

# The transcription factor network of *E. coli* steers global responses to shifts in RNAP concentration

Bilena L.B. Almeida<sup>1,\*</sup>, Mohamed N. M. Bahrudeen<sup>1,†</sup>, Vatsala Chauhan<sup>1,†</sup>,  
Suchintak Dash<sup>1,†</sup>, Vinodh Kandavalli<sup>2</sup>, Antti Häkkinen<sup>3</sup>, Jason Lloyd-Price<sup>4</sup>,  
Palma S.D. Cristina<sup>1</sup>, Ines S.C. Baptista<sup>1</sup>, Abhishekh Gupta<sup>5</sup>, Juha Kesseli<sup>6</sup>, Eric Dufour<sup>7</sup>,  
Olli-Pekka Smolander<sup>8,9</sup>, Matti Nykter<sup>6</sup>, Petri Auvinen<sup>9</sup>, Howard T. Jacobs<sup>10</sup>,  
Samuel M.D. Oliveira<sup>11</sup> and Andre S. Ribeiro<sup>1,12,\*</sup>

<sup>1</sup>Laboratory of Biosystem Dynamics, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland, <sup>2</sup>Department of Cell and Molecular Biology, Uppsala University, Uppsala, Sweden, <sup>3</sup>Research Program in Systems Oncology, Research Programs Unit, Faculty of Medicine, University of Helsinki, FI-00014 Helsinki, Finland, <sup>4</sup>Google LLC, 111 8th Ave, New York, NY 10010, USA, <sup>5</sup>Center for Quantitative Medicine and Department of Cell Biology, University of Connecticut School of Medicine, 263 Farmington Av., Farmington, CT 06030-6033, USA, <sup>6</sup>Prostate Cancer Research Center, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland; Tays Cancer Center, Tampere University Hospital, Tampere, Finland, <sup>7</sup>Mitochondrial bioenergetics and metabolism, BioMediTech, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland, <sup>8</sup>Department of Chemistry and Biotechnology, Tallinn University of Technology, Tallinn, Estonia, <sup>9</sup>Institute of Biotechnology, University of Helsinki, Viikinkaari 5D, 00790 Helsinki, Finland, <sup>10</sup>Faculty of Medicine and Health Technology, FI-33014 Tampere University, Finland; Department of Environment and Genetics, La Trobe University, Melbourne, Victoria 3086, Australia, <sup>11</sup>Department of Electrical and Computer Engineering, Boston University, Boston, MA, USA and <sup>12</sup>Center of Technology and Systems (CTS-Uninova), NOVA University of Lisbon, 2829-516 Monte de Caparica, Portugal

Received March 07, 2022; Revised June 02, 2022; Editorial Decision June 06, 2022; Accepted June 14, 2022

## ABSTRACT

The robustness and sensitivity of gene networks to environmental changes is critical for cell survival. How gene networks produce specific, chronologically ordered responses to genome-wide perturbations, while robustly maintaining homeostasis, remains an open question. We analysed if short- and mid-term genome-wide responses to shifts in RNA polymerase (RNAP) concentration are influenced by the *known* topology and logic of the transcription factor network (TFN) of *Escherichia coli*. We found that, at the gene cohort level, the magnitude of the single-gene, mid-term transcriptional responses to changes in RNAP concentration can be explained by the absolute difference between the gene's numbers of activating and repressing input transcription factors (TFs). Interestingly, this difference is strongly positively correlated with the number of input TFs of the gene. Meanwhile, short-term responses showed

only weak influence from the TFN. Our results suggest that the global topological traits of the TFN of *E. coli* shape which gene cohorts respond to genome-wide stresses.

## INTRODUCTION

Gene regulatory networks (GRNs) receive, process, act upon and send out information, while being robust to random fluctuations. How signals targeting one to a few genes are processed is relatively well understood (1,2). Meanwhile, many cellular environments fluctuate (sometimes unpredictably) in nutrient availability, pH, temperature, salts, community of other cells or species they live with, etc., which may cause genome-wide stresses. We investigate how GRNs produce chronologically ordered responses to genome-wide perturbations, while robustly maintaining homeostasis.

Evidence suggests that genome-wide stresses initially perturb hundreds to thousands of genes (3) but are quickly

\*To whom correspondence should be addressed. Tel: +358 2945211; Email: andre.sanchesribeiro@tuni.fi

Correspondence may also be addressed to Bilena L.B. Almeida. Tel: +358 2945211; Email: bilena.limadebritoalmeida@tuni.fi

†The authors wish it to be known that, in their opinion, the second, third and fourth authors should be regarded as Joint Second Authors.

processed. As a result, after a transient period, only specific gene cohorts of tens to a few hundred genes (4,5) (usually sharing common feature(s)) participate in the responsive short-, mid- and long-term transcriptional programs (6). For example, when *Escherichia coli* suffers a cold shock, a specific cohort exhibits a fast, short-term response ( $\sim 70$  genes), while another has a longer-term response ( $\sim 35$  genes). Other genes remain relatively passive (7,8). Since cells exhibit predictable, temporally ordered, beneficial phenotypic changes, these response programs have likely been positively selected during evolution.

It has been shown that global regulators (GRs) (9–12), DNA supercoiling (13) and small RNAs (sRNAs) (14), among other, can select large cohorts of stress-specific, responsive genes. It was also reported that 60–90% of *E. coli* genes respond to changing growth conditions following a constant global scaling factor (15). Further, measurements using fluorescent reporters and small circuits (16,17) showed that the effects of RNA polymerase (RNAP), and other GRs, can be separate from the effects of input transcription factors (TFs) during genome-wide responses. Nevertheless, establishing whether and how the global topology and logic of transcription factor networks (TFNs) affect genome-wide responses remains challenging (18), despite successes in establishing that gene-gene interactions generate gene-gene correlated dynamics (19–22).

To investigate the influence of the topology and logic of TFNs on large transcriptional programs, we study what occurs following genome-wide perturbations. We consider that, in *E. coli*, the concentration of the key genome-wide regulator, the RNAP, naturally differs with medium composition (23). Also, it is well established how transcription kinetics differs according to RNAP concentration (24,25). We expect that these changes can have genome-wide effects. We thus obtain increasingly dilute media to alter systematically and rapidly (26,27) the abundance of RNAP (illustrated in Figure 1A), and measure the genome-wide, short- and mid-term changes in transcript abundances.

Decreases in RNAP abundance should, in the short-term, cause quick genome-wide decreases in transcription rates, and thus in RNA abundances (Figure 1B<sub>1</sub>, Supplementary Results section *Expected effects of shifting RNA polymerase concentration on a gene's transcription dynamics*). Such shifts, likely diverse in magnitudes, should then cause downshifts in the corresponding protein abundances. Thus, in the case of input TFs, their 'output' genes will, later on, be affected as well (Figure 1B<sub>2</sub>), causing further (here named 'mid-term') changes in RNA abundances (Figure 1B<sub>4</sub>). Meanwhile, the short-term changes most likely are only affected by the genes' individual features affecting their responsiveness to RNAP, since the protein abundances have not yet changed significantly.

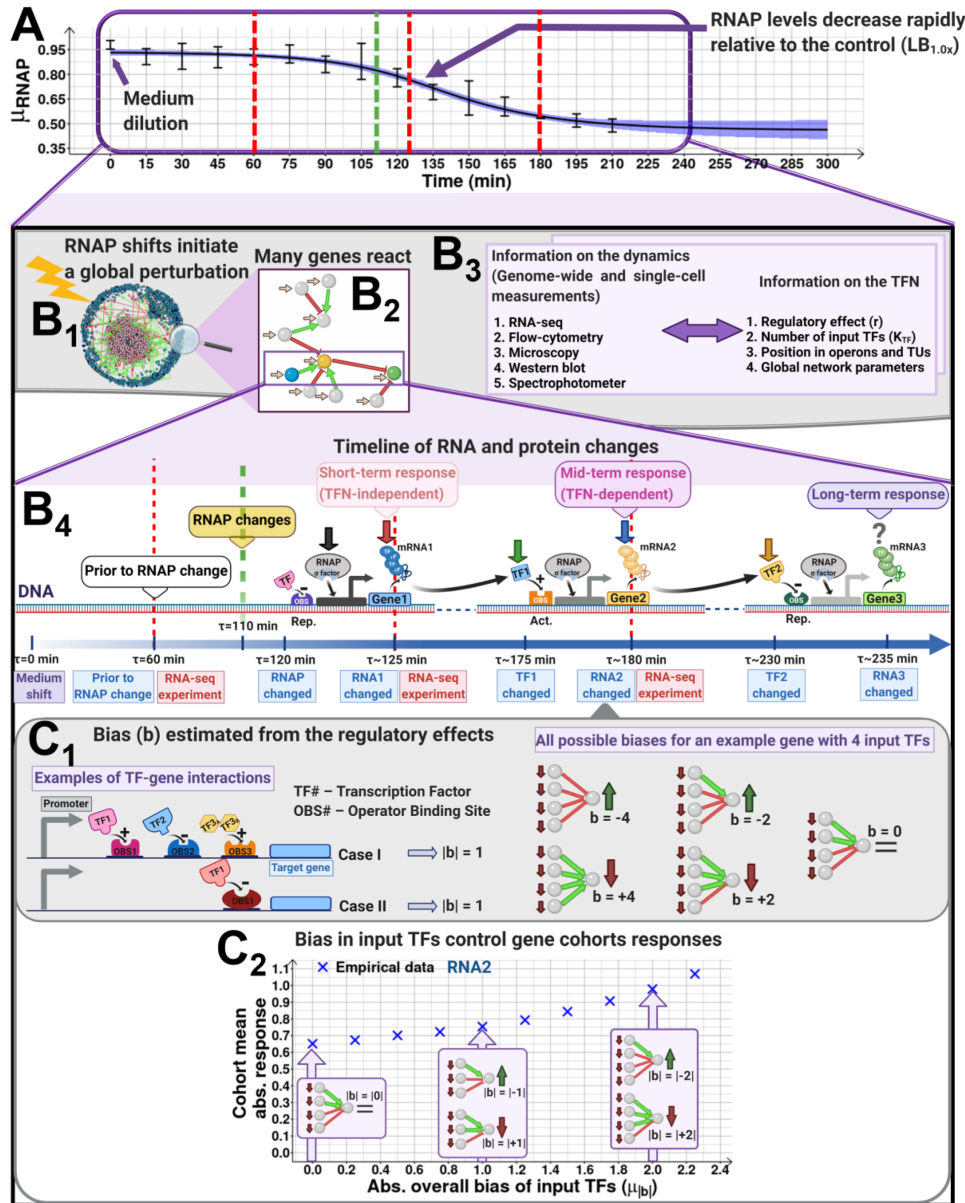
We focus on the mid-term changes in single-gene expression levels. Specifically, we hypothesize that, in the short-term, a reduction in RNAP concentration will be followed by a reduction in most genes' expression levels. Subsequently, in the mid-term, genes with one input TF will have their RNA abundance either further decreased or, instead, increased, depending on whether their input TF is an ac-

tivator or a repressor, respectively. Meanwhile, the average magnitude of the mid-term response of genes with multiple input TFs should be correlated with the difference between the numbers of their activator and repressor input TFs. This difference is here named 'bias' in the regulatory effect (i.e. activation or repression) of the input TFs of a gene. A schematic and a predictive model of this regulatory mechanism of the genome-wide single-gene mid-term responses are shown in Figure 1B and C, respectively.

At the single-gene level, the magnitude of the mid-term changes in RNA abundances should be influenced by the magnitude of the shift in RNAP and in input TFs concentrations, as well as by the specific features of each gene and its input TFs (bindings affinities, etc.). However, many of these features are largely unknown. As such, here we only study empirically if the average responses of gene cohorts can be explained by the mean bias in the regulatory effect of their input TFs, along with the magnitude of the shift in RNAP concentration (Figure 1C<sub>2</sub>). In detail, we interpret the data on the genome-wide kinetics based on the information on the TFN structure (logic and topology) (Figure 1B<sub>3</sub>).

We use *E. coli* to validate this hypothesis since its gene expression mechanisms have been largely dissected and the kinetics of transcription, translation, and RNA and protein degradation are well known (26,32,33). Also, its TFN is extensively mapped, with RegulonDB (34) informing on  $\sim 4700$  TF interactions between  $\sim 4500$  genes (and on their activating or repressing regulatory roles). Consequently, since we know the regulatory network *a priori*, instead of using the data on gene expression for network inference, we use it solely to quantify the genes' responsiveness with respect to the TFN topology and logic. We then investigate whether the mid-term responsiveness to shifting RNAP concentration is in accordance with the *presently known* topology and logic of the TFN, as hypothesized (Figure 1B<sub>1</sub>–B<sub>4</sub>). Supplementary Table S1 has a description of the variables used throughout the manuscript.

In summary, here we show that changes in RNAP concentration due to medium dilutions are followed by short- and mid-term genome-wide changes in RNA abundances (obtained by time-lapsed RNA-seq). These RNA changes, globally, cannot be explained by potential influences from GRs,  $\sigma$  factors, (p)ppGpp, non-coding RNAs, or post-translation regulators (e.g. due to lack of correlation with their output genes, lack of RNA changes, etc.). Instead, we find that genes directly linked by TF interactions show correlated changes in RNA abundances. Further, the average magnitude of the mid-term responses of gene cohorts can be explained by the mean bias in the regulatory effect of their genes' input TFs (obtained from RegulonDB and shown to be correlated to the number of input TFs,  $K_{TF}$ ). Also influential is the magnitude of the shift in RNAP concentration (obtained by flow-cytometry and western blot) and operons and transcription units (TUs) organization (from RegulonDB). Meanwhile, short-term responses are less influenced by TFs. Finally, we show the same phenomenon for opposite shifts in medium concentration that cause the same shifts in RNAP concentration.



**Figure 1.** Expected short- and mid-term effects of quick downshifts of the RNAP abundance on the TFN of *E. coli*. (A) Example changes in mean RNAP ( $\mu_{RNAP}$ ) and 68% CB (shadow) relative to control ( $LB_{1.0x}$ ) after diluting the medium ( $LB_{0.5x}$ ). Vertical red lines mark when the transcriptome measurements at 60, 125 and 180 min. Given the RNAP levels and the kinetics of RNA and protein abundances, these moments are named ‘prior to RNAP changes’ and ‘short-’, and ‘mid-term’ changes in RNA abundances. (B<sub>1</sub>) Known TF-gene interactions (red and green lines, if repressing and activating, respectively) and genes with (pink) and without (blue) input TFs of *E. coli*. (B<sub>2</sub>) Schematics of the expected effects of a local topology of activating (green) and repressing (red) input TFs on mid-term responses. Genes (balls) are coloured (blue, yellow, and green) according to the events in B<sub>4</sub>. (B<sub>3</sub>) Data collected on the genome-wide kinetics as well as data collected on the TFN structure. (B<sub>4</sub>) Following a medium dilution, intracellular RNAP concentrations (black arrow) decrease after a time lag, and RNA abundances (red arrow) will decrease accordingly. Compared to when at  $\sim 0$  min, the RNAP at  $\sim 120$  min and corresponding RNAs at  $\sim 125$  min should be lower (25,28). Given translation times ( $\sim 50$  min (29–31)), at  $\sim 175$  min, the protein abundances, including input TFs, coded by the perturbed RNAs (green arrow) should differ as well. Fluctuations in these input TFs abundances will then propagate to nearest neighbour ‘output’ genes, further shifting their RNA abundances (blue arrow) depending on whether the input TF is an activator or a repressor. Finally, the yellow arrow represents (not measured) long-term changes ( $\sim 230$  min or longer). We performed RNA-seq at  $\sim 60$  min (prior to RNAP changes),  $\sim 125$  min (short-term RNA changes), and  $\sim 180$  min (mid-term RNA changes, affected by input TFs). Finally, the green dashed line marks when the RNAP level already differs significantly from the control (see example Figure 1A). (C<sub>1</sub>) Predictive model of the expected biases in sets of input TFs of individual genes. Considering TF-gene interactions as either repressions (regulatory effect of -1) or activations (regulatory effect of +1), the overall effect of a set of input TFs during these stresses should be predictable from the sum of the regulatory effect of the input TFs, named ‘bias’, ( $b$ ). Regulatory effects obtained from RegulonDB. (C<sub>2</sub>) Example average response ( $\mu_{|b|}$ ) from RNA-seq at 180 min of gene cohorts with a given  $\mu_{|b|}$  and how they are expected to relate to the biases. Figures created with BioRender.com.

