Towards Elucidation of the *Escherichia coli* K-12 Unknowneome

Yukako Tohsato, Natsuko Yamamoto, Toru Nakayashiki, Rikiya Takeuchi, Barry L. Wanner, and Hirotada Mori

Abstract Advances in genome sequencing have revolutionized biology by providing the molecular blueprints for thousands of living organisms. Yet, the functions of a large fraction, as much as one-half, of the component parts remain unknown even for the best understood organisms, including *Escherichia coli*, *Bacillus subtilis*, and *Saccharomyces cerevisiae*. Here, we describe our development of comprehensive genomic resources (ORFeome clone sets and mutant libraries) for systematic functional analysis of *E. coli*, summaries on our use of these resources, the GenoBase information resource for handling high-throughput experimental data obtained with them, and our creation of user workspaces at our *Pr*otein *F*unction *El*uc*idation T*eam (www.PrFEcT.org) website.

Defining the Unknowneome

Understanding the functionality of the protein components of living cells demands application of next-generation biology approaches. Not only are current annotations of genes encoding proteins incomplete and often times inaccurate, but up to 30% of the genes of newly sequenced bacteria cannot be annotated (or even recognized) using our current knowledge about proteins [1].

The development of comprehensive genomic resources for *Escherichia coli* K-12 [2, 3] have not only made systematic functional analyses feasible [4] but also have opened up new avenues for protein function elucidation, e.g., [5, 6], which otherwise cannot even be considered. The ability to carry out systematic analyses has even led biologists to re-consider the definition of biological function. *E. coli* K-12, which is one the best studied model organisms, is not exceptional in this regard. Similar genome-wide functional genomics approaches are underway in yeast ([7]; http://www.yeastgenome.org/), *Acinetobacter* species [8], *Bacillus subtilis* ([9]; http://www.genoscope.cns.fr/agc/microscope/home/index.php), *Pseudomonas*

Y. Tohsato (✉)
Department of Bioscience and Bioinformatics, Ritsumeikan University, Kusatsu, Shiga 525-8577, Japan
e-mail: yukako@sk.ritsumei.ac.jp

aeruginosa ([10]; http://www.pseudomonas.com/), and many others (e.g., http://pfgrc.jcvi.org/index.php/gateway_clones/about_libraries.html). As a proto-type of what can now be achieved through the use of genomic resources, we describe approaches that have already been proven to be useful in systematic functional analyses of *E. coli* K-12 with the Keio single-gene deletion library [2, 4, 11].

The Need for Gene Ontologies

Gene functions have traditionally been defined with natural language terms much like how humans speak. However, just like colloquialisms are specific to a region, gene terminologies often relate only to specific species, which creates difficulties when making comparisons between species. Biochemical, physiological, and phe-notypic functions of proteins can differ dramatically even for seemingly similar proteins. Clear examples exist for particular enzyme and crystalline protein families, e.g., argininosuccinate lyase and δ-crystallin, enolase and τ-crystallin, glutathione S-transferase and SIII-crystallin, and lactate dehydrogenase and ε-crystallin [12]. On the basis of DNA and amino acid sequence similarities, these enzyme and crys-talline families are closely related, however their physiological functions are quite different.

The rapid generation of new genome sequences has led to recognizing an urgent need for consistent descriptions of proteins across different organisms. The Gene Ontology (GO) Consortium (http://www.geneontology.org/) was launched to consolidate efforts for development of systematic and standardized terminologies [13]. The GO Consortium was initiated as a collaboration among three model organism databases [14]: the *Saccharomyces* Genome Database (SGD; http://www.yeastgenome.org/); a Database of Drosophila Genes and Genomes (FlyBase; http://flybase.org/); and Mouse Genome Database/Informatics (MGD/MGI; http://www.informatics.jax.org). GO uses a controlled ontology in which proteins are described in a species-independent manner in terms of a biological process (es), cellular com-ponent(s) and molecular function(s) (Table 1). The GO consortium has since been expanded to include a wide range of eucaryotic organisms, as well as many bacteria and Archaea [15] and *E. coli* K-12 among its twelve "reference genomes" [16].

Systematic Screening for Gene Functions Using the Keio Collection Single-Gene Deletion Library

Our development of the Keio collection single-gene deletion library [2], in which each of nearly 4000 of the ca. 4300 *E. coli* genes is individually deleted, has allowed testing for mutant phenotypes on a genome-wide scale. Because all mutants are in the same genetic background, genome-wide screening is expected to show effects resulting from loss of the respective gene. Questions are how to detect the effect(s) of a gene deletion and how to elucidate the function based on phenotypic changes. Two issues require mention. First, most single-gene deletions showed no observable

Table 1 Gene ontology terms[a]

Cellular component
   The cellular component ontology describes locations, at the levels of subcellular structures and macromolecular complexes. As is true for the other ontologies, not all terms are applicable to all organisms; the set of terms is meant to be inclusive. Cellular component includes such terms as ATP synthase, cell envelope, inner membrane, periplasm, and ribosome, where many gene products are found.

Biological process
   A biological process is series of events accomplished by one or more ordered assemblies of molecular functions. Examples are metabolic processes, regulation, response to stimulus, signaling, and transport and transporters.

Molecular function
The biochemical activity (including specific binding to ligands or structures) of a gene product. This definition also applies to the capability that a gene product (or gene product complex) carries as a potential. It describes only what is done without specifying where or when the event actually occurs.

[a]Gene Ontology (GO) Consortium, http://www.geneontology.org/

phenotype during growth on a rich (LB) medium. Second, many that did show an effect on rich medium grew poorly, thus making it difficult to distinguish primary and secondary mutational effects.

   Many research groups have now used the Keio collection to study different biological processes, like osmolarity, salt stress, heat stress, DNA repair, antibiotic sensitivity, etc. In general, these groups have used the Keio collection to perform genome-wide screens in two ways: (i) by screening mutants for ones that display a novel phenotype under a particular growth condition or (ii) by testing mutants for ones that display an altered cellular behavior. In these ways, systematic and comprehensive screening of the Keio collection [4] have led to finding genes whose loss affects antibiotic hypersensitivity [17, 18], swarming motility [19], biofilm formation [20], growth in human blood [21], recipient ability in conjugation [22], cysteine tolerance and production [23], colicin import and cytotoxicity [24], deethylation of 7-ethoxycoumarin [25], and glycogen metabolism [26]. In most cases, several single-gene deletion mutants were identified that affected the biological process of interest. For example, Samant et al. [21] discovered that purine and pyrimidine biosynthesis is critical for growth of *E. coli* in human serum; they uncovered 17 of the 22 Keio mutants that have deletions of pyrimidine or purine biosynthetic genes among mutants to grow in human serum.

