1   **Forecasting bacterial survival-success and adaptive evolution through multi-omics stress-**
2   **response mapping, network analyses and machine learning**
3
4
5
6   Zeyu Zhu[1†], Defne Surujon[1†], Aidan Pavao[1], José Bento[2], and Tim van Opijnen[1*]
7
8
9   [1] Biology Department, Boston College, Chestnut Hill, MA, USA
10   [2] Computer Science Department, Boston College, Chestnut Hill, MA, USA
11
12
13
14
15
16   † Equal contribution
17   * Corresponding author
18
19
20
21
22
23   Keywords:
24   Tn-Seq
25   RNA-Seq
26   machine learning
27   experimental evolution
28   network analysis

1

ABSTRACT

Whether a bacterial pathogen establishes an infection and/or evolves antibiotic resistance depends on successful survival while experiencing stress from for instance the host immune system and/or antibiotics. Predictions on bacterial survival and adaptive outcomes could thus have great prognostic value. However, it is unknown what information is required to enable such predictions. By developing a novel network-based analysis method, a bacterium's phenotypic and transcriptional response can be objectively quantified in temporal 3D-feature space. The resulting trajectories can be interpreted as a degree of coordination, where a focused and coordinated response predicts bacterial survival-success, and a random uncoordinated response predicts survival-failure. These predictions extend to both antibiotic resistance and in vivo infection conditions and are applicable to both Gram-positive and Gram-negative bacteria. Moreover, through experimental evolution we show that the degree of coordination is an adaptive outcome - an uncoordinated response evolves into a coordinated response when a bacterium adapts to its environment. Most surprisingly, it turns out that phenotypic and transcriptional response data, network features and genome plasticity data can be used to train a machine learning model that is able to predict which genes in the genome will adapt under nutrient or antibiotic selection. Importantly, this suggests that deterministic factors help drive adaptation and that evolution is, at least partially, predictable. This work demonstrates that with the right information predictions on bacterial short-term survival and long-term adaptive outcomes are feasible, which underscores that personalized infectious disease diagnostics and treatments are possible, and should be developed.

INTRODUCTION

50

51 The ability to predict whether a bacterial pathogen is successfully establishing an infection, will

52 adapt to the stress it encounters in the host and/or progress to cause disease could have great

53 diagnostic value. However, it is unknown whether such predictions are entirely possible and

54 what information they would require. As a consequence, most diagnostics today come from a

55 physician's deductive reasoning, which can lead to sub-optimal antibiotic treatments and may

56 contribute to the emergence and spread of antibiotic resistance [1, 2]. Alternatively, in cancer

57 diagnostics transcriptional changes in specific genes of cancerous tissue, in addition to changes

58 in the host response, are used to provide prognostic information beyond standard clinical

59 assessment [3-6]. Moreover, integration of systems-level data, machine learning, and various

60 network/graph-based approaches have been employed to classify cancer subtypes and identify

61 subtype-specific drug targets, enhancing the diagnostic power of current approaches and leading

62 to more effective treatment options [7, 8]. With analogy to cancer diagnostics, a systems-wide

63 understanding of the state of a bacterial infection and how the infection may possibly progress

64 under pressure of the host-immune system and/or other stresses, could similarly aid in providing

65 targeted and personalized infectious-disease treatments.

66

67 Our previous work has indicated that advanced infectious-disease prognostics may be possible

68 by combining bacterial stress-response monitoring with network analyses [9]. A commonly

69 applied approach for characterizing bacterial stress responses is through RNA-Seq, which

70 measures genome-wide transcriptional changes upon an environmental perturbation. With the

71 advent of transposon-insertion sequencing (Tn-Seq), it has now also become relatively easy to

72 determine, on a genome-wide scale, the phenotypic importance of a gene, i.e. a gene's

73 contribution to fitness in a specific environment [10, 11]. Importantly, direct comparisons

74 between data from these different omics-approaches has shown, contrary to expectations, that

75 genes that change in transcription are poor indicators of what matters phenotypically. In other

76 words, phenotypically important and transcriptionally important genes (PIGs and TIGs) rarely

77 overlap [9, 12-18]. However, when integrated into a network, highly coordinated patterns

78 between PIGs and TIGs surface when the organism is challenged with an evolutionarily familiar

79 stress (i.e. one that has been experienced for many generations, e.g. nutrient depletion), while the

80 response becomes less coordinated when the bacterium is challenged with and responds to a

81   relatively new stress (e.g. antibiotics) [9]. This means that the degree of network coordination
82   between PIGs and TIGs originates from the bacterium's 'adaptive past' and should thus be
83   indicative of the degree to which the bacterium is adapted to a specific stress and will survive the
84   challenge (short-term survival outcome). Moreover, since evolution is a continuing process,
85   survival outcome - influenced by past adaptation - is ultimately related to future adaptive
86   outcomes; i.e. network coordination is indicative of where and how stress is experienced in the
87   genome, while selection drives adaptive evolution to resolve this stress. Thus, it may be possible
88   to predict where in the network innovation (adaptation) is most likely to occur to optimize
89   network coordination and increase survival success (long-term adaptive outcome).

90

91   Here we develop a novel integrated approach that combines genome-wide profiling, network
92   analyses and machine learning, which enables predictions on bacterial short-term survival and
93   long-term adaptive outcomes. As our model system, we use the respiratory pathogen
94   *Streptococcus pneumoniae*, which on a yearly basis causes ~1 million fatalities worldwide [19]
95   and ~4 million disease episodes in the US alone, among which ~40% are caused by strains that
96   are resistant to at least one antibiotic [20]. To develop this predictive strategy, we first establish
97   the transcriptionally and phenotypically important genes using temporal RNA-Seq and Tn-Seq
98   respectively in different *S. pneumoniae* strains that have different survival outcomes under
99   nutrient stress conditions and in the presence of antibiotics. By overlaying data onto newly
100   developed strain-specific networks and applying network analyses, we find that distinct network
101   patterns emerge that can be depicted as temporal trajectories that move through a specially
102   constructed feature space. Importantly, these patterns are predictive of whether or not a
103   bacterium is successfully surviving in its environment. Moreover, we apply the approach to *in
104   vitro* and *in vivo* data from *Pseudomonas aeruginosa,* highlighting its generalizability and the
105   possibility to predict bacterial survival-success in the host. Lastly, the development of a support
106   vector machine (SVM) leads to the ability to predict which genes acquire adaptive mutations
107   while adapting to nutrient stress or while evolving antibiotic resistance. This study shows that
108   infectious-disease prognostics is feasible through the implementation of different omics-
109   approaches, network analyses and machine learning, enabling the prediction of whether a
110   bacterium will survive or not under a given stress and where in the genome it is most likely to
111   adapt.

112 **RESULTS AND DISCUSSION**

113 **Strain specific metabolic networks are insufficient in defining nutrient dependency or**

114 **predicting survival outcomes in three strains of *S. pneumoniae.***

115 *Streptococcus pneumoniae* on average contains 2100 genes and harbors considerable genetic

116 diversity, with two strains differing on average by 250 genes (presence and absence), and a pan-

117 genome (collection of all genes across all strains) that is approximately double the size of the

118 genome of any given strain. *S. pneumoniae* designates ~30% of its genome to metabolic

119 functions, which enables growth on different carbon sources and in the presence and absence of

120 different substrates (e.g. amino acids, lipids). This 'strategy' all but guarantees the bacterium's

121 survival in a variety of host-niches, including the nasopharynx, inner-ear and lungs. Since

122 different host niches have different nutrient availability [21], nutrient depletion is evolutionarily

123 an important stress to the obligate non-motile human pathogen *S. pneumoniae* and has shaped its

124 genetic composition. We thus reasoned that strain-specific nutrient dependencies must exist and

125 that such dependencies can be used as a testing-ground to predict whether a strain will survive in

126 a specific environment and what information is needed to make such predictions.

127

128 Three strains (TIGR4 [T4], Taiwan-19F [19F] and D39) that differ in ~7% of their genetic

129 content (presence or absence of genes; [9]), were assayed to identify essential nutrients for

130 growth. Single nutrients were sequentially removed from a chemically defined medium (CDM)

131 and the effect on the growth rate was calculated. A nutrient is defined as essential if its removal

132 causes a >70% reduction in the bacterium's growth rate, and important if the reduction is

133 between 50-70% (Supplementary Figure 1, detailed explanation of definitions in this study can

134 be found in Supplementary Information). In total, four amino acids are essential to all three

135 strains: (L-Arginine, L-Cysteine, L-Histidine, and L-Leucine; Supplementary Figure 1A), while

136 6 nutrients have strain-specific requirements: 1) three amino acids (Glycine, L-Isoleucine and L-

137 Valine) and the nucleobase uracil are essential to D39; 2) Pantothenate is important to T4; 3) L-

138 Glutamine is important for T4 and D39 (Supplementary Figure 1A). At least two possible

139 explanations for this strain-specific nutrient dependency are that a strain either lacks certain

140 genes that are required to synthesize the nutrient or the respective metabolic network is

141 differentially wired. For instance, a metabolic gene might encode isoforms of an enzyme that

142 catalyze different reactions in different strains [22]. To determine the origin of the strain-specific

5

143   nutrient dependency we expanded the *S. pneumoniae* metabolic model we previously built for T4

144   [9] with two additional strain-specific models for D39 and 19F (Supplementary Figure 2;

145   Supplementary File 2). The three models are highly conserved, sharing 96% of all metabolic

146   genes across all strains (i.e. 431 metabolic genes/868 metabolic reactions), however, neither the

147   presence of strain-specific metabolism genes nor differences in the metabolic network topology

148   can sufficiently explain the observed strain-specific nutrient requirements.

149

150   **Genome-wide profiling reveals distinct transcriptional patterns between a nutrient**

151   **dependent strain and an independent strain.**

152   Genomic content and network architecture are thus not enough to consistently predict bacterial

153   survival and growth in a certain environment. We previously demonstrated that the degree of

154   network coordination between phenotypic and transcriptional responses distinguishes

155   evolutionarily familiar stresses from relatively novel ones [9]. Such network patterns could thus

156   be key to predicting whether a bacterium is successfully surviving in a specific environment.

157

158   To uncover genes that are phenotypically important (PIGs), we performed Tn-Seq on T4 in the

159   absence of either uracil, L-Valine or Glycine (i.e. nutrients essential for D39 but not T4). Tn-Seq

160   measures, in a highly quantitative fashion and on a genome-wide scale, which genes and

161   pathways are important for growth in a specific environment [11, 23]. By comparing fitness in

162   the presence and absence of a nutrient, genes that are important for T4's survival in the absence

163   of the nutrient are identified, which leads to a total of 134 PIGs that contribute to growth of T4

164   (15 genes for Glycine, 75 genes for uracil, 44 genes for L-Valine). All of these genes have

165   homologs in D39 and thus do not directly explain the different dependencies between T4 and

166   D39. Subsequently, we profiled the manner in which T4 and D39 transcriptionally respond to the

167   absence of the D39-specific essential nutrients. Genome-wide transcriptional responses were

168   determined by temporal RNA-Seq for T4 (the nutrient-independent strain) and D39 (the nutrient-

169   dependent strain) at 30 and 90 min after nutrient depletion (Supplementary Table 1).

