eSGA: *E. coli* synthetic genetic array analysis

Gareth Butland[1,2,7,8], Mohan Babu[1,8], J Javier Díaz-Mejía[1,3], Fedyshyn Bohdana[1], Sadhna Phanse[1], Barbara Gold[2], Wenhong Yang[2], Joyce Li[1], Alla G Gagarinova[4], Oxana Pogoutse[1], Hirotada Mori[5], Barry L Wanner[5], Henry Lo[1], Jas Wasniewski[1], Constantine Christopoulos[1], Mehrab Ali[4], Pascal Venn[1], Anahita Safavi-Naini[1], Natalie Sourour[1], Simone Caron[1], Ja-Yeon Choi[1], Ludovic Laigle[1], Anaies Nazarians-Armavil[1], Avnish Deshpande[1], Sarah Joe[1], Kirill A Datsenko[6], Natsuko Yamamoto[5], Brenda J Andrews[1,4], Charles Boone[1,5], Huiming Ding[1], Bilal Sheikh[1], Gabriel Moreno-Hagelsieb[3], Jack F Greenblatt[1,4] & Andrew Emili[1,4]

Physical and functional interactions define the molecular organization of the cell. Genetic interactions, or epistasis, tend to occur between gene products involved in parallel pathways or interlinked biological processes. High-throughput experimental systems to examine genetic interactions on a genome-wide scale have been devised for *Saccharomyces cerevisiae, Schizosaccharomyces pombe, Caenorhabditis elegans* and *Drosophila melanogaster,* but have not been reported previously for prokaryotes. Here we describe the development of a quantitative screening procedure for monitoring bacterial genetic interactions based on conjugation of *Escherichia coli* deletion or hypomorphic strains to create double mutants on a genome-wide scale. The patterns of synthetic sickness and synthetic lethality (aggravating genetic interactions) we observed for certain double mutant combinations provided information about functional relationships and redundancy between pathways and enabled us to group bacterial gene products into functional modules.

Measuring the phenotypes of double mutants on a genome-wide scale can help to explain the biological functions of individual gene products[1], reveal the components of biological pathways and protein complexes[2], and illuminate the modular arrangement of functional neighborhoods[3]. Several high-throughput screening methods have been developed to detect genetic interactions in a genome-wide manner in eukaryotic model systems. Most of these technologies are array-based, which allows thousands of interactions to be measured simultaneously.

The archetypal approach, termed synthetic genetic array (SGA) analysis[4], makes use of robotic automation to construct double mutants in the yeast *Saccharomyces cerevisiae* by mating. A 'query' strain, with a drug-resistance marker replacing a gene of interest, is crossed to an arrayed collection of single-gene deletion strains of opposite mating type marked with a different selectable marker. After meiosis, sporulation and selection for double mutants, synthetic genetic interactions are identified when specific combinations of mutations cause cell death (synthetic lethality) or retard growth (synthetic sickness). Such aggravating genetic interactions often occur when two nonessential gene products impinge on the same essential cellular process or function within different pathways such that one pathway can functionally compensate for defects in the other. Hence, identification of epistasis on a large scale can provide a global map of the functional relationships between genes and pathways[5]. Gene clustering based on the similarity of genetic interaction patterns can reveal novel components of pathways and the roles of various subunits of a particular protein complex[5–7].

Bacteria are hardy organisms, which suggests that extensive functional redundancy occurs in prokaryotes as well. Indeed, only 303 genes had been identified as essential for growth of *E. coli* K12 laboratory strain, a leading model microbe, in rich medium[8]. Experimental examples of epistasis have been documented in *E. coli* and other prokaryotes[9]. However, the extent of functional cross-talk among the ~4,200 genes of *E. coli* remains uncertain as no systematic unbiased genome-wide surveys of genetic interactions have been reported.

The principle of meiotic assortment is not applicable to bacteria. However, *E. coli* can use conjugation to exchange genetic information. Certain strains (F^+) harbor a transmissible low-copy episome, the fertility (F) factor[10], which promotes its own replication and transfer via conjugation into an F^- bacterial cell. At a low frequency, the F factor integrates into the host chromosome, giving rise to high frequency of recombination (Hfr) strains[10]. In this scenario, concomitant chromosomal DNA transfer occurs in a unidirectional

[1]Banting and Best Department of Medical Research, Terrence Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, 160 College Street, Toronto M5S 3E1, Canada. [2]Life Sciences Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, California 94720, USA. [3]Department of Biology, Wilfrid Laurier University, 75 University Avenue West, Waterloo N2L 3C5, Canada. [4]Department of Molecular Genetics, University of Toronto, 1 King's College Circle, Toronto M5S 1A8, Canada. [5]Graduate School of Biological Sciences, Nara Institute of Science and Technology, 8916-5 Takayama, Ikoma, Nara 630-0101, Japan. [6]Department of Biological Sciences, Purdue University, 915 W. State Street, West Lafayette, Indiana 47907, USA. [7]Present address: Life Sciences Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, California 94720, USA. [8]These authors contributed equally to this work. Correspondence should be addressed to J.F.G. (jack.greenblatt@utoronto.ca) or A.E. (andrew.emili@utoronto.ca).

