Elucidation of regulatory modes for five two-component systems in *Escherichia coli* reveals novel relationships

Kumari Sonal Choudhary[1], Julia A. Kleinmanns[1], Katherine Decker[1], Anand V Sastry[1], Ye Gao[1], Richard Szubin[1], Yara Seif[1], Bernhard O. Palsson[1,2,3]

1. Department of Bioengineering, University of California, San Diego, San Diego-92122, CA
2. Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, 2800 Lyngby, Denmark
3. Address correspondence to Bernhard O. Palsson, **E-mail:** palsson@ucsd.edu

Running Title: Transcriptional regulatory network of TCSs in *E. coli*

Kumari Sonal Choudhary, Julia A. Kleinmanns and Katherine Decker contributed equally. Author order was determined based on contributions to experiments and analysis.

Abstract

Escherichia coli uses two-component systems (TCSs) to respond to environmental signals. TCSs affect gene expression and are parts of *E. coli*'s global transcriptional regulatory network (TRN). Here, we identified the regulons of five TCSs in *E. coli* MG1655: BaeSR and CpxAR, which were stimulated by ethanol stress; KdpDE and PhoRB, induced by limiting potassium and phosphate, respectively; and ZraSR, stimulated by zinc. We analyzed RNA-seq data using independent component analysis (ICA). ChIP-exo data was used to validate condition-specific target gene binding sites. Based on this data we (1) identify the target genes for each TCS; (2) show how the target genes are transcribed in response to stimulus; and (3) reveal novel relationships between TCSs, which indicate non-cognate inducers for various response regulators, such as BaeR to iron starvation, CpxR to phosphate limitation, and PhoB and ZraR to cell envelope stress. Our understanding of the TRN in *E. coli* is thus notably expanded.

Importance

E. coli is a common commensal microbe found in human gut microenvironment; however, some strains cause diseases like diarrhea, urinary tract infections and meningitis. *E. coli's* two-component system (TCS) modulates target gene expression, specially related to virulence, pathogenesis and anti-microbial peptides, in response to environmental stimuli. Thus, it is of utmost importance to understand the transcriptional regulation of the TCSs to infer its environmental adaptation and disease pathogenicity. Utilizing a combinatorial approach integrating RNAseq, independent component analysis, ChIP-exo and data mining, we show that TCSs have five different modes of transcriptional regulation. Our data further highlights non-

cognate inducers of TCSs emphasizing cross-regulatory nature of TCSs in *E. coli* and suggests that TCSs may have a role beyond their cognate functionalities. In summary, these results when further incorporated with genome scale metabolic models can lead to understanding of metabolic capabilities of bacteria and correctly predict complex phenotype under diverse conditions.

Keywords

Two-component systems, *E. coli*, independent component analysis, transcriptomics, ChIP-exo, transcriptional regulatory network, gene targets

Introduction

Bacterial survival and resilience across diverse conditions relies upon environmental sensing and a corresponding response. One pervasive biological design towards this goal consists of a histidine kinase unit to sense the environment and a related response regulator unit to receive the signal and translate it into gene expression changes. This signaling process is known as a two-component system (TCS) (1). In the case of *Escherichia coli (E. coli)* strain K12 MG1655, there are 30 histidine kinases and 32 response regulators involved in 29 complete two-component systems that mediate responses to various environmental stimuli such as metal sensing, cell envelope stress, acid stress, and pH stress (2). The cell envelope is an important barrier between bacteria and their surrounding environment and is exposed to a variety of stresses and stimuli. In *E. coli* MG1655, BaeSR and CpxAR are two TCSs that are each instrumental in regulating the response to envelope stress (3). KdpDE and ZraSR are key

regulators in maintaining homeostasis for potassium and zinc, respectively, and PhoRB has a role in bacterial pathogenesis and phosphate homeostasis (4). Understanding these elements is important for developing an overarching knowledge of bacterial regulatory networks. A detailed reconstruction of the transcriptional regulatory network (TRN) contribution of these systems in response to a specific stimulus will help us discover different modes of transcriptional regulation. Furthermore, this can unravel cross-regulation and complex relationships amongst TCS systems.

Multiple tools exist to investigate the regulons and activities of transcription factors and TCSs. In this study, we focused on two data-types: chromatin immunoprecipitation coupled with exonuclease treatment (ChIP-exo) binding peaks, and gene expression profiles from RNA sequencing (RNA-seq). ChIP-exo identifies genome-wide binding locations for response regulators with high precision (5). Although ChIP methods provide direct binding evidence for the interactions between TFs and the genome, they also tend to identify off-target, nonspecific binding events that may not affect downstream gene expression (6). In order to reduce the false positive rate from ChIP-exo, gene expression profiles can identify differentially expressed genes (DEGs) in response to appropriate stimuli or response regulator knock-outs. However, some DEGs may be affected by indirect regulation. To circumvent this issue, we also used independent component analysis (ICA), an unsupervised machine learning algorithm that decomposes gene expression data, to find independently modulated sets of genes, or iModulons (Figure 1A), that closely resemble known regulons (7).

A previous application of ICA to PRECISE (a high-quality RNA-seq gene expression compendium) (7) identified nine iModulons that corresponded to eleven TCS regulons. We

noticed that two sets of TCS pairs were grouped together: YpdAB and BtsRS, and HprSR and CusSR (7). Interestingly, YpdAB and BtsRS respond to the same inducer (pyruvate). Conversely, HprSR and CusSR both regulate the same set of genes but have different inducing signals. This second observation is in agreement with previous studies (8), and is indicative of cross-regulation between the two TCSs wherein they have the same regulon. ICA generates iModulon activity levels to determine gene expression changes relative to a specific condition. Indeed, these previous analyses showed that ICA can be aptly used to describe condition-specific activities (YpdAB and BtsRS) and provide a detailed understanding of tightly cross-regulated TRNs (HprSR and CusSR).

In this study, we aimed to characterize global TRN contributions of these five TCSs (BaeSR, CpxAR, KdpDE, PhoRB, and ZraSR) in *E. coli* MG1655. We used three methods (ICA, ChIP-exo, and Differential Expression of Genes (DEGs)) to reveal which genes were regulated by the TCS and characterized the modes of transcriptional regulation into five categories: 1) Functional direct binding, 2) Indirect targets, 3) Non-functional/spurious binding, 4) Cross-regulated genes, and 5) Hypothetically functional binding. Finally, we used the iModulon activities to characterize TCS activation across hundreds of experimental conditions to identify interconnectedness among TCSs.

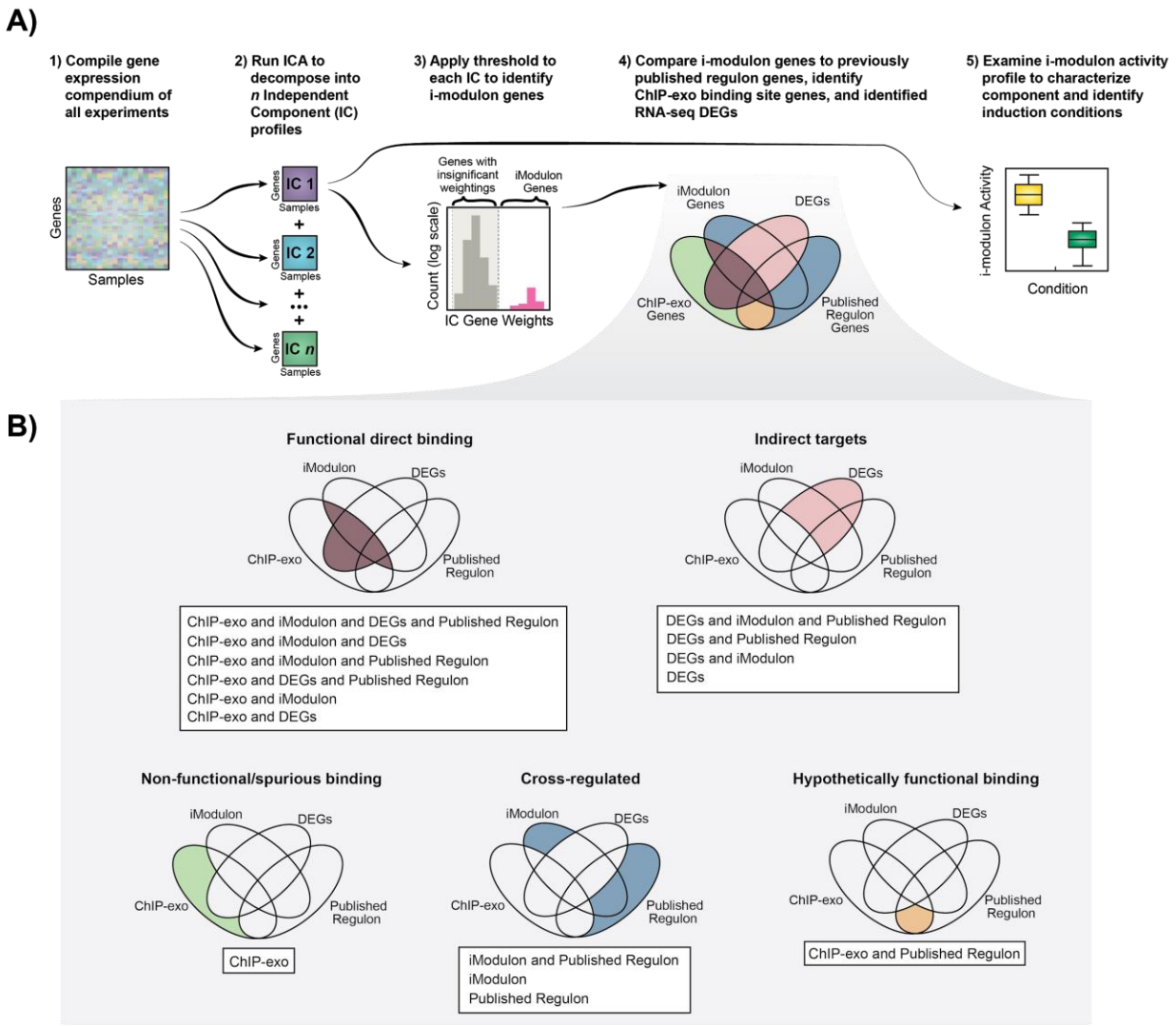


Figure 1: A) Workflow used for each gene expression dataset. B) Categories of regulation by two-component systems in *E. coli*.

Results

We generated knockout mutants for each of the five two-component systems (BaeSR, CpxAR, KdpDE, PhoRB, and ZraSR) in *E. coli* K-12 strain MG1655 (Table 1), and collected RNA-seq data for both wild-type and knockout mutants under stimulated and unstimulated conditions, using specific stimuli for each TCS. We combined these expression profiles with 278 additional expression profiles previously generated (called PRECISE (7)), and applied the ICA algorithm to the combined dataset (See Methods) to identify 101 iModulons. Of these 101 iModulons, 90 iModulons mapped to the previous set of 92 iModulons from PRECISE (7). Eleven iModulons were specifically activated by the inclusion of the TCS-related expression profiles, four of which represented TCSs (BaeSR, KdpDE, PhoRB, and ZraSR). The CpxR iModulon was among the 90 iModulons previously detected in PRECISE, and was retained upon addition of the new expression profiles. In total, 15 of the 101 iModulons represented TCSs (Supplemental Dataset S1). For further analysis, we focused on the five target TCSs (BaeSR, CpxAR, KdpDE, PhoRB, and ZraSR).

Strains	**Phenotype**	**Stimulation condition**	**Reference**	**Unstimulated condition**	**Reference**
E. coli K12 MG1655	Wild type	1) LB medium + 5% Ethanol 2) Tris maleic acid minimal medium (TMA) + 0.1 mM KCl 3) M9 minimal medium without phosphate (M9-P) 4) LB medium + 1 mM $ZnCl_2$	This study	1) LB medium 2) Tris maleic acid minimal medium (TMA) + 115 mM KCl 3) M9 minimal medium	This study
⊿baeR	*baeR* knockout mutant of *E. coli*	LB medium + 5% Ethanol	This study	LB medium	This study

	K12 MG1655				
⊿cpxR	*cpxR* knockout mutant of *E. coli* K12 MG1655	LB medium + 5% Ethanol	This study	LB medium	This study
⊿kdpE	*kdpE* knockout mutant of *E. coli* K12 MG1655	Tris maleic acid minimal medium (TMA) + 0.1 mM KCl	This study	Tris maleic acid minimal medium (TMA) + 115 mM KCl	This study
⊿phoB	*phoB* knockout mutant of *E. coli* K12 MG1655	M9 minimal medium without phosphate (M9-P)	This study	M9 minimal medium	*(7)*
⊿zraR	*zraR* knockout mutant of *E. coli* K12 MG1655	LB medium + 1 mM $ZnCl_2$	This study	LB medium	This study

Table 1: List of strains used in this study

To identify the true regulons for each TCS, we compared four sets of genes: (1) Genes in the iModulon associated with the respective TCS, (2) Genes with ChIP-exo determined upstream binding sites, (3) Genes that are differentially expressed between wild-type and knockout mutants under stimulated conditions, and (4) Genes in previously published regulons in RegulonDB (9) (Figure 1B). Each method alone is prone to false positives and biases, but together provide multidimensional information that can be probed to identify a complete, high-confidence regulon (Figure 1B). For example, differentially expressed genes (DEGs) without direct ChIP-exo evidence are likely targets of downstream, indirect regulation wherein one TF may regulate a set of genes which in turn may lead to regulation of a second set of genes. On the other hand, ChIP-exo binding sites upstream of genes that are absent from both the iModulon and the DEG set are likely non-functional. Such integrated analysis led us to describe the

regulatory network of each selected two-component system into five categories (Figure 1B): **1) Functional direct binding:** genes which were observed in either iModulons, published regulons, or DEGs, and a ChIP-exo binding peak was observed upstream; **2) Indirect targets**: genes that were differentially expressed under each specific condition, but no binding peaks were observed upstream; **3) Non-functional/spurious binding:** peaks upstream of genes which were neither differentially expressed nor had iModulons and/or published regulons associated to them. Such peaks may have arisen due to off-target recognition or due to hot spots which tend to be recognized by many TFs; **4) Cross-regulated genes**: genes which were observed in iModulons and/or published regulons but were not part of functional direct targets and indirect targets; and **5) Hypothetically functional binding**: genes which were found in the published regulon and showed ChIP-exo binding peaks but were not differentially expressed. These genes may be a part of a direct regulatory network, but the data was not supported by transcriptomics; therefore, we categorized them into hypothetical functional binding.