   In many cases, mutants were recovered which had deletions of genes of unknown function, which made interpretations difficult. In other cases, mutants were found which had deletions of genes that appeared to be unrelated to the process being investigated. For instance, a genome-wide screening for mutants altered in swarming motility revealed mutants with deletions of unrelated functions, such as translation or DNA replication as well as ones of unknown functions, which caused strong repression of swarming [19]. These results show how strongly biological processes are interconnected within the cell, as well as how complicated the link can be between cause and effect. That is, in some cases, a deletion that directly affects one

process can in turn indirectly affect a second process that is coordinately regulated with the first process. Such examples are well known in transcriptional regulatory networks where a global regulator(s) can affect the expression of genes beyond those that are directly controlled by the said regulator. Complex biochemical pathways are arranged in hubs with many connectivities. Hence, mutants lacking a hub protein are likely to display many more phenotypes than mutants lacking a protein with fewer connectivities. Understanding how various cellular networks interact requires much further studies in next-generation biology.

We recently screened the Keio collection for hydroxyurea (HU)-sensitive mutants. HU is believed to interfere with DNA replication by inhibiting ribonucleotide reductase (RNR, encoded by *nrd*), which is required for conversion of NTPs to dNTPs [27]. High-throughput screening was done using robotics as illustrated in
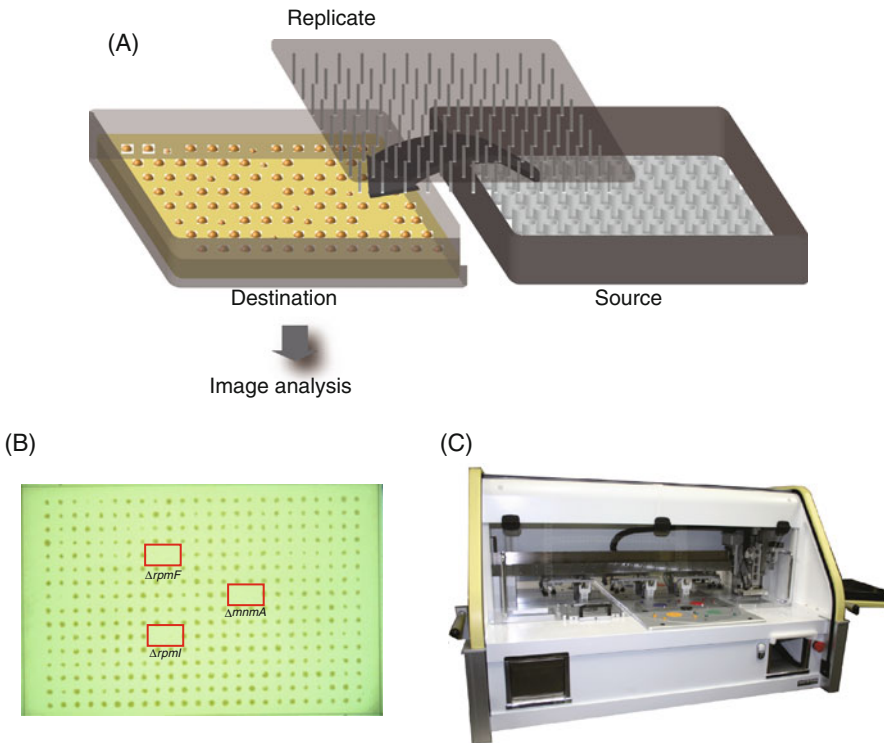

Fig. 1 High-throughput screening of the Keio collection. A schematic view of our screening protocol is shown. (**a**) The Keio collection is maintained as frozen glycerol stocks in 96-well microplates (Source). Portions are replicated in 384-spot format onto the surfaces of LB agar plates without or with 10 m*M* HU (Destination) such that duplicate replicas are juxtaposed horizontally. (**b**) Growth is measured by imaging plates at various times with a CCD camera. *Red rectangles* show candidate single-gene deletion mutants unable to grow on HU-containing agar for which both replicas grew only in the absence of HU. (**c**) Stamping is done with a Singer RoToR (Singer Instruments, UK) colony pinning robot (shown) or until recently with a Biomek FX robot (Beckman Coulter, Brea, CA)

Table 2 Hydroxyurea-sensitive mutants in the Keio collection[a]

Gene	Encoded protein	GO biological process
ahpC	alkyl hydroperoxide reductase, C22 subunit	6805 xenobiotic metabolic process
dnaT	DNA biosynthesis protein	6261 DNA-dependent DNA replication
fis	global DNA-binding transcriptional dual regulator	6310 DNA recombination
hda[b]	ATPase regulatory factor involved in DnaA inactivation	6261 DNA-dependent DNA replication
holC	DNA polymerase III, X subunit	6261 DNA-dependent DNA replication
iscS	cysteine desulfurase	9451 RNA modification
mnmA[b]	tRNA-methyltransferase	9451 RNA modification
nusB	transcription antitermination protein	transcription
priA[b]	primosome factor n′ (replication factor Y)	6261 DNA-dependent DNA replication
rplA	50S ribosomal subunit protein L1	6412 translation
rpmF	50S ribosomal subunit protein L32	6412 translation
rpmJ	50S ribosomal subunit protein L36	6412 translation
rrmJ	23S rRNA methyltransferase	6364 rRNA processing
sirA (*tusA*)	2-thiolation step of mnm^5s^2U34-tRNA synthesis	6400 tRNA modification
yheL (*tusB*)	2-thiolation step of mnm^5s^2U34-tRNA synthesis	6400 tRNA modification
yheM (*tusC*)	2-thiolation step of mnm^5s^2U34-tRNA synthesis	6400 tRNA modification
yheN (*tusD*)	2-thiolation step of mnm^5s^2U34-tRNA synthesis	6400 tRNA modification
yfaE	ferredoxin involved with ribonucleotide reductase cofactor	6124 ferredoxin metabolic process

[a]Annotations were found by using the PrFEcT WebSearch (www.prfect.org) and include results from the EcoGene (www.EcoGene.org), EcoliWiki (www.EcoliWiki.org) and PEC (http://www.shigen.nig.ac.jp/ecoli/pec/index.jsp) databases
[b]*hda*, *mnmA*, and *priA* are annotated as essential in the PEC database, yet we isolated single-gene *hda*, *mnmA*, and *priA* deletion mutants [2, 3]

Fig. 1. HU-sensitive mutants were identified as ones unable to form colonies on rich (LB) agar containing 10 m*M* HU. We uncovered 18 different HU-sensitive mutants (Table 2). Unexpectedly, ten HU-sensitive mutants (*iscS*, *mnmA*, *rplA*, *rpmF*, *rpmJ*, *rrmJ*, *sirA* (*tusA*), *yheL* (*tusB*), *yheM* (*tusC*), and *yheN* (*tusD*)) are deleted of genes connected to translation, including six (*iscS*, *mnmA*, *sirA* (*tusA*), *yheL* (*tusB*), *yheM* (*tusC*), and *yheN* (*tusD*)) in a specific tRNA modification, the thiolation step of mnm^5s^2U34-tRNA synthesis. Further analysis revealed that these mutants showed strongly reduced synthesis of Nrd (Fig. 2).

