Genome-wide analysis of cell-gene interactions

S. Cardinale[1, 2, *]

[1] University of California-Berkeley, Department of Bioengineering, Berkeley, CA 94720 – USA

[2] Current address: Technical University of Denmark · Novo Nordisk Foundation Center for Biosustainability · Building 220 · DK-2800 Kgs. Lyngby · Denmark.

* Correspondence: stefca@biosustain.dtu.dk

Keywords

Gene knockout, gene expression, noise, pleiotropic interaction.

Subject Category: Synthetic Biology

Abstract

The study presents an analysis of how different cellular functions link cell size to the expression of synthetic genes in *E. coli*. The Size-Expression interaction was mapped with a two-gene genetic probe across 3800 single-gene deletion strains. Through regression analysis, expression-specific effects and gene-specific effects were derived from size effects and generic expression effects, respectively. The entire compendium of cell functions broadly mapped to four systems of distinct primary influence on the Size-Expression map. Specifically, membrane structural components primarily affected size, whereas protein and RNA stability primarily affected gene expression. In addition, major Size-Expression shifts showed no substantial gene-specific effects unless they were mediated by key components of the protein synthesis apparatus.

Background

The goal of synthetic biology to standardize the engineering of biology while scaling the design of novel cell functions to increasing complexities critically relies on the reliability and predictability of synthetic biological components [1,2]. Challenged by constantly mutating endogenous interactions, designed synthetic molecules with specific functions can be quickly made unstable or non-functional by variations in host physiology or genetic makeup. For these reasons, tools are being constructed to shield the functions or predict the behaviors of synthetic genes in the cell. These tools include the design of robustness to evolutionary instability [3] and molecular context [4] in addition to the application of computational algorithms in the design of molecular parts to ensure robust function [5].

37   Genome-wide mapping of the gene-to-phenotype relationship has enabled the effective

38   identification of genetic targets to improve complex traits in bacteria, such as tolerance to ethanol

39   [6], cellulosic hydrolysate and isobutanol [7]. However, knowledge of the genotype-phenotype

40   relationship does not necessarily allow engineering of reliable heterologous pathways in these

41   strains. To achieve this capability, a systematic understanding of cell-function-to-pathway

42   interactions enabling accurate whole-cell models of the biochemical networks is necessary [8].

43   Here, the cell function-to-gene expression interaction was investigated for all non-essential

44   genes in the *E. coli* genome. Specifically, both the global and specific effects of the lack of a

45   cellular component on cell biomass and synthetic gene expression were mapped. The global

46   effects may involve both the ability of the cell to grow and the cellular amount of a synthetic

47   component. Perturbations in the synthetic circuit dynamics of this type of effect may include

48   growth feedback and have been described mathematically [9]. Alternatively, a cell function can

49   have a localized effect, such as an effect on the output of synthetic genes but not on cell size.

50   Therefore, to investigate the sequence-specificity of the effects of cell function loss, the present

51   study used a genetic construct with two identically expressed reporter genes.

52   The results indicated that most functional disruptions affected both cell size and synthetic

53   gene expression and caused an analogous positive or negative shift. Phenotypic patterns of the

54   cell-expression interaction were mapped to 4 major systems in the cell: membrane function,

55   RNA-ribosome-protein stability, biosynthetic and energy-motility functions. Although both sides

56   of the interaction were affected in general, the results suggested that each system predominantly

57   affected one side and indirectly affected the other. In particular, the impairment of cell

58   invagination caused larger cells and indirectly resulted in higher cellular reporter concentrations,

59   whereas defective motility and nutrient uptake yielded mainly smaller cells and secondarily

60   decreased gene expression. Protein folding and RNA stability mainly affected synthetic gene

61   expression while indirectly affecting cell size, particularly if the disruption involved structural

62   ribosome components.

63

64   **Results**

65   **Quantification of cell size and reporter expression across the KEIO collection**

66   The aim of this study was to map cell-expression interactions by quantifying the effects of

67   systematic removal of each non-essential *E. coli* gene on the constitutive expression of two

68   synthetic reporter genes. To achieve this, a genetic probe containing the mVenus and mCherry

69   genes expressed from identical promoter-5'-UTR sequences [10] was introduced in 3823 gene

70   knockout strains of the KEIO collection [11]. Each of the 3823 strains carrying the probe was grown

71   from a mixture of 2-3 single colonies from agar plates to avoid colony-to-colony variability of

72   growth on agar and inoculation. Single-cell measurements of mVenus and mCherry fluorescence

73   along with other parameters were acquired with a flow cytometer at the mid-log phase of growth.

74   The variability across biological replicates and data collected at different times was tested for 180

75   strains from three different plates and was found to be $5*10^{-4}$ for mVenus and $1.3*10^{-3}$ for

76   mCherry. These values were approximately 1 order of magnitude smaller than the variance across

77   KEIO strains for the measured variables, which was between $3.6*10^{-3}$ and $1*10^{-2}$ (Sup. Info).

78   Although forward scattered light (abbreviated FSC) can be effectively used to measure microbial

79   cell size in cytometry [12], this measure does not scale linearly with particle sizes for all instruments

80   [13,14]. In this study, the cytometer FSC measure scaled linearly with cell volume within the size

81   range typical for an *E. coli* cell (~2μm, Sup. Fig. 1); therefore, this measure was used as a proxy

82   for cellular size (**S**).

