset, regulon discovery, and poorly-matched



**Fig. 1.** Independent component analysis (ICA) of the *putida*PRECISE321 transcriptome compendium. (a) Overall data processing pipeline for independent component analysis (ICA) of *putida*PRECISE321. By performing ICA, the gene expression profiles [X] (5564 genes by 321 conditions) were deconvoluted into an [M] matrix containing weights of 5564 genes for each of the 84 iModulon identified, and an [A] matrix containing iModulon activities for each condition. A part of this diagram was adapted from Sastry et al. (2021). (b) A treemap of *P. putida*'s 84 iModulons. The size of each rectangle indicates the explained variances in the data sets attributable to each iModulon. (c) Bottom-up vs top-down assessment of co-regulated genes (i.e., bio-molecular vs data-driven). The scatter plot shows Regulon Recalls (RR) and iModulon Recalls (IR) for 39 regulatory iModulons. For iModulons with names consisting of two regulators, RR and IR are the values of the first regulator. Colors indicate iModulon types. (d) An example of a well-matched case: HexR. Gene weights of the HexR iModulon. (e) Condition-dependent HexR iModulon activities with the supplementation of different carbon sources. Asterisks indicate the reference conditions (glucose minimal media). It should be noted that each project (separated by gray lines) was normalized to its unique control condition (Supplementary Data) to remove batch effects, so activity values from different projects should not be directly compared. (f) Enriched motif (top) from the upstream of HexR-regulating genes and previously reported HexR binding site (bottom, obtained from CollectTF (Kiliç et al., 2014)).

(Rychel et al., 2020; Sastry et al., 2019). 12 regulatory iModulons (both MR and RR  $\geq 0.6$ ), including HexR, GlcC, AcoR, and BenR, were “well-matched” with linked regulons (Supplementary Fig. 4b), indicating that ICA successfully recapitulated previously reported

regulation. Regulon subset iModulons (MR  $\geq 0.6$  and RR  $< 0.6$ ) indicate that a subgroup of a linked regulon shows an independent signal, suggesting different strength of regulation within regulon genes or another layer of regulation by unstudied co-regulator(s). Larger established

regulons often contain multiple iModulons (Rychel et al., 2020; Sastry et al., 2019), underlining the fact that regulons have a ‘static’ definition (i.e., TF binding) and iModulons have a ‘dynamic’ definition (i.e., are computed from expression profiles).

We identified ten regulon subset iModulons associated with TFs, including FleQ (PP\_4373), GlnG (PP\_5048), and PedR1 (PP\_2665), known to regulate relatively large numbers of genes. In addition, we also obtained eight “regulon discovery” iModulons ( $MR < 0.6$  and  $RR \geq 0.6$ ). These iModulons include most reported target genes of linked TFs but also contain additional genes whose regulation had not been previously reported. Given the similar expression pattern of genes in the same iModulons, these additional genes are potentially new target genes of a linked TF. Alternatively, entire iModulon genes can be co-regulated by a less-studied or yet unidentified regulator, closely associated with a linked TF. Lastly, we obtained eight “poorly-matched” iModulons (both  $MR$  and  $RR < 0.6$ ) which have small overlaps with linked closest regulons. The relatively low regulon overlap of these iModulons may be explained by additional, unknown TFs, weak or variable binding strengths for TFs with large regulons, or incompletely described regulons.

The 84 iModulons reproduce many known details of the *P. putida* KT2440 TRN, including the TFs and their target genes, regulatory modes, and binding sites. For example, the HexR iModulon has an  $MR$  and  $RR$  of 1, indicating that it precisely captured the previously reported regulation of HexR (PP\_1021), a glucose metabolism regulator (Dad-daoua et al., 2009) for nine target genes (*zwfA*, *pgl*, *gapA*, *eda*, *edd*, *gltR-II*, *glk*, PP\_1013, *hexR*, Fig. 1d). Since HexR represses the expression of itself and the 8 other genes (Del Castillo et al., 2008), ICA assigned a negative gene weight to *hexR* and a positive gene weight to the remaining genes. In addition, the activity of the HexR iModulon was found to be decreased when non-glucose carbon sources were provided (Fig. 1e), consistent with its expected function. Lastly, a binding site of HexR can be also computationally predicted by enriching a consensus motif in the upstreams of iModulon member genes (Bailey et al., 2009). With a maximum window of 30 bp, we found the “GTWWTGTGKTKGWTTACWASATTATYCM” motif is present in the upstream regions of all four transcriptional units (TUs,  $e$ -value of  $7.5 \times 10^{-15}$ , Fig. 1f, see Supplementary Note 2 for each symbol), which perfectly includes a previously reported motif. Similar characterization can be extended to other TFs from iModulons; we provided summaries and identified motifs of selected iModulons in Supplementary Tables 1 and 2, respectively. The agreement between well-matched regulons and iModulons shows commonality in the output of the complementary bottom-up (biomolecular) and top-down (data analytics) approaches.

## 2.2. iModulons allows data-driven curation of the TRN in *P. putida* KT2440

Regardless of iModulon types, differences between iModulons and corresponding regulons indicate the presence of gaps between the ICA-defined gene groups and existing knowledge and thus allow further data-driven curation of *P. putida* KT2440’s TRN. As examples, we provided detailed characterizations of FleQ/Fur, Zur, PP\_0204/CysB iModulons, a representative iModulon of each ‘regulon subset,’ ‘regulon discovery,’ and ‘poorly-matched’ iModulons.

**A regulon subset iModulon shows co-regulation in the FleQ regulon by multiple TFs** - ICA identified a regulon subset iModulon (originally named as FleQ-1 but renamed to FleQ/Fur) in the FleQ regulon ( $MR$  and  $RR$  of 0.72 and 0.09, Fig. 2a). FleQ (PP\_4373) is a c-di-GMP(cyclic-di-GMP)-dependent TF, known to regulate up to 300 genes, involved in iron transport, flagellar synthesis, and biofilm formation (Blanco-Romero et al., 2018; Molina-Henares et al., 2017). This iModulon contains 36 genes, mostly involved in inorganic ion transport and metabolism (Supplementary Figs. 5a and 5b) and regulated by FleQ. Its activity increased when *P. putida* was grown in a relatively less-aerated condition, in a medium containing acetate, or in cells which were adapted to

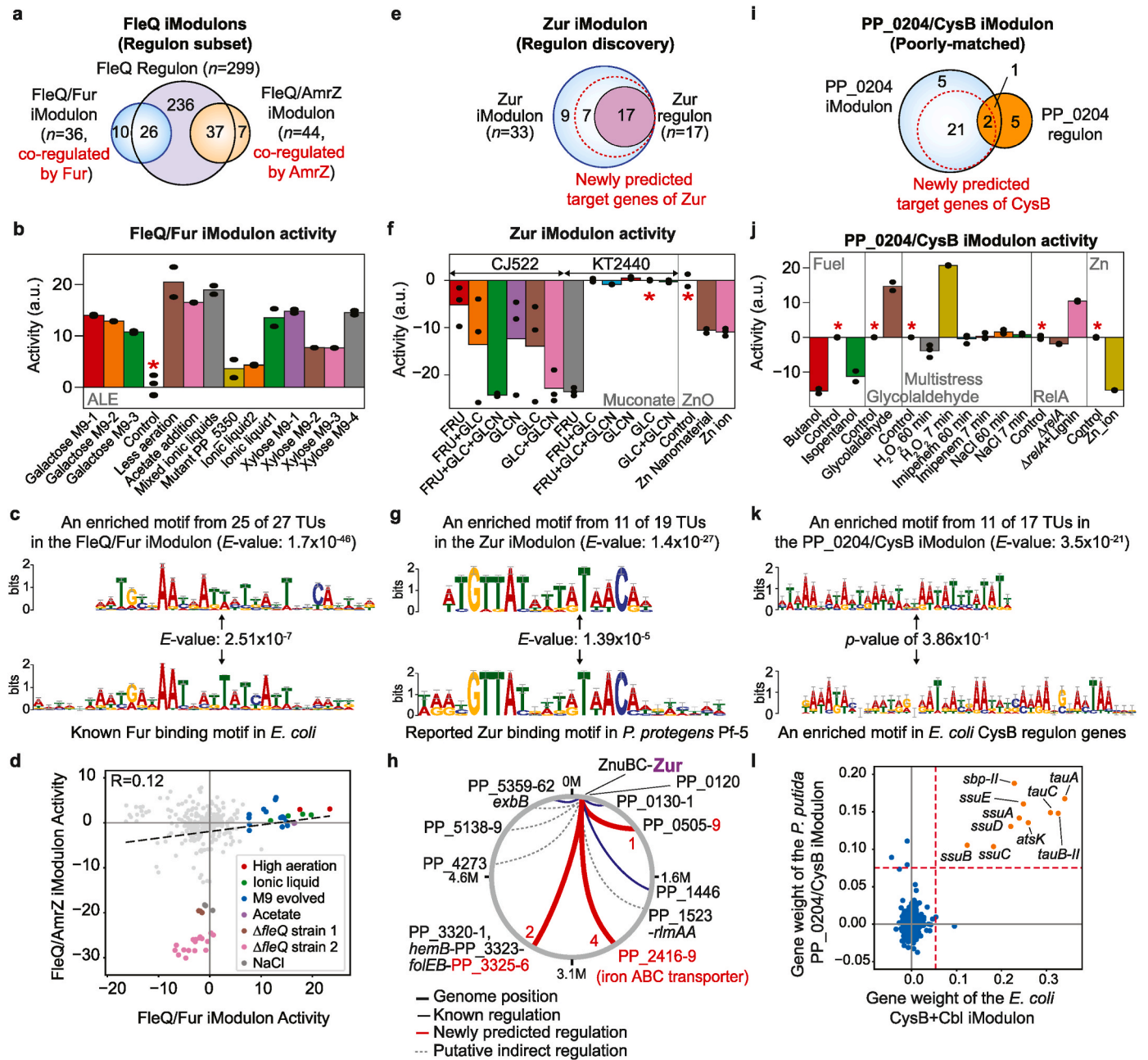
minimal media conditions; conditions which likely affected oxygen or iron availability (Fig. 2b). We hypothesized that this independent signal within the FleQ regulon is owing to the presence of another regulator(s) for genes.