## MATERIALS AND METHODS

### Bacterial strains, media, growth conditions and curves, and intracellular concentrations

We used wild type MG1655 cells as a base strain to study the transcriptome. In addition, we used an RL1314 strain with RpoC endogenously tagged with GFP (generously provided by Robert Landick) to measure RNAP levels, and 20 YFP fusion strains with genes endogenously tagged with the YFP coding sequence (25) to measure single-cell protein levels (Supplementary Table S4). Further, we used a strain carrying an *rpoS::mCherry* gene (generously provided by James Locke), shown to track RpoS (35), to measure RpoS levels. In addition, we measured the protein levels of the *spoT* gene, which is one of the genes responsible for (p)ppGpp synthesis (3), using the YFP fusion library. Finally, we measured single-cell levels of the *crl* gene using a low-copy plasmid fusion library of fluorescent (GFP) reporter strain (36).

From glycerol stocks (at  $-80^{\circ}\text{C}$ ), cells were streaked on lysogeny broth (LB) agar plates with antibiotics and kept at  $37^{\circ}\text{C}$  overnight. Next, a single colony was picked, inoculated into fresh LB medium and, kept at  $30^{\circ}\text{C}$  overnight with appropriate antibiotics and aeration at 250 rpm. From overnight cultures (ONC), cells were diluted to 1:1000 in tailored LB media (see below) with antibiotics, incubated at  $37^{\circ}\text{C}$  with aeration, and allowed to grow until reaching an optical density of  $\approx 0.4$  at 600 nm ( $\text{OD}_{600}$ ).

Using this protocol, to attain cells with different intracellular RNAP concentration, starting from LB, we used tailored media, denoted as 'LB<sub>1.0x</sub>', 'LB<sub>0.75x</sub>', 'LB<sub>0.5x</sub>', 'LB<sub>0.25x</sub>', 'LB<sub>1.5x</sub>', 'LB<sub>2.0x</sub>' and 'LB<sub>2.5x</sub>' specifically, as in (26). Their composition for 100 ml (pH of 7.0) are, respectively: (LB<sub>1.0x</sub>) 1 g tryptone, 0.5 g yeast extract and 1 g NaCl; (LB<sub>0.75x</sub>) 0.75 g tryptone, 0.375 g yeast extract and 1 g NaCl; (LB<sub>0.5x</sub>) 0.5 g tryptone, 0.25 g yeast extract and 1 g NaCl; and (LB<sub>0.25x</sub>) 0.25 g tryptone, 0.125 g yeast extract and 1 g NaCl; (LB<sub>1.5x</sub>) 1.5 g tryptone, 0.75 g yeast extract and 1 g NaCl; (LB<sub>2.0x</sub>) 2 g tryptone, 1 g yeast extract and 1 g NaCl; (LB<sub>2.5x</sub>) 2.5 g tryptone, 1.25 g yeast extract and 1 g NaCl.

To measure cell growth curves and rates, ONC of the RL1314 strain were diluted to an initial optical density at 600 nm ( $\text{OD}_{600}$ ) of  $\approx 0.05$  into independent fresh media (LB<sub>1.0x</sub>, LB<sub>0.75x</sub>, LB<sub>0.5x</sub>, LB<sub>0.25x</sub>, LB<sub>1.5x</sub>, LB<sub>2.0x</sub> and LB<sub>2.5x</sub>). The cultures were aliquoted in a 24-well flat bottom transparent plate and incubated at  $37^{\circ}\text{C}$  with continuous shaking in a Biotek Synergy HTX Multi-Mode Reader. Growth was monitored every 10 min for 10 h.

Finally, the RNAP and  $\sigma$  factor concentrations were estimated by measuring their average abundances from single-cell fluorescence levels (of RpoC-GFP and *rpoS::mCherry*, respectively, by flow-cytometry). Next, we divided that abundance by the mean single-cell area (from phase-contrast microscopy images of cell populations in the same condition), used as a proxy for cell volumes, to obtain concentrations in fluorescence intensity per pixel (not shown in the figures). In all cases, for each condition, we obtained images from, on average, 2500 cells (from three biological replicates). The images are provided in Supplementary Data.

### Microscopy

To measure single-cell RNAP levels, ONC RL1314 cells were pre-inoculated into LB<sub>1.0x</sub>, LB<sub>0.75x</sub>, LB<sub>0.5x</sub> and LB<sub>0.25x</sub> media. Upon reaching mid-exponential growth phase, cells were pelleted by quick centrifugation (10000 rpm for 1 min), and the supernatant was discarded. The pellet was re-suspended in 100  $\mu\text{l}$  of the remaining medium. Next, 3  $\mu\text{l}$  of cells were placed in between 2% agarose gel pad and a coverslip and imaged by confocal microscopy with a 100 $\times$  objective (example images in Supplementary Figure S1). GFP fluorescence was measured with a 488 nm laser and a 514/30 nm emission filter. Phase-contrast images were simultaneously acquired. MG1655 cells were imaged to measure cell size in LB<sub>1.0x</sub>, LB<sub>0.75x</sub>, LB<sub>0.5x</sub>, LB<sub>0.25x</sub>, LB<sub>1.5x</sub>, LB<sub>2.0x</sub> and LB<sub>2.5x</sub> media. Finally, MG1655 cells was also imaged in LB<sub>1.0x</sub> during stationary growth. Finally, we imaged cells of the YFP strain library to assess if their morphology and physiology were consistent with healthy cells during measurements.

### Flow-cytometry

We performed flow-cytometry of RL1314 cells to measure single-cell RNAP over time. ONC were diluted at 1:1000 into respective fresh media (LB<sub>1.0x</sub>, LB<sub>0.75x</sub>, LB<sub>0.5x</sub> and LB<sub>0.25x</sub>) and grown as described in Methods section *Bacterial strains, media, growth conditions and curves, and intracellular concentrations*. Flow-cytometry data was recorded every 30 min (three biological replicates), up to 210 min. Data was also captured in the mid-exponential phase (at 180 min), in the media studied (LB<sub>1.0x</sub>, LB<sub>0.75x</sub>, LB<sub>0.5x</sub>, LB<sub>0.25x</sub>, LB<sub>1.5x</sub>, LB<sub>2.0x</sub> and LB<sub>2.5x</sub>), with 3 biological replicates each. We used a similar protocol to perform flow-cytometry of several strains of the YFP library (25) in LB<sub>1.0x</sub> and LB<sub>0.25x</sub> (three biological replicates, Supplementary Table S4), including to measure single-cell SpoT levels in LB<sub>1.0x</sub>, LB<sub>0.75x</sub>, LB<sub>0.5x</sub>, LB<sub>0.25x</sub>, LB<sub>1.5x</sub>, LB<sub>2.0x</sub> and LB<sub>2.5x</sub> (three biological replicates each).

Meanwhile, we measured single-cell levels of the *crl* gene in LB<sub>0.5x</sub> at 0 and 180 min, using a strain from the GFP-promoter fusion library. Further, to measure RpoS levels, we performed flow-cytometry of cells of the MGmCherry strain in LB<sub>1.0x</sub>, LB<sub>0.75x</sub>, LB<sub>0.5x</sub> and LB<sub>0.25x</sub> during the exponential (180 min) and stationary growth phases (LB<sub>1.0x</sub>, 14 h after pre-inoculation). In these measurements, as well as the measurements above, we recorded FSC-H, SSC-H and Width, to be used as proxies for cell size and density (i.e. composition), as they are positively correlated with these features (37).

In addition, data from measurements of MG1655 cells were used to discount background fluorescence from cells of the MGmCherry and the YFP strains. Similarly, measurements of the W3110 strain were used to discount the background fluorescence from the RL1314 strain.

For performing flow-cytometry, 5  $\mu\text{l}$  of cells were diluted in 1 ml of PBS, and vortexed. In each condition, 50000 events were recorded. Prior to the experiments, QC was performed as recommended by the manufacturer. Measurements were conducted using an ACEA NovoCyte Flow Cytometer (ACEA Biosciences Inc., San Diego, USA) equipped with yellow and blue lasers.

For detecting the GFP and YFP signals, we used the FITC channel (-H parameter) with 488 nm excitation, 530/30 nm emission, and 14  $\mu$ l/min sample flow-rate with a core diameter of 7.7  $\mu$ m. PMT voltage was set to 550 for FITC and kept the same for all conditions. Similarly, to detect the mCherry signal, we used PE-Texas Red channel (-H parameter) having an excitation of 561 nm and emission of 615/20 nm and sample flow-rate of 14  $\mu$ l/min, with a core diameter of 7.7  $\mu$ m. PMT voltage was set to 584 for PE-Texas Red and kept the same for all conditions. To remove background signal from particles smaller than bacteria, the detection threshold was set to 5000. All events were collected by Novo Express software from ACEA Biosciences Inc.

### Protein isolation and western blotting

Western blotting was used to quantify relative RNAP levels of MG1655 cells and GFP levels of RL1314 RNAP-GFP cells (Supplementary Figure S2 and Table S2). Briefly, cells were diluted from ONC into respective fresh media and incubated at 37 °C with aeration and grown until reaching an  $OD_{600} \approx 0.4$ . Next, cells were harvested by centrifugation (8000 rpm for 5 min) and pellets were lysed with B-PER bacterial protein extraction reagent, added with a protease inhibitor for 10 min at room temperature (RT). Following lysis, centrifugation was done at 14 000 rpm for 10 min and the supernatant was collected. Next, the supernatant was diluted in 4X Laemmli buffer with  $\beta$ -mercaptoethanol and samples were boiled at 95 °C for 5 min.

Samples with  $\sim 30 \mu$ g of soluble total proteins were loaded on 4–20% TGX stain-free precast gels (Biorad). These proteins were then separated by electrophoresis and transferred on PVDF membrane using TurboBlot (Biorad). Next, membranes were blocked with 5% non-fat milk at room temperature (RT) for 1 h and probed with primary antibodies at 1:2000 dilutions (Biolegend) at 4 °C overnight. The antibodies used for the MG1655 strain were against RpoC ( $\beta$  prime subunit of RNAP), while for the RL1314 strain it was used antibodies against GFP. As a control we also subjected MG1655 cells to antibodies against GFP (Supplementary Figure S2B<sub>2</sub>). HRP-secondary antibody (1:5000) treatment was then done (Sigma Aldrich) for 1 h at RT. Excess antibodies were removed by washing with buffer. The membrane was treated with chemiluminescence reagent (Biorad) for band detection. Images were obtained by the Chemidoc XRS system (Biorad) and band quantification was done using the Image Lab software (v.5.2.1).

### RNA-seq

**Sample preparation.** RNA-seq was performed thrice, for decreasing [LB<sub>0.75x</sub>, LB<sub>0.5x</sub>, and LB<sub>0.25x</sub>, at 180 min; LB<sub>0.5x</sub> at 60 and 125 min] and for increasing (LB<sub>1.5x</sub>, LB<sub>2.0x</sub> and LB<sub>2.5x</sub>, at 180 min) medium richness relative to a control (LB<sub>1.0x</sub>) (an independent control was used for each three sets of conditions). Cells from 3 independent biological replicates of MG1655 in each modified medium were treated with RNA protect bacteria reagent (Qiagen, Germany), to prevent degradation of RNA, and their total RNA was extracted using RNeasy kit (Qiagen). RNA was treated twice

with DNase (Turbo DNA-free kit, Ambion) and quantified using Qubit 2.0 Fluorometer RNA assay (Invitrogen, Carlsbad, CA, USA). Total RNA abundance was determined by gel electrophoresis, using a 1% agarose gel stained with SYBR safe (Invitrogen). RNA was detected using UV with a Chemidoc XRS imager (Biorad).

### Sequencing.

**Part 1: For shifts from LB<sub>1.0x</sub> to LB<sub>0.75x</sub>, LB<sub>0.5x</sub> and LB<sub>0.25x</sub>, at 180 min.** Sequencing was performed by Acobiom (Montpellier, France). The RNA integrity number (RIN) of the samples was obtained with the 2100 Bioanalyzer (Agilent Technologies, Palo Alto, USA) using Eukaryotic Total RNA 6000 Nano Chip (Agilent Technologies). Ribosomal RNA depletion was performed using Ribo-Zero removal kit (Bacteria) from Illumina. RNA-seq libraries were constructed according to the Illumina's protocol. Samples were sequenced using a single-index, 1  $\times$  75 bp single-end configuration ( $\sim 10$ M reads/library) on an Illumina MiSeq instrument. Sequencing analysis and base calling were performed using the Illumina Pipeline. Sequences were obtained after purity filtering.

**Part 2: For shifts from LB<sub>1.0x</sub> to LB<sub>1.5x</sub>, LB<sub>2.0x</sub> and LB<sub>2.5x</sub> at 180 min, and from LB<sub>1.0x</sub> to LB<sub>0.5x</sub> at 60 and 125 min.** Sequencing was performed by GENEWIZ, Inc. (Leipzig, Germany). The RIN of the samples was obtained with the Agilent 4200 TapeStation (Agilent Technologies, Palo Alto, CA, USA). Ribosomal RNA depletion was performed using Ribo-Zero Gold Kit (Bacterial probe) (Illumina, San Diego, CA, USA). RNA-seq libraries were constructed using NEBNext Ultra RNA Library Prep Kit (NEB, Ipswich, MA, USA). Sequencing libraries were multiplexed and clustered on 1 lane of a flow-cell.

For shifts from LB<sub>1.0x</sub> to LB<sub>1.5x</sub>, LB<sub>2.0x</sub> and LB<sub>2.5x</sub> at 180 min, samples were sequenced using a single-index, 2  $\times$  150 bp paired-end (PE) configuration ( $\sim 350$ M raw paired-end reads per lane) on an Illumina HiSeq 4000 instrument. Image analysis and base calling were conducted with HiSeq Control Software (HCS). Raw sequence data (.bcl files) were converted into fastq files and de-multiplexed using Illumina bcl2fastq v.2.20. One mismatch was allowed for index sequence identification.

For shifts from LB<sub>1.0x</sub> to LB<sub>0.5x</sub> at 60 and 125 min, samples were sequenced using a single-index, 2  $\times$  150 bp paired-end (PE) configuration ( $\sim 10$ M raw paired-end reads per lane) on an Illumina NovaSeq 6000 instrument. Image analysis and base calling were conducted with NovaSeq Control Software v1.7. Raw sequence data (.bcl files) were converted into fastq files and de-multiplexed using Illumina bcl2fastq v.2.20. One mismatch was allowed for index sequence identification.