170

171   Three distinct transcriptional patterns emerge that differentiate a nutrient-dependent from an

172   independent strain: 1) A dependent strain tends to trigger a greater number of expression changes

173   under nutrient depletion (Supplementary Table 2). For instance, in the absence of L-Valine or

174   Glycine, D39 triggers significantly more TIGs than T4 at both the early and the late time points

175   (two proportion Z-test, p<0.01) (Supplementary Table 2). Additionally, in the absence of uracil,

176   D39 and T4 trigger similar numbers of TIGs at 30min, however at 90min, the number of TIGs in

177   T4 decrease (from 22 to 13), while in D39 the number of differentially expressed genes increases

178   to 857 (nearly 40% of the genome); 2) In each single nutrient-depletion condition, magnitude

179   distributions of differential expression are significantly wider in D39 than in T4 (Figure 2A,

180   Kolmogorov-Smirnov test, p<0.01, Supplementary Table 2), indicating that the extent of

181   genome-wide transcriptional change is much larger in the dependent strain; 3) A functional

182   distribution analysis of TIGs shows that at 30 and 90 min after the depletion of Glycine or L-

183   Valine, and at 90 min after the depletion of uracil more TIGs per functional tag are differentially

184   regulated in the dependent strain (Figure 2B; Supplementary File 3). Furthermore, the TIGs are

185   distributed across more functional categories indicating that nutrient depletion has a greater

186   impact on most cellular systems of the dependent strain (Figure 2C; Supplementary File 3). If we

187   directly compare the TIGs of the independent with the dependent strain, it turns out that the T4-

188   TIGs (both early and late) are also TIGs in D39. This suggests that the dependent strain can raise

189   a similar 'appropriate response' as the independent strain to the endured stress. To obtain slightly

190   higher temporal resolution we additionally profiled 60 min after uracil depletion, which triggers

191   20 TIGs in D39, the majority of which are involved in uracil uptake (uracil permease SP_1286)

192   and the metabolic pathway that generates the pyrimidine precursor uridine monophosphate

193   (UMP) (SP_0701-0702, SP_0963-0964, SP_1275-1278, SP_1288; Supplementary File 3). These

194   exact uracil-related genes are also up-regulated in T4 and form the majority of T4's response to

195   uracil depletion at both early and late time points (Supplementary File 3, Figure 2D).

196   Importantly, this further shows that D39 is actually able to generate an appropriate

197   transcriptional response, but only over a limited amount of time. Instead, somewhere between 60

198   and 90 minutes D39's response is washed out by a rapidly expanding genome-wide

199   dysregulation (Figure 2E-F).

200

201   **Network analyses of the transcriptional and phenotypic responses can be visualized in a**

202   **temporal feature space and define survival as a coordinated response.**

203   To enable detailed network analyses and determine the degree of network coordination, the

204   strain-specific metabolic network models were converted into genetic networks where each gene

7

205   is represented as a node, and two gene nodes are connected if the proteins encoded by these

206   genes are involved in the same or in subsequent reactions. Overlaying TIGs and PIGs on the

207   network shows very little overlap and when genome-wide fitness is plotted against expression

208   change most genes distribute along the horizontal and vertical axes (Supplementary Figure 3).

209   This means that genes that change in expression rarely change in fitness, indicating that

210   transcriptional importance is a poor indicator of what matters phenotypically (Supplementary

211   Table 2, Supplementary Figure 3), which is consistent with our previous observations [9]. When

212   the independent and dependent strains' responses are plotted on a network, visual inspection

213   suggests that the independent response remains contained to a specific part of the network over

214   time (Figure 3A), while the dependent strain's response becomes increasingly scattered across

215   the entirety of the network (Figure 3B). In order to objectively quantify these responses we

216   devised three types of measurements that capture the defining network characteristics of a

217   response:

218   1) Connectedness (CC): the number of connected components is calculated by removing all

219   nodes from the network that are neither PIGs nor TIGs. This leaves a collection of sub-networks

220   (or components) that are separated and unreachable from one another. In a network sense, this

221   means that information may flow within a component but not between components due to

222   missing connections. The number of components thus explains the cohesiveness and continuity

223   of the response. For instance, in the absence of uracil in T4 we observe one large component

224   which corresponds to the UMP biosynthesis pathway, and several small (single-node)

225   components (Figure 3C). In contrast, the dependent D39-uracil at 90 min response is defined by

226   a large number of small components consisting of 1 or 2 genes (Figure 3D), however a large

227   dominating component consisting of 121 TIGs and PIGs is also observed (Figure 3D). This large

228   component potentially results from the presence of few highly connected "hub" genes. It is thus

229   important to evaluate whether the number of connected components formed in an observed

230   response are significantly different from a random response, which is achieved by permutation

231   testing (see Methods).

232   2) Closeness (CN): while a small number of components may indicate that a response is

233   contained to a few network modules, it is equally important to take into account the relative

234   position, or closeness, of the components, where highly related (sub)pathways are generally

235   closer to each other than unrelated pathways. This measure thus explains whether components

8

236   are functionally related and a response is targeted. For instance, out of the 13 components in the

237   T4 uracil depletion response 12 are only 2-3 edges away from their nearest component (Figure

238   3E), which is significantly smaller than the distances between randomized responses (obtained

239   through permutation testing). This indicates that the response, while not fully connected, is

240   contained and targeted in a relatively small area of the full metabolic network. In contrast, the

241   distances between the components of the D39-uracil response are not significantly smaller than a

242   random response (Figure 3F).

243   3) Representation (RE): while our network is limited to metabolism, the observed TIGs and PIGs

244   are genome-wide (Figure 2). For instance, D39's response to Glycine depletion is significantly

245   connected, however the metabolic portion of the response comprises only ~20% of the full

246   response. Importantly, since only the part of the response that falls on the network is considered,

247   the majority of the response in this case is thus ignored. This heavily skewed off-network

248   response is problematic because while the 20% on-network may give an indication of being

249   connected and/or close, in reality the true response could be random. This is illustrated with

250   respect to the earlier observation that even though the dependent strain may trigger an

251   appropriate transcriptional response that suggests survival-success, when the entire response is

252   considered it becomes clear that the transcriptional dysregulation is scattered across many other

253   non-metabolic pathways, processes and genes that are overwhelming the "appropriate response"

254   (Figure 2). To account for this, the RE is calculated, which defines a response as "metabolically

255   represented" if a significant proportion of the responsive genes fall on the metabolic network

256   (see Methods).

257   Lastly, to incorporate the manner in which the response changes over-time the log-transformed

258   p-values for CC, CN and RE calculated from each time point are plotted in a feature space,

259   where each of the three characteristics are placed along separate axes (Figure 3G; Supplementary

260   Figure 4). In this scheme, the region around the origin (grey box, Figure 3G) represents a

261   response that is non-significant in terms of CC, CN, and RE.

262

263   For all three depletion conditions (L-Valine, Glycine and uracil), the response of the nutrient-

264   independent strain (T4) tends to move away from the origin over time, and the responses are

265   characterized by significant CC, CN and/or RE (Figure 3H-J). In contrast, the nutrient-dependent

266   D39's responses are mostly confined to the non-significant regions near the origin (Figure 3H-J).

267   This is especially well illustrated by the uracil depletion experiment, where T4 and D39 strains

268   are situated at a very similar location at 30 min (Figure 3J). However, while the independent

269   strain T4 moves towards a higher CC, CN and RE, D39 moves in the opposite direction and into

270   the non-significant space. Thus, coordination between the transcriptional and phenotypic

271   response is maintained and strengthened over time in strains that can tolerate and survive in a

272   particular environment but weakened in strains that cannot (Figure 3H-J). Importantly, this

273   trajectory reinforces quantitatively what was suggested by the transcriptional response where

274   both strains start out in a very similar manner, and while the T4 response remains targeted, the

275   D39 response ends in uncoordinated dysregulation (Figure 2-uracil depletion). The temporal

276   trajectory formed by three network parameters (CC, CN, RE) thus characterizes the stress-

277   response of a strain as coordinated or uncoordinated, corresponding to survival success or

278   failure.

279

280   **Experimental evolution of a sensitive strain reverts nutrient dependencies and rewires**

281   **stress responses into a coordinated response.**

282   In order to test whether a dependent strain that becomes adapted to the absence of a nutrient (i.e.

283   it becomes independent) acquires network coordination, two short-term evolution experiments

284   were designed in which D39 was adapted to grow in the absence of uracil or L-Valine separately.

285   Four replicate populations were established for each experiment and cultured by serial passaging

286   in CDM in which either nutrient was decreased by approximately 15% every 3 days until

287   populations were obtained that are able to robustly grow in the absence of either nutrient (~40

288   generations each; Supplementary Figure 1C; Figure 4A).

289

290   To determine the adapted strains' transcriptional response, temporal RNA-Seq was performed on

291   a uracil-adapted (aD39-uracil) and a L-Valine-adapted strain (aD39-val) in the presence and

292   absence of the respective nutrient. Similar to the original independent strain T4, the two adapted

293   D39 strains now exhibit only a small number of differentially expressed genes (Supplementary

294   Table 2; Supplementary File 3), the magnitude of differential expression has a narrow

295   distribution, and TIGs in the adapted strains show specific function distributions similar to the

296   'original' independent strain T4 (Figure 4C and Figure 2B). On a network level, coordination

297   profiles and trajectories arise that are highly similar to T4 (Figure 4D and E). For instance, the

298 trajectory of aD39-val tracks along a higher RE and CC, resembling T4 (Figure 4D), and the

299 trajectory of aD39-uracil moves in the opposite direction of D39 with higher CC, CN and RE

300 and is almost indistinguishable from T4 (Figure 4E). Our analyses thus show that adaptation to

301 nutrient depletion stress leads to transcriptional rewiring and that adapted strains gain highly

302 targeted and coordinated responses, predictive of their ability to survive in an environment.

303

304 **Rewiring of genome-wide transcriptional and phenotypic responses to achieve coordination**

305 **extends to the evolution of antibiotic resistance.**

306 To test if network trajectories can also predict survival outcomes in a more complex

307 environment, we extended our approach to the evolution of antibiotic-resistance by challenging

308 T4 with vancomycin. Vancomycin is often used in treating infections caused by beta lactam-

309 resistant *S. pneumoniae* especially during sepsis and meningitis [24, 25]. The MIC of T4 is

310 0.24ug/mL (Supplementary Figure 1C) and in order to obtain a vancomycin-adapted strain a

311 short-term evolution experiment was performed. Four replicate populations were adapted to

312 vancomycin for ~70 generations (Supplementary Figure 1C), and an adapted strain (aT4-vanc)

313 was isolated, which can grow at 1xMIC with a relative fitness of $W_{aT4-vanc} = 0.88$ compared to the

314 no drug control (Figure 5A; i.e. a 12% relative growth defect). Fluorescence microscopy on T4

315 (wild-type) and aT4-vanc reveal significantly longer cell chains for T4 in the absence of

316 vancomycin (p<0.0001 in t-test; Figure 5B and C). After one-hour exposure to vancomycin

317 (1xMIC), the wildtype loses the long chain morphology and often exhibits a bulging phenotype

318 (Figure 5B), which is in agreement with previous reports [26], while aT4-vanc cells under

319 vancomycin treatment are indistinguishable from untreated cells (Two-sample t-test, p=0.6001)

320 confirming their adapted state.

321

322 The transcriptional response of T4 and aT4-vanc was determined with RNA-Seq at six time

323 points post-vancomycin treatment (10, 20, 30, 45, 60, and 90 min at 1xMIC). Overall, the

324 distinct patterns that are observed under nutrient-depletion are observed in the presence of

325 vancomycin as well: 1) aT4-vanc triggers fewer differential expression than T4 (Supplementary

326 Table 2); 2) aT4-vanc has significantly narrower magnitude distributions of differential

327 expression (Figure 5D, Supplementary Table 2; Kolmogorov-Smirnov test, p<0.02); 3) aT4-vanc

11

328   triggers significantly fewer TIGs in most functional tags (Figure 5E and Supplementary Table 2;

329   p<0.002 with Bonferroni correction for multiple testing).

