Recent developments in high-throughput reverse genetics have revolutionized our ability to map gene function and interactions. The power of these approaches depends on their ability to identify functionally associated genes, which elicit similar phenotypic changes across several perturbations (chemical, environmental or genetic) when knocked out. However, owing to the large number of perturbations, these approaches have been limited to growth or morphological readouts. Here we use a high-content biochemical readout, thermal proteome profiling, to measure the proteome-wide protein abundance and thermal stability in response to 121 genetic perturbations in *Escherichia coli*. We show that thermal stability, and therefore the state and interactions of essential proteins, is commonly modulated, raising the possibility of studying a protein group that is particularly inaccessible to genetics. We find that functionally associated proteins have coordinated changes in abundance and thermal stability across perturbations, owing to their co-regulation and physical interactions (with proteins, metabolites or cofactors). Finally, we provide mechanistic insights into previously determined growth phenotypes that go beyond the deleted gene. These data represent a rich resource for inferring protein functions and interactions.

Understanding the function of genes is one of the main goals of molecular biology. Although genetic approaches provide insights into protein function and interactions, biochemical readouts bring us closer to their molecular mechanism. Mass spectrometry has enabled a view of the entire proteome and, when coupled with traditional biochemistry tools, such as affinity purification or size exclusion chromatography, can directly detect protein–protein interactions. Although powerful, these approaches are performed after cell lysis, which can alter the protein environment and interactions. Recently, thermal proteome profiling (TPP) has been developed, which couples the cellular thermal shift assay (CETSA) with multiplexed quantitative proteomics. Protein thermal stability offers insights into protein state in situ, because it reflects interactions with metabolites, other proteins and nucleic acids, as well as the post-translational makes-up. Here, we combine reverse genetics with TPP in *Escherichia coli* to profile the effect of genetic perturbations on protein abundance and thermal stability.

### High-throughput thermal proteome profiling

We used two-dimensional thermal proteome profiling (2D-TPP) in 121 *E. coli* strains (most of which were single-gene deletion mutants from the Keio library) to perturb diverse cellular processes (Extended Data Fig. 2a), by leveraging chemical genetics data. Each mutant was grown in duplicate to study changes in protein abundance and thermal stability across 10 temperatures (Fig. 1a). This generated 2,420 samples (121 mutants × 2 replicates × 10 temperatures) that were multiplexed with tandem mass tags and measured by quantitative mass spectrometry-based proteomics (Supplementary Data 1). In total, we detected 2,586 proteins with at least 2 unique peptides (Extended Data Fig. 1a, Supplementary Data 2). For each protein in each mutant, we calculated the ratio of the signal intensity to the median signal intensity of the respective protein in the same mass spectrometry run (Supplementary Data 3). Measurements were largely consistent across biological replicates (Extended Data Fig. 1b–d), with differences in clone behaviour reflecting biological phenomena, such as mutations that activate the flagellar master regulator (FlhDC) in only one of the clones (Extended Data Fig. 1e, f, Supplementary Data 4).

Abundance (corresponding to the average changes at the two lowest temperatures and thermal stability (corresponding to changes remaining at higher temperatures after correcting for protein abundance changes) were determined for 1,764 proteins across the genetic background (Fig. 1b, Supplementary Data 5). We observed significant changes in 1,213 proteins in at least one mutant (|z-score| > 1.96 and q-value < 0.05) (Fig. 1c), with 840 proteins affected in abundance, 886 proteins in thermal stability, and 513 proteins in both. However, abundance and thermal stability were only weakly anti-correlated (r = −0.12) (Extended Data Fig. 1g).

Several mechanisms can lead to these changes, such as the deletion of protein complex members leading to the thermal destabilization of other proximal complex members (Extended Data Fig. 2b), or regulatory
mechanisms of envelope stress responses (Extended Data Fig. 3, Supplementary Discussion). Proteins were also affected in the absence of cofactors, as illustrated by the thermal destabilization of iron-sulfur cluster binding proteins in ΔiscA, ΔiscS and ΔiscU (Fig. 1d, Extended Data Fig. 4a), or the thermal destabilization of the periplasmic copper oxidase CueO in ΔtatB (Extended Data Fig. 4b). CueO is translocated from the cytosol to the periplasm after recognition of its signal peptide by the Tat system. By deleting the signal peptide (Δ28-CueO), we could trap CueO in the cytosol in wild-type cells (Extended Data Fig. 4c), and phenocopy the thermal destabilization observed in ΔtatB (Extended Data Fig. 4d–f). Notably, Δ28-CueO was thermostabilized by the addition of copper in lysate, which suggests that the lack of copper in the cytoplasm prevents CueO from being thermostabilized (Extended Data Fig. 4g, Supplementary Discussion).

In summary, we generated a comprehensive dataset of changes in protein abundance and thermal stability in more than 100 E. coli mutants that can be explored at http://ecoliTPP.shiny.embl.de. Nearly 70% of the proteins were altered in at least one perturbation (Extended Data Fig. 1h, i), and abundance and thermal stability were largely orthogonal. The proteome changes observed can help in investigating the physiological state of each mutant.

**Essential proteins change in thermal stability**

Because the function of essential genes (that is, genes that cannot be deleted) is difficult to study by genetic approaches, we explored how essential genes behaved across the genetic perturbations included in this study. We observed that proteins coded by essential genes were generally more abundant (Fig. 2b), but less often altered in their abundance than those coded by non-essential genes (Fig. 2a). However, essential proteins were more often hits in thermal stability than non-essential ones (Fig. 2a), which suggests that their activity or interactions might be modulated in some genetic backgrounds.

**Fig. 1** TPP of 121 E. coli mutants. a, Experimental layout for TPP experiments. Two clones of each mutant were grown to exponential phase and subjected to a short heat treatment. Cells were lysed and the soluble protein fraction at each temperature was analysed by mass spectrometry (MS)-based quantitative proteomics, using the multiplexing strategy depicted in panel at the right (Methods). b, Heat map of abundance and thermal stability of each protein (rows) in each mutant (columns). c, Rarefaction analysis of the fraction of the proteome affected as a function of the number of genetic perturbations probed. Accumulation curves obtained after 50 random subsamples without replacement, in which the line represents the mean of the permutations and shaded area the standard deviation. d, Zoomed inset from b demonstrates thermal destabilization of iron-sulfur cluster containing proteins in iron-sulfur cluster biosynthesis mutants.
To gain insights into the consequences of changes in thermal stability of essential proteins, we used CRISPRi31,32 to reduce the levels of FtsK (a cell division DNA translocase) and ParC (a subunit of topoisomerase IV) in different genetic backgrounds. Reducing the levels of FtsK (by around sixfold) or ParC (by eightfold) (Fig. 2c) only mildly affected cell growth in wild-type cells (Fig. 2d, e). However, cells could not tolerate the depletion of the essential proteins in mutants in which these were affected in thermal stability (for example, ΔclpS for ParC, or ΔphoP for both FtsK and ParC) (Fig. 2d, e, Extended Data Fig. 5). Notably, in mutants for which we did not observe changes in thermal stability in the essential proteins, the growth phenotype was similar to wild-type cells (with the exception of ΔamiA and ΔenvC, both affecting cell division, and the latter being a genetic perturbation that by itself causes growth defects) (Fig. 2d, e). We confirmed by proteomics that the essential protein downregulation remained similar in all mutants tested (Fig. 2c).