**Data analysis.** Regarding the RNA-seq data analysis pipeline: (i) RNA sequencing reads were trimmed to remove possible adapter sequences and nucleotides with poor quality with Trimmomatic (38) v.0.36 (for data from sequencing part 1) and v.0.39 (for data from sequencing part 2). (ii) Trimmed reads were then mapped to the reference genome, *E. coli* MG1655 (NC\_000913.3), using the

Bowtie2 v.2.3.5.1 aligner, which outputs BAM files (39). (iii) Then, *featureCounts* from the Rsubread R package (v.1.34.7) was used to calculate unique gene hit counts (40). Genes with less than five counts in more than three samples, and genes whose mean counts are less than 10 were removed from further analysis. (iv) Unique gene hit counts were then used for the subsequent differential expression analysis. For this, we used the DESeq2 R package (v.1.24.0) (41) to compare gene expression between groups of samples and calculate p-values and log<sub>2</sub> of fold changes (LFC) of RNA abundances using Wald tests (function *nbinomWaldTest*). P-values were adjusted for multiple hypotheses testing (Benjamini–Hochberg, BH procedure, (42)) and genes with adjusted P-values (False discovery rate (FDR)) < 0.05 were selected to be further tested as being differentially expressed (DE) (Methods section *RNA-seq d*).

For logistical reasons, the sequencing platform for the RNA-seq data in Methods section *RNA-seq b* differ from one another. Consequently, the data sets first mentioned in the Results sections *Genome-wide mid-term responses correlate with shifts in RNAP concentration* and *Further increases in medium richness do not decrease RNAP concentration* and RNA numbers also do not change, respectively, cannot be compared quantitatively nor be used to infer gene-specific conclusions.

Finally, to analyse the data from LB<sub>1.0x</sub> and LB<sub>0.5x</sub> at 60 and 125 min and compare its results with the results from the data of Methods section *RNA-seq b Part 1* at 180 min, their raw count matrices were merged. Also, only genes that passed the filtering were studied. The filtering removed genes with less than 5 counts in more than 6 samples, and genes whose mean counts were less than 10.

Moreover, we expect the overall sums of LFCs from each perturbation to equal zero since, in DESeq2, the median-of-ratios normalization calculates the normalizing size factors assuming a symmetric differential expression across conditions (i.e. same number of up- and down-regulated genes) (43). Further, it fits a zero-centered normal distribution to the observed distribution of maximum-likelihood estimates (MLEs) of LFCs over all genes (41). Both steps (perhaps related) force the mean LFC to be 0.

*LFC criteria for differentially expressed genes.* From past methods (44–46), we classified genes as statistically significantly DE following perturbations, by setting a maximum FDR threshold for adjusted P-values (Methods section *RNA-seq c*) and a minimum threshold for the absolute LFC of RNA numbers of individual genes (|LFC|).

From the  $\mu_{|LFC|}$  of genes whose FDR > 0.05, named  $\mu_{|LFC|}(FDR > 0.05)$ , we identified DEGs (DE Genes) as those that, in addition to having FDR < 0.05, also have |LFC| >  $\mu_{|LFC|}(FDR > 0.05)$ . Specifically, we added the conditions: |LFC| > 0.4248 for LB<sub>0.75x</sub>, > 0.4085 for LB<sub>0.5x</sub>, > 0.4138 for LB<sub>0.25x</sub>, > 0.2488 for LB<sub>1.5x</sub>, > 0.2592 for LB<sub>2.0x</sub>, and > 0.2711 for LB<sub>2.5x</sub>, for accepting a gene as being significantly DE. Meanwhile, for the data in LB<sub>0.5x</sub> at 60 and 125 min, we added the conditions: |LFC| > 0.2171 for LB<sub>0.5x</sub> 60 min, and > 0.2977 for LB<sub>0.5x</sub> 125 min. This allows removing genes whose FDR < 0.05 but have a negligible LFC. Noteworthy, in no condition did we remove from the set of DEG more than 5 genes by applying this rule.

*RNA-seq vs Flow-cytometry.* RNA and protein abundances are expected to be positively correlated in bacteria, since transcription and translation are mechanically bound (47–49). Further, most regulation occurs during transcription initiation (50), which is the lengthiest sub-process (24).

To validate that this relationship holds during the genome-wide stresses, we randomly selected a set of genes whose LFC's, as measured with RNA-seq, cover nearly the entire spectrum of LFCs observed genome-wide. Next, we measured their LFC in protein levels, using the YFP strain library (25) (Methods section *Bacterial strains, media, growth conditions and curves, and intracellular concentrations*) and flow-cytometry (Methods section *Flow-cytometry*), at 180 min after shifting the medium. The list of selected genes is shown in Supplementary Table S4. For the fold change of 1/8x, 1/4x, 1/2x, 1x, 2x, 4x and 8x, we selected three genes whose LFC in RNA abundances is closest to that value (except for the 8x fold change, since only two genes were available). This range of values covers nearly the whole LFC spectrum observed by RNA-seq (Supplementary Figure S9).

### Transcription factor network of *Escherichia coli*

We assembled a directed graph of the network of TF interactions between the genes present in our RNA-seq data, based on the data in RegulonDB v10.9 (34), as of 28 January 2022. We used all reported TF-gene, TF–TF, TF–operon and TF–TU interactions. These equally contribute to our network of gene-gene directed interactions. In detail, a TF or regulatory protein is a complex protein that activates/represses transcription of a TU upon binding to specific DNA sites. A TU is one or more genes transcribed from a single promoter. Similarly, an operon are one or more genes and associated regulatory elements, transcribed as a single unit.

The TFN graph was analysed using MATLAB (2021b) and Network Analyzer v.3.7.2 plug-in in cytoscape (51) to extract the following network parameters: number of nodes and directed edges, number of connected components, number of isolated nodes and self-loops, and single-gene in- and out-degree, edge-count, clustering coefficient, eccentricity, average minimum path length, betweenness and stress centrality, and neighbourhood connectivity. The statistics considered are shown in Supplementary Tables S5 and S19.

### Statistical tests

*2-Sample T-test, 2-sample KS-test and 1-sample Z-test.* The 2-sample T-test evaluates the null hypothesis that the two samples come from independent random samples from normal distributions with equal means and unequal and unknown variances. For this, we have set a significance level of 10% significance level (P-value < 0.10) when applying the MATLAB function *ttest2*.

The 2-sample KS-test returns a test decision for the null hypothesis that the data from two data sets are from the same continuous distribution, using the two-sample Kolmogorov–Smirnov test. As above, we have set the null hypothesis at 10% significance level (P-value < 0.10).

The one-sample Z-test tests for the null hypothesis that the sample is from a normal distribution with mean *m* and

a standard deviation  $\sigma$ . In this case,  $m$  and  $\sigma$  are estimated from the genes with  $K_{TF} = 0$ . As above, we have set the null hypothesis at 10% significance level ( $P$ -value  $< 0.10$ ).

**Fisher test.** The Fisher test evaluates the null hypothesis that there is no association between the two variables of a contingency table. We reject the null hypothesis at 10% significance level ( $P$ -value  $< 0.1$ ), meaning that the variables are significantly associated.

**Correlations between data sets.** The correlation between two data sets with known uncertainties (standard error of the mean (SEM) in each data point) was obtained by performing linear regression fitting using Ordinary Least Squares. The best fitting line along with its 68.2% confidence interval/bounds (CB) and statistics was obtained as described in Supplementary Materials and Methods 1.4 of (52). In short, the uncertainty of each of the  $N$  empirical data points was represented by  $m$  points, resulting in  $n = N \times m$  points. Each of these points is obtained by random sampling from a normal distribution whose mean ( $\mu$ ) and standard deviation ( $\sigma$ ) equal the mean and error of the empirical data point, respectively. It was set  $m = 1000$ , as it was sufficient to represent the error bars of the actual data points. We obtained the coefficient of determination ( $R^2$ ) and the root mean square error (RMSE) of the fitted regression line, and the  $p$ -values of the regression coefficients. The  $P$ -value of  $x$  ( $P$ -value<sub>1</sub>) was obtained of a  $T$ -test under the null hypothesis that the data is best fit by a degenerate model consisting of only a constant term. If  $P$ -value<sub>1</sub> is smaller than 0.1, we reject the null hypothesis that the line is horizontal, i.e. that one variable does not linearly correlate with the other. When there are more than three data points, we also calculated regression coefficient of  $x^2$  ( $P$ -value<sub>2</sub>) of a  $T$ -test under the null hypothesis that the second order polynomial fit is no better than lower order polynomial fit, i.e. coefficient of  $x^2 = 0$ . If  $P$ -value<sub>2</sub> is smaller than 0.1, we reject the linear model favouring the quadratic.

To obtain the overall best non-linear fit (and its 68.2% CB) for the empirically measured datasets with uncertainties, Monte Carlo simulations (1000 iterations) were performed. In particular, to obtain Figure 2B, on each iteration, we randomly sampled each data point from a normal distribution whose mean and standard deviation are equal to the mean (actual value) and SEM of the corresponding empirical data point, respectively. Then a sigmoid (logistic) curve fitting (R P (2020). `sigm_fit` ([https://www.mathworks.com/matlabcentral/fileexchange/42641-sigm\\_fit](https://www.mathworks.com/matlabcentral/fileexchange/42641-sigm_fit)), MATLAB Central File Exchange. Retrieved 6 August 2020) was used to obtain the best fitting curve and its 68.2% CB for each iteration. Finally, the best fitting curve along with their 68.2% CB is obtained by averaging the respective values from the 1000 iterations.

Finally, to create null-models of how variable  $X$  affects variable  $Y$ , we performed random sampling without replacement of both  $X$  and  $Y$  datapoints. The number of samplings and the sampling size (number of samples in each sampling) are set to the maximum array size possible to us ( $\sim 45980 \times 45980$ , 15.8 GB). The sampling size is set to 5% of the number of datapoints ( $\text{size}_{XY}$ ) and the number of samplings ( $K$ ) is set according to

$\text{Max\_size}/(0.05 \times \text{size}_{XY})$  where  $\text{Max\_size} = 45980/2$ . Next, for both  $X$  and  $Y$ , we combine the sampled datapoints in a vector (`sample_X`, `sample_Y`) and calculate the correlation between `sample_X` and `sample_Y` by linear regression fitting using ordinary least squares. To correct for over-representation of the original datapoints, we corrected the degrees of freedom to be  $(\text{size}_{XY} - C)$ , where  $C$  is the number of parameters. In detail, for the linear regression fitting,  $C$  equals to 2 (intercept and slope of best fitting line).

**ANCOVA test to evaluate if two lines can be distinguished.** To evaluate if two lines are statistically different, we performed the analysis of covariance (ANCOVA) test (53). ANCOVA is an extension of the one-way ANOVA to incorporate a covariate. This allows comparing if two lines are statistically distinct in either slope or intercept. This is done by evaluating the significance of the  $T$ -test under the null hypothesis that both the slopes and intercepts are equal.

## Figures

Figures were produced in R (v.3.6.0) using the packages ‘`ggplot2`’ (v.3.2.0), ‘`pheatmap`’ (v.1.0.12), ‘`VennDiagram`’ (v.1.6.20) along with ‘`grid`’ (v.3.6.0), ‘`gridExtra`’ (v.2.3), ‘`gplots`’ (v.3.0.1.1), ‘`R.matlab`’ (v.3.6.2), ‘`dplyr`’ (v.1.0.2), ‘`scales`’ (v.1.0.0), ‘`Metrics`’ (v.0.1.4) and ‘`fitdistrplus`’ (v.1.0–14).

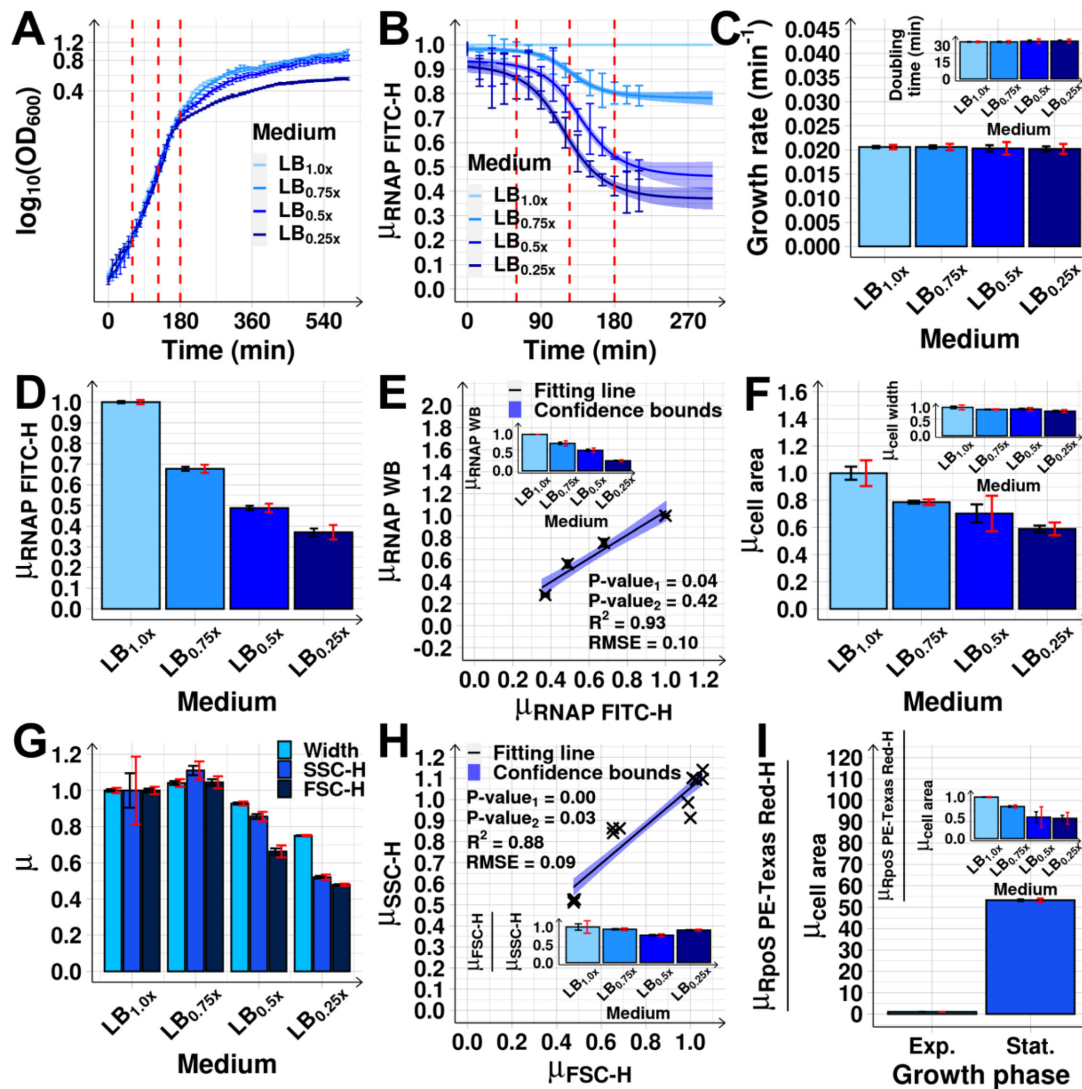
## RESULTS

### Effects of medium dilution on cell growth, morphology, and RNAP concentration

We first studied how the RNAP concentration changes with medium dilutions. Concentration of RNAP (as well as of other molecular species) was obtained as described in Methods section *Bacterial strains, media, growth conditions and curves, and intracellular concentrations*. From a control medium ( $\text{LB}_{1.0\times}$ ), we moved cells to diluted media ( $\text{LB}_{0.75\times}$ ,  $\text{LB}_{0.5\times}$ , and  $\text{LB}_{0.25\times}$ , Methods section *Bacterial strains, media, growth conditions and curves, and intracellular concentrations*). RNAP levels start changing  $\sim 75$  min later, based on a as yet to be identified mechanism, stabilizing at  $\sim 165$  min (Figure 2B). Given this timing of events, measurements to assess the effects on the RNA population should be performed after  $\sim 165$  min.

