

# High-resolution temporal profiling of *E. coli* transcriptional response

Received: 22 July 2023

Accepted: 2 November 2023

Published online: 22 November 2023

 Check for updates

Arianna Miano <sup>1</sup>✉, Kevin Rychel <sup>1</sup>, Andrew Lezia<sup>1</sup>, Anand Sastry<sup>1</sup>, Bernhard Palsson<sup>1,2</sup> & Jeff Hasty<sup>1,3,4</sup>

Understanding how cells dynamically adapt to their environment is a primary focus of biology research. Temporal information about cellular behavior is often limited by both small numbers of data time-points and the methods used to analyze this data. Here, we apply unsupervised machine learning to a data set containing the activity of 1805 native promoters in *E. coli* measured every 10 minutes in a high-throughput microfluidic device via fluorescence time-lapse microscopy. Specifically, this data set reveals *E. coli* transcriptome dynamics when exposed to different heavy metal ions. We use a bioinformatics pipeline based on Independent Component Analysis (ICA) to generate insights and hypotheses from this data. We discovered three primary, time-dependent stages of promoter activation to heavy metal stress (fast, intermediate, and steady). Furthermore, we uncovered a global strategy *E. coli* uses to reallocate resources from stress-related promoters to growth-related promoters following exposure to heavy metal stress.

Over the past few decades, various high-throughput technologies have emerged to investigate how cellular transcription changes in response to environmental perturbations<sup>1,2</sup>. Currently, RNA sequencing (RNA-seq) is the most widely used method to determine the relative abundance of each mRNA transcript in a cellular population<sup>3</sup>. However, RNA-seq requires the destruction of cells to harvest RNA, making it challenging to obtain the necessary time resolution to study the dynamics of transcription. Although sampling RNA from parallel cultures at different times can partially overcome this limitation, many protocols only sample mRNA levels at a few time points resulting in a lack of information regarding transcriptional regulatory network (TRN) dynamics<sup>4,5</sup>.

To overcome these limitations, we previously introduced Dynamics, a microfluidic platform for monitoring *Escherichia coli* promoter activity in real-time using fluorescence time lapse microscopy<sup>6</sup>. This technology allows us to grow more than 2000 unique bacterial strains continuously in small cell “traps” in the presence of various nutrients and stressors. To monitor the dynamics of the *E. coli* TRN, we combined a pre-existing *E. coli* promoter library with

Dynamics. The promoter library, created by Zaslaver et al., contains approximately 2000 unique *E. coli* strains where each strain has a different, native promoter driving the expression of green fluorescent protein (GFP) on a low-copy number plasmid<sup>7</sup>. This library enables highly accurate dynamic measurements of each promoter in the genome as initially demonstrated in an experiment looking at the response of *E. coli* during a diauxic growth shift<sup>7</sup>. By combining this library with our microfluidic device and a custom imaging platform, we were able to record transcriptional cellular responses to environmental perturbations every 10 minutes for up to 14 days. Using this technology, we observed the dynamics of all *E. coli* promoters with unprecedented temporal resolution.

In the original Dynamics study, Graham et al. explored how *E. coli* responds to heavy metal exposure. Heavy metal contamination, due to industrial wastewater discharge, has become a significant threat to the environment<sup>8</sup>. In response to the presence of these ions, many microorganisms have evolved strategies to cope with high concentrations of toxic heavy metals. For instance, some bacteria express enzymes that can reduce heavy metal ions, lowering their toxicity<sup>9</sup>. The

<sup>1</sup>Department of Bioengineering, University of California San Diego, 9500 Gilman Dr, La Jolla, CA, USA. <sup>2</sup>Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kemitorvet, Building 220, 2800 Kgs, Lyngby, Denmark. <sup>3</sup>Department of Molecular Biology, School of Biological Sciences, University of California San Diego, 9500 Gilman Dr, La Jolla, CA, USA. <sup>4</sup>Synthetic Biology Institute, University of California San Diego, 9500 Gilman Dr, La Jolla, CA, USA. ✉e-mail: [armiano@ucsd.edu](mailto:armiano@ucsd.edu)

breadth of metal-responsive promoters and genes found throughout the microbial world have made bacteria a popular choice for engineering heavy metal biosensors<sup>6</sup> and bioremediation strategies<sup>10</sup>. While the potential applications of studying bacterial responses to heavy metals have received significant attention, we are interested in the fundamental biological implications of the timing of transcriptional responses to different types of heavy metals. This analysis has the potential to shed light on TRN structure and to help hypothesize on the cellular responses to specific metals based on the temporal patterns of gene activations observed.

Transcriptomic methods such as RNAseq and Dynamics generate complex, high dimensional data sets that require sophisticated analysis techniques to interpret<sup>11</sup>. Independent component analysis (ICA), an unsupervised machine learning technique developed to deconvolute mixed signals into individual sources<sup>12</sup>, has proven particularly successful at extracting biologically relevant transcriptional modules across the phylogenetic tree from a variety of transcriptomic data sets<sup>13-17</sup>. In fact, ICA outperformed 42 similar methods in reproducing known transcriptional modules<sup>18</sup>. Using ICA, we can identify groups of genes, called iModulons, which are co-regulated throughout the data and are associated with promoter activity levels in each sample. Generating iModulons has been useful for understanding TRNs and low-resolution dynamics in RNAseq and microarray data<sup>19</sup>. For instance, ICA was able to quantitatively summarize the major steps in *Bacillus subtilis* sporulation by grouping relevant genes into specific iModulons<sup>13</sup>. Moreover, ICA consistently identifies similar gene groupings, even between RNAseq and microarray data sets<sup>19</sup>.

In this study, we apply the ICA workflow to the gene expression dataset originally generated by Graham et al., thereby leveraging pre-existing data to derive insights into the temporal patterns of activation of bacterial promoters in response to external heavy metal stress. The data was previously generated using environmentally relevant concentrations of heavy metals, typically several orders of magnitude smaller than the minimum inhibitory concentration (MIC) known for *E. coli*<sup>20</sup>. Our results confirm previously known patterns of promoter activation while also identifying additional gene associations. Importantly, to our knowledge, this is the first time ICA has been applied to a time-series transcriptomic data set with measurements taken at regular short time intervals across multiple days. Combining ICA with high temporal-resolution data enabled us to generate iModulons that reveal the response patterns of different gene groups as a function of time. This approach generates interesting hypotheses regarding transcriptional dynamics.

## Results

### Dynamic transcriptional response to heavy metal exposure

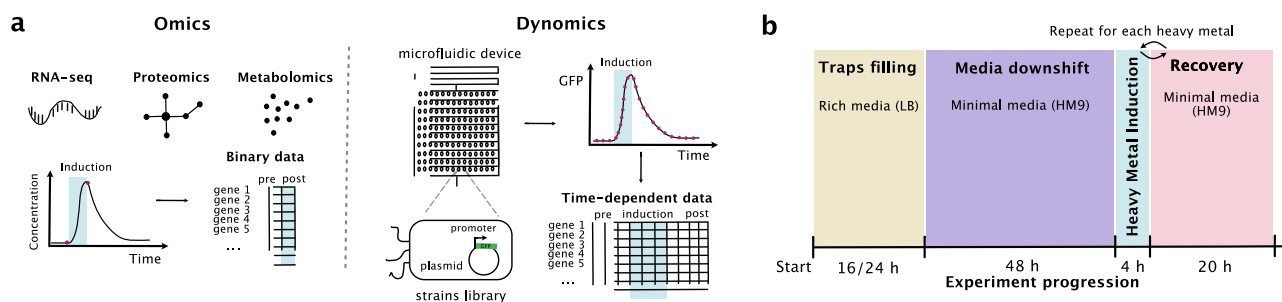
The protocol used to acquire the GFP promoter library data analyzed in this study is described in detail in the original Dynamics

publication<sup>6</sup>. In brief, each strain was induced with a heavy metal ion for four hours and then left to recover in minimal media for 20 h (Fig. 1). A custom fluorescence imaging system was used to record the average GFP expression every 10 min. The raw fluorescence data were background-subtracted, median-filtered, and normalized using the mean GFP intensity of a control strain. The fold change was calculated as the log base 2 of the ratio of fluorescence values at selected time points within the induction window to the fluorescence values at the start of induction. To minimize noise while retaining temporal information, six time points spaced forty minutes apart were selected for each induction window. This data was then organized into a matrix, *X*, where each of 1805 rows represented a different promoter and each of 36 columns represented the log<sub>2</sub> of the fold change for each time point during the six heavy metal inductions. Thus, a total of 36 conditions were analyzed for each promoter (6 different heavy metals multiplied by 6 different induction time windows) (Fig. 2a).