83   At the time of library construction [11], some bias was introduced by the investigators while

84   arraying strains in plates. This bias resulted in the occasional presence of genes with similar

85   functions together in the same plate, such as genes encoding chemotactic and flagellar proteins in

86   plate #45 (Sup. Fig. 2). This bias, however, did not appear to significantly affect the mean

87   fluorescence of the plate in relation to the dataset distribution (Sup. Fig. 3-6). Therefore, no

88   further data normalization was performed. However, in light of this bias, the per-plate distribution

89   of strains with significant phenotypes was further assessed (see below).

90

91   The average fluorescence of mVenus and mCherry varied approximately four-fold and was

92   strongly correlated across the 3823 transformed KEIO strains (r = 0.90, Fig. 1A) and with the

93   FSC measure (correlation 0.67 and 0.61 for mVenus and mCherry, respectively) (Fig. 1A). A

94   change in cell size may globally affect cell physiology and indirectly affect heterologous gene

95   expression in cases in which proteins are not sufficiently divided between daughter cells. This

96   scenario was supported by the observed positive correlation between FSC and fluorescence

97   output (Fig. 1A). We sought to assess the effects of gene knockout on heterologous gene

98   expression after normalizing for the influence of variations in cell size (FSC). Fluorescence

99   measurements were regressed against FSC to obtain the residuals for mCherry and mVenus, and

100   the average of the two values was used as the FSC-normalized measure of heterologous gene

101   expression **E**. To obtain a measure of the differential effect of KEIO knockouts on the individual

102   reporter genes, the residuals of the FSC-regressed fluorescence using **E** as predictor were

103   quantified. The resulting set of regressed mCherry and mVenus expression values were highly

104   correlated with **E** (cor=0.94-0.98) and used as proxy for gene-specific effects (**G$_{spec}$**).

105

E and G_{spec} features are not affected by the same cell functions

107   The top and bottom 5% quantiles in each distribution were used to select genes with

108 significant feature values. These extreme values were homogenously distributed across the

109 dataset (Fig. 1A, cyan dots), thus indicating that the regression did not introduce noticeable bias.

110 The number of genes in each group with higher or lower **S** or **E** values denoting S_{high}/S_{low} and

111 E_{high}/E_{low} phenotypes was 153 and the number of genes with a G_{spec} phenotype was 384. Some

112 co-localization of genes in KEIO plates was observed: genes with a S_{high} phenotype were

113 concentrated in the range between plate #17 and #43 (Sup. Fig. 7), several E_{high} genes were found

114 in plates #33, #37 and #39, and E_{low} genes were found in plate #45 (Sup. Fig. 8). In addition,

115 several genes with the G_{spec} phenotype were found in plates #61 and #89 (Sup. Fig. 9). There did

116 not appear to be a clear link between the co-localization of genes in a certain plate and significant

117 shifts in multiple features. For instance, although a number of genes with the E_{high} - S_{high}

118 phenotype were found in the same #37 and #39 plates (Sup. Fig. 7-8), other plates, such as plate

119 #33, contained a number of genes with E_{high} but insignificant **S** or **G** phenotypes.

120

121   In general, a large number of genes had a significant value only in one feature. A S_{high}

122 phenotype combined with an $E_{high/low}$ shift was found only in ~7-8% of the genes, while a S_{low}

123 phenotype shared ~13% of the genes with the E_{low} phenotype and a higher (~20%) fraction with

124 the E_{high} phenotype ($p < 10^{-4}$, Fig. 1B). **E** and G_{spec} phenotypes also shared 14-19% of genes.

125 Notably, 33% (53/159) of genes with a S_{low} shift also had a G_{spec} phenotype compared with only

126 7.5% of those with a S_{high} phenotype. This result may indicate that a decrease in cell size may

127 disrupt the balance in the expression of two synthetic genes more significantly than an increase in

128 cell size (Fig. 1B).

129   Next, the presence of functional enrichments in gene groups with specific **E** or **G** phenotypes

130 was assessed. The E_{low} group was enriched in knockouts in Enterobacterial Common Antigen

131 (ECA) biosynthesis and flagella assembly (Bonferroni corrected $p < 10^{-2}$, Table 1 - bottom). We

132 also found that knockouts for other membrane-associated functions, such as chemotaxis, cell

133 adhesion, membrane transport and Two-Component Signal (TCS) transduction (KEGG pathways,

134 $p < 10^{-1}$), were enriched in this group. Genes with the E_{high} phenotype were characterized by an

135 enrichment in nucleotide biosynthesis activity for both purine and pyrimidine nucleotides (Table

136 1 - top). In this group, enrichment for homologous recombination functions, including the key

137 genes *ruvA*, *priC* and *holD*, were also found. Purine nucleotide biosynthesis was also enriched

138 among genes with E_{high} – G_{spec} phenotypes ($p < 0.5$) (Table 1). In this group, there were also

139 KEIO strains of important housekeeping genes, such as *priB* and *priC,* which are involved in

140 DNA replication. In addition, genes with key functions in mRNA (*pnp*) and protein stability

141 (*cpxA*) as well as protein folding and export (*tatBC, dnaK, dmsD*) were identified (Table 1).

142 Importantly, *tatBC* and *dmsD* physically interact in the cell and share a $\mathbf{E_{high}} - \mathbf{G_{spec}}$ phenotype,

143 but these KEIO knockouts were located in different KEIO plates (#89, #71 and #61) indicating

144 that significant functional gene enrichment is not plate-biased.

145   Genes with a $\mathbf{G_{spec}}$ phenotype belonged to a diverse range of cell functions including

146 transcription factors and enzymes involved in central carbon metabolism and amino acid

147 biosynthesis ($p < 10^{-2}$). Overall, $\mathbf{E}$ and $\mathbf{G_{spec}}$ phenotypes did not share substantial sets of cellular

148 functions beyond genes involved in purine nucleotide biosynthesis.