Overall, we observed that the levels of essential proteins are high and rarely modulated, consistent with their housekeeping roles. Although bacterial cells can tolerate fluctuations in the levels of these proteins, they seem to prefer maintaining them above the levels required for optimal growth33. By contrast, the thermal stability of essential proteins, a trait that is impervious to expression proteomics, was regularly affected. Notably, cells became more vulnerable to changes in levels of essential proteins in conditions that affected their thermal stability. Hence, cells may maintain higher levels of essential proteins to buffer changes in their activity across different conditions. Overall, this synthetic lethality of essential and non-essential genes could provide new paths for combinatorial drug therapies.

**Functionally related proteins are co-regulated**
We assessed which pairs of proteins co-changed across the 121 genetic perturbations at the 10 different temperatures, by calculating Spearman’s rank correlation (rS) for all protein pairs (Fig. 3a, Supplementary Data 6). As previously shown for gene and protein co-expression analysis14,34–37, we observed an enrichment of strong correlations for proteins with known biological associations (Extended Data Fig. 6a). Notably, our ability to measure protein thermal stability further contributed to capturing functional associations. For example, the essential core subunits of RNA polymerase (RNAP; RpoA, RpoB and RpoC) were correlated (rS = 0.68) (Fig. 3b) mostly due to changes at higher temperatures (Fig. 3e). Although it is currently unclear how these changes in thermal stability link to RNAP states (in eukaryotes, increase in RNAP II thermal stability correlates with DNA-bound active holoenzyme15), these changes were unrelated to upregulation of the flagellar sigma factor (FlaA) in a large number of mutants38—as we did not observe a correlation between flagellar protein abundance (for example, FlIC) and RNAP thermal stability (for example, RpoB; rS = −0.22, n = 121 mutants). Other functionally related proteins also clustered closely, such as all the enzymes of the t-histidine biosynthesis pathway (Fig. 3c) or proteins involved in protein folding (Fig. 3d). A receiver operating characteristic (ROC) analysis revealed that strongly correlated protein pairs captured previously described functional associations (Fig. 3f), particularly for proteins expressed from the...
Fig. 3 | Co-changes in protein abundance and thermal stability are strong identifiers of functional relationships. a. Heat map of Spearman’s rank correlation of all protein pairs using all the acquired data across the 121 genetic perturbations at the ten different temperatures. b – d. Zoomed insets demonstrate co-clustering of functionally-related proteins, for members of the RNAP (b), l-histidine biosynthesis (c) and proteins involved in protein folding (d). e. Example of protein pair (two subunits of RNA polymerase) co-changing in its thermal stability across all the genetic perturbations profiled in this study. Each data point corresponds to the log₂-transformed fold change relative to control in one of the genetic perturbations at one temperature (colour-coded). f. ROC analysis based on the decreasing absolute Spearman’s rank correlation compared to known operons, protein complexes and metabolic pathways.

same operon (area under the ROC (AUROC) = 0.86; strongly driven by protein abundance changes) (Extended Data Fig. 6c), part of the same protein complex (AUROC = 0.81; mostly driven by protein abundance changes, as 38% of proteins that belong to the same complex are also in the same operon) (Extended Data Fig. 6d), or belonging to the same metabolic pathway (AUROC = 0.70; driven to a large extent by thermal stability) (Extended Data Fig. 6e–h, Supplementary Discussion). Furthermore, we compared our data with STRING associations and found that the higher the confidence of interactions in STRING the better they were recapitulated by our data (Extended Data Fig. 6b). For the highest confidence interactions (combined STRING score ≥ 0.999; n = 1,493), we obtained an AUROC of 0.90, recovering 47% true-positive interactions at 1% false-positive rate (corresponding to $r_S \geq 0.45$).

In addition to the overall strong correlation of proteins belonging to the same complexes or metabolic pathways, we also captured complex (Extended Data Fig. 7; see Supplementary Discussion for examples of the ribosome, ATP synthase and respiratory complex I) or pathway substructures (next section; Fig. 4a). For protein complexes, strongly correlating subunits were generally at a shorter physical distance from each other (Extended Data Fig. 7h), confirming that physically interacting proteins melt coherently across perturbations. Therefore, the data presented here might aid future structural biology efforts for other protein complexes, by constraining which subunits should be spatially close to each other.

Having established that our data recapitulated known biology, we looked into our ability to provide new insights into the function of proteins of unknown function (orphan proteins). To facilitate this, we performed Gene Ontology (GO) enrichment of the highly correlated proteins ($r_S \geq 0.45$) for each protein in our dataset (Supplementary Data 7). In total, 140 orphan proteins could be associated with known biological processes. For several of these, we found corroborating evidence that they are involved in the process we link them to (Extended Data Fig. 8, Supplementary Discussion).

Overall, we demonstrate that co-changes in protein abundance and thermal stability are strong identifiers of functional associations in the cell, and provide organizational insights into large protein complexes. Notably, many functional associations identified by us are not previously described (only 6,116 of the 16,995 correlations with $r_S \geq 0.45$ are reported in STRING). This could uncover new cellular links between proteins of known function, and provide leads for the function of orphan genes (see Supplementary Discussion on how integrating data from this study can be used to suggest molecular mechanisms).
Glucose-SdhB

Hierarchically clustered heat map of Spearman's rank correlation (as in Fig. 3a) different from zero using a bootstrap test. NS, not significant.

pairs of metabolite–enzyme for each distribution; Extended Data Fig. 9a, b).

wild-type cells using targeted metabolomics.

thermal stability of enzymes that directly interact with the metabolite (Fig. 4).

correlation coefficients for metabolite levels in each mutant and abundance or products. Thus, TPP captures enzymatic activity in vivo, offering a unique view into the metabolic state of the cell and the ability to generate metabolic pathway associations.

Proteome changes explain mutant phenotypes

We investigated whether proteome changes could explain growth phenotypes of the mutants in different chemical and environmental stresses. For this, we used data from chemical genetics studies, in which the fitness of all E. coli single-gene deletion mutants has been measured in nearly 1,000 conditions. In general, mutants with a larger proportion of the proteome affected had a larger number of phenotypes in chemical genetics screens (r = 0.57, P < 0.001) (Extended Data Fig. 10a).

To gain insights into possible causal effects, we correlated protein abundance or thermal stability of each detected protein across all mutant backgrounds with the fitness of the same mutants in all chemical genetic conditions. This highlighted examples of proteome changes that explain growth phenotypes that are not solely related to the deleted gene (Supplementary Data 9), such as the abundance of the multidrug efflux pump MdtK explaining resistance to metformin (Extended Data Fig. 10b–d) or the abundance of the DNA repair protein RecR explaining sensitivity to UV (Extended Data Fig. 10e–g, Supplementary Discussion).