ICA was applied to the matrix, *X*, resulting in an *M* matrix of promoter coefficients and an *A* matrix of activity coefficients (Supplementary Data 1-3). The *M* matrix has dimensions of 1805 by 15, which correspond to 1805 promoters and 15 independently modulated sets of genes referred to as iModulons as suggested in previous studies<sup>14</sup>. The activity profiles and corresponding gene weights for all 15 iModulons obtained from the bioinformatics pipeline are provided in the Supplementary Information (Figs. S1 to S15). Out of the 15 iModulons, six (iModulons 0, 2, 4, 5, 9, and 13) were discarded because they were dominated by a single high coefficient promoter. These iModulons are uninformative and are usually the result of noisy promoters in the data.

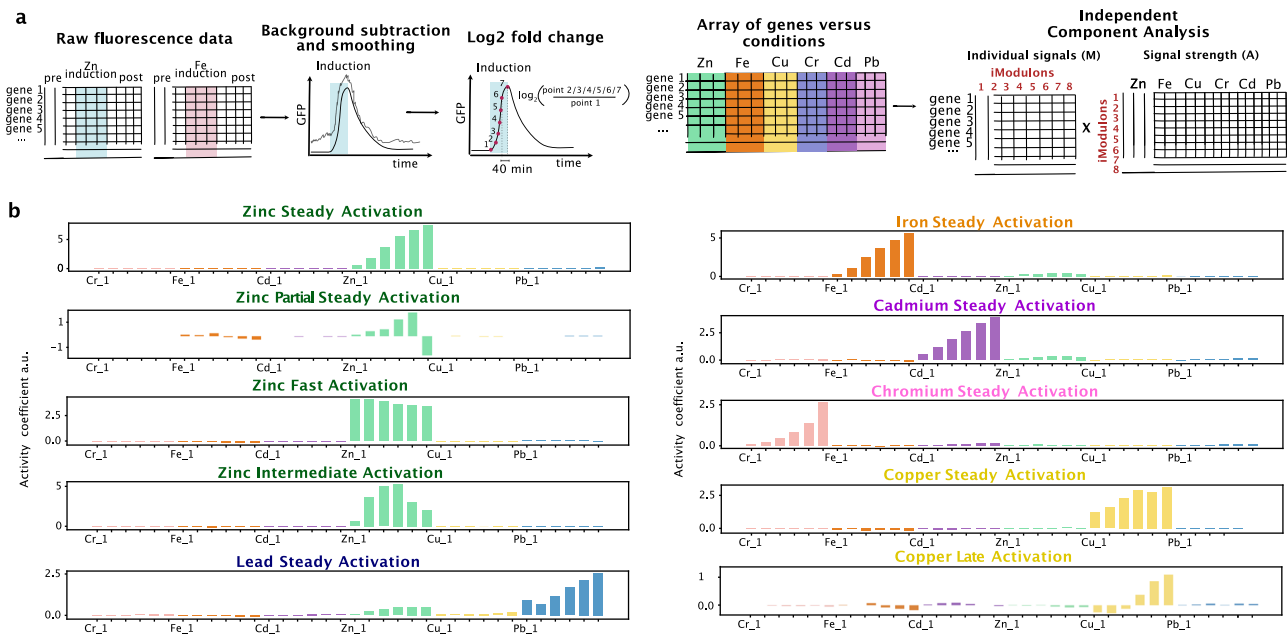
While the *M* matrix shows how much each promoter contributes to each iModulon, the *A* matrix shows the relative contribution of each heavy metal induction window (e.g., zinc induction and time window two) to each iModulon. We named each iModulon according to its shape and the predominant environmental condition associated with it (Fig. 2b). From the *A* matrix, we see that each iModulon was largely specific to a single heavy metal suggesting that the underlying biology is highly sensitive to which heavy metal is present (Supplementary Fig. S52).

For lead, iron, cadmium, and chromium, a single iModulon per metal was sufficient to capture the observed response, and these iModulons represent promoters that became steadily activated over the course of the induction. Zinc, on the other hand, had four separate iModulons that captured unique activation patterns. We classified these patterns as “Steady Activation” for promoters whose activity steadily increased over the induction window, “Fast Activation” for promoters that activated quickly after the beginning of the induction, “Intermediate Activation” for promoters that were mostly active in the middle of the activation window, and lastly “Partial Steady Activation” for promoters that were steadily activated at the beginning of the



**Fig. 1 | Dynamics experimental set up.** **a** Illustration of the differences between standard -omics technologies (left) and the Dynamics experimental set up using a 2176-strain microfluidic device (right). **b** Diagram to illustrate the experiment steps. In chronological order the steps are: growth on rich media (LB) until cells

completely fill the traps, downshift to minimal media (M9) for 48 h, induction with the heavy metal of choice for 4 h, recovery with minimal media. After 20 h on minimal media another cycle of induction and recovery can be performed.



**Fig. 2 | Independent Component Analysis (ICA) to analyze heavy metal inductions data from Dynamics experiments.** **a** Illustration of the different steps of our data analysis pipeline. The analysis starts with raw fluorescence data which are processed with a background signal removal algorithm, normalized by promoterless strains, and smoothed by median filtering. The log<sub>2</sub> is taken to convert to fold change, and the data is formatted as a matrix of genes versus conditions (heavy metal inductions). ICA is applied to this matrix to obtain the M (promoter

coefficients) and A (activity coefficients) matrices respectively. **b** iModulon activity plots representing different response dynamics to the heavy metals investigated. The x-axis represents five time points spaced 40 min apart for each of the heavy-metal shown. For illustration purposes only the first time point of each metal is labeled on the x-axis. The y-axis is unitless as it represents the log<sub>2</sub> of the fold change of the normalized fluorescence data.

induction window and then suddenly repressed in the last part of the induction window. A summary of the different activation patterns is included in Supplementary Fig. S52. Copper exposure was associated with two iModulons (“Steady” and “Late Activation” iModulons) based on their activation profile pattern. The remaining heavy metal iModulons exhibited a single “Steady Activation” profile.