149

150

151 **Cell function disruption triggers specific changes in S and E features**

152

153   To gain a better understanding of the role of different cell functions on $\mathbf{S}$ and $\mathbf{E}$, KEIO

154 knockouts populating all observed combinations of $\mathbf{S_{high/low}}$ and $\mathbf{E_{high/low}}$ patterns were

155 investigated. In addition, the presence of a $\mathbf{G_{spec}}$ effect, which quantifies expression imbalance

156 between the two reporter genes, was assessed in each of the observed $\mathbf{S\text{-}E}$ patterns. The final

157 matrix consisted of 17 combinations populated by 400 genes with significant feature values (Fig.

158 2).

159   In general, the majority of genes (344 or 82%) showed a single-feature phenotype rather than

160 significant changes in both features. Because there is a degree of co-localization of knockouts of

161 genes with similar functions within the KEIO collection (see above), the plate aggregation of sets

162 of genes sharing a specific $\mathbf{S\text{-}E}$ phenotype was investigated. Generally, strains from the entire

163 KEIO collection populated each pattern (Sup. Fig. 10A-C and info). However, we found some co-

164 localization of knockouts of genes involved in chemotaxis and flagellum biosynthesis in addition

165 to a set of 4 protein chaperones (Group 15) on plate #45 (Sup. Fig. 10D). These genes shared a

166 $\mathbf{E_{low}}$ phenotype together with 7 other chemotactic/motility genes and the proteases ClpP/X, which

167 resided in different KEIO plates. In conclusion, although we could not completely rule out a

168 plate-effect for these functions, their influence on the Size-Expression interaction was supported

169 by a larger number of members from the entire dataset.

170

171   The features-matrix was re-arrayed to observe the gradient of the shift with $\mathbf{S_{high}}$ and $\mathbf{E_{low}}$

172 placed at the top, bottom and all other possible combinations in between. This arrangement

173 displayed a pattern that suggested that most genes were characterized by a similar up/down shift

174   in the S and E features. Only 30 genes presented a mixed phenotype (for example $S_{down} - E_{high}$),

175   representing a total of only 7.5% of all cases. This result indicated a lack of substantial protein

176   'concentration' effect, likely a result of the regression against the FSC measure.

177   Next, all phenotypes with >20 members were assessed for functional enrichment, while gene

178   lists were reported for smaller groups. Although a large number of genes presenting only a single-

179   feature phenotype were observed earlier (Fig. 1B), the effect of disrupting major cell functions

180   generally affected both S and E phenotypes simultaneously. Only the impairment in amino acid

181   biosynthesis, and particularly in aromatic amino acids, appeared to specifically lead to a S_{high}

182   phenotype (Bonferroni corrected $p < 10^{-1}$) (Fig. 2 group 1 brown text). Many genes with key

183   housekeeping cellular functions presented a significant S_{high} or E_{high} phenotype either alone or in

184   combination (Fig. 2, groups 2-6 orange text, see also Table 1). Intriguingly, as discussed above

185   (Table 1), the effect of impairing nucleotide biosynthesis appeared to primarily lead to the E_{high}

186   phenotype, with or without a S_{low} or G_{spec} effect (Fig. 2, groups 4, 5, 8 brown text). The

187   observation that the deletion of amino acids and nucleotide biosynthetic enzymes showed an

188   exclusively defined effect of increasing cell size or generic synthetic gene expression,

189   respectively, is noteworthy and further discussed below.

190

191   Disruption in carbon metabolism and protein synthesis/folding appeared to dominantly be

192   associated with S_{low} and/or E_{low} phenotypes. Knockout strains sharing a S_{low} phenotype were

193   involved in nutrient and metal ion uptake, including phosphate (*pstA*, *pstC*), sulfur (*cysC*, *cysN*)

194   and zinc (*ZnuA*, *ZnuB*) (Fig. 2 – groups 8-10 red text). The S_{low}- E_{low} pattern was enriched with

195   genes associated with carbohydrate catabolism (*sucA*, *sucC*) and a critical regulator of stationary

196   phase onset (*dksA*). The lack of several cellular chaperones (Gene Ontology class GO:0006457,

197   Bonferroni corrected $p < 10^{-1}$) presented an exclusive E_{low} phenotype, thus indicating that a lack

198   of cellular protein folding functions negatively affected heterologous gene expression (Fig. 2

199   group 14, discussed below). Notably, these knockouts did not present a G_{spec} effect; therefore, the

200   effect on gene expression appeared to be global and did not affect individual genes differentially.

201

202   A S_{low} - E_{low} phenotype was also found for knockout strains for various genes involved in

203   bacterial chemotaxis (*cheY*, *motA*, *motB*). However, disruptions in flagellum assembly (GO:

204   0006935 and KEGG pathway flagellar biosynthesis) were linked to an exclusive E_{low} phenotype

205   (Group 14, Fig 2, Bonferroni corrected $p < 10^{-1}$). These findings indicated that cell motility and

206   global gene expression are co-regulated in *E. coli*. Indeed, there is compelling evidence that

207   widespread regulatory protein acetylation links nutrient scavenging and cell motility [15].

208   A disruption of ECA synthesis also showed an E_{low} phenotype, but, in contrast to chemotaxis

209   genes, this effect was combined and resulted in a S_{high} shift (Fig. 2 – Group 16). Because generic

210   cell growth delay appeared to lead to a S_{high} - E_{high} shift (Groups 2-3), the effects of deletions in

211   ECA structural components were likely to have occurred through a different mechanism

212   (discussed below) that led to defined and unique changes in the S-E features.