To test this hypothesis, we computationally enriched a consensus sequence from the upstream sequences of 27 transcriptional units in this iModulon using MEME (Bailey et al., 2009). Notably, one motif (AAT-GYKAAAYRATWHTBAYTHBCAWTWD, Fig. 2c), distinctively different from a known FleQ motif (Supplementary Fig. 5a), was enriched with high confidence (an  $e$ -value of  $0.7 \times 10^{-46}$ ) from 25 TUs (34 of the total 36 genes, 94%, shown in bold in Supplementary Fig. 5a). By comparing this enriched motif to known TF binding sites in public databases using TOMTOM (Gupta et al., 2007) (See Methods), we found that the motif is highly similar to the binding site of *E. coli* Fur (a ferric uptake regulator) at an  $e$ -value of  $2.51 \times 10^{-7}$ . This result strongly suggests the co-regulation by Fur (PP\_4730) in addition to FleQ. Additionally, we obtained another FleQ regulon subset iModulon ( $MR$  of 0.86 and  $RR$  of 0.12, originally named as FleQ-2 but renamed to FleQ/AmrZ) containing 44 genes, mostly encoding machinery for cell motility such as flagella synthesis (Fig. 2a and Supplementary Fig. 5c). Its activity decreased when *fleQ* was deleted or cells were osmotically stressed (Fig. 2d). Despite their common regulation by FleQ, activities of the two iModulons did not have a significant correlation (Pearson coefficient  $< 0.12$ ), suggesting the presence of another co-regulator. Indeed, from the upstream regions of 13 out of the 18 TUs (30 of the total 44 genes, 68.2%), another motif (GGTDTTGGYSWTGGCVTYGGCCWGS), similar to the binding site of *P. aeruginosa* AmrZ (PA3385, alginate and motility regulator Z), was commonly enriched at an  $e$ -value of  $5.2 \times 10^{-7}$  (Supplementary Table 2). This observation suggests that most of the member genes in this iModulon are possibly co-regulated by AmrZ. Collectively, these subset iModulons clearly indicate multiple roles of FleQ, controlling together with other TFs, in inorganic ion uptake and cell motility and provide a baseline for future regulation study for an in-depth understanding of associated TFs.

**ICA suggested an expanded regulatory role of Zur for metal uptake genes** - ICA identified the co-expression of 33 genes which were down-regulated when an excess amount of zinc was included in a medium or different carbon sources were utilized (Fig. 2e and f). This iModulon, named Zur ( $MR$  of 0.52 and  $RR$  of 1.00, “regulon discovery”), included all seventeen genes (*znuBC*, *zur*, *znuA*, PP\_0506-8, PP\_1446, PP\_3320-3, *folEB*, PP\_5359-62), known to involve in zinc uptake and to be regulated by Zur (PP\_0119) (Lim et al., 2013). Interestingly, 16 new genes which do not have any reported regulation by Zur, were additionally detected by ICA to form an iModulon that is larger than the corresponding bottom-up defined regulon.

To investigate the possibility of common regulation by Zur, we enriched a common motif from the upstream sequence of all 19 TUs of the Zur iModulon genes. Notably, a common motif “WRRYGT-TATRWTRTAACAWKWHWHW” was found in the upstream of 11 TUs including 22 genes (67%) at an  $e$ -value of  $1.4 \times 10^{-27}$  (Fig. 2g). This enriched motif is highly consistent with a previously known binding motif (TTGTTATAAGATAACAT) of Zur in *P. protegens* Pf-5 at an  $E$ -value of  $1.39 \times 10^{-5}$  (Guo and Houghton, 1999; Lim et al., 2013), supporting their regulation by Zur. The 22 genes include seven newly predicted target genes (PP\_0509, PP\_2416-9, PP\_3325-6, Fig. 2h). In particular, the PP\_2416-9 operon was annotated to encode an iron ABC transporter, which indicates that Zur has an expanded regulatory role in iron uptake, not limited to zinc. The activity changes depending on carbon sources (Fig. 2f), which may suggest different levels of metal requirements; further related studies are warranted. Nevertheless, these results show how iModulons in combination with binding motifs are useful to identify additional target genes of a given regulator.

**A ‘poorly-matched’ iModulon suggests conserved CysB regulation in *P. putida* KT2440** - We obtained a poorly-matched iModulon, PP\_0204/CysB (originally named as PP\_0204), which contains 29 genes in 17 TUs (Fig. 2i,  $MR$  of 0.10 and  $RR$  of 0.38). These iModulon genes were



**Fig. 2.** iModulons allow data-driven curation of the TRN in *P. putida*. (a–d) Characteristics of the FleQ regulon subset iModulons. (a) A Venn diagram of the two FleQ iModulons, FleQ/Fur and FleQ/AmrZ, and the FleQ regulon. (b) An activity plot of the FleQ/Fur iModulon in selected conditions. (c) Comparison of a motif found in the upstream regions of the FleQ/Fur iModulon genes and previously known binding motifs of Fur in *E. coli* (obtained from CollectTF (Kiliç et al., 2014)). The common motif was enriched from 25 out of 27 TUs (94.4% of the genes) in the FleQ/Fur iModulon. (d) A scatter plot of activities of the FleQ/Fur and FleQ/AmrZ iModulons for all conditions. The Pearson correlation of the two iModulon activities was 0.12. Selected samples with high activities were colored and growth conditions were indicated in the inset. Activities of the FleQ/Fur iModulon changed when *fleQ* was deleted or an excess amount of NaCl was added. (e–h) Characteristics of the Zur regulon discovery iModulon. (e) A Venn diagram of the Zur iModulon and related Zur regulon. (f) An activity plot of the Zur iModulon in selected conditions. High activities of the Zur iModulon were observed when zinc ion or nanomaterial was added. The CJ522 strain is an engineered *P. putida* KT2440 strain  $\Delta catRBCA::Ptac:cata \Delta pcaHG::Ptac:aroY:ecdB::asbF \Delta pykA::aroG^{D146N}:aroY:ecdB::asbF \Delta pykF \Delta ppc \Delta pgi-1 \Delta pgi-2 \Delta gcd$  (Johnson et al., 2019). (g) Comparison with an enriched motif from 11 TUs in the Zur iModulon (66.7% of the total genes) and a previously reported *P. protegens* Pf-5 Zur binding motif (obtained from CollectTF (Kiliç et al., 2014)). (h) A schematic diagram of regulation by Zur. Previously known regulations were colored in black while newly predicted regulations were colored in red. Potential indirect regulations were indicated with dashed gray lines. (i–l) Characteristics of the PP\_0204 poorly-matched iModulon. (i) A Venn diagram of the PP\_0204/CysB iModulon and PP\_0204 regulon. (j) An activity plot of the PP\_0204 iModulon in selected conditions. Asterisks indicate the reference conditions. (k) Comparison with an enriched motif from 11 TUs in the PP\_0204 iModulon and an enriched motif from 14 TUs in the *E. coli* CysB regulon (obtained from RegulonDB (Santos-Zavaleta et al., 2019)). (l) Gene weight correlation of the PP\_0204/CysB iModulon and the Cbl+CysB iModulon of *E. coli*. The dashed lines indicate the threshold values (0.054 and 0.075) for *E. coli* and *P. putida* to determine iModulon member genes. See Methods for the calculation. (a, e, and i) Gene numbers were shown in these diagrams. (b, f, and j) Asterisks indicate the reference conditions.

annotated to involve transportation of sulfur-containing compounds (e.g., *sbp-II*, *tauABC*, *ssuEADCB* encoding sulfate or sulfonate transporters) and showed high expression in the presence of hydrogen peroxide or glycolaldehyde, known to induce oxidative stress in microorganisms (Bojanović et al., 2017) (Fig. 2j). Although PP\_0204 (a GntR family TF) was originally enriched as a regulator, this TF was previously suggested to be a local regulator for itself and neighboring genes (e.g., PP\_0205 and PP\_0206 encoding an oxidoreductase and 4Fe–4S ferredoxin) (Szkłarczyk et al., 2019).

We hypothesized that there might be another major TF regulating these iModulon genes with a close interaction to PP\_0204. This hypothesis was supported by a motif found from 11 of the total 17 TUs (23 genes of the 29 genes, 79%) at an *e*-value of  $3.5 \times 10^{-21}$  (Fig. 2k). To infer a potential main regulator of this iModulon, we compared its member gene weights with those of previously reported iModulons in *E. coli* (Sastry et al., 2019). Notably, the comparison showed that this iModulon is highly consistent with the Cbl+CysB iModulon in *E. coli* (Lamoureux et al., 2021) (Fig. 2l). Among the two *E. coli* TFs, we found that the *cysB* gene is also present in *P. putida* KT2440. In addition, CysB in the two strains shows high similarity (Supplementary Fig. 6), suggesting this TF is the main regulator. We further compared a previously reported binding motif of *E. coli* CysB with the enriched motif. Although the significance was relatively low (*p*-value of  $3.86 \times 10^{-1}$ ), the two motifs commonly share AT-rich sequences (Fig. 2k), also supporting common regulation for these genes by CysB in *P. putida* KT2440. Consequently, this iModulon was renamed as PP\_0204/CysB based on these observations and this example shows that a poorly-matched iModulon can be evidence of regulation by an unstudied TF closely related to a linked TF.

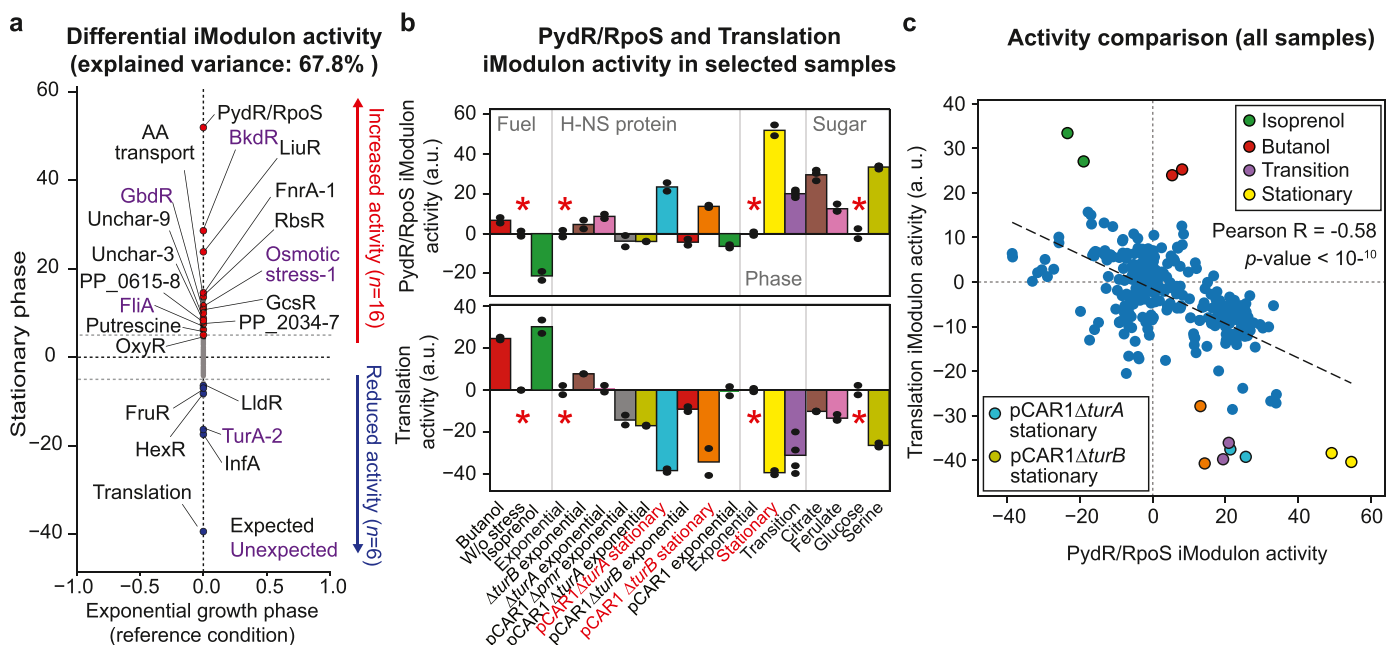
### 2.3. iModulons provide a comprehensive description of transcriptome changes after the transition from the exponential growth phase to the stationary phase

Since iModulons represent biologically meaningful groups of co-regulated genes, complex transcriptomic changes can be simplified

and comprehensively interpreted by monitoring iModulon activities rather than investigating the roles of huge numbers of differentially expressed genes (DEGs) (Poudel et al., 2020; Sastry et al., 2019). To demonstrate this capability, we analyzed changed iModulon activities after the transition from the exponential growth phase to the stationary phase (project *phase*, Supplementary Data). When we applied the conventional DEG analysis, we obtained 1454 genes ( $p_{adj} < 0.05$ ) which correspond to 26.1% of the total number of genes (Supplementary Fig. 7).