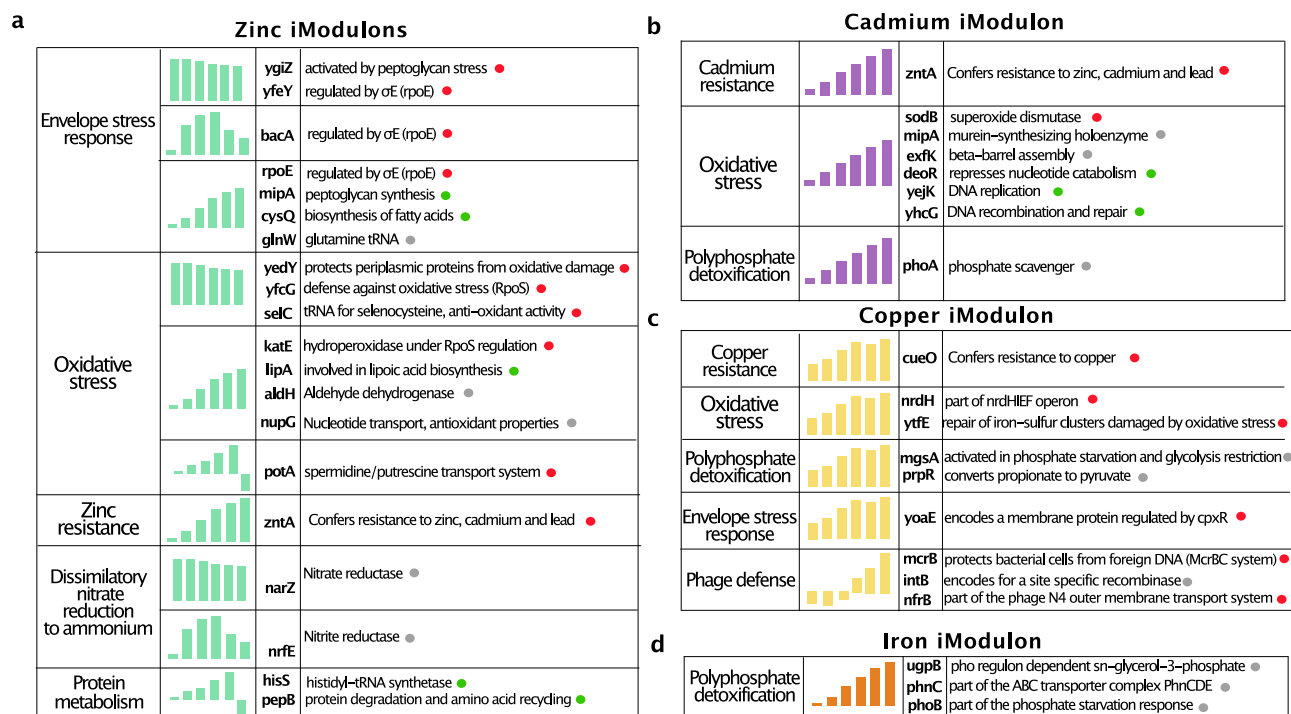
To identify the most significant promoters in each iModulon, we employed a thresholding method. In this method, we iteratively remove promoters with the highest absolute weighting from an iModulon until the D’Agostino K2 statistic of normality of the remaining distribution falls below a predetermined cutoff. This cutoff distinguishes high-coefficient promoters from those with negligible influence that cluster around zero. By using a fixed threshold value across all iModulons, we minimized bias and ensured a reliable analysis. More details can be found in the original publications that used this thresholding method<sup>21,22</sup>. After applying the calculated cutoffs, we extracted and analyzed the promoters with the highest coefficients for each iModulon. To better understand the overall bacterial response to each heavy metal, we classified the promoters by phenotype (i.e., the functions associated with each promoter based on existing literature) and activation profile (i.e., how and when the promoter is activated) (Fig. 3). In the following sections describing the features of the iModulons, when we state that a certain gene belongs to an iModulon, we mean that the promoter activity for that gene matched the activation profile for that iModulon. We will generally refer to genes as belonging to iModulons for brevity and clarity with the understanding that it is actually the activity of the gene promoters that we are analyzing.

### Zinc iModulons

We observed that the zinc iModulons were associated with several functions, including envelope stress response, oxidative stress response, zinc resistance, the dissimilatory nitrate reduction to ammonium pathway, and protein metabolism (Fig. 3a). Within this iModulon, we identified a set of genes directly related to the envelope

stress response (*yfeY*, *baca*, *ropE*) and to peptidoglycan stress that were present in both fast and steady activation profiles, consistent with previous literature<sup>23–25</sup>. Notably, we also found that two genes associated with steady activation (*mipA*, *cysQ*) are directly involved in peptidoglycan synthesis. This iModulon also includes the activation of the *glnW* promoter, which transcribes glutamine tRNA. Given that glutamine is important for fatty acid synthesis, we speculate that the upregulation of *glnW* may help restore membrane damage caused by zinc exposure. Our ICA analysis also revealed multiple genes involved in the oxidative stress response, consistent with former studies<sup>26</sup>, and distinguished between fast activation (*yedY*, *yfcG*, *selC*) and steady activation (*katE*, *lipA*, *aldH*, *nupG*). Interestingly, the promoters for *yedY* and *yfcG* specifically regulate genes that combat oxidative damage<sup>27,28</sup>, indicating a targeted response against cellular damage. However, the genes associated with steady activation are involved in cellular processes unrelated to defense response. These include lipote biosynthesis<sup>29</sup>, nucleoside transport<sup>30</sup>, and hydrogen peroxide detoxification<sup>31</sup>.

Our results also confirm the activation of the *zntA* gene which is associated with zinc export as a detoxification mechanism when excess levels of zinc are detected<sup>32</sup>. Interestingly, our analysis is also able to detect the activation of two genes (*narZ* and *nrfE*) belonging to the dissimilatory nitrate reduction to ammonium pathway (DNRA)<sup>32</sup>. In particular, we were able to observe the temporal dynamics of the pathway activation since *narZ* (nitrate reductase) belonged to the fast activation iModulon and *nrfE* (nitrite reductase) was associated with the intermediate activation profile. The order of activation matches the order required for nitrogen respiration, with nitrate first being reduced to nitrite by *narZ*, then nitrite being reduced to ammonium by *nrfE*<sup>33</sup>. This mechanism follows the “just-in-time” transcription program in metabolic pathways that was previously suggested in the literature<sup>34</sup>. To our knowledge, this is the first time that the activation of DNRA pathway in *E. Coli* is linked to the presence of elevated zinc concentrations. We hypothesize that in this context, DNRA activation



**Fig. 3 | Illustration of significant promoters enriched by the ICA analysis organized by function and activation profile.** Green dots indicate promoters that drive genes related to transcription/translation/synthesis functions. Red dots indicate promoters related to stress/damage responses. Grey dots indicate promoters in neither of those categories. **a** Significant promoters associated to the

induction of zinc. Promoters are classified according to the physiological function they are associated with. **b** Significant promoters associated to the induction of cadmium. **c** Significant promoters associated to the induction of copper. **d** Significant promoters associated to the induction of iron.

is beneficial to the cells due to the generation of an electron sink that can be used for NADH re-oxidation into NAD<sup>+</sup> which is a key metabolite in counteracting DNA damage and oxidative stress<sup>35,36</sup>. Future experiments will be needed to validate this claim. Finally, we also detect a “Partial Steady Activation” iModulon, characterized by an initial steady response which is followed by a sudden repression in the last phase of the induction window. Interestingly, we find that the promoter which transcribes for the gene *potA* is enriched. This gene is involved in the spermidine and putrescine transport system. Previous studies found that putrescine and spermidine added to the culture medium significantly increased expression of oxidative-stress related genes *oxyR* and *katG*<sup>37</sup>. In line with these findings, we observe that zinc exposure leads to the activation of both *potA* and *katE* promoters (Fig. 3a). The same iModulon enriched for two genes (*hisS*, *pepB*) which are both related to protein metabolism. The *hisS* gene encodes histidyl-tRNA synthetase, an essential enzyme in protein synthesis that charges tRNA with the amino acid histidine<sup>38</sup>. The *pepB* gene encodes endopeptidase PepB or leucyl aminopeptidase, involved in intracellular protein degradation and amino acid recycling<sup>39</sup>.

### Cadmium iModulons

The cadmium iModulon was characterized by a steady increase in promoter activation and included promoters that are mainly associated with cadmium resistance and oxidative stress (Fig. 3b). As expected from the literature<sup>32</sup>, we found that the *zntA* gene was up-regulated, which confers cadmium tolerance as well as zinc tolerance as previously described. Furthermore, we detected the activation of the *sodB* promoter, which encodes a superoxide dismutase that is a key factor in defending the cell against oxidative stress through decomposition of superoxide radicals. This finding agrees with previous results reported in the literature, which showed that superoxide dismutases (SODs) can protect *E. coli* from heavy metal toxicity, particularly from cadmium exposure<sup>40</sup>.

The iModulon also captures the cell’s response to two of the main effects of oxidative stress: lipid peroxidation and DNA damage. In particular, we report the activation of the promoter for *mipA*, which encodes a scaffolding protein for murein synthesizing machinery<sup>41</sup>, and the promoter for *exfK* (also known as *bama*), which is involved in the assembly and insertion of beta-barrel proteins in the outer membrane<sup>42</sup>. Additionally, we identify three promoters which are known to be involved in nucleotide catabolism repression (*deoR*<sup>43</sup>), DNA replication (*yejK*<sup>44</sup>), and DNA recombination and repair (*yhcG*<sup>45</sup>) respectively. Finally, we detect the activation of the *phoA* promoter, which is responsible for the breakdown of organic phosphate esters<sup>46</sup>.