Discussion

We systematically measured the abundance and thermal stability of nearly 1,800 proteins in 121 mutants of E. coli. We detected significant changes in more than 1,200 proteins, with thermal stability and abundance measurements being largely orthogonal. Only 61 of the 273 (22%) detected essential proteins changed in their abundance, most of them being altered in a single mutant. Recently, CRISPRi has provided a way to knockdown genes. However, levels of knockdown and polar effects still present complications, especially for bacterial genomes. Because we detected changes in the thermal stability of 164 (60%) essential
proteins, our approach provides a unique view into their regulation and activity. Inspired by our ability to probe protein state and activity, we confirmed the power of our data to identify functional associations and identified more than 10,000 potentially new interactions, with 3,655 of these interactions involving 253 orphan proteins. These could provide new hints for the function of these orphan proteins. With the largest perturbation dataset for TPP, we also investigated the underlying reasons for why proteins change melting behaviour in living cells. It has previously been observed that protein thermal stability can be affected by drug16,18,25,39, nucleic acid21 and metabolite19 binding, as well as protein interactions16,26 and post-translational modifications25-27. Here, we show that protein thermal stability can also be affected by levels of cofactors and metabolites that directly bind the protein. TPP thus provides a way for surveying metabolic activity. Finally, we combined our data with existing large-scale phenotyping data25,46 to gain mechanistic insights into the causes of conditional growth phenotypes that lie beyond the knocked out gene, providing foundational information for thousands of such causal protein-phenotype connections. In conclusion, the dataset here presented can be used to gain insights into protein function and associations, and the approach is readily expandable to other organisms.

Online content
Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-3002-5.
Methods

No statistical methods were used to predetermine sample size. The experiments were not randomized and investigators were not blinded to allocation during experiments and outcome assessment.

Strains

All the *E. coli* mutants used in this study come directly from the Keio collection\(^{26}\), with the exception of *bamA\(^{41}\), *ftsA\(^{44}\), *bamD\(^{44}\) and *lptD* mutants\(^{44}\) (Supplementary Data 1; also previously described\(^{43}\)). All mutants used have been made in the *E. coli* BW25113 strain background\(^{42}\). When possible, we used two independent clones from the Keio collection to maximize variability and to spot effects that might originate from secondary mutations. For CRISPRi experiments, we transferred the chromosomal dCas9 expression cassette previously described\(^{33}\) into the BW25113 strain using PI transduction. For all follow-up work involving specific mutants, the gene deletions were retransduced into the wild-type or CRISPRi strain.

Mutant selection and multiplexing for thermal proteome profiling

To select 121 *E. coli* mutants that target diverse cellular processes, we first calculated the Pearson correlation coefficient of the chemical genetics fingerprint (5-scores across hundreds of chemical and environmental stresses\(^{27}\)) of all pairs of mutants. We then clustered the mutants based on their correlation coefficient profile, cut the tree at 11 clusters, and manually selected approximately 11 mutants per cluster—using previous knowledge of gene function to guide our selection.

For thermal proteome profiling experiments, we used tandem mass tags that allow multiplexing of 11 different conditions (TMT11plex). We decided not to use the wild-type strain in each mass spectrometry run to maximize the number of genetic perturbations tested. Instead, we considered that, in most mutants, proteins will not change abundance or thermal stability and hence the median of the 11 perturbations on each protein would work as control for each mass spectrometry run (see below for details). Therefore, it was important that the cellular processes perturbed within each mass spectrometry run were as diverse as possible. For this, the 121 mutants were randomly sampled to an 11 x 11 matrix, with the aim of running the first biological replicate of the mutants row-wise, and the second biological replicate column-wise—in a 11 x 11 matrix, with the aim of running the first biological replicate of the mutants row-wise, and the second biological replicate column-wise—in this way, each perturbation was probed against a background of 20 different perturbations (Supplementary Fig. 1). The Pearson correlation coefficient of the chemical genetic fingerprints of each mutant against the 20 different background perturbations was calculated as a proxy for the processes targeted (mutants with weak correlation target different processes\(^{5,6}\)). The randomization procedure was repeated 1,000 times and the solution in which the sum of all absolute correlation coefficients between the mutants within an experiment was minimal was considered optimal (Supplementary Fig. 1, Supplementary Data 1).

Thermal proteome profiling

Thermal proteome profiling was performed as previously described\(^{20}\). In brief, each mutant was streaked out from two independent glycerol stocks on lysogeny broth (LB) agar plates and incubated overnight at 37 °C. The next day, single colonies were picked and incubated in 2 ml LB for around 6 h, after which 50 μl of bacterial culture were transferred to 5 ml of LB and further incubated at 37 °C overnight (approximately 16 h). Overnight cultures were diluted to OD\(_{578}\) 0.001 in 50 ml LB medium and further incubated at 37 °C, 220 rpm until OD\(_{578}\) of 0.1 (range: 0.084–0.176) (Supplementary Data 1). Cells were pelleted at 4,000g for 5 min, washed with 10 ml PBS, and resuspended in PBS in a volume in ml equal to 12 x OD\(_{578}\) (equivalent to resuspending to an OD\(_{578}\) of 4). The cell suspension (100 μl) was then aliquoted to ten wells of a PCR plate, which was centrifuged at 4,000g for 5 min. Most of the supernatant (80 μl) was removed and cells were subjected to a thermal gradient (42 °C, 45.4 °C, 49 °C, 51.9 °C, 54.8 °C, 57.9 °C, 60.5 °C, 63.6 °C, 67 °C, 71.3 °C) for 3 min in a PCR machine (Agilent SureCycler 8800) followed by 3 min at room temperature. Cells were lysed with 30 μl lysis buffer (final concentration: 50 μg ml\(^{-1}\) lysozyme, 0.8% NP-40, 1 x protease inhibitor (Roche), 250 μM benzamidine, and 1 mM MgCl\(_2\) in PBS) for 20 min, shaking at room temperature, followed by three freeze–thaw cycles (freezing in liquid nitrogen, followed by 1 min at 25 °C in a PCR machine and vortexing). The plate was then centrifuged at 2,000g for 5 min to remove cell debris, and the supernatant was filtered at 500g for 5 min through a 0.45-μm 96-well filter plate (Millipore, ref: MSVHN4530) to remove protein aggregates. The flow-through was mixed 1:1 with 2 x sample buffer (180 mM Tris pH 6.8, 4% SDS, 20% glycerol, 0.1 g bromophenol blue) and kept at −20 °C until prepared for mass spectrometry analysis. To verify the effect of the heat treatment, the soluble protein concentration at each temperature for each experiment was determined using the BCA assay, according to the manufacturer’s instructions (ThermoFisher Scientific).

Mass spectrometry-based proteomics

Proteins were digested according to a modified SP3 protocol\(^{34,48}\). In brief, approximately 2 μg of protein (4 μl of frozen samples) was added to 16 μl of water and added to the bead suspension (10 μg of beads (Thermo Fisher Scientific, Sera-Mag Speed Beads, 4515-2105-05250, 6515-2105-05250) in 10 μl 15% formic acid and 30 μl ethanol). After a 15-min incubation at room temperature with shaking, beads were washed four times with 70% ethanol. Next, proteins were digested overnight by adding 40 μl of digest solution (5 mM chloroacetic acid, 1.25 mM TCEP, 200 ng trypsin, and 200 ng LysC in 100 mM HEPES pH 8). Peptides were eluted from the beads, dried under vacuum, reconstituted in 10 μl of water, and labelled for 1 h at room temperature with 17 μg of TMT11plex (Thermo Fisher Scientific) dissolved in 4 μl of acetonitrile (the label used for each experiment can be found in Supplementary Data 1). The reaction was quenched with 4 μl of 5% hydroxylamine, and experiments belonging to the same mass spectrometry run were combined. Samples were desalted with solid-phase extraction by loading the samples onto a Waters Oasis HLB μElution Plate (30 μm), washing them twice with 100 μl of 0.05% formic acid, eluting them with 100 μl of 80% acetonitrile, and drying them under vacuum. Finally, samples were fractionated onto 29 fractions on a reversed-phase C18 system running under high pH conditions. This consisted of an 85 min gradient (mobile phase A: 20 mM ammonium formate (pH10) and mobile phase B: acetonitrile) at a 0.1 ml min\(^{-1}\) starting at 0%, followed by a linear increase to 35% B from 2 min to 60 min, with a subsequent increase to 85% B from up to 62 min and holding this up to 68 min, which was followed by a linear decrease to 0% B up to 70 min, finishing with a hold at this level until the end of the run. Fractions were collected every two minutes from 12 min to 70 min and every sixth fraction was pooled together.