1) Reconstruction of CpxR and BaeR regulatory responses under ethanol stress in *E. coli*

BaeR and CpxR are activated by a wide range of stressors (10–14) and play a regulatory role in the envelope stress response (ESR) system (10, 15–17). We stimulated the wild-type, *ΔbaeR* and *ΔcpxR* strains with 5% ethanol stress, which has been identified as a particularly effective inducer (18) to elucidate their regulons.

BaeR

BaeR is involved in multidrug resistance through regulation of the MdtABC efflux pump (19–21). However, BaeR has also been shown to impact genes related to many other functions, such as signal transduction, chemotactic responses, flagellar biosynthesis, and maltose transport (22). Despite this range of functionality, there are only eight genes in the previously published BaeR regulon. The BaeR iModulon exhibited 37.5% (3 out of 8) overlap with these published regulon genes, consisting of *spy*, *baeR*, and *mdtA* genes (Table 2). Five of the published regulon genes (*mdtB, mdtC, mdtD/iceT, baeS,* and *acrD*) were not identified in the BaeR iModulon. This discrepancy was most likely due to differences in BaeR induction conditions across the published results (10–13), or the possibility that the three iModulon genes are much more significantly regulated by BaeR than the other five genes in the published regulon. To test this claim, ChIP-exo results were analyzed (Supplemental Dataset S2) then compared to the iModulon. ChIP-exo results indicated 21 BaeR binding peaks upstream of 17 operons consisting of 32 genes. We identified binding sites upstream of *spy* and *mdtA* genes (*mdtABCD-baeSR* operon). Since a peak was identified upstream of *mdtA* gene, which forms a part of *mdtABCD-baeSR* operon (*see methods*), we considered all the genes in that operon to be under the direct regulatory network of BaeR. In total, we designated six genes as functional direct targets of BaeR. No binding peaks were identified upstream of *acrD*, indicating that it is indirectly regulated by BaeR.

Differentially expressed genes (DEGs) were analyzed to further expand the global regulatory network of BaeR. A total of 328 genes were found to be differentially expressed due to the *baeR* knockout under 5% ethanol stress condition (Supplemental Dataset S3). ChIP-exo

peaks were detected upstream of nine of these genes, expanding the number of direct targets from seven to nine. Of these nine targets, two genes, *intF* and *tnaA*, have not been previously identified as a part of the BaeR regulon. However, *tnaA* has previously been shown to contribute to BaeR activity under indole stress (23), as it is involved in metabolizing tryptophan to indole. The remaining 319 differentially expressed genes could either be indirect targets of BaeR (Figure 1, Supplemental Dataset S4) or other transcription factors responsive to ethanol such as CpxR. Functional analysis of indirect targets through COG categories confirmed BaeR's manifold activity (Figure 2).

Overall, 2.74% of DEGs were under the direct regulatory network of BaeR. The remaining 23 genes found through ChIP-exo results could be classified as non-functional targets or cases of spurious binding (Figure 1, Supplemental Dataset S5). In most of the cases we did identify genes being regulated by BaeR knockout but they either did not pass the P-value threshold (< 0.05) or log 2 fold change threshold (1.5). Others may be the result of off-target binding.

Response Regulator	**iModulon Genes**	**Targets from RegulonDB**	**Direct targets**
BaeR	***baeR***, ***mdtA***, ***spy***	*acrD*, ***baeR***, *baeS*, *mdtD*, ***mdtA***, *mdtB*, *mdtC*, ***spy***	*baeR*, *baeS*, *mdtD*, *mdtA*, *mdtB*, *mdtC*, *spy*, *intF*, *tnaA*
CpxR	*alx*, ***baeR***, ***cpxP***, ***cpxR***, ***dgcZ***, ***ftnB***, ***ldtC***, *mdtJ*, *raiA*, ***tomB***, *yagU*, ***yccA***,	*acrD*, *aroG*, *bacA*, ***baeR***, *baeS*, *bamE*, *cheA*, *cheW*, *cpxA*, ***cpxP***, ***cpxR***, *csgA*, *csgB*, *csgC*, *csgD*, *csgE*, *csgF*, *csgG*, *degP*, ***dgcZ***, *dsbA*, *dsbC*, *efeU_2*, *fabZ*, ***ftnB***, *hha*, *iceT*, ***ldtC***, *ldtD*, *lpxA*, *lpxD*, *marA*, *marB*, *marR*, *mdtA*, *mdtB*, *mdtC*, *motA*, *motB*, *mscM*, *mzrA*, *ompC*, *ompF*, *ppiA*, *ppiD*, *psd*, *rpoE*,	*cpxA*, *cpxP*, *cpxR*, *yagU*, *yccA*, *fimA*, *fimI*, *lldD*, *lldR*, *ppiA*, *yceI*

	yebE, *yjfN*, *yncJ*, *yobB*	*rpoH*, *rseA*, *rseB*, *rseC*, *sbmA*, *skp*, *slt*, *spy*, *srkA*, ***tomB***, *tsr*, *ung*, *yaiW*, ***yccA***, ***yebE***, *ygaU*, *yidQ*, *yqaE*, *yqjA*	
KdpE	***kdpA***, ***kdpB***, ***kdpC***, *kdpD*, *kdpE*	***kdpA***, ***kdpB***, ***kdpC***, *kdpF*	*kdpA*, *kdpB*, *kdpC*, *adeP*, *uxaA*, *uxaC*
PhoB	***phoB***, ***phoR***, ***pstA***, ***pstB***, ***pstC***, ***pstS***	*adiC*, *amn*, *argP*, *asr*, *cra*, *cusA*, *cusB*, *cusC*, *cusF*, *cusR*, *cusS*, *eda*, *feaR*, *gadW*, *gadX*, *hiuH*, *mipA*, *ompF*, *phnC*, *phnD*, *phnE_1*, *phnE_2*, *phnF*, *phnG*, *phnH*, *phnI*, *phnJ*, *phnK*, *phnL*, *phnM*, *phnN*, *phnO*, *phnP*, *phoA*, ***phoB***, *phoE*, *phoH*, *phoQ*, ***phoR***, *phoU*, *pitB*, *prpR*, *psiE*, *psiF*, ***pstA***, ***pstB***, ***pstC***, ***pstS***, *rspR*, *sbcC*, *sbcD*, *tktB*, *ugpA*, *ugpB*, *ugpC*, *ugpE*, *ugpQ*, *waaH*, *yegH*, *yhjC*	*phoB*, *phoR*, *pstA*, *pstB*, *pstC*, *pstS*, *phoU*
ZraR	***zraP***, ***zraR***	***zraP***, ***zraR***, *zraS*	*zraP*, *zraR*, *zraS*, *mgtA*

Table 2: List of genes in each iModulon and published regulon, and respective direct targets. Genes in bold text in "iModulon Genes" and "Targets from RegulonDB" columns indicate common genes.

CpxR

CpxR is among the most extensively studied of the response regulators, which is unsurprising when considering its host of functions. CpxR has been shown to have a regulatory role in the envelope stress response (ESR) system (10, 15–17), protein folding and degradation (24–27), pilus assembly and expression (28–30), secretion (15), motility and chemotaxis (15, 31), biofilm development (14, 32), adherence (33, 34), multidrug resistance and efflux (12, 35), porins (36), and copper response (37, 38), among others. Two iModulons were linked to the CpxR regulon: CpxR and CpxR-KO. The CpxR iModulon consists of genes that were regulated specifically under wild type conditions, and CpxR_KO consists of genes differentially expressed due to CpxR deletion. Collectively, both CpxR iModulons consist of 16 genes that cover diverse

functions such as motility, inorganic ion transport and metabolism, and carbohydrate metabolism, among others. Nine of the 16 iModulon genes were previously published regulon genes, which is only a fraction (14%) of CpxR's previously-identified 66 genes (9) (Table 2). This provides extra incentive to eventually validate the CpxR iModulon with different induction conditions. Interestingly, the CpxR iModulon consisted of seven new genes that were not previously identified to be part of the CpxR regulon (Table 2). These seven new genes may indicate expansion of the CpxR regulon specific to ethanol stress. To validate this hypothesis, ChIP-exo peaks were examined. We were able to identify 44 binding peaks upstream of 16 operons (consisting of 30 genes). Of these genes, five were in the CpxR iModulon (*cpxP*, *cpxR*, *raiA*, *yagU*, and *yccA*). Two genes (*raiA* and *yagU)* in the CpxR iModulon with associated binding peaks had not been previously identified as part of the CpxR regulon. The five overlapping ChIP-exo/iModulon/regulon genes were classified as direct targets of CpxR under ethanol. For the remaining five iModulon genes that are not previously published regulon genes and do not have binding peaks identified, 60% are part of the uncharacterized "y-ome" (39). These genes may be the result of cross-regulation between different pathways responding to ethanol and put under the category of *cross-regulatory network*.

We inspected DEGs to further understand the nature of the five new iModulon genes and to expand the global regulatory network of CpxR. Differential gene expression analysis expanded the indirect targets of CpxR to include 368 genes. Comparison of ChIP-exo binding peaks and differentially expressed genes of the *cpxR* knockout vs. wild type under 5% ethanol stress condition revealed seven additional genes (*cpxA, fimA, fimI, lldD, lldR, ppiA,* and *yceI*) that fall under the direct regulatory network of CpxR, expanding this network to 12 genes. The

244 remaining 357 DEGs formed the indirect regulon of CpxR, including eight genes from the
245 previously published regulon (Figure 2, Supplemental Dataset S4). A diverse group of COG
246 categories corroborated CpxR's regulatory role.
247
248 Interestingly, among the 368 differentially expressed genes (DEGs), 194 were
249 differentially expressed in both knockout strains (*baeR and cpxR*), suggesting non-specific
250 targets of BaeR and CpxR and high cross-regulation between these two pathways. The remaining
251 17 ChIP-exo binding peaks overlapped with one previously published regulon gene (*ompC*)
252 which was not part of either the DEGs or iModulons. This suggested that *ompC* was neither
253 directly nor indirectly affected by *cpxR* knockout under ethanol stress, and therefore is
254 considered to be part of the "hypothetically functional binding" network (Supplemental Dataset
255 S6). Five iModulon genes (*alx, mdtJ, yjfN, yliI,* and *yobB*) did not form part of direct and indirect
256 targets but instead formed part of the network which is involved in cross-regulation. The
257 remaining 52 previously published regulon genes which did not have binding sites upstream and
258 were not differentially expressed also formed part of a *cross-regulatory network* (Supplemental
259 Dataset S7).
260

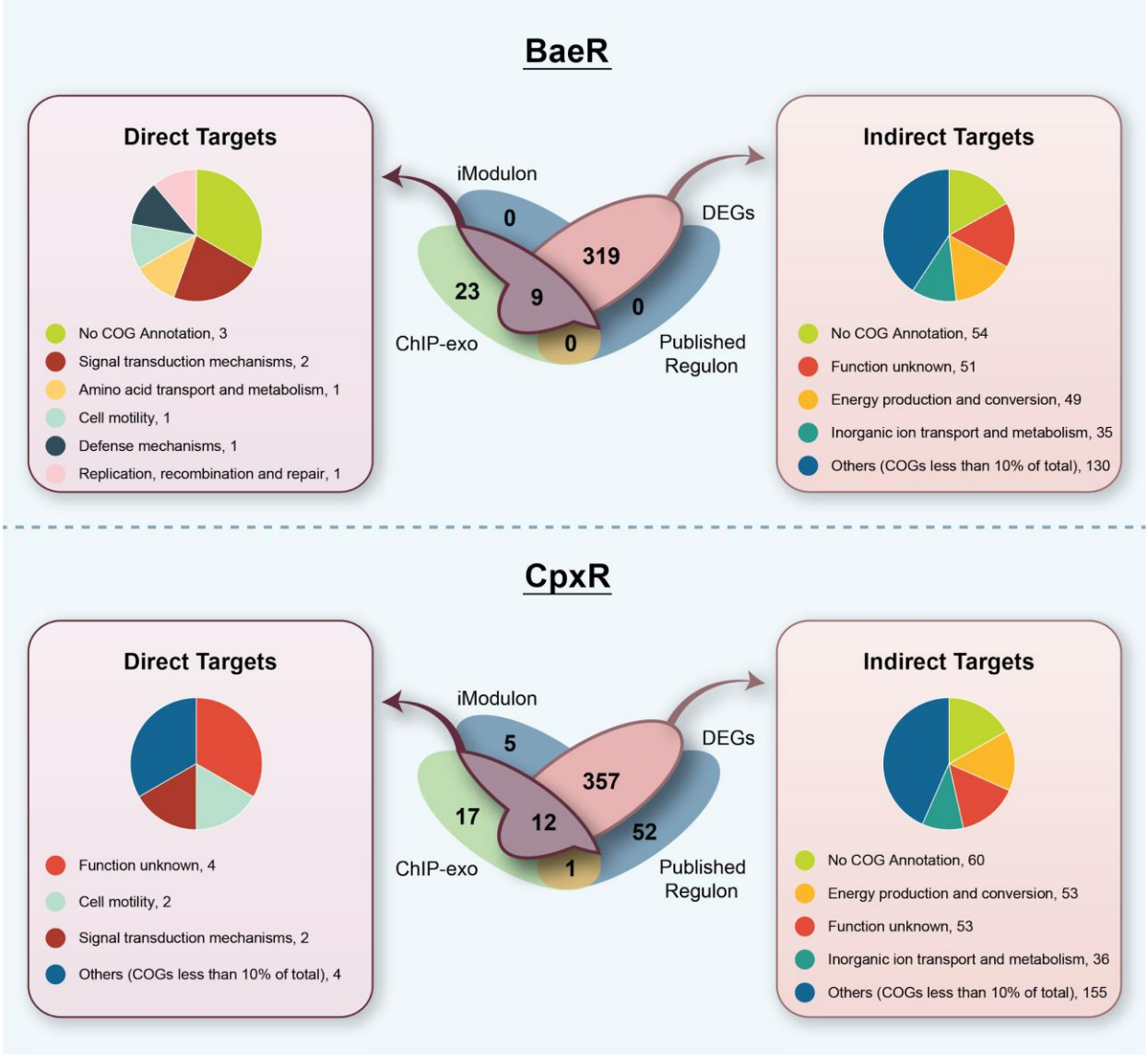


Figure 2: Comparison of regulatory network in BaeR and CpxR and functional characterization of direct and indirect targets.