RECEIVED 11 MARCH; ACCEPTED 19 JUNE; PUBLISHED ONLINE 1 AUGUST 2008; CORRECTED ONLINE 17 AUGUST 2008 (DETAILS ONLINE); DOI:10.1038/NMETH.1239

manner, starting with loci adjacent to the episome origin of transfer[10], *oriT*, until mating is interrupted. Once inside the recipient, the donor DNA is integrated into the recipient chromosome by homologous recombination by endogenous host factors[11].

We exploited this capacity to develop an array-based high-throughput bacterial genetic interaction assay, *E.* coli synthetic genetic array (eSGA), for comprehensive bacterial genetic-interaction screens. We showed that conjugation can drive efficient genetic exchange of a marked query gene mutation from an *E. coli* Hfr donor strain into a genome-wide high-density arrayed collection of *E. coli* F⁻ recipient single gene mutants. We then used robotic pinning, dual marker selection and colony imaging to identify quantitative colony growth defects in the resulting double mutants, revealing functional dependencies and pathway redundancy. Hierarchical clustering of the genetic interaction profiles obtained for 39 independent genome-scale screens revealed the modular architecture of *E. coli* pathways synthesizing iron-sulfur (Fe-S) clusters and suggested new biological relationships.

RESULTS

Establishment of the eSGA system

A schematic flowchart illustrating the eSGA procedure is shown in **Figure 1**. For each screen, we constructed a query gene deletion mutant by replacing the target gene with a chloramphenicol-resistance marker (*cat*) in an *E. coli* Hfr Cavalli donor strain bearing an integrated temperature-inducible λ-Red high-efficiency homologous recombination system[12] (**Supplementary Fig. 1** online).

The recipient mutant strains (**Supplementary Table 1** online) arrayed in twenty-five 384-well plates included the Keio single gene deletion mutant strain collection[8] covering 3,968 nonessential single gene replacements marked with a kanamycin-resistance cassette (*kan*). Of these, 3,956 were represented by two independent single gene deletion isolates, for a total of 7,924 single gene deletion mutants[8].

Hereafter, we refer to each of these two independent mutants as 'isolate 1' and 'isolate 2'. To this collection, we also added 149 F⁻ potentially hypomorphic *kan*-marked strains with essential genes involved in conserved bacterial processes[13] tagged with a gene encoding a C-terminal sequential peptide affinity tag (*SPA*).

Then we constructed double mutants via conjugation of the query mutant strain with recipient F⁻ mutant strains by pinning both strains onto solid Luria-Bertani (LB) medium (**Supplementary Methods** online) where the chloramphenicol resistance–marked query gene on the donor chromosome is transferred into the recipient strain and integrated into the recipient chromosome via homologous recombination. Subsequently, we selected the resulting chloramphenicol- and kanamycin-resistant double mutants on LB containing both kanamycin and chloramphenicol. Finally, we digitally imaged the plates and, for whole genome screens, scored colony growth and determined fitness statistics (**Supplementary Methods**). The eSGA screening process was automated by performing the replica transfer and pinning steps robotically.

Parameter testing and optimization

The frequency of conjugative exchange is related to the genetic distance of the query locus relative to *oriT*, the conjugation time and other parameters[14]. In Hfr Cavalli, *oriT* is located at 12.2 min[15] and the chromosome is transferred counterclockwise (**Fig. 2a**). We