From the point of view of cellular function, tRNA modification and dNTP supply would seem to be quite distinct biological processes. Although it is well known that DNA replication and protein synthesis are coordinated in bacteria [28], the precise mechanism is unknown. Conducting global analyses with the Keio collection may shed new light on many undiscovered or poorly understood cellular networks.
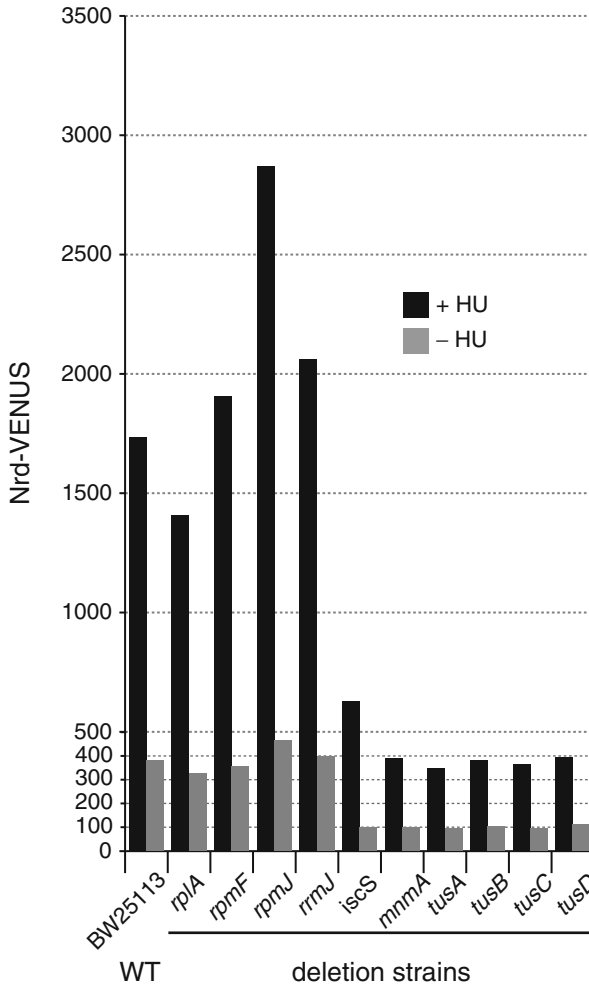
Fig. 2 Effect of HU on Nrd synthesis. Nrd synthesis was monitored by measuring fluorescence in cells carrying a low-copy plasmid with an *nrd*-Venus fusion during growth in LB without (–HU) or with 5 m*M* HU (+HU). Fluorescence was measured by flow cytometry (FACScan; BD, Franklin Lakes NJ). *E. coli* K-12 BW25113, the parent of the Keio collection [2], was used as the "wild-type (WT) control"

Screening the Keio collection for effects on many different biological processes is expected to provide new insights into the elucidation of functions of proteins belonging to uncharacterized or poorly characterized protein families – a critical challenge in the post-genomics era.

In the case of HU-sensitive screening, mutants were categorized into several groups. First, they are divided groups based on whether HU inhibited growth or killed the cells. Second, the latter were tested for whether cell death resulted from a membrane stress response or non-membrane stress response [29]. Classifying the mutants by suitable methods should yield new clues regarding function.

In addition to the Keio collection, the construction of a second single-gene deletion library, the ASKA deletion collection, is underway. Among other new features of the ASKA collection, mutants belonging to this library contain a 20-nt molecular barcode. This feature allows identification of individual mutants within a mixed population of all single-gene deletion mutants. Studying mixed populations has several benefits over studying individual cultures or stamping protocols. The selection process is closer to natural environmental conditions. Because competition occurs among the mutants, one can quantitatively assess growth advantage(s) and disadvantage(s) of all mutants simultaneously under various culture conditions. It is also beneficial for examining effects of drugs and inhibitors because much smaller amounts are required for mixed cultures, which is especially important when the drug or inhibitor of interest is expensive or hard to purify. Depending upon the purpose, one needs to choose the most appropriate procedure (simple screening, stamping, or competition). Regardless of purpose, the availability of genome-wide single-gene deletion mutants provides many advantages over traditional methods: (1) by conducting systematic and comprehensive genome-wide screens, one can quickly determine whether all (non-essential) genes for a biological process have been identified; (2) genome-wide screening can provide clues of value for identification of unknown gene functions; and (3) Genome-wide screening can help elucidate complex intracellular networks. Due to the vast amount of detailed knowledge already available for *E. coli*, developing deeper understanding of "intracellular networks" will be especially informative towards construction of a whole-cell model of unicellular organisms.

Systematic Screening of Single-Gene Deletion Mutants for Phenotypes Using Phenotype MicroArray[TM] Technology

We employed Phenotype MicroArray[TM] (PM) technology to perform systematic phenotype screening of selected single-gene deletion mutants [30, 31] (Fig. 3). PM technology was originally developed as a method for finding unique traits of individual organisms and for recognizing traits common to groups of organisms, such as species, and has been expanded as a high-throughput tool for global analysis of cellular phenotypes in post-genomic era [32]. This system monitors cellular respiration during growth in 96-well microplates under 1536 different chemical environments over a period of 24 h. Growth in each well is detected colorimetrically by quantifying the generation of purple colored formazan from tetrazolium which corresponds to the intracellular reducing state by NADH simultaneously. The effect of single-gene deletions on this screen provides information on the importance of the corresponding protein in response to diverse chemical stresses, as well as its contribution to a wide variety of different metabolic pathways. This high-throughput assay provides direct information on the contribution of the protein to the environmental fitness of the organism.