### Copper iModulons

The copper iModulon was characterized by a “Steady” activation profile and a “Late” activation profile (Fig. 3c). As expected, the Steady iModulon contains the *cueO* promoter which encodes for a multi-copper oxidase involved in copper tolerance under aerobic conditions<sup>47</sup>. Furthermore, we detect the activation of promoters known to be linked to the oxidative stress response (*nrdH*<sup>48</sup>, *ytfE*<sup>49</sup>) and to the envelope stress response (*yoeE*<sup>50</sup>). Interestingly, we also detect the activation of promoter *mgsA* which is known to be activated during phosphate starvation and glycolysis restriction<sup>51</sup>. Similarly, we report the activation of promoter *prpR* which is part of the *Pho* regulon which we also detected in the iron iModulon<sup>52</sup>.

The Late Activation iModulon is characterized by a late response with increased promoter activity starting roughly halfway through the induction window. Interestingly, we find that this iModulon was enriched for multiple genes related to the cell’s response against phages including *mcrB*, involved in a restriction-modification system for defense against foreign DNA<sup>53</sup>; *intB*, encoding an integrase enzyme for site-specific recombination<sup>54</sup>; *prpR*, a transcriptional regulator controlling propionate catabolism<sup>55</sup>; and *nfrB*, part of the N4 phage outer membrane transport system involved in phage DNA entry during

infection<sup>56</sup>. It is interesting to note that three of these genes (*mcrB*, *intB*, and *nfrB*) appear to be related to phage defense and interaction, suggesting a potential coordinated response to phage infection or other stressors that could impact the bacterial cell's interaction with foreign DNA. We hypothesize that the activation of genes related to phage defense and interaction might reflect a broader stress response aimed at protecting the cell from additional threats, such as foreign DNA or phage infection, under conditions where cellular processes and genomic integrity are already compromised by copper-induced stress.

### Iron iModulons

The iron iModulon is also characterized by a steady state activation profile (Fig. 3d). Interestingly, we found three enriched genes (*ugpB*, *phnC* and *phoB*) which all belong to the Pho regulon, a group of genes in *E. coli* that are involved in the regulation of phosphate metabolism<sup>57</sup>. In particular, *phoB* is the response regulator in a two component regulatory system with *phoR* (or *creC*) and regulates inorganic phosphate (P<sub>i</sub>) uptake. The gene *phnC* is part of the ABC transporter complex PhnCDE involved in phosphonates, phosphate esters, phosphite and phosphate import. Finally, gene *ugpB* codes for a binding protein-dependent sn-glycerol-3-phosphate transport system which is under the control of the Pho regulon. In particular, *ugp*-dependent G3P transport activity is present only after growth at limiting concentrations of P<sub>i</sub><sup>58</sup>. Previous literature indicates that excess iron can cause the formation of toxic reactive oxygen species (ROS) through Fenton chemistry<sup>59</sup>. Both iron stress and oxidative DNA damage are successfully prevented by polyphosphates which are polymers formed by covalently linked inorganic phosphates<sup>59</sup>. Therefore, we hypothesize that the genes enriched in the Iron iModulon reflect the need for the cell to import additional phosphate in order to form polyphosphates which can be used as a defense mechanism against iron. When excess iron is present, *E. coli* can transport it into the cell and store it within the polyphosphate granules, which act as a sink for the iron<sup>60</sup>. Based on the lack of specific iron toxicity genes in the iModulon, we propose that the concentration of iron used was too low to induce specific responses. Nonetheless, it is still informative that phosphate chemistry accounts for the ROS-related effects of an increase in iron.

### Transcriptional response of the recovery post-induction

The primary advantage of the Dynamics technology is the ability to track promoter activity over an extended period of time. This enabled us to analyze not only the effect of the heavy metal exposure during the induction window, but also to explore how the bacteria responded during the recovery period post-induction. In this case, we calculated the log<sub>2</sub> of the fold change with respect to the end of the induction window (Fig. 4a). We took into consideration 20 time points spaced 40 minutes apart (Supplementary Data 4). Applying the ICA algorithm to this dataset we found 35 iModulons (Figs. 4b, S16 to 50, Supplementary Data 5 and 6). We focus our analysis on a subset of iModulons whose activity seemed particularly interesting and biologically significant (Fig. 4c, Methods).

### Zinc iModulons

We report four iModulons whose activity profiles show activation following zinc exposure. In line with the induction response, we identify a few promoters associated with stress responses such as the SOS response (*recX*), peptoglycan stress (*yjfQ*, *helD*, *skp*) and oxidative stress (*kateE*). We also report the activation of promoters which are related to transcription and translation processes such as RNA processing and decay (*rne*), purine metabolism (*yneF*), proline tRNAs (*proL*), methionine synthesis (*metF*), and DNA ligation (*ligB*). Additionally, we find that one of the zinc-related iModulons (iModulon 16) transcribes mRNA for monothiol glutaredoxin (*ydhD*, also known as *grxD*) which is involved in the biogenesis of iron-sulfur clusters<sup>61</sup>. This

confirms previous studies that have shown that excess zinc is associated with the disruption of iron-sulfur clusters in *E. coli*<sup>62</sup>. Interestingly, the same iModulon contains the *edd* promoter which encodes the enzyme phosphogluconate dehydratase which is the first reaction in the Entner-Doudoroff pathway and uses the [4Fe-4S] iron-sulfur cluster as a cofactor<sup>63</sup>.

### Cadmium iModulons

The ICA analysis also identifies two iModulons whose activity profiles are mostly active post cadmium exposure (Fig. 4c). Similar to the trend we observed for the zinc iModulons, we find that there are several promoters associated with biosynthesis and protein translation processes. This bolsters our prediction that the cells mainly activate stress response processes during the induction itself and switch to recovery and growth processes when the inducer is removed. In particular, for the post-induction cadmium iModulon we report the activation of promoter *asd* which is involved in L-lysine, L-methionine, and L-threonine biosynthesis, promoter *tufB* which codes for an elongation factor for protein biosynthesis and *ykgM* which encodes a ribosomal protein. Interestingly, *ykgM* was previously found to be upregulated upon zinc starvation<sup>64,65</sup>. This physiological response is further supported by the activation of promoter *znuC* which drives expression of a zinc import ATP-binding protein. Our results therefore align with previous studies based on the analysis of genome-wide temporal gene expression data which suggested that the molecular mechanisms of cadmium toxicity could be partially explained by the disruption in the transcription of genes encoding ribosomal proteins and zinc-binding proteins<sup>66</sup>.

### Iron iModulons

Furthermore, we report two iModulons containing promoters that activate following iron induction. We find that one of the promoters enriched in these iModulons is *ompW* which transcribes mRNA for an outer membrane protein whose expression has been previously observed to be down-regulated in iron limited growth conditions<sup>67</sup>. Our results show that the expression of *ompW* is greatly increased post-induction when the toxin is removed from the media. Additionally, we find a second iron iModulon with a steeper activation profile which is associated with the activation of two genes: *rfbB* (involved in biogenesis of the bacterial outer membrane) and *ydcl* (uncharacterized lipoprotein).