We also considered that at  $\sim 180$  min (Figure 2A) the cells are at late mid-log phase. Thus, measuring the effects of changing RNAP should occur *prior* to  $\sim 180$  min, since leaving the mid-log phase will involve significant, unrelated genome-wide changes in RNA abundances (54–57). From the point of view of cell division, from the moment when the RNAP starts changing, up to the moment when we measure the short- and the mid-term changes in RNA abundances, on average, less than one cell cycle and less than two cell cycles should have passed, respectively.

Interestingly, this time moment ( $\sim 180$  min) matches our predictions of when, on average, RNA abundances have changed due to changes in the abundances of both RNAP as well as direct input TFs. In detail, from the timing of the changes in RNAP (Figure 2B) and from known rates of RNA and protein production and degradation in *E. coli*



**Figure 2.** Cell growth and morphology, and RNAP abundance after medium dilutions. (A) Growth curves from  $\text{OD}_{600}$  measured every 10 min (Methods section *Bacterial strains, media, growth conditions and curves, and intracellular concentrations*). The vertical dashed red lines mark when RNA-seq was performed. After  $\sim 180$  min, cells subject to different dilutions ( $\text{LB}_{m \times}$ ) start differing in growth rates. (B) Mean single-cell RNAP-GFP fluorescence relative to the control ( $\text{LB}_{1.0 \times}$ ),  $\mu_{\text{RNAP FITC-H}}$ , measured every 15 min for 210 min by flow-cytometry (FITC-H channel). The mean cellular background fluorescence in each condition was subtracted (Methods section *Flow-cytometry*). The vertical dashed red lines mark when RNA-seq was performed. (C) Growth rates at 180 min after medium dilution. The inset shows the corresponding doubling times. (D) Mean single-cell RNAP levels ( $\mu_{\text{RNAP FITC-H}}$ ) at 180 min relative to the control (Methods section *Flow-cytometry*). (E)  $\mu_{\text{RNAP FITC-H}}$  plotted against  $\mu_{\text{RNAP WB}}$  (RNAP levels measured by Western Blot, Methods section *Protein isolation and western blotting*). The inset shows  $\mu_{\text{RNAP WB}}$  alone. (F) Mean cell area relative to the control, extracted from phase-contrast images ( $\sim 2000$  cells per condition) (Methods section *Microscopy*). The inset shows the mean cell width relative to the control. (G) Mean (relative to the control) Width, FSC-H and SSC-H obtained by flow-cytometry (Methods section *Flow-cytometry*). (H) Mean (relative to the control) FSC-H versus SSC-H in each condition, obtained from 3 biological replicates. The inset shows the mean ratio between the relative FSC-H and SSC-H. (I) Mean mCherry-tagged RpoS ( $\mu_{\text{RpoS PE-Texas Red-H}}$ ) concentration in the stationary growth phase relative to the exponential growth (set to 1), as measured by mean single-cell fluorescence (PE-Texas Red channel, Methods section *Flow-cytometry*) over mean cell area ( $\mu_{\text{cell area}}$ ) (Methods section *Microscopy*), after subtracting mean background fluorescence(s). The inset shows the same, but after each medium dilution. Measurements in (D)–(I) taken 180 min after medium dilution. Data points are from 3 biological replicates (except for (A) and (B), where 6 replicates were used).  $\mu$  stands for mean relative to the control. In (A)–(C) error bars represent the SEM. In (B) and (D)–(I), black error bars are the SEM and red error bars are the 95% CB of the SEM. In (C), (E) and (H), the best fitting lines and their 68% CB and statistics ( $R^2$  and RMSE), and P-values at 10% significance level were obtained as described in Methods section *Statistical tests c*.



(25,28–31), widespread heterogeneous short-term changes in RNA abundances should occur, on average, at  $\sim 120$ – $135$  min after shifting the medium (at which moment the RNAP has already changed significantly). Changes in the corresponding protein abundances should then occur tenths of minutes later, i.e. at  $\sim 160$ – $175$  min (29–31).

Soon after, we expect additional changes in RNA abundances, now due to changes in direct input TFs abundances. This second stage of events, here classified as ‘mid-term’, should occur between  $\sim 165$  and  $180$  min. This is also when cells are in the late mid-log phase (Figure 2A), while cell growth rates do not yet differ between conditions (Figure 2C) and cell sizes only differ slightly (Figure 2F–H and Supplementary Figures S3 and S4, Methods sections *Microscopy* and *Flow-cytometry*). This is relevant since growth rates affect protein concentrations due to dilution in growth and division (58,59).

Finally, at  $180$  min, the  $\sigma^{38}$  concentration is lower than at  $0$  min (Figure 2I inset and Supplementary Figure S5), in agreement with previous reports (27,35,60,61), suggesting that the cells are not committed to the stationary growth phase. The same is observed for the CrI protein (Supplementary Figure S6). This protein contributes to the expression of genes whose promoter is recognized by  $\sigma^{38}$  and is known to be at higher abundance during stationary phase (62), as confirmed here (Supplementary Figure S6).

Given the above, to capture the average mid-term effects of RNAP shifts, we measured the transcriptome at  $180$  min (Figure 2A). This timing should allow discerning the average genes’ behaviour under the influence of their local network of TF interactions, albeit the diversity in RNA and protein production and decay kinetics, etc. RNAP levels at that moment are shown in Figure 2D (flow-cytometry data) and 2E (flow-cytometry versus western blot data), Supplementary Figures S2A<sub>1</sub> and S2A<sub>2</sub> (western blot of RNAP) and Table S2 (absolute values extracted from western blot). Similar RNAP downshifts have been observed in natural conditions (63) and described in (23,26,27).

Finally, we performed an additional western blot to measure RNAP-GFP using antibodies against GFP alone. From Supplementary Figures S2B<sub>1</sub> and S2B<sub>2</sub> and Table S2, the RNAP-GFP does not appear to be significantly degraded by cleavage, with the strongest bands being observed for molecular weights between  $150$  and  $250$  kDa. As such, these strongest bands should correspond to GFP (known to be  $27$  kDa (64)) fused with the  $\beta'$  unit (known to be  $155$  kDa (65,66)). Moreover, no clear bands appear in the region between  $150$  and  $250$  kDa for the MG1655 strain. Meanwhile, the weak band just above  $25$  kDa in some samples from RL1314 cells (particularly in LB<sub>0.75x</sub>) might correspond to GFP that has been cleaved off from the chimeric protein but, given that it is only a small fraction compared to the amount of RpoC-GFP in the same cells, one can conclude that its contribution to the total cell fluorescence signals is negligible.

### Genome-wide mid-term responses correlate with shifts in RNAP concentration

Transcription rates are expected to follow the free RNAP concentration in a cell, rather than the total RNAP con-

centration (which is the sum of the free RNAP with the RNAP engaged with the DNA). We here measured the total RNAP concentration. However, within the range of conditions studied, the fractions of free and DNA-bound RNAP remain rather constant (26). Therefore, the total RNAP is a good proxy for the free RNAP. Specifically, using modified strains and plasmids controlled by lac and tet mutant promoters (67–69), whose regulatory mechanisms have been dissected, it was shown that their transcription rates are linearly correlated with the total RNAP concentration (26). From here on, when mentioning RNAP concentration, we refer to the total RNAP concentration.

The increasing medium dilution and corresponding decreases in RNAP concentration (Figure 3A) cause increasingly broad distributions of single-gene LFCs at  $180$  min (Supplementary Figures S7 and S8A–C and Table S3). Specifically, the mean absolute LFC ( $\mu_{|LFC|}$ ) of the 4045 genes and the number of DEGs increased with medium dilution (Figures 3B and C).

These RNAs changes correlate with subsequent changes in proteins levels (Supplementary Figure S9, Methods sections *Flow-cytometry* and *RNA-seq*). This suggests that no significant translational or post-translational regulation is taking place in between the perturbation and the measurements, that would alter proteins abundances significantly.

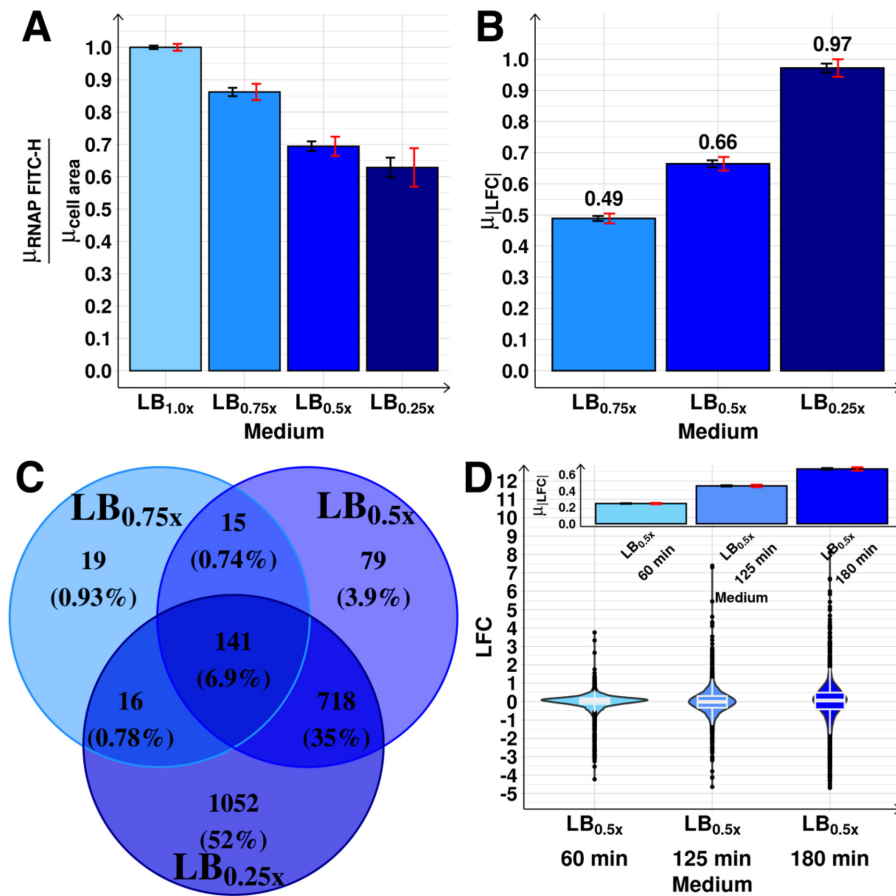
Interestingly, while both  $\mu_{|LFC|}$  and DEGs numbers follow the RNAP concentration (Supplementary Figures S8D and S10B), these relationships are not strictly linear (p-value of  $0.29$ , Supplementary Figure S8D). This suggests that, in addition to RNAP, the direct input TFs are also influential. In this regard, we note that the assumption of linearity in the absence of the influence of input TFs (observed and discussed in (26)) is only expected to occur within a narrow range of parameter values.

Notably, some of the genes may be also influenced by sources other than RNAP and direct input TFs, such as supercoiling buildup. Also, some input TFs other than the direct input TFs maybe be influential. However, we show evidence below that this does not affect the average results (Figure 4C and Supplementary Figure S18).

We also performed RNA-seq *prior* to when most signals, generated by the shift in RNAP, propagated in the TFN. First, we measured LFCs at  $60$  min after diluting the medium (Figure 1). From Figure 2B, at this moment, RNAP abundances have not yet changed relative to the control. In agreement, the genome-wide  $\mu_{|LFC|}$  is very weak (Figure 3D). We further performed RNA-seq at  $125$  min. At this moment, RNAP levels have already reduced significantly (Figure 2B), but we do not expect input TF abundances to have changed significantly given protein production times (Figure 1). In agreement,  $|LFC|$ s at  $125$  min are stronger than at  $60$  min, but much weaker than at  $180$  min (Figure 3D). We conclude that the mid-term changes in the TFN have not occurred yet (further evidence is provided below). Given this, from here onwards, we focus on the state of the TFN at  $180$  min.

### Influences from regulators other than RNAP

We investigated whether other factors influenced the global response of the TFN. We considered GRs,  $\sigma$  factors,



**Figure 3.** Genome-wide effects on the transcriptome of diluting the medium. (A) RNAP concentration estimated from the ratio between the RNAP measured by FITC-H ( $\mu_{RNAP\ FITC-H}$ ) at 180 min (Methods section *Flow-cytometry*), and the mean cell area ( $\mu_{cellarea}$ , used as a proxy for cell volume) obtained by phase-contrast microscopy (Methods section *Microscopy*). Values relative to the control (LB<sub>1.0x</sub>). (B)  $\mu_{LFC}$  in each medium. (C) Venn diagram of the number (and percentage relative to the total number of genes) of DEGs. (D) Violin plot with the maximum, minimum, median, interquartile ranges, and probability density of the distributions prior to RNAP changes (LB<sub>0.5x</sub> at 60 min) and the subsequent short- (LB<sub>0.5x</sub> at 125 min) and mid-term (LB<sub>0.5x</sub> at 180 min) responses to shifting RNAP. The inset shows  $\mu_{LFC}$  of the distributions. In (A), (B) and (C), black error bars are the SEM, while red error bars are the 95% CB of the SEM.

(p)ppGpp and non-coding RNAs. We assumed the classification of GR in (70,71) as an input TF that regulates a large number of genes that rarely regulate themselves and participate in metabolic pathways. Meanwhile, we did not account for promoters' close proximity (e.g. tandem formation), since a recent study (72) showed that, under similar stress, while close proximity causes transcription interference, it does not influence the genes' input TF regulation.

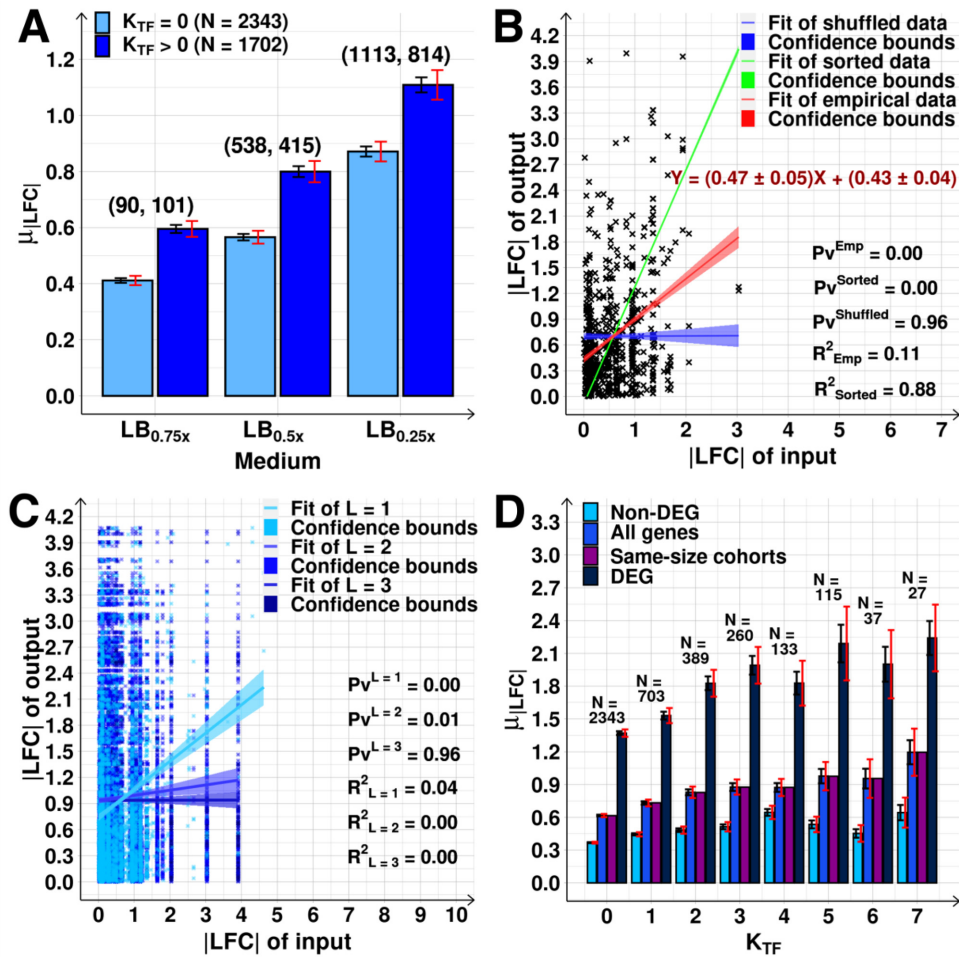
First, the RNA-seq shows large numbers of DEGs (> 1000 for the two strongest dilutions (Supplementary Figure S10A)) as well as a linear correlation between these numbers and  $\mu_{LFC}$  (Supplementary Figure S10C). Thus, we argue that the responsive genes are not constrained to a specific cluster, such as genes responding to a GR other than RNAP (the most influential is, arguably,  $\sigma^{70}$  with 1555 genes recognizing it, while other GRs control less than 510 genes each (34)).