213

214   **A map of the Size-Expression interaction**

215   Most cell functions played a role in the Size-Expression interaction defined by the S and E

216   features. In this interaction, both features largely changed correspondingly (Fig 3, color gradient).

217   Several KEIO knockouts with a S_{high} – E_{high} phenotype carried a disruption in key bacterial

218   growth functions, including cell division, protein folding and DNA replication (Fig. 3 top-

219   middle). However, a S_{high} – E_{high} phenotype was not common among all 32 knockouts of genes

220   involved in cell division (GO:0051301), but among those responsible for outer membrane

221   invagination (cytokinesis), in which the Tol-Pal system plays an important role [16] (Sup. Fig. 11).

222   Bacterial DNA replication relies on a complex series of enzymatic and regulatory mechanisms. A

223   significant number of KEIO strains of genes associated with the *E. coli* GO class 'DNA-

224   dependent DNA replication' (GO:0006261) were found to be associated with a S_{high} – E_{high}

225   phenotype ($p < 10^{-4}$, 5 of 9 Sup. Fig. 12). The relationship between cytokinesis and chromosome

226   replication is not completely clear, but it is reasonable to conjecture that growth delay (S_{high}) may

227   indirectly increase cellular fluorescence (E_{high}) in these mutants. However, knocking out *seqA*, a

228   negative modulator of chromosome replication, led to a S_{low} – E_{high} phenotype, thus suggesting a

229   more direct mechanism.

230

231   For a number of strains, cell size and global gene expression did not vary together. For

232   example, some KEIO knockouts with a S_{low} – E_{high} phenotype indicated an important role in the

233   cellular SOS response (Fig. 2 Groups 8-9). However, a lack of other members associated with this

234   DNA repair function did not result in a univocal S-E phenotype ($p = 0.17$, Sup. Fig. 13). Other

235   knockouts with a S_{low} – E_{high} pattern did not have ribosome structural components (*rpsT*, *rpsU*,

236   *dbpA*) or important steps of RNA translation (*prfC*, *efp*) (Fig 3 – mid right). Twelve knockouts in

237   ribosome structural genes were present in the dataset and, significantly, 5 of them – *rpsU*, *rpsT*,

238   *rpmJ*, *rpmE*, *rplA* - showed a S_{low} – E_{high} phenotype ($p = 0.02$, Sup. Info.). After the analysis was

239   extended to all genes important for cellular translation (GO:0006412, 26 genes in dataset), the

240   link to a S_{low} – E_{high} pattern was even more substantial ($p < 10^{-2}$) (Sup. Fig. 14). Importantly,

241   18/26 of these genes also showed a significant G_{spec} phenotype ($p < 10^{-4}$), thus suggesting that

242   major gene expression disruption also differentially affects the expression of heterologous genes

243   apart from causing a small-cell phenotype.

244

245   The second important aspect of the Size-Expression interaction was that most cell functions

246   affected both sides. Indeed, only a limited set of related KEIO gene knockouts showed an

247   exclusive **S** or **E** phenotype. For example, a number of structural flagellar components and

248   various cellular chaperone proteins showed a predominantly $\mathbf{E_{low}}$ phenotype (Fig 3 – cyan). An

249   extension to all members of the Gene Ontology term class 'chemotaxis' (GO:0006935, 23 genes)

250   and 'bacterial-type flagellum' (GO:0009288, 23 genes) also yielded a significant decrease in **E** (p

251     $< 10^{-4}$ and $p < 0.005$, respectively, calculated by bootstrapping) (Sup. Fig. 15). Only a few

252   knockouts in this group (e.g., *flhE*, *flgG*, *fliQ*, Sup. Fig. 15) had an associated $\mathbf{S_{high}}$ value, a

253   pattern reminiscent of that of knockouts of genes involved in ECA biosynthesis (Fig 2 Group 16)

254   (Fig. 3 middle left). The inclusion of the 13 other gene members of the 'Enterobacterial Common

255   Antigen Biosynthetic Process' in the dataset further supported the phenotype for this group of

256   genes (GO:0009246) ($p < 10^{-4}$, Sup. Fig. 15).

257   In contrast, KEIO knockouts probably leading to increased RNA stability, such as those in

258   ribonuclease II (*rnb*) and polynucleotide phosphorylase (PNPase), which degrade various types of

259   mRNA [17], presented an exclusive $\mathbf{E_{high}}$ phenotype (Fig. 3 – dark blue). Finally, the lack of a

260   number of amino acid biosynthetic genes, and particularly those for aromatic amino acids,

261   presented a unique $\mathbf{S_{high}}$ phenotype (Fig. 3 top), which was confirmed after the analysis was

262   extended to all genes associated with the 'Cellular amino acid biosynthetic process' GO class

263   (GO:0008652, $p = 0.011$ calculated by bootstrapping, 86 genes in the dataset) (Fig. 3 – dark red).