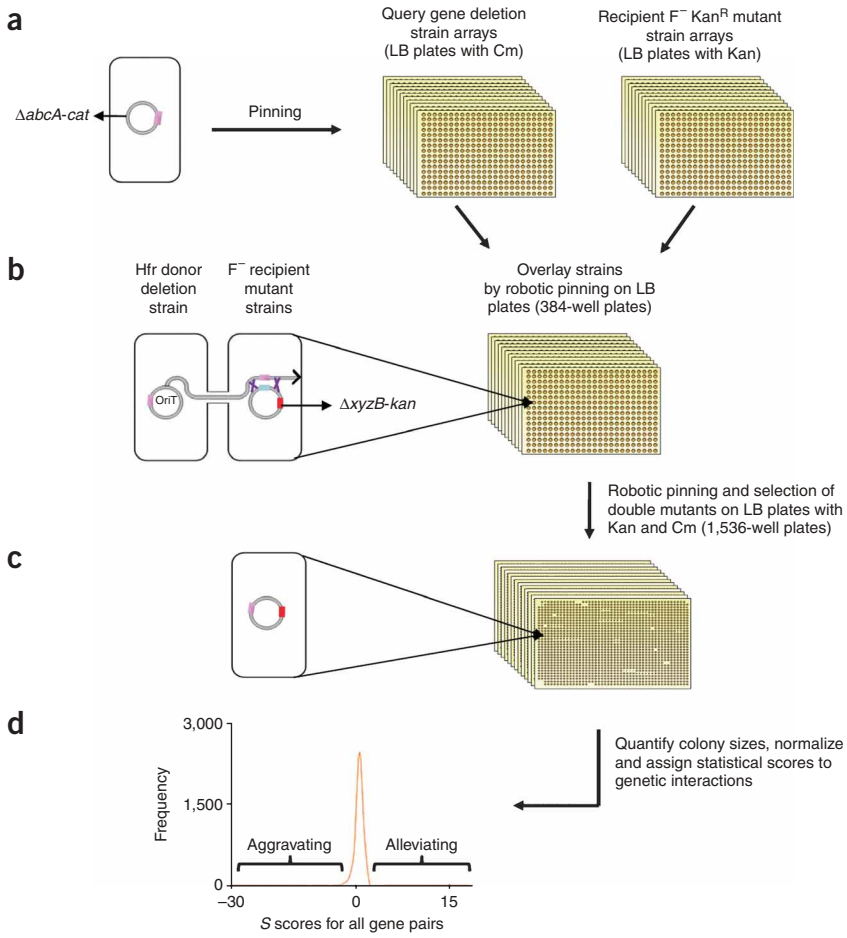

Figure 1 | eSGA outline. (**a**) An Hfr strain with a query gene deletion mutation marked with *cat* (chloramphenicol-resistance gene; pink box) is grown overnight in liquid LB with chloramphenicol (Cm) and pinned onto LB-Cm plates. Simultaneously, the recipient F⁻ mutant array strains marked with *kan* (kanamycin-resistance gene; red box) are pinned onto LB-kanamycin (Kan) plates. (**b**) After overnight growth the Hfr query strain colonies and recipient array colonies are pinned over each other on LB plates. Conjugation ensues, with transfer initiating at a specific origin of transfer, *oriT*, and proceeding via a rolling-circle replication. The donor chromosome undergoes homologous recombination with the recipient chromosome. (**c**) The resulting colonies are pinned onto plates containing both kanamycin and chloramphenicol for selection of double mutants. (**d**) The double mutant plates are then imaged and colony sizes are scored to identify aggravating and alleviating interactions.

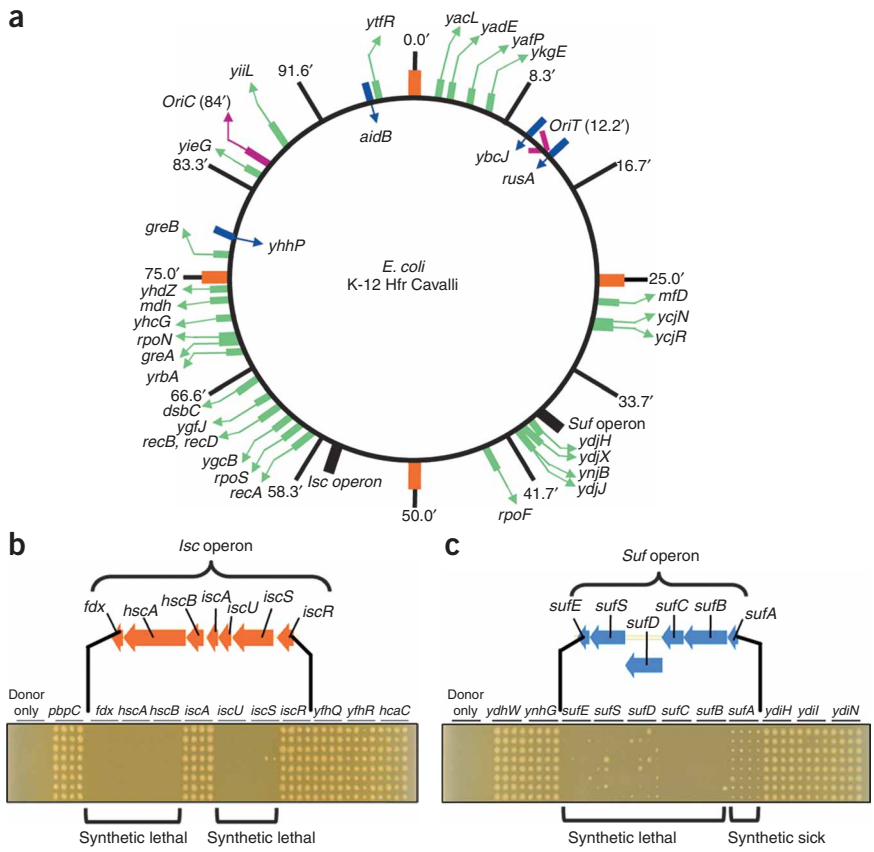

Figure 2 | Map of the query mutations and detection of aggravating genetic interactions by eSGA. (**a**) Chromosomal positions of the query mutations used in this study. Green arrows indicate the deletions (marked with *cat*) used as queries in genome-wide screens, and blue arrows indicate deletions (marked with *cat*) used as queries in mini-screens. Chromosome coordinates are indicated in minutes (′). *OriT* is the F origin of transfer in Hfr Cavalli, and *oriC* the chromosomal origin of replication. Query locus *ybcJ* located close to *oriT* transfers early, whereas *rusA* on the other side of *oriT* transfers late[18]. (**b**,**c**) Mini-screens showing synthetic lethal or synthetic sick relationships between components of the functionally redundant Isc and Suf pathways. In **b**, a query strain bearing a deletion of *sufC* marked with *cat* is crossed against recipient strains bearing individual gene deletions in and around the *isc* operon. In **c**, results of a converse experiment are shown, in which a Δ*iscU-cat* donor strain was conjugated with strains bearing null mutations in and around the *suf* operon. "Donor only" represents a negative control with no recipient strain.

first performed small-scale experiments to determine the feasibility of conjugation as a mechanism of marker transfer to create double mutants and determine optimum parameters for conjugation (**Supplementary Results** online).