Samples were analysed with liquid chromatography coupled to tandem mass spectrometry, as previously described\(^{20}\). In brief, peptides were separated using an UltiMate 3000 RSLCnano system (Thermo Fisher Scientific) equipped with a trapping cartridge (Precolumn; C18 PepMap100, 5 μm, 300 μm i.d. x 5 mm, 100 Å) and an analytical column (Waters nanoEase HSS C18 T3, 75 μm x 25 cm, L. μm, 100 Å). Solvent A was 0.1% formic acid in LC-MS grade water and solvent B was 0.1% formic acid in LC-MS grade acetonitrile. Peptides were loaded onto the trapping cartridge (30 μl min\(^{-1}\) of solvent A for 3 min) and eluted with a constant flow of 0.3 μl min\(^{-1}\) using 90 min of analysis time (with a 2–28% B elution, followed by an increase to 40%, B, a washing step up to 90% B, followed by re-equilibration to initial conditions). The LC system was directly coupled to a Q Exactive Plus mass spectrometer (Thermo Fisher Scientific) or a Fusion Lumos Tribid mass spectrometer (Thermo Fisher Scientific) using a Nanospray-Flex ion source and a Pico-Tip Emitter 360 μm OD x 20 μm ID; 10 μm tip (New Objective). The mass spectrometer was operated in positive ion mode with a spray voltage of 2.3 kV and capillary temperature of 320 °C. Full-scan mass
**Article**

spectrum was a mass range of 375–1,200 m/z were acquired in profile mode using a resolution of 70,000 (maximum fill time of 250 ms or a maximum of 3 × 10^6 ions (automatic gain control, AGC)). Fragmentation was triggered for the top 10 peaks with charge 2−4 on the MS scan (data-dependent acquisition) with a 30-s dynamic exclusion window (normalized collision energy was 32), and MS/MS spectra were acquired in profile mode with a resolution of 35,000 (maximum fill time of 120 ms or an AGC target of 2 × 10^7 ions).

**Protein identification and quantification**

Mass spectrometry data were processed as previously described. In brief, raw MS files were processed with isobarQuant, and the identification of peptide and protein was performed with Mascot 2.4 (Matrix Science) against the E. coli (strain K12) UniProt FASTA (Proteome ID: UP000000625), modified to include known contaminants and the reversed protein sequences (search parameters: trypsin; missed cleavages 3; peptide tolerance 10 ppm; MS/MS tolerance 0.02 Da; fixed modifications were carbamidomethyl on cysteines and TMT10plex on lysine; variable modifications included acetylation on protein N-terminus, oxidation of methionine, and TMT10plex on peptide N-termini).

**Abundance and thermal stability score calculation**

We calculated abundance and thermal stability scores for every protein in every mutant by combining the data from the two replicates similarly to previously described, using R (v. 3.6.1). In brief, the overall distribution of signal sum intensities was normalized with vsn to compensate for slight differences in protein amounts from each tandem mass tag channel. Then, for every protein, we calculated the ratio of the signal sum intensity of each mutant to the median signal sum of the same protein in all the mutants in the same mass spectrometry experiments (that is, this yielded a fold change relative to control for every protein in each mutant at each temperature). The abundance score of each protein in each mutant was calculated as the average log-transformed fold change at the two lowest temperatures weighted for the number of temperatures in which the protein was identified for each replicate (requiring that there was data for the two biological replicates in at least one of the two temperatures). The thermal stability score of each protein in each mutant was then calculated by subtracting the abundance score from the log-transformed fold changes of all temperatures, and summing the resulting fold changes weighted for the number of temperatures in which the protein was identified for each replicate (requiring that there were at least ten data points to calculate this score). To assess the significance of abundance and thermal stability scores, we used a limma analysis instead of a previously described bootstrap approach, followed by an FDR analysis, using the fdrtool package. Abundance and thermal stability scores for all mutants were separately transformed to z-scores. Proteins with calculated |z-score| >1.96 (corresponding to a global P < 0.05 for the effect size) and with q-value <0.05 were considered significantly changed.

**Highly variable protein analysis**

We evaluated which proteins showed consistently different values between the two biological replicates. For this, we calculated the difference between replicates for all log2 fold changes of each protein at each temperature (Extended Data Fig. 1b). We extracted all the proteins that were in the top 5% of absolute difference (that is, 2.5% of each side of the distribution) and counted how many times each protein appeared (that is, from multiple mutants and multiple temperatures). We considered the top 10% of these proteins as highly variable proteins (Supplementary Data 4). GO enrichment was performed as described below.

**flhDC upstream sequence determination**

The promoter of flhDC was amplified by PCR using the forward primer 5′-GTAACCGCAACAGCGACAAG-3′ and the reverse primer 5′-CAATCAACGCTGTGCAAGTAG-3′ and the product was run on a 1% agarose gel.

**CRISPRi experiments**

We first designed guide RNAs for each gene that we wanted to knockdown. The guides comprised sequences of 20 nucleotides with perfect complementarity towards the open reading frame of the target gene and located next to a protospacer adjacent motif (NGG). The guides were designed with the guidelines from Cui et al. in mind and using the CRISPOR tool. We synthesized the following nucleotides: 5′-TTCGGGCGAACCTTCAAAGCAGCAGCTCGGTGCAACTTTTC (flhDC) 5′-TACCTCGAGCGCTCGGTGCAACTTTTC (parC) 5′-AAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAA (parE) 5′-TTCGGGCCCAAGCTTCAAAAAAAGCACCGACTCGGTGCCACTTTTTC (parA) 5′-CAATCAAACGCTGTGCAAGTAG-3′ and the product was run on a 1% agarose gel.

**Protein correlation profiling and ROC analysis**

For each protein, we averaged the log2-transformed fold change at each temperature and in each mutant (only if data were available for the two biological replicates), resulting in a maximum of 1,210 data points. We calculated the Spearman’s rank correlation for all protein pairs that overlapped by at least 242 data points (a minimum of 20% of the possible data). In this analysis, we kept proteins changing in abundance or thermal stability in only one of the two mutant clones (for example, flagella proteins above), since these were consistently co-regulated within the replicate.

These data were then used to perform a ROC analysis using the pROC package for R, ignoring the sign of the correlation (that is, the absolute correlation coefficient was used). Data were benchmarked against data from Ecocyc (operons, protein complexes, and metabolic pathways) and STRING.