264

Discussion

266   The present study describes how specific cellular activities influence the relationship

267   between a cell's ability to grow and its synthesis of neutral heterologous proteins. We found that

268   bacterial cell functions could be broadly subdivided into 4 systems of clearly defined phenotypes:

269   Membrane system, Ribosome-RNA-Protein stability (RRP) system, the Biosynthetic system and

270   the Energy-Motility system. Mutations in the Membrane system mainly lead to an increase in cell

271   size and are likely to have indirect effects on protein concentration, which manifest as changes in

272   gene expression. The RRP stability system primarily affects gene expression, with or without a

273   global (size) effect that depends on the structural nature (e.g., ribosome) of the disruption. The

274   lack of certain steps in the Biosynthetic system, particularly those for amino acids, are likely to

275   have a global impact simulating the bacterial stringent response with a broad remodeling of

276   cellular protein composition [18]. Finally, disruptions in the Energy-Motility system, which is

277   highly interconnected in the stationary phase by the global regulators ppGpp and dksA [19], may

278   exert a negative feedback on the ability of the cell to search for nutrients or increase gene

279   expression, for example via global protein acetylation [20,21], which leads to a small-cell low-

280   expression phenotype.

281   This study also shows that most mutations affecting cell functions, including cell motility,

282   membrane structure, chromosomal DNA replication, repair and homologous recombination, do

283   not differentiate between identically expressed synthetic genes (no G_{spec}). Not surprisingly,

284   however, gene knockouts with differential effects on the two reporter genes were primarily

285   related to protein expression and were associated with functions including ribosome structure,

286   translation (*efp*), protein folding (*cpxA*, *dnaK*) and transport (membrane TAT complex).

287

288   An important challenge in both metabolic engineering and synthetic biology is to precisely

289   understand how the introduction of engineered or non-native components into a biochemical

290   network influences the behavior of the entire system.

291   This study mapped how the removal of each cell function in *E. coli* influences the cell,

292   the synthetic genes introduced, or both. This accurate systematic analysis of the cellular context

293   of synthetic gene expression may facilitate metabolic engineering workflows and systems-level

294   modeling of this model prokaryote that serves as a key industrial workhorse organism.

295

Methods

296

Strains, plasmids and media

297

298   Single-gene knockout strains were obtained from the KEIO collection and wild type laboratory E. coli from

299   the Joint Bio-Energy Institute (JBEI, Emeryville-CA). For construction of plasmid library, strains were

300   cultivated in LB media supplemented with appropriate antibiotics and subjected to chemical ($CaCl_2$)

301   transformation. For flow cytometry and further analysis, all strains were cultured in Neidhardt's MOPS-

302   based Rich defined medium (Teknova), supplemented with 0.5% glucose and antibiotics Ampicillin or

303   Kanamycin (50-70 □g/ml). The construction of pEZ8-123 synthetic genetic probe has been previously

304   described [10].

305

Flow Cytometry

306

307   All data were acquired with a Guava Flow Cytometer (FC) as follows: 24-36 strain cultivations at a time

308   (2-3 rows for each 96 well plate) were inoculated at a 1:80 dilution from overnight-cultures in warm MOPS

309   rich medium supplemented with 0.5% glucose, and grown for exactly 1h15', in a shaker incubator at 37C.

310   At the end of the incubation, the OD of the culture was read with the help of a microtiter plate reader to

311   determine the number of cultures at mid-exponential phase (a 0.3-1.0 range corresponded to a

312   spectrophotometer OD reading of ~1.3-2.0). These cultures were diluted 1:200 in PBS + translation

313   inhibitor, and single-cell readings acquired via flow cytometry.

314

315   **Statistical and computational analysis**

316   Programming software R was utilized for statistical and other type of data analysis. Please refer to

317   supporting information for a detailed description of methodologies and functions.

318

319   ## Supporting information

320   Refer to the web version for supplementary material and information. This file provides in depth

321   description of experimental and statistical methodology, as well as supporting figures and tables.

322

323

324   ## Acknowledgments

325   I would like to thank Prof. Adam Arkin (University of California-Berkeley, USA) for financial

326   and intellectual support throughout the study. Without inspirational and extensive discussions

327   with Prof. Arkin over the years this study could not have been completed. I also want to thank Dr.

328   Marcin Joachimiak for protracted dialogues on the appropriate analysis and results significance.

329

330   This work was funded by the National Science Foundation as part of the Synthetic Biology

331   Engineering Research Center grant number 04570/0540879.

332

333   **Competing financial interests**

334   The authors have no competing financial interests to declare.

335

336

337   ## Bibliography

338   (1) Arkin, A. P. (2013) A wise consistency: engineering biology for conformity, reliability, predictability.

339   *Curr. Opin. Chem. Biol. 17*, 893–901.

340   (2) Cardinale, S., and Arkin, A. P. (2012) Contextualizing context for synthetic biology--identifying causes

341   of failure of synthetic biological systems. *Biotechnol. J. 7*, 856–66.

342   (3) Sleight, S. C., Bartley, B. A., Lieviant, J. A., and Sauro, H. M. (2010) Designing and engineering

343   evolutionary robust genetic circuits. *J. Biol. Eng. 4*, 12.

344   (4) Mutalik, V. K., Guimaraes, J. C., Cambray, G., Lam, C., Christoffersen, M. J., Mai, Q.-A., Tran, A. B.,

345   Paull, M., Keasling, J. D., Arkin, A. P., and Endy, D. (2013) Precise and reliable gene expression via

346   standard transcription and translation initiation elements. *Nat. Methods 10*, 354–60.

347   (5) Wu, C.-H., Lee, H.-C., and Chen, B.-S. (2011) Robust synthetic gene network design via library-based

348   search method. *Bioinformatics 27*, 2700–6.