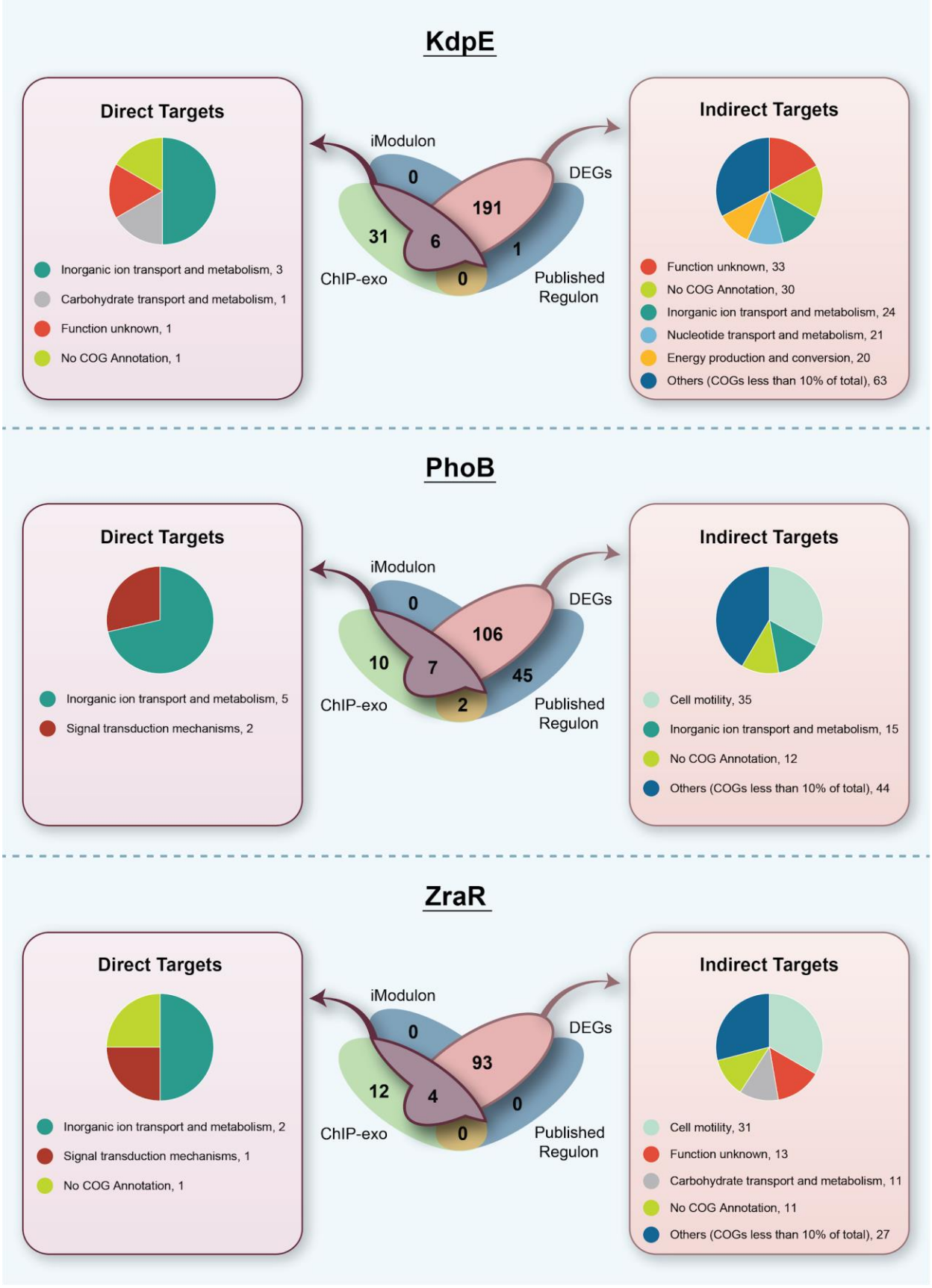
2) Reconstruction of KdpE, PhoB, and ZraR transcriptional regulatory network in *E. coli*

Decoding the role of metal and nutrient sensors in the transcriptional regulatory network may provide a deeper understanding of how bacteria utilize these systems to sense nutrients and ionic strength of the environment. These mechanisms play a crucial role in the organism's ability to adapt to various environmental niches. The KdpE, PhoB, and ZraR two-component systems were selected for further analysis in this category.

KdpE

The KdpE response regulator's sole target is the *kdpFABC* operon, which has been shown to activate the potassium uptake system during low potassium conditions (40, 41) or salt stress (42, 43). To induce KdpE, a small amount of potassium (0.1 mM KCl) was added to the cells in tris maleic acid (TMA) minimal medium, which aligns with previous experiments where successful induction was achieved (44, 45). The KdpE iModulon included the expected *kdpA*, *kdpB*, and *kdpC* potassium uptake genes; however, the iModulon does not include the fourth gene in the *kdpFABC* operon, *kdpF* (Table 2). This result provides validation of previous studies that have instead described the target operon as *kdpABC* (43, 46). *kdpF* may function only as a stabilizer of the transporter complex (47), and the gene may also just be a less vital component of the regulon. Apart from the *kdpABC* operon, the only other KdpE iModulon genes include those that code for the TCS itself, *kdpD* and *kdpE*. The small KdpE iModulon confirms the existing knowledge of the response regulator as targeting a single locus in the bacterial genome. Despite the small iModulon size, there were still 31 ChIP-exo binding peaks upstream of 37 genes including *kdpABC* genes from KdpE iModulon observed. Surprisingly, no binding peaks were observed upstream of two iModulon genes, *kdpDE*; however, *kdpDE* is downstream of the

292 *kdpFABC* operon, which may indicate that *kdpDE* is in the same transcription unit as *kdpABC*.
293 Consistent with this, when comparing 197 DEGs in response to *kdpE* knockout under potassium
294 limited condition (0.1 mM KCl), we did find *kdpDE* to be differentially expressed. DEGs further
295 added three extra genes (*adeP*, *uxaA*, and *uxaC*) to the direct regulatory network of KdpE,
296 amounting to six genes under direct regulation of KdpE. The remaining 191 DEGs can be
297 interpreted as potassium-dependent genes under indirect KdpE regulation. As expected, the COG
298 category 'inorganic ion transport' was overrepresented. In addition, the categories 'energy
299 production and conversion' and 'metabolism and nucleotide transport and metabolism' were also
300 over represented. A wide array of COG categories suggests that KdpDE may indirectly regulate
301 genes involved in metal efflux pump and electron transport systems to maintain homeostasis
302 (Figure 3).
303
304 Notably, there was evidence of *kdpF* being down-expressed in the KdpE knockout (log
305 fold change of -3.22) but due to an insignificant p-value, this was not regarded as differentially
306 expressed. Therefore, *kdpF* appears as the sole gene from the KdpE published regulon that was
307 not identified by DEG analysis in this study (placing it in the category of "cross-regulated"
308 genes). Apart from *kdpF* and genes categorized as direct or indirect targets for KdpDE, there
309 were also 31 genes identified with ChIP-exo that did not overlap with other datasets and were
310 thus categorized as potential spurious binding genes.



311

Figure 3: Comparison of regulatory network in metal sensors and functional characterization of direct and indirect targets.

PhoB

The PhoB response regulator has been demonstrated to regulate phosphate uptake and metabolism under phosphate-limiting conditions (48–51) and is linked to virulence in pathogenic *E. coli (4)*. To induce PhoB via phosphate deficiency, cells were first grown in M9 minimal medium and then washed with and further grown in M9 minimal medium without phosphate (M9-P). The PhoB iModulon consisted of the expected *pstSCAB* operon for regulation of the phosphate uptake system (52–54) and the genes coding for PhoRB itself. Contrary to the modest six gene iModulon, the previously published PhoB regulon is considerably larger and is thought to consist of 60 genes (55, 56), including all six of the iModulon genes. For the binding site analysis, we found 12 ChIP-exo binding peaks upstream of 17 genes. Six of these genes were included in the iModulon (the *pstSCAB phoBR* operons), indicating direct response to *phoB* knockout. When comparing the 113 DEGs under phosphate limited condition to the ChIP-exo binding peaks, an additional gene, *phoU* -phosphate specific transporter (which was also part of the previously published regulon), was added to the direct regulatory network of PhoB. The remaining 106 DEGs can be identified as phosphate-dependent genes, including six previously-determined regulon genes (*cusB, phnC, phnD, phnN, phoA,* and *ugpB*) or indirect targets of PhoB. Interestingly, the COG category ‘cell motility’ was more pronounced than ‘inorganic ion transport and metabolism’, indicative of the involvement of PhoRB in cell migration in addition to adapting to a phosphate limited environment. Aside from the direct and indirect targets, the

335 remaining 57 genes found to be associated with PhoB included 10 genes that were identified as a
336 potential result of spurious binding, two genes that fell under the hypothetically functional
337 binding category, and 45 genes that were possibly related to the KdpE network via cross-
338 regulation.
339
340 **ZraR**
341
342 Finally, we chose to examine ZraR (formerly called HydG), which had been previously
343 characterized as a σ^{54}-dependent response regulator that activates chaperone ZraP to provide
344 tolerance to high zinc concentrations (57–59). More recently, it has been proposed that ZraR is
345 not directly involved in zinc resistance (60) but rather is simply activated by zinc to achieve its
346 broader role in the envelope stress response (ESR), similar to CpxAR (61). Since the
347 commonality of these differing theories centers on zinc, we chose to study ZraR regulatory
348 activity on LB medium supplemented with 1 mM $ZnCl_2$ (57). The ZraR iModulon included only
349 two of the three previously published regulon genes, *zraR* and *zraP*. One of the regulon genes,
350 *zraS*, was not identified by the iModulon. However, ChIP-exo binding peaks were observed
351 upstream of all three regulon genes (*zraSR* and *zraP*), confirming these genes to be under the
352 direct control of ZraR. The ChIP-exo results only indicated one additional direct target gene,
353 *mgtA*, and the remaining 12 binding peaks might be attributed to spurious binding. DEGs were
354 examined to identify indirect targets and expand direct targets (if any) of ZraR. In the presence
355 of 1 mM $ZnCl_2$, 97 genes were found to be differentially expressed between the *zraR* knockout
356 and wild-type strain. Four of these genes (*zraSR*, *zraP* and *mgtA)* had corresponding upstream
357 binding peaks found by ChIP-exo or were included in the ZraR iModulon. The remaining 93

DEGs can be attributed to indirect ZraR regulation. Therefore, ZraSR seems to be involved in cell migration and carbohydrate metabolism (Figure 3) similar to CpxAR, further strengthening the cross-regulation of these two TCSs.

Overall, integrating iModulons with ChIP-exo and DEGs enabled us to expand the global regulatory network contribution of each selected TCS and categorize genes in different modes (direct or indirect) of transcription regulation.

3) ICA identifies potential novel inducers

To identify potential novel inducers and inter-connection between TCS pathways, we compared the activity levels of each TCS iModulon across all PRECISE experiments. To this end, the iModulon activities of each TCS were analyzed across their expected induction condition(s) in our experiments, as well as the previous conditions of PRECISE experiments. By calculating each TCS iModulon activity for the experiments under control conditions (e.g. LB medium for BaeSR, CpxAR, and ZraSR), a baseline was established for expected TCS behavior in the absence of an induction condition. From there, iModulon activities were calculated for all other experimental conditions across PRECISE and the other TCS experiments of this study, and then these activities were evaluated against the control condition iModulon activity. While many experimental conditions yielded iModulon activities indiscernible from those of the control, several key environmental inducers were shown to have significant differential activity for each TCS (Figure 4). These results are important for understanding which types of environments are most likely to trigger TCS action. The ability to succinctly examine the activity levels of an

iModulon over an ever-growing compendium of conditions enables the broad characterization of the iModulon's response (or lack of response) to each of the conditions. Supplemental Figure S2 shows the activity profiles for each of the five TCS response regulators (RR) across all conditions, including PRECISE conditions and the new conditions tested in this study.

When observing the activity levels of ethanol sensors (BaeR and CpxR), we noted that a high average activity level occurred when the wild-type strain was cultured on LB + 5% ethanol. Interestingly, both the BaeR and CpxR iModulons exhibited increased activity for a previous set of experiments that studied the effect of osmotic stress on *E. coli* K-12 MG1655 in relation to the OmpR regulon (62). Although Bury-Moné 2009 (18) had suggested that 0.6 M NaCl did not activate a CpxR and BaeR target gene reporter, our results suggest that osmotic stress (via the addition of 0.3 M NaCl) may induce BaeR and CpxR to some extent. A prior study attempted to identify cross-regulation between CpxAR and OmpR-EnvZ systems (63). In addition to osmotic stress induction, previous PRECISE iron starvation conditions (64) also yielded a significant increase in BaeR iModulon activity, which indicates that BaeR may also be involved in response to iron deficiency (Figure 4). For CpxR, iModulon activity peaks also occurred across experiments testing for phosphate starvation. The array of conditions inducing the BaeR and CpxR iModulons is reasonable considering the response regulators' extensive functionalities and roles in responding to cell envelope stress (Figure 4).