We further calculated the Spearman’s rank correlation based on z-scores of abundance or thermal stability, resulting in a maximum of 121 data points (requiring protein pairs to have their abundance or thermal stability quantified in a minimum of 60 overlapping mutants).

For GO enrichment of protein partners, we selected proteins with |z| ≥ 0.45 for each protein and performed GO enrichments as described below.
Physical distance in protein complexes
From the PDB files of the ribosome (PDB: 4YBB), the ATP synthase (PDB: 5T4O) and the respiratory complex I (PDB: 4HEA) structures, we retrieved the coordinates of every atom along with its subunit identity. We used this information to calculate the centre-of-mass (assuming the same mass for every atom) of each subunit using the package SDMTools for R.

Mass spectrometry-based metabolite quantification
Cells were grown and collected as described in 'Thermal proteome profiling' to keep the two experiments as similar as possible. After washing with PBS, cells were resuspended in a volume in millilitres equal to 12 × OD_{600} (equivalent to resuspending to an OD_{600} of 4) with an acetonitrile:methanol:water (40:40:20) mixture with 100 ng ml⁻¹ of creatinine-(methyl-³¹C) and 100 ng ml⁻¹ phosphoenolpyruvic acid-²-¹³C potassium salt—used as internal standards. Samples were subjected to five freeze–thaw cycles (freezing in liquid nitrogen, followed by 5 min at 25 °C in while vortexing), centrifuged at 20,000g for 15 min at 4 °C to remove cell debris, and the supernatant was collected and kept at −80 °C until analysis.

All samples and standards were analysed on a Vanquish UHPLC system coupled to a Q Exactive Plus HRMS (Thermo Scientific) in HESI negative mode. The separation of metabolites was carried out on an XBridge BEH Amide column (100 × 2.1mm; 2.5μm) at a flow rate of 0.3 ml min⁻¹ at 40 °C. The mobile phase consisted of solvent A (7.5 mM ammonium acetate with 0.05% ammonium hydroxide) and solvent B (acetonitrile). A 16 min chromatographic run comprised a linear gradient from 2 to 12 min starting at 85% of solvent B and ending at 10%, followed by a hold from 12 to 14 min and a linear gradient to return to the initial conditions at 14.1 min.

Metabolites were analysed in HRMS full scan mode at a resolution of 35,000, an AGC target of 1×10⁶ ions, and a maximum IT of 100 ms, in the mass range of 60–900 m/z. The mass spectrometer was operated with a spray voltage of 3.5 kV, sheath gas 30 and auxiliary gas 5 units, S-Lens 65 eV, capillary temperature 320 °C, and vaporization temperature of auxiliary gas 250 °C.

Prior to the sample analysis, metabolite standards (d-glucose 6-phosphate sodium salt, d-fructose 6-phosphate disodium salt hydrate, phosphoenolpyruvic acid monopotassium salt, sodium pyruvate, suc-cinic acid, alpha-ketogutaric acid disodium salt hydrate, and malic acid) and a dilution series of a QC sample (prepared by mixing equal volumes of each sample) were analysed on the LC–MS system to determine retention times and an injection volume allowing the detection of all metabolites of interest within a linear range. For the sample analysis, blank and multiple QC samples were injected at the beginning of the sample analysis sequence in order to stabilize the LC–MS system. Samples (8 μl injection volume) were randomized during LC–MS analysis and a QC sample was injected after every five samples to track the stability of the instrument and analytical method throughout the analysis sequence. Peak areas of the deprotonated M-H metabolite ions for each metabolite were quantified on the smoothed extracted ion chromatograms (15 smoothing points) using the XCalibur Quan Browser software (Thermo Scientific) with a mass tolerance of 7 ppm. Internal standards were used to detect procedural errors, not for data normalization.

Peak area ratios were calculated for each metabolite in each mutant replicate by dividing the peak area of the metabolite in the mutant by the average of the wild-type samples of the same mass spectrometry batch. For each mutant, the average of the log₂-transformed peak area ratios was compared to abundance and thermal stability z-scores of enzymes that directly consume or produce each metabolite in glycolysis or citric acid cycle. Correlation coefficients were also calculated for random metabolite–enzyme pairs (from the pool of the same enzymes)—with this procedure being repeated 1,000 times to generate a distribution of the median of correlation coefficients. The real median of correlation coefficients was compared to the bootstrapped distribution, with the P value corresponding to the fraction of times the bootstrapped median was higher than the real median—that is, the probability that the real median is higher than zero.

CueO experiments
We amplified and Flag-tagged cueO from the E. coli BW25113 strain genome by PCR using the forward primer 5’-TTCTATCCCGGGATGTCACCGTGATTCTTTAAATATTCCG-3’ (for full length CueO) or 5’-TTCTATCCCGGGATGTCACCGTGATTCTTTAAATATTCCG-3’ for Δ28 CueO and the reverse primer 5’-TTCTATCGATTCTATCTACCTAGCTCACTTCTCTGACGCGGCGCTACCCGTAACCCGATACATCATC-3’. The PCR products were digested with Xmal and HindIII restriction enzymes, and ligated to pBAD24 plasmid—digeested with the same restriction enzymes. Plasmids were sequenced with Sanger sequencing to check that no mutations were introduced in cueO, and transformed into ΔcueO::FRT or ΔcueO::FRT ΔatakB::kan.

For periplasm extraction experiments, cells (50 ml) were grown to OD_{578} of approximately 0.5, as described in 'Thermal proteome profiling', with the exception that LB medium contained 100 μg ml⁻¹ of ampicillin and 0.2% arabinose. Cells were resuspended in wash buffer (10 mM Tris-Cl, 150 mM NaCl, pH 7.3) in a volume in millilitres equal to 2 × OD_{578} (equivalent to resuspending to an OD_{578} of 25). An aliquot (500 μl) was transferred to a new tube, cells were centrifuged at 4,000g for 5 min, and the supernatant was discarded. The pellet was resuspended in 300 μl SET buffer (0.5 M sucrose, 200 mM Tris-Cl, 1 mM EDTA, pH 7.3), followed by the addition of 100 μl of 3 mg ml⁻¹ lysozyme and 300 μl of ice cold water. Cells were incubated for 20 min at 37 °C without shaking. After incubation, a 50 μl aliquot (whole cells) was collected, cells were centrifuged at 10,000g for 30 s, and a 100 μl aliquot of supernatant (periplasm) was collected. Benzonase (final concentration: 250 U ml⁻¹) and MgCl₂ (final concentration: 1 mM) were added to the samples, and samples were incubated for 10 min at room temperature. Sample buffer (180 mM Tris pH 6.8, 4% SDS, 20% glycerol, 0.1 g bromophenol blue) was added to the samples, and samples were kept at −20 °C until analysis by western blot (described below).