330

331   To generate the phenotypic response and enable network analyses Tn-Seq was performed in the

332   presence of vancomycin, which, as expected, reveals little overlap between PIGs and TIGs

333   (Supplementary Table 2). The CC, CN and RE trajectories for T4 and aT4-vanc at 1xMIC start at

334   very similar coordinates in the feature space with high RE (Figure 6A). However, T4 rapidly

335   transitions to a less-represented space, displaying an erratic trajectory that ends in a non-

336   significant and uncoordinated response, indicative of survival-failure. On the other hand, aT4-

337   vanc moves away from the origin, to a state where it is significantly connected, close and

338   represented over the first 30 minutes. Between 30 and 90 minutes, aT4-vanc then follows an arc

339   where it gradually becomes less represented, close or connected, and eventually ends just below

340   the significance threshold for all three characteristics (Figure 6A). Thus, while aT4-vanc can

341   maintain a highly coordinated response for at least 60 minutes, this coordination is still partially

342   lost at the 90-minute time point, most likely because aT4-vanc is not fully adapted to

343   vancomycin, displaying a detectable growth defect in the presence of 1xMIC compared to the no

344   drug control (Figure 5A). We reasoned that at a higher vancomycin concentration, aT4-vanc

345   would start to behave more similarly to the sensitive T4 at 1xMIC. When challenged with

346   1.4xMIC of vancomycin, aT4-vanc initially shows a similar trajectory to 1xMIC (Figure 6A) but

347   traverses the same arc faster, i.e. at 1xMIC aT4-vanc traverses an arc over 60 minutes whereas at

348   1.4xMIC the traversal of the same arc is completed in 30-45 minutes. Finally, at 1.4xMIC,

349   between 45 and 90 minutes, the trajectory stays near the non-significant space. Thus, aT4-vanc at

350   1.4xMIC displays similarities to both the wild-type and aT4-vanc at 1xMIC where it has a

351   coordinated response at earlier time points but loses its coordination relatively fast (over fewer

352   number of line segments) and behaves erratically (similar to T4) at the later time points. This

353   means that, similar to nutrient-depletion, the direction of the trajectory but also the shape and the

354   speed at which it moves along a trajectory has predictive value concerning short-term survival

355   success under antibiotic exposure.

356

357   **Network coordination is predictive of survival outcome in other bacterial pathogens.**

12

358   In order to determine whether our findings are applicable to other bacterial species, network

359   analyses were extended to the evolutionarily distant opportunistic pathogen *Pseudomonas*

360   *aeruginosa* [27]. Tn-Seq and RNA-Seq data collected for strain PAO1 tested against 14

361   antimicrobials [28] and for strain PA14 tested in 2 *in vivo* wound infections (chronic and acute)

362   [17] were overlaid onto their respective strain-specific metabolic models [29]. In none of the 14

363   antimicrobial conditions PA14 elicits a coordinated response, i.e. CC, CN and RE of the PIGs

364   and TIGs are never significant (Figure 6B). On the other hand, during an infection, the

365   transcriptional and phenotypic responses of PAO1 are significant in RE on the metabolic

366   network, and in the case of an acute infection the response is also significant in CN (Figure 6B).

367   The higher coordination in the acute infection suggests that the pathogen is more likely to

368   survive in this condition. Indeed, acute burn infections tend to spread and deteriorate rapidly

369   [30], indicating a more successful outcome (at least with respect to short-term bacterial survival)

370   for the pathogen *P. aeruginosa,* and thus suggesting that network analyses can be applied to infer

371   disease progression, although more time-points would most likely be more informative.

372

373   **Integration of machine learning, genome-wide profiles, and network characteristics enables**

374   **prediction of adaptive evolution.**

375   Network analyses thus reveal where on the genetic network stress is experienced, while the level

376   of coordination is indicative of how stress is processed. Importantly, adaptive mutations are

377   generally localized in genetic regions that resolve (part of) the experienced stress. It may thus be

378   possible, that with the right information (e.g. where is stress experienced in the genome, how

379   evolvable is that part of the genome, how is it connected in a network context), we can predict

380   which parts of the genome are most likely to contribute to adaptive evolution. Since there are no

381   obvious patterns in our data (e.g. TIGs, PIGs, network connectivity) that are predictive of

382   adaptation we test this hypothesis by training a support vector machine (SVM) - one of the most

383   established supervised classifiers in machine learning [31], with the goal to develop a model that

384   is able to predict which genes will acquire adaptive mutations.

385

386   Adaptive mutations are defined as non-synonymous mutations in coding regions that went to

387   fixation or reached a frequency > 50% during experimental evolution in the absence of uracil and

388   L-Valine and in the presence of vancomycin, determined through whole-genome sequencing on

13

389 the adapted populations. In total, four mutations (in three genes) were identified in uracil-adapted
390 populations, three mutations (in two genes) in L-Valine-adapted populations, and seven
391 mutations (in five genes) in vancomycin-adapted populations (indicated by radial lines in
392 lavender in Figure 7A-C). The mutations' high frequency and condition-specificity are indicative
393 of their adaptive nature. Furthermore, in the nutrient (uracil and L-Valine) adapted populations
394 the mutated genes are involved in the metabolic pathways of the depleted nutrient
395 (Supplementary Table 3). Additionally, in the vancomycin adapted populations, mutated genes
396 are involved in capsule metabolism (SP_0350/*cps4E*), cell division/cell-wall synthesis
397 (SP_1067/*ftsW*), stringent response (SP_1645/*relA*), membrane transport (SP_1796), and
398 carbohydrate metabolism (SP_2107/*malM*). Although few cases of vancomycin
399 resistance/tolerance have been reported in *S. pneumoniae,* the capsule influences sensitivity to
400 this antibiotic [24, 32, 33], while reduced sensitivity to vancomycin has been reported in *relA*
401 mutants of other Gram-Positive cocci, including *Enterococcus faecalis* [34], vancomycin-
402 resistant *E.faecium* [35], *Staphylococcus aeurus* [36], and cell wall modifications (e.g.
403 thickening) are common features for vancomycin resistance [37, 38].

404

405 Genotypes of the mutated genes were compared to their homologous genes in 371 *S. pneumoniae*
406 strains that cover the variation present in the pan-genome [39]. Interestingly, the adaptive
407 mutations that arose in the nutrient experiments always resulted in the acquisition of the nutrient-
408 insensitive T4 genotype at these loci (Supplementary Table 3), which is also the shared genotype
409 among the majority of the pan-genome strains, indicative of most strains being tolerant to
410 nutrient deprivation of uracil and L-Valine. In contrast, adaptation to vancomycin results in the
411 acquisition of novel genotypes; i.e. none of the 371 strains carry any of the aT4-vanc mutations,
412 indicative of the fact that very few vancomycin-resistant/tolerant clinical strains have been
413 reported for *S. pneumoniae*. Despite this difference between nutrient and antibiotic adaptive
414 patterns, there are common features to all mutations from all three conditions. For instance, they
415 appear in highly conserved genes, i.e. core genes with high sequence similarity. Next, adaptation
416 data was overlaid with genome-wide profiles and sequence conservation data (Figure 7A-C) in
417 order to visually inspect whether adapted genes overlap with drastic phenotypic changes,
418 transcriptional changes and/or sequence conservation. For example, *carA* (SP_1275) is an
419 adapted gene in the uracil adaptation experiment and it also has both transcriptional and

14

420   phenotypic importance. While this is a suggestive pattern, at a genome-wide level it is hard to

421   detect such consistent patterns across all three experiments that could be indicative of other

422   likely candidates for adaptive evolution (Figure 7A-C).

423

424   To generate a classifier that is able to separate adapted genes (AGs) from non-adapted genes by

425   uncovering hidden patterns in our data, an SVM was built on the Tn-Seq and RNA-Seq profiles,

426   the network characteristics as well as the species-wide sequence conservation data. Importantly,

427   the latter datatype is included because sequence conservation is indicative of genomic plasticity,

428   i.e. it gives insight into the genomic regions that change the most/least and thereby potentially

429   influences the adaptability of each gene. Subsequently, the SVM was trained on the aggregation

430   of all adaptation experiments (uracil, L-Valine and vancomycin), with oversampling of the AGs

431   (see Methods). A total of 1409 data points and 18 features were used, with 10-fold cross-

432   validation and no parameter tuning. In total, 5 out of 6 adapted genes that are on our network are

433   successfully identified as adapted with 3 false positives and 1 false negative (Supplementary

434   Table 3). In cases where one class dominates the dataset (e.g. here we have >99% non-AGs) a

435   high accuracy can even be achieved by a naïve classifier that only selects the more numerous

436   class. Therefore, the observed accuracy of the classifier (99.69%) is compared to a naïve

437   classifier, which performs significantly worse (98.91%, Cohen's kappa=0.7128, p=0).

438   Furthermore, the sensitivity of the SVM (true positive rate: the proportion of true AGs that are

439   correctly identified) is 83.33%, the specificity (true negative rate: the proportion of true non-AGs

440   that are correctly identified) is 99.77% and the classifier achieved an AUROC (Area Under

441   Receiver Operating Characteristic curve, representing the tradeoff between true positive and

442   false positive rates) of 0.9978, which significantly outperforms a random classifier

443   (AUROC=0.5) and thus indicates that AGs are successfully distinguished from non-AGs (Figure

444   7D). Importantly, this means that adapted genes indeed share certain common features that are

445   not immediately obvious but can be detected using machine learning. Prior studies of adaptive

446   evolution focus on interpreting adaptive mutations only after they have been acquired, and these

447   interpretations are very specific to the selective pressure under which adaptation has happened in

448   a particular study [40-42]. Instead, the classifier presented here can make *a priori* predictions on

449   which genes will adapt under stress/selective pressure, regardless of the nature of this stress.

15

450   Thus, we demonstrate that incorporation of different data-types reveals that deterministic factors

451   exist that shape adaptive evolution thereby making it predictable.

452

453 **CONCLUSIONS**

454   An important goal here is to determine what type of data is needed to predict a bacterium's

455 chances of surviving in its environment. We show that a comparison of transcriptional responses

456 between a stress-sensitive and insensitive strain by itself shows stark differences in the number,

457 magnitude and functional tags that are involved in responding to the environment, which are

458 suggestive for their differences in survival-success. The full response thus carries important

459 information; however, a more granular analysis of the response is no less interesting. For

460 instance, both T4 and D39 respond very similarly early on to uracil depletion by 'appropriately'

461 upregulating expression of the UMP-pathway, and while T4 maintains a similar response over

462 time, D39's response is overwhelmed by genome-wide differential expression, resulting in

463 chaos. In addition, components of the stringent response (which is not understood in detail in *S.*

464 *pneumoniae*) such as genes involved in purine biosynthesis (SP_0044-0056) are down-regulated

465 in both T4 and D39 under amino acid depletion (L-Valine and Glycine; Supplementary file 3).

466 While this shows that particular response mechanisms are activated under stress, it turns out that

467 this is only a partial view. We show that by extending our focus and by taking the temporal

468 genome-wide response into account, it is possible to paint a global and detailed picture of how

469 the organism senses and processes stress. Moreover, we showed previously that it is important to

470 interrogate a bacterial response at both the transcriptional and phenotypic level to uncover

471 network patterns [9], and also here we find that PIGs are critical in enhancing our network

472 coordination analyses, especially when there are a few TIGs (Supplementary File 4). Overall our

473 strategy demonstrates that by integrating temporal transcriptional and phenotypic changes into

474 strain-specific networks, distinct patterns emerge that can be depicted as trajectories in feature

475 space. These temporal trajectories are composed of three types of measurements, Connectedness,

476 Closeness and Representation (CC, CN, RE) that capture the defining network characteristics of

477 a response and objectively quantify a strain's response into a degree of coordination that reflects

478 survival success. Importantly, we show that the degree of coordination is an evolvable trait;

479 when strains evolve the ability to grow in the absence of a nutrient, or when antibiotic resistance

480 emerges, the network is rewired, increasing coordination and unfolding a focused and targeted

481 response. In other words, selective pressure optimizes a strain's network coordination, which in

482 turn increases survival success; explaining why network coordination can be used to predict

483 short-term survival outcome. Past adaptation and future adaptive outcome are thereby intricately

17

484 linked, leading to the possibility of predicting where innovation (adaptation) in the network is
485 most likely to occur. Indeed, we show that by developing a support vector machine that
486 incorporates a wide array of data-types, genes that adapt can be distinguished from those that do
487 not. This indicates that with the right information, adaptation becomes a predictable process.
488