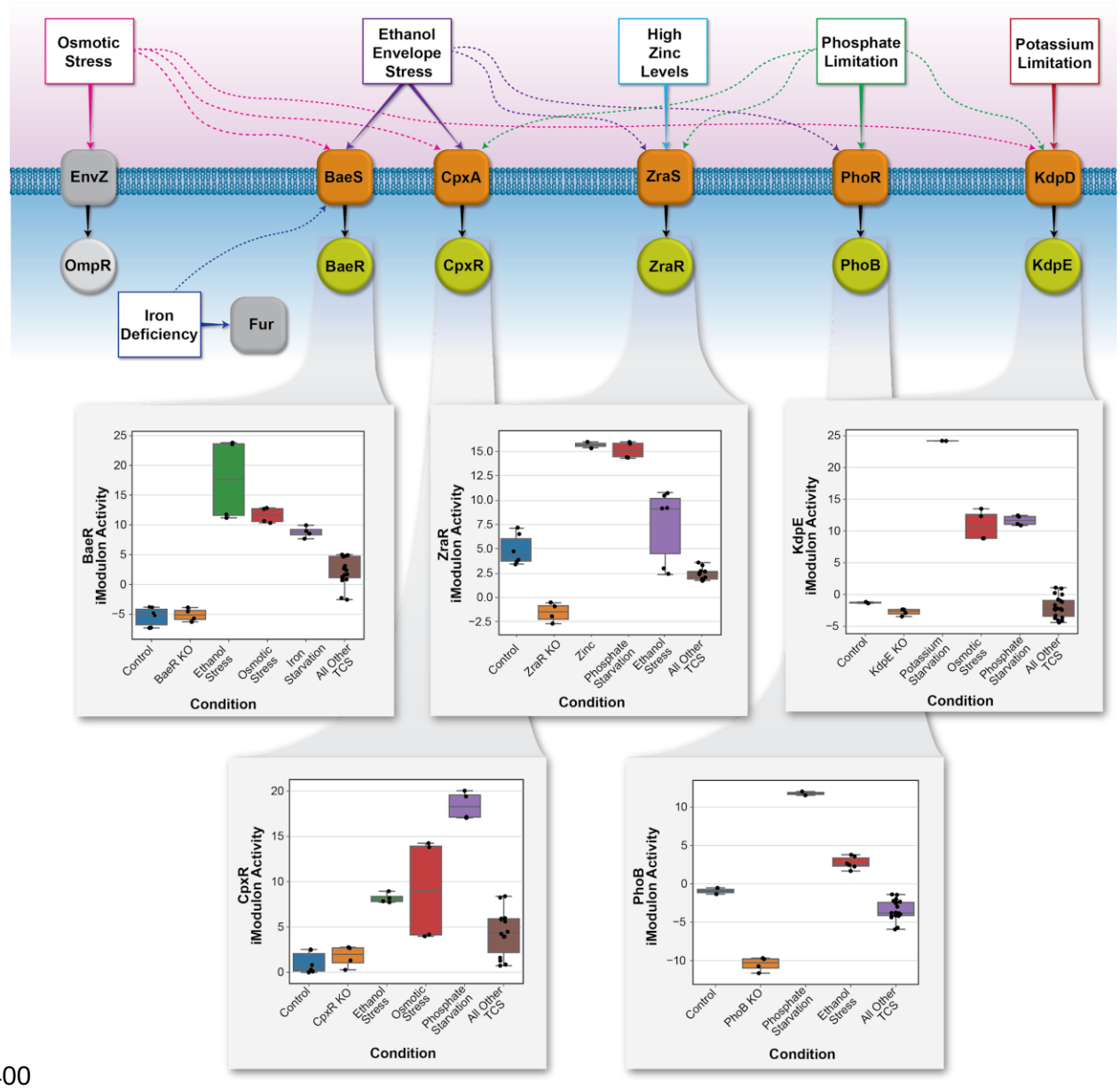


Figure 4: Cross-regulation among TCS and iModulon activities across selected conditions

KdpE was shown to be directly phosphorylated by PhoR under simultaneous potassium and phosphate limitation (45). Consistent with this, we observed high peak activity of the KdpE iModulon for our TCS experiments conducted on phosphate-depleted M9 medium, in addition to the expected induction of the wild-type strain on TMA + 0.1 mM KCl (Figure 4). Another

notable iModulon activity increase was observed for a subset of PRECISE experiments that were conducted under osmotic stress (62), which is consistent with KdpE's documented induction under salt stress (42, 43). Compared to KdpE, the PhoB iModulon was active in several of our other TCS experimental conditions, with an especially notable trend of activity increases for all our samples on LB + 5% ethanol stress. Although PhoB is not generally thought to be linked to the envelope stress response, these iModulon activity increases could potentially be linked to previous research that has found ethanol to induce calcium phosphate crystallization (65, 66) that could limit the amount of freely available phosphate and mirror starvation conditions. For the PRECISE experiments, most iModulon activities were negative or only minimally positive, which is consistent with the absence of phosphate starvation conditions in any of the previous experiments (Figure 4).

Notably, the ZraR iModulon includes only the two expected regulon genes, *zraR* and *zraP*, yet this iModulon exhibits significantly heightened activity over a surprisingly wide range of experimental conditions apart from LB + 1 mM $ZnCl_2$ (Supplemental Figure S2). For this study's experiments, the ZraR iModulon activity was high for all experiments conducted in phosphate-depleted M9 medium and most experiments conducted in LB + 5% ethanol. The high activity of the iModulon across diverse conditions supports the view of ZraR as a broader regulator past simple zinc resistance; specifically, ethanol's activation of the iModulon could be indicative of ZraR's relation to ESR, as previously suggested (61) (Figure 4).

Overall, combining our new dataset with PRECISE and applying ICA enabled us to identify potential cross-regulation between TCS pathways, which allowed the identification of

novel or non-cognate TCS-inducer pairs. Both the known induction conditions validated by this study and the novel induction conditions illuminated by this study are summarized in Table 3 below.

Response Regulator	ICA-identified Induction Conditions also found in Previous Studies	Novel ICA-identified Induction Conditions
BaeR	Ethanol Stress; Osmotic Stress	Iron Starvation
CpxR	Ethanol Stress; Osmotic Stress	Phosphate Starvation
KdpE	Potassium Starvation; Osmotic Stress	Phosphate Starvation
PhoB	Phosphate Starvation	Ethanol Stress
ZraR	Zinc; Ethanol Stress	Phosphate Starvation

Table 3: Response regulator induction across various conditions using ICA iModulon activity levels.

4) Validation of Results

Before using the various gene datasets to scrutinize the regulatory network of each TCS, the quality of the newly generated data was first evaluated. Namely, the peaks found with ChIP-exo were shown to contain the published consensus motif of each TCS, and the RNA-seq data was analyzed to verify TCS regulon induction under their respective stimulation conditions.

Verifying ChIP-exo Binding Peaks

To validate the binding peaks found by the ChIP-exo analysis, the peaks associated with direct targets from this study were compared to known binding peak motifs using AME (Analysis of Motif Enrichment) from the MEME Suite (67). The experimentally identified peaks that were found to match the consensus motif as represented by RegulonDB (68), are summarized below in Table 4.

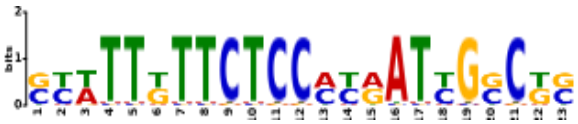
TCS	Peak ID of Match	Operons Associated with Peak	PWM Score	Consensus Motif
BaeR	peak_5	*mdtABCD-baeSR*	3550730	
	peak_2	*spy*	32	
CpxR	peak_10	*cpxPQ / cpxRA*	12053	
KdpE	peak_2	*kdpFABC*	34971900	
PhoB	peak_7	*pstSCAB- phoU*	3910090	
	peak_0	*phoBR*	6496	
ZraR	peak_9	*zraSR*	2.42e+17	

Table 4 : ChIP-exo peaks for direct target genes that were identified as a match to the consensus motif by AME

All five of the examined TCSs were shown to have at least one direct target ChIP-exo peak that matches the consensus motif using AME. In accordance with expectations, the matching peaks pinpointed by AME included canonical primary targets of each TCS response regulator. A visualization of the peak genomic location and sequence match to the consensus motif is shown below in Figure 5 for the *pstSCAB- phoU* peak (id: peak_7) that was identified as a match to the PhoB consensus motif.

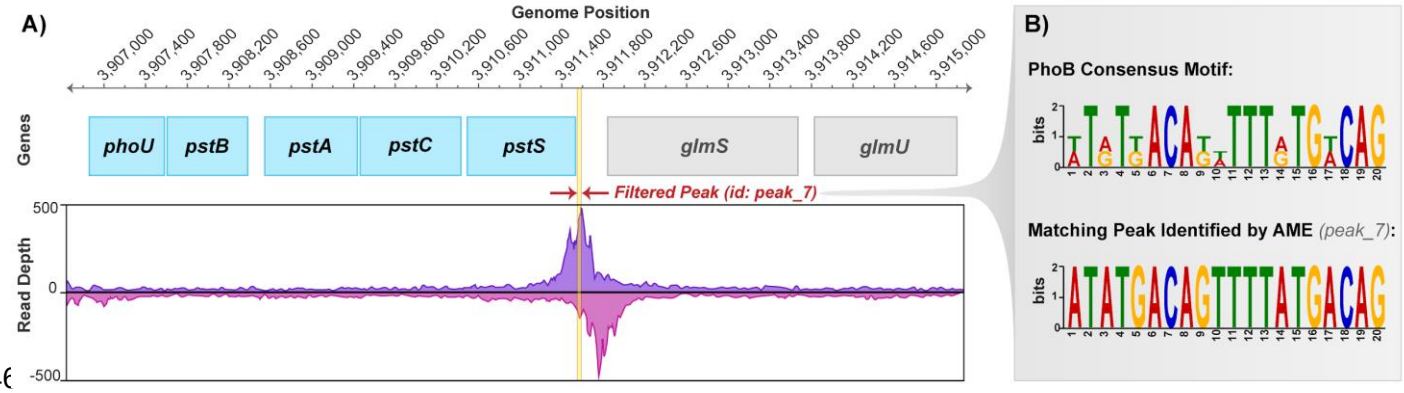


Figure 5: A) Example of ChIP-exo peak identification with PhoB peak_7 (*pstSCAB-phoU* operon). B) Comparison of PhoB binding consensus motif to peak_7, which was identified as a match by AME.

Examining Induction with RNA-seq Differential Gene Expression

The results of this study rely upon the adequate stimulation of the various TCSs under their various induction conditions. As a verification of these induction conditions, a closer look is given to differential gene expression in both the unstimulated and stimulated conditions. For the histidine kinase and response regulator genes of each TCS, the response regulator knockout versus wild-type RNA-seq gene expression is compared under both the unstimulated and

stimulated conditions to confirm successful induction (Table 5). A broader comparison of the response regulator knockout versus wild-type expression for the entire direct regulon of each TCS is available in Supplemental Dataset S8.

TCS	TCS Gene	Unstimulated Condition: Knockout versus Wild-Type		Stimulated Condition: Knockout versus Wild-Type	
		Log_2 Fold Change	p value	Log_2 Fold Change	p value
BaeR	*baeS*	-0.41	2.96E-01	-4.62	4.22E-15
	baeR	-8.83	6.47E-09	-9.43	4.09E-10
CpxR	*cpxA*	-2.20	5.29E-94	-3.93	4.95E-77
	cpxR	-12.55	3.53E-17	-11.35	2.65E-14
KdpE	*kdpD*	-0.22	1.28E-01	-6.78	≅ 0.00E+00
	kdpE	-8.09	1.54E-06	-13.86	1.51E-20
PhoB	*phoR*	-1.52	1.42E-12	-3.16	1.41E-151
	phoB	-9.95	1.45E-10	-15.61	8.47E-26
ZraR	*zraS*	-0.02	9.45E-01	-3.18	4.00E-139
	zraR	-10.67	5.38E-13	-13.33	5.44E-19

Table 5: Summary of differential expression of RR knockout versus wild-type strains under unstimulated and stimulated conditions for each TCS gene

Both the histidine kinase and response regulator genes were shown to be differentially expressed at their respective stimulation condition with a log fold change greater than 1.5 and a p value less than 0.05 (Table 5). This observation confirms that each of the TCS were more expressed in the wild-type strain versus the RR KO strain under induction. For the respective unstimulated conditions, some of the TCS genes were also differentially expressed, including all the response regulator genes and the *cpxA* and *phoR* histidine kinase genes. This result shows that all the response regulator genes and two of the histidine kinase genes were being expressed

in the wild type even without induction. However, all of the histidine kinase genes and all of the response regulator genes other than *cpxR* exhibited a greater magnitude of log fold change in the stimulated condition versus the unstimulated condition. Overall, this seems to indicate successful induction of the histidine kinase elements to the respective stimuli.

Discussion

E. coli uses two-component systems (TCSs) to respond to environmental signals by modulating gene expression. Each TCS employs a signal transduction mechanism to modulate target gene expression when a certain stimulus is received. These signal transduction systems form a part of the global transcriptional regulatory network (TRN). In this study, we identified the TRN contributions of five TCSs in *E. coli*: BaeSR and CpxAR, which were stimulated by ethanol stress, and KdpDE, PhoRB, and ZraSR, which were stimulated by potassium limitation, phosphate limitation, and added zinc, respectively. The recent identification of iModulons in *E. coli's* transcriptome (7) enables new capabilities to further explore its response to environmental stimuli. Here, this ICA-driven iModulon determination process was implemented in combination with ChIP-exo data analysis to study the TRN contribution of five TCSs. This technique delivered three key results: 1) pinpointing and validation of the target genes for each TCS; 2) determination of modes of transcriptional regulation of the target genes for each TCS; and 3) the elucidation of TCS activity across varying conditions and novel relationships among TCSs.