349   (6) Woodruff, L. B. A., Boyle, N. R., and Gill, R. T. (2013) Engineering improved ethanol production in

350   Escherichia coli with a genome-wide approach. *Metab. Eng. 17*, 1–11.

351   (7) Zeitoun, R. I., Garst, A. D., Degen, G. D., Pines, G., Mansell, T. J., Glebes, T. Y., Boyle, N. R., and Gill,

352   R. T. (2015) Multiplexed tracking of combinatorial genomic mutations in engineered cell populations. *Nat.*

353   *Biotechnol. 33*.

354   (8) Brunk, E., George, K. W., Alonso-Gutierrez, J., Thompson, M., Baidoo, E., Wang, G., Petzold, C. J.,

355   McCloskey, D., Monk, J., Yang, L., O'Brien, E. J., Batth, T. S., Martin, H. G., Feist, A., Adams, P. D.,

356   Keasling, J. D., Palsson, B. O., and Lee, T. S. (2016) Characterizing Strain Variation in Engineered E. coli

357   Using a Multi-Omics-Based Workflow. *Cell Syst. 2*, 335–346.

358   (9) Klumpp, S., Zhang, Z., and Hwa, T. (2009) Growth Rate-Dependent Global Effects on Gene Expression

359   in Bacteria. *Cell 139*, 1366–1375.

360   (10) Cardinale, S., Joachimiak, M. P., and Arkin, A. P. (2013) Effects of genetic variation on the E. coli

361   host-circuit interface. *Cell Rep. 4*, 231–7.

362   (11) Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M.,

363   Wanner, B. L., and Mori, H. (2006) Construction of Escherichia coli K-12 in-frame, single-gene knockout

364   mutants: the Keio collection. *Mol. Syst. Biol. 2*, 2006.0008.

365   (12) Ormerod, M. (2000) A Practical Approach, in *Flow Cytometry* 3rd ed. Oxford University Press.

366   (13) Shapiro, H. (2003) Practical Flow Cytometry 4th ed. Wiley-Liss.

367   (14) Robertson, B., Button, D., and Koch, A. (1998) Determination of the biomasses of small bacteria at

368   low concentrations in a mixture of species with forward light scatter measurements by flow cytometry.

369   *Appl. Environ. Microbiol. 64*, 3900–9.

370   (15) Wolfe, A. J. (2005) The Acetate Switch. *Microbiol. Mol. Biol. Rev. 69*, 12–50.

371   (16) Gerding, M. A., Ogata, Y., Pecora, N. D., Niki, H., and De Boer, P. A. J. (2007) The trans-envelope

372   Tol-Pal complex is part of the cell division machinery and required for proper outer-membrane invagination

373   during cell constriction in E. coli. *Mol. Microbiol. 63*, 1008–1025.

374   (17) Arraiano, C. M., Yancey, S. D., and Kushner, S. R. (1988) Stabilization of discrete mRNA breakdown

375   products in ams pnp rnb multiple mutants of Escherichia coli K-12. *J. Bacteriol. 170*, 4625–4633.

376   (18) Scott, M., Gunderson, C. W., Mateescu, E. M., Zhang, Z., and Hwa, T. (2010) Interdependence of Cell

377   Growth. *Science (80-.). 330*, 1099–1102.

378   (19) Magnusson, L. U., Gummesson, B., Joksimović, P., Farewell, A., and Nyström, T. (2007) Identical,

379   independent, and opposing roles of ppGpp and DksA in Escherichia coli. *J. Bacteriol. 189*, 5193–5202.

380   (20) Castaño-cerezo, S., Bernal, V., Post, H., Fuhrer, T., Cappadona, S., and Nerea, C. (2014) Protein

381   acetylation affects acetate metabolism , motility and acid stress response in Escherichia coli. *Mol. Syst.*

382   *Biol.* 1–15.

383   (21) Liang, W., Malhotra, A., and Deutscher, M. P. (2011) Short Article Acetylation Regulates the Stability