In contrast, we obtained only 22 iModulons with differentially changed activities at an FDR < 0.05 and an activity threshold of 5; 67.8% of the total variance was explained by using iModulons (Fig. 3a). Among them, a total of sixteen iModulons showed increased activities while the activities of six iModulons were decreased. Changes of many iModulons were consistent with previous literature or reasonable in view of decreased metabolism and stress response due to nutrient depletion (Supplementary Table 3). For example, increased activities of PydR/RpoS for starvation responses, LiuR for ribose and lipid catabolism were identified whereas decreased activities of translation for protein synthesis, and HexR for glucose metabolism were observed. Additionally, there were some iModulons (e.g., BkdR, Osmotic stress-1, GbdR, TurA-2) that showed unexpected activity changes; they can be targeted for detailed future studies.

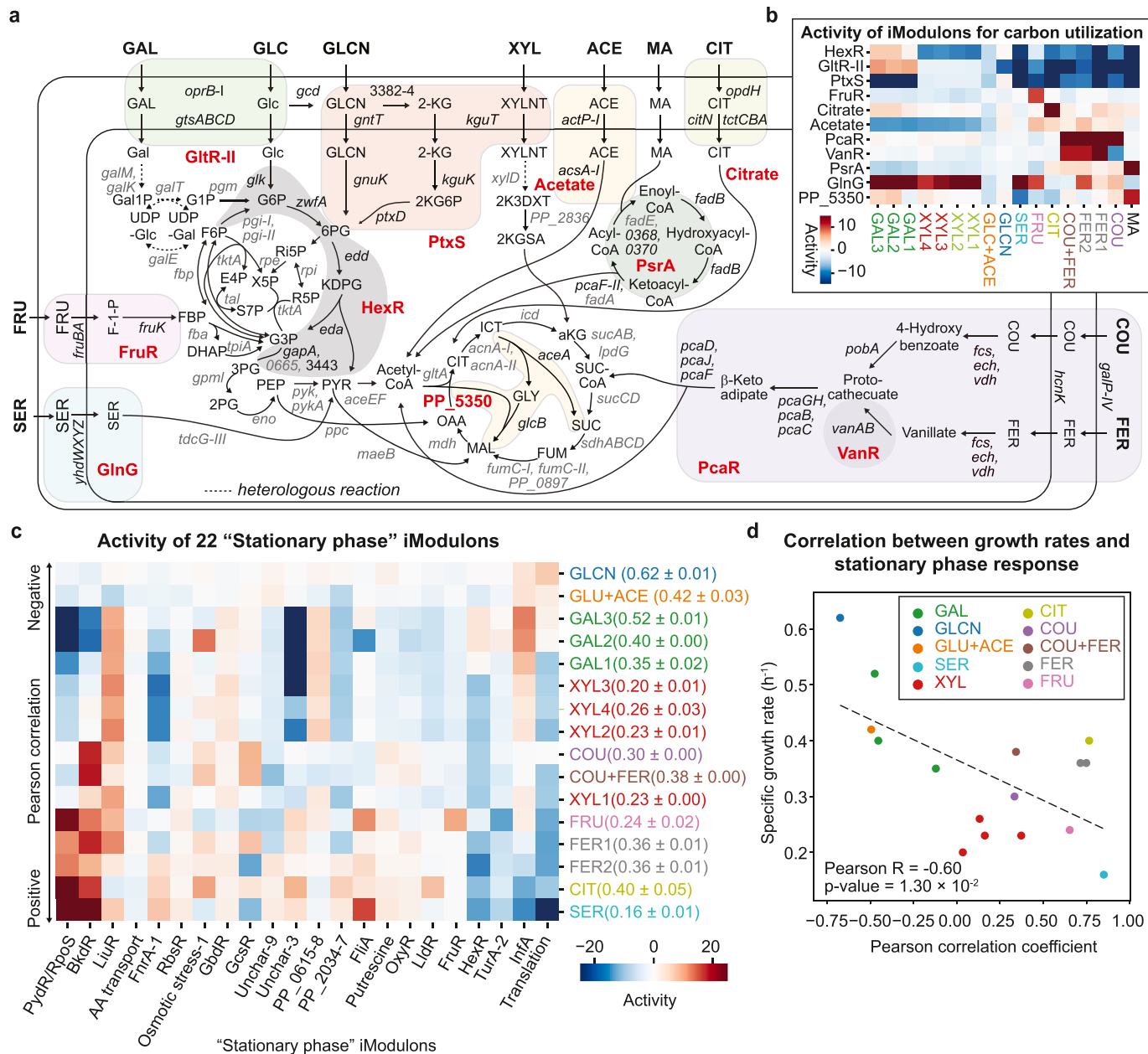
The characteristics of PydR/RpoS and Translation iModulons were further investigated since they exhibited the greatest activity changes (Supplementary Figs. 8a and 8b). PydR/RpoS, a regulatory iModulon (MR and RR of 0.17 and 1, respectively, regulon discovery), contains 30 genes presenting at various locations on the genome and explains the highest fraction of the total variance in our compendium (Supplementary Fig. 3d). Increased activities of the PydR/RpoS iModulon were consistently observed in all samples collected at the stationary phase in another project (*project*: H-NS protein, Fig. 3b). Its activity also changed depending on the carbon source (citrate, ferulate, or serine) or the presence of a stressor in the media. Originally, PydR was enriched as a major regulator; this TF is known to regulate the expression of four nucleobase degradation genes (Rinas et al., 1995) which display high



**Fig. 3.** iModulons describe global transcriptional changes after the transition from the growth phase to the stationary phase. (a) Differentially changed iModulon activities during the transition from the exponential growth phase to the stationary phase at a threshold of 5 and an FDR of 0.05. Unexpected changes were colored in purple. (b) Activities of the PydR/RpoS and Translation iModulons in selected samples. Project names are colored in gray. Asterisks indicate the reference conditions for each project. (c) Activity correlation of the PydR/RpoS and Translation iModulons. Activity of these two iModulons showed a negative correlation across samples (Pearson correlation coefficient *R* of  $-0.58$ , *p*-value <  $1 \times 10^{-10}$ ).

gene weights in this iModulon: *pydXA*, *hyuC*, and *pydB* (Supplementary Fig. 8a). However, additional member genes for carbohydrate transporters (e.g., PP\_1121), stress-mitigation (e.g., *katG* encoding catalase peroxidase and PP\_1121 encoding OmpA family protein), and energy generation (e.g., *icd* encoding NADP<sup>+</sup>-specific isocitrate dehydrogenase) were identified. Notably, a motif was enriched from all 27 TUs, including the PydR regulon genes, at an *e*-value of  $2.8 \times 10^{-7}$

(Supplementary Table 2), suggesting that not only are the four genes regulated by a common regulator, but all the other genes are as well. In many microorganisms including *E. coli*, a similar stationary response was shown to be governed by RpoS and a RpoS iModulon explained the largest variance in a corresponding gene expression compendium (Poudelet et al., 2020; Rychel et al., 2020; Sastry et al., 2019). Furthermore, its activity also showed a high correlation with the expression



**Fig. 4.** iModulons show transcriptional responses to the utilization of different substrates. (a) Metabolic pathway associated with iModulons for substrate utilization in *P. putida*. Boundaries of iModulons (names in red) were indicated by colored rounded shapes. Names of genes belonging to iModulons were colored in black. Subcellular localization information was obtained from Pseudomonas.com (Winsor et al., 2016). (b) Activities of the carbon utilization iModulons depending on substrates. For a clear comparison, the minimum and maximum activity values were set to  $-15$  and  $15$ , respectively. (c) The activities of the 22 iModulons, that showed differential activities during the transition to the stationary phase from the growth phase, during utilization of each substrate. For clear visualization, the minimum and maximum activities were set to  $-25$  and  $25$ , respectively. (b and c) Multiple RNA-Seq samples with the same substrates were differentiated by adding numbers. iModulon activities in the glucose condition are zero. (d) A scatter plot of the growth rates and Pearson correlation coefficients of activities of the “stationary response” in iModulons in the group. Abbreviations: GAL, galactose; GLC, glucose; GLCN, gluconate; XYL, xylose; ACE, acetate; MA, myristic acid; CIT, citrate; FRU, fructose; SER, serine; 2-KG, 2-ketogluconate; XYLNT, xylonate; 2K3DXT, 2-keto-3-deoxyxylonate; 2KGSA, 2-keto-glutaric semialdehyde; G6P, glucose-6-phosphate (P); 6 PG, 6-phosphogluconate; 2KG6P, 2-ketogluconate-6-P; Gal1P, galactose-1-P; F-1-P, fructose-1-P; UDP-Glc, uridine diphosphate-glucose; UDP-Gal, uridine diphosphate galactose; G1P, glucose-1-P; KDPG, 2-dehydro-3-deoxy-phosphogluconate; PYR, pyruvate; G3P, glyceraldehyde-3-P; 3 PG, glycerate-3-P; 2 PG, 2-phosphoglycerate; PEP, phosphoenolpyruvate; acetyl-CoA, acetyl coenzyme A (CoA), CIT, citrate; ICT, isocitrate;  $\alpha$ KG,  $\alpha$ -ketoglutarate; SUC-CoA, succinyl-CoA; SUC, succinate; FUM, fumarate; MAL, malate; OAA, oxaloacetate; GLY, glyoxylate.

level of *rpoS* rather than *pydR* (Supplementary Figs. 9a and 9b). Collectively, this iModulon may represent the regulation by RpoS (PP\_1623) in *P. putida* KT2440 and thus renamed to PydR/RpoS.

On the other hand, the Translation iModulon, a functional iModulon with an unknown regulator, includes 45 genes that are mostly ribosomal proteins and clustered together on the genome (Supplementary Fig. 8b). Additionally, it includes genes encoding RNA polymerases (*rpoB* and *rpoC*), elongation factors (e.g., *tufA*, *fusA*), cytochrome oxidases (*cyoC* and *cyoD*), and ATP synthase subunits (*atpG* and *atpA*), important for protein synthesis and energy production. Therefore, its reduced activity at the stationary phase indicates decreased biomass production. Interestingly, a global inverse correlation of the PydR and Translation iModulon activities were observed across samples regardless of growth phases (Pearson coefficient of  $-0.58$  at a  $p$ -value  $< 10^{-10}$ , Fig. 3c), except a few conditions (e.g., butanol addition).