We performed PM analysis on ca. 300 single-gene mutants from the Keio collection and clustered them as reported in part previously deletion strains of *E. coli*
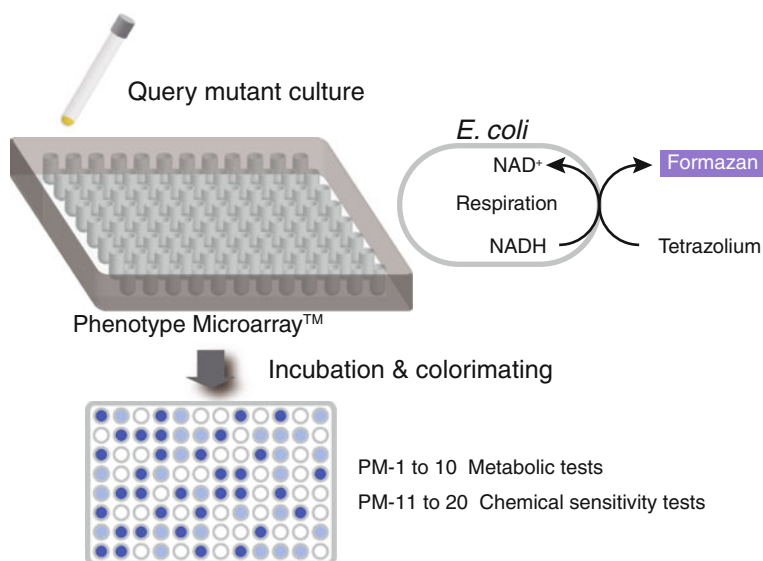
Fig. 3 Schematic view of Phenotype MicroArray^{TM} analysis. Respiration is measured by quantifying the generation of NAD$^+$ by formation of purple-colored formazan from tetrazolium. A culture of the query mutant is dispensed into BIOLOG Phenotype MicroArray^{TM} plate. Color development is automatically quantified during incubation in an *OmniLog®* instrument. PM-1 to PM-10 include 1 blank culture well and PM-11-20 have different chemical concentrations, resulting in a total of 1536 different chemical environments (1920 conditions). More precise information is available at http://www.biolog.com

and the clustering analysis using the part of the results was reported previously [30, 31]. Based on the entire PM results, we show here statistically the effectiveness of systematic phenotype screening using PM technology. The precise method of statistical measurement of PM will be reported elsewhere [31]. As shown in Fig. 4, 709 (36.9%) of 1920 conditions showed no respiration in our control strain and 12 (0.6%) of the conditions are negative (water) and positive (LB medium) controls. Only 8 of the remaining 1199 conditions had no significant phenotype change in any of the 300 mutants tested.

Figure 5 shows the medium condition effects and environmental dependencies of the ca. 300 mutants tested. On average, 45 mutants showed significant phenotypic changes under each condition (Fig. 5a). Further, each mutant showed differences under 183 conditions (Fig. 5b). It is worth mentioning that our PM tests were done in duplicate and showed a high degree of reproducibility (94.5%).

Although PM technology is a powerful tool for functional screening, it alone provides insufficient sensitivity to identify functions for proteins of unknown function. The most likely causes are robustness and the existence of unknown alternative metabolic pathways. *E. coli* K-12 has 98 genes encoding isozymes, including 80 encoding pairs of isozymes based on annotations in EcoCyc version 13.5 [22]. In some cases, expression patterns of these isozymes differ; integrating results from
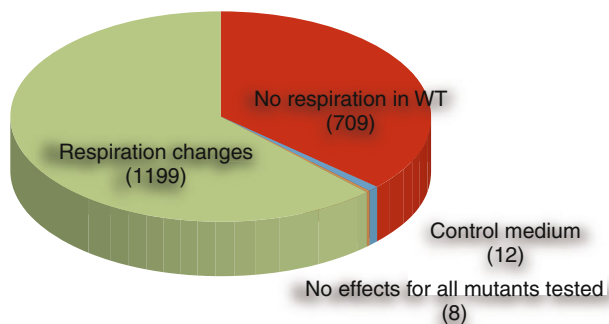
Fig. 4 Classification of 1920 medium conditions by color change. The number of conditions showing no respiration in *E. coli* K-12 BW25113 (WT), respiration changes, and no effects for all mutants tested are given in *parentheses*. Conditions included negative (water) and positive (LB) control medium
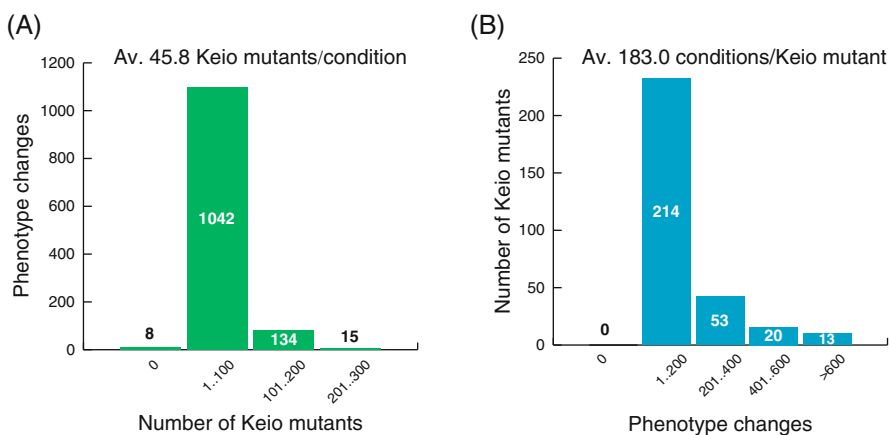

Fig. 5 Medium condition effects and environmental dependencies. (**a**) *Abscissa* and ordinate axes show number of single-gene deletions resulting in significant phenotype changes using 100 gene deletions as the bin size. (**b**) *Abscissa* and ordinate axes show the number of phenotype changes observed in the single-gene deletion mutants using 200 medium conditions as the bin size

transcriptome analysis from DNA microarrays, or preferably RNA-seq, with PM analysis may provide deeper insight into physiological function.