Second, from the RNA-seq, we analysed the relative abundances of GRs,  $\sigma$  factors and of their output genes. From Supplementary Figures S26A and S26C, apart from

rpoS (an input TF recognized by 321 genes) and flhC (an input TF recognized by 75 genes), GRs and  $\sigma$  factors did not change significantly (Supplementary Figure S26). Further, those two changes (rpoS and flhC) were positively correlated with the RNAP concentration (Figure 2I inset and Supplementary Figure S5), not allowing to separate their effects. Noteworthy, alternative  $\sigma$  factors did not change significantly relative to  $\sigma^{70}$  (Supplementary Figure S26E), which would have changed the competition for RNAP binding.

We thus failed to find evidence that the  $\sigma$  factors and GRs were influential, globally, in the mid-term responses. Supplementary Table S15 lists the conclusion for each specific GR and  $\sigma$  factor and Supplementary Figure S27 shows these results at 125 min.

We then investigated if (p)ppGpp could be influential since, under some nutrient starvation conditions, they affect ~ 1000 genes by binding RNAP and altering its affinity for their promoters (3). Reports suggest that the effects are rapid (5–10 min (3)). In agreement, genes responsive



**Figure 4.** Genome-wide propagation of the effects of shifting RNAP in the TFN. (A)  $\mu_{|LFC|}$  of  $N$  genes with and without input TFs ( $K_{TF} > 0$  and  $= 0$ , respectively). On the top of each bar is the number of DEGs in each set. (B)  $|LFC|$  of genes with  $K_{TF} = 1$  versus the  $|LFC|$  of genes coding their direct input TFs. Data from the  $LB_{0.5x}$  shift. The red line is the best fit. The blue line is the null-model fitting lines and was obtained as described in Methods section *Statistical tests c*. The green line is the best fit after sorting the input-output pair values to maximize the correlation. Shadows are their 68% CB. The equations of the red fitting lines with ‘ $\pm$ ’ inform on the standard error of the slope. (C) Scatter plots between  $|LFC|$  of output and input genes distanced by a minimum path length  $L$  of 1, 2 and 3 input TFs (edges) in the TFN, respectively (data from  $LB_{0.5x}$ ). Only for  $L = 1$  do the activities of output and input genes correlate ( $P\text{-value}_1 > 0$  and  $R^2 > 0$ ). (D)  $\mu_{|LFC|}$  of all genes, DEG, non-DEG, and cohorts of randomly selected genes of the same size (‘same sized cohorts’) for  $K_{TF} = 0$  to 7, using merged data from all shifts ( $LB_{0.75x}$ ,  $LB_{0.5x}$  and  $LB_{0.25x}$ ). Black error bars are the SEM and red error bars are the 95% CB of the SEM. Best fitting lines and 68% CB obtained using *FITLM* (MATLAB). p-values, obtained using the null hypothesis that the data is best fit by a horizontal line, are not rejected at 10% significance level. (B) and (C) do not include a few data points to facilitate visualization. See Supplementary Figures S14 and S18 for complete data.

to (p)ppGpp (3) exhibited abnormal short-term responses (Supplementary Table S20). However, their mid-term responses at 180 min were no longer atypical and, instead, followed the RNAP changes. The expression of *spoT*, one of the genes responsible for ppGpp synthesis, also followed the RNAP (Supplementary Figure S28). As such, we could not establish a long-lasting global influence from (p)ppGpp in response to growth-medium dilution. Nevertheless, the LFCs of the 14 out of the 22 genes coding for rRNAs listed in RegulonDB did reveal atypical behaviors (Supplementary Table S22).

Finally, we searched for unique behaviors in sRNAs by analyzing the LFC of the 93 sRNAs reported in RegulonDB. Their behavior was not atypical, neither at 180 min after the perturbations (Supplementary Table S21), nor at

125 min. Further, we analysed if their output genes followed their behaviour. We found that the LFCs of genes directly regulated by the sRNAs were not correlated with their input TFs, neither at 125 min, nor at 180 min after the medium shifts. Specifically, of the 93 sRNAs, 37 of them have known output genes (in a total of 145 outputs). The RNA-seq data provided information on the LFC of 40 of the 145 outputs. Finally, we searched for linear correlations between the pairs of LFCs of sRNAs and their output genes, respectively, in the short-term (125 min in  $LB_{0.5x}$ ) and in the mid-term (180 min in  $LB_{0.5x}$ ). We found an  $R^2$  of 0.03 ( $P$  value = 0.18) at 125 min and an  $R^2$  of 0.05 ( $P$  value = 0.10) at 180 min, respectively. We thus cannot conclude that sRNAs were influential during the short- and mid-term responses to the stresses.

### Input TFs influence the transcriptional response

If the TFN influences the genes' mid-term response to the shift in RNAP concentration, this should cause genes with and genes without input TFs to behave differently. Particularly, since the latter should only be affected by the RNAP abundances.

In agreement, genes with input TFs had higher  $\mu_{|LFC|}$  than genes without input TFs (Figure 4A, Supplementary Figure S13 and Table S6). Also, the |LFC| of output genes and of genes coding for their direct input TFs correlate statistically (Figure 4B, Supplementary Figure S14 and Table S7). Therefore, on average, TF–gene interactions affected the single-gene, mid-term responses as hypothesized (Figure 1).

### Input TFs influence all genes within operons

When considering the TFN topology, we have accounted for TF–gene interactions both between the input TF and the first gene of an operon or TU. Further, we have also accounted for the interactions between the same input TF and the other genes of the operon or TU (illustration of TUs and operons in Supplementary Figure S11B, which follows the standard definition of a group of two or more genes transcribed as a polycistronic unit (1)).

If we had not account for all these interactions, we would have failed to correlate the activities of genes interacting with each other. For example, consider an operon consisting of genes  $X_1$  and  $X_2$ . Then, assume that gene A represses  $X_1$  and  $X_2$ , by repressing their common promoter. If  $X_1$  is an input TF to gene C, while  $X_2$  is an input TF to gene D, then gene A should indirectly affect both genes C and D. If we had ignored the interaction between A and  $X_2$ , because it is not the first gene in its operon, we would explain why A affects C, but fail to explain why A affects D.

Further, many operons contain sets of genes whose RNAs code for subunits of the same protein complex (73,74). However, the opposite is also true. Moreover, the fraction of complexes encoded by proteins from different TU's is higher than those encoded from the same operon (75). Thus, we need to track interactions between input TFs and genes in any position in an operon or TU.

We tested if the positioning of the genes in the operon influenced their responsiveness to their input TFs. As a case study, we considered operons with 3 genes. These account for ~ 21% of all operons with more than 1 gene (34). We found that the genes' positioning did not affect how they relate to the input TFs (Supplementary Figure S16). We obtained similar results for TUs (Supplementary Figure S17). The tests of statistical significance are shown in Supplementary Tables S9-S12.

### Genes expressing TFs are correlated with their nearest neighbour output genes

Consider the interval between the shift in RNAP levels and the sampling for RNA-seq (Figure 1). From these, we hypothesized that, on average, at 180 min (i.e. ~ 70 min after the RNAP changed relative to the control), mostly only genes directly linked by input TFs should exhibit correlated responses. Nevertheless, the genome-wide diversity in the

kinetics of gene expression and in RNA and protein lifetimes will allow for correlations between genes more distanced in the TFN. The number of such correlations should decrease rapidly with the path length between the gene pairs considered.

Results in Figure 4C support this. Genes distanced by 1 input TF ( $L = 1$ , i.e. directly linked) have related |LFCs|, while genes distanced by two input TFs in the TFN have much less correlated responses (albeit still statistically significant). Finally, we found no correlations between the |LFCs| of genes distanced by three input TFs (Supplementary Figure S18).

Noteworthy, the lack of correlation between genes separated by  $L > 1$  could also be partially due to interference from the TFs of the 'intermediary' genes between the gene pairs. However, this is only a possibility when all input TFs involved can change in abundance in less than 60 min, which is likely uncommon in *E. coli*. This is supported by the RNA-seq data at 125 min after medium dilution (Supplementary Figure S19), when even direct input TFs and output genes are weakly correlated. This suggests that this shorter time interval was insufficient for most signals to have propagated between nearest neighbours (Supplementary Figure S19).

### The number of input TFs, $K_{TF}$ , of a gene correlates to the magnitude of its transcriptional response

We investigated if the genes mid-term responses are sensitive to their  $K_{TF}$  (Supplementary Figure S20A). When averaging the results from the three perturbations (Figure 4D), we found that the average of the absolute LFCs,  $\mu_{|LFC|}$ , increases with  $K_{TF}$  (Supplementary Figure S22 shows the goodness of the linear fits). The result was the same whether considering all genes or just the DEGs (Figure 4D, Supplementary Figure S21 and Table S13). Further, it holds true even for non-DEGs (Figure 4D), which justifies also considering these genes when studying the genome-wide effects. In agreement, we found no trend in the fraction of DEGs when plotted against  $K_{TF}$  (Supplementary Figure S23). For comparison, neither at 60 min nor 125 min do the genes' response and their  $K_{TF}$  correlate (Supplementary Figures S14 and S15 and Tables S7 and S8).

We verified that the relationship between  $\mu_{|LFC|}$  and  $K_{TF}$  at 180 min is not an artifact caused by a decrease in cohort size with  $K_{TF}$ . We used bootstrapping to obtain cohorts of randomly sampled genes with increasing  $K_{TF}$  (10000 cohorts). We imposed a cohort size equal to the number of genes with  $K_{TF} = 7$  (27 genes). The new, estimated  $\mu_{|LFC|}$  was always within the SEM of the  $\mu_{|LFC|}$  of the cohorts of all genes (Figure 4D). Finally, we again verified that considering only the first gene of each operon does not affect how  $\mu_{|LFC|}$  and  $K_{TF}$  relate (Supplementary Figure S25).

### The correlation between input and output genes responses decreases with the number of input TFs

Most input TFs discernibly affect the output genes (Supplementary Figure S14), except when  $K_{TF} > 5$  (perhaps due to saturation). Nevertheless, the correlation between input and output genes appears to be decreasing with  $K_{TF}$ .

Namely, the average slopes of the fitted lines between  $|\text{LFCl}|$  of the output and  $|\text{LFCl}|$  of each input (Supplementary Figure S14) decreased with the  $K_{\text{TF}}$  of the output gene (Supplementary Figure S20B). Also decreasing was the  $R^2$  between input-output pairs (Supplementary Table S7). This could explain why, when plotting  $|\text{LFCl}|$  against the RNAP concentration, there is a weak trend towards increased slope with  $K_{\text{TF}}$  (Supplementary Figures S20C and S21).

### The variability in single-gene absolute LFC increases with $K_{\text{TF}}$

We also investigated if the variability in  $|\text{LFCl}|$ s, as quantified by its standard deviation  $\sigma_{|\text{LFCl}|}$ , relates with  $K_{\text{TF}}$ . There should exist (at least) four sources of this variability: (a) RNA-seq measurement noise (76,77); (b) intrinsic and (c) extrinsic noise in gene expression (78,79), and (d) TF and non-TF dependent regulatory mechanisms.

Overall, we observed that, from a genome-wide perspective,  $\sigma_{|\text{LFCl}|}$  increases with  $K_{\text{TF}}$  (Supplementary Figure S24A) similarly to  $\mu_{|\text{LFCl}|}$ , and the two values are also related (Supplementary Figure S24B). Examples of the variability are shown in Supplementary Figures S24C (genes with null  $K_{\text{TF}}$ ), S24F (genes with two GRs, FNR and ArcA) and S24D and S24E (genes controlled by the GRs FIS or CRP) (see also Supplementary Table S14).

### Other topological features of the TFN do not influence mid-term responses

Globally, the TFN of *E. coli* has in- and out-degree distributions that are well fit by power laws (Supplementary Figures S12E<sub>1</sub>, S12E<sub>2</sub>, S12F<sub>1</sub> and S12F<sub>2</sub>) (80,81). This may explain its relatively short mean path length (Supplementary Figure S12G and Table S5).

Having established a relationship between the response kinetics and the indegree of the TFN, we next searched for correlations between  $|\text{LFCl}|$  and other single-gene topological traits (Methods section *Transcription Factor Network of Escherichia coli*). We considered average shortest path length, betweenness, closeness and stress centrality, clustering coefficient, eccentricity, out-degree, neighbourhood connectivity and edge-count (51). Of these, only the clustering coefficient was statistically correlated with the  $|\text{LFCl}|$  ( $P$ -value  $< 0.1$ ) (Supplementary Table S20). However, it should not be influential, since the corresponding  $R^2$  is nearly zero ( $R^2 = 0.01$ ).

### The numbers of activating and repressing input TFs differ in most genes

In our original hypothesis, the mid-term response ( $|\text{LFCl}|$ ) of a gene should follow from the bias in the numbers of activators and repressors in its set of input TFs (Figure 1B<sub>4</sub> and Supplementary Figure S11A). In detail, we predicted that if the sum of the regulatory effect ( $r$ ) of the input TFs (i.e. bias  $b = |\sum r|$ ) is null (unbiased), then the gene should have weak or zero mid-term LFC. Also, the  $|\text{LFCl}|$  should increase with  $b$ .

We tested this hypothesis by extracting information on the input TFs and corresponding  $r$  values for each gene

from RegulonDB. We set  $r$  of an input TF to +1 if it is activating, to -1 if it is repressing, and to 0 if it is unknown (Supplementary Figure S12B). Then, we obtained the absolute sum of the regulatory effect of the input TFs for each gene:  $|b|$ .

From the data in RegulonDB, while the gene-TF interactions that are repressions and activations exist in similar numbers, the numbers of repressor TFs exist in larger numbers (Supplementary Figures S12A–C). Also, of the genes with input TFs, most ( $\sim 85\%$ ) have a non-zero  $|b|$  (Supplementary Figure S12D and Table S16).

This can explain why so many are mid-term responsive (Figure 3C), even though the genome-wide numbers of *activation* and *repression interactions* are similar (Supplementary Figure S12B). It may also explain why genes with  $K_{\text{TF}} \geq 1$  have higher  $|\text{LFCl}|$  than genes with  $K_{\text{TF}} = 0$  (Figure 4A).

### The bias in the input TFs follows the number of input TFs

Using information from RegulonDB, we found that the mean bias,  $\mu_{|b|}$ , increases with  $K_{\text{TF}}$  (Figure 5A, light blue), except for  $K_{\text{TF}} > 5$ , which includes only  $\sim 64$  out of 4045 genes (Supplementary Table S17). The same is observed if considering only the first gene of each operon (Supplementary Figure S29).