The integrative approach of combining ICA, Chip-exo, and RNA-seq datasets provided us with the strong evidence to expand the global regulatory network of each of the five TCSs (BaeSR, CpxAR, KdpDE, PhoRB, and ZraSR) and decipher different modes of transcriptional

509 regulation. With this study, we extend the scope of TRNs beyond direct regulations by defining
510 indirect targets, cross-regulatory networks, and hypothetically functional binding genes. Apart
511 from the classically studied direct targets, genes in cross-regulatory networks and hypothetically
512 functional binding genes fall into new categories that are of particular interest. For example,
513 *ompC* was characterized as a hypothetically functional binding gene target of CpxR, since it was
514 previously reported as part of the CpxR regulon in response to high acetyl phosphate (36) and we
515 identified ChIP-exo binding peaks upstream of it. However, we did not observe any differential
516 expression of this gene or iModulon activity. The lack of response in *ompC* expression despite
517 the potential binding of CpxR suggests that some other regulator may be repressing the action of
518 CpxR on *ompC*. Since *ompC* already belongs to the published CpxR regulon, it is possible that
519 the specific conditions of this study (i.e. ethanol induction on LB media) triggered a repression
520 response that is not always present. Therefore, we classified this gene under the "hypothetically
521 functional binding gene" category, suggesting a possible regulation of this gene via CpxR despite
522 the lack of substantiation with RNA-seq expression results. Similarly, two genes (*sbcC* and
523 *sbcD*) from the PhoRB network were characterized as hypothetically functional binding genes,
524 which could also indicate that these genes are potential binding targets of PhoB.
525
526 The "cross-regulatory network" category included genes which were identified in the
527 iModulon and/or published regulon but could not be confirmed using ChIP-exo methods and/or
528 DEGs. This category of genes are of interest because they show the interwoven and signal-
529 specific nature of the TRN. It could be postulated that cases where the genes were just identified
530 by iModulons were due to a more stringent threshold for calculating DEGs. However, it may be
531 more accurate to consider the unique, albeit complementary, perspectives offered by iModulons

and DEGs on the effects of gene knockouts. Specifically, iModulons identify genes that are co-expressed, and likely co-regulated, across over 300 expression conditions, whereas DEGs offer a targeted analysis of the effect of a single regulator knock-out. Due to the larger dataset feeding ICA, patterns can be deciphered that may appear insignificant by DEG analysis alone. For example, we observed five genes in the cross-regulatory network of the CpxR regulon (*alx*, *mdtJ*, *yjfN*, *yliI*, and *yobB*) that were identified by iModulons but not by DEGs because of low p-value significance (> 0.05 threshold). This example highlights the capability of ICA in deciphering the global TRN.

In addition to introducing new ways to categorize RNA-seq and ChIP-exo results, we show that applying ICA to transcriptomics data provides a thorough understanding of TCS TRN contributions. Even further, we are able to propose novel relationships between different regulatory elements. For example, we identified that apart from the previously known ethanol stress as an inducer of BaeR, osmotic stress and iron-deficiency may also modulate BaeR target gene binding; however, additional experimentation would be necessary to verify this relationship. Other novel relationships were indicated by the increase in CpxR and ZraR iModulon activities under phosphate starvation conditions, which suggests that phosphate limiting conditions may also act as an inducer for CpxAR and ZraSR two-component systems. Conversely, we also identified a potential involvement of PhoRB and ZraSR in the envelope stress response. One possible explanation of the link between phosphate starvation and envelope stress response could be the role of each of the stimuli in biofilm formation. Specifically, it has been suggested that the phosphate starvation of *E. coli* results in the modification of lipopolysaccharides to increase biofilm formation (69), and CpxAR has been proposed to be

555 induced by biofilm formation due to its sensing of cell adhesions (34). Therefore, the phosphate
556 starvation in this study may have stimulated biofilm formation, which may have in turn induced
557 CpxAR and broader ESR targets. We were also able to confirm previously documented
558 relationships between KdpDE and PhoRB systems (45) wherein the activity of the KdpE
559 iModulon was found to be high under phosphate limited conditions, in addition to potassium and
560 osmotic stress. Taken together, this study identified several new potential relationships and
561 provided evidence towards suspected relationships by focusing on only five out of the 29
562 complete TCSs. As the scope increases past these five TCSs, there is potential for identifying
563 additional novel relationships to enhance our knowledge of TRN interconnectedness.
564
565 Finally, we validated the genesets falling in each category by comparing the known
566 consensus motifs using AME. We observed that at least one direct target ChIP-exo peak for each
567 TCS matched the consensus motif using AME. Interestingly, two genes (*sspA* and *sspB*) that
568 were categorized as potentially spurious binding peaks for BaeR were also found to match the
569 consensus BaeR motif. This result could indicate that the *sspAB* operon should be reclassified as
570 a novel binding target of BaeR. The gene *sspA* has been shown to be essential for survival during
571 acid-induced stress (70), so a relationship to the envelope stress-mediating BaeSR TCS does not
572 seem improbable. However, further experimentation would be necessary to validate *sspAB* as a
573 potential binding target of BaeR, especially with the lack of RNA-seq substantiation in this
574 study.
575
576 In summary, we suggest that integrating ChIP-exo and DEG analyses with ICA can
577 provide a deeper understanding of the transcriptional regulatory network of *E. coli*. In particular,

ICA provides a broad view of TCS activity over diverse conditions, enables the detection of non-cognate inducers to response regulators (this study), and can identify redundancy in TF actions (HprSR and CusSR) (our previous study (7)) by showing that different external signals cause expression of the same set of genes. This robust validation of the five TCSs' contributions to the *E. coli* TRN expands the current knowledge base by identifying new potential direct and indirect targets of each TCS RR and by confirming previously predicted or inferred TCS target genes that were found in earlier studies. Looking forward, the horizon of bacterial 'awareness' can be examined by incorporating these results into a genome-scale reconstruction of two-component systems which can reliably predict metabolic capabilities under different environmental conditions.

Materials and Methods

Bacterial strains and growth conditions

Strains used in this study are *E. coli* K-12 MG1655 and its derivatives (deletion strains and myc-tagged strains). As previously described (71), strains retaining 8-myc were generated by a λ red-mediated site-specific recombination system targeting the C-terminal region of each selected response regulator. Knock-out mutants (*ΔbaeR, ΔcpxR, ΔkdpE, ΔphoB* and *ΔzraR*) were generated according to the procedure from Datsenko and Wanner (72), using pKD13 and pKD46 as suggested by the authors (please refer to Supplemental Table S1 for full oligonucleotide list). All knockouts are complete deletions, except for KdpE; the KdpE locus could potentially produce an 18 bp peptide, since the sequence after the stop codon contains too many A/T-rich repeats for proper oligonucleotide annealing hence the binding site had to be

shifted into the coding region (please see supplemental table for oligonucleotide list). If not stated otherwise, cells were grown in pre-cultures prepared from glycerol stocks, overnight and then transferred to fresh medium the next morning. For ethanol sensors (BaeR and CpxR): Cells were grown at 37 °C in liquid LB medium to an OD600=0.5, then ethanol was added to a final concentration of 5% w/v. To ensure proper binding of RRs to their target genes, cells were grown for 30 min in the presence of ethanol before being collected for ChIP-exo and RNA-seq. This growth condition was chosen according to Bury-Moné 2009 (see Table 1 in Bury-Moné (18)), who showed, that 5% ethanol led to the strongest induction of β-galactosidase reporters for CpxR and BaeR among all tested conditions (i.e. 2 mM and 4 mM Indole, 3% and 5% ethanol, 0.5 mM Dibucaine, 5 mM EDTA, and 0.6 M NaCl). For RNA seq controls, cells were grown in liquid LB medium and collected at an OD600=0.5. The KdpDE TCS induces expression of the high affinity K^+ transporter KdpFABC under K-limiting conditions (Schramke et al., 2016). According to Schramke et al., KdpD acts as a phosphatase on KdpE-P and prevents production of the high-affinity K^+ transporter at high extracellular K+ concentration (>5 mM). When environmental levels of K^+ fall below the threshold for autokinase activation, *kdpFABC* expression is initiated; however, as long as the intracellular K^+ concentration remains high, the KdpD phosphatase activity remains stimulated (Schramke et al., 2016). Therefore, to induce KdpDE, cells were grown in K-sufficient conditions, washed and then grown under K-limiting conditions to ensure drops in intracellular K^+ levels below the threshold level and thus proper induction of the TCS: Cells were grown at 37 °C overnight in liquid Tris maleic acid minimal medium (TMA) (44, 45) supplemented with 115 mM KCl and 0.4 % w/v Glucose, then washed twice with TMA containing 0.1 mM KCl and 0.4 % w/v Glucose. Cells were inoculated in TMA with 0.1 mM KCl and 0.4 % w/v Glucose and collected at an OD600=0.5 for ChIP-exo and

RNA-seq analysis. For RNA seq controls, cells were grown in liquid TMA supplemented with 115 mM KCl and 0.4 % w/v Glucose, and collected at an OD600=0.5. For induction of the PhoRB TCS we chose to use M9 minimal medium without any phosphate, to ensure proper induction of the system. To our knowledge, this condition was not used before. To induce phosphate limiting conditions, cells were grown in liquid M9 minimal medium (containing phosphate) until OD600=0.5. Growing cells until the mid exponential growth phase was necessary, because the lack of phosphate led to a growth arrest. Then cells were washed three times with M9 minimal medium without phosphate (M9-P; without Na_2HPO_4 and KH_2PO_4) and incubated in M9-P for 60 min (about the doubling time of MG1655 in M9 minimal medium) at 37 °C. Cells were then collected for ChIP-exo and RNA-seq analysis. For RNA seq controls, *ΔphoB* cells were grown in liquid M9 minimal medium and collected at an OD600=0.5 (WT samples were taken from (7), see Table 1). For ZraSR: Cells were grown at 37 °C in liquid LB medium containing 1 mM $ZnCl_2$, according to Leonhartsberger et al., 2001 (57) but without the addition of glucose to the LB medium. At an OD600=0.5 cells were collected for ChIP-exo and RNA-seq analysis. For RNA seq controls, cells were grown in liquid LB medium and collected at an OD600=0.5. Samples for ChIP-exo and RNA-seq were taken independently.

RNA-seq

For RNA-seq, 3 mL of culture were mixed with 6 mL of RNAprotect bacteria reagent (Qiagen) and processed according to the manufacturer's instructions. Cell pellets were frozen and stored at -80 °C until processed. RNA was extracted using the Zymo Research Quick RNA fungal/bacterial microprep kit, according to the manufacturer's instructions. Ribosomal RNA was removed from total RNA preparations using RNase H. First, traces of genomic DNA were removed with a DNase I treatment. Then, secondary structures in the ribosomal RNA were

removed by heating to 90 °C for one second. A set of 32-mer DNA oligonucleotide probes complementary to 5S, 16S, and 23S rRNA subunits and spaced nine bases apart were then annealed at 65 °C followed by digestion with Hybridase (Lucigen), a thermostable RNase H. Hybridase was added at 65 °C, the reaction was incubated for 20 minutes at that temperature, then heated again to 90 °C for one second to remove remaining secondary structures, and finally returned to 65 °C for 10 minutes. The reaction was quickly quenched by the addition of guanidine thiocyanate while still at 65 °C before purifying the mRNA with a Zymo Research RNA Clean and Concentrator kit using their 200 nt cutoff protocol. Carryover oligos were removed with a DNase I digestion which was started at room temperature and gradually increased to 42 °C over a half hour. This was followed up with another column purification as stated above.

Paired-end library preparation was done using the KAPA RNA HyperPrep kit following the manufacturer's instructions with an average insert size of 300 bp. The libraries were then analyzed on an Agilent Bioanalyzer DNA 1000 chip (Agilent). RNA-seq libraries were sequenced on the NextSeq 500 (Illumina) at the Salk Institute for Biological Studies, La Jolla, CA, USA. RNA-seq experiments were performed in duplicates.

ChIP-exo experiments

To identify binding sites for each response regulator under their respective induction conditions (see above and Table 1), ChIP-exo experiments were performed as described previously (64). Specifically, cells were crosslinked in 1% Formaldehyde and DNA bound to each response regulator was isolated by chromatin immunoprecipitation (ChIP) with the specific antibodies that recognize myc tag (9E10, Santa Cruz Biotechnology), and Dynabeads Pan Mouse

IgG magnetic beads (Invitrogen). This was followed by stringent washing as described earlier (73). To perform on-bead enzymatic reactions of the ChIP-exo experiments, ChIP materials (chromatin-beads) were used (5). The method has been described in detail previously (64, 74). ChIP-exo experiments were performed in duplicates. ChIP-exo libraries were sequenced on the NextSeq 500 (Illumina) at the Sanford Burnham Prebys Medical Discovery Institute, La Jolla, CA, USA, or on the NextSeq 500 (Illumina) at the Salk Institute for Biological Studies.

ChIP-exo processing and peak calling

The ChIP-exo sequence reads were mapped onto the reference genome (*E. coli* MG1655 NC_000913.3) using bowtie (75) with default options as also described previously (64, 74). This generated SAM output files. Peak calling from biological duplicates for each experimental condition was done using the MACE program (76). Two level filtering of ChIP-exo data was done: first, we removed any peaks with S/N ratio of less than 1.5 to reduce false positive peaks. Top 5% of the signals at genomic positions was set as noise level as previously described (64, 74). MetaScope (https://sites.google.com/view/systemskimlab/software) was used to visualise the peaks. To further reduce the false positive peaks, a cut-off analysis was done for each regulator on the signal-to-noise (S/N) ratio (Supplemental Figure S1). We removed any peaks that had S/N ratio below the threshold. Thereafter, the peak was assigned to the nearest operon on both strands. Operons located within 500 bp of the peak were considered. Specifically, if a peak was upstream of an operon, we considered all the genes in that operon to be regulated by that response regulator.

Analysis of Motif Enrichment (AME) for Consensus in ChIP-exo Peaks

After the final S/N thresholding and 500 bp cutoffs were applied to filter the ChIP-exo data, the resulting binding peaks were buffered with a margin of 20 base pairs on either end of their sequence then compared to the published consensus motif to identify matching sequences (Table 4). The search was performed by AME from the MEME suite (67) using a one-tailed Fisher's exact test method to evaluate motif enrichment. The fraction of maximum log-odds for scoring a potential match was specified as 0.25, and the E-value for reporting was arbitrarily set at 10 (no TCS had more than 10 identified sequences). A 0-order Markov sequence model was created by AME to normalize for biased distributions; additionally, AME created a control sequence by shuffling the letters in each of the input sequences while preserving 2-mers.

Differential Gene Expression

RNA-seq sequence reads were mapped onto the reference genome (*E. coli* MG1655 NC_000913.3) using bowtie 1.1.2 (75) with the following options "–X 1000 –n 2 –3 3", where X is maximum insert size, n is number of mismatches, and –3 3 denotes trimming of three base pairs at 3'–end. Read count was performed using *summarizeOverlaps* from the R *GenomicAlignments* package, using options "mode="IntersectionStrict", singleEnd=FALSE, ignore.strand=FALSE, preprocess.reads=invertStrand" (77). Differential gene expression was identified by *DESeq2* (78). Genes with $\log_2$ fold change > 1.5 and a false discovery rate (FDR) value < 0.05 were considered as differentially expressed genes. Genes with p values assigned 'NA' based on extreme count outlier detection were not considered as potential DEGs.