Proof of concept: interactions between *isc* and *suf* pathways

To confirm that aggravating genetic interactions can be detected by our eSGA method, we monitored genetic interactions between the components of two well-studied, functionally redundant pathways, encoded by the *isc* and *suf* operons, which jointly mediate the essential and highly conserved process of Fe-S cluster biosynthesis[16]. Whereas null mutation(s) of single components yield viable strains[17], inactivation of the Isc and Suf pathways simultaneously results in synthetic sick or lethal (SSL) mutants[18,19].

In small-scale experiments we transferred deletions of *sufC* and *iscU*, key genes in the Suf and Isc pathways, respectively, into recipient strains bearing null mutations in and around the chromosomal region of the respective buffering pathways (**Fig. 2b**,**c**). As expected, viability was abolished for double mutants inactivating core components of both pathways, but not for combinations involving unrelated adjacent genes. The Δ*iscU-cat* and Δ*sufA-kan* combination caused a synthetic-sick phenotype, consistent with recent work indicating partial complementation of *sufA* loss by *iscA*[20].

We observed similar genetic interaction patterns with other *isc* and *suf* gene donors (data not shown), confirming that functionally informative SSL genetic interactions between components of parallel pathways are readily detected by eSGA.

Development of a genome-wide assay and data normalization

We devised assay configurations for high-density, genome-wide robotic screening which minimized spurious technical variation (**Supplementary Methods**). Conjugation was performed on solid medium in a 384-spot configuration using four biological replicate recipient colonies, two for each of isolate 1 and isolate 2. Two-rounds of selective outgrowth in 1,536-colony format then produced consistent double mutant colony sizes (**Supplementary Fig. 2** online), allowing for more accurate quantification of growth. We digitally imaged the final selective plates and quantified the data using an automated image processing system originally devised for yeast[6].

Although synthetic lethality is straightforward to score visually (that is, no viable colony is detected), our automated quantitative scoring system greatly improved the informational content per screen as the recorded genetic interactions reflected gradations in fitness rather than arbitrary thresholds. The expectation is that, as for yeast[21], the growth rate of a double mutant involving two functionally unrelated genes should be approximately the product of the relative growth rates of the individual single mutants. Double mutants that grow considerably more slowly therefore represent an aggravating interaction, implying the genes may be in redundant parallel pathways, whereas double mutants that grow more rapidly than expected represent alleviating interactions, implying the two genes belong to the same pathway or that one mutation suppresses the mutant phenotype associated with the other.

Ultimately, we inferred fitness defects as a function of normalized growth rates relative to unperturbed reference controls. However, because systematic biases, such as inter-plate variation and plate-edge effects[6], and stochastic growth artifacts owing to pinning defects, could lead to spurious fluctuations in colony sizes, we normalized the raw data from each screen and then analyzed them statistically to take into account replicate reproducibility and

deviations from median measurements. To reduce any residual error, we then also normalized the results obtained for unrelated donors based on the median double mutant fitness recorded for each screen (**Supplementary Methods**). We then calculated an interaction score (*S*) to quantify the strength and confidence of genetic interaction determined for each mutant gene pair. Negative *S* scores correspond to putative aggravating interactions and positive *S* scores, to putative alleviating interactions. Moreover, to provide a biological measure of relative fitness between double mutants, we also measured the magnitude of the fitness defect for each significant ($P < 0.0001$) mutant gene pair by calculating the $\log_2(Q/R)$ ratio, where *Q* represents normalized colony size of each double mutant, and *R* represents the average normalized median colony size of all double mutants arising from the same donor strain screen.

Global profiling with 39 genome-wide screens

To validate the effectiveness of the eSGA assay, we performed 39 independent genome-wide screens (**Supplementary Table 2** online). We first screened 13 functionally related genes encoded by the *iscRSUA-hscBA-fdx* (Isc system[18]) and *sufABCDSE* (Suf system[19]) operons as a benchmark because of extensive documentation of functional redundancy in the literature. We next queried an additional 26 functionally unrelated genes primarily to control for biological and technical false positives and to provide a greater breadth of data with which to normalize potentially locus-specific effects (that is, uniqueness of the profiles). These additional donors contained mutations affecting RNA polymerase (RNAP)-associated factors (*mfd*, *fliA*, *greA*, *greB*, *rpoS*, *rpoN* and *yacL*), homologous recombination (*recD*), disulphide isomerase (*dsbC*), malate dehydrogenase (*mdh*) and 16 genes of unknown function (*yadE*, *yafP*, *ydjH*, *ydjX*, *ygfJ*, *yhdZ*, *yieG*, *yhcG*, *yiiL*, *ycjR*, *ynjB*, *ygcB*, *yrbA*, *ykgE*, *ycjN* and *ytfR*).

Evaluation of data quality

We assessed reproducibility between replicate screens by calculating the Pearson correlation coefficient of the averaged, normalized colony sizes. Typically, this coefficient for replicate screens was 0.5–0.7 (**Fig. 3a**), and we repeated and rescored screens with a coefficient <0.5. As expected, the results obtained for isolate-1 and isolate-2 deletion strains were highly correlated (data not shown), although occasionally the scores differed markedly (**Supplementary Fig. 3** online), suggesting a defect in one strain.