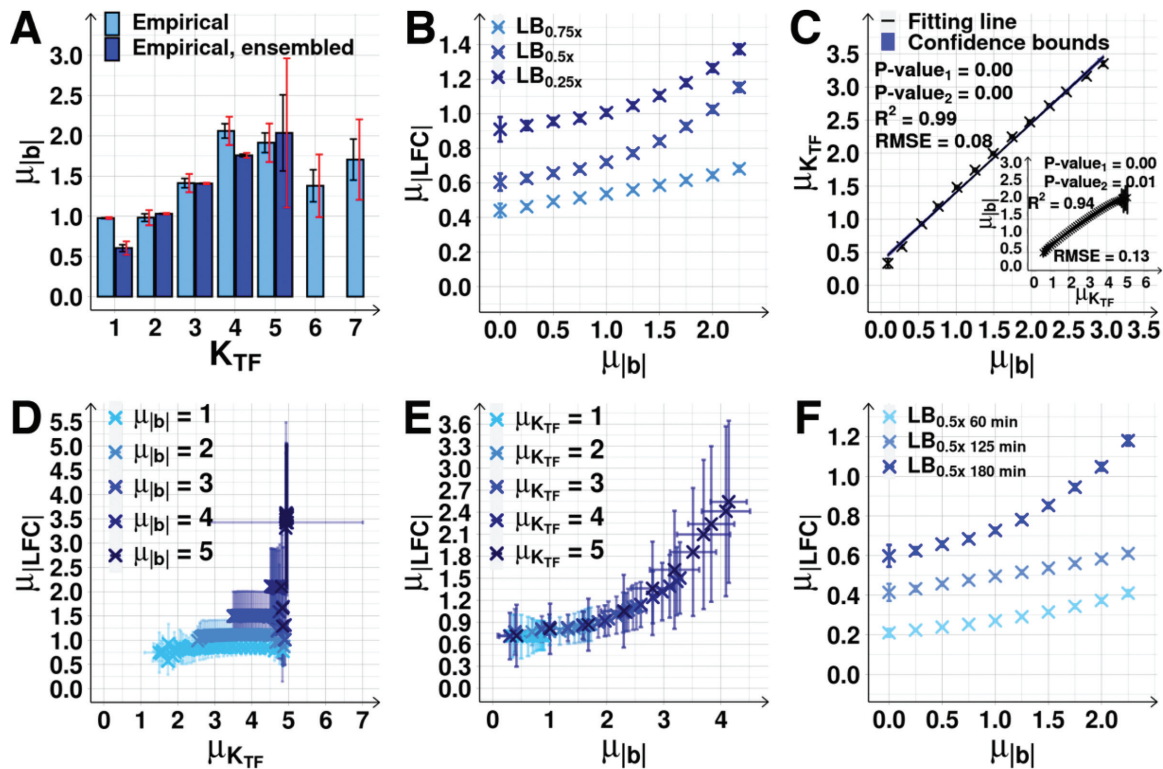
To test if these results were affected by local topological specificities, we employed an ensemble approach (Supplementary Results section *Estimation of the expected  $\mu_{K_{\text{TF}}}$  and  $\mu_{|b|}$  using an ensemble approach*), to reduce their influence (82). We sampled genes (with replacement) to form cohorts with a given average  $K_{\text{TF}}$  (from 1 to 5, due to insufficient samples for higher  $K_{\text{TF}}$ ). This made the relationship between  $\mu_{|b|}$  and  $K_{\text{TF}}$  more stable (Figure 5A). Thus, from here onwards, we use the ensemble approach to study the influence of the global logical and topological features on the response's dynamics to the RNAP shifts.

### The bias of the sets of input TFs can explain the mid-term responses of individual genes

From the data in RegulonDB, using the ensemble approach (Supplementary Results section *Estimation of the expected  $\mu_{K_{\text{TF}}}$  and  $\mu_{|b|}$  using an ensemble approach*), we formed random cohorts of genes with an imposed average  $|b|$ . Next, from the mid-term RNA-seq data, we calculated the average  $\mu_{|\text{LFCl}|}$  of the set of cohorts with a given  $\mu_{|b|}$ . We found that  $\mu_{|\text{LFCl}|}$  increases with  $\mu_{|b|}$  (Figure 5B).

Interestingly,  $\mu_{|b|}$  and  $\mu_{K_{\text{TF}}}$  are strongly correlated in the TFN of *E. coli* (Figure 5C). To assert which one controls  $\mu_{|\text{LFCl}|}$ , we assembled cohorts differing in  $\mu_{K_{\text{TF}}}$ , but not in  $\mu_{|b|}$ . In these,  $\mu_{|\text{LFCl}|}$  does not increase with  $\mu_{K_{\text{TF}}}$  (Figure 5D). We also assembled cohorts differing in  $\mu_{|b|}$ , but not in  $\mu_{K_{\text{TF}}}$ . In these,  $\mu_{|\text{LFCl}|}$  increases with  $\mu_{|b|}$  (Figure 5E). Thus, the increase of  $\mu_{|b|}$  with  $\mu_{K_{\text{TF}}}$  (Figure 5C), is what explains the increase in  $\mu_{|\text{LFCl}|}$  with  $K_{\text{TF}}$  (Figure 4D).

Finally, for comparison, we also investigated the relationship between  $\mu_{|b|}$  and  $\mu_{K_{\text{TF}}}$  prior to the perturbation and in the short-term (at 60 min and at 125 min after shifting the medium, respectively Figure 1A). From Figure 5F, first, the



**Figure 5.** Effect of biases  $\mu_{|b|}$  on the magnitude of the response of output genes. (A)  $\mu_{|b|}$  as a function of  $K_{TF}$  (light blue) of gene cohorts with all genes (light blue) and of gene cohorts assembled using the ensemble approach (dark blue). Supplementary Table S17 shows the fractions of genes with equal  $|b|$  and  $K_{TF}$ . Black error bars are the SEM, and red error bars are the 95% CB of the SEM. Dark blue bars not shown for  $K_{TF} > 5$  due to small sample sizes. (B) Mid-term  $\mu_{|LFC|}$  as a function of  $\mu_{|b|}$ , obtained using the ensemble approach (Supplementary Results section *Estimation of the expected  $\mu_{K_{TF}}$  and  $\mu_{|b|}$  using an ensemble approach*, Supplementary Figures S30 and S31 and Table S18). Each blue cross is the average outcome from up to 24750 cohorts of 10 genes.  $\mu_{|b|}$  plotted against the corresponding  $\mu_{K_{TF}}$ , mean of  $K_{TF}$  of the cohorts in (B). The inset shows the inverse correlation plot for the cohorts in Supplementary Figure S30, assembled based on  $\mu_{K_{TF}}$  (Supplementary Results section *Estimation of the expected  $\mu_{K_{TF}}$  and  $\mu_{|b|}$  using an ensemble approach*). Shown are best fitting lines and 68% CB (shadow areas, barely visible),  $R^2$ , RMSE, and P-value (Methods section *Statistical tests c*). (D)  $\mu_{|LFC|}$  of gene cohorts with increasing  $\mu_{K_{TF}}$ , but constant  $\mu_{|b|}$  (from 1 to 5) (Supplementary Results section *Estimation of the expected  $\mu_{K_{TF}}$  and  $\mu_{|b|}$  using an ensemble approach*). (E)  $\mu_{|LFC|}$  of gene cohorts with increasing  $\mu_{|b|}$ , but constant  $\mu_{K_{TF}}$  (from 1 to 5) (Supplementary Results section *Estimation of the expected  $\mu_{K_{TF}}$  and  $\mu_{|b|}$  using an ensemble approach*). (F)  $\mu_{|LFC|}$  as a function of  $\mu_{|b|}$  prior to RNAP changes (60 min) as well as the short-term (125 min) and the mid-term responses (180 min) to RNAP changes when shifting to  $LB_{0.5x}$ . Each blue cross is the average outcome from up to 24829 cohorts of 10 genes. In (D) and (E), the data is merged from the three conditions corresponding to (B). In all figures, the error bars are the SEM. Since the three conditions differ slightly in mean values (Figure 5B), the SEM is larger than when observing each condition separately.

$\mu_{|LFC|}$  at 125 min is stronger than at 60 min. This agrees with the expectation that shifts in RNAP suffice to shift the  $|LFC|$  of many genes. Second, the  $\mu_{|LFC|}$  at 180 min is stronger than at 125 min. This agrees with our expectation that, at 125 min, input TFs numbers have not yet changed significantly in order to enhance the  $|LFC|$  of their output genes (Figure 1).

### RNA numbers follow the RNAP concentration, not the medium composition

We next increased growth medium richness, instead of diluting it (Methods section *Bacterial strains, media, growth conditions and curves, and intracellular concentrations*). As before, we limited this to not alter growth rates significantly in the first 180 min (Figure 6A and B), while altering RNAP levels (Figure 6C).

As before (Figure 4C), at mid-term, only genes directly linked by input TFs showed correlation in their  $|LFC|$  (Figure 6D<sub>1</sub>–D<sub>3</sub> and Supplementary Figure S32). This supports the previous assumption on the kinetics of transcription,

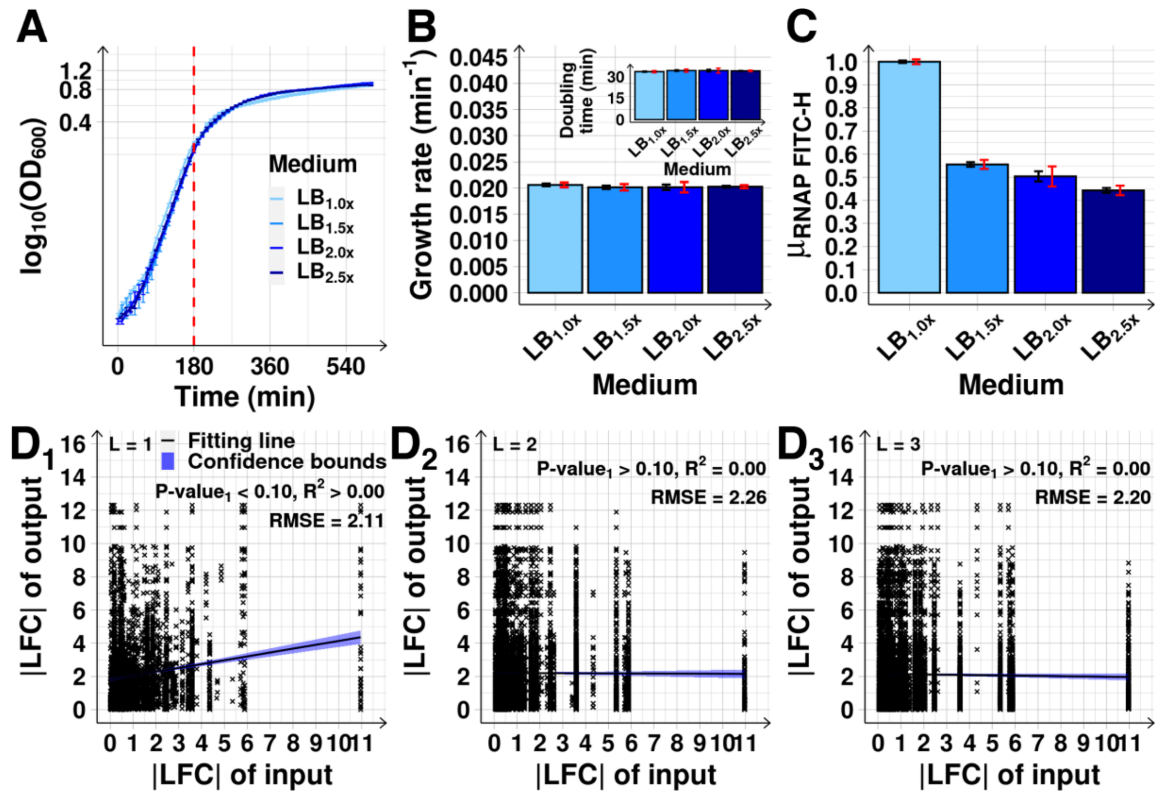
translation, and signals propagation via shifts in input TFs numbers (Figure 1).

Meanwhile, in contrast to above, shifting cells from  $LB_{1.0x}$  to the richer  $LB_{1.5x}$  medium was accompanied by a decrease in the RNAP concentration (Figure 7A). This was followed by substantial alterations in the RNA populations, with a large number of DEGs and high  $\mu_{|LFC|}$  (Figure 7B and C, respectively). As previously, in the mid-term, genes with input TFs reacted more strongly (Figure 7C).

These results support the initial assumption that the changes in RNA abundances follow the RNAP concentration, rather than the medium richness.

### Further increases in medium richness do not decrease RNAP concentration and RNA numbers also do not change

Finally, we further increased growth-medium richness (to  $LB_{2.0x}$  and to  $LB_{2.5x}$ ). This caused no significant change in RNAP levels and concentration (Figures 6C and 7A). Here, we also did not observe significant changes in DEGs or  $\mu_{|LFC|}$  at mid-term, when compared with the  $LB_{1.5x}$  con-



**Figure 6.** RNAP levels following increasing medium richness and corresponding relationships between ILFCs of pairs of genes separated by specific path lengths,  $L$ . (A) Growth curves from  $OD_{600}$  assessed every 10 min (Methods section *Bacterial strains, media, growth conditions and curves, and intracellular concentrations*), following each medium shift. (B) Growth rates at 180 min after medium enrichment. The inset shows the corresponding doubling times. (C) Mean RNAP levels relative to the control estimated from single-cell RNAP-GFP fluorescence intensities (FITC-H) ( $\mu_{RNAP\ FITC-H}$ ). (D<sub>1</sub>–D<sub>3</sub>) Scatter plots between absolute LFC (ILFC) of outputs and corresponding input genes distanced by  $L$  (path length) of 1, 2 and 3 transcription factors, respectively. Data from the  $LB_{2.5x}$  condition. Shown are the best fitting line and its 68% CB (blue shadow), and the  $R^2$  and RMSE of the fitted regression line, along with its  $P$ -value at 10% significance level under the null hypothesis that this line is horizontal. From (A) to (C), the black error bars are the SEM and red error bars represent the 95% CB of the SEM.

dition (Figure 7B and C, respectively). This is in agreement with the assumption that the shifts in the RNAP concentration caused the short-term changes in RNA abundances, which then caused the mid-term changes.

Finally, as before,  $\mu_{|LFC|}^{TF}$  follows  $\mu_{|b|}$  (Figure 7D and Supplementary Figure S33). Moreover, it does so almost identically in the three perturbations, as expected from the original assumptions (Figure 1).