Independent component analysis (ICA)

To identify independent sources of gene expression control, ICA was applied to the combined dataset that consisted of all 278 RNAseq gene expression data from PRECISE (7) and 32 datasets from this study which included gene expression data of knockout mutants and wild types under stimulated and unstimulated conditions. Genes with less than 10 fragments mapped per million reads across the entire dataset were removed from the RNA-seq datasets to reduce noise. FastICA was performed 100 times with random seeds and a convergence tolerance of 10^{-6}. For each iteration, we constrained the number of components to the number that generates 99% variance in principal component analysis (PCA). We further used DBSCAN to cluster the resulting components to identify a set of robust independent components using a minimum cluster seed size of 50. We repeated this process five times and only the consistent components that occurred in each run were taken. iModulons were extracted as previously described (7).

iModulon Characterization

We checked for the significant genes in each iModulon and mapped it to the set of genes regulated by a specific regulator or transcription factor. This comparison was done using one-sided Fisher's Exact Test (FDR < 10^{-5}). The regulon/regulator with lowest p-value was given the name of the iModulon. The ones which were not characterized by this method were further analyzed through Gene Ontology (GO) enrichment. In this case, significant genes in each iModulon were compared to genes in each GO term and a p-value was assigned using the one-sided Fisher's Exact Test (FDR < 10^{-5}). The GO term having lowest p-value was assigned as the name of the iModulon.

Data Availability

The ChIP-exo and RNA-seq datasets have been deposited to GEO with the accession number of GSE143856. All other data is available in the supplementary data.

Figure Titles and legends

Figure 1: A) Workflow used for each gene expression dataset. B) Categories of regulation by two-component systems in *E. coli*.

Figure 2: Comparison of regulatory network in BaeR and CpxR and functional characterization of direct and indirect targets.

Figure 3: Comparison of regulatory network in metal sensors and functional characterization of direct and indirect targets.

Figure 4: Cross-regulation among TCS and iModulon activities across selected conditions.

Figure 5: A) Example of ChIP-exo peak identification with PhoB peak_7 (*pstSCAB-phoU* operon). B) Comparison of PhoB binding consensus motif to peak_7, which was identified as a match by AME.

Tables

Table 1: List of strains used in this study

754 **Table 2:** List of genes in each iModulon and published regulon, and respective direct targets.
755 Genes in bold text in "iModulon Genes" and "Targets from RegulonDB" columns indicate
756 common genes.

757 **Table 3:** Response regulator induction across various conditions using ICA iModulon activity
758 levels.

759 **Table 4 :** ChIP-exo peaks for direct target genes that were identified as a match to the consensus
760 motif by AME

761 **Table 5:** Summary of differential expression of RR knockout versus wild-type strains under
762 unstimulated and stimulated conditions for each TCS gene

763

764 Supplementary materials file list

765 Figures

766 **Supplemental Figure S1:** Signal-to-noise (S/N) ratio for ChIP-exo data of BaeR, CpxR, KdpE,
767 PhoB, and ZraR.

768 **Supplemental Figure S2:** ICA Activity profiles for BaeR, CpxR, KdpE, PhoB, and ZraR
769 iModulons.

770 Tables

771 **Supplemental Table S1:** Oligonucleotides used in this study to add a 8-myc tag or to knock out
772 genes.

Datasets

Supplemental Dataset S1: Genes and their weighting for each of the 15 identified TCS RR iModulons.

Supplemental Dataset S2: ChIP-exo binding peaks for BaeR, CpxR, KdpE, PhoB, and ZraR.

Supplemental Dataset S3: Differentially Expressed Genes (DEG) between wild-type and knockout strains for BaeR, CpxR, KdpE, PhoB, and ZraR.

Supplemental Dataset S4: Indirect target genes of BaeR, CpxR, KdpE, PhoB, and ZraR.

Supplemental Dataset S5: Potentially spurious ChIP-exo binding genes of BaeR, CpxR, KdpE, PhoB, and ZraR.

Supplemental Dataset S6: Hypothetically functional binding genes of BaeR, CpxR, KdpE, PhoB, and ZraR.

Supplemental Dataset S7: Cross-regulated genes of BaeR, CpxR, KdpE, PhoB, and ZraR.

Supplemental Dataset S8: Differential expression of TCS RR regulons under unstimulated and stimulated conditions.

Acknowledgements

We thank Irina Rodionova for helpful discussion. This work was funded by the Novo Nordisk Foundation Grant Number NNF10CC1016517. We thank Marc Abrams for reviewing the manuscript and providing constructive suggestions.

Author contribution

KSC and BOP designed the study. JAK designed, generated the strains and performed the experiments. YG, RS assisted in generating RNA-seq and ChIP-exo data. KSC and KD processed the RNA-seq and ChIP-exo data and performed computational analysis. AVS and YS contributed in data analysis. KSC, KD, AVS and BOP wrote the manuscript, with contributions from all other authors.

Competing financial interests

The authors declare no financial competing interests.

REFERENCES:

1. Zschiedrich CP, Keidel V, Szurmant H. 2016. Molecular Mechanisms of Two-Component Signal Transduction. J Mol Biol 428:3752–3775.

2. Mizuno T. 1997. Compilation of all genes encoding two-component phosphotransfer signal transducers in the genome of Escherichia coli. DNA Res 4:161–168.

811 3. Macritchie DM, Raivio TL. 2009. Envelope Stress Responses. EcoSal Plus 3.

812 4. Crépin S, Chekabab S-M, Le Bihan G, Bertrand N, Dozois CM, Harel J. 2011. The Pho
813 regulon and the pathogenesis of Escherichia coli. Vet Microbiol 153:82–88.

814 5. Rhee HS, Pugh BF. 2012. ChIP-exo Method for Identifying Genomic Location of DNA-
815 Binding Proteins with Near-Single-Nucleotide Accuracy. Curr Protoc Mol Biol 100:135.

816 6. Spivakov M. 2014. Spurious transcription factor binding: non-functional or genetically
817 redundant? Bioessays 36:798–806.

818 7. Sastry AV, Gao Y, Szubin R, Hefner Y, Xu S, Kim D, Choudhary KS, Yang L, King ZA,
819 Palsson BO. 2019. The Escherichia coli transcriptome mostly consists of independently
820 regulated modules. Nat Commun 10:5536.

821 8. Urano H, Umezawa Y, Yamamoto K, Ishihama A, Ogasawara H. 2015. Cooperative
822 regulation of the common target genes between H_2O_2-sensing YedVW and Cu^{2+}-sensing
823 CusSR in Escherichia coli. Microbiology 161:729–738.

824 9. Santos-Zavaleta A, Sánchez-Pérez M, Salgado H, Velázquez-Ramírez DA, Gama-Castro S,
825 Tierrafría VH, Busby SJW, Aquino P, Fang X, Palsson BO, Galagan JE, Collado-Vides J.
826 2018. A unified resource for transcriptional regulation in Escherichia coli K-12
827 incorporating high-throughput-generated binding data into RegulonDB version 10.0. BMC
828 Biology.

829 10. Raffa RG, Raivio TL. 2002. A third envelope stress signal transduction pathway in
830 Escherichia coli. Mol Microbiol 45:1599–1611.

831 11. Yamamoto K, Ogasawara H, Ishihama A. 2008. Involvement of multiple transcription
832 factors for metal-induced spy gene expression in Escherichia coli. J Biotechnol 133:196–
833 200.

834 12. Hirakawa H, Inazumi Y, Masaki T, Hirata T, Yamaguchi A. 2005. Indole induces the
835 expression of multidrug exporter genes in Escherichia coli. Mol Microbiol 55:1113–1126.

836 13. Srivastava SK, Lambadi PR, Ghosh T, Pathania R, Navani NK. 2014. Genetic regulation of
837 spy gene expression in Escherichia coli in the presence of protein unfolding agent ethanol.
838 Gene 548:142–148.

839 14. Dorel C, Lejeune P, Rodrigue A. 2006. The Cpx system of Escherichia coli, a strategic
840 signaling pathway for confronting adverse conditions and for settling biofilm communities?
841 Res Microbiol 157:306–314.

842 15. De Wulf P, McGuire AM, Liu X, Lin ECC. 2002. Genome-wide profiling of promoter
843 recognition by the two-component response regulator CpxR-P in Escherichia coli. J Biol
844 Chem 277:26652–26661.

845 16. Zahrl D, Wagner M, Bischof K, Koraimann G. 2006. Expression and assembly of a
846 functional type IV secretion system elicit extracytoplasmic and cytoplasmic stress responses
847 in Escherichia coli. J Bacteriol 188:6611–6621.

848 17. Fleischer R, Heermann R, Jung K, Hunke S. 2007. Purification, reconstitution, and
849 characterization of the CpxRAP envelope stress system of Escherichia coli. J Biol Chem
850 282:8583–8593.

851 18. Bury-Moné S, Nomane Y, Reymond N, Barbet R, Jacquet E, Imbeaud S, Jacq A, Bouloc P.
852 2009. Global analysis of extracytoplasmic stress signaling in Escherichia coli. PLoS Genet
853 5:e1000651.

854 19. Baranova N, Nikaido H. 2002. The baeSR two-component regulatory system activates
855 transcription of the yegMNOB (mdtABCD) transporter gene cluster in Escherichia coli and
856 increases its resistance to novobiocin and deoxycholate. J Bacteriol 184:4168–4176.

857 20. Hirakawa H, Nishino K, Hirata T, Yamaguchi A. 2003. Comprehensive studies of drug
858 resistance mediated by overexpression of response regulators of two-component signal
859 transduction systems in Escherichia coli. J Bacteriol 185:1851–1856.

860 21. Hirakawa H, Nishino K, Yamada J, Hirata T, Yamaguchi A. 2003. Beta-lactam resistance
861 modulated by the overexpression of response regulators of two-component signal
862 transduction systems in Escherichia coli. J Antimicrob Chemother 52:576–582.

863 22. Nishino K, Honda T, Yamaguchi A. 2005. Genome-wide analyses of Escherichia coli gene
864 expression responsive to the BaeSR two-component regulatory system. J Bacteriol
865 187:1763–1772.

866 23. Leblanc SKD, Oates CW, Raivio TL. 2011. Characterization of the induction and cellular
867 role of the BaeSR two-component envelope stress response of Escherichia coli. J Bacteriol
868 193:3367–3375.

869 24. Pogliano J, Lynch AS, Belin D, Lin EC, Beckwith J. 1997. Regulation of Escherichia coli
870 cell envelope proteins involved in protein folding and degradation by the Cpx two-
871 component system. Genes Dev 11:1169–1182.

872 25. Danese PN, Snyder WB, Cosma CL, Davis LJ, Silhavy TJ. 1995. The Cpx two-component
873 signal transduction pathway of Escherichia coli regulates transcription of the gene
874 specifying the stress-inducible periplasmic protease, DegP. Genes Dev 9:387–398.

875 26. Dartigalongue C, Raina S. 1998. A new heat-shock gene, ppiD, encodes a peptidyl-prolyl
876 isomerase required for folding of outer membrane proteins in Escherichia coli. EMBO J
877 17:3968–3980.

878 27. Danese PN, Silhavy TJ. 1997. The sigma(E) and the Cpx signal transduction systems
879 control the synthesis of periplasmic protein-folding enzymes in Escherichia coli. Genes Dev
880 11:1183–1193.

881 28. Nevesinjac AZ, Raivio TL. 2005. The Cpx envelope stress response affects expression of
882 the type IV bundle-forming pili of enteropathogenic Escherichia coli. J Bacteriol 187:672–
883 686.

884 29. Hernday AD, Braaten BA, Broitman-Maduro G, Engelberts P, Low DA. 2004. Regulation
885 of the pap epigenetic switch by CpxAR: phosphorylated CpxR inhibits transition to the
886 phase ON state by competition with Lrp. Mol Cell 16:537–547.

887 30. Gubbins MJ, Lau I, Will WR, Manchak JM, Raivio TL, Frost LS. 2002. The positive
888 regulator, TraJ, of the Escherichia coli F plasmid is unstable in a cpxA* background. J
889 Bacteriol 184:5781–5788.

890 31. De Wulf P, Kwon O, Lin EC. 1999. The CpxRA signal transduction system of Escherichia
891 coli: growth-related autoactivation and control of unanticipated target operons. J Bacteriol
892 181:6772–6778.

893 32. Prigent-Combaret C, Brombacher E, Vidal O, Ambert A, Lejeune P, Landini P, Dorel C.
894 2001. Complex Regulatory Network Controls Initial Adhesion and Biofilm Formation in
895 Escherichia coli via Regulation of the csgD Gene. Journal of Bacteriology.

896 33. Jubelin G, Vianney A, Beloin C, Ghigo J-M, Lazzaroni J-C, Lejeune P, Dorel C. 2005.
897 CpxR/OmpR interplay regulates curli gene expression in response to osmolarity in
898 Escherichia coli. J Bacteriol 187:2038–2049.

899 34. Otto K, Silhavy TJ. 2002. Surface sensing and adhesion of Escherichia coli controlled by
900 the Cpx-signaling pathway. Proc Natl Acad Sci U S A 99:2287–2292.

901 35. Weatherspoon-Griffin N, Yang D, Kong W, Hua Z, Shi Y. 2014. The CpxR/CpxA two-
902 component regulatory system up-regulates the multidrug resistance cascade to facilitate
903 Escherichia coli resistance to a model antimicrobial peptide. J Biol Chem 289:32571–
904 32582.

905 36. Batchelor E, Walthers D, Kenney LJ, Goulian M. 2005. The Escherichia coli CpxA-CpxR
906 envelope stress response system regulates expression of the porins ompF and ompC. J
907 Bacteriol 187:5723–5731.

908 37. Yamamoto K, Ishihama A. 2006. Characterization of copper-inducible promoters regulated
909 by CpxA/CpxR in Escherichia coli. Biosci Biotechnol Biochem 70:1688–1695.

910 38. Gupta SD, Lee BT, Camakaris J, Wu HC. 1995. Identification of cutC and cutF (nlpE)
911 genes involved in copper tolerance in Escherichia coli. J Bacteriol 177:4207–4215.

912 39. Ghatak S, King ZA, Sastry A, Palsson BO. 2019. The y-ome defines the 35% of Escherichia

913 coli genes that lack experimental evidence of function. Nucleic Acids Res 47:2446–2454.