Systematic Screening for Genetic Interactions

Synthetic lethal screens are an effective experimental approach for revealing mechanisms of cellular robustness [33]. We have developed a strategy to screen comprehensively for effects of double gene deletions in *E. coli* [34, 35] (Fig. 6). Preliminary results using such strategies have been reported [5, 6]. Even though
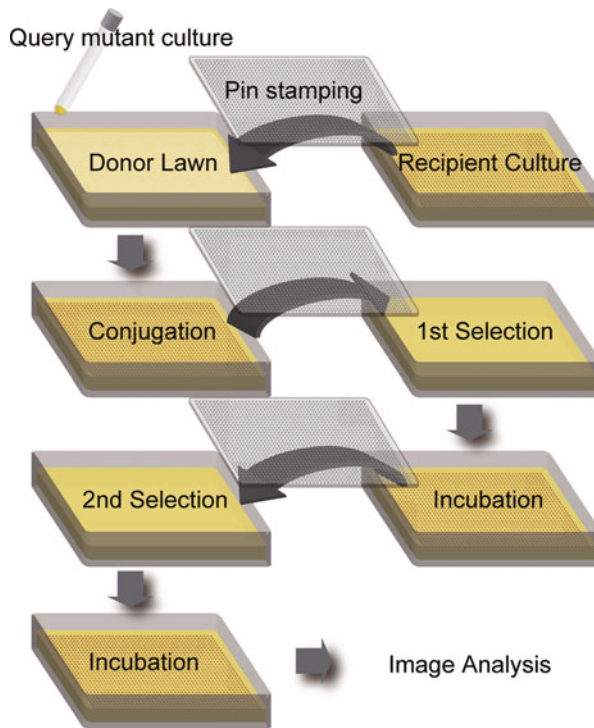
Fig. 6 Schematic view of genetic interaction analysis by double gene deletion. A schematic view of the protocol for creating and examining double-gene deletion mutants in a high-throughput manner by conjugation is shown. The query gene mutant, which serves as an Hfr donor, is evenly spread on LB agar to form a donor lawn. Single-gene deletion library, which serves as recipient culture and stored as frozen glycerol stocks or colonies on agar in high-density format, is replicated onto the donor lawn by robotic pin stamping. Following growth to allow conjugation to occur, pin stamping is used to replicate from the conjugation surface onto the 1st selection. These plates are incubated for 6 h and then replicated by pin stamping on the 2nd selection plate, which is necessary to eliminate background growth. The 2nd selection plate is imaged with a CCD camera over time, and image analysis is done to identify double mutants growing poorer or better than control matings. Details will be reported elsewhere

the genome sequencing and large-scale genetic analyses have revealed the enormous amount of genetic information of the target organisms, our knowledge of cellular system is still very limited. A major challenge is to understand physiological networks of genes in a living cell. As described above, single-gene deletion mutants generally show limited phenotype changes because of the redundancy or compensatory pathways. This phenomenon is called robustness; there can be many mechanisms that can lead to re-construction of physiological steps or gene product networks. The structure of the cellular network may not be rigid but rather be dynamically changing according to the environment, which can result not only from the extracellular environment but also by genetic alterations (mutation or deletion).
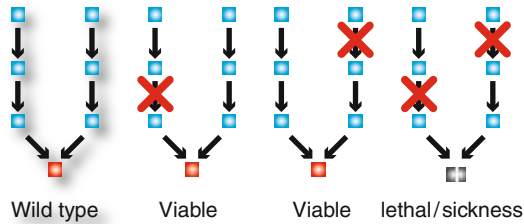
Fig. 7 Concept of synthetic lethality analysis by a double-gene deletion strategy. Normally organisms may have multiple pathways to generate essential substrates. In such cases, elimination of one pathway step by mutation is without effect because the other pathway can provide the missing step. The cell is unable to survive only when both pathways are disrupted simultaneously. The consequence of the corresponding double mutation leads to synthetic lethality (or sickness), which reveals a genetic interaction

Robustness in a cellular network is similar to the operation of a transportation network. Although a shortest path exists, an alternative longer detour pathway(s) is often available if the shortest one is blocked. The concept of the synthetic lethality analysis by a double-gene deletion strategy is shown in Fig. 7. When two genes are found to interact such that loss of one is without (major) effect but loss of both results in a new mutant phenotype, the genetic interaction is called epistasis. Epistasis effects can cause many kinds of effects; only a subset cause cell lethality or sickness. However, those causing severe growth effects are easiest to score. Large-scale genetic interaction studies have provided the basis for defining gene function and gene networks. Recent results from comprehensive genetic interaction analyses have greatly accelerated deeper insight into physiological gene functions and networks from bacteria to humans [33].
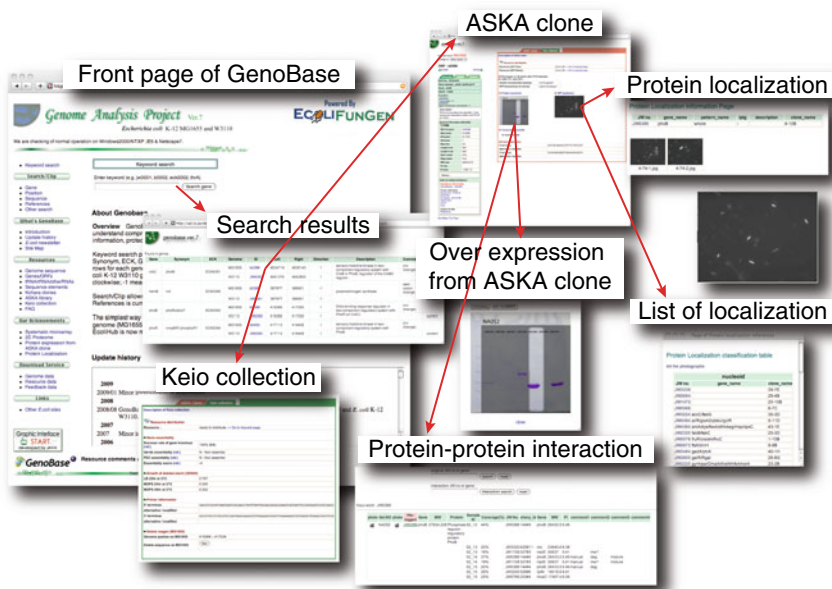
Information Resources

GenoBase was originally developed for the *E. coli* genome project, which was launched in Japan in 1989, to select phage clones from the ordered Kohara library [36] for sequencing [37]. Because *E. coli* is one of the best studied organisms, many gene sequences had already been accumulated at the time. To facilitate selection of the target phage clone, we constructed the GenoBase database to help distinguish sequenced and non-sequenced regions of the chromosome. First, we collected all of the *E. coli* sequence information from publicly available databases, such as GenBank, EMBL and DDBJ, and made consensus sequence by assembling these followed by mapping them onto the chromosome to identify sequenced regions. Upon completion of genome sequencing [38–42], GenoBase was further developed for supporting systematic functional genomics and systems analyses for *E. coli*. The development of biological resources for systematic studies of *E. coli* K-12, like the single-gene deletion Keio collection [2] and the ASKA ORFeome clone library [43], has proven especially valuable worldwide. The advent of technologies

for acquisition of high-throughput data types (series of comprehensive network analyses, e.g., transcriptome, proteome, interactome and genetic interaction) has created need to preserve and share information.

GenoBase originally displayed information only for the W3110 strain of *E. coli* K-12, which was the target in the Japanese *E. coli* genome project [27, 28], whereas the MG1655 strain was the target in the Wisconsin genome sequencing project in the USA [29]. GenoBase version 7, which was developed in collaboration with Purdue University (www.PrFEcT.org/GenoBase), has been enhanced to permit the user to choose displaying information for *E. coli* K-12 MG1655 or W3110. GenoBase ver. 7 has also been enhanced to support image data and other high-throughput data.