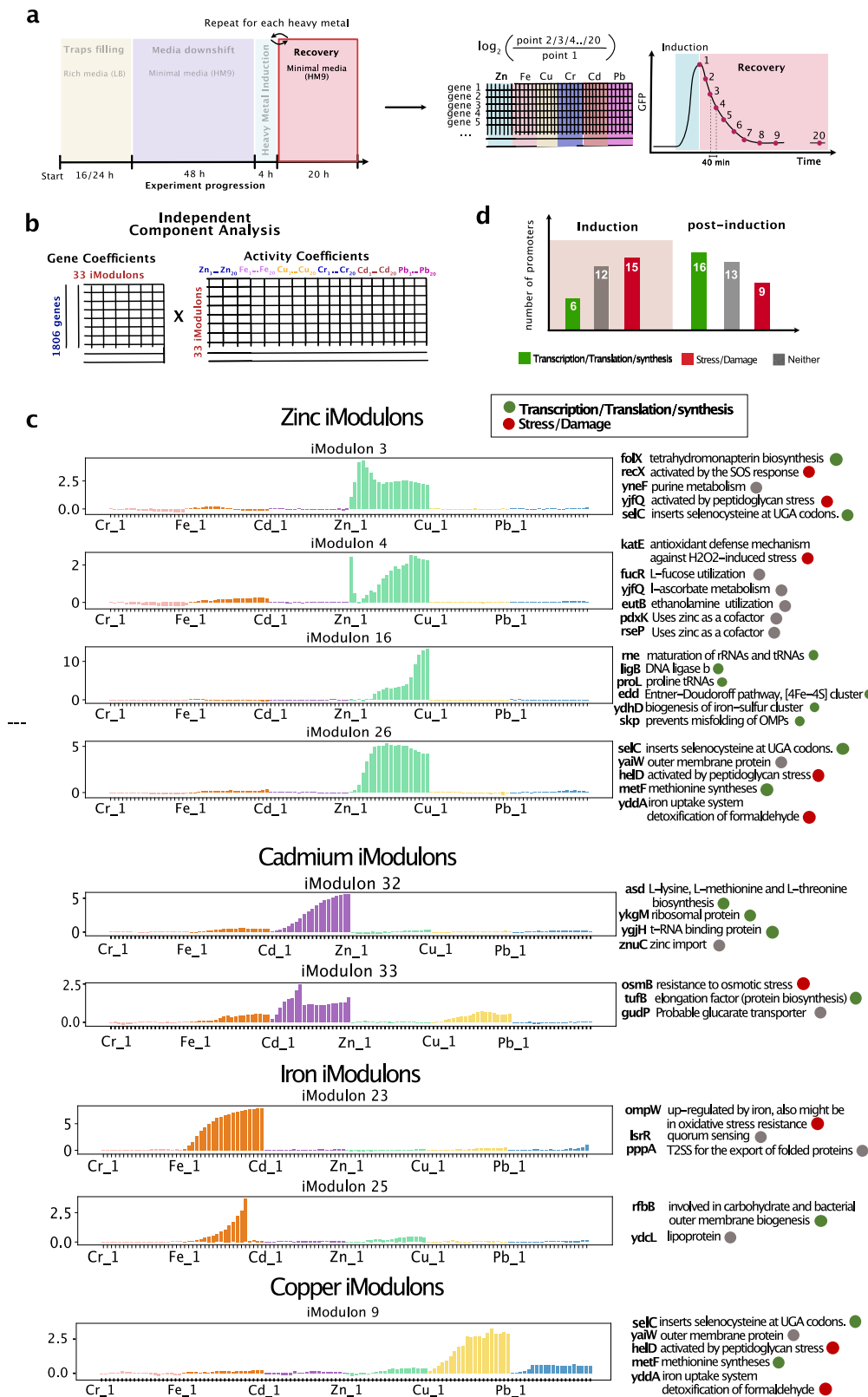
### Copper iModulons

Finally, we report a copper recovery iModulon linked to the activation of five genes, two of which are involved in amino acid biosynthesis (*metF*) and DNA replication/transcription processes (*helD*). We categorized the promoters according to their function in three main groups which are identified with dots of different colors in Fig. 4.

Analyzing the data across the induction and post-induction window, we report a clear shift from an enrichment of promoters driving genes related to stress and damage responses (45% during induction versus 23% post induction) to promoters associated with transcription, translation and synthesis functions (18% during induction versus 42% post induction)(Fig. 4d).

## Discussion

This study presents the integration of a genome-scale platform for monitoring temporal gene expression with the analytical power of ICA<sup>6,14</sup>. Previous studies showed that ICA is highly effective at extracting biologically significant transcriptional modules (called iModulons) from a wide range of datasets, but these datasets lacked information on the time-dependent activation of genes. Instead, most studies have relied on static data obtained from RNA-seq, proteomics, and metabolomics experiments<sup>3,19,68,69</sup>. The Dynamics microfluidic platform enabled sampling of promoter activity every 10 min for



**Fig. 4 | ICA analysis of the recovery period following heavy metal inductions.** **a** Diagram illustrating the steps involved in the analysis pipeline. The ICA was applied to the data collected from the end of the induction window. We considered 20 time points at a distance of 40 min each. **b** Matrix decomposition after applying ICA. Overall, the ICA analysis identifies 33 iModules, a selection of which is shown in panel C. **c** Activity plots and related significant genes for a selection of iModules associated to the post-induction response. The x-axis represents five time points

spaced 40 min apart for each of the heavy-metal shown. For illustration purposes only the first time point of each metal is labeled. The y-axis is unitless as it represents the log<sub>2</sub> of the fold change of the normalized fluorescence data. Green dots indicate promoters that drive genes related to transcription/translation/synthesis functions. Red dots indicate promoters related to stress/damage responses. Grey dots indicate promoters in neither of those categories. **d** Comparison of number of promoters grouped by function during and after heavy metal induction.

multiple days, which is unrivaled by any transcriptome-scale analysis method to our knowledge.

By applying ICA to this data set, we demonstrated the importance of time-dependent analysis in providing insights into the dynamic nature of gene expression in response to environmental stressors. By splitting different heavy metal inductions into separate 40 minute time windows, we were able to apply ICA to time-series transcriptomic data for the first time. We observed the richest response for the zinc inductions. Specifically, we found four different iModulons which differentiate fast responders (genes that are activated at the start of the induction window) from intermediate responders (genes that are maximally active in the middle of the induction window), steady responders (genes whose expression steadily increases throughout the induction window) and partially steady responders (genes that steadily increase over time until they are repressed in the last window). The zinc data demonstrates the ability of this analysis method to resolve the activation sequence of promoters involved in the same metabolic pathway. We detect the activation of promoter *narZ* as an early responder and promoter *nrfE* as a late responder which are involved in the first and second step of dissimilatory nitrate reduction to ammonium metabolic pathway. This result is a clear example of the power of this platform when used for metabolic pathway reconstruction which is a topic of great interest in systems biology<sup>70</sup>.

Additionally, the data in this study were generated using environmentally relevant heavy metal concentrations which are significantly lower than the concentration range known to be toxic for *E. coli*<sup>71</sup>. This allowed us to study the bacterial response to elevated, but non-toxic, levels of iron in the environment. Our ICA analysis suggests that in this scenario the bacteria activate several genes in the Pho regulon (which regulates phosphate uptake and metabolism) in response to elevated concentrations of iron. We hypothesize that bacteria accumulate phosphate as an early response to excess iron in preparation for polyphosphate synthesis. Several studies have shown that polyphosphates can sequester iron in *E. coli* and other microorganisms, limiting its bioavailability and protecting cells from iron toxicity<sup>59,72</sup>. Further studies need to be conducted to experimentally verify this hypothesis.

Lastly, this study expands our understanding of the recovery process *E. coli* following the removal of stressors such as heavy metals. Our findings indicate a marked shift in the cellular functions of enriched promoters during and after heavy metal induction. We quantified a transition from the activation of promoters associated with stress defense mechanisms and detoxification processes to the activation of promoters involved in ribosome biosynthesis, tRNA synthesis, mRNA processing and decay, amino acid biosynthesis, and replication.

While the library of 1805 *E. coli* promoters employed in this study represents the most comprehensive collection currently available and allows for substantial insights into gene expression dynamics, we acknowledge that it encompasses less than half of the total known promoters in *E. coli*. This limitation signifies that our analysis might not fully capture all the potential regulatory mechanisms at play, and we advocate for future studies to incorporate a more expansive promoter library to offer a more encompassing view of *E. coli* gene expression.