914 40. Walderhaug MO, Polarek JW, Voelkner P, Daniel JM, Hesse JE, Altendorf K, Epstein W.
915 1992. KdpD and KdpE, proteins that control expression of the kdpABC operon, are
916 members of the two-component sensor-effector class of regulators. J Bacteriol 174:2152–
917 2159.

918 41. Altendorf K, Voelkner P, Puppe W. 1994. The sensor kinase KdpD and the response
919 regulator KdpE control expression of the kdpFABC operon in Escherichia coli. Res
920 Microbiol 145:374–381.

921 42. Heermann R, Weber A, Mayer B, Ott M, Hauser E, Gabriel G, Pirch T, Jung K. 2009. The
922 universal stress protein UspC scaffolds the KdpD/KdpE signaling cascade of Escherichia
923 coli under salt stress. J Mol Biol 386:134–148.

924 43. Nakashima K, Sugiura A, Momoi H, Mizuno T. 1992. Phosphotransfer signal transduction
925 between two regulatory factors involved in the osmoregulated kdp operon in Escherichia
926 coli. Mol Microbiol 6:1777–1784.

927 44. Weiden PL, Epstein W, Schultz SG. 1967. Cation transport in Escherichia coli. VII.
928 Potassium requirement for phosphate uptake. J Gen Physiol 50:1641–1661.

929 45. Schramke H, Laermann V, Tegetmeyer HE, Brachmann A, Jung K, Altendorf K. 2017.
930 Revisiting regulation of potassium homeostasis in Escherichia coli: the connection to
931 phosphate limitation. Microbiologyopen 6.

932 46. Rhoads DB, Laimins L, Epstein W. 1978. Functional organization of the kdp genes of

933 Escherichia coli K-12. J Bacteriol 135:445–452.

934 47. Gassel M, Möllenkamp T, Puppe W, Altendorf K. 1999. The KdpF subunit is part of the
935 K(+)-translocating Kdp complex of Escherichia coli and is responsible for stabilization of
936 the complex in vitro. J Biol Chem 274:37901–37907.

937 48. Baek JH, Lee SY. 2006. Novel gene members in the Pho regulon of Escherichia coli. FEMS
938 Microbiol Lett 264:104–109.

939 49. Wanner BL. 1993. Gene regulation by phosphate in enteric bacteria. J Cell Biochem 51:47–
940 54.

941 50. VanBogelen RA, Olson ER, Wanner BL, Neidhardt FC. 1996. Global analysis of proteins
942 synthesized during phosphorus restriction in Escherichia coli. J Bacteriol 178:4344–4366.

943 51. Marzan LW, Shimizu K. 2011. Metabolic regulation of Escherichia coli and its phoB and
944 phoR genes knockout mutants under phosphate and nitrogen limitations as well as at acidic
945 condition. Microb Cell Fact 10:39.

946 52. Amemura M, Makino K, Shinagawa H, Kobayashi A, Nakata A. 1985. Nucleotide sequence
947 of the genes involved in phosphate transport and regulation of the phosphate regulon in
948 Escherichia coli. J Mol Biol 184:241–250.

949 53. Surin BP, Rosenberg H, Cox GB. 1985. Phosphate-specific transport system of Escherichia
950 coli: nucleotide sequence and gene-polypeptide relationships. J Bacteriol 161:189–198.

951 54. Luecke H, Quiocho FA. 1990. High specificity of a phosphate transport protein determined
952 by hydrogen bonds. Nature 347:402–406.

953 55. Makino K, Amemura M, Kawamoto T, Kimura S, Shinagawa H, Nakata A, Suzuki M.
954 1996. DNA binding of PhoB and its interaction with RNA polymerase. J Mol Biol 259:15–
955 26.

956 56. Yang C, Huang T-W, Wen S-Y, Chang C-Y, Tsai S-F, Wu W-F, Chang C-H. 2012.
957 Genome-wide PhoB binding and gene expression profiles reveal the hierarchical gene
958 regulatory network of phosphate starvation in Escherichia coli. PLoS One 7:e47314.

959 57. Leonhartsberger S, Huber A, Lottspeich F, Böck A. 2001. The hydH/G Genes from
960 Escherichia coli code for a zinc and lead responsive two-component regulatory system. J
961 Mol Biol 307:93–105.

962 58. Lee LJ, Barrett JA, Poole RK. 2005. Genome-wide transcriptional response of chemostat-
963 cultured Escherichia coli to zinc. J Bacteriol 187:1124–1134.

964 59. Appia-Ayme C, Hall A, Patrick E, Rajadurai S, Clarke TA, Rowley G. 2012. ZraP is a
965 periplasmic molecular chaperone and a repressor of the zinc-responsive two-component
966 regulator ZraSR. Biochem J 442:85–93.

967 60. Petit-Härtlein I, Rome K, de Rosny E, Molton F, Duboc C, Gueguen E, Rodrigue A, Covès
968 J. 2015. Biophysical and physiological characterization of ZraP from Escherichia coli, the
969 periplasmic accessory protein of the atypical ZraSR two-component system. Biochem J
970 472:205–216.

971 61. Rome K, Borde C, Taher R, Cayron J, Lesterlin C, Gueguen E, De Rosny E, Rodrigue A.
972 2018. The Two-Component System ZraPSR Is a Novel ESR that Contributes to Intrinsic
973 Antibiotic Tolerance in Escherichia coli. J Mol Biol 430:4971–4985.

974 62. Seo SW, Gao Y, Kim D, Szubin R, Yang J, Cho B-K, Palsson BO. 2017. Revealing
975 genome-scale transcriptional regulatory landscape of OmpR highlights its expanded
976 regulatory roles under osmotic stress in Escherichia coli K-12 MG1655. Sci Rep 7:2181.

977 63. Albert Siryaporn MG. 2008. Cross-talk Suppression between the CpxA-CpxR and EnvZ-
978 OmpR Two-Component Systems in E. coli. Mol Microbiol 70:494.

979 64. Seo SW, Kim D, Latif H, O'Brien EJ, Szubin R, Palsson BO. 2014. Deciphering Fur
980 transcriptional regulatory network highlights its complex role beyond iron metabolism in
981 Escherichia coli. Nat Commun 5:4910.

982 65. Tung MS, O'Farrell TJ. 1996. Effect of ethanol on the formation of calcium phosphates.
983 Colloids Surf A Physicochem Eng Asp 110:191–198.

984 66. Larsen MJ, Thorsen A, Jensen SJ. 1985. Ethanol-induced formation of solid calcium
985 phosphates. Calcif Tissue Int 37:189–193.

986 67. McLeay RC, Bailey TL. 2010. Motif Enrichment Analysis: a unified framework and an
987 evaluation on ChIP data. BMC Bioinformatics 11:165.

988 68. Santos-Zavaleta A, Salgado H, Gama-Castro S, Sánchez-Pérez M, Gómez-Romero L,
989 Ledezma-Tejeida D, García-Sotelo JS, Alquicira-Hernández K, Muñiz-Rascado LJ, Peña-
990 Loredo P, Ishida-Gutiérrez C, Velázquez-Ramírez DA, Del Moral-Chávez V, Bonavides-
991 Martínez C, Méndez-Cruz C-F, Galagan J, Collado-Vides J. 2019. RegulonDB v 10.5:
992 tackling challenges to unify classic and high throughput knowledge of gene regulation in E.
993 coli K-12. Nucleic Acids Res 47:D212–D220.

994 69. Vogeleer P, Vincent AT, Chekabab SM, Charette SJ, Novikov A, Caroff M, Beaudry F,
995 Jacques M, Harel J. 2019. Escherichia coli O157:H7 responds to phosphate starvation by
996 modifying LPS involved in biofilm formation. bioRxiv.

997 70. Hansen A-M, Qiu Y, Yeh N, Blattner FR, Durfee T, Jin DJ. 2005. SspA is required for acid
998 resistance in stationary phase by downregulation of H-NS in Escherichia coli. Mol
999 Microbiol 56:719–734.

1000 71. Cho B-K, Knight EM, Palsson BO. 2006. PCR-based tandem epitope tagging system for
1001 Escherichia coli genome engineering. Biotechniques 40:67–72.

1002 72. Datsenko KA, Wanner BL. 2000. One-step inactivation of chromosomal genes in
1003 Escherichia coli K-12 using PCR products. Proc Natl Acad Sci U S A 97:6640–6645.

1004 73. Cho B-K, Kim D, Knight EM, Zengler K, Palsson BO. 2014. Genome-scale reconstruction
1005 of the sigma factor network in Escherichia coli: topology and functional states. BMC Biol
1006 12:4.

1007 74. Gao Y, Yurkovich JT, Seo SW, Kabimoldayev I, Dräger A, Chen K, Sastry AV, Fang X,
1008 Mih N, Yang L, Eichner J, Cho B-K, Kim D, Palsson BO. 2018. Systematic discovery of
1009 uncharacterized transcription factors in Escherichia coli K-12 MG1655. Nucleic Acids Res
1010 46:10682–10696.

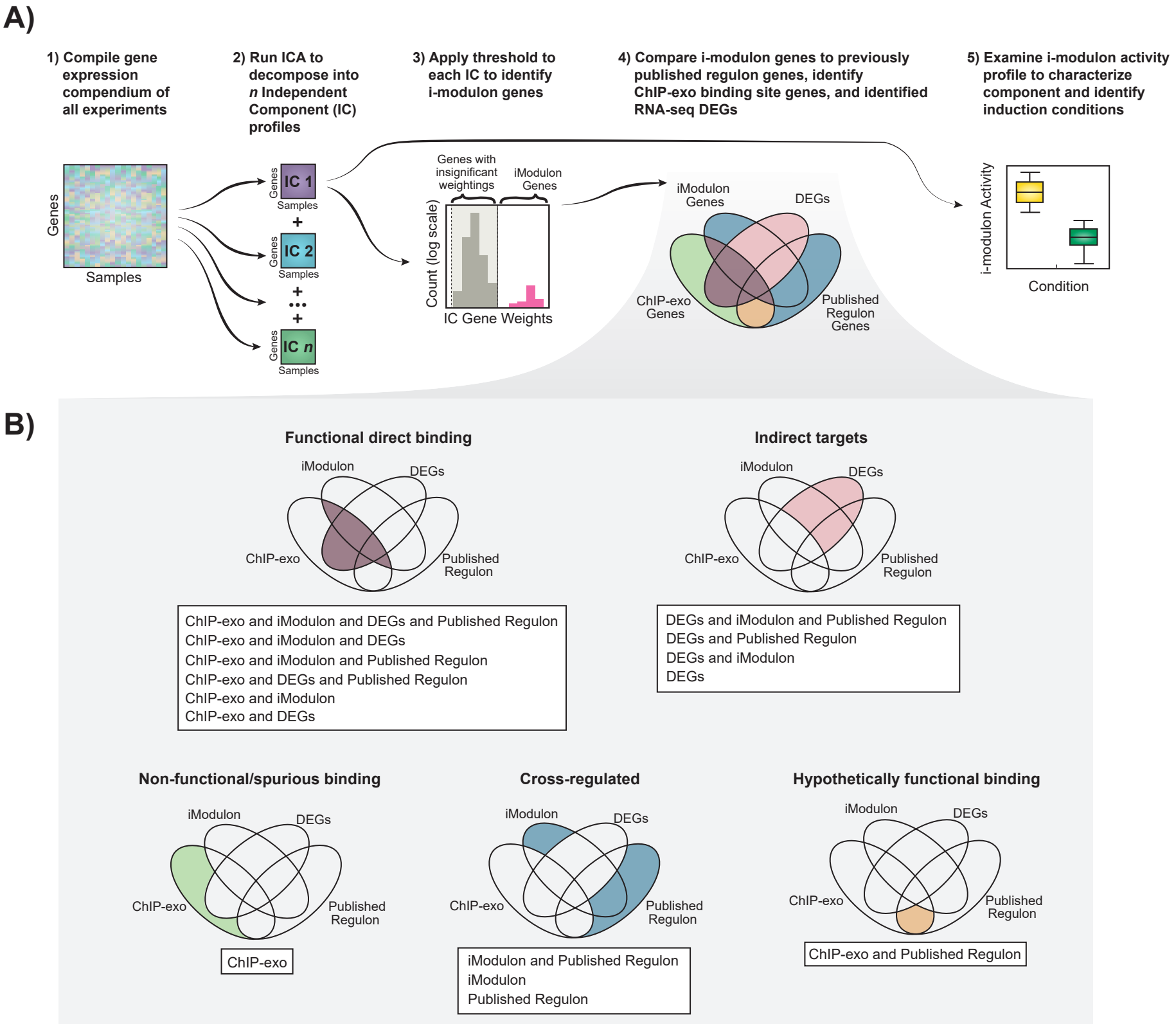
1011 75. Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient
1012 alignment of short DNA sequences to the human genome. Genome Biol 10:R25.

1013 76. Wang L, Chen J, Wang C, Uusküla-Reimand L, Chen K, Medina-Rivera A, Young EJ,

Zimmermann MT, Yan H, Sun Z, Zhang Y, Wu ST, Huang H, Wilson MD, Kocher J-PA, Li W. 2014. MACE: model based analysis of ChIP-exo. Nucleic Acids Res 42:e156.

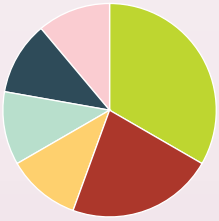
77. Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, Morgan MT, Carey VJ. 2013. Software for computing and annotating genomic ranges. PLoS Comput Biol 9:e1003118.

78. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:550.