Closely linked genes should exhibit negative *S* scores because they recombine with lower frequencies and thus fail to create double mutants[22]. Indeed, this effect was evident in eSGA and was inversely proportional to genetic distance (**Supplementary Fig. 4** online). An example of the linkage effect in a systematic genome-wide screen using *sufA* as query gene is shown in **Figure 3b**. Such linkage suppression, also evident in yeast SGA screens[6], has been used to confirm the correctness of query mutations[6]. However, linkage suppression was generally undetectable 30 kbp from the query.

The *S* scores generated with the 39 query strains approximated a normal distribution (**Fig. 3c**). Hence, given the possibility of some residual noise, we estimated the statistically significant fraction of observed *S* scores (that is, deviating from random expectation) by permutation testing (**Supplementary Methods**). This false discovery rate assessment revealed that in the experimental screens a substantially greater number of interactions in the tail of the distribution represented aggravating genetic interactions (**Fig. 3c**).

Functional enrichment

As *S* scores deviating significantly ($P < 0.0001$) from the mean represent candidates for functional associations[23], we evaluated the extent to which existing knowledge about functionally related gene pairs was accurately reflected at various *S* score thresholds (**Fig. 3d**).

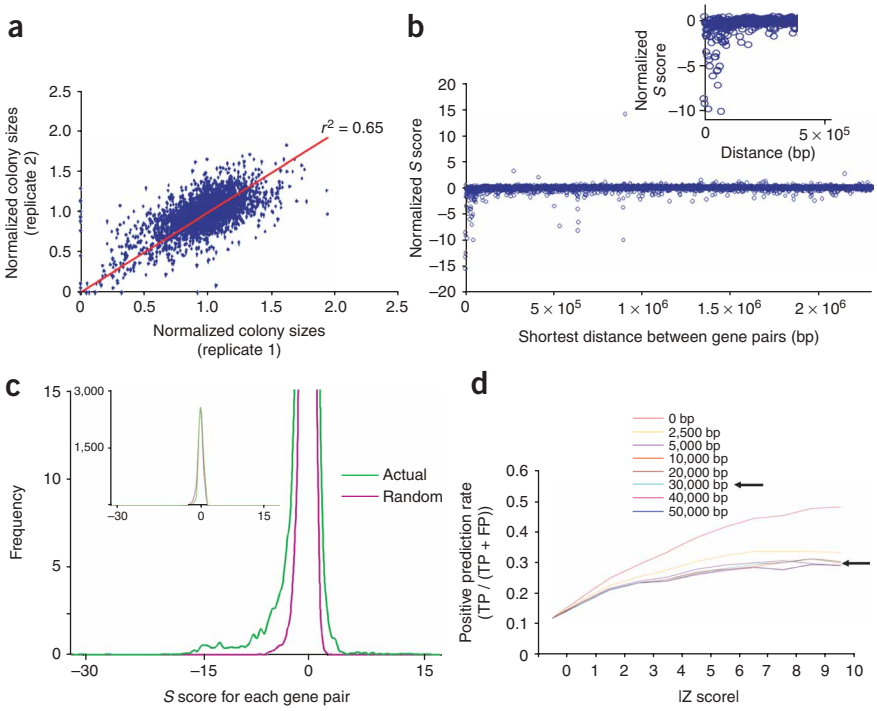

Figure 3 | Evaluation of data quality and statistical significance of interaction (*S*) scores from genome-wide screens. (**a**) An example showing biological reproducibility between replicate experiments from Δ*sufA-cat* genome-wide screens. Pearson correlation coefficient is given for normalized colony sizes from the replicate screens. Red line represents best fit. (**b**) Effect of linkage on recombination efficiency using Δ*sufA-cat* as a query mutation. The expanded region shows closely linked genes that generate viable recombinants with reduced frequency. Negative *S* scores correspond to aggravating interactions, and positive *S* scores indicate putative alleviating interactions. (**c**) Distribution plots of average *S* scores from all genes used as queries (actual), as well as from randomized datasets (random) for all screens. The overall distribution is shown in the inset. (**d**) Significance of quantitative genetic interactions was calculated using standard *Z* score transformation. Functional association was also determined for all interactions after removing adjacent linked genes at the indicated distances from the query loci. The arrow indicates that all recipient genes located within 30 kbp of the query gene were removed from consideration.