Previously, a similar negative correlation between stationary-phase induced genes and growth-related genes was described as a major transcriptomic trade-off in *E. coli* (Sastray et al., 2019). The observation in *P. putida* KT2440 indicates a similar trade-off is conserved across species for diverting resources from growth to maintenance. Collectively, the changed iModulon activities clearly illustrate the transcriptional changes that occurred during the transition from the growth phase to the stationary phase.

#### 2.4. iModulons identify substrate-specific gene groups and show increased stationary phase response during utilization of poor carbon sources

We examined how iModulon activities reflect the utilization of different substrates. Although catabolic efficiencies vary depending on substrates, resulting in global changes in growth rates, understanding of transcriptome changes in *P. putida* KT2440 depending on substrates has been limited. *putida*PRECISE321 includes seventeen gene expression profiles with 11 different carbon source conditions: glucose (as a reference), gluconate, fructose, galactose, xylose, citrate, coumarate, ferulate, serine, acetate, and myristic acid. In addition, it should be noted that gene expression profiles with xylose and galactose utilization were obtained from engineered *P. putida* KT2440 strains with heterologous gene expression (Lim et al., 2021).

ICA identified multiple iModulons (e.g., HexR, GltR-II, FruR, PsrA, PP\_5350) containing genes responsible for transport and catabolism for each substrate (Fig. 4a). Except for the GlnG iModulon, the activity of most iModulons specifically showed increased or decreased activity depending on catabolic routes for each carbon source (Fig. 4b). For example, the PsrA and PP\_5350 iModulons responsible for  $\beta$ -oxidation and glyoxylate shunt pathway, respectively, showed high activities only when myristic acid, a C-14 fatty acid, was utilized as a sole carbon source. These observations indicate that the expression of catabolic genes is tightly regulated by multiple related TFs. Additionally, high activities of the GlnG iModulon were observed in most non-glucose conditions (e.g., galactose, xylose, fructose) in addition to the serine utilization condition. This iModulon is the regulon subset of GlnG (PP\_5048, also called NtrC) is known to control the expression of 63 nitrogen metabolism genes (Hervás et al., 2009). Although it is not clear, this observation implied that intracellular nitrogen availability could be affected depending on carbon sources.

We further hypothesized that poor carbon sources, which result in low growth rates, induce the stationary phase-like response in the transcriptome even if the culture is still in the exponential growth phase. To test this hypothesis, we investigated the activity of the “stationary phase” iModulons (i.e., the 22 iModulons that showed differential activity during the transition from the growth phase to the stationary phase) with specific growth rates ( $0.16 \text{ h}^{-1}$  to  $0.62 \text{ h}^{-1}$ , Fig. 4c and d). Notably, we observed that the activity of the 22 iModulons in cells growing on gluconate, a mixture of glucose and acetate, and galactose clearly showed a negative correlation (up to a Pearson correlation of  $-0.83$  at a  $p$ -value of  $8.54 \times 10^{-4}$ ) whereas cells growing on xylose,

aromatic compounds (i.e., coumarate and ferulate), citrate, and serine showed positively correlated iModulon activity (up to a Pearson correlation of  $0.89$  at a  $p$ -value of  $0.96 \times 10^{-4}$ ). We found that the correlation of a sample’s activity levels with stationary phase activity levels is itself correlated with growth rate, indicating that stationary phase iModulon activities are a fairly conserved transcriptomic signature for slowed growth. Taken together, these results show that global correlations in transcriptomes can be efficiently found by analyzing iModulon activities in multiple datasets, revealing tradeoffs in transcriptome composition.

#### 2.5. iModulons unveil evolutionary reallocation of the transcriptome to achieve fast growth rates

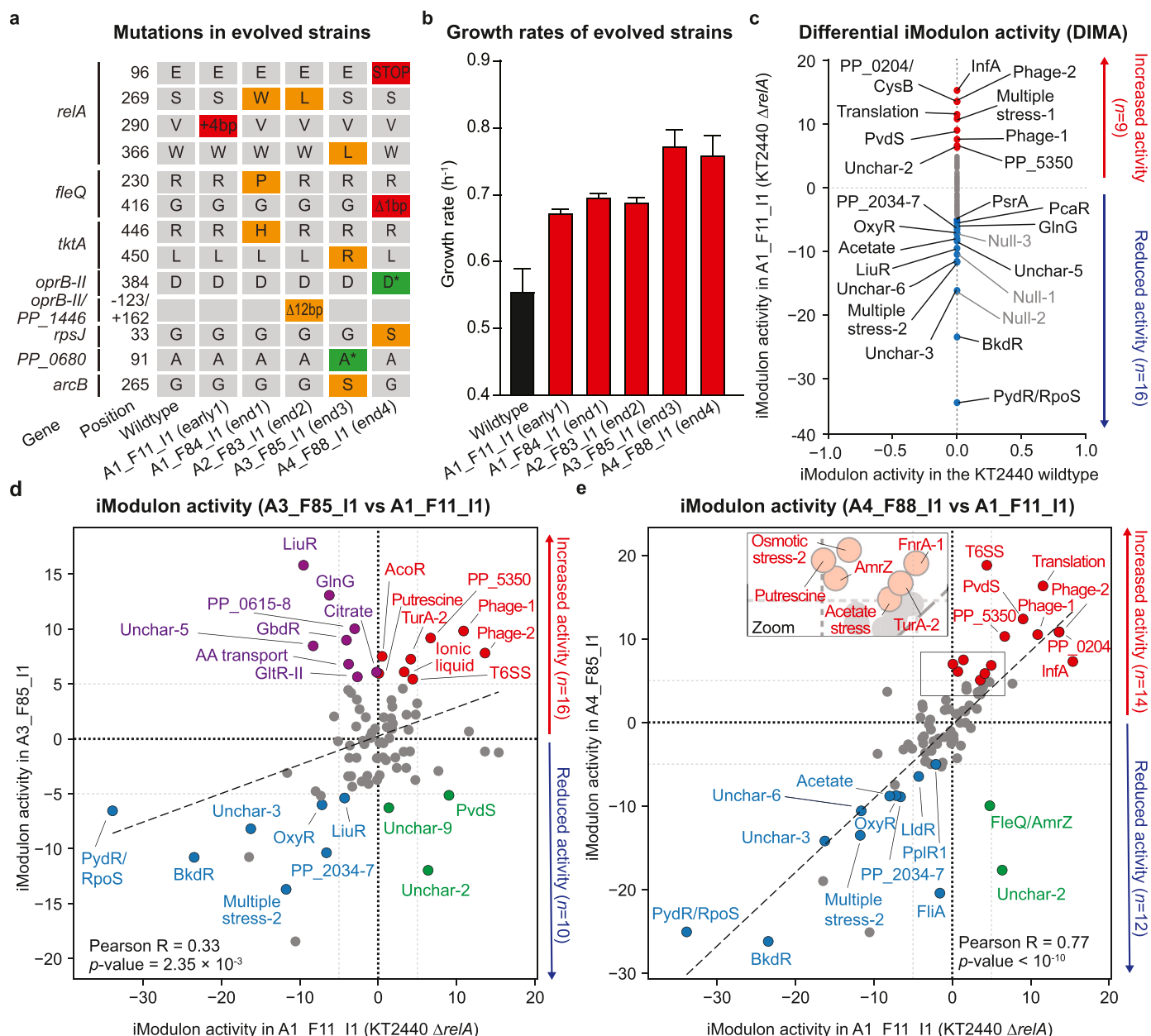
Once an iModulon structure for an organism is established, it can be used to interpret new data sets by inferring the activities of each iModulon based on the gene weights from the [M] matrix of the original compendium. Thus, we examined the utility of the *P. putida* KT2440 iModulon structure to unveil an evolutionary strategy for achieving high fitness in a glucose-minimal medium. We analyzed iModulon activity in three evolved *P. putida* KT2440 strains (A1\_F11\_I1, A3\_F85\_I1, and A4\_F88\_I1, see Supplementary Note 3 for naming) (Lim et al., 2020), displaying higher growth rates ( $0.64 \text{ h}^{-1}$ ,  $0.77 \text{ h}^{-1}$ , and  $0.76 \text{ h}^{-1}$ , respectively) over the wildtype ( $0.58 \text{ h}^{-1}$ , Fig. 5a and b). These strains were independently evolved but all acquired mutations in *relA*, which encodes one of two guanosine tetraphosphate (ppGpp) synthetases in *P. putida*, suggesting its critical role in the improved fitness. The A1\_F11\_I1 strain which was isolated from an early flask (11th) in a laboratory evolution study, only has a single mutation - a frame-shift in *relA*, whereas the two latter strains have different additional mutations in other genes in addition to a single amino acid change or early termination mutation in *relA*.

Initially, we investigated changes in iModulon activities in the A1\_F11\_I1 strain (KT2440 with the loss-of-function mutation in *relA*) to understand the effect of reduced ppGpp synthesis (Fig. 5c). We found 22 iModulons with large activity changes (using a threshold of 5). Nine iModulons including Translation and InfA responsible for cell growth showed increased activities whereas thirteen iModulons including starvation iModulons (PydR and BkdR) and stress-hedging iModulons (OxyR, Multiple stress-2) showed decreased activities. These observations were consistent with previous studies that ppGpp has a pivotal role to determine whether cells actively grow or show quiescence phenotypes (Diez et al., 2020). In addition, it also shows efficient reallocation of cellular resources to obtain robust cell growths in the evolution condition.

We further compared iModulon activities in the A3\_F85\_I1 and A4\_F88\_I1 strains against the A1\_F11\_I1 strain (Fig. 5d and e). These two strains showed even higher growth rates ( $0.77 \text{ h}^{-1}$  and  $0.76 \text{ h}^{-1}$ ) over the A1\_F11\_I1 strain. However, no genes other than *relA* were commonly mutated in these strains, suggesting different evolutionary strategies for improving growth rates. When the two strains were compared with the A1\_F11\_I1 strain, their transcriptomes have an overall positive correlation (Pearson R values of  $0.33$  and  $0.77$  at  $p$ -values of  $2.35 \times 10^{-3}$  and less than  $10^{-10}$ , respectively), likely owing to mutations commonly occurred in *relA*. Interestingly, we identified several iModulons with inversely correlated activity changes in the two strains. In the A3\_F85\_I1 strain, eight iModulons (colored in purple in Fig. 5d, e.g., LiuR, GlnG, PP\_0615) showed high activities, whereas three iModulon (Uncharacterized-2, Uncharacterized-3, PvdS, colored in green) showed decreased activities. On the other hand, in the A4\_F88\_I1 strain, only two iModulons (FleQ/AmrZ and Uncharacterized-2, colored in green in Fig. 5e) showed decreased activity.