GenoBase (Fig. 8) is especially rich in experimental resources (mutants and plasmids) and experimental data from a large *E. coli* functional genomics project in Japan, which far exceeds all other resources combined. Information in GenoBase is public or private (password accessible), depending upon whether the data have been published. Current resources include (1) two types of ASKA ORFeome libraries [30], including one with a C-terminal GFP tag and one without; and (2) the single-gene deletion library known as the Keio collection [2]. Comprehensive experimental

http://ecoli.naist.jp
http://www.PrFEcT.org/GenoBase

Fig. 8 The GenoBase Information Resource. GenoBase version 6 is fully operational at http://ecoli.naist.jp, while GenoBase version 7 is at www.PrFEcT.org/Genobase. Version 8 is now under construction. Once development is completed, mirroring will be deployed to maintain synchrony between these sites. Querying from home page gives search results in a table with links to pages for resources and experimental results based on the use of these resources

resources and data generated by systematic analysis using those resources are continuously growing. To facilitate both systems and individual research approaches using *E. coli* K-12 as a model system, integrative databases provide essential information.

GenoBase activities have not only involved the collection of high-throughput experimental data but also improvement in the quality of K-12 genome annotation. One example was the re-confirmation of the *E. coli* genome sequence, which resulted in correction of sequencing errors of previously published W3110 and MG1655 sequences [28]. Correction of the K-12 genome sequence provided major stimulus for cooperative re-annotation of the K-12 genome at international annotation workshops held in Woods Hole in 2003 and 2005 [31].

GenoBase is a searchable database devoted to systems biology of *E. coli* K-12. Querying GenoBase is done from the home page (Fig. 8). Any term, such as an id, gene name, product name, is accepted. Searching results are displayed in a tabular format with links to gene pages, which show additional information about the target gene. Contents show genome annotation information together with the biological resources and systematic analysis data using those resources. GenoBase is based on the predicted genes and all data are stored associated with the genes.

High-throughput systematic experimental data currently includes three large data sets:

- Protein–protein interaction data are based on using His-tagged ASKA ORF clone library without GFP [32]. All of the interaction data including data produced from TOF-MAS analysis is stored and specific partner candidates as prey proteins are available from each target protein as bait.
- DNA microarray data from analysis of single-gene deletion mutants were generated using full length cDNA type arrays, which were made with PCR-amplified fragments from ASKA clone library. Quantitative data were generated by ImaGene for about 150 deletion mutants, mostly for ones lacking transcription factors.
- Protein localization data are displayed for transformants carrying the GFP-tagged ASKA ORFeome clones. Transformants were analyzed by confocal microscopy of transformants expressing each protein at a low basal level in absence of an inducer to avoid misfolding of the target protein that can accompany protein overproduction. Images captured with a CCD camera are stored in our database.

Future Perspectives

(A) Quality control of resources
Our group continues to improve the quality of the biological resources created in Japan. For example, one quality control issue for the Keio collection has been the occasional discovery of partial duplications. Accordingly, the entire collection has now been validated. Upon publication, these will be open to the public through GenoBase.

(B) New experimental resources

Continuous efforts are underway to improve current resources and to construct new resources to expand systematic studying of *E. coli*. A new single gene deletion library with a different antibiotic resistance marker has been created for construction of double mutants to test for genetic interactions. The same library is bar coded to permit population studies. A second new resource near completion is a Gateway-fitted ASKA clone entry. All precise information on these new resources will also be stored in GenoBase.

(C) Systematic approaches using resources

As described above, systematic analyses, such as protein–protein interaction and protein localization, were performed and the data stored in GenoBase. Recently, we reported high-throughput systems for studying genetic interactions [5, 6] and the results from these analyses will be stored in GenoBase or partner databases.

References

1. Roberts, R.J. Identifying protein function-a call for community action. PLoS. Biol. **2**: E42 (2004).
2. Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L., Mori, H. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol. Syst. Biol. **2**: 2006 (2006).
3. Yamamoto, N., Nakahigashi, K., Nakamichi, T., Yoshino, M., Takai, Y., Touda, Y., Furubayashi, A., Kinjyo, S., Dose, H., Hasegawa, M., Datsenko, K.A., Nakayashiki, T., Tomita, M., Wanner, B.L., Mori, H. Update on the Keio collection of *Escherichia coli* single-gene deletion mutants. Mol. Syst. Biol. **5**: 335 (2009).
4. Baba, T., Huan, H.C., Datsenko, K., Wanner, B.L., Mori, H. The applications of systematic in-frame, single-gene knockout mutant collection of *Escherichia coli* K-12. Methods Mol. Biol. **416**: 183–194 (2008).
5. Butland, G., Babu, M., Diaz-Mejia, J.J., Bohdana, F., Phanse, S., Gold, B., Yang, W., Li, J., Gagarinova, A.G., Pogoutse, O., Mori, H., Wanner, B.L., Lo, H., Wasniewski, J., Christopolous, C., Ali, M., Venn, P., Safavi-Naini, A., Sourour, N., Caron, S., Choi, J.Y., Laigle, L., Nazarians-Armavil, A., Deshpande, A., Joe, S., Datsenko, K.A., Yamamoto, N., Andrews, B.J., Boone, C., Ding, H., Sheikh, B., Moreno-Hagelseib, G., Greenblatt, J.F., Emili, A. eSGA: *E. coli* synthetic genetic array analysis. Nat. Methods **5**: 789–795 (2008).
6. Typas, A., Nichols, R.J., Siegele, D.A., Shales, M., Collins, S.R., Lim, B., Braberg, H., Yamamoto, N., Takeuchi, R., Wanner, B.L., Mori, H., Weissman, J.S., Krogan, N.J., Gross, C.A. High-throughput, quantitative analyses of genetic interactions in *E. coli*. Nat. Methods **5**: 781–787 (2008).
7. Giaever, G., Chu, A.M., Ni, L., Connelly, C., Riles, L., Veronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., Andre, B., Arkin, A.P., Astromoff, A., El-Bakkoury, M., Bangham, R., Benito, R., Brachat, S., Campanaro, S., Curtiss, M., Davis, K., Deutschbauer, A., Entian, K.D., Flaherty, P., Foury, F., Garfinkel, D.J., Gerstein, M., Gotte, D., Guldener, U., Hegemann, J.H., Hempel, S., Herman, Z., Jaramillo, D.F., Kelly, D.E., Kelly, S.L., Kotter, P., LaBonte, D., Lamb, D.C., Lan, N., Liang, H., Liao, H., Liu, L., Luo, C., Lussier, M., Mao, R., Menard,

P., Ooi, S.L., Revuelta, J.L., Roberts, C.J., Rose, M., Ross-Macdonald, P., Scherens, B., Schimmack, G., Shafer, B., Shoemaker, D.D., Sookhai-Mahadeo, S., Storms, R.K., Strathern, J.N., Valle, G., Voet, M., Volckaert, G., Wang, C.Y., Ward, T.R., Wilhelmy, J., Winzeler, E.A., Yang, Y., Yen, G., Youngman, E., Yu, K., Bussey, H., Boeke, J.D., Snyder, M., Philippsen, P., Davis, R.W., Johnston, M. Functional profiling of the Saccharomyces cerevisiae genome. Nature **418**: 387–391 (2002).