Using concentrations that are several orders of magnitude lower than the minimum inhibitory concentrations (MICs) for *E. coli* led us to observe distinct gene expression patterns compared to those reported in previous studies<sup>66,71,73,74</sup>. We speculate that by utilizing lower heavy metal concentrations, we are better able to reveal the subtle and nuanced responses of bacteria to these stressors, which might otherwise be overshadowed by the more pronounced effects at higher concentrations. This, in turn, contributes to a more comprehensive understanding of bacterial behavior and adaptation mechanisms under more environmentally relevant conditions, expanding our knowledge of how bacteria cope with heavy metal exposure in real-world scenarios. Overall, we believe that combination of Dynamics

promoter activity data with our approach to applying ICA to time-series data will continue to serve as a valuable tool for generating hypotheses on how cells respond to various external stimuli, uncovering fundamental principles of transcriptional regulation.

## Methods

### Data collection

Detailed description of the Dynamics experimental set up can be found in the original study<sup>6</sup>. In brief, data was obtained from fluorescence values extraction from flat-field corrected images gathered using a custom optical set up. The first step of the experimental set up is arraying the cells using a Singer ROTOR robot so that they could be spotted onto the microfluidic device before glass bonding. Once the device was ready, it was set up inside a custom box kept at 37 °C for imaging. For media flow, the inlet and outlet were connected to 140 mL syringes. The concentrations of the heavy metals tested in this study can be found in Fig. S51.

### Data processing

Detailed information on how the raw data was processed can be found in the original study<sup>6</sup>. In summary, the data was first processed by subtracting the local background signal and then dividing the result by the background signal again in order to create a measure of the amplification of the signal over the background. Specifically, fluorescence values measured at a location outside the cell trap were subtracted from those obtained within the cell-containing regions. This method effectively removes signal noise not associated with the cells, thereby providing a more reliable representation of the cellular fluorescence signals. Then, the data were passed through a median filter (`scipy.signal.medfilt`, `kernel_size=11`) and normalized by subtracting and dividing the average expression values of the promoterless strains. The code used to process data from the original study is available on GitHub. The original data files post-processing for each heavy metal are also available on GitHub. The data were further processed by calculating the log base 2 of the fold change of the ratio of six time points (spaced 40 min apart) with respect to the start of the induction window for the data plotted in Figs. 2 and 3. On the other hand, for the dataset used to produce the results shown in Fig. 4 the data was converted in the log base 2 of the fold change of the ratio between 20 points (spaced 40 minutes apart) with respect to the end of the induction window. All python scripts used to produce these datasets are available in the "ICA\_dynamics" GitHub repository. The final dataset which was fed to the ICA algorithm consisted of a matrix where the rows represented all the different promoters (1805) and the columns represented the different conditions (i.e. the type of metal and the specific time point). Therefore, for the analysis in Figs. 1 and 2 the final dataset had a total of 36 columns representing 6 heavy metal inductions, each with 6 time points. Similarly, the dataset behind the results of Fig. 4 had dimensions of 1805 rows and 100 columns representing 6 heavy metal inductions, each with 20 points.

### Independent component analysis

We used the pipeline for ICA implementation that has been described in previous studies<sup>14,22</sup>. In brief, we run the *Scikit-learn*70 (v0.19.2) implementation of the FastICA algorithm 100 times with random seeds, a convergence tolerance of  $10^{-6}$ . The number of components in each iteration was set to the number of components that reconstruct 99% of the variance as calculated by principal component analysis. The resulting source components (M) from all runs were clustered using the Scikit-learn implementation of the DBSCAN algorithm which does not require predetermination of the number of clusters. In our DBSCAN analysis, we used the following parameters: DISTANCE parameter, analogous to the epsilon parameter in traditional DBSCAN applications, was set to its default value of 0.1, determining the maximum distance between two points to be considered as in the same

cluster. Meanwhile, the MIN\_FRAC parameter, which specifies the minimum fraction of total data points required to form a valid cluster, was maintained at its default setting of 0.5. This implies that each cluster in our analysis contains at least 50% of the total data points present in the dataset. The final independent components were defined as the centroid of each cluster in M, and the weightings were defined as the centroid of their corresponding weighting vectors in A. To ensure that the final components were consistent across multiple runs, we computed the clustered components 100 times, and selected the components that were identified in every run. The previously published code used to compute robust independent components is publicly available at [github.com/SBRG/precise-db](https://github.com/SBRG/precise-db). We also added the scripts specifically used in this paper in the “precise\_db” folder within the “ICA\_dynamics” GitHub repository.

### Determination of the gene coefficient threshold

The dataset M contains all the genes coefficients associated with each iModulon. Most of these coefficients have values close to zero which indicates they are not significantly enriched for that iModulon. In order to extract the genes that belong to each iModulon we computed the D’Agostino  $K^2$  test statistic which is a measure of the skew and kurtosis of a sample distribution. Genes with the largest absolute value were iteratively removed and the D’Agostino  $K^2$  test statistic was computed for the resulting distribution<sup>14</sup>. The statistic cutoff was kept fixed at 800 for the entire analysis. We only considered positively correlated genes in this analysis. The code to calculate the threshold was originally developed by the Palsson lab at UCSD and can be found at the GitHub repository [pymodulon](https://github.com/pymodulon). We also included the scripts used specifically for this paper in the [ICA\\_dynamics](https://github.com/ICA_dynamics) repository.

### Identification of primary heavy metal associated to each iModulon

We used a simple computational strategy that involved the analysis of the activation matrix data of each iModulon to associate them with the heavy metal that induced the highest aggregate expression over the six time points within the corresponding induction window. This approach allowed us to pinpoint the condition under which each iModulon exhibited the most pronounced change, thereby facilitating a data-driven classification. The script used to implement this classification can be found in the [ICA\\_dynamics](https://github.com/ICA_dynamics) GitHub repository.

### Classification of iModulons based on patterns in activity coefficients

In our study, we employed a simple computational method to classify the iModulons based on the shape of their activity coefficients within the induction window of heavy metal enrichment. This method involved calculating the center of mass (CM) and identifying the highest peak (HP) within the specified window for each iModulon, wherein the center of mass represents the mean position of a given function. The iModulons were categorized into three groups based on these calculations: Steady ( $CM \leq 5$  and  $HP \geq 4$ ), Intermediate ( $CM \leq 4$  and  $HP \leq 4$ ), and Fast ( $CM \leq 4$  and  $HP \leq 3$ ). iModulons that did not conform to these parameters were labeled as “other,” with plans for further characterization on a case-by-case basis. We have made the corresponding code block available in the “iModulon\_identification” script within our GitHub repository, and illustrated the calculated values and resultant classifications in Supplementary Fig. S54. It’s noteworthy that the methodology applied is tailored to the specific shapes encountered in our study. For future studies, especially where the shape of activity coefficients is unknown or difficult to categorize, we advocate for exploring more generalized approaches. Machine learning methodologies, particularly those centered around shape recognition, could offer a robust and adaptable solution for classifying transcriptional profiles regardless of the inherent patterns within the data.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

All data is available in the paper, Supplementary Materials and on [Figshare](https://www.figshare.com).

### Code availability

The code used is available on [GitHub](https://github.com) with identifier *ICA\_dynamics*. Citation: Miano, A. (2023). High-Resolution Temporal Profiling of E. coli Transcriptional Response. [Zenodo](https://zenodo.org/record/7811111). We also link the other repositories which were used for different parts of data processing: •[precise-db—https://github.com/SBRG/precise-db](https://github.com/SBRG/precise-db) •[pymodulon—https://github.com/SBRG/pymodulon](https://github.com/SBRG/pymodulon) •[dynamics\\_public—https://github.com/GarrettCGraham/dynamics\\_public](https://github.com/GarrettCGraham/dynamics_public).