489 To improve on the short-term survival outcome and long-term adaptive outcome predictions, it is
490 likely that additional types of data as well as genome-wide networks will be beneficial. For
491 instance, epistatic and regulatory interactions have been shown to influence adaptive evolution
492 [43-46]. It is also possible to include information pertaining to the external environment that the
493 pathogen experiences into a predictive framework. The simultaneous transcriptomic profiling of
494 the host via dual RNA-Seq [47] and cytokine profiling (e.g. determining the state of the host
495 response can allow us to infer the magnitude of host-associated stress the pathogen is
496 experiencing) could also be informative and is something we are currently exploring. Along with
497 the host-response, the infection-causing pathogen potentially experiences competition or
498 participates in cooperation with the resident microbiota of the infection site, which can influence
499 the effectiveness of a given antimicrobial treatment [48]. Therefore, metagenomic profiling of
500 the microbiota from the site of infection may also aid in predicting the survival of a specific
501 pathogen.
502

503 To conclude, we demonstrate that network analyses and machine learning make short-term
504 survival outcome and long-term adaptive outcome predictable. Most importantly, the approach is
505 generalizable with respect to the applicability to Gram-positive and Gram-negative bacteria, the
506 emergence of antibiotic resistance, and the applicability to *in vivo* host infection. Thus, our
507 approach offers a primary gateway towards the development of highly accurate infectious
508 disease prognostics.
509
510

MATERIALS AND METHODS

Bacterial strains, culture media and growth curve assays

S. pneumoniae strain TIGR4 (T4; NC_003028.3) is a serotype 4 strain originally isolated from a Norwegian patient [49, 50], Taiwan-19F (19F; NC_012469.1) is a multi-drug resistant strain [51, 52] and D39 (NC_008533) is a commonly used serotype 2 strain originally isolated from a patient about 90 years ago [53]. All gene numbers refer to the T4 genome. Correspondence between homologous genes among the three strains and gene function annotations are described in Supplementary File 3. Unless otherwise specified, *S. pneumoniae* strains were cultivated in Todd Hewitt medium with 5% yeast extract (THY) with 5uL/mL oxyrase (Oxyrase, Inc) or on sheep's blood agar plates (Northeastern Laboratories) at $37^{o}C$ with 5% CO2. Tn-Seq and RNA-Seq experiments under nutrient-depletion and vancomycin conditions were performed in chemically defined medium (CDM; [9]) and semi-defined minimal medium (SDMM; [21]), respectively. Single strain growth assays were performed at least three times using 96-well plates by taking OD_{600} measurements on a Tecan Infinite 200 PRO plate reader.

Tn-Seq experiments, sample preparation and analysis

Six independent transposon libraries were constructed in T4 using transposon Magellan 6 as previously described [10, 11, 21]. Tn-Seq experiments under single nutrient depletion conditions were performed in CDM in the presence or absence of one of the three nutrients: Glycine, uracil and L-Valine. Vancomycin Tn-Seq experiment were performed in SDMM in the presence or absence of 0.1ug/mL vancomycin (MP Biomedicals).

Library preparation, Illumina sequencing, data processing and fitness calculations (W_i; representing the growth rate) were performed as previously described [10, 11, 21]. Genes with significant fitness change must satisfy three criteria: 1) Fitness of a gene must be calculated from at least three insertion mutants in both control and experimental conditions. 2) A gene must have a fitness difference greater than 15% ($|W_{Control}-W_{Experimental}|>0.15$). 3) $W_{Control}$ and $W_{Experimental}$ must significantly differ in a one sample t-test with Bonferroni correction for multiple testing.

Temporal RNA-Seq sample collection, preparation and analysis

19

541   In nutrient RNA-Seq experiments, T4, D39 and adapted D39 were collected at 30 and 90min

542   after depletion of D39-essential nutrients (Supplementary Table 1). In vancomycin RNA-Seq

543   experiment, T4 and adapted T4 were collected at 10, 20, 30, 45, 60 and 90min post-vancomycin

544   (1x MIC) treatment. Cell pellets were collected by centrifugation at 4000 rpm at 4°C and snap

545   frozen and stored at -80°C until RNA isolation by RNeasy Mini Kit (Qiagen). 400ng of total

546   RNA from each sample was used for generating cDNA libraries following the RNAtag-Seq

547   protocol [54] as previously described [9]. PCR amplified cDNA libraries were sequenced on an

548   Illumina NextSeq500 generating a high sequencing depth of ~7.5 million reads per sample [55].

549   RNA-Seq data was analyzed using an in-house developed analysis pipeline. In brief, raw reads

550   are demultiplexed by 5' and 3' indices [54], trimmed to 59 base pairs, and quality filtered (96%

551   sequence quality>Q14). Filtered reads are mapped to the corresponding reference genomes using

552   bowtie2 with the --very-sensitive option (-D 20 –R 3 –N 0 –L 20 –i S, 1, 0.50) [56]. Mapped

553   reads are aggregated by featureCount and differential expression is calculated with DESeq2 [57,

554   58]. In each pair-wise differential expression comparison, significant differential expression is

555   filtered based on two criteria: |log2foldchange| > 1 and adjusted p-value (padj) <0.05. All

556   differential expression comparisons are made between the presence and absence of the nutrient at

557   the same time point.

558

Experimental evolution and whole-genome sequencing

560   D39 and T4 were used as parental strains in nutrient-depletion and vancomycin evolution

561   experiments, respectively. Four replicate populations were grown in fresh CDM with decreasing

562   concentration of uracil or L-Val for nutrient adaptation populations, or increasing concentration

563   of vancomycin for antibiotic adaptation populations. Four replicate populations were serial

564   passaged in CDM as controls for background adaptations in nutrient adaptation experiments.

565   When populations had adapted a single colony was picked from each experiment, checked for its

566   adaptive phenotype by growth curve experiments. Genomic DNA was isolated from adapted

567   populations and single strains using a DNase Blood and Tissue kit (Qiagen), concentrations of

568   genomic DNA were measured on a Qubit 3.0 fluorometer (Invitrogen) and diluted to 5ng/uL for

569   library preparation using a Nextera kit (Illumina). Libraries were sequenced on an Illumina

570   NextSeq500 and reads were mapped to their corresponding reference genomes. Mutations were

571   identified using the breseq pipeline with polymorphism mode for populations and consensus

20

572   mode for adapted strains [59]. Adaptive mutations in each experiment are determined based on

573   the following criteria: 1) mutation frequency is greater than 50% in at least one replicate

574   population, and 2) this mutation is not present in any CDM-background adapted populations and

575   3) the mutation is a nonsense or missense mutation.

576

577   **Determination of relative minimal inhibitory concentration (MIC) by microdilution**

578   1 to 5 x 10^5 CFU of mid-exponential T4 in 100uL was diluted with 100uL of fresh medium with

579   vancomycin to achieve a gradient of final concentrations from 0 to 0.5ug/mL in 96-well plates.

580   Each concentration was tested in triplicates. Growth was monitored on a Tecan Infinite 200 PRO

581   plate reader at 37^oC for 16 hours. MIC is determined as the lowest concentration that abolishes

582   bacterial growth (Supplementary Figure 1C).

583

584   **Fluorescent microscopy**

585   Wild-type and vancomycin adapted T4 were grown to mid-exponential phase. Half of the culture

586   was left untreated, while the other half was exposed to 0.24ug/mL of vancomycin for 60 minutes.

587    $1x10^8$ CFUs were collected by centrifugation, resuspended in 20uL of PBS and stained with

588   Syto9 (DNA stain) and FM4-64 (cell membrane stain) for 10 minutes at room temperature. 1uL

589   of stained cells were imaged on an Olympus IX83 microscope system with an ORCA-Flash4.0

590   camera (Hamamatsu) and a 60x oil immersion objective. Phase contrast and fluorescence images

591   through GFP and RFP channels were taken for each sample. Microscopy of each sample was

592   repeated with at least three technical replicates. Images were modified for publication using Fiji

593   [60]. Cell numbers per chain was visually quantified based on 1000 *S. pneumoniae* chains from

594   each treatment group using at least three technical replicate micrographs.

595

596   **Strain-specific metabolic model construction**

597   Thirty-six reactions were manually added to the previously described T4 model [9] using the

598   COBRA toolbox based on updated information from three databases (NCBI, KEGG and BiGG)

599   and literature [61, 62]. Metabolite and reaction IDs were cross-referenced to follow the BiGG

600   naming convention [63]. Gene-reaction associations in the updated T4 metabolic model were

601   adjusted into three strain-specific models based on the correspondence table (Supplementary

602   Table 1). For visualization of the metabolic models, a map of the *S. pneumoniae* metabolism was

603   constructed using Escher [64] referencing the KEGG pathway base (Supplementary Figure 2).

604

Spectral Clustering of the *S. pneumoniae* pan-genome

606   Complete, annotated genomes from 22 reference strains (RefSeq 58) and contigs from 350

607   clinical strains [39] were assembled for data analysis. The contigs were annotated using the

608   PATRIC Genome annotation service to identify coding sequences [65]. A total of 820,754 amino

609   acid sequences from the 372 strains were assembled. In order to reduce redundancy and expedite

610   clustering, representative sequences were selected using a boundary forest algorithm [66], with

611   Smith-Waterman distance as the similarity measure. This decreased the number of sequences to

612   17,000 representatives. Pairwise distances between representatives were computed to generate a

613   sequence similarity matrix (S). The gaussian kernel of S was thresholded and transformed to an

614   adjacency matrix. Spectral clustering with normalization of the Laplacian was performed to

615   generate sequence clusters [67]. Since we had no prior knowledge of what the most appropriate

616   number of clusters would be, we scanned the range of 1000 to 10,000 clusters, and computed the

617   sum of squared errors (SSE) on all clusters, for each cluster set. SSE was minimized at 4300

618   clusters, therefore, we determined this to be the appropriate number of clusters of homologous

619   genes in the *S. pneumoniae* pan-genome. Sequences in each gene cluster were aligned using

620   Clustal Omega [68], and the average pairwise Smith-Waterman distance within each cluster was

621   computed. In the case of large clusters (containing >50 sequences), 50 random sequences were

622   selected for pairwise distance calculation. We define gene conservation as -log(mean(distance))

623   within a cluster, and count (number of strains that share the gene) of sequences in each cluster.

624

Network coordination analysis

626   We define 3 criteria for metabolic coordination: connectedness (CC), closeness (CN) and

627   representation (RE) in the metabolic network. Number of connected components (NCC) is used

628   as the metric for connectedness. For each experiment, connected components were determined

629   using the components function in the igraph package [69]. Since the expected NCC heavily

630   depends on the number of nodes selected, and the network architecture, in order to test whether

631   the observed NCC is significantly lower than expected, we apply permutation testing on random

632   selection of nodes on the network as follows: In an experiment with M responsive genes on the

22

633 network, we generate 1000 sets of M random genes, and compute the NCC for each permutation.

634 The empirical one-tailed p-value for this experiment is the proportion of permutations in which

635 we observed fewer NCC than the responsive genes in the experiment. A response is connected if

636 the empirical p-value for the NCC permutation testing is <0.01. To determine closeness of

637 responsive genes, the average length of shortest paths is computed for each pair of genes. Since

638 biological pathways may appear as long chains with few branches, it is possible to have a

639 connected component of TIGs and/or PIGs arranged in a line, with a high average pairwise

640 distance. In order to avoid such skew, we considered any responsive gene pairs that appear in the

641 same component to be at distance 0 to each other by assigning each edge on the network a

642 weight of 0 if it connects two responsive genes, and 1 otherwise. If there is no path connecting

643 the two components, the distance between this pair is replaced by the diameter of the network+1

644 (i.e. 21 in our network), to avoid infinite values. Similar to connectedness evaluation,

645 permutation testing is applied to the average network distance. A response is "close" if the

646 empirical p-value for the distance permutation testing is <0.01. To assess whether TIGs and PIGs

647 were significantly highly represented in the metabolic network we consider N, the total number

648 of responsive genes, and M, the subset of N that appear on the network. The probability of

649 observing M or more genes on the network, given N total responsive genes in the genome

650 ($p(m>M|N)$) is computed assuming a hypergeometric distribution. A response is metabolically

651 well-represented if this probability is <0.01.