From these iModulon activity changes, two distinctive evolution strategies could be inferred. In the A3\_F85\_I1 strain, three mutations in *tktA* (a gene encoding transketolase A), PP\_0680 (a gene encoding ATP-dependent protease-like protein), and *arcB* (a gene encoding ornithine carbamoyltransferase) likely activated nutrient re-assimilation or



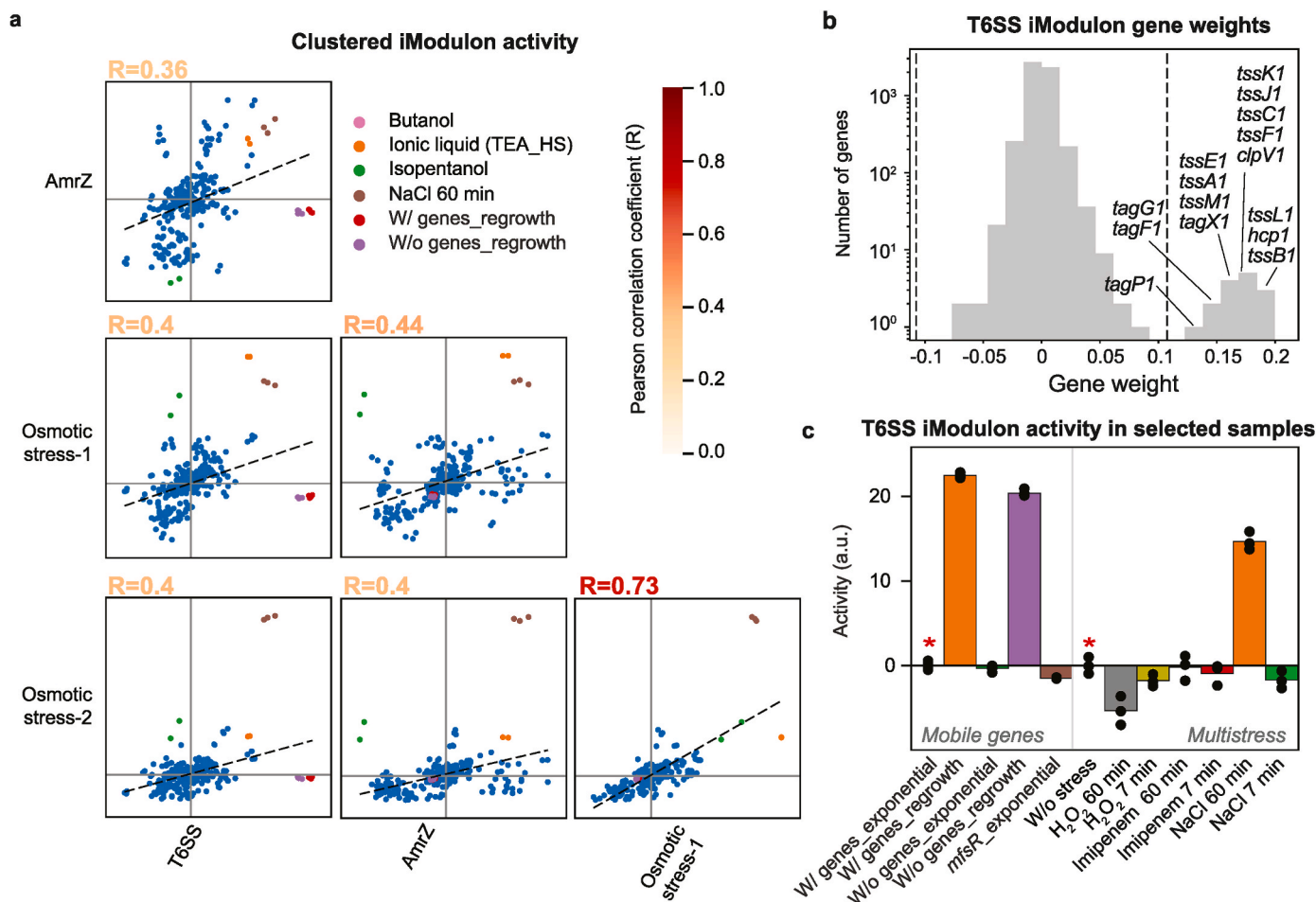


**Fig. 5.** Analyzing iModulon activities unveil evolutionary strategies to achieve high fitness in a glucose minimal medium condition. (a) Mutations identified in the wildtype, A1\_F11\_I1 (early1), A1\_F84\_I1 (end1), A2\_F83\_I1 (end2), A3\_F85\_I1 (end3), and A4\_F88\_I1 (end4) strains (Lim et al., 2020). Strain names indicate passage numbers during an evolution study (see Supplementary Note 3). Colors: orange, single amino acid changes or an in-frame deletion; green, nonsynonymous mutation; red, early termination or frame-shift mutations (b) Maximum specific rates ( $h^{-1}$ ) of the six strains in a glucose minimal medium. (c) Differential iModulon activity plot for the A1\_F11\_I1 and wildtype strains. (d and e) Differential iModulon activity plot for the A1\_F11\_I1 versus either (d) A3\_F85\_I1 or (e) A4\_F88\_I1 strain. iModulon activities in the non-wild type strains were inferred from the [M] matrix of the original iModulon activity. Dashed gray lines indicate a threshold value of 5 and solid gray lines indicate a 45-degree line of each plot. Null iModulons were colored in gray even with changes greater than 5.

utilization; their effects were shown as increased activities of iModulons for Ribose and lipid catabolism (LiuR), amino acid metabolism (GlnG, AA transport), carbon metabolism (GltR-II). Differently, the decreased activity of FleQ/AmrZ containing flagella synthesis genes implies that the A4\_F88\_I1 strain achieved efficient biomass formation by minimizing wasteful synthesis of extracellular proteins from the frame-shift mutation in *fleQ* (Smith and Chapman, 2010). Taken together, multiple evolutionary strategies were identified by analyzing iModulon activities. This method has significant potential to be broadly applicable to identify meaningful information from new gene expression profiles.

## 2.6. Coordinated iModulon activity changes define a stimulon and underlie the activation of the Type VI secretion system (T6SS) by osmotic stress

Activities of iModulons may be correlated, especially if they share master regulators or common activating conditions. Therefore, clustering the activity levels can elucidate underlying relationships or hierarchical transcriptional regulation and define a stimulon (i.e., genes with changed expression by an external stimulus (Barer, 2012)). As an example, we found that activities of four iModulons (Osmotic stress-1, Osmotic stress-2, AmrZ, T6SS) showed high correlations across samples (Fig. 6a). The Osmotic stress-1 and Osmotic stress-2 iModulons



**Fig. 6.** Coordinated iModulon activity changes underlie the activation of the Type VI secretion system by osmotic stress and define a stimulon. (a) Activity correlation of the T6SS, Osmotic stress-1, Osmotic stress-2, and AmrZ iModulons. Activities of all these four iModulons showed a positive correlation. Font colors indicate Pearson correlation coefficients. Samples displaying high activities were colored: pink (fuel:butanol), green (fuel:isopentanol), orange (ALE:IL\_TEA\_HS), brown (multistress:NaCl\_60min), red (mobile\_genes:2737\_reg-w/mobile genes at the regrowth phase), and purple (mobile\_genes:3227\_reg-w/mobile genes at the regrowth phase). Dashed lines are the best fits on each plot. (b) Gene weights of the T6SS iModulon. The dashed line indicates a threshold value of 0.11 (See Methods). (c) T6SS iModulon activities in selected samples from projects: *mobile genes* and *multistress* (Bojanović et al., 2017). See Supplementary Data for details of the experimental conditions. Asterisks indicate the reference conditions.

included 46 and 31 genes, respectively, commonly showing high activities under osmotic-stressed conditions but displaying different strengths depending on the types of osmotic stress (i.e., NaCl or protic ionic liquid addition). The AmrZ iModulon included 12 genes for alginate synthesis. The high correlation of the first three iModulons recapitulates previous findings that *P. putida* increases the synthesis of alginate, which is one of the biofilm components, to create a micro-hydrated environment under high-osmotic stress conditions (Chang et al., 2007; Nielsen et al., 2011). Meanwhile, the T6SS iModulon includes 15 machinery genes for the Type VI secretion system (T6SS, Fig. 6b) required for its antibacterial activity against phytopathogens (e.g., *Xanthomonas campestris*), which functions by injecting toxins (Bernal et al., 2017). In particular, this iModulon showed high activities in samples collected under regrowth phase (i.e., promoting cell growth by additional supplementation of a carbon source after its depletion) and NaCl-treated samples (Fig. 6c). Given that the T6SS iModulon was clustered together, it can be suggested that the secretion system is induced by osmotic changes. Indeed, a close relationship between the activity of T6SS and biofilm formation was reported in *P. aeruginosa* (Chen et al., 2020) and the activation of T6SS by OmpR, a global osmotic stress regulator, was previously suggested for *Klebsiella pneumoniae* (Barbosa and Lery, 2019). Overall, our iModulon clustering provides strong evidence that the T6SS system is activated by osmotic changes in

a surrounding environment, and reflects the classical definition of a stimulon.

### 3. Discussion

In this study, we elucidated the TRN of *P. putida* KT2440 by obtaining 84 iModulons using ICA of 321 high-quality RNA-seq samples. 39 Regulatory iModulons with linked TFs, recapitulated previously reported TF-gene interactions and identified knowledge gaps (e.g., the existence of a potential co-regulator or a new regulation target) via a systematic comparison of iModulon gene membership and existing TRNs. In addition, we identified 24 groups of genes as functional iModulons and suggested them as excellent targets for future studies to elucidate unreported regulators, given their co-expression upon environmental changes (e.g., osmotic stress). Compared to traditional regulon discovery approaches which investigate activities in a few experimental conditions, ICA simultaneously analyzes various gene expression profiles and identifies multiple gene groups likely under the same regulation in a high-throughput manner.

We provided strong evidence of co-regulation of iModulon member genes by identifying common motifs from the upstream of TUs and their comparison with previously reported motifs for many prokaryote bacteria. We obtained a total of 32 common motifs from 76% of the total

TUs in an iModulon on average (Supplementary Table 2), indicating that most member genes in the same iModulon are likely regulated by a common regulator(s). Furthermore, we found that eight best motifs for the FleQ/Fur, FnrA-2, HexR, Multiple stress-1, PvdS, PydR/RpoS, TCA cycle, Zur iModulons showed similarity to known TF binding motifs for *E. coli* and other *Pseudomonas* species. From this cross-species comparison, regulations by linked TFs (i.e., Fur, FnrA, HexR, PvdS, RpoS, Zur) were identified; also, potential regulators (i.e., FnrA, for the Multiple stress-1 and TCA cycle) were suggested.

iModulons are powerful to visualize key changes in gene expression profiles. In contrast, DEG analysis is not suitable for the identification of global changes in multiple gene expression profiles containing thousands of genes. Although the 26 categories of Clusters of Orthologous Groups of proteins (COGs) (Galperin et al., 2015) are often utilized, the category size is relatively small, and such categories do not consider individual characteristics of a specific microorganism of interest. Conversely, as shown in this work, iModulons were defined from gene expression profiles of a given microorganism and thus they are specific to the corresponding microorganism. This strain-specific TRN is particularly important for characterizing non-model microorganisms, including *P. putida* KT2440, relatively lacking knowledge and genetic engineering tools compared to model microorganisms. Moreover, the number of iModulons is determined after a screening of the optimal dimensionality to avoid both under-decomposition and over-decomposition of a gene expression compendium (McConn et al., 2021). With the optimum reduction in complexity, various transcriptomes, even newly generated, can be comprehensively interpreted.