### References

- Joyce, A. R. & Palsson, B. Ø. The model organism as a system: integrating ‘omics’ data sets. *Nat. Rev. Mol. Cell Biol.* **7**, 198–210 (2006).
- O’Donnell, S. T., Ross, R. P. & Stanton, C. The progress of multi-omics technologies: determining function in lactic acid bacteria using a systems level approach. *Front. Microbiol.* **10**, 3084 (2020).
- Wang, Z., Gerstein, M. & Snyder, M. Rna-seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **10**, 57–63 (2009).
- Miller, C. et al. Dynamic transcriptome analysis measures rates of mrna synthesis and decay in yeast. *Mol. Syst. Biol.* **7**, 458 (2011).
- Jäger, D., Förstner, K. U., Sharma, C. M., Santangelo, T. J. & Reeve, J. N. Primary transcriptome map of the hyperthermophilic archaeon *thermococcus kodakarensis*. *BMC Genomics* **15**, 1–15 (2014).
- Graham, G. et al. Genome-scale transcriptional dynamics and environmental biosensing. *Proc. Natl Acad. Sci. USA* **117**, 3301–3306 (2020).
- Zaslaver, A. et al. A comprehensive library of fluorescent transcriptional reporters for escherichia coli. *Nat. Methods* **3**, 623–628 (2006).
- Mustafa, G. & Komatsu, S. Toxicity of heavy metals and metal-containing nanoparticles on plants. *Biochim. Biophys. Acta* **1864**, 932–944 (2016).
- Bisht, J. & Harsh, N. S. K. Utilizing *Aspergillus niger* for bioremediation of tannery effluent. *Oct. Jour. Env. Res.* **2**, 77–81 (2014).
- Kapahi, M. & Sachdeva, S. Bioremediation options for heavy metal pollution. *J. Health Pollut.* **9**, 191203 (2019).
- Yamada, R., Okada, D., Wang, J., Basak, T. & Koyama, S. Interpretation of omics data analyses. *J. Hum. Genet.* **66**, 93–102 (2021).
- Comon, P. Independent component analysis, a new concept? *Signal Process.* **36**, 287–314 (1994).
- Rychel, K., Sastry, A. V. & Palsson, B. O. Machine learning uncovers independently regulated modules in the *Bacillus subtilis* transcriptome. *Nat. Commun.* **11**, 1–10 (2020).
- Sastry, A. V. et al. The *Escherichia coli* transcriptome mostly consists of independently regulated modules. *Nat. Commun.* **10**, 1–14 (2019).
- Poudeh, S. et al. Revealing 29 sets of independently modulated genes in *Staphylococcus aureus*, their regulators, and role in key physiological response. *Proc. Natl Acad. Sci. USA* **117**, 17228–17239 (2020).
- Lamoureux, C. R. et al. A multi-scale expression and regulation knowledge base for *Escherichia coli*. *Nucleic Acids Res.* **51**, 10176–10193 (2023).
- Yoo, R. et al. Machine learning of all *Mycobacterium tuberculosis* H37Rv RNA-seq data reveals a structured interplay between metabolism, stress response, and infection. *MSphere* **7**, e00033-22 (2022).



18. Saelens, W., Cannoodt, R. & Saeys, Y. A comprehensive evaluation of module detection methods for gene expression data. *Nat. Commun.* **9**, 1–12 (2018).
19. Sastry, A. V. et al. Independent component analysis recovers consistent regulatory signals from disparate datasets. *PLoS Comput. Biol.* **17**, e1008647 (2021).
20. Akhavan Sepahy, A., Sharifian, S., Zolfaghari, M., Khalily Dermany, M. & Rashedi, H. Study on heavy metal resistant fecal coliforms isolated from industrial, urban wastewater in Arak, Iran. *Int. J. Environ. Res.* **9**, 1217–1224 (2015).
21. Sastry, A. V. et al. Machine learning of bacterial transcriptomes reveals responses underlying differential antibiotic susceptibility. *Msphere* **6**, e00443–21 (2021).
22. Sastry, A. V. et al. Mining all publicly available expression data to compute dynamic microbial transcriptional regulatory networks. Preprint at *bioRxiv* <https://doi.org/10.1101/2021.07.01.450581> (2021).
23. Nies, D. H. & Silver, S. *Molecular Microbiology of Heavy Metals*, vol. 6 (Springer Science & Business Media, 2007).
24. Xue, Y., Osborn, J., Panchal, A. & Mellies, J. L. The rpoE stress response pathway mediates reduction of the virulence of enteropathogenic *Escherichia coli* by zinc. *Appl. Environ. Microbiol.* **81**, 3766–3774 (2015).
25. Mitchell, A. M. & Silhavy, T. J. Envelope stress responses: balancing damage repair and toxicity. *Nat. Rev. Microbiol.* **17**, 417–428 (2019).
26. Babai, R. & Ron, E. Z. An *Escherichia coli* gene responsive to heavy metals. *FEMS Microbiol. Lett.* **167**, 107–111 (1998).
27. Gennaris, A. et al. Repairing oxidized proteins in the bacterial envelope using respiratory chain electrons. *Nature* **528**, 409–412 (2015).
28. Kanai, T., Takahashi, K. & Inoue, H. Three distinct-type glutathione S-transferases from *Escherichia coli* important for defense against oxidative stress. *J. Biochem.* **140**, 703–711 (2006).
29. Reed, K. & Cronan Jr, J. Lipoic acid metabolism in *Escherichia coli*: zc sequencing and functional characterization of the lipa and lipb genes. *J. Bacteriol.* **175**, 1325–1336 (1993).
30. Xie, H. et al. Purification and properties of the *Escherichia coli* nucleoside transporter nugg, a paradigm for a major facilitator transporter sub-family. *Mol. Membr. Biol.* **21**, 323–336 (2004).
31. Jozefczuk, S. et al. Metabolomic and transcriptomic stress response of *Escherichia coli*. *Mol. Syst. Biol.* **6**, 364 (2010).
32. Beard, S. J., Hashim, R., Membrillo-Hernández, J., Hughes, M. N. & Poole, R. K. Zinc (ii) tolerance in *Escherichia coli* k-12: evidence that the znta gene (o732) encodes a cation transport atpase. *Mol. Microbiol.* **25**, 883–891 (1997).
33. Wrage, N., Velthof, G. L., Van Beusichem, M. L. & Oenema, O. Role of nitrifier denitrification in the production of nitrous oxide. *Soil Biol. Biochem.* **33**, 1723–1732 (2001).
34. Zaslaver, A. et al. Just-in-time transcription program in metabolic pathways. *Nat. Genet.* **36**, 486–491 (2004).
35. Tiedje, J. M. Ecology of Denitrification and Dissimilatory Nitrate Reduction to Ammonium. *Biology of Anaerobic Microorganisms* **717**, 179–244 (1988).
36. Xie, N. et al. Nad<sup>+</sup> metabolism: pathophysiologic mechanisms and therapeutic potential. *Signal. Transduct. Target. Ther.* **5**, 1–37 (2020).
37. Tkachenko, A. & Nesterova, L. Y. Polyamines as modulators of gene expression under oxidative stress in *Escherichia coli*. *Biochemistry* **68**, 850–856 (2003).
38. Eisenbeis, S. J. & Parker, J. The nucleotide sequence of the promoter region of hisS, the structural gene for histidyl-trna synthetase. *Gene* **18**, 107–114 (1982).
39. Foglino, M., Gharbi, S. & Lazdunski, A. Nucleotide sequence of the pepn gene encoding aminopeptidase n of *Escherichia coli*. *Gene* **49**, 303–309 (1986).
40. Geslin, C., Llanos, J., Prieur, D. & Jeanthon, C. The manganese and iron superoxide dismutases protect *Escherichia coli* from heavy metal toxicity. *Res. Microbiol.* **152**, 901–905 (2001).
41. Vollmer, W., von Rechenberg, M. & Holtje, J.-V. Demonstration of molecular interactions between the murein polymerase pbp1b, the lytic transglycosylase mlta, and the scaffolding protein mipa of *Escherichia coli*. *J. Biol. Chem.* **274**, 6726–6734 (1999).
42. Hagan, C. L., Kim, S. & Kahne, D. Reconstitution of outer membrane protein assembly from purified components. *Science* **328**, 890–892 (2010).
43. Mortensen, L., Dandanell, G. & Hammer, K. Purification and characterization of the deor repressor of *Escherichia coli*. *EMBO J.* **8**, 325–331 (1989).
44. Lee, C. & Marians, K. J. Characterization of the nucleoid-associated protein YejK. *J. Biol. Chem.* **288**, 31503–31516 (2013).
45. Kosinski, J., Feder, M. & Bujnicki, J. M. The pd-(d/e) xk superfamily revisited: identification of new members among proteins involved in dna metabolism and functional predictions for domains of (hitherto) unknown function. *BMC Bioinforma.* **6**, 1–13 (2005).
46. Agrawal, D. K. & Wanner, B. L. A phoA structural gene mutation that conditionally affects formation of the enzyme bacterial alkaline phosphatase. *J. Bacteriol.* **172**, 3180–3190 (1990).
47. Djoko, K. Y., Chong, L. X., Wedd, A. G. & Xiao, Z. Reaction mechanisms of the multicopper oxidase cueO from *Escherichia coli* support its functional role as a cuprous oxidase. *J. Am. Chem. Soc.* **132**, 2005–2015 (2010).
48. Monje-Casas, F., Jurado, J., Prieto-Álamo, M.-J., Holmgren, A. & Pueyo, C. Expression analysis of the nrdHIEF operon from *Escherichia coli*: conditions that trigger the transcript level in vivo. *J. Biol. Chem.* **276**, 18031–18037 (2001).
49. Justino, M. C., Almeida, C. C., Teixeira, M. & Saraiva, L. M. *Escherichia coli* di-iron ytfE protein is necessary for the repair of stress-damaged iron-sulfur clusters. *J. Biol. Chem.* **282**, 10352–10359 (2007).
50. Raha, M., Kawagishi, I., Müller, V., Kihara, M. & Macnab, R. *Escherichia coli* produces a cytoplasmic alpha-amylase, amya. *J. Bacteriol.* **174**, 6644–6652 (1992).
51. Saadat, D. & Harrison, D. H. Identification of catalytic bases in the active site of *Escherichia coli* methylglyoxal synthase: cloning, expression, and functional characterization of conserved aspartic acid residues. *Biochemistry* **37**, 10074–10086 (1998).
52. Yang, C. et al. Genome-wide phob binding and gene expression profiles reveal the hierarchical gene regulatory network of phosphate starvation in *Escherichia coli*. *PLoS ONE* (2012).
53. Panne, D., Müller, S. A., Wirtz, S., Engel, A. & Bickle, T. A. The McrBC restriction endonuclease assembles into a ring structure in the presence of G nucleotides. *EMBO J.* **20**, 3210–3217 (2001).
54. Hochhut, B. et al. Role of pathogenicity island-associated integrases in the genome plasticity of uropathogenic *Escherichia coli* strain 536. *Mol. Microbiol.* **61**, 584–595 (2006).
55. Lee, S. K., Newman, J. D. & Keasling, J. D. Catabolite repression of the propionate catabolic genes in *Escherichia coli* and salmonella enterica: evidence for involvement of the cyclic amp receptor protein. *J. Bacteriol.* **187**, 2793–2800 (2005).
56. Mutalik, V. K. et al. High-throughput mapping of the phage resistance landscape in *E. coli*. *PLoS Biol.* **18**, e3000877 (2020).
57. Taschner, N. P., Yagil, E. & Spira, B. A differential effect of  $\sigma$  on the expression of the pho regulon genes of *Escherichia coli*. *Microbiology* **150**, 2985–2992 (2004).
58. Schweizer, H., Grussenmeyer, T. & Boos, W. Mapping of two ugp genes coding for the pho regulon-dependent sn-glycerol-3-phosphate transport system of *Escherichia coli*. *J. Bacteriol.* **150**, 1164–1171 (1982).
59. Beaufay, F. et al. Polyphosphate functions in vivo as an iron chelator and fenton reaction inhibitor. *Mbio* **11**, e01017–20 (2020).