652

653 **Support Vector Machine Classification of Adapted Genes**

654 A support vector machine (SVM) using a gaussian kernel is trained and cross-validated using the

655 fitcsvm function in MATLAB to distinguish whether a gene will contain adaptive mutations or

656 not. The model was trained on network parameters (degree, transitivity, centrality), TnSeq,

657 RNAseq and sequence conservation (count, or number of occurrences across the pan-genome,

658 and sequence similarity) of each gene. Data from the dependent (parental) strains from the uracil

659 (D39), L-Valine (D39) and vancomycin (T4) experiments were assembled into a set of 1283 data

660 points with 18 features that were standardized. Genes that were not represented on the metabolic

661 network were excluded. Each observation was then labeled as AG or non-AG. Because the

662 number of AGs is very small (6 out of 1283), we applied synthetic minority oversampling [70]

663 until 10% of the observations were AGs. The SVM was trained on a total of 1409 data points

664 (1283 experimental and 126 synthetic) using 10-fold cross-validation, and report the average

665 accuracy, kappa, precision and recall on the 10 cross-validation sets.

666

667 **Statistical analysis**

668 Quantification and statistical analysis are described in the above Method Details section,

669 Supplementary Table2 and in figure legends (Figures 2, 3, 5, S4).

670

671 **List of abbreviations**

672 AG: adapted gene

673 AUROC: area under receiver operating characteristic curve

674 CC: connectedness

675 CDM: chemically defined medium

676 CFU: colony forming unit

677 CN: closeness

678 MIC: minimum inhibitory concentration

679 NCC: number of connected components

680 PIG: phenotypically important gene

681 RE: representation

682 RNA-Seq: RNA-Sequencing

683 SDMM: semi-defined minimal medium

684 SSE: sum of squared errors

685 SVM: support vector machine

686 TIG: transcriptionally important gene

687 Tn-Seq: transposon insertion sequencing

688 UMP: uridine monophosphate

689

690 **Accession numbers.** The datasets generated during the current study are available as
691 Supplementary Files and in the Sequence Read Archive (SRX2039176, SRX2039177,
692 SRP156493 and SRP156489).

693

695 **Authors' contributions**. TvO devised the study, ZZ performed the wet-lab experiments, DS

696 performed the network analyses, clustering and machine learning. JB consulted on the clustering

697 and machine learning. AP constructed the metabolic maps. ZZ, DS and TvO analyzed data,

698 interpreted results and wrote the manuscript. ZZ, DS, AP, JB and TvO approved the final

699 manuscript.

700

701 **Acknowledgements**. DNA sequencing was performed at the Boston College Sequencing Core.

702 The authors wish to thank Jon Anthony for running the Aerobio sequencing analyses pipeline.

703

704 **Conflict of interest.**

705 The authors declare that they have no conflict of interest.

REFERENCES

1. Lopez-Gongora S, Puig I, Calvet X, Villoria A, Baylina M, Munoz N, Sanchez-Delgado J, Suarez D, Garcia-Hernando V, Gisbert JP: Systematic review and meta-analysis: susceptibility-guided versus empirical antibiotic treatment for Helicobacter pylori infection. *J Antimicrob Chemother* 2015, 70:2447-2455.

2. Scholte JB, Duong HL, Linssen C, Van Dessel H, Bergmans D, van der Horst R, Savelkoul P, Roekaerts P, van Mook W: Empirical antibiotic therapy for pneumonia in intensive care units: a multicentre, retrospective analysis of potentially pathogenic microorganisms identified by endotracheal aspirates cultures. *Eur J Clin Microbiol Infect Dis* 2015, 34:2295-2305.

3. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, et al: Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* 2001, 98:10869-10874.

4. van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, et al: Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002, 415:530-536.

5. van't Veer LJ, Paik S, Hayes DF: Gene expression profiling of breast cancer: a new tumor marker. *J Clin Oncol* 2005, 23:1631-1635.

6. Pentheroudakis G, Raptou G, Kotoula V, Wirtz RM, Vrettou E, Karavasilis V, Gourgioti G, Gakou C, Syrigos KN, Bournakis E, et al: Immune response gene expression in colorectal cancer carries distinct prognostic implications according to tissue, stage and site: a prospective retrospective translational study in the context of a hellenic cooperative oncology group randomised trial. *PLoS One* 2015, 10:e0124612.

7. Agrahari R, Foroushani A, Docking TR, Chang L, Duns G, Hudoba M, Karsan A, Zare H: Applications of Bayesian network models in predicting types of hematological malignancies. *Sci Rep* 2018, 8:6951.

8. Zaman N, Li L, Jaramillo ML, Sun Z, Tibiche C, Banville M, Collins C, Trifiro M, Paliouras M, Nantel A, et al: Signaling network assessment of mutations and copy number variations predict breast cancer subtype-specific drug targets. *Cell Rep* 2013, 5:216-223.

9. Jensen PA, Zhu Z, van Opijnen T: Antibiotics Disrupt Coordination between Transcriptional and Phenotypic Stress Responses in Pathogenic Bacteria. *Cell Rep* 2017, 20:1705-1716.

10. van Opijnen T, Bodi KL, Camilli A: Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nat Methods* 2009, 6:767-772.

11. van Opijnen T, Camilli A: Transposon insertion sequencing: a new tool for systems-level analysis of microorganisms. *Nat Rev Microbiol* 2013, 11:435-442.

12. Birrell GW, Brown JA, Wu HI, Giaever G, Chu AM, Davis RW, Brown JM: Transcriptional response of Saccharomyces cerevisiae to DNA-damaging agents does not identify the genes that protect against these agents. *Proc Natl Acad Sci U S A* 2002, 99:8778-8783.

13. Smith JJ, Sydorskyy Y, Marelli M, Hwang D, Bolouri H, Rachubinski RA, Aitchison JD: Expression and functional profiling reveal distinct gene classes involved in fatty acid metabolism. *Mol Syst Biol* 2006, 2:2006 0009.

14. Deutschbauer A, Price MN, Wetmore KM, Shao W, Baumohl JK, Xu Z, Nguyen M, Tamse R, Davis RW, Arkin AP: Evidence-based annotation of gene function in Shewanella oneidensis MR-1 using genome-wide fitness profiling across 121 conditions. *PLoS Genet* 2011, 7:e1002385.

15. Giaever G, Chu AM, Ni L, Connelly C, Riles L, Veronneau S, Dow S, Lucau-Danila A, Anderson K, Andre B, et al: Functional profiling of the Saccharomyces cerevisiae genome. *Nature* 2002, 418:387-391.

16. Price MN, Deutschbauer AM, Skerker JM, Wetmore KM, Ruths T, Mar JS, Kuehl JV, Shao W, Arkin AP: Indirect and suboptimal control of gene expression is widespread in bacteria. *Mol Syst Biol* 2013, 9:660.

17. Turner KH, Everett J, Trivedi U, Rumbaugh KP, Whiteley M: Requirements for Pseudomonas aeruginosa acute burn and chronic surgical wound infection. *PLoS Genet* 2014, 10:e1004518.

18. Keren L, Hausser J, Lotan-Pompan M, Vainberg Slutskin I, Alisar H, Kaminski S, Weinberger A, Alon U, Milo R, Segal E: Massively Parallel Interrogation of the Effects of Gene Expression Levels on Fitness. *Cell* 2016, 166:1282-1294 e1218.

19. O'Brien KL, Wolfson LJ, Watt JP, Henkle E, Deloria-Knoll M, McCall N, Lee E, Mulholland K, Levine OS, Cherian T, et al: Burden of disease caused by Streptococcus pneumoniae in children younger than 5 years: global estimates. *Lancet* 2009, 374:893-902.

20. CDC: Antibiotic Resistance Threats in the Unites States. Centers for Disease Control and Prevention Threat Report. 2013, 1:1-114.

21. van Opijnen T, Camilli A: A fine scale phenotype-genotype virulence map of a bacterial pathogen. *Genome Res* 2012, 22:2541-2551.

22. Gisk B, Wiethaus J, Aras M, Frankenberg-Dinkel N: Variable composition of heme oxygenases with different regiospecificities in Pseudomonas species. *Arch Microbiol* 2012, 194:597-606.

23. van Opijnen T, Lazinski DW, Camilli A: Genome-Wide Fitness and Genetic Interactions Determined by Tn-seq, a High-Throughput Massively Parallel Sequencing Method for Microorganisms. *Curr Protoc Mol Biol* 2014, 106:7 16 11-24.

24. Moscoso M, Domenech M, Garcia E: Vancomycin tolerance in clinical and laboratory Streptococcus pneumoniae isolates depends on reduced enzyme activity of the major LytA autolysin or cooperation between CiaH histidine kinase and capsular polysaccharide. *Mol Microbiol* 2010, 77:1052-1064.

25. Haas W, Kaushal D, Sublett J, Obert C, Tuomanen EI: Vancomycin stress response in a sensitive and a tolerant strain of Streptococcus pneumoniae. *J Bacteriol* 2005, 187:8205-8210.

26. Daly KE, Huang KC, Wingreen NS, Mukhopadhyay R: Mechanics of membrane bulging during cell-wall disruption in gram-negative bacteria. *Phys Rev E Stat Nonlin Soft Matter Phys* 2011, 83:041922.

27. Zavascki AP, Carvalhaes CG, Picao RC, Gales AC: Multidrug-resistant Pseudomonas aeruginosa and Acinetobacter baumannii: resistance mechanisms and implications for therapy. *Expert Rev Anti Infect Ther* 2010, 8:71-93.

28. Murray JL, Kwon T, Marcotte EM, Whiteley M: Intrinsic Antimicrobial Resistance Determinants in the Superbug Pseudomonas aeruginosa. *MBio* 2015, 6:e01603-01615.

796   29.   Bartell JA, Blazier AS, Yen P, Thogersen JC, Jelsbak L, Goldberg JB, Papin JA:
797   Reconstruction of the metabolic network of Pseudomonas aeruginosa to interrogate
798   virulence factor synthesis. *Nat Commun* 2017, 8:14631.

799   30.   McManus AT, Mason AD, Jr., McManus WF, Pruitt BA, Jr.: Twenty-five year review of
800   Pseudomonas aeruginosa bacteremia in a burn center. *Eur J Clin Microbiol* 1985, 4:219-
801   223.

802   31.   Hastie T, Tibshirani, R., Friedman, J: *The Elements of Statistical Learning: Data Mining,*
803   *Inference, and Prediction.* 2 edn. New York: Springer-Verlag; 2009.

804   32.   Fernebro J, Andersson I, Sublett J, Morfeldt E, Novak R, Tuomanen E, Normark S,
805   Normark BH: Capsular expression in Streptococcus pneumoniae negatively affects
806   spontaneous and antibiotic-induced lysis and contributes to antibiotic tolerance. *J Infect*
807   *Dis* 2004, 189:328-338.

808   33.   Henriques Normark B, Kalin M, Ortqvist A, Akerlund T, Liljequist BO, Hedlund J,
809   Svenson SB, Zhou J, Spratt BG, Normark S, Kallenius G: Dynamics of penicillin-
810   susceptible clones in invasive pneumococcal disease. *J Infect Dis* 2001, 184:861-869.

811   34.   Abranches J, Martinez AR, Kajfasz JK, Chavez V, Garsin DA, Lemos JA: The molecular
812   alarmone (p)ppGpp mediates stress responses, vancomycin tolerance, and virulence in
813   Enterococcus faecalis. *J Bacteriol* 2009, 191:2248-2256.

814   35.   Honsa ES, Cooper VS, Mhaissen MN, Frank M, Shaker J, Iverson A, Rubnitz J, Hayden
815   RT, Lee RE, Rock CO, et al: RelA Mutant Enterococcus faecium with Multiantibiotic
816   Tolerance Arising in an Immunocompromised Host. *MBio* 2017, 8.

817   36.   Geiger T, Kastle B, Gratani FL, Goerke C, Wolz C: Two small (p)ppGpp synthases in
818   Staphylococcus aureus mediate tolerance against cell envelope stress conditions. *J*
819   *Bacteriol* 2014, 196:894-902.