8. de Berardinis, V., Vallenet, D., Castelli, V., Besnard, M., Pinet, A., Cruaud, C., Samair, S., Lechaplais, C., Gyapay, G., Richez, C., Durot, M., Kreimeyer, A., Le, F.F., Schachter, V., Pezo, V., Doring, V., Scarpelli, C., Medigue, C., Cohen, G.N., Marliere, P., Salanoubat, M., Weissenbach, J. A complete collection of single-gene deletion mutants of *Acinetobacter baylyi* ADP1. Mol. Syst. Biol. **4**: 174 (2008).

9. Kobayashi, K., Ehrlich, S.D., Albertini, A., Amati, G., Andersen, K.K., Arnaud, M., Asai, K., Ashikaga, S., Aymerich, S., Bessieres, P., Boland, F., Brignell, S.C., Bron, S., Bunai, K., Chapuis, J., Christiansen, L.C., Danchin, A., Debarbouille, M., Dervyn, E., Deuerling, E., Devine, K., Devine, S.K., Dreesen, O., Errington, J., Fillinger, S., Foster, S.J., Fujita, Y., Galizzi, A., Gardan, R., Eschevins, C., Fukushima, T., Haga, K., Harwood, C.R., Hecker, M., Hosoya, D., Hullo, M.F., Kakeshita, H., Karamata, D., Kasahara, Y., Kawamura, F., Koga, K., Koski, P., Kuwana, R., Imamura, D., Ishimaru, M., Ishikawa, S., Ishio, I., Le, C.D., Masson, A., Mauel, C., Meima, R., Mellado, R.P., Moir, A., Moriya, S., Nagakawa, E., Nanamiya, H., Nakai, S., Nygaard, P., Ogura, M., Ohanan, T., O'Reilly, M., O'Rourke M, Pragai, Z., Pooley, H.M., Rapoport, G., Rawlins, J.P., Rivas, L.A., Rivolta, C., Sadaie, A., Sadaie, Y., Sarvas, M., Sato, T., Saxild, H.H., Scanlan, E., Schumann, W., Seegers, J.F., Sekiguchi, J., Sekowska, A., Seror, S.J., Simon, M., Stragier, P., Studer, R., Takamatsu, H., Tanaka, T., Takeuchi, M., Thomaides, H.B., Vagner, V., van Dijl, J.M., Watabe, K., Wipat, A., Yamamoto, H., Yamamoto, M., Yamamoto, Y., Yamane, K., Yata, K., Yoshida, K., Yoshikawa, H., Zuber, U., Ogasawara, N. Essential *Bacillus subtilis* genes. Proc. Natl. Acad. Sci. USA **100**: 4678–4683 (2003).

10. Jacobs, M.A., Alwood, A., Thaipisuttikul, I., Spencer, D., Haugen, E., Ernst, S., Will, O., Kaul, R., Raymond, C., Levy, R., Chun-Rong, L., Guenthner, D., Bovee, D., Olson, M.V., Manoil, C. Comprehensive transposon mutant library of *Pseudomonas aeruginosa.* Proc. Natl. Acad. Sci. USA **100**: 14339–14344 (2003).

11. Baba, T., Mori, H. The construction of systematic in-frame, single-gene knockout mutant collection in *Escherichia coli* K-12. Methods Mol. Biol. **416**: 171–181 (2008).

12. Wistow, G., Piatigorsky, J. Recruitment of enzymes as lens structural proteins. Science **236**: 1554–1556 (1987).

13. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., Sherlock, G. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. **25**: 25–29 (2000).

14. The Gene Ontology Consortium: Creating the gene ontology resource: design and implementation. Genome Res. **11**: 1425–1433 (2001).

15. Carbon, S., Ireland, A., Mungall, C.J., Shu, S., Marshall, B., Lewis, S. AmiGO: online access to ontology and annotation data. Bioinformatics **25**: 288–289 (2009).

16. Reference Genome Group of the Gene Ontology Consortium. The Gene Ontology's Reference Genome Project: a unified framework for functional annotation across species. PLoS Comput. Biol. **5**: e1000431 (2009).

17. Tamae, C., Liu, A., Kim, K., Sitz, D., Hong, J., Becket, E., Bui, A., Solaimani, P., Tran, K.P., Yang, H., Miller, J.H. Determination of antibiotic hypersensitivity among 4,000 single-gene-knockout mutants of *Escherichia coli.* J. Bacteriol. **190**: 5981–5988 (2008).

18. Liu, A., Tran, L., Becket, E., Lee, K., Chinn, L., Park, E., Tran, K., Miller, J.H. Antibiotic sensitivity profiles determined with an *Escherichia coli* gene knockout collection: generating an antibiotic bar code. Antimicrob. Agents Chemother. **54**: 1393–1403 (2010).

19. Inoue, T., Shingaki, R., Hirose, S., Waki, K., Mori, H., Fukui, K. Genome-wide screening of genes required for swarming motility in *Escherichia coli* K-12. J. Bacteriol. **189**: 950–957 (2007).

20. Niba, E.T., Naka, Y., Nagase, M., Mori, H., Kitakawa, M. A genome-wide approach to identify the genes involved in biofilm formation in *E. coli*. DNA Res. **14**: 237–246 (2007).

21. Samant, S., Lee, H., Ghassemi, M., Chen, J., Cook, J.L., Mankin, A.S., Neyfakh, A.A. Nucleotide biosynthesis is critical for growth of bacteria in human blood. PLoS Pathog. **4**: e37 (2008).

22. Perez-Mendoza, D., de la, C.F. *Escherichia coli* genes affecting recipient ability in plasmid conjugation: are there any? BMC Genomics **10**: 71 (2009).

23. Wiriyathanawudhiwong, N., Ohtsu, I., Li, Z.D., Mori, H., Takagi, H. The outer membrane TolC is involved in cysteine tolerance and overproduction in *Escherichia coli*. Appl. Microbiol. Biotechnol. **81**: 903–913 (2009).

24. Sharma, O., Datsenko, K.A., Ess, S.C., Zhalnina, M.V., Wanner, B.L., Cramer, W.A. Genome-wide screens: novel mechanisms in colicin import and cytotoxicity. Mol. Microbiol. **73**: 571–585 (2009).