60. Santos-Beneit, F. The pho regulon: a huge regulatory network in bacteria. *Front. Microbiol.* **6**, 402 (2015).
61. Yeung, N. et al. The *E. coli* monothiol glutaredoxin Grxd forms homodimeric and heterodimeric FeS cluster containing complexes. *Biochemistry* **50**, 8957–8969 (2011).
62. Li, J. et al. Zinc toxicity and iron-sulfur cluster biogenesis in *Escherichia coli*. *Appl. Environ. Microbiol.* **85**, e01967–18 (2019).
63. Egan, S. E. et al. Molecular characterization of the entner-doudoroff pathway in *Escherichia coli*: sequence analysis and localization of promoters for the *edd-eda* operon. *J. Bacteriol.* **174**, 4638–4646 (1992).
64. Panina, E. M., Mironov, A. A. & Gelfand, M. S. Comparative genomics of bacterial zinc regulons: enhanced ion transport, pathogenesis, and rearrangement of ribosomal proteins. *Proc. Natl Acad. Sci. USA* **100**, 9912–9917 (2003).
65. Graham, A. I. et al. Severe zinc depletion of *Escherichia coli*: roles for high affinity zinc binding by Zint, zinc transport and zinc-independent proteins. *J. Biol. Chem.* **284**, 18377–18389 (2009).
66. Wang, A. & Crowley, D. E. Global gene expression responses to cadmium toxicity in *Escherichia coli*. *J. Bacteriol.* **187**, 3259–3266 (2005).
67. Zhang, P. et al. *OmpW* is positively regulated by iron via *fur*, and negatively regulated by *soxS* contribution to oxidative stress resistance in *Escherichia coli*. *Microb. Pathogenesis* **138**, 103808 (2020).
68. McConn, J. L., Lamoureux, C. R., Poudel, S., Palsson, B. O. & Sastry, A. V. Optimal dimensionality selection for independent component analysis of transcriptomic data. *BMC Bioinforma.* **22**, 1–13 (2021).
69. Tan, J. et al. Independent component analysis of *E. coli*'s transcriptome reveals the cellular processes that respond to heterologous gene expression. *Metab. Eng.* **61**, 360–368 (2020).
70. Saha, R., Chowdhury, A. & Maranas, C. D. Recent advances in the reconstruction of metabolic models and integration of omics data. *Curr. Opin. Biotechnol.* **29**, 39–45 (2014).
71. Sütterlin, S. et al. Heavy metal susceptibility of *Escherichia coli* isolated from urine samples from Sweden, Germany, and Spain. *Antimicrob. Agents Chemother.* **62**, e00209–18 (2018).
72. Kulakovskaya, T. Inorganic polyphosphates and heavy metal resistance in microorganisms. *World J. Microbiol. Biotechnol.* **34**, 1–8 (2018).
73. Kershaw, C. J., Brown, N. L., Constantinidou, C., Patel, M. D. & Hobman, J. L. The expression profile of *Escherichia coli* K-12 in response to minimal, optimal and excess copper concentrations. *Microbiology* **151**, 1187–1198 (2005).
74. Lee, L. J., Barrett, J. A. & Poole, R. K. Genome-wide transcriptional response of chemostat-cultured *Escherichia coli* to zinc. *J. Bacteriol.* **187**, 1124–1134 (2005).

## Acknowledgements

We would like to acknowledge Garrett Graham and Nicholas Csicsery for their help and mentorship in the initial stages of this project.

Additionally, we would like to thank Elizabeth Stasiowski and Gregoire Thouvenin for kindly sharing the data used in this project and Professor Terence Hwa for his helpful comments and advice on the project. This work was supported by the National Institute of General Medical Sciences of the National Institutes of Health grant No. R01GM069811 (J.H.).

## Author contributions

Conceptualization: A.M. and J.H. Methodology: A.M., K.R., A.S., B.P., and J.H. Analysis: A.M., K.R., and A.S. Writing—original draft: A.M. Writing—review & editing: A.M., A.L., and K.R.

## Competing interests

J.H. is a co-founder of GenCirq Inc., which focus on cancer therapeutics. He is on the Board of Directors and has equity in GenCirq. His spouse is employed part time for bookkeeping and to support employees with Human Resources. The remaining authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-023-43173-7>.

**Correspondence** and requests for materials should be addressed to Arianna Miano.

**Peer review information** *Nature Communications* thanks Julio Freyre-González and Alexandra Lee for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

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

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

© The Author(s) 2023