By using iModulons, we comprehensively explained transcriptome changes in various conditions that are closely related to industry-relevant bioprocesses (i.e., carbon-depletion induced stationary phase, utilization of different carbon sources with different catabolic efficiencies, evolutionary adaptation to a glucose minimal medium condition). Although understanding the stationary phase response is essential, its understanding at a glance was limited owing to an absence of tools. In this regard, we identified 22 iModulons with differentially changed activities after transitioning to the stationary phase from the exponential growth phase. These iModulons indicate how *P. putida* KT2440 reallocates its transcriptome under a nutrient-depleted environment for its survival. Moreover, it was shown that slow growth rates on poor substrates induce the stationary phase-like response even at the exponential growth phase. Finally, multiple evolutionary strategies to achieve high growth rates were unveiled by using iModulons. Similar iModulon-based interpretation of transcriptome changes that occurred during diverse bioprocesses or adaptive evolution will aid a better understanding of cellular responses and subsequent strain designing for diverse biotechnology applications.

The iModulon results can be utilized in a variety of future studies. They can be subjected to detailed experimental validation via RNA-seq of TF-deleted strains or ChIP-sequencing (Gao et al., 2021) to confirm suggested TF-gene relationships. Regulators for functional or uncharacterized iModulons could be identified by using the motif information via DNA-pull down assay. In addition, since iModulons allow efficient comparison across multiple conditions, they can generate many hypotheses from the activity comparison. Based on the activity clustering, we suggested that the activity of the T6SS is closely related to osmotically stressed conditions by monitoring gene expressions in samples. In a similar vein, new hypotheses can be generated and subsequently tested to expand our understanding of *P. putida* KT2440. The functions of uncharacterized genes (i.e., unknown function genes) can be also inferred from active conditions. Additionally, since iModulons provide boundaries of genes required for a given functionality (e.g., aromatics catabolism) with statistical significance, identified gene sets can be studied for reinforcement of functions or even heterologously transferred to another microorganism for engineering.

Lastly, the explanatory power of iModulons can be further increased by performing ICA for a larger scoped compendium to achieve a more

detailed structure of the TRN. For example, the second round of ICA was performed for 815 *E. coli* RNA-seq samples (Lamoureux et al., 2021) to identify 218 iModulons, which were expanded from 92 iModulons reported in the first version (Sastry et al., 2019); the expansion covering diverse genetic perturbations and environmental conditions generated more detailed iModulon structures and new regulon discoveries (Lamoureux et al., 2021). Efforts can be similarly extended to identify a more detailed *P. putida* KT2440 iModulon structure underlying its TRN, given a growing number of new RNA-seq samples uploaded on public repositories.

Taken together, this study provides the first global TRN analysis for *P. putida* KT2440, and we believe that the *P. putida* KT2440 iModulons will be useful for advancing our fundamental understanding and engineering of this microorganism for many different purposes.

## 4. Methods

### 4.1. Generation of a high-quality *P. putida* RNA-Seq compendium

To perform ICA, a total of 541 *P. putida* RNA-Seq samples were collected; 435 samples were obtained from the NCBI Sequence Read Archive (SRA, <https://www.ncbi.nlm.nih.gov/sra>, published before Aug 31, 2020); 9 samples were obtained from a previous study (Bentley et al., 2020); 97 samples were newly generated in this study (Supplementary Data). RNA-seq samples, generated in this study, were prepared as described in Supplementary Method 1. Raw sequencing files were processed by the prokaryotic RNA-seq processing pipeline (Sastry et al., 2021) (<https://github.com/avsastry/modulome-workflow>) implemented to use Nextflow v20.01.0 (Di Tommaso et al., 2017) on Amazon Web Services (AWS) as previously described (Sastry et al., 2021). Briefly, the pipeline utilizes fasterq-dump (<https://github.com/ncbi/sra-tools/wiki/HowTo-fasterq-dump>), Trim Galore ([https://www.bioinformatics.babraham.ac.uk/projects/trim\\_galore](https://www.bioinformatics.babraham.ac.uk/projects/trim_galore)), FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) for raw data processing. Sequencing reads were aligned to the reference genome (AE015451.2) using Bowtie (Langmead et al., 2009). Read counts were generated by using RSEQC (Wang et al., 2012) and featureCounts (Liao et al., 2014) and converted into log<sub>2</sub> transcripts per million (TPM).

To ensure the quality of the compendium (Supplementary Fig. 1), MultiQC (Ewels et al., 2016) was performed and failed samples were discarded based on per base sequence quality, per sequence quality scores, per base *n* content, and adapter content. Hierarchical clustering was also utilized to identify and discard datasets not displaying a typical expression profile. Furthermore, metadata for each sample was manually curated by pulling information from literature and the registry and samples with low correlation within biological replicates ( $R^2 < 0.95$ ) were discarded. Additionally, non-*P. putida* KT2440 samples were excluded for consistency of gene expression profiles. For 321 samples that passed the aforementioned QC metrics, unique project and condition identifiers were given (*project:condition*, Supplementary Data).

### 4.2. Independent component analysis (ICA)

ICA (Hyvärinen, 1999; Sastry et al., 2021) was performed as introduced in previous studies (Sastry et al., 2019, 2021) to decompose the gene expression profile [X] into independent components (i.e., iModulons, [M]) and their activities [A] in each sample.

$$X = M \times A$$

Briefly, each gene expression profile was normalized by dividing the log TPM values by the values of a reference condition within each project to remove potential batch effects. Then, the scikit-learn (v0.23.2) (Pedregosa et al., 2011) implementation of FastICA (Hyvärinen, 1999) was performed 100 times with a convergence tolerance of  $10^{-7}$  to obtain robust independent components (ICs). The resulting ICs were clustered

by using DBSCAN (Ester et al., 1996) to identify robust ICs from each random restart. For the clustering, an epsilon of 0.1 and a minimum cluster seed size of 50 were used. Because identical ICs can have opposite signs, distance ( $d$ ) between components was calculated from the absolute value of the Pearson correlation ( $\rho$ ) as follows:

$$d_{x,y} = 1 - |\rho_{x,y}| \quad (1)$$

To choose the optimal dimension (McConn et al., 2021), we performed the clustering to the gene expression profile multiple times for dimensions between 10 and 290 with the step size of 10. The optimal dimension of 220 was chosen where the number of non-single gene ICs was equal to the number of final components. Finally, the M and A matrices for 84 robust ICs (i.e., iModulons) were obtained. From the obtained [M] matrix, iModulon member genes for each iModulon were determined by choosing an optimal threshold, as explained in Supplementary Method 2.

#### 4.3. Characterization of iModulons

The 84 iModulons were characterized by the Pymodulon package (<https://github.com/SBRG/pymodulon>) (Sastry et al., 2021). Functions of the iModulons were inferred by enriching known regulators, Kyoto Encyclopedia of Genes and Genomes (KEGG, <https://www.genome.jp/kegg/>) annotations of member genes. In particular, we collected 1993 TF-gene interactions (i.e., regulon) in *P. putida* (and some in *P. aeruginosa*) from RegPrecise 3.0 (Novichkov et al., 2013), BioCyc Database Collection (Karp et al., 2019), and literature. All interactions and their sources are listed in Supplementary Data. Subsequently, Fisher's Exact Test was performed with an FDR of  $<10^{-4}$  using the Benjamini-Hochberg correction. KEGG annotations and Cluster of Orthologous Groups (COG) information were obtained by running eggNOG-mapper (version 5) (Huerta-Cepas et al., 2019). Fisher's Exact test was also utilized to enrich GO and KEGG annotations with an FDR of  $<10^{-2}$ . Common motifs in upstreams of iModulons were enriched by using the find\_motifs function that uses MEME (Bailey et al., 2009) with a maximum window of 30 bp. When motifs were compared, the web-based TOMTOM (Gupta et al., 2007) was utilized. As a statistical significance,  $E$ -values were provided as adjusted  $p$ -values when more than one common sequence was identified or compared. iModulons were named with their enriched TFs from the regulons or functional characteristics. When additional TF was suggested from a motif search or a comparison of iModulons in other microorganisms, iModulon names were changed accordingly. Additionally, agglomerative activity clustering of iModulons was performed by using the cluster\_activities function. iModulon activities in new expression profiles were inferred by using the infer\_activities function.

#### 4.4. Generating iModulonDB web page

The *P. putida* iModulonDB (Rychel et al., 2021) page was generated by using the imodulondb\_export function in the Pymodulon package (Sastry et al., 2021) and it is accessible at <https://imodulondb.org>. The generated page includes gene information from the Pseudomonas Genome DB (<https://pseudomonas.com>) (Winsor et al., 2016).

#### Data availability

All RNA-seq raw files are available at GEO with accession numbers listed in Supplementary Data. Source codes for iModulon analysis and figures are available at [https://github.com/SBRG/modulome\\_ppu](https://github.com/SBRG/modulome_ppu).

#### Author contributions

Hyun Gyu Lim: Conceptualization, Data Curation, Investigation, Writing - Original Draft, Visualization Kevin Rychel: Methodology,

Software, Writing - Original Draft Anand V. Sastry: Conceptualization, Methodology, Software, Writing - Original Draft Gayle J. Bentley: Investigation, Resources Joshua Mueller: Investigation, Resources Heidi S. Schindel: Investigation, Resources Philip D. Laibleg: Investigation, Resources Peter E. Larseng: Investigation, Resources Adam M. Guss: Investigation, Resources, Funding acquisition Wei Niu: Investigation, Resources, Funding acquisition Christopher W. Johnson: Investigation, Resources Gregg T. Beckham: Investigation, Resources, Funding acquisition, Writing - Review & Editing Adam M. Feist: Conceptualization, Writing - Review & Editing, Funding acquisition Bernhard O. Palsson: Conceptualization, Writing - Review & Editing, Funding acquisition, Project administration, Supervision.

#### Declaration of competing interest

All authors declare no competing interests.

#### Acknowledgments

This work conducted by the Joint BioEnergy Institute was supported by the Office of Science, Office of Biological and Environmental Research, of the U.S. Department of Energy (DOE) under Contract No. DE-AC02-05CH11231. This work was authored in part by Alliance for Sustainable Energy, LLC, the manager and operator of the National Renewable Energy Laboratory for the U.S. DOE under Contract No. DE-AC36-08GO28308. This work was also partially authored by Oak Ridge National Laboratory, which is managed by UT-Battelle, LLC, for the U.S. DOE under contract DE-AC05-00OR22725. Funding was provided by the U.S. DOE Office of Energy Efficiency and Renewable Energy Bioenergy Technologies Office (BETO) for the Agile BioFoundry. This work was also supported by the Nebraska Center for Energy Science Research at the University of Nebraska - Lincoln (Cycle-11 grant to W.N.). We also thank members of the Agile BioFoundry, Richard Szubin, Ying Hutchison, and Marc K. Abrams for helpful discussions.

#### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ymben.2022.04.004>.

#### References

- Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., Noble, W.S., 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 37, W202–W208. <https://doi.org/10.1093/nar/gkp335>.
- Barbosa, V.A.A., Lery, L.M.S., 2019. Insights into *Klebsiella pneumoniae* type VI secretion system transcriptional regulation. *BMC Genom.* 20, 506. <https://doi.org/10.1186/s12864-019-5885-9>.
- Barer, M.R., 2012. Bacterial growth, physiology and death. In: *Medical Microbiology*. Elsevier, pp. 39–53. <https://doi.org/10.1016/B978-0-7020-4089-4.00019-6>.
- Bartlett, A., O'Malley, R.C., Huang, S.-S.C., Galli, M., Nery, J.R., Gallavotti, A., Ecker, J. R., 2017. Mapping genome-wide transcription-factor binding sites using DAP-seq. *Nat. Protoc.* 12, 1659–1672. <https://doi.org/10.1038/nprot.2017.055>.
- Bentley, G.J., Narayanan, N., Jha, R.K., Salvachúa, D., Elmore, J.R., Peabody, G.L., Black, B.A., Ramirez, K., De Capite, A., Michener, W.E., Werner, A.Z., Klingeman, D. M., Schindel, H.S., Nelson, R., Foust, L., Guss, A.M., Dale, T., Johnson, C.W., Beckham, G.T., 2020. Engineering glucose metabolism for enhanced muconic acid production in *Pseudomonas putida* KT2440. *Metab. Eng.* 59, 64–75. <https://doi.org/10.1016/j.ymben.2020.01.001>.
- Bernal, P., Allsopp, L.P., Filloux, A., Llamas, M.A., 2017. The *Pseudomonas putida* T6SS is a plant warden against phytopathogens. *ISME J.* 11, 972–987. <https://doi.org/10.1038/ismej.2016.169>.
- Blanco-Romero, E., Redondo-Nieto, M., Martínez-Granero, F., Garrido-Sanz, D., Ramos-González, M.I., Martín, M., Rivilla, R., 2018. Genome-wide analysis of the FleQ direct regulon in *Pseudomonas fluorescens* F113 and *Pseudomonas putida* KT2440. *Sci. Rep.* 8, 13145. <https://doi.org/10.1038/s41598-018-31371-z>.
- Bojanović, K., D'Arrigo, I., Long, K.S., 2017. Global transcriptional responses to osmotic, oxidative, and imipenem stress conditions in *Pseudomonas putida*. *Appl. Environ. Microbiol.* 83 <https://doi.org/10.1128/AEM.03236-16>.
- Chang, W.-S., van de Mortel, M., Nielsen, L., Nino de Guzman, G., Li, X., Halverson, L.J., 2007. Alginate production by *Pseudomonas putida* creates a hydrated microenvironment and contributes to biofilm architecture and stress tolerance under

- water-limiting conditions. *J. Bacteriol.* 189, 8290–8299. <https://doi.org/10.1128/JB.00727-07>.
- Chauhan, S.M., Poudel, S., Rychel, K., Lamoureux, C., Yoo, R., Al Bulushi, T., Yuan, Y., Palsson, B.O., Sastry, A.V., 2021. Machine learning uncovers a data-driven transcriptional regulatory network for the Crenarchaeal Thermoacidophile *Sulfolobus acidocaldarius*. *Front. Microbiol.* 12, 753521. <https://doi.org/10.3389/fmicb.2021.753521>.
- Checovich, W.J., Bolger, R.E., Burke, T., 1995. Fluorescence polarization—a new tool for cell and molecular biology. *Nature* 375, 254–256. <https://doi.org/10.1038/375254a0>.
- Chen, L., Zou, Y., Kronfl, A.A., Wu, Y., 2020. Type VI secretion system of *Pseudomonas aeruginosa* is associated with biofilm formation but not environmental adaptation. *Microbiol. Open* 9, e991. <https://doi.org/10.1002/mbo3.991>.
- Comon, P., 1994. Independent component analysis, A new concept? *Signal Process.* 36, 287–314. [https://doi.org/10.1016/0165-1684\(94\)90029-9](https://doi.org/10.1016/0165-1684(94)90029-9).
- Daddaoua, A., Krell, T., Ramos, J.-L., 2009. Regulation of glucose metabolism in *Pseudomonas*: the phosphorylative branch and enter-doudoroff enzymes are regulated by a repressor containing a sugar isomerase domain. *J. Biol. Chem.* 284, 21360–21368. <https://doi.org/10.1074/jbc.M109.014555>.
- Del Castillo, T., Duque, E., Ramos, J.L., 2008. A set of activators and repressors control peripheral glucose pathways in *Pseudomonas putida* to yield a common central intermediate. *J. Bacteriol.* 190, 2331–2339. <https://doi.org/10.1128/JB.01726-07>.
- Di Tommaso, P., Chatzou, M., Floden, E.W., Barja, P.P., Palumbo, E., Notredame, C., 2017. Nextflow enables reproducible computational workflows. *Nat. Biotechnol.* 35, 316–319. <https://doi.org/10.1038/nbt.3820>.
- Diez, S., Ryu, J., Caban, K., Gonzalez, R.L., Dworkin, J., 2020. The alarmones (pppGpp) directly regulate translation initiation during entry into quiescence. *Proc. Natl. Acad. Sci. U.S.A.* 117, 15565–15572. <https://doi.org/10.1073/pnas.1920013117>.
- Ester, Martin, Kriegel, Hans-Peter, Sander, Jörg, Xu, Xiaowei, 1996. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In *Proceedings of 2nd International Conference on KDD* 226–231.
- Ewels, P., Magnusson, M., Lundin, S., Käller, M., 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32, 3047–3048. <https://doi.org/10.1093/bioinformatics/btw354>.
- Fang, X., Lloyd, C.J., Palsson, B.O., 2020. Reconstructing organisms in silico: genome-scale models and their emerging applications. *Nat. Rev. Microbiol.* 18 (12), 731–743. <https://doi.org/10.1038/s41579-020-00440-4>.
- Galperin, M.Y., Makarova, K.S., Wolf, Y.I., Koonin, E.V., 2015. Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Res.* 43, D261–D269. <https://doi.org/10.1093/nar/gku1223>.
- Gao, Y., Lim, H.G., Verkler, H., Szubin, R., Quach, D., Rodionova, I., Chen, K., Yurkovich, J.T., Cho, B.-K., Palsson, B.O., 2021. Unraveling the functions of uncharacterized transcription factors in *Escherichia coli* using ChIP-exo. *Nucleic Acids Res.* 49 (17), 9696–9710. <https://doi.org/10.1093/nar/gkab735>.
- Guo, Z., Houghton, J.E., 1999. PcaR-mediated activation and repression of *pca* genes from *Pseudomonas putida* are propagated by its binding to both the -35 and the -10 promoter elements. *Mol. Microbiol.* 32, 253–263. <https://doi.org/10.1046/j.1365-2958.1999.01342.x>.
- Gupta, S., Stamatoyannopoulos, J.A., Bailey, T.L., Noble, W.S., 2007. Quantifying similarity between motifs. *Genome Biol.* 8, R24. <https://doi.org/10.1186/gb-2007-8-2-r24>.
- Hellman, L.M., Fried, M.G., 2007. Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. *Nat. Protoc.* 2, 1849–1861. <https://doi.org/10.1038/nprot.2007.249>.
- Herrgård, M.J., Covert, M.W., Palsson, B.O., 2004. Reconstruction of microbial transcriptional regulatory networks. *Curr. Opin. Biotechnol.* 15, 70–77. <https://doi.org/10.1016/j.copbio.2003.11.002>.
- Hervás, A.B., Canosa, I., Little, R., Dixon, R., Santero, E., 2009. NtrC-dependent regulatory network for nitrogen assimilation in *Pseudomonas putida*. *J. Bacteriol.* 191, 6123–6135. <https://doi.org/10.1128/JB.00744-09>.
- Huang, H., Shao, X., Xie, Y., Wang, T., Zhang, Y., Wang, X., Deng, X., 2019. An integrated genomic regulatory network of virulence-related transcriptional factors in *Pseudomonas aeruginosa*. *Nat. Commun.* 10, 2931. <https://doi.org/10.1038/s41467-019-10778-w>.
- Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S.K., Cook, H., Mende, D.R., Letunic, I., Rattai, T., Jensen, L.J., von Mering, C., Bork, P., 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* 47, D309–D314. <https://doi.org/10.1093/nar/gky1085>.
- Hyvärinen, A., 1999. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans. Neural Network.* 10, 626–634. <https://doi.org/10.1109/72.761722>.
- Jayakody, L.N., Johnson, C.W., Whitham, J.M., Giannone, R.J., Black, B.A., Cleveland, N.S., Klingemann, D.M., Michener, W.E., Oldstad, J.L., Vardon, D.R., Brown, R.C., Brown, S.D., Hettich, R.L., Guss, A.M., Beckham, G.T., 2018. Thermochemical wastewater valorization via enhanced microbial toxicity tolerance. *Energy Environ. Sci.* 11, 1625–1638. <https://doi.org/10.1039/C8EE00460A>.
- Johnson, C.W., Salvachúa, D., Rorrer, N.A., Black, B.A., Vardon, D.R., St John, P.C., Cleveland, N.S., Dominick, G., Elmore, J.R., Grundl, N., Khanna, P., Martinez, C.R., Michener, W.E., Peterson, D.J., Ramirez, K.J., Singh, P., VanderWall, T.A., Wilson, A.N., Yi, X., Biddy, M.J., Bomble, Y.J., Guss, A.M., Beckham, G.T., 2019. Innovative chemicals and materials from bacterial aromatic catabolic pathways. *Joule*. <https://doi.org/10.1016/j.joule.2019.05.011>.
- Karp, P.D., Billington, R., Caspi, R., Fulcher, C.A., Latendresse, M., Kothari, A., Keseler, I.M., Krummenacker, M., Midford, P.E., Ong, Q., Ong, W.K., Paley, S.M., Subhraveti, P., 2019. The BioCyc collection of microbial genomes and metabolic pathways. *Briefings Bioinf.* 20, 1085–1093. <https://doi.org/10.1093/bib/bbx085>.
- Kiliç, S., White, E.R., Sagitova, D.M., Cornish, J.P., Erill, I., 2014. CollecTF: a database of experimentally validated transcription factor-binding sites in bacteria. *Nucleic Acids Res.* 42, D156–D160. <https://doi.org/10.1093/nar/gkt1123>.
- Lamoureux, C.R., Decker, K.T., Sastry, A.V., McConn, J.L., Gao, Y., Palsson, B.O., 2021. PRECISE 2.0: an expanded high-quality RNA-seq compendium for *Escherichia coli* K-12 reveals high-resolution transcriptional regulatory structure. *bioRxiv*. <https://doi.org/10.1101/2021.04.08.439047>.
- Langmead, B., Trapnell, C., Pop, M., Salzberg, S.L., 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10, R25. <https://doi.org/10.1186/gb-2009-10-3-r25>.
- Lei, L., Chen, J., Liao, W., Liu, P., 2020. Determining the different mechanisms used by *Pseudomonas* species to cope with minimal inhibitory concentrations of zinc via comparative transcriptomic analyses. *Front. Microbiol.* 11, 573857. <https://doi.org/10.3389/fmicb.2020.573857>.
- Liao, Y., Smyth, G.K., Shi, W., 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930. <https://doi.org/10.1093/bioinformatics/btt656>.
- Liebermeister, W., 2002. Linear modes of gene expression determined by independent component analysis. *Bioinformatics* 18, 51–60. <https://doi.org/10.1093/bioinformatics/18.1.51>.
- Lim, C.K., Hassan, K.A., Penesyan, A., Loper, J.E., Paulsen, I.T., 2013. The effect of zinc limitation on the transcriptome of *Pseudomonas protegens* Pf-5. *Environ. Microbiol.* 15, 702–715. <https://doi.org/10.1111/j.1462-2920.2012.02849.x>.
- Lim, H.G., Eng, T., Banerjee, D., Alarcon, G., Lau, A.K., Park, M.-R., Simmons, B.A., Palsson, B.O., Singer, S.W., Mukhopadhyay, A., Feist, A.M., 2021. Generation of *Pseudomonas putida* KT2440 strains with efficient utilization of xylose and galactose via adaptive laboratory evolution. *ACS Sustain. Chem. Eng.* 9 (34), 11512–11523. <https://doi.org/10.1021/acssuschemeng.1c03765>.
- Lim, H.G., Fong, B., Alarcon, G., Magurudeniya, H.D., Eng, T., Szubin, R., Olson, C.A., Palsson, B.O., Gladden, J.M., Simmons, B.A., Mukhopadhyay, A., Singer, S.W., Feist, A.M., 2020. Generation of ionic liquid tolerant *Pseudomonas putida* KT2440 strains via adaptive laboratory evolution. *Green Chem.* 22, 5677–5690. <https://doi.org/10.1039/D0GC01663B>.
- Lundblad, J.R., Laurance, M., Goodman, R.H., 1996. Fluorescence polarization analysis of protein-DNA and protein-protein interactions. *Mol. Endocrinol.* 10, 607–612. <https://doi.org/10.1210/mend.10.6.8776720>.
- McConn, J.L., Lamoureux, C.R., Poudel, S., Palsson, B.O., Sastry, A.V., 2021. Optimal dimensionality selection for independent component analysis of transcriptomic data. *BMC Bioinf.* 22, 584. <https://doi.org/10.1186/s12859-021-04497-7>.
- Miyazaki, R., Yano, H., Sentchilo, V., van der Meer, J.R., 2018. Physiological and transcriptome changes induced by *Pseudomonas putida* acquisition of an integrative and conjugative element. *Sci. Rep.* 8, 5550. <https://doi.org/10.1038/s41598-018-23858-6>.
- Molina-Henares, M.A., Ramos-González, M.I., Daddaoua, A., Fernández-Escamilla, A.M., Espinosa-Urgel, M., 2017. FleQ of *Pseudomonas putida* KT2440 is a multimeric cyclic diguanylate binding protein that differentially regulates expression of biofilm matrix components. *Res. Microbiol.* 168, 36–45. <https://doi.org/10.1016/j.resmic.2016.07.005>.
- Nakamura, T., Suzuki-Minakuchi, C., Kawano, H., Kanesaki, Y., Kawasaki, S., Okada, K., Nojiri, H., 2020. H-NS family proteins drastically change their targets in response to the horizontal transfer of the catabolic plasmid pCARI. *Front. Microbiol.* 11, 1099. <https://doi.org/10.3389/fmicb.2020.01099>.
- Nielsen, L., Li, X., Halverson, L.J., 2011. Cell-cell and cell-surface interactions mediated by cellulose and a novel exopolysaccharide contribute to *Pseudomonas putida* biofilm formation and fitness under water-limiting conditions. *Environ. Microbiol.* 13, 1342–1356. <https://doi.org/10.1111/j.1462-2920.2011.02432.x>.
- Nikel, P.I., de Lorenzo, V., 2018. *Pseudomonas putida* as a functional chassis for industrial biocatalysis: from native biochemistry to trans-metabolism. *Metab. Eng.* 50, 142–155. <https://doi.org/10.1016/j.ymben.2018.05.005>.
- Novichkov, P.S., Kazakov, A.E., Ravcheev, D.A., Leyn, S.A., Kovaleva, G.Y., Sutorin, M.A., Kazanov, M.D., Riehl, W., Arkin, A.P., Dubchak, I., Rodionov, D.A., 2013. RegPrecise 3.0—a resource for genome-scale exploration of transcriptional regulation in bacteria. *BMC Genom.* 14, 745. <https://doi.org/10.1186/1471-2164-14-745>.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É., 2011. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* 12 (85), 2825–2830.
- Pobre, V., Graça-Lopes, G., Saramago, M., Ankenbauer, A., Takors, R., Arraiano, C.M., Viegas, S.C., 2020. Prediction of novel non-coding RNAs relevant for the growth of *Pseudomonas putida* in a bioreactor. *Microbiology* 166, 149–156. <https://doi.org/10.1099/mic.0.000875>.
- Poudel, S., Tsunemoto, H., Seif, Y., Sastry, A.V., Szubin, R., Xu, S., Machado, H., Olson, C.A., Anand, A., Pogliano, J., Nizet, V., Palsson, B.O., 2020. Revealing 29 sets of independently modulated genes in *Staphylococcus aureus*, their regulators, and role in key physiological response. *Proc. Natl. Acad. Sci. U.S.A.* 117, 17228–17239. <https://doi.org/10.1073/pnas.2008413117>.
- Rajput, A., Tsunemoto, H., Sastry, A.V., Szubin, R., Rychel, K., Sugie, J., Pogliano, J., Palsson, B.O., 2021. Machine Learning of *Pseudomonas aeruginosa* transcriptomes identifies independently modulated sets of genes associated with known transcriptional regulators. *bioRxiv*. <https://doi.org/10.1101/2021.07.28.454220>.
- Rinas, U., Hellmuth, K., Kang, R., Seeger, A., Schlieker, H., 1995. Entry of *Escherichia coli* into stationary phase is indicated by endogenous and exogenous accumulation of nucleobases. *Appl. Environ. Microbiol.* 61, 4147–4151. <https://doi.org/10.1128/aem.61.12.4147-4151.1995>.