For cellular thermal shift assay (CETSA) experiments, cells were grown as described in 'Thermal proteome profiling', with the exception that LB medium contained 100 μg ml⁻¹ of ampicillin and 0.2% arabinose. The CETSA experiment with 4 mM CuCl₂ was performed in lysate of ΔcueO::FRT cells transformed with Δ28-CueO plasmid. Lysate was prepared as previously described. In brief, cells were grown to OD_{578} of approximately 0.5 in 100 ml LB medium containing 100 μg ml⁻¹ of ampicillin and 0.2% arabinose, washed with PBS, and resuspended in lysis buffer (without NP40) in a volume in millilitres equal to 2 × OD_{578} (equivalent to resuspending to an OD_{578} of 50). Aliquots of lysate (200 μl) were treated with 4 mM CuCl₂ (2 μl of 400 mM CuCl₂) or water (2 μl), aliquoted (20 μl) to 10 wells of a PCR plate, and subjected to the temperature gradient described in 'Thermal proteome profiling', NP-40 was then added to a final concentration of 0.8%, and samples were processed as described in 'Thermal proteome profiling' .

Samples were run on SDS–PAGE and CueO was detected by western blot using mouse monoclonal anti-Flag antibody (F3165, Merck, dilution 1:10,000) and goat anti-mouse IgG-HRP (sc-2005, Santa Cruz Biotechnology, dilution 1:5,000). As a loading control, rabbit anti-LpoB antibody was used. The immunoblots were scanned and visualized using the Storm image analysis system (Molecular Dynamics).

Metformin and UV sensitivity
Plasmids p-empty, p-mdtk, p-ahpc, p-cpxA were purified from the Transbac library. Plasmids p-empty, p-recR, and p-ybaB were purified from the pMORB library. These were transformed to wildtype, Δahpc::kan, ΔcpxA::kan, Δmdtk::FRT, Δmdtk::FRT Δahpc::kan, and Δmdtk::FRT ΔcpxA::kan strains for metformin experiments, or wild-type, ΔybaB::kan, and ΔrecR::kan strains for UV experiments.
For metformin experiments, strains were grown to early stationary phase at 37 °C in the presence of 10 μg ml⁻¹ tetracycline. Cells were then diluted to OD₅₇₈ = 0.5, serially diluted in tenfold steps in LB, and spotted on LB agar plates containing 10 μg ml⁻¹ tetracycline, 0.1 mM IPTG, and metformin to the desired concentration. Plates were incubated at 37 °C overnight and imaged in the morning.

For UV sensitivity experiments, strains were grown to early stationary phase at 37 °C in the presence of 50 μg ml⁻¹ ampicillin. Cells were then diluted to OD₅₇₈ = 0.1, serially diluted in tenfold steps in LB, and spotted on LB agar plates containing 50 μg ml⁻¹ ampicillin and 0.1 mM IPTG. Plates were exposed to UV with a total energy of 85 mJ cm⁻² in a Spectrolinker XL-1500 UV crosslinker. Plates were incubated at 37 °C overnight and imaged in the morning.

**GO enrichments**

GO enrichments were performed using the Fisher's exact test and corrected for multiple comparison with the Benjamini-Hochberg procedure.

**Reporting summary**

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

**Data availability**

The thermal proteome profiling data are available at http://ecoliTPP.shiny.embl.de. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD016589. The mass spectrometry metabolomics data have been deposited to the MassIVE repository with the dataset identifier MSV000084632. Data for protein complexes, pathways, and operons were retrieved from Ecocyc v21.1 (https://ecocyc.org/). STRING database v10.5 was used. Data referring to protein localization were retrieved from STEPdb v1.0 (http://stepdb.eu/). Cellular processes targeted by mutants in this study were derived from Clusters of Orthologous Groups (COG) database (https://www.ncbi.nlm.nih.gov/research/cog-project/). Gene ontology annotations (release: 2020-01-01) were downloaded from http://geneontology.org.

**Code availability**

The code to process raw mass spectrometry data (available at PRIDE partner repository with the dataset identifier PXD016589) and to calculate abundance and thermal stability scores and q-values (Supplementary Data 3) is available at https://github.com/fstein/EcoliTPP.

61. Orfanoudaki, G. & Economou, A. Proteome-wide subcellular topologies of E. coli polypeptides database (STEPdb). Mol. Cell. Proteomics 13, 3674–3687 (2014).
62. Tatusov, R. L., Galperin, M. Y., Natale, D. A. & Koonin, E. V. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res. 28, 33–36 (2000).

**Acknowledgements**

We thank P. Phapale for help with metabolomics analysis, and H. Link and D. Bikard for strains and plasmids for the dCas9 work. This work was supported by the European Molecular Biology Laboratory. A.M. and K.M. were supported by a fellowship from the EMBL Interdisciplinary Postdoc (EIPPOD) programme under Marie Sklodowska-Curie Actions COFUND (grant number 664726). C.V.G. is recipient of an EMBO long-term postdoctoral fellowship and an add-on fellowship from the Christiane Nüsslein-Volhard-Stiftung. AT is supported by an ERC consolidator grant, uCARE.

**Author contributions**

A.M., N.K., F.S., M.M.S. and A.T. designed the study. A.M. and J.H. performed the thermal proteome profiling experiments. A.M., J.H. and D.H. performed the proteomics mass spectrometry analysis. A.M. and K.M. performed the metabolomics mass spectrometry analysis. A.M., J.B., M.S., C.V.G. performed follow-up molecular work. flhDC (A.M.), CueO (A.M. and J.B.), CRISPRi (A.M. and M.S.), MdtK and RecR (A.M., M.S. and J.B.), other genetics and biochemistry (A.M., J.B. and C.V.G.). A.M., N.K. and F.S. performed the data analysis. A.M., A.T. and M.M.S. drafted the manuscript, which was reviewed and edited by all authors. A.T. and M.M.S. supervised the study.

**Competing interests**

The authors declare no competing interests.

**Additional information**

Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-3002-5.

**Correspondence and requests for materials** should be addressed to A.T. or M.M.S.

**Peer review information** Nature thanks the anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

**Reprints and permissions information** is available at http://www.nature.com/reprints.
Extended Data Fig. 1 | Biological replicates show good reproducibility, with differences revealing biological phenomena. **a**, Rarefaction analysis of the proteome coverage (proteins with at least two unique peptides in each mass spectrometry run) as a function of the number of mass spectrometry runs. **b**, Distribution of log₂-transformed fold change differences between the two biological replicates. **c**, Scatter plot of protein fold changes between all biological replicate measurements (\(n = 1,512,475\); all proteins, all temperatures, all mutants). \(r\) depicts Pearson correlation. **d**, Reproducibility of protein fold changes between biological replicate measurements at each temperature. **e**, Examples of replicate correlation for specific mutants, highlighting that flagellar proteins are common outliers in one of the two clones (\(n_{\Delta\text{hemX}} = 13,150\), \(n_{\Delta\text{ybaB}} = 12,313\), \(n_{\Delta\text{clpA}} = 12,950\), \(n_{\Delta\text{mrcB}} = 12,604\), \(n_{\Delta\text{fur}} = 12,543\), \(n_{\Delta\text{mlaA}} = 12,559\), \(n_{\Delta\text{lpp}} = 12,719\); all proteins, all temperatures). **f**, Polymerase chain reaction of the promoter region of the \(\text{flhDC}\) operon (schematic on top) demonstrates the presence of insertions in mutant clones (gel on bottom, \(n = 1\); for gel source data see Supplementary Fig. 2) with high flagellar protein expression (FliC fold-changes at the two lowest temperatures of each mutant replicate used as a proxy for abundance). **g**, Scatter plot of abundance and thermal stability z-scores of all proteins in all mutants (\(n = 170,150\)). \(r\) depicts Pearson correlation. **h**, Distribution of the number of mutants in which a protein is significantly altered (\(n = 1,764\) proteins). Box plots as in Fig. 2a. **i**, Distribution of the number of proteins that are significantly altered in each mutant (\(n = 121\) mutants). Box plots as in Fig. 2a.
Extended Data Fig. 2 | Cellular processes targeted in this study and changes in thermal stability reflect protein complex architecture in *E. coli* mutants.