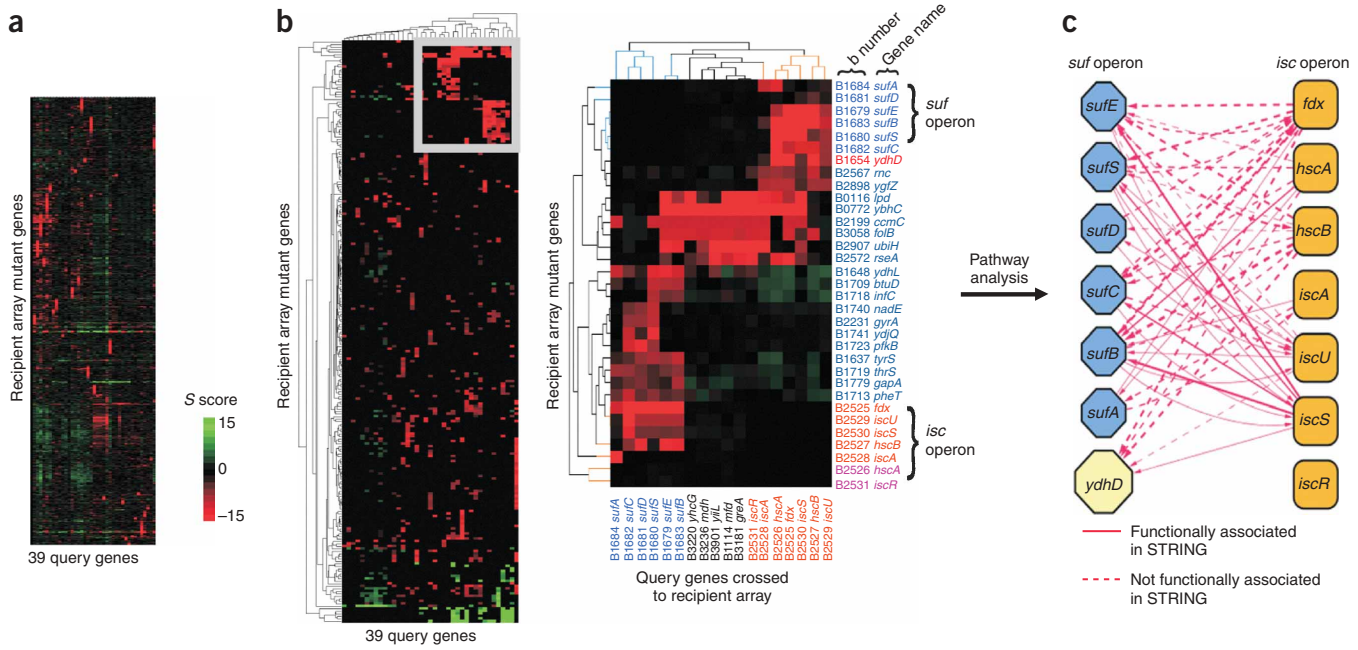

Figure 4 | Clustering of genetic interaction profiles from genome-wide screens. (**a**) Heat map of the *S* scores generated from 39 genome-wide screens. (**b**) Two dimensional hierarchical clustergram of the high confidence genetic interactions, generated using a stringent cut-off with |Z score| ≥ 4 ($P < 0.0001$) and re-clustering the profiles shown in **a** after removing interactions with genes located within 30 kbp of the query genes. Red represents aggravating (negative *S* score), green alleviating (positive *S* score) and black the absence of genetic interactions. A subset of the high-confidence genetic interaction network (gray box) highlighting the known Isc (orange) and Suf (blue) pathways is expanded for easier visualization. The recipient strains with the supposed *iscR* and *hscA* (magenta) mutations were defective (data not shown). The *ydhD* gene (red) displays interactions with the Isc pathway. (**c**) A schematic genetic interaction network of *isc* and *suf* pathway genes and their functional association in the STRING database. Line thickness reflects the strength of the interaction scores between gene pairs. The *ydhD* gene has SSL interactions with the members of the Isc system.

We applied standard *Z* score transformation to calculate significance and calculated functional enrichment, expressed as the positive prediction rate (PPR), for each value of |Z score|. For functional associations between *E. coli* genes, we obtained annotated and predicted interactions from the Search Tool for the Retreival of Interacting Genes/Proteins (STRING) database[24] using a medium (0.4) confidence cut-off value.

The PPR increased with increasing |Z score|, approaching 0.5 at high |Z score| with no linkage correction (**Fig. 3d**). Because functionally linked genes are often in the same operon on bacterial chromosomes, we progressively removed closely linked genes from consideration to mitigate this effect. Although the PPR was reduced to 0.28 once we removed from consideration genes within 30 kbp of the query locus, the functional enrichment was still substantially greater than the positive prediction rate for |Z score| = 0.

Projecting pathway relationships and functional modules

Genes in the same pathway should display closely correlated patterns of aggravating interactions with genes in parallel, functionally redundant pathways[5]. Hence, we performed two-dimensional hierarchical clustering of the *S* scores to analyze the roles of individual genes in various pathways and processes (**Fig. 4a**). Based on the functional-enrichment analysis (**Fig. 3d**), we reclustered the profiles after accounting for linkage (30-kbp window) and used a stringent |Z score| cut-off of ≥ 4 ($P < 0.0001$) to create a high-confidence dataset of 1,288 genetic interactions (**Fig. 4b** and **Supplementary Table 3** online). Among these, we detected 799 genetic interactions for nonessential genes, with 730 categorized as aggravating interactions and only 69 as alleviating interactions. We then compared the *S* scores of the high-confidence interaction dataset for nonessential genes to the $\log_2(Q/R)$ for each pair of mutants. Although all the significant *S* score mutant gene pairs with aggravating interactions exhibited a similar interaction trend with the $\log_2(Q/R)$ (**Supplementary Table 3**), only 41% (28/69) of the alleviating interactions were concordant, suggesting these tend to be biologically subtle. As expected, the genes of the Isc and Suf pathways formed distinct clusters (**Fig. 4b**). Moreover, mutations of individual Isc pathway components (*iscS*, *iscU*, *hscA*, *hscB* and *fdx)* all resulted in aggravating interactions when combined with mutations in the Suf system. Exceptions were the *iscS-sufD* and *hscB-sufD* combinations, which have weaker SSL interactions. Using *iscA* as query, we observed an aggravating interaction with *sufA*, but only very weak or no SSL interactions with other members of the Suf operon. This observation is in good agreement with a recent report demonstrating a synthetic sick *iscA-sufA* combination[20], implying that *sufA* partially complements loss of *iscA* function.