820   37.   Cui L, Ma X, Sato K, Okuma K, Tenover FC, Mamizuka EM, Gemmell CG, Kim MN,
821   Ploy MC, El-Solh N, et al: Cell wall thickening is a common feature of vancomycin
822   resistance in Staphylococcus aureus. *J Clin Microbiol* 2003, 41:5-14.

823   38.   Courvalin P: Vancomycin resistance in gram-positive cocci. *Clin Infect Dis* 2006, 42
824   Suppl 1:S25-34.

825   39.   Cremers AJ, Mobegi FM, de Jonge MI, van Hijum SA, Meis JF, Hermans PW, Ferwerda
826   G, Bentley SD, Zomer AL: The post-vaccine microevolution of invasive Streptococcus
827   pneumoniae. *Sci Rep* 2015, 5:14952.

828   40.   Marvig RL, Damkiaer S, Khademi SM, Markussen TM, Molin S, Jelsbak L: Within-host
829   evolution of Pseudomonas aeruginosa reveals adaptation toward iron acquisition from
830   hemoglobin. *MBio* 2014, 5:e00966-00914.

831   41.   Chen FJ, Lauderdale TL, Ho M, Lo HJ: The roles of mutations in gyrA, parC, and
832   ompK35 in fluoroquinolone resistance in Klebsiella pneumoniae. *Microb Drug Resist*
833   2003, 9:265-271.

834   42.   Tremblay PL, Summers ZM, Glaven RH, Nevin KP, Zengler K, Barrett CL, Qiu Y,
835   Palsson BO, Lovley DR: A c-type cytochrome and a transcriptional regulator responsible
836   for enhanced extracellular electron transfer in Geobacter sulfurreducens revealed by
837   adaptive evolution. *Environ Microbiol* 2011, 13:13-23.

838   43.   MacLean RC: Predicting epistasis: an experimental test of metabolic control theory with
839   bacterial transcription and translation. *J Evol Biol* 2010, 23:488-493.

840   44.   Trindade S, Sousa A, Xavier KB, Dionisio F, Ferreira MG, Gordo I: Positive epistasis
841   drives the acquisition of multidrug resistance. *PLoS Genet* 2009, 5:e1000578.

842 45. Aldana M, Balleza E, Kauffman S, Resendiz O: Robustness and evolvability in genetic
843   regulatory networks. *J Theor Biol* 2007, 245:433-448.
844 46. Conant GC, Wagner A: Convergent evolution of gene circuits. *Nat Genet* 2003, 34:264-
845   266.
846 47. Aprianto R, Slager J, Holsappel S, Veening JW: Time-resolved dual RNA-seq reveals
847   extensive rewiring of lung epithelial and pneumococcal transcriptomes during early
848   infection. *Genome Biol* 2016, 17:198.
849 48. Cookson W, Cox MJ, Moffatt MF: New opportunities for managing acute and chronic
850   lung infections. *Nat Rev Microbiol* 2018, 16:111-120.
851 49. Aaberge IS, Eng J, Lermark G, Lovik M: Virulence of Streptococcus pneumoniae in
852   mice: a standardized method for preparation and frozen storage of the experimental
853   bacterial inoculum. In *Microbial Pathogenesis*, vol. 18. pp. 141-152; 1995:141-152.
854 50. Tettelin H, Ciammola A, Rigamonti D, Leavitt BR, Goffredo D, Conti L, MacDonald
855   ME, Friedlander RM, Silani V, Hayden MR, et al: Complete Genome Sequence of a
856   Virulent Isolate of Streptococcus pneumoniae. In *Science*, vol. 293. pp. 493-498:
857   American Association for the Advancement of Science; 2001:493-498.
858 51. McGee L, McDougal L, Zhou J, Spratt BG, Tenover FC, George R, Hakenbeck R,
859   Hryniewicz W, Lefevre JC, Tomasz A, Klugman KP: Nomenclature of major
860   antimicrobial-resistant clones of Streptococcus pneumoniae defined by the pneumococcal
861   molecular epidemiology network. In *Journal of Clinical Microbiology*, vol. 39. pp. 2565-
862   2571: American Society for Microbiology; 2001:2565-2571.
863 52. Shi Z-y, Enright MC, Wilkinson P, Griffiths D, Spratt BG: Identification of Three Major
864   Clones of Multiply Antibiotic- Resistant. In *Journal of Clinical Microbiology*. pp. 1-6;
865   1998:1-6.
866 53. Lanie JA, Ng WL, Kazmierczak KM, Andrzejewski TM, Davidsen TM, Wayne KJ,
867   Tettelin H, Glass JI, Winkler ME: Genome Sequence of Avery's Virulent Serotype
868   2 Strain D39 of Streptococcus pneumoniae and Comparison with That of Unencapsulated
869   Laboratory Strain R6. In *J Bacteriol*, vol. 189. pp. 38-51; 2006:38-51.
870 54. Shishkin AA, Giannoukos G, Kucukural A, Ciulla D, Busby M, Surka C, Chen J,
871   Bhattacharyya RP, Rudy RF, Patel MM, et al: Simultaneous generation of many RNA-
872   seq libraries in a single reaction. *Nat Methods* 2015, 12:323-325.
873 55. Haas BJ, Chin M, Nusbaum C, Birren BW, Livny J: How deep is deep enough for RNA-
874   Seq profiling of bacterial transcriptomes? *BMC Genomics* 2012, 13:734.
875 56. Langmead B, Salzberg SL: Fast gapped-read alignment with Bowtie 2. *Nat Methods*
876   2012, 9:357-359.
877 57. Liao Y, Smyth GK, Shi W: featureCounts: an efficient general purpose program for
878   assigning sequence reads to genomic features. *Bioinformatics* 2014, 30:923-930.
879 58. Love MI, Huber W, Anders S: Moderated estimation of fold change and dispersion for
880   RNA-seq data with DESeq2. *Genome Biol* 2014, 15:550.
881 59. Deatherage DE, Barrick JE: Identification of mutations in laboratory-evolved microbes
882   from next-generation sequencing data using breseq. *Methods Mol Biol* 2014, 1151:165-
883   188.
884 60. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S,
885   Rueden C, Saalfeld S, Schmid B, et al: Fiji: an open-source platform for biological-image
886   analysis. *Nat Methods* 2012, 9:676-682.

887   61.   Denapaite D, Bruckner R, Hakenbeck R, Vollmer W: Biosynthesis of teichoic acids in
888   *Streptococcus pneumoniae* and closely related species: lessons from genomes. *Microb*
889   *Drug Resist* 2012, 18:344-358.
890   62.   Becker SA, Feist AM, Mo ML, Hannum G, Palsson BO, Herrgard MJ: Quantitative
891   prediction of cellular metabolism with constraint-based models: the COBRA Toolbox.
892   *Nat Protoc* 2007, 2:727-738.
893   63.   Schellenberger J, Park JO, Conrad TM, Palsson BO: BiGG: a Biochemical Genetic and
894   Genomic knowledgebase of large scale metabolic reconstructions. *BMC Bioinformatics*
895   2010, 11:213.
896   64.   King ZA, Drager A, Ebrahim A, Sonnenschein N, Lewis NE, Palsson BO: Escher: A
897   Web Application for Building, Sharing, and Embedding Data-Rich Visualizations of
898   Biological Pathways. *PLoS Comput Biol* 2015, 11:e1004321.
899   65.   Wattam AR, Davis JJ, Assaf R, Boisvert S, Brettin T, Bun C, Conrad N, Dietrich EM,
900   Disz T, Gabbard JL, et al: Improvements to PATRIC, the all-bacterial Bioinformatics
901   Database and Analysis Resource Center. *Nucleic Acids Res* 2017, 45:D535-D542.
902   66.   Mathy CD, N; Bento, J; Rosenthal, J; Yedidia, J: The Boundary Forest Algorithm for
903   Online Supervised and Unsupervised Learning. *Proceedings of the 29th National*
904   *Conference on Artificial Intelligence (AAAI)* 2015.
905   67.   Yu SX, Shi J: Segmentation given partial grouping constraints. *IEEE Trans Pattern Anal*
906   *Mach Intell* 2004, 26:173-183.
907   68.   Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H,
908   Remmert M, Soding J, et al: Fast, scalable generation of high-quality protein multiple
909   sequence alignments using Clustal Omega. *Mol Syst Biol* 2011, 7:539.
910   69.   Csardi GN, T: The igraph software package for complex network research. *InterJournal,*
911   *Complex Systems* 2006, 1695.
912   70.   Taft LM, Evans RS, Shyu CR, Egger MJ, Chawla N, Mitchell JA, Thornton SN, Bray B,
913   Varner M: Countering imbalanced datasets to improve adverse drug event predictive
914   models in labor and delivery. *J Biomed Inform* 2009, 42:356-364.
915

916 **FIGURE LEGENDS**

917 **Figure 1. Study overview. A.** Differential survival outcomes under nutrient depletion, antibiotic

918 exposure and *in vivo* conditions from *Streptococcus pneumoniae* and *Pseudomonas aeruginosa*

919 are investigated in this study. Experimental evolution is performed on stress-sensitive strains

920 (red) to achieve adapted strains (blue). **B.** Temporal RNA-Seq data are collected from the stress-

921 insensitive (green), stress-sensitive and adapted *S. pneumoniae* strains; Tn-Seq data are collected

922 from the stress-insensitive strain. RNA-Seq and Tn-Seq data of *P. aeruginosa* are obtained from

923 published datasets (Murray *et al.*, 2015, Turner *et al.*, 2014). **C.** Data obtained from (**B.**) are

924 subjected to genome-wide analyses, network coordination analyses and machine learning to

925 generate predictive patterns of survival outcomes for the stress-sensitive, insensitive and adapted

926 strains; and adaptive outcomes for the stress-sensitive strains.

927

928 **Figure 2**. **Distinct patterns characterize the transcriptional response of nutrient-dependent**

929 **and nutrient-independent strains. A.** The magnitude of genome-wide differential expression

930 shows significantly different distributions between D39 (red) and T4 (green) in the absence of

931 Glycine (ΔGly), uracil (ΔUracil) or L-Valine (L-Val) at 30min and 90min in a Kolmogorov-

932 Smirnov test. **B.** D39 triggers significantly more TIGs in each functional tag than T4, compared

933 in a Z test for two population proportions with Bonferroni correction for multiple testing. **C.**

934 Genome-wide functional category distribution of TIGs in D39 and T4 after 90 minutes of uracil

935 depletion. **D.** Functional tag distribution of TIGs in D39 after 60 minutes of uracil depletion

936 resembles T4. Genome-wide differential expression of D39 under uracil depletion shows time-

937 dependent increase in magnitude (**E.**) and function distribution (**F.**). For **A-B**, *: $0.001 < p < 0.02$;

938 **: $0.0001 < p < 0.001$; ***: $p < 0.0001$. See in-figure legend for color-coding schemes of

939 functional tags and categories in B- D, F.

940

941 **Figure 3**. **Network coordination analyses can be visualized in a feature space and define**

942 **survival as a coordinated response**. PIGs (red), TIGs (green) and PIG/TIG overlaps (blue)

943 from the uracil depletion experiment (at 90 minutes) are overlaid on the metabolic network for

944 TIGR4 (**A.**) and D39 (**B.**), highlighting differences in network response. Connected components

945 (CC) formed by PIGs and TIGs and the shortest path distances between CC are calculated for

946 TIGR4 (**C.** and **E.**) and D39 (**D.** and **F.**). **C-F**. Inset histograms show the expected results

31

947 (permutation testing) in comparison with experimental observations (red lines). The p-value is

948 the proportion of permutations that are more extreme than the observation. **G.** Example of the

949 integration of the three-coordination metrics (CC, CN, and RE) for an experiment (blue point) by

950 plotting the -log(p-value) in a 3-dimensional feature space. The gray box represents the

951 significance threshold for each p-value. A coordinated response is typically far away from the

952 origin. **H-J.** Response trajectories for D39 (red) and T4 (green) from 30 to 90 minutes in the

953 absence of L-Valine, Glycine or uracil, respectively. In (**J.**) the D39 trajectory includes the 60-

954 minute time point. For all three graphs the dependent strain D39 remains close to the origin

955 (uncoordinated response), while the independent strain T4 moves away from the origin

956 (coordinated response). An alternative visualization of the degree of coordination of each

957 individual data point can be found in Supplementary Figure 4.