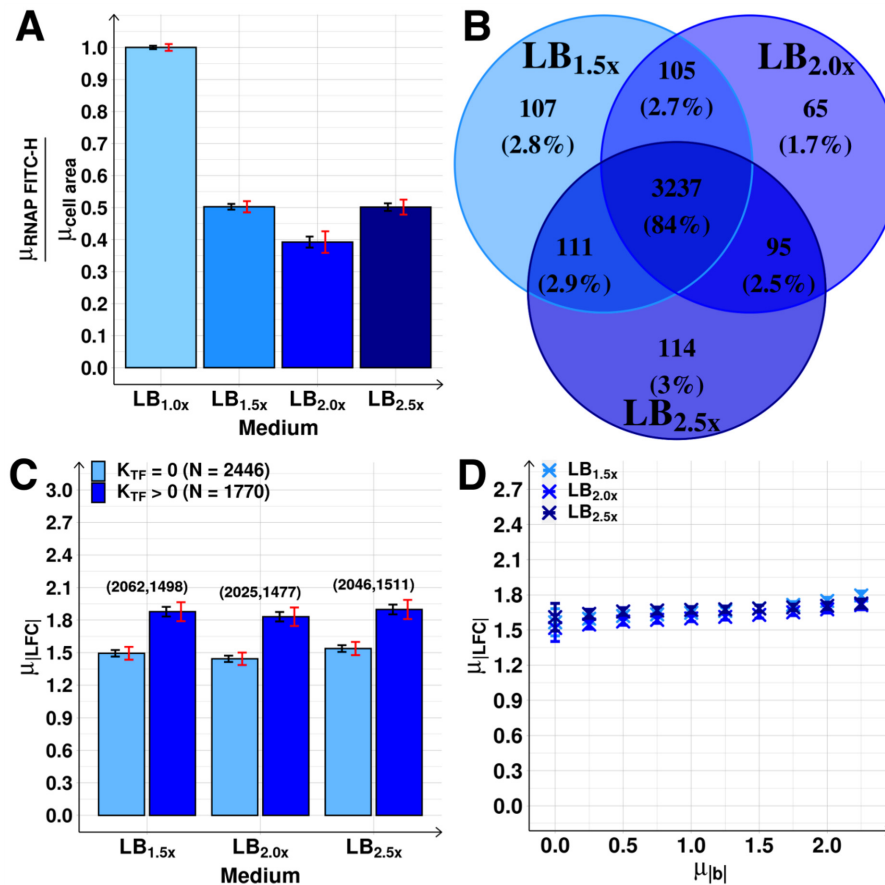
## DISCUSSION

We investigated if the mid-term responses to genome-wide perturbations of *E. coli*'s TFN are mediated by its topology and logic. We diluted LB medium since this dramatically and reproducibly affects the RNAP concentration (26,27). The increasingly strong nature of the dilutions facilitated the verification of how the RNAP concentration and single-gene, mid-term ILFCs related. We focused on mid-term transcriptional responses (Figure 1), since short-term responses are unlikely to have been influenced by the TFN due to protein folding and maturation times, etc. Meanwhile, long-term responses were most likely affected by the TFN. However, dissecting them would have been onerous, due to the complicating effects of loss, backpropagation, and co-

alescence of possibly dozens of signals from origins other than direct input TFs.

We lack information on the affinity between each gene and their input TFs, on how the input TFs operate, and on how the *de novo* presence of an input TF alters the binding or activity of other input TFs on the same promoter. Thus, we would have failed to predict the behaviour of individual genes with accuracy. Instead, we predicted the responses of gene cohorts, since their behaviour is less influenced by particular single-gene features (other than cohort-specific features), which should average out at the cohort level. Further, as in (18), we were only able to correlate *absolute* LFCs of input and output genes (Figure 4B), likely due to limitations in RNA-seq technology and the analysis, and/or missing information on the TFN. Nevertheless, the present information on input TFs and their regulatory effect sufficed to relate the TFN with the genes' response.

From the RNA-seq data on three time points, we provided evidence that both the TFN and the RNAP affect the results at mid-term ( $\sim 180$  min), and not before that. In addition, while other factors also influenced genes' behaviour at mid-term, including single-gene features, they only had minor, local effects. In detail, first, we could not find evidence of GRs (including  $\sigma^{38}$ ) and (p)ppGpp being material in the global mid-term behaviour (although



**Figure 7.** Genome-wide effects of increasing medium richness. (A) RNAP concentrations relative to the control, estimated from  $\mu_{RNAPFITC-H}$  divided by mean cell area ( $\mu_{cellarea}$ , used as a proxy for cell volume). (B) Venn diagrams of the DEG. (C)  $\mu_{|LFC|}$  of  $N$  genes with  $K_{TF}$  equal to and larger than 0, following each medium shift. Above each bar are the number of DEG. (D)  $\mu_{|LFC|}$  as a function of  $\mu_{|b|}$  after the growth-medium shifts.  $\mu_{|LFC|}$  obtained using the ensemble approach (Supplementary Results section *Estimation of the expected  $\mu_{K_{TF}}$  and  $\mu_{|b|}$  using an ensemble approach*, Supplementary Figure S33). Each blue cross is the average outcome from up to 24536 cohorts of 10 genes. In (A) and (C), the black error bars are the SEM and the red error bars are the 95% CB of the SEM. In (D), the small error bars are the SEM (most not visible).

(p)ppGpp may be significant in the short-term response). Second, we excluded the medium as directly influencing RNA abundances. Third, we excluded global network parameters, other than  $K_{TF}$ , as being influential as well (none of them correlated to single-gene responses). Fourth, we did not find evidence for significant translational or post-translational regulation. Namely, RNA and protein abundances correlated well, and so did the RNA levels of input TFs and of output genes. Finally, sRNAs did not respond atypically to the RNAP shifts neither in the short-term, nor in the mid-term.

We have made six key observations on the influence of the logic and topology of the TFN on the mid-term response. First, genes without input TFs were less responsive. Second, the  $|LFC|$  of input and output genes correlated positively. Thus, we argue that, on average, input TFs enhanced the  $|LFC|$  of individual genes. Third, only nearest neighbour genes in the TFN consistently correlated in  $|LFC|$ 's. Thus, either the effects of the shift in RNAP only reached nearest neighbour genes or they 'dissipated' beyond that. Since the correlations between nearest neighbours were weaker in the short-term than in the mid-term, the first possibility is more

likely. This observation also suggests that there is a degree of genome-wide homogeneity in how long input TF abundance take to change (likely due to physical limitations on the rates constants controlling bacterial gene expression). This agrees with the constraints on timing variability reported in (6). Fourth, the behaviour was orderly (rather than chaotic), with most genes responsive to the weak perturbations also responding to the stronger perturbations. This suggests the existence of features (on genes and/or the TFN) affecting the responsiveness (Supplementary Figure S34). Similarly, there is a good overlap between the sets of genes responsive in the short- and in the mid-term, but weak overlap to those responsive prior to the perturbation (Supplementary Figure S35). Fifth, on average, as  $K_{TF}$  increased, the correlation between the input and each output gene decreased. This is likely unavoidable and may be a limiting factor in how many input TFs genes can have. Finally, it is  $\mu_{|b|}$  that (partially) controls the genes' responsiveness to the stress, while the apparent relationship between  $\mu_{K_{TF}}$  and  $\mu_{|LFC|}$  is due to the linear correlation between  $\mu_{|b|}$  and  $\mu_{K_{TF}}$ . Nevertheless, the possible values of  $\mu_{|b|}$  are limited by the values of  $\mu_{K_{TF}}$ .



These observations provide direct empirical evidence that the genome-wide, mid-term, transcriptional stress responsiveness of *E. coli* depends on a global topological feature of its TFN. Namely, by exploiting the known features of the TFN of *E. coli*, we showed that the bias in the regulatory effect of input TFs of gene cohorts,  $\mu_{|b|}$ , acts as a major determinant of their mean response to a stress. So far, the existence of influence of global topological features has only been supported by theoretical models, e.g. (82,83), and by indirect empirical evidence, i.e. by the observation that the abundance of input TFs correlates with the activity of their direct output genes (18,20–22,70,84). Because of these observations, it has been assumed that the topology and logic of TFNs should play a role. Here, we provided direct evidence of this, i.e. that the TFN topology and logic are major contributors to a global mid-term response. Our approach, relating genome-wide dynamics and global topological features, should now be applied to other genome-wide stresses, as well as to the genomes of other organisms, when data on their TFN becomes sufficient to permit such an analysis.

Expanding this research may inform on how to improve the robustness and plasticity of synthetic circuits. Further, as suggested in (20), bacteria subjected to stress, rather than under optimal conditions, may be a better proxy of their state when infecting a host. Thus, imposing stresses may be a valuable strategy to identify new target genes for antibiotics for disrupting bacterial adaptability to new conditions. The use of medium dilution as a genome-wide stress is a good proxy for nutrient imbalance, and we identified ~ 900 responsive genes, even for moderate nutritional stress, of which only 58 are essential under optimal conditions. Plausibly, some of the responsive genes, particularly those responsive to all 3 medium dilutions, may be essential to adapt to poorer media, and thus are potential new drug targets. Conversely, it may be possible to tune these genes to assist in the performance of metabolic tasks, without disturbing the basic biology of the cells. As such, they are promising targets for modifications that could improve the yield and sustainability of bio-industrial processes.

Finally, our findings can be used to develop new large-scale, dynamic models of gene networks. These models should be able to predict short- and mid-term transcriptional responses of gene cohorts to genome-wide perturbations of transcription activities. The short-term responses should be mostly controlled by single-gene features (e.g. RNAP-promoter binding affinity). Meanwhile, the mid-term responses should be heavily influenced by the topology and logic of the TFN. The new dynamic models could be developed starting from the schematic and predictive models in Figures 1B and C, respectively, using empirical kinetic parameters of bacterial RNA and protein production and degradation.

Such dynamic models could then be used to predict how natural TFNs perform complex transcriptional programs, and how these programs can be modified to achieve desired goals. Interesting models involving many genes include models of programs responsive to environmental shifts, antibiotics, etc. These models could assist in identifying critical elements of the TFN during short-, mid-, and potentially long-term stress responses. These efforts should be facilitated by the ongoing information gathering

on single-gene features (34,85–87), including on microorganisms other than *E. coli*.

## DATA AVAILABILITY

RNA-seq \*.fastq data (trimmed) and processed RNA-seq data are deposited in NCBI GEO with accession code GSE178281 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE178281>). The raw data of the control condition (180 min, LB<sub>1.0x</sub>) was also used in (72) (GSE183139). In Dryad, we deposited a package with flow-cytometry, microscopy, spectrophotometry, and western blot data (DOI: 10.5061/dryad.wh70rxwnp). The package also has two \*.xlsx files informing on the genes TFN topological and logical parameters, RNA-seq expression, global network topological features, gene-gene interactions, and the lists of pairs of genes separated by path lengths from 1 to 8.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

*Author contributions:* A.S.R., B.L.B.A. and S.M.D.O. conceived the study and A.S.R. and B.L.B.A. directed it. B.L.B.A. planned, directed and executed data analysis, to which C.S.D.P. and I.S.C.B. contributed to. M.N.M.B., A.S.R. and B.L.B.A. interpreted and integrated the data. V.C., S.D. and V.K. performed measurements. A.H., H.T.J. and E.D. assisted the data analysis. A.H., J.L.P., O.P.S., A.G., M.N., J.K. and P.A. assisted on the RNA-seq data analysis. B.L.B.A. and A.S.R., assisted by H.T.J., drafted the manuscript, supplementary information, figures and tables, which were revised by all co-authors. M.N.M.B., V.C. and S.D. contributed equally as second author and have the right to list their name first.

## FUNDING

Jane and Aatos Erkkö Foundation [10-10524-38 to A.S.R.]; Finnish Cultural Foundation [50201300 to S.D., and 00200193, 00212591 to I.S.C.B.]; EDUFI Fellowship [TM-19-11105 to S.D.]; Suomalainen Tiedekatemia (to C.S.D.P.); Tampere University Graduate Program [to V.C., M.N.M.B., B.L.B.A.]; TTU development program 2016–2022 [2014-2020.4.01.16-0032 to O.P.S.]; Academy of Finland [307856 and 323576 to P.A., 322927 to A.H., and 272376 and 256615 to H.T.J.]. Funding for open access and for article processing charges provided by Tampere University Library [to B.L.B.A.]. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

*Conflict of interest statement.* None declared.

## REFERENCES

- Jacob, F. and Monod, J. (1961) Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.*, **3**, 318–356.
- Mejía-Almonte, C., Busby, S.J.W., Wade, J.T., van Helden, J., Arkin, A.P., Stormo, G.D., Eilbeck, K., Palsson, B.O., Galagan, J.E. and Collado-Vides, J. (2020) Redefining fundamental concepts of transcription initiation in bacteria. *Nat. Rev. Genet.*, **21**, 699–714.