25. Zhou, Y., Minami, T., Honda, K., Omasa, T., Ohtake, H. Systematic screening of *Escherichia coli* single-gene knockout mutants for improving recombinant whole-cell biocatalysts. Appl. Microbiol. Biotechnol. **87**: 647–655 (2010).

26. Montero, M., Eydallin, G., Viale, A.M., Almagro, G., Munoz, F.J., Rahimpour, M., Sesma, M.T., Baroja-Fernandez, E., Pozueta-Romero, J. *Escherichia coli* glycogen metabolism is controlled by the PhoP-PhoQ regulatory system at submillimolar environmental Mg2+ concentrations, and is highly interconnected with a wide variety of cellular processes. Biochem. J. **424**: 129–141 (2009).

27. Krakoff, I.H., Brown, N.C., Reichard, P. Inhibition of ribonucleoside diphosphate reductase by hydroxyurea. Cancer Res. **28**: 1559–1565 (1968).

28. Maaloe, O. The control of normal DNA replication in bacteria. Cold Spring Harb. Symp. Quant. Biol. **26**: 45–52 (1961).

29. Davies, B.W., Kohanski, M.A., Simmons, L.A., Winkler, J.A., Collins, J.J., Walker, G.C. Hydroxyurea induces hydroxyl radical-mediated cell death in *Escherichia coli*. Mol. Cell **36**: 845–860 (2009).

30. Tohsato, Y., Mori, H. Phenotype profiling of single gene deletion mutants of *E. coli* using Biolog technology. Genome Inform. **21**: 42–52 (2008).

31. Tohsato, Y., Baba, T., Mazaki, Y., Ito, M., Wanner, B.L., Mori, H. Environmental dependency of gene knockouts on phenotype microarray analysis in *Escherichia coli*. J. Bioinform. Comput. Biol. **8**(Suppl 1): 83–99.

32. Bochner, B.R. Global phenotypic characterization of bacteria. FEMS Microbiol. Rev. **33**: 191–205 (2009).

33. Dixon, S.J., Costanzo, M., Baryshnikova, A., Andrews, B., Boone, C. Systematic mapping of genetic interaction networks. Annu. Rev. Genet. **43**: 601–625 (2009).

34. Typas, A., Nichols, R.J., Siegele, D.A., Shales, M., Collins, S.R., Lim, B., Braberg, H., Yamamoto, N., Takeuchi, R., Wanner, B.L., Mori, H., Weissman, J.S., Krogan, N.J., Gross, C.A. High-throughput, quantitative analyses of genetic interactions in E. coli. Nat. Methods **5**(9): 781–787 (2008).

35. Butland, G., Babu, M., Díaz-Mejía, J.J., Bohdana, F., Phanse, S., Gold, B., Yang, W., Li, J., Gagarinova, A.G., Pogoutse, O., Mori, H., Wanner, B.L., Lo, H., Wasniewski, J., Christopolous, C., Ali, M., Venn, P., Safavi-Naini, A., Sourour, N., Caron, S., Choi, J.Y., Laigle, L., Nazarians-Armavil, A., Deshpande, A., Joe, S., Datsenko, K.A., Yamamoto, N., Andrews, B.J., Boone, C., Ding, H., Sheikh, B., Moreno-Hagelseib, G., Greenblatt, J.F., Emili, A. eSGA: E. coli synthetic genetic array analysis. Nat. Methods **5**(9): 789–795 (2008).

36. Kohara, Y., Akiyama, K., Isono, K. The physical map of the whole *E. coli* chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library. Cell **50**: 495–508 (1987).

37. Yura, T., Mori, H., Nagai, H., Nagata, T., Ishihama, A., Fujita, N., Isono, K., Mizobuchi, K., Nakata, A. Systematic sequencing of the *Escherichia coli* genome: analysis of the 0–2.4 min region. Nucleic Acids Res. **20**: 3305–3308 (1992).
38. Blattner, F.R., Plunkett, G., III, Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B., Shao, Y. The complete genome sequence of *Escherichia coli* K-12. Science **277**: 1453–1474 (1997).
39. Aiba, H., Baba, T., Hayashi, K., Inada, T., Isono, K., Itoh, T., Kasai, H., Kashimoto, K., Kimura, S., Kitakawa, M., Kitagawa, M., Makino, K., Miki, T., Mizobuchi, K., Mori, H., Mori, T., Motomura, K., Nakade, S., Nakamura, Y., Nashimoto, H., Nishio, Y., Oshima, T., Saito, N., Sampei, G., Horiuchi, T. A 570-kb DNA sequence of the *Escherichia coli* K-12 genome corresponding to the 28.0–40.1 min region on the linkage map. DNA Res. **3**: 363–377 (1996).
40. Itoh, T., Aiba, H., Baba, T., Hayashi, K., Inada, T., Isono, K., Kasai, H., Kimura, S., Kitakawa, M., Kitagawa, M., Makino, K., Miki, T., Mizobuchi, K., Mori, H., Mori, T., Motomura, K., Nakade, S., Nakamura, Y., Nashimoto, H., Nishio, Y., Oshima, T., Saito, N., Sampei, G., Seki, Y., Horiuchi, T. A 460-kb DNA sequence of the *Escherichia coli* K-12 genome corresponding to the 40.1–50.0 min region on the linkage map. DNA Res. **3**: 379–392 (1996).
41. Oshima, T., Aiba, H., Baba, T., Fujita, K., Hayashi, K., Honjo, A., Ikemoto, K., Inada, T., Itoh, T., Kajihara, M., Kanai, K., Kashimoto, K., Kimura, S., Kitagawa, M., Makino, K., Masuda, S., Miki, T., Mizobuchi, K., Mori, H., Motomura, K., Nakamura, Y., Nashimoto, H., Nishio, Y., Saito, N., Horiuchi, T. A 718-kb DNA sequence of the *Escherichia coli* K-12 genome corresponding to the 12.7–28.0 min region on the linkage map. DNA Res. **3**: 137–155 (1996).
42. Hayashi, K., Morooka, N., Yamamoto, Y., Fujita, K., Isono, K., Choi, S., Ohtsubo, E., Baba, T., Wanner, B.L., Mori, H., Horiuchi, T. Highly accurate genome sequences of *Escherichia coli* K-12 strains MG1655 and W3110. Mol. Syst. Biol. **2**: 2006 (2006).
43. Kitagawa, M., Ara, T., Arifuzzaman, M., Ioka-Nakamichi, T., Inamoto, E., Toyonaga, H., Mori, H. Complete set of ORF clones of *Escherichia coli* ASKA library (a complete set of *E. coli* K-12 ORF archive): unique resources for biological research. DNA Res. **12**: 291–299 (2005).