The Isc components interacted not only with the Suf pathway, but also with certain genes outside the Suf system, including *ydhD* (monothiol glutaredoxin), *ygfZ* (folate-binding, tRNA-modifying enzyme), *gntY* (paralogue of *iscA* and *sufA)*, *hemF* (oxygen-dependent coproporphyrinogen oxidase) and *yrbA* (a *bolA*-like protein) (**Fig. 4b**). Some interactions also appear to be highly specific, such as *iscS* with *truA* (tRNA pseudouridine synthase), both of which were previously reported to be involved in tRNA modification[25].

To validate the reliability of the novel interactions, we confirmed epistasis by reanalyzing individual donor-recipient combinations using a custom mini-array (**Supplementary Results**). For example, we confirmed the strong SSL interactions of a donor lacking *ydhD* when conjugated to *iscSUA*, *hscAB* and *fdx* deletion recipient strains (**Supplementary Fig. 5** online). Conversely, we isolated viable double mutants when conjugating *ydhD* with strains deleted for components of the Suf pathway.

Interaction properties of essential genes

Over one-third (489/1,288) of our high-confidence significant interactions involved a hypomorphic allele of an essential gene (394 aggravating, 95 alleviating), suggesting that essential bacterial genes have more functional links than nonessential genes, as is also evident in yeast[26]. Virtually all (392/394) of the significant ($P < 0.0001$) aggravating interactions exhibited a similar interaction trend with $\log_2(Q/N)$, whereas only half (46/95) of the alleviating interactions did so (**Supplementary Table 3**). The observed SSL relationships included interactions of query deletion of *greA* and *greB*, which encode closely related transcription factors, with the *rpoA-SPA*, which encodes an SPA-tagged core subunit of RNAP, as well as with *rpoN*, which encodes the RNAP σ^N factor; and a query deletion of *yacL*, which encodes a putative elongation factor, interacted with *rpoC-SPA*, which encodes SPA-tagged RNAP β′ subunit. Interpretation of SSL interactions involving potential hypomorphs is, however, more complex than those involving null alleles because in the majority of cases the nature of the observed hypomorphic defect is unknown.

DISCUSSION

We authenticated our eSGA method by verifying known pathway relationships between the Suf and Isc systems[18,19], which contain many paralogous components (**Supplementary Discussion** online). After data normalization and filtering, we recorded an average of approximately 20 high-confidence genetic interactions per query gene for nonessential genes, which is on par with the ~30 typically observed for nonessential yeast genes[5]. However, our 39 screens may not be representative of *E. coli* genes in general, with results likely representing minimum estimates for two reasons: first, we used a stringent |Z score| cutoff to minimize false positives, which may have eliminated genuine interactions; second, even though functionally related genes tend to cluster in the bacterial chromosome, we removed genetically linked genes (within 30 kbp) from consideration to eliminate the linkage effect.

In our data assessment, the *S* score and $\log_2(Q/R)$ value agreed to a greater degree for aggravating interactions than for alleviating interactions, where more than 50% of the significant alleviating interactions (*S* score) diverged dramatically from the $\log_2(Q/R)$ value, suggesting the possibility of misinterpretation of certain large-scale genetic interaction data when we used only one model. Subsequently, at least two independent measurement models may be necessary to determine the extent of functional relationship.

Our assay identified many highly significant ($P < 0.0001$) interactions, some of which define previously unidentified functional relationships even within two heavily studied Fe-S cluster pathways (**Supplementary Discussion**). For example, *ydhD* recipient mutant strain displayed aggravating interactions with all members of the *isc* operon, including *iscA*, but no discernable interactions with the Suf system (**Fig. 4b,c** and **Supplementary Fig. 5**). Reports of YdhD facilitating redox reactions by the thioredoxin TrxAB reductase system, also used for *in vitro* Fe-S cluster assembly[27], as well as of *sufABCDSE* operon[19] induction and increased *ydhD* amounts under iron depletion conditions[28], correlate with our findings and, together with our results, implicate YdhD as a previously unrecognized factor working with the Suf system.

Using the eSGA assay we detected both previously characterized and new functional relationships, supporting the efficacy of this approach for systematic, genome-wide, quantitative, double-deletion surveys. Despite the advantages of eSGA, potential artifacts may have been introduced by the use of nonisogenic donor strains (see **Supplementary Discussion**). Although we suspect that, in practice, screening artifacts rarely occur, nevertheless an Hfr donor strain isogenic with the BW25113 KEIO gene deletion collection strain background has been developed recently (H. Mori and B. Wanner; unpublished data). This strain is fully compatible with the screening strategy reported here and will be used in future experiments to define new components of existing pathways as well as the mechanisms underlying genetic robustness and phenotypic buffering in *E. coli*. The large number of genetic interactions detected for the relatively small subset of *E. coli* mutant gene pairs examined to date suggests that the concept of a 'minimal genome'[29] must accommodate the notion of epistasis. Given that conjugation is widespread in prokaryotes, donor-recipient pairs need not necessarily be restricted to *E. coli* (**Supplementary Discussion**). Thus, eSGA could potentially be extended to examine genetic interactions in other bacterial species.