3. Sanchez-Vazquez,P, Dewey,C.N., Kitten,N., Ross,W. and Gourse,R.L. (2019) Genome-wide effects on escherichia coli transcription from ppGpp binding to its two sites on RNA polymerase. *Proc. Natl. Acad. Sci. U.S.A.*, **116**, 8310–8319.
4. Kao,K.C., Yang,Y.L., Boscolo,R., Sabatti,C., Roychowdhury,V. and Liao,J.C. (2004) Transcriptome-based determination of multiple transcription regulator activities in escherichia coli by using network component analysis. *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 641–646.
5. Jozefczuk,S., Klie,S., Catchpole,G., Szymanski,J., Cuadros-Inostroza,A., Steinhäuser,D., Selbig,J. and Willmitzer,L. (2010) Metabolic and transcriptomic stress response of escherichia coli. *Mol. Syst. Biol.*, **6**, 364.
6. Mitosch,K., Rieckh,G. and Bollenbach,T. (2019) Temporal order and precision of complex stress responses in individual bacteria. *Mol. Syst. Biol.*, **15**, e8470.
7. Phadtare,S. and Inouye,M. (2004) Genome-wide transcriptional analysis of the cold shock response in wild-type and cold-sensitive, quadruple-csp-deletion strains of escherichia coli. *J. Bacteriol.*, **186**, 7007–7014.
8. Dash,S., Palma,C.S.D., Baptista,I.S.C., Bahrudeen,M.N.M., Almeida,B.L.B., Chauhan,V., Jagadeesan,R. and Ribeiro,A.S. (2021) Positive supercoiling buildup is a trigger of *e. coli*'s short-term response to cold shock. bioRxiv doi: <https://doi.org/10.1101/2021.12.22.473827>, 23 December 2021, preprint: not peer reviewed.
9. Rau,M.H., Calero,P., Lennen,R.M., Long,K.S. and Nielsen,A.T. (2016) Genome-wide escherichia coli stress response and improved tolerance towards industrially relevant chemicals. *Microb. Cell Fact.*, **15**, 176.
10. Kochanowski,K., Gerosa,L., Brunner,S.F., Christodoulou,D., Nikolaev,Y.V. and Sauer,U. (2017) Few regulatory metabolites coordinate expression of central metabolic genes in escherichia coli. *Mol. Syst. Biol.*, **13**, 903.
11. Engl,C., Jovanovic,G., Brackston,R.D., Kotta-Loizou,I. and Buck,M. (2020) The route to transcription initiation determines the mode of transcriptional bursting in *e. coli*. *Nat. Commun.*, **11**, 2422.
12. Deter,H.S., Hossain,T. and Butzin,N.C. (2021) Antibiotic tolerance is associated with a broad and complex transcriptional response in *e. coli*. *Sci. Rep.*, **11**, 6112.
13. Lal,A., Dhar,A., Trostel,A., Kouzine,F., Seshasayee,A.S. and Adhya,S. (2016) Genome scale patterns of supercoiling in a bacterial chromosome. *Nat. Commun.*, **7**, 11055.
14. Storz,G., Vogel,J. and Wassarman,K.M. (2011) Regulation by small RNAs in bacteria: expanding frontiers. *Mol. Cell*, **43**, 880–891.
15. Keren,L., Zackay,O., Lotan-Pompan,M., Barenholz,U., Dekel,E., Sasson,V., Aidelberg,G., Bren,A., Zeevi,D., Weinberger,A. et al. (2013) Promoters maintain their relative activity levels under different growth conditions. *Mol. Syst. Biol.*, **9**, 701.
16. Gerosa,L., Kochanowski,K., Heinemann,M. and Sauer,U. (2013) Dissecting specific and global transcriptional regulation of bacterial gene expression. *Mol. Syst. Biol.*, **9**, 658.
17. Berthoumieux,S., de Jong,H., Baptist,G., Pinel,C., Ranquet,C., Ropers,D. and Geiselmann,J. (2013) Shared control of gene expression in bacteria by transcription factors and global physiology of the cell. *Mol. Syst. Biol.*, **9**, 634.
18. Larsen,S.J., Röttger,R., Schmidt,H.H.H.W. and Baumbach,J. (2019) *E. coli* gene regulatory networks are inconsistent with gene expression data. *Nucleic Acids Res.*, **47**, 85–92.
19. Brooks,A.N., Reiss,D.J., Allard,A., Wu,W.J., Salvanha,D.M., Plaisier,C.L., Chandrasekaran,S., Pan,M., Kaur,A. and Baliga,N.S. (2014) A system-level model for the microbial regulatory genome. *Mol. Syst. Biol.*, **10**, 740.
20. Côté,J.P., French,S., Gehrke,S.S., MacNair,C.R., Mangat,C.S., Bharat,A. and Brown,E.D. (2016) The genome-wide interaction network of nutrient stress genes in escherichia coli. *Mbio*, **7**, e01714-16.
21. Fang,X., Sastry,A., Mih,N., Kim,D., Tan,J., Yurkovich,J.T., Lloyd,C.J., Gao,Y., Yang,L. and Pálsson,B.O. (2017) Global transcriptional regulatory network for escherichia coli robustly connects gene expression to transcription factor activities. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, 10286–10291.
22. Urcheguía,A., Galbusera,L., Chauvin,D., Bellement,G., Julou,T. and van Nimwegen,E. (2021) Genome-wide gene expression noise in escherichia coli is condition-dependent and determined by propagation of noise through the regulatory network. *PLoS Biol.*, **19**, e3001491.
23. Bremer,H.D.P.P. and Dennis,P.P. (1996) Modulation of chemical composition and other parameters of the cell by growth rate. In: Neidhardt,F.C. (ed). *Escherichia Coli and Salmonella: Cellular and Molecular Biology*. ASM Press, Washington, D.C., pp. 1553–1569.
24. McClure,W.R. (1985) Mechanism and control of transcription initiation in prokaryotes. *Annu. Rev. Biochem.*, **54**, 171–204.
25. Taniguchi,Y., Choi,P.J., Li,G.W., Chen,H., Babu,M., Hearn,J., Emili,A. and Xie,X.S. (2010) Quantifying *e. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science*, **329**, 533–538.
26. Lloyd-Price,J., Startceva,S., Kandavalli,V., Chandraseelan,J.G., Goncalves,N., Oliveira,S.M., Häkkinen,A. and Ribeiro,A.S. (2016) Dissecting the stochastic transcription initiation process in live escherichia coli. *DNA Res.*, **23**, 203–214.
27. Kandavalli,V.K., Tran,H. and Ribeiro,A.S. (2016) Effects of  $\sigma$  factor competition are promoter initiation kinetics dependent. *Biochim. Biophys. Acta*, **1859**, 1281–1288.
28. Bernstein,J.A., Khodursky,A.B., Lin,P.H., Lin-Chao,S. and Cohen,S.N. (2002) Global analysis of mRNA decay and abundance in escherichia coli at single-gene resolution using two-color fluorescent DNA microarrays. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 9697–9702.
29. Maurizi,M.R. (1992) Proteases and protein degradation in escherichia coli. *Experientia*, **48**, 178–201.
30. Heibisch,E., Knebel,J., Landsberg,J., Frey,E. and Leisner,M. (2013) High variation of fluorescence protein maturation times in closely related escherichia coli strains. *PLoS One*, **8**, e75991.
31. Balleza,E., Kim,J.M. and Cluzel,P. (2018) Systematic characterization of maturation time of fluorescent proteins in living cells. *Nat. Methods*, **15**, 47–51.
32. deHaseth,P.L., Zupancic,M.L. and Record,M.T. (1998) RNA polymerase-promoter interactions: the comings and goings of RNA polymerase. *J. Bacteriol.*, **180**, 3019–3025.
33. Mazumder,A. and Kapanidis,A.N. (2019) Recent advances in understanding  $\sigma$ 70-dependent transcription initiation mechanisms. *J. Mol. Biol.*, **431**, 3947–3959.
34. Santos-Zavaleta,A., Salgado,H., Gama-Castro,S., Sánchez-Pérez,M., Gómez-Romero,L., Ledezma-Tejeda,D., García-Sotelo,J.S., Alquicira-Hernández,K., Muñoz-Rascado,L.J., Peña-Loredo,P. et al. (2019) RegulonDB v 10.5: tackling challenges to unify classic and high throughput knowledge of gene regulation in *e. coli* K-12. *Nucleic Acids Res.*, **47**, D212–D220.
35. Patange,O., Schwall,C., Jones,M., Villava,C., Griffith,D.A., Phillips,A. and Locke,J.C.W. (2018) Escherichia coli can survive stress by noisy growth modulation. *Nat. Commun.*, **9**, 5333.
36. Zaslaver,A., Bren,A., Ronen,M., Itzkovitz,S., Kikoin,I., Shavit,S., Liebermeister,W., Surette,M.G. and Alon,U. (2006) A comprehensive library of fluorescent transcriptional reporters for escherichia coli. *Nat. Methods*, **3**, 623–628.
37. Chantzoura,E. and Kaji,K. (2017) Chapter 10 - flow cytometry. In: Jalali,M., Saldanha,F.Y. and Jalali,M. (eds). *Basic Science Methods for Clinical Researchers*. Academic Press, Boston, pp. 173–189.
38. Bolger,A.M., Lohse,M. and Usadel,B. (2014) Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics*, **30**, 2114–2120.
39. Langmead,B. and Salzberg,S.L. (2012) Fast gapped-read alignment with bowtie 2. *Nat. Methods*, **9**, 357–359.
40. Liao,Y., Smyth,G.K. and Shi,W. (2019) The r package rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Res.*, **47**, e47.
41. Love,M.I., Huber,W. and Anders,S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, **15**, 550.
42. Benjamini,Y. and Hochberg,Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B (Methodological)*, **57**, 289–300.
43. Evans,C., Hardin,J. and Stoebel,D.M. (2018) Selecting between-sample RNA-Seq normalization methods from the perspective of their assumptions. *Brief. Bioinform.*, **19**, 776–792.
44. Spaniol,V., Wyder,S. and Aebi,C. (2013) RNA-Seq-based analysis of the physiologic cold shock-induced changes in moraxella catarrhalis gene expression. *PLoS One*, **8**, e68298.
45. Hazen,T.H., Daugherty,S.C., Shetty,A., Mahurkar,A.A., White,O., Kaper,J.B. and Rasko,D.A. (2015) RNA-Seq analysis of isolate- and

- growth phase-specific differences in the global transcriptomes of enteropathogenic *Escherichia coli* prototype isolates. *Front Microbiol.*, **6**, 569.
46. Yung, P.Y., Grasso, L.L., Mohidin, A.F., Acerbi, E., Hinks, J., Seviour, T., Marsili, E. and Lauro, F.M. (2016) Global transcriptomic responses of *Escherichia coli* K-12 to volatile organic compounds. *Sci. Rep.*, **6**, 19899.
  47. Yanofsky, C. (1981) Attenuation in the control of expression of bacterial operons. *Nature*, **289**, 751–758.
  48. Proshkin, S., Rahmouni, A.R., Mironov, A. and Nudler, E. (2010) Cooperation between translating ribosomes and RNA polymerase in transcription elongation. *Science*, **328**, 504–508.
  49. Dahan, O., Gingold, H. and Pilpel, Y. (2011) Regulatory mechanisms and networks couple the different phases of gene expression. *Trends Genet.*, **27**, 316–322.
  50. Albert, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. and Walter, P. (2008) In: *Molecular Biology of the Cell*. Garland Science, NY.
  51. Doncheva, N.T., Assenov, Y., Domingues, F.S. and Albrecht, M. (2012) Topological analysis and interactive visualization of biological networks and protein structures. *Nat. Protoc.*, **7**, 670–685.
  52. Bahrudeen, M.N.M., Chauhan, V., Palma, C.S.D., Oliveira, S.M.D., Kandavalli, V.K. and Ribeiro, A.S. (2019) Estimating RNA numbers in single cells by RNA fluorescent tagging and flow cytometry. *J. Microbiol. Methods*, **166**, 105745.
  53. McDonald, J.H. (2009) In: *Analysis of covariance. In: Handbook of Biological Statistics*. Sparky House Publishing, Baltimore, pp. 232–238.
  54. Farewell, A., Kvint, K. and Nyström, T. (1998) Negative regulation by rpos: a case of sigma factor competition. *Mol. Microbiol.*, **29**, 1039–1051.
  55. Chang, D.E., Smalley, D.J. and Conway, T. (2002) Gene expression profiling of *Escherichia coli* growth transitions: an expanded stringent response model. *Mol. Microbiol.*, **45**, 289–306.
  56. Tani, T.H., Khodursky, A., Blumenthal, R.M., Brown, P.O. and Matthews, R.G. (2002) Adaptation to famine: a family of stationary-phase genes revealed by microarray analysis. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 13471–13476.
  57. Baptista, I.S.C., Kandavalli, V., Chauhan, V., Bahrudeen, M.N.M., Almeida, B.L.B., Palma, C.S.D., Dash, S. and Ribeiro, A.S. (2022) Sequence-dependent model of genes with dual  $\sigma$  factor preference. *Biochim. Biophys. Acta Gene Regul. Mech.*, **1865**, 194812.
  58. Klumpp, S., Zhang, Z. and Hwa, T. (2009) Growth rate-dependent global effects on gene expression in bacteria. *Cell*, **139**, 1366–1375.
  59. Huh, D. and Paulsson, J. (2011) Non-genetic heterogeneity from stochastic partitioning at cell division. *Nat. Genet.*, **43**, 95–100.
  60. Lange, R. and Hengge-Aronis, R. (1994) The cellular concentration of the sigma s subunit of RNA polymerase in *Escherichia coli* is controlled at the levels of transcription, translation, and protein stability. *Genes Dev.*, **8**, 1600–1612.
  61. Gruber, T.M. and Gross, C.A. (2003) Multiple sigma subunits and the partitioning of bacterial transcription space. *Annu. Rev. Microbiol.*, **57**, 441–466.
  62. Pratt, L.A. and Silhavy, T.J. (1998) Crl stimulates RpoS activity during stationary phase. *Mol. Microbiol.*, **29**, 1225–1236.
  63. Nyström, T. (2004) Growth versus maintenance: a trade-off dictated by RNA polymerase availability and sigma factor competition. *Mol. Microbiol.*, **54**, 855–862.
  64. Buhler, J.M., Riva, M., Mann, C., Thuriaux, P., Memet, S., Micouin, J.Y., Treich, I., Mariotte, S., Sentenac, A., Reznikoff, W.S. et al. (1987) In: *RNA Polymerase and the Regulation of Transcription*. Elsevier Science Publishing Co., NY, pp. 25–36.
  65. Nishiuchi, Y., Inui, T., Nishio, H., Bódi, J., Kimura, T., Tsuji, F.I. and Sakakibara, S. (1998) Chemical synthesis of the precursor molecule of the aequorea green fluorescent protein, subsequent folding, and development of fluorescence. *Proc. Natl. Acad. Sci. U.S.A.*, **95**, 13549–13554.
  66. Severinov, K., Mooney, R., Darst, S.A. and Landick, R. (1997) Tethering of the large subunits of *Escherichia coli* RNA polymerase. *J. Biol. Chem.*, **272**, 24137–24140.
  67. Lutz, R. and Bujard, H. (1997) Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and araC/11-12 regulatory elements. *Nucleic Acids Res.*, **25**, 1203–1210.
  68. Gardner, T.S., Cantor, C.R. and Collins, J.J. (2000) Construction of a genetic toggle switch in *Escherichia coli*. *Nature*, **403**, 339–342.
  69. Lutz, R., Lozinski, T., Ellinger, T. and Bujard, H. (2001) Dissecting the functional program of *Escherichia coli* promoters: the combined mode of action of lac repressor and AraC activator. *Nucleic Acids Res.*, **29**, 3873–3881.
  70. Martínez-Antonio, A. and Collado-Vides, J. (2003) Identifying global regulators in transcriptional regulatory networks in bacteria. *Curr. Opin. Microbiol.*, **6**, 482–489.
  71. Balderas-Martínez, Y.I., Savageau, M., Salgado, H., Pérez-Rueda, E., Morett, E. and Collado-Vides, J. (2013) Transcription factors in *Escherichia coli* prefer the holo conformation. *PLoS One*, **8**, e65723.
  72. Chauhan, V., Bahrudeen, M.N.M., Palma, C.S.D., Baptista, I.S.C., Almeida, B.L.B., Dash, S., Kandavalli, V. and Ribeiro, A.S. (2022) Analytical kinetic model of native tandem promoters in *E. coli*. *PLoS Comput. Biol.*, **18**, e1009824.
  73. Mushegian, A.R. and Koonin, E.V. (1996) Gene order is not conserved in bacterial evolution. *Trends Genet.*, **12**, 289–290.
  74. Dandekar, T., Snel, B., Huynen, M. and Bork, P. (1998) Conservation of gene order: a fingerprint of proteins that physically interact. *Trends Biochem. Sci.*, **23**, 324–328.
  75. Wells, J.N., Bergendahl, L.T. and Marsh, J.A. (2016) Operon gene order is optimized for ordered protein complex assembly. *Cell Rep.*, **14**, 679–685.
  76. Ozsolak, F. and Milos, P.M. (2011) RNA sequencing: advances, challenges and opportunities. *Nat. Rev. Genet.*, **12**, 87–98.
  77. Haas, B.J., Chin, M., Nusbbaum, C., Birren, B.W. and Livny, J. (2012) How deep is deep enough for RNA-Seq profiling of bacterial transcriptomes. *BMC Genomics*, **13**, 734.
  78. Elowitz, M.B., Levine, A.J., Siggia, E.D. and Swain, P.S. (2002) Stochastic gene expression in a single cell. *Science*, **297**, 1183–1186.
  79. Engl, C. (2019) Noise in bacterial gene expression. *Biochem. Soc. Trans.*, **47**, 209–217.
  80. Dobrin, R., Beg, Q.K., Barabási, A.L. and Oltvai, Z.N. (2004) Aggregation of topological motifs in the *Escherichia coli* transcriptional regulatory network. *BMC Bioinf.*, **5**, 10.
  81. Martínez-Antonio, A. (2011) *Escherichia coli* transcriptional regulatory network. *Network Biology*, **1**, 21.
  82. Ribeiro, A.S., Kauffman, S.A., Lloyd-Price, J., Samuelsson, B. and Socolar, J.E. (2008) Mutual information in random boolean models of regulatory networks. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.*, **77**, 011901.
  83. Samuelsson, B. and Socolar, J.E. (2006) Exhaustive percolation on random networks. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.*, **74**, 036113.
  84. Gutiérrez-Ríos, R.M., Rosenblueth, D.A., Loza, J.A., Huerta, A.M., Glasner, J.D., Blattner, F.R. and Collado-Vides, J. (2003) Regulatory network of *Escherichia coli*: consistency between literature knowledge and microarray profiles. *Genome Res.*, **13**, 2435–2443.
  85. Reed, J.L., Vo, T.D., Schilling, C.H. and Palsson, B.O. (2003) An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). *Genome Biol.*, **4**, R54.
  86. Ishihama, A., Shimada, T. and Yamazaki, Y. (2016) Transcription profile of *Escherichia coli*: genomic SELEX search for regulatory targets of transcription factors. *Nucleic Acids Res.*, **44**, 2058–2074.
  87. Gao, Y., Lim, H.G., Verkler, H., Szubin, R., Quach, D., Rodionova, I., Chen, K., Yurkovich, J.T., Cho, B.K. and Palsson, B.O. (2021) Unraveling the functions of uncharacterized transcription factors in *Escherichia coli* using chip-exo. *Nucleic Acids Res.*, **49**, 9696–9710.