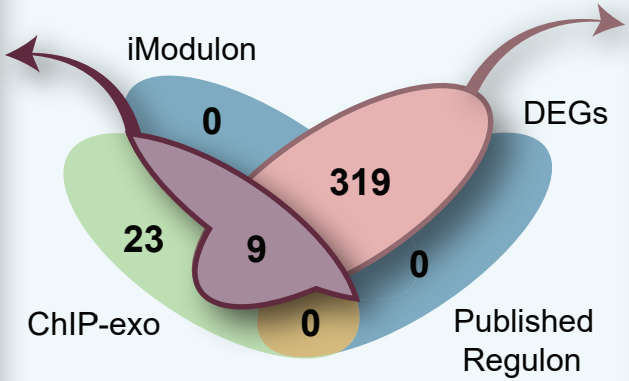

A)
1) Compile gene expression compendium of all experiments
2) Run ICA to decompose into n Independent Component (IC) profiles
3) Apply threshold to each IC to identify i-modulon genes
4) Compare i-modulon genes to previously published regulon genes, identify ChIP-exo binding site genes, and identified RNA-seq DEGs
5) Examine i-modulon activity profile to characterize component and identify induction conditions
Genes
Samples
IC 1
IC 2
IC n
Genes with insignificant weightings
iModulon Genes
Count (log scale)
IC Gene Weights
iModulon Genes
DEGs
ChIP-exo Genes
Published Regulon Genes
i-modulon Activity
Condition
B)
Functional direct binding
iModulon
DEGs
ChIP-exo
Published Regulon
ChIP-exo and iModulon and DEGs and Published Regulon
ChIP-exo and iModulon and DEGs
ChIP-exo and iModulon and Published Regulon
ChIP-exo and DEGs and Published Regulon
ChIP-exo and iModulon
ChIP-exo and DEGs
Indirect targets
iModulon
DEGs
ChIP-exo
Published Regulon
DEGs and iModulon and Published Regulon
DEGs and Published Regulon
DEGs and iModulon
DEGs
Non-functional/spurious binding
iModulon
DEGs
ChIP-exo
Published Regulon
ChIP-exo
Cross-regulated
iModulon
DEGs
ChIP-exo
Published Regulon
iModulon and Published Regulon
iModulon
Published Regulon
Hypothetically functional binding
iModulon
DEGs
ChIP-exo
Published Regulon
ChIP-exo and Published Regulon

BaeR

Direct Targets



- No COG Annotation, 3
- Signal transduction mechanisms, 2
- Amino acid transport and metabolism, 1
- Cell motility, 1
- Defense mechanisms, 1
- Replication, recombination and repair, 1

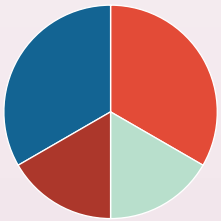


Indirect Targets

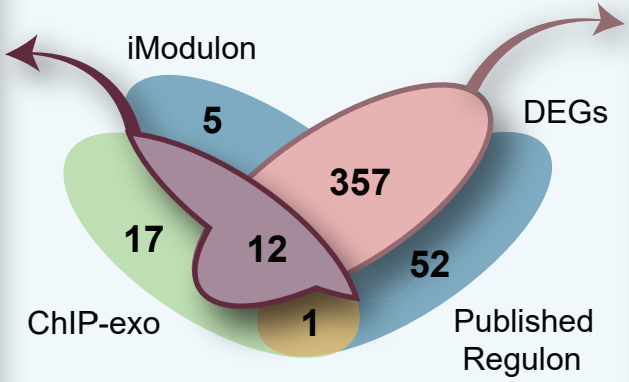
- No COG Annotation, 54
- Function unknown, 51
- Energy production and conversion, 49
- Inorganic ion transport and metabolism, 35
- Others (COGs less than 10% of total), 130

CpxR

Direct Targets



- Function unknown, 4
- Cell motility, 2
- Signal transduction mechanisms, 2
- Others (COGs less than 10% of total), 4

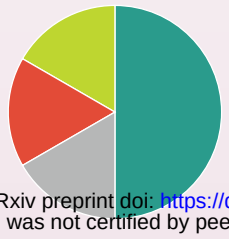


Indirect Targets

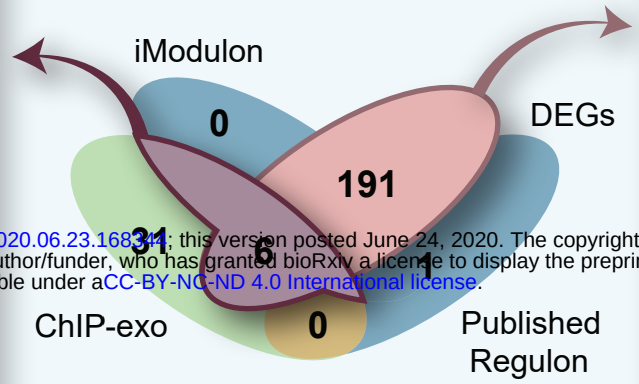
- No COG Annotation, 60
- Energy production and conversion, 53
- Function unknown, 53
- Inorganic ion transport and metabolism, 36
- Others (COGs less than 10% of total), 155

KdpE

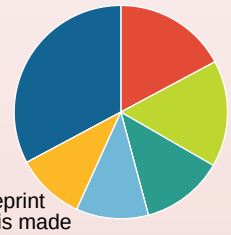
Direct Targets



- Inorganic ion transport and metabolism, 3
- Carbohydrate transport and metabolism, 1
- Function unknown, 1
- No COG Annotation, 1



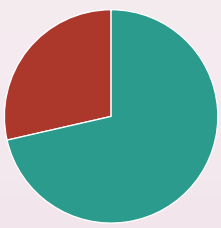
Indirect Targets



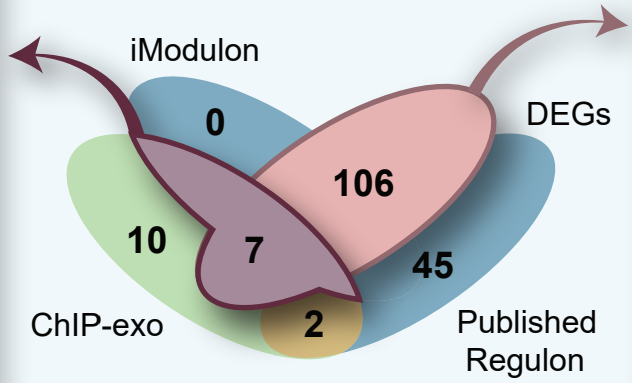
- Function unknown, 33
- No COG Annotation, 30
- Inorganic ion transport and metabolism, 24
- Nucleotide transport and metabolism, 21
- Energy production and conversion, 20
- Others (COGs less than 10% of total), 63

PhoB

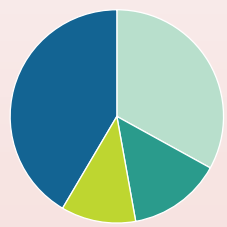
Direct Targets



- Inorganic ion transport and metabolism, 5
- Signal transduction mechanisms, 2



Indirect Targets

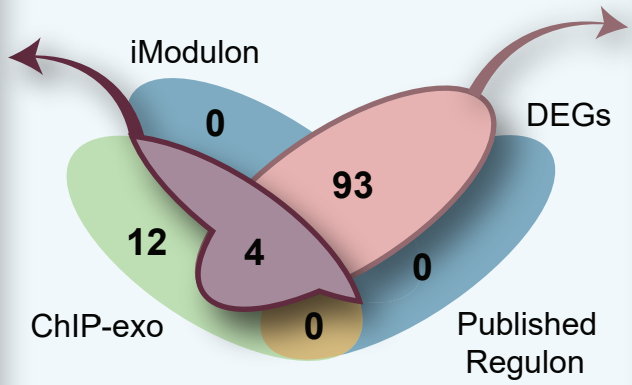


- Cell motility, 35
- Inorganic ion transport and metabolism, 15
- No COG Annotation, 12
- Others (COGs less than 10% of total), 44

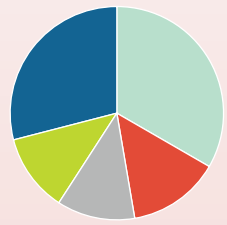
ZraR

Direct Targets

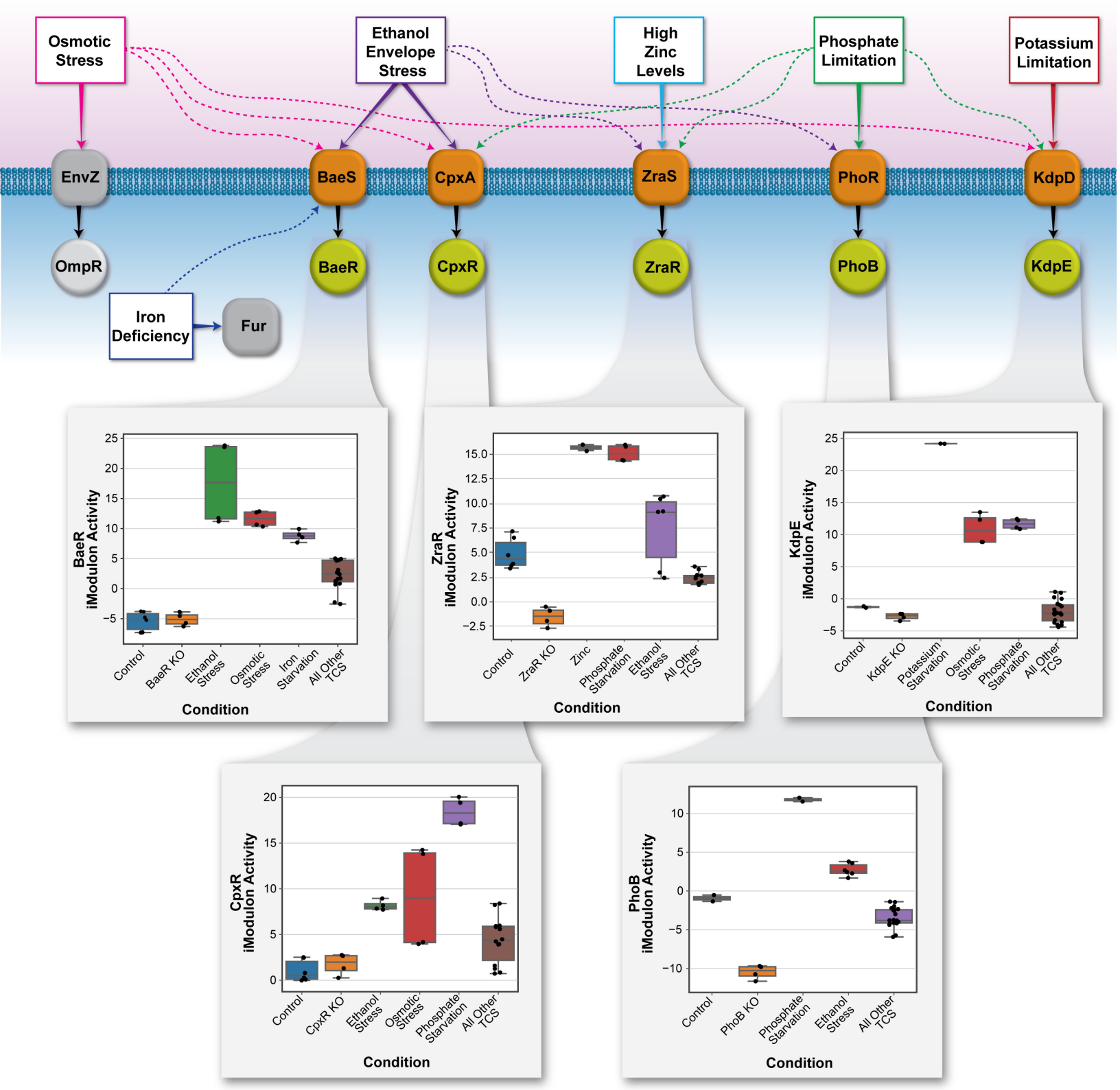
- Inorganic ion transport and metabolism, 2
- Signal transduction mechanisms, 1
- No COG Annotation, 1

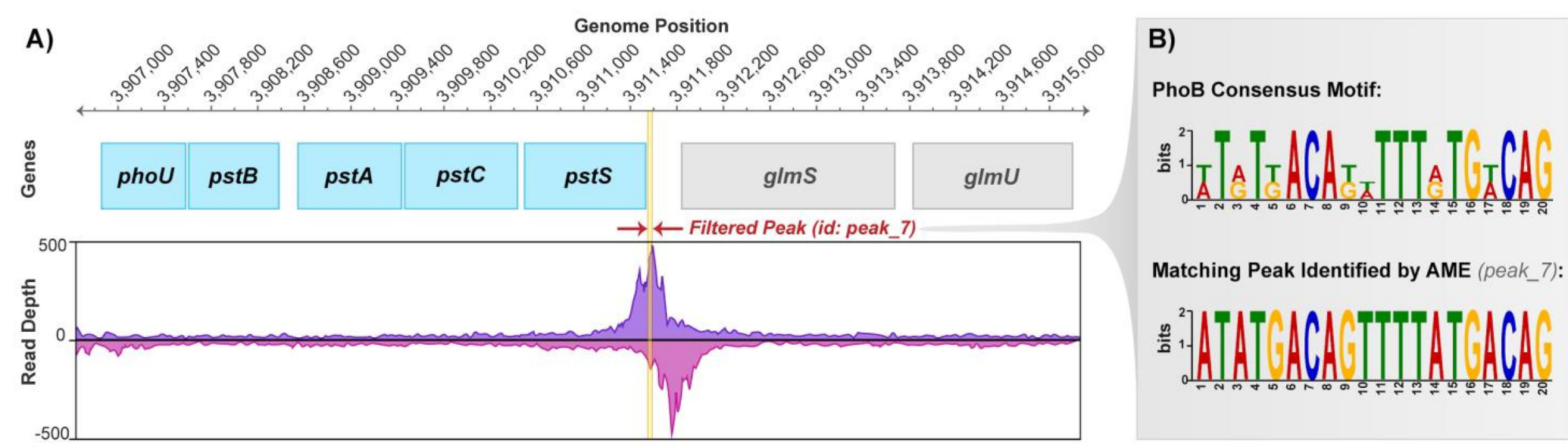


Indirect Targets



- Cell motility, 31
- Function unknown, 13
- Carbohydrate transport and metabolism, 11
- No COG Annotation, 11
- Others (COGs less than 10% of total), 27




A)
Genome Position
3,907,000
3,907,400
3,907,800
3,908,200
3,908,600
3,909,000
3,909,400
3,909,800
3,910,200
3,910,600
3,911,000
3,911,400
3,911,800
3,912,200
3,912,600
3,913,000
3,913,400
3,913,800
3,914,200
3,914,600
3,915,000
Genes
phoU
pstB
pstA
pstC
pstS
glmS
glmU
Filtered Peak (id: peak_7)
Read Depth
500
0
-500
B)
PhoB Consensus Motif:
bits
Matching Peak Identified by AME (peak_7):
bits