958

959 **Figure 4. Experimental evolution revert nutrient dependencies and rewires stress responses**

960 **into a coordinated response. A.** Adapted D39 strains recover growth in the absence of uracil

961 (top; aD39-uracil) or L-Valine (bottom; aD39-val). **B.** Differential expression magnitude

962 distributions are narrower in aD39-uracil and aD39-val compared to D39 and resemble T4

963 (Figure 2A). **C.** Functional tag distribution of TIGs in aD39-uracil and aD39-val at 90min after

964 uracil or L-Valine depletion are narrower compared to D39 and resemble T4 (Figure 2B).

965 Network trajectories of aD39-uracil (**D.** blue) and aD39-val (**E.** blue) show an increase in

966 coordination from 30 to 90 minutes that are similar to T4 (**D.** and **E.** green) and dissimilar to

967 wild-type D39 (**D.** and **E.** red).

968

969 **Figure 5**. **Adapted *S. pneumoniae* exhibits reduced sensitivity, changed morphology and a**

970 **rewired transcriptional response under vancomycin treatment.** Growth phenotypes (**A.**) and

971 morphology (**B.**) of wild-type (WT) and adapted T4 were examined in the absence and presence

972 of vancomycin (1xMIC) in SDMM. **B.** Cells were stained with Syto9 (green) and fm464 (red).

973 White arrowheads highlight bulging cells, typical of vancomycin sensitivity. **C.** Cell numbers per

974 chain were quantified from 1000 cell chains, indicating the adapted strain has a shorter chain-

975 length phenotype, comparable to the vancomycin-treated WT. **D.** Genome-wide differential

976 expression shows a significantly wider magnitude distribution in WT compared to adapted T4 at

977 30min post-vancomycin treatment in a Kolmogorov-Smirnov test. **E.** WT triggers significantly

32

978   more TIGs than adapted T4 in most functional tags in both early and late vancomycin response

979   in a Z-test for two population proportions with Bonferroni correction for multiple testing in (**E.**).

980   n.s.: p>0.02, *:0.001< p<0.02; **:0.0001<p<0.001, ***: p<0.0001

981

982   **Figure 6.** **Network coordination defines antibiotic resistance in S.** *pneumoniae* **and**

983   **antimicrobial and** *in vivo* **responses in** *P. aeruginosa*. **A.** Temporal network trajectories of the

984   vancomycin response for vancomycin-sensitive (wild-type T4, blue) and vancomycin-adapted

985   (aT4-vanc, red) strains profiled at 10, 20, 30, 45, 60 and 90 minutes after 1xMIC vancomycin

986   treatment. In addition, aT4-vanc is also profiled under 1.4xMIC vancomycin (green). All three

987   trajectories start at a significantly represented state, however the T4 response quickly becomes

988   uncoordinated and erratic. In contrast, aT4-vanc demonstrates a gradual trajectory that mainly

989   moves through significantly coordinated intermediate time points. N.B the speed at which a

990   trajectory is traversed is determined by the number of line segments, and not by the lengths of

991   segments, as each line is a separate time point. **B.** Network coordination analyses extended to *P.*

992   *aeruginosa* distinguishes between uncoordinated responses to antimicrobials (red), and

993   coordinated responses in *in vivo* wound infection models (blue).

994

995   **Figure 7.** **Prediction of adaptive evolution through the integration of machine learning,**

996   **genome-wide profiles, network characteristics and pan-genome sequence conservation.**

997   Pan-genome-wide sequence conservation, RNA-Seq, Tn-Seq and adaptation data are assembled

998   for the uracil (**A.**), L-Valine (**B.**) and vancomycin (**C.**) experiments and visualized by circular

999   plots: 1) Green bar plots represent expression change of parental (the innermost circles) and

1000 adapted strains (outside the orange trace); each circle represents a time-point. 2) The orange

1001 scatter plot indicates sequence conservation score, while the orange trace is a count of strains that

1002 share a gene; 3) Red arrows mark essential genes; 4) Red bar plot represents Tn-Seq fitness

1003 change; 5) Blue scatter plot indicates the mutation frequencies, with adapted genes marked by

1004 purple arrows and black lines. **D**. Receiver-operator curve (ROC) for SVM classifier. An SVM is

1005 trained to distinguish adapted genes from non-adapted genes with high accuracy. Cohen's kappa,

1006 precision, recall, and AUROC are reported.

1007

1008 **Supplementary File 1**: Supplementary information

1009

1010   **Supplementary File 2**: iSP16 consensus model.

1011

1012   **Supplementary File 3**: Tn-Seq and temporal RNA-Seq data in this study.

1013

1014   **Supplementary File 4**: Network analysis with TIGs and PIGs, and only TIGs.