384   of a Bacterial Protein □: Growth Stage-Dependent Modification of RNase R. *Mol. Cell 44*, 160–166.
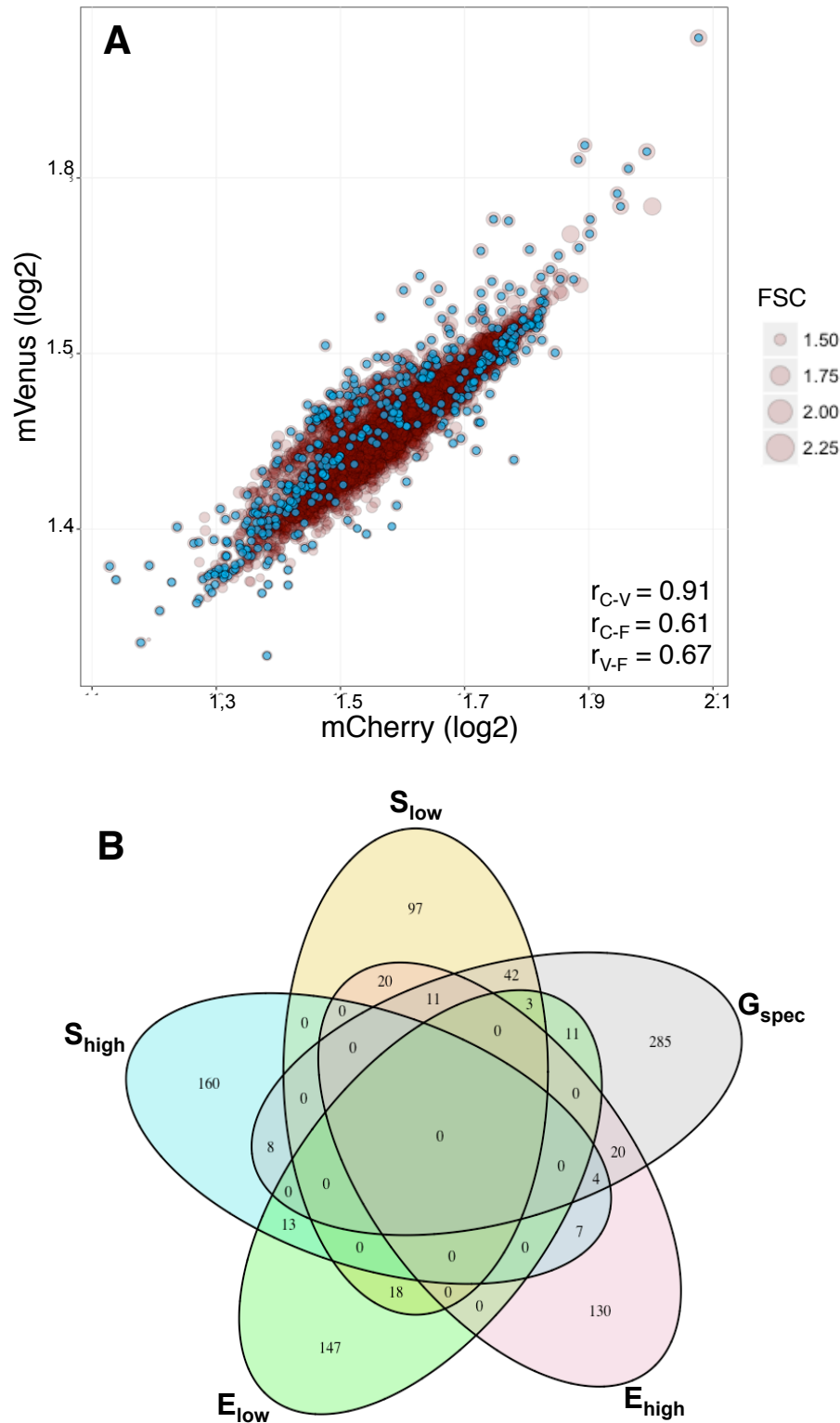
385

Figure 1

Figure 1. (A) Scatterplot of the mean population single-cell measure of cell volume (**FSC**) and fluorescence of mVenus and mCherry reporters (cyan dots: significant knockouts after FSC regression; r: Pearson correlation coefficients). **(B)** Venn diagram of the overlap between $S_{high/low}$, $E_{high/low}$ and G_{spec} extreme value genes.

	DAVID enrichment
E$_{\text{high}}$	*ruvA, priC, holD (hom. Rec.)* *holD, guaA, purA, puF, gpt* (Pur. Metabolism $5.0*10^{-1}$) *holD, pyrB, cmK, upp (pyr. Met. $4.6*10^{-1}$)*
E$_{\text{high}}$ **G**$_{\text{spec}}$	*purH, guaB, pgm, purL, pnp* (Pur. Metabolism $3.0*10^{-'}$) *tatB, tatC* (Tat complex transport $9.2*10^{-1}$)
G$_{\text{spec}}$	*metF, purE, gmhB, serA, atpB, atpE, atpA, nanK, manX, galM, asnB, coaE, pyrD, fdoI, glnA, hpt, icd, idi, mazG, edd, purD, nadC, astD, miaA, ubiE* (Met. Path. $4.8*10^{-1}$) *aroD, sdaAB, argC, yfbQ, trpE, tyrA, metC, cysK, cysM, lysA, proB, thrB, hisH, trpB* (AA Bios. $3.8*10^{-3}$) *sdaAB, metC, cysK, cysM, cysC, cysN, cysJ* (Cys and Met. Met.) *aroD, gmd, ansB, sdaAB, argC, pntA, nuoN, ugd, acs, add, cysC, dadX, yfbQ, trpE, tyrA, ybcF, metC, cysK, cysM, lysA, gatC, proB, glk, gshB, glgA, thrB, hisH, ybhC, cysG, cysN, cysJ, putA, trpB* (Met. Path. $3.6*10^{-1}$) *yeiL, basR, dsdC, galR, glcC, mraZ, fnr, nikR, glnB, yfhA, uvrY, slyA, putA* (Transr. DNA-templated)
E$_{\text{low}}$	*wzxE, rffA, rfe, rffE, rffM* (ECA Bios. Proc. $2.4*10^{-2}$) *flgH, flhB, fliK, motA, motB, fliK, fliG, fliI, flhD* (flagellar assembly $5.6*10^{-3}$) *motA, motB, fliG, trg, tap* (chemotaxis $4.4*10^{-2}$) *ahpC, katE, osmC* (response to oxidative stress $3.5*10^{-1}$) *rcsA, cpxR, phoP, ampC, motA, trg, tap, basS, zraS, creC, flhD* (TCS $3.3*10^{-1}$) *oppA, ydgI, ygiS, eutL* (AA transport and metabolism $3.5*10^{-1}$) *fimA, fimF, fimH, sfmF, ydeQ* (cell adhesion $7.4*10^{-1}$)

Table 1 Significantly enriched GO Biological classes or KEGG pathways (Bonferroni-corrected p-value) within groups of genes with a significant shift in the **E** or **G**$_{\text{spec}}$ values.