**a**, Distribution of cellular processes targeted in this study compared to the general distribution of the *E. coli* genome using Clusters of Orthologous Groups (COG).

**b**, **c**, Schematic representation of protein complexes targeted by genetic perturbations in this study. Protein missing (encoded by gene deleted) is highlighted by a dashed line and other complex members are coloured according to their thermal stability (b) or abundance (c) in that mutant. *|Z-score| >1.96 and with *p*-value ≤ 0.05. *ΔtolC* data are from ref. 20.
Extended Data Fig. 3 | Protein co-expression patterns provide insight into gene expression regulation. 

a, Correlation of DegP and OmpF log$_2$-transformed fold changes to control in each of the genetic perturbations probed here ($n = 120$, as OmpF is not detected in ΔompF) at each temperature (colour coded; $n = 10$). Mutants that lead to cell envelope stress (highlighted), and therefore activation of stress response (see also b) lead to upregulation of DegP and downregulation of OmpF. 

b, Schematic representation of regulation of degP and ompF genes. CpxAR two-component system regulates both genes, while EnvZ/OmpR regulates only ompF. Heatmap shows Spearman’s rank correlation (calculated as in Fig. 3a) for proteins involved in regulation of degP and ompF.
Extended Data Fig. 4 | Cofactor binding leads to changes in protein thermal stability.

The figure shows the distribution of thermal stability z-scores of all proteins in the iron-sulfur cluster biosynthesis mutants, ΔiscA, ΔiscS and ΔiscU according to their Gene Ontology annotation as iron-sulfur cluster binding proteins (nΔiscA = 41, nΔiscS = 41, nΔiscU = 40) or not (nΔiscA = 1,400, nΔiscS = 1,415, nΔiscU = 1,314). Box plots as in Fig. 2a. Significance assessed with two-sided Wilcoxon signed-rank test (PΔiscA = 3.9 × 10−5, PΔiscS = 9.5 × 10−11, PΔiscU = 7.7 × 10−5).

Extended Data Table 1 summarizes the changes in thermal stability of CueO in different conditions.

- **Extended Data Table 1**
  | Condition | ΔiscA | ΔiscS | ΔiscU |
  |-----------|-------|-------|-------|
  | WT | 0.10 | 0.12 | 0.13 |
  | Δ28 | 0.07 | 0.09 | 0.10 |
  | Δ28 tatB | 0.06 | 0.08 | 0.09 |
  | Δ28 ΔcueO | 0.05 | 0.07 | 0.08 |

Coomassie to ensure that periplasmic extraction was successful (n = 1).

- **Extended Data Table 2**
  | Condition | ΔiscA | ΔiscS | ΔiscU |
  |-----------|-------|-------|-------|
  | WT | 0.10 | 0.12 | 0.13 |
  | Δ28 | 0.07 | 0.09 | 0.10 |
  | Δ28 tatB | 0.06 | 0.08 | 0.09 |
  | Δ28 ΔcueO | 0.05 | 0.07 | 0.08 |

Coomassie to ensure that periplasmic extraction was successful (n = 1).

- **Extended Data Table 3**
  | Condition | ΔiscA | ΔiscS | ΔiscU |
  |-----------|-------|-------|-------|
  | WT | 0.10 | 0.12 | 0.13 |
  | Δ28 | 0.07 | 0.09 | 0.10 |
  | Δ28 tatB | 0.06 | 0.08 | 0.09 |
  | Δ28 ΔcueO | 0.05 | 0.07 | 0.08 |

Coomassie to ensure that periplasmic extraction was successful (n = 1).
Extended Data Fig. 5 | Changes in thermal stability of essential proteins.  

**a.** log₂-transformed fold change of FtsK protein levels in each mutant compared to control at each temperature. FtsK is strongly thermally destabilized in the ΔphoP mutant and the ftsK knockdown is synthetically lethal with the phoP deletion (Fig. 2d).  