Figure 1
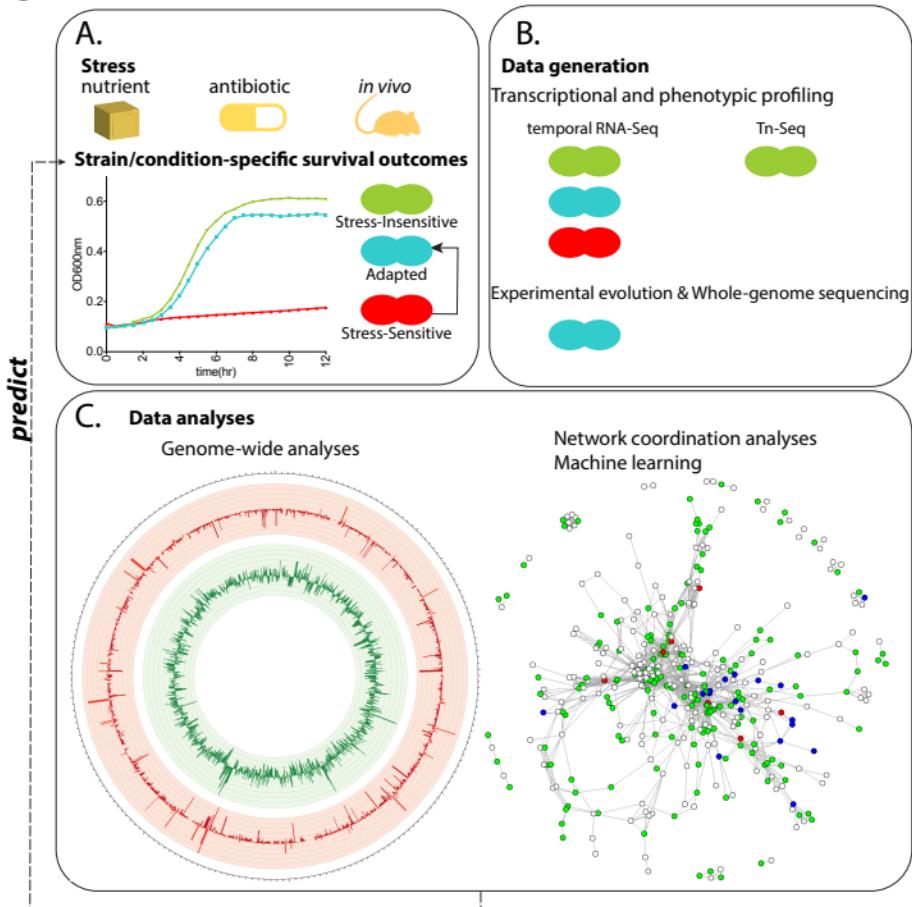
Figure 2
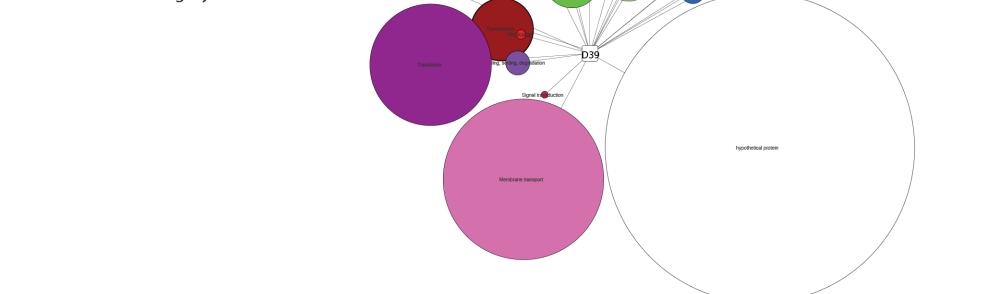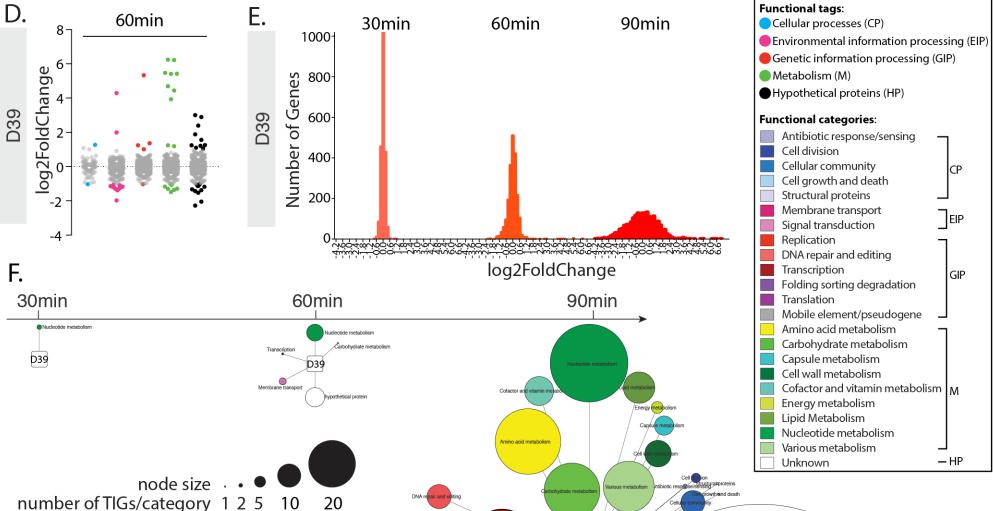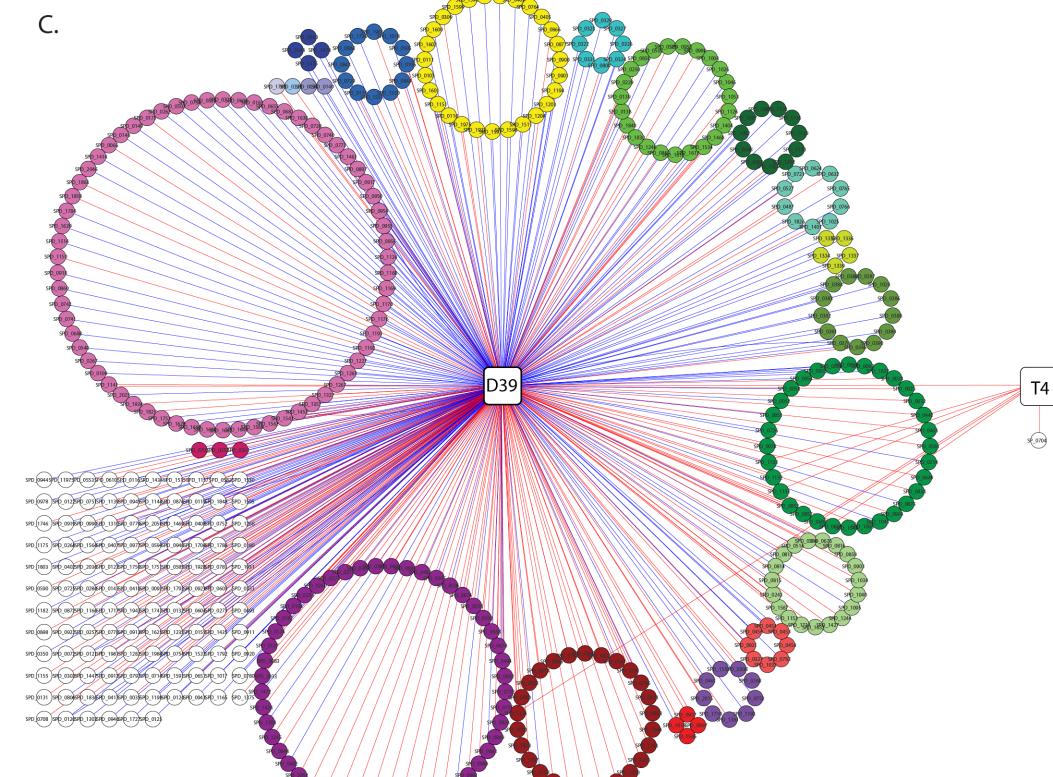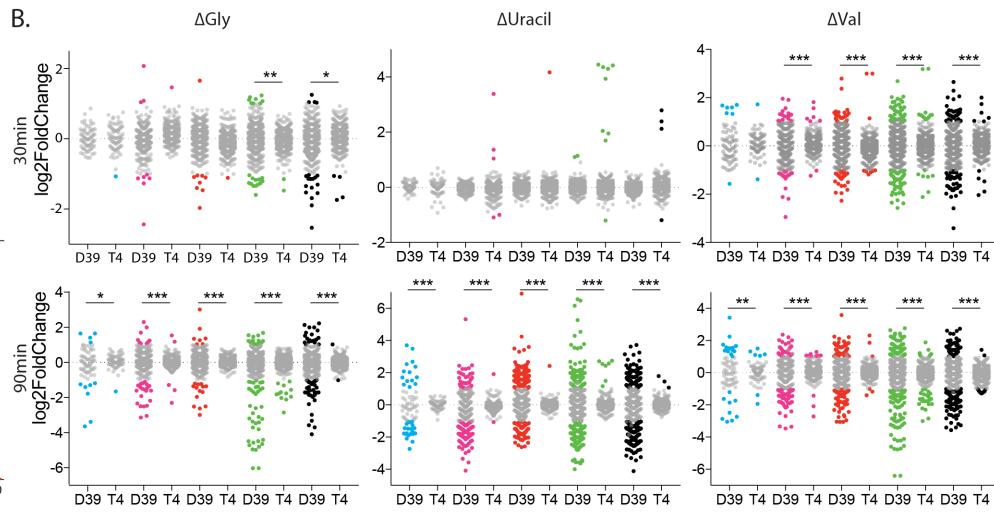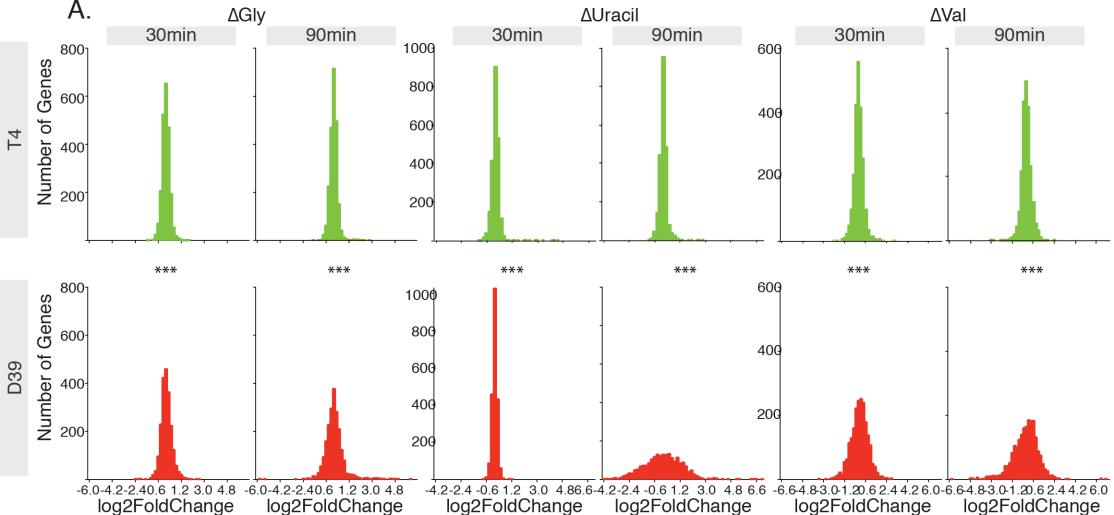
Figure 3
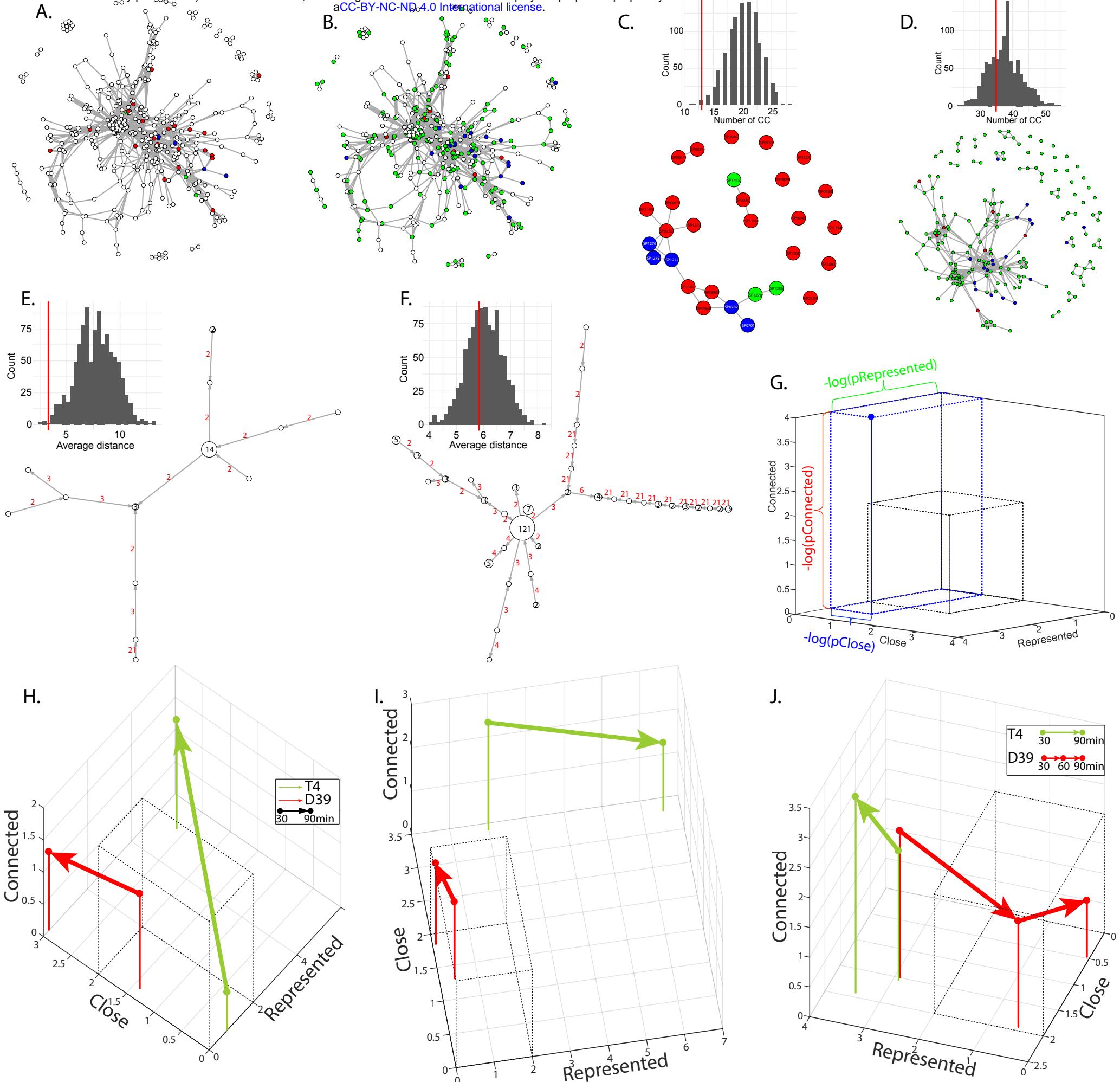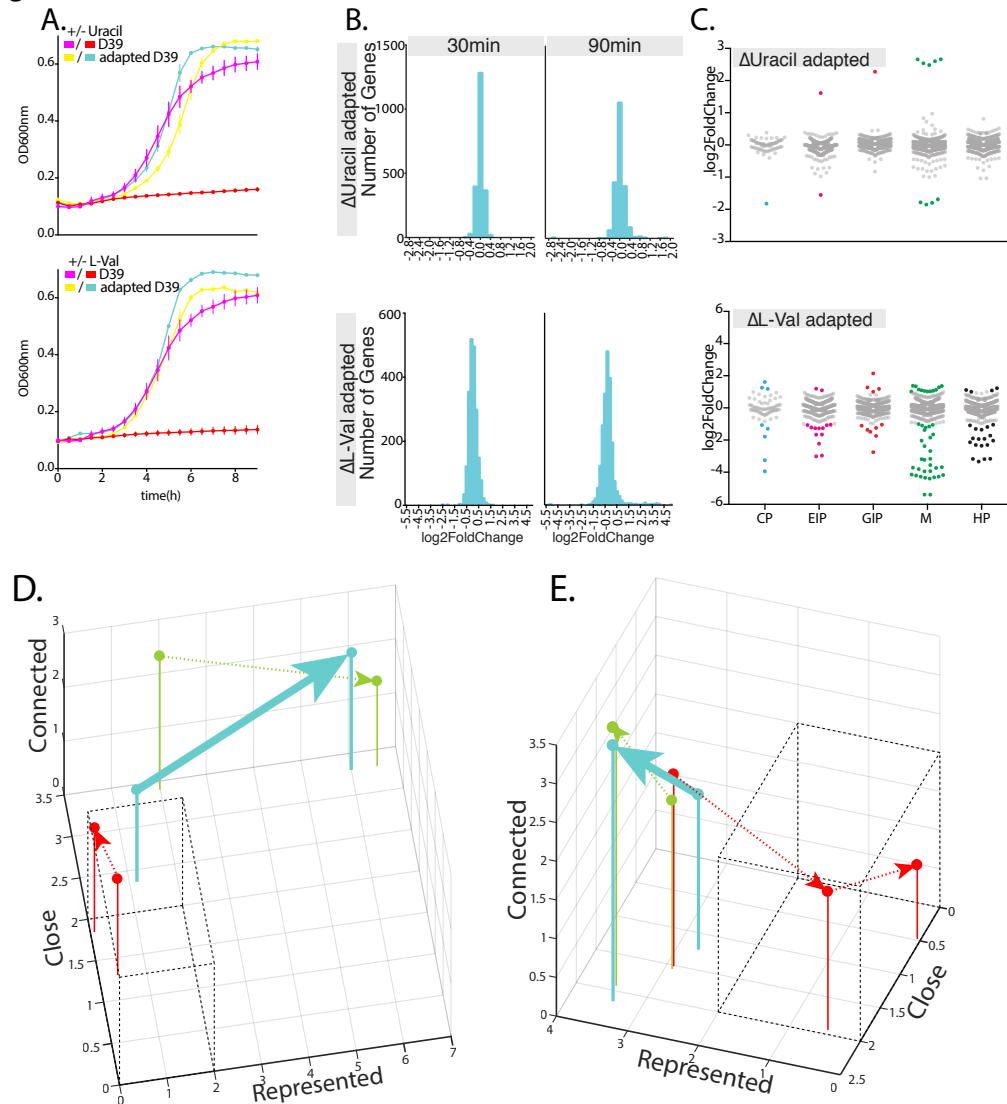
Figure 4

Figure 5

Figure 6

A.

B.

Figure 7

A.

B.

C.

D.

AUC: 0.9978
Precision: 0.6250
Recall: 0.8333
Accuracy: 0.9969
Kappa: 0.7128

True positive rate

False positive rate

AUC (random) 0.5