- Rychel, K., Decker, K., Sastry, A.V., Phaneuf, P.V., Poudel, S., Palsson, B.O., 2021. iModulonDB: a knowledgebase of microbial transcriptional regulation derived from machine learning. *Nucleic Acids Res.* 49, D112–D120. <https://doi.org/10.1093/nar/gkaa810>.
- Rychel, K., Sastry, A.V., Palsson, B.O., 2020. Machine learning uncovers independently regulated modules in the *Bacillus subtilis* transcriptome. *Nat. Commun.* 11, 6338. <https://doi.org/10.1038/s41467-020-20153-9>.
- Saelens, W., Cannoodt, R., Saey, Y., 2018. A comprehensive evaluation of module detection methods for gene expression data. *Nat. Commun.* 9, 1090. <https://doi.org/10.1038/s41467-018-03424-4>.
- Santos-Zavaleta, A., Salgado, H., Gama-Castro, S., Sánchez-Pérez, M., Gómez-Romero, L., Ledezma-Tejeda, D., García-Sotelo, J.S., Alquicira-Hernández, K., Muñoz-Rascado, L. J., Peña-Loredo, P., Ishida-Gutiérrez, C., Velázquez-Ramírez, D.A., Del Moral-Chávez, V., Bonavides-Martínez, C., Méndez-Cruz, C.-F., Galagan, J., Collado-Vides, J., 2019. RegulonDB v 10.5: tackling challenges to unify classic and high throughput knowledge of gene regulation in *E. coli* K-12. *Nucleic Acids Res.* 47, D212–D220. <https://doi.org/10.1093/nar/gky1077>.
- Sastry, A.V., Gao, Y., Szubin, R., Hefner, Y., Xu, S., Kim, D., Choudhary, K.S., Yang, L., King, Z.A., Palsson, B.O., 2019. The *Escherichia coli* transcriptome mostly consists of independently regulated modules. *Nat. Commun.* 10, 5536. <https://doi.org/10.1038/s41467-019-13483-w>.
- Sastry, A.V., Poudel, S., Rychel, K., Yoo, R., Lamoureux, C.R., Chauhan, S., Haiman, Z.B., Al Bulushi, T., Seif, Y., Palsson, B.O., 2021. Mining all publicly available expression data to compute dynamic microbial transcriptional regulatory networks. *bioRxiv*. <https://doi.org/10.1101/2021.07.01.450581>.
- Smith, D.R., Chapman, M.R., 2010. Economical evolution: microbes reduce the synthetic cost of extracellular proteins. *mBio* 1. <https://doi.org/10.1128/mBio.00131-10>.
- Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N.T., Morris, J.H., Bork, P., Jensen, L.J., Mering, C. von, 2019. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 47, D607–D613. <https://doi.org/10.1093/nar/gky1131>.
- Wackett, L.P., 2003. *Pseudomonas putida*—a versatile biocatalyst. *Nat. Biotechnol.* 21, 136–138. <https://doi.org/10.1038/nbt0203-136>.
- Wang, L., Wang, S., Li, W., 2012. RSeQC: quality control of RNA-seq experiments. *Bioinformatics* 28, 2184–2185. <https://doi.org/10.1093/bioinformatics/bts356>.
- Winsor, G.L., Griffiths, E.J., Lo, R., Dhillon, B.K., Shay, J.A., Brinkman, F.S.L., 2016. Enhanced annotations and features for comparing thousands of *Pseudomonas* genomes in the *Pseudomonas* genome database. *Nucleic Acids Res.* 44, D646–D653. <https://doi.org/10.1093/nar/gkv1227>.
- Xiao, Y., Zhu, W., Liu, H., Nie, H., Chen, W., Huang, Q., 2018. FinR regulates expression of nicC and nicX operons, involved in nicotinic acid degradation in *Pseudomonas putida* KT2440. *Appl. Environ. Microbiol.* 84 <https://doi.org/10.1128/AEM.01210-18>.
- Yoo, R., Rychel, K., Poudel, S., Al-bulushi, T., Yuan, Y., Chauhan, S., Lamoureux, C., Palsson, B., Sastry, A.V., 2022. Machine learning of all *Mycobacterium tuberculosis* H37Rv RNA-seq data reveals a structured interplay between metabolism, stress response, and infection. *mSphere*, e0003322. <https://doi.org/10.1101/2021.07.01.450045>.