**b.** As in a for ParC. ParC is strongly thermally stabilized in the ΔclpS mutant and thermally destabilized in the ΔphoP mutant and the parC knockdown is synthetically lethal with both. Synthetic lethality is also apparent in the ΔahpC, ΔamiA and ΔenvC mutants, despite the absence in changes in ParC thermal stability (Fig. 2e).
Extended Data Fig. 6 | See next page for caption.
Extended Data Fig. 6 | Protein correlation profiling recapitulates known biological interactions with abundance and thermal stability data having different contribution to functional associations. a, Distribution of Spearman’s rank correlation of all protein pair comparisons compared to known operons, protein complexes, and metabolic pathways. Distribution statistics refer to all protein pairs. b, ROC analysis based on the decreasing absolute Spearman’s rank correlation compared to interactions in STRING database at different cut-offs of the combined STRING score. c–e, Spearman’s rank correlation of protein pairs belonging to the same operon (c), protein complex (d), or metabolic pathway (e) using solely changes in abundance (x axis) or changes in thermal stability (y axis). Protein pairs belonging to the same operon are highlighted in purple. Distribution of Spearman’s rank correlation are shown outside the axes. n = 446 for operons, n = 348 for protein complexes, and n = 801 for metabolic pathways. Proteins belonging to the same operon or complex mostly have coordinated abundance changes, while proteins belonging to the same pathway have also often coordinated thermal stability. f, Schematic representation of UDP-N-acetylmuramoyl-pentapeptide biosynthesis pathway. g, Example of protein pair (DdlA and MurC) co-changing in their thermal stability (r = 0.79), but not abundance (r = −0.13) across 81 genetic perturbations. Each data point corresponds to the abundance or thermal stability z-score in one of the genetic perturbations (colour-coded). h, Heat map of Spearman’s rank correlation of all quantified members of UDP-N-acetylmuramoyl-pentapeptide biosynthesis pathway based on co-changes in abundance (top triangle) or thermal stability alone (bottom triangle).
Extended Data Fig. 7 | Protein correlation profiling reflects substructures of protein complexes. a, Heat map of Spearman’s rank correlation (lower triangle; based on protein abundance and thermal stability data across 121 mutants, as in Fig. 3a) and the physical distance (upper triangle; based on ribosome structure, PDB: 4YBB, and using the centres of mass of each protein) between the ribosome members. At the bottom, 30S and 50S ribosomal subunits are shown in purple and green, respectively, and lower triangle data are clustered hierarchically. b, c, High-resolution structure of the ribosome coloured according to the heat map clusters from a (b) or 30S and 50S ribosomal subunits (c). d–g, ATP synthase members (d, e: PDB: 5T4O) and respiratory complex I (f, g: PDB: 4HEA), as in a–c. h, Closely located members of protein complexes are more likely to be similarly regulated across different conditions. Spearman’s rank correlation plotted against the distance between complex subunits for the three complexes represented in the figure, with an apparent negative correlation. Box plots are as in Fig. 2a.
Extended Data Fig. 8 | GO enrichments of co-changing partners of proteins of unknown function can reveal their function. Examples of links between proteins of unknown function and GO terms that their co-changing proteins are enriched in. Some of these links are supported by external evidence (node colour, see Supplementary Discussion). Edges are coloured according to the enrichment $P$-value using the Fisher’s exact test after correction for multiple comparison with the Benjamini–Hochberg procedure.
Extended Data Fig. 9 | Metabolite levels correlate with thermal stability of enzyme producing or using the metabolite. a, b, Scatter plot of metabolite log₂ fold-changes in mutant compared to wild-type strain (y axis) and protein abundance (a) or thermal stability (b) in each mutant for enzymes that directly interact with the metabolite (x axis) (n = 19 mutants, except for G6P/F6P–PhoA (n = 7), 2-oxoglutarate–SucA (n = 18), succinate–SdhD (n = 12), malate–FumA (n = 6), and malate–FumB (n = 12)). r depicts the Pearson correlation coefficient for each metabolite-enzyme pair. Black line represents the linear fit and grey shades the 95% confidence interval of the fit. c, Twenty strains used for targeted metabolomics analysis. d, Distribution of Pearson correlation coefficients for metabolite levels in each mutant and abundance or thermal stability of enzymes that directly interact with the metabolite (upstream and downstream of metabolite, as in a and b). Box plots are as in Fig. 2a. With all data represented on top of the box plots (n_G6P/F6P = 6, n_PEP = 5, n_Pyruvate = 8, n_2-oxoglutarate = 4, n_succinate = 6, n_malate = 9).
Extended Data Fig. 10 | Changes in protein abundance and thermal stability explain growth phenotypes of *E. coli* mutants. a, Scatter plot of number of significantly affected proteins (abundance or thermal stability) in each mutant (x axis) and the number of significant growth phenotypes of the same mutant (y axis; data from ref. 12). Pearson’s correlation coefficient is $r_p = 0.56$ ($p=4.2 \times 10^{-11}$, n=119). b, Scatter plot of MdtK abundance in mutants profiled in this study and their sensitivity to 80 mM metformin ($r = 0.44$; n=119 mutants). c, d, Spot assay for the indicated strains overexpressing mdtK, ahpC or cpxA, or a control empty plasmid in plates containing 0–80 mM metformin. Cells were diluted to OD$_{578} = 0.5$, serially diluted in tenfold steps, and spotted on LB agar plates containing 10 µg ml$^{-1}$ tetracycline (to maintain plasmid), 0.1 mM IPTG (to induce expression of encoded gene), and metformin as indicated. e, As in b, but showing correlation of RecR abundance and UV exposure for 18 s ($r = 0.53$; n=99 mutants). f, Schematic representation of the *ybaB*-recR operon and protein abundance scores in the Δ*ybaB* mutant. g, Spot assay for the indicated strains overexpressing *ybaB*, *recR*, or a control empty plasmid after exposure to UV with a total energy of 85 mJ cm$^{-2}$ or control non-exposed plate. Cells were diluted to OD$_{578} = 0.1$ and then serially diluted in tenfold steps, and spotted on LB agar plates containing 50 µg ml$^{-1}$ ampicillin (to maintain plasmid) and 0.1 mM IPTG (to induce expression of encoded gene).
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- n/a
- Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
- Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. mean) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted Give P values as exact values wherever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen’s d, Pearson’s r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection
Proteomics raw data files were processed using isobarQuant (https://github.com/proctode/isob), Mascot (ver. 2.4, Matrix Science). Metabolomics raw data files were processed using XCalibur Quan Browser software (ver. 4.4.28.14).

Data analysis
R (ver. 3.6.1), vsn for R (ver. 3.9), limma for R (ver. 3.9), fdtool for R (ver. 1.2.15), pROC for R (ver. 1.15.5), SDMtools for R (ver. 1.1-221.1) were used for data analysis. Code to analyze the data is available at https://github.com/fsstek/EcoilTPP.

For manuscripts utilizing custom algorithms or software that are not yet described in published literature, software must be made available to editors/reviewers.

We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD016589. The mass spectrometry metabolomics data have been deposited to the MassIVE repository with the dataset identifier MSV000084632.

Data for protein complexes, pathways, and operons was retrieved from Ecocyc v21.1 (https://ecocyc.org/). STRING database v10.5 was used (https://string-db.org/). Data referring to protein localization was retrieved from STEPDbl v1.0 (http://stepdb.eu/). Cellular processes targeted by mutants in this study were derived from Clusters of Orthologous Groups (COG) database (https://www.ncbi.nlm.nih.gov/research/cog-project/). Gene ontology annotations (release: 2020-01-01) were downloaded from http://geneontology.org.
Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences  ☐ Behavioural & social sciences  ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/hr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| Sample size | All experiments were performed in at least two independent biological replicates, with the exact numbers for each experiment given in the figure legends. No sample size calculation was performed, since for most experiments we were limited by the available two independently generated strains of the Kxo collection. |
|-------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Data exclusions | No data were excluded from the analyses. |
| Replication | We performed the analysis on at least two biological replicates. When possible, these originated from two independent biological samples to maximize variability (see Supplementary Data 1 for cases in which this was not possible), and with the samples being prepared on separate days to maximize variability. |
| Randomization | Samples for proteomics analysis were randomized to mass spectrometry runs in a way that minimized their phenotypic correlation [highlighted in Extended Data Figure 1]. Samples for metabolomics analysis were randomized during LC-MS analysis as described in methods. For all other experiments, randomization is not applicable, since we included samples that we wanted to directly compare in the same gel/ blot or agar plate. |
| Blinding | No blinding was performed, since it was necessary to keep track of which samples are used in each experiment. Controls were always processed in parallel to samples being compared. |

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

| n/a | Involved in the study |
|-----|-----------------------|
| ☐   | Antibodies            |
| ☑   | Eukaryotic cell lines |
| ☑   | Palaeontology         |
| ☑   | Animals and other organisms |
| ☑   | Human research participants |
| ☑   | Clinical data         |

Methods

| n/a | Involved in the study |
|-----|-----------------------|
| ☑   | ChiP-seq              |
| ☑   | Flow cytometry        |
| ☑   | MRI-based neuroimaging |

Antibodies

Antibodies used

- Mouse monoclonal anti-FLAG antibody (F3165, Merck)
- Goat anti-mouse IgG-HRP [sc-2005, Santa Cruz Biotechnology]
- Rabbit anti-Lox8 antibody [Typhas et al., Cell 2010]
- Goat anti-rabbit IgG-HRP [sc-2004, Santa Cruz Biotechnology]

Validation

Antibodies were commercially available or validated in the publication in which they were developed:

- Mouse monoclonal anti-FLAG antibody [https://www.sigmadirich.com/catalog/product/sigma/f3165]
- Goat anti-mouse IgG-HRP [https://www.scbt.com/p/goat-anti-mouse-igg-hrp]
- Rabbit anti-Lox8 antibody [Typhas et al., Cell 2010, PMID: 21183073]
- Goat anti-rabbit IgG-HRP [https://www.scbt.com/p/goat-anti-rabbit-igg-hrp]

Additionally, we always confirmed the molecular weight of the protein detected by western blot.