Figure 2

Gr	C	E	G	n.	
1	1	0		67	(*aroB*, *aroC*, *aroL* GO:0009073) (*gpmI*, *leuA*, *visC*, *aroB*, *panD*, *ushA*, *aroC*, *entC*, *aroL* (Bios. Of Sec. Met.)
2	1	1		5	*dedD*, *holD*, nhaA, *gidB*, *rluD*
3	1	0		4	fdoI, *tolR*, hflD, *rpmG*
4	0	1		54	(*purA*, *purF*, *pyrB* GO:0044205) (*hupA*, *hupB* GO:0030261)
5	0	1		13	acnB, *cpxA*, *rnb*, *priB*, *aspC*, *purL*, exbD, *guaB*, *purH*, *ddlB*, *pnp*, oxyR, *envC*
6	1	1		4	*slt*, *tatB*, *tatC*, *dnaK*
7	−1	1		3	speB, dhaR, *dmsD*
8	−1	1		14	*rpe*, ubiF, *ruvA*, brnQ, *cmk*, *seqA*, *ptsH*, *upp*, *guaA*, *prfC*, wcaC, *rpsT*, *rpsU*, csiE
9	−1	0		20	*dbpA*, *argC*, hfq, *pphA*, cysG, nlpI, *metC*, *pntA*, *ppx*, ugpQ, *cysC*, *cysN*, ugd, *acs*, *phoU*, endA, *pstA*, *pstC*, *exoX*, slyA
10	−1	1		3	*tktA*, *pgm*, mscS
11	−1	0		63	*ego*, *metI*, *sapD*, *artI*, *artJ*, *proW*, *znuC*, *znuA*, *livJ*, *modF* (ABC transporters)
12	−1	0		5	fadR, *edd*, *znuB*, asnB, *efp*
13	−1	−1		2	*dksA*, *sucC*
14	−1	−1		15	*sucA*, *cheY*, helD, *clpP*, *rffG*, *motB*, *motA*, narJ, *sgcA*, *dppF*, *alsE*, *ulaB*, *ulaD*, *kdgT*, glxK
15	0	−1		112	*fliL*, *fliG*, *trg*, *tap* (GO:0006935), *flgH*, *flhB*, *fliK*, *fliG*, *fliI*, *flhD* (Flagellar assembly). ahpC, katE, osmC (GO:0006979), *clpX*, *hscA*, *cbpA*, *htpG* (GO:0006457)
16	1	−1		10	*rffH*, *spr*, *rfe*, *rffA*, *tdk*, *sfmF*, idnD, *rffE*, *wzxE*, *mpl*
17	0	−1		6	fadD, hdeD, *rsmB*, astD, *bioH*, cysB, *mltA*, malE

Figure 2. Distribution of genes in 17 phenotypic patterns (groups, **Gr**) with significantly (>2 st.dev.) increased (+1/pink) or decreased (-1/cyan) cell size (**S**), generic synthetic gene expression (**E**) and gene-specific effects (**G**). For each combination of **S**, **G** and **E** patterns listed are either the individual extreme genes (for <20 genes), or the functional enrichments (DAVID) of Gene Ontology classes (bold/GO) or KEGG pathways (underscored).

(Text color: brown = AA biosynthesis; orange = important to cell growth; red = nutrient uptake and catabolic reactions; cyan = motility and chemotaxis function; dark green = cell membrane structural component).
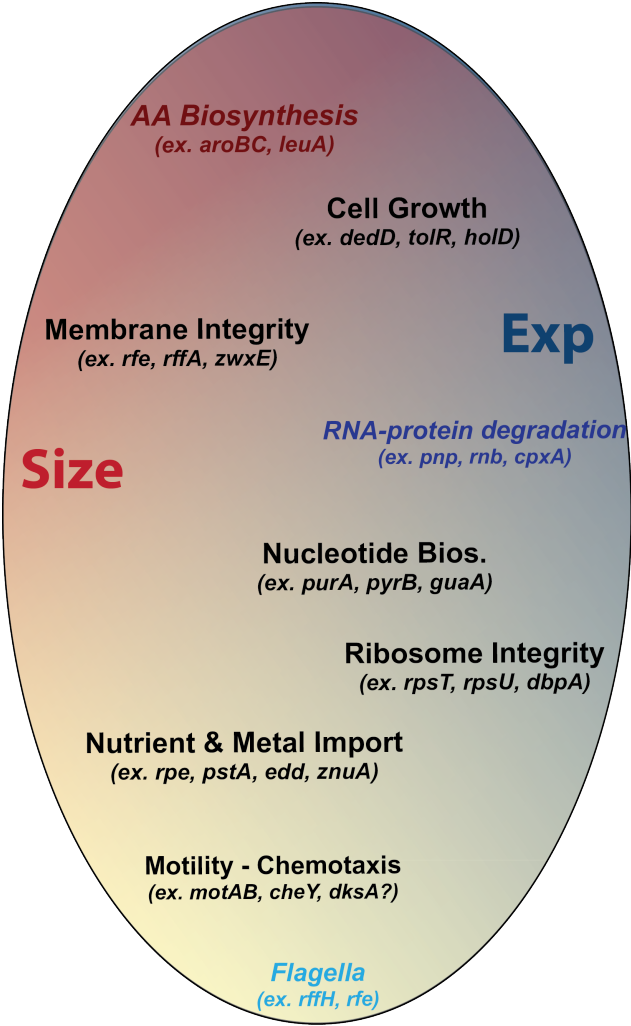
Figure 3

Figure 3. A map of the effect of major cellular systems on the interaction between cell **Size** (red) and synthetic gene **Exp**ression (blue). The impact of each cell functional category can be to increase (darker shade) or decrease (lighter shade) predominantly one side of the interaction or both (represented via the positioning towards respectively the edge or the inner of the oval).