METHODS

Bacterial strains, plasmids, media and reagents. Hfr Cavalli was obtained from the *E. coli* Genetic Stock Centre. The Keio strain collection has been described previously[8]. Standard LB agar media were used to grow the *E. coli* deletion strains with or without antibiotics at concentrations of 34 μg/ml chloramphenicol and 50 μg/ml kanamycin (Sigma). Double mutants were cultured at 32 °C on rectangular plates.

Construction of *E. coli* double mutants using automated strain arraying. We developed a 6-day strain-handling procedure using a replica-pinning RoToR-HDA benchtop robot (Singer Instruments) to automate double mutant construction. The complete step-wise protocols are described in **Supplementary Methods**.

Functional enrichment. The raw and normalized colony sizes used to generate the *S* scores from the genome-wide screens are shown in **Supplementary Table 4** online. Significance of quantitative genetic interactions was calculated using a standard *Z* score formula. To calculate the positive prediction rate [TP / (TP + FP)], true positives (TP) were defined as functional associations predicted by eSGA that have annotation links in STRING and false positives (FP) were defined as eSGA hits without supporting evidence. A |Z score| cut-off ≥ 4, corresponding to $P < 0.0001$, where the interaction score (*S*) is on average ≤ –2.78 for aggravating and ≥ 2.61 for alleviating interactions, was considered statistically significant. Two-dimensional clustering was performed according to the similarity of genetic interaction profiles as calculated using an uncentered correlation distance metric with average linkage implemented in Cluster 3.0 (ref. 30).

Note: Supplementary information is available on the Nature Methods website.

ACKNOWLEDGMENTS

We thank M. Costanzo for comments on the manuscript. This work was supported by a Canadian Institute of Health Research (CIHR) grant (82852) to J.F.G. and A.E., by a US National Institutes of Health grant to B.L.W. (GM62662), by Grant-in-Aid for Scientific Research on Priority Areas from the Ministry of Education, Sports, Science and Technology of Japan Core Research for Evolutional Science and Technology, and Japan Science and Technology grants to H.M., by a CIHR grant (GSP-41567) to B.J.A. and C.B., by a partial postdoctoral fellowship from the Mexican Science and Technology Research Council to J.J.D.-M., and by a Laboratory Directed Research and Development grant to G.B. The phage-λ Red system was kindly provided by D.L. Court (National Cancer Institute).

AUTHOR CONTRIBUTIONS

M.B. and G.B. coordinated experimental design and data analysis. M.B., G.B., H.L., J.W., P.V., C.C., L.L., M.A., K.A.D., N.Y., H.M. and B.L.W. performed optimization studies and constructed recipient arrays. J.L., A.S.-N., W.Y., A.G.G., O.P., G.B. and M.B. generated query deletion knockouts. M.B., F.B., N.S., S.C., J.-Y.C. and A.N.-A. performed genome-wide screens. M.B., S.P., A.D., S.J., B.S., H.D., B.J.A. and C.B. designed and performed quantitative image analysis. J.J.D.-M., M.B., S.P. and G.M.-H. performed informatics analyses. G.B., B.G. and W.Y. performed validation experiments. G.B., M.B., J.F.G. and A.E. drafted the manuscript. J.F.G. and A.E. conceived, designed and directed the project.

Published online at http://www.nature.com/naturemethods/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Boone, C., Bussey, H. & Andrews, B.J. Exploring genetic interactions and networks with yeast. *Nat. Rev. Genet.* **8**, 437–449 (2007).
2. Costanzo, M., Giaever, G., Nislow, C. & Andrews, B. Experimental approaches to identify genetic networks. *Curr. Opin. Biotechnol.* **17**, 472–480 (2006).
3. Beyer, A., Bandyopadhyay, S. & Ideker, T. Integrating physical and genetic maps: from genomes to interaction networks. *Nat. Rev. Genet.* **8**, 699–710 (2007).
4. Tong, A.H. *et al.* Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science* **294**, 2364–2368 (2001).
5. Tong, A.H. *et al.* Global mapping of the yeast genetic interaction network. *Science* **303**, 808–813 (2004).
6. Collins, S.R., Schuldiner, M., Krogan, N.J. & Weissman, J.S. A strategy for extracting and analyzing large-scale quantitative epistatic interaction data. *Genome Biol.* **7**, R63 (2006).
7. Collins, S.R. *et al.* Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map. *Nature* **446**, 806–810 (2007).
8. Baba, T. *et al.* Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2** 2006.0008 (2006).
9. Kogoma, T. Stable DNA replication: interplay between DNA replication, homologous recombination and transcription. *Microbiol. Mol. Biol. Rev.* **61**, 212–238 (1997).
10. Ippen-Ihler, K.A. & Minkley, E.G. Jr. The conjugation system of F, the fertility factor of *Escherichia coli*. *Annu. Rev. Genet.* **20**, 593–624 (1986).
11. Kowalczykowski, S.C., Dixon, D.A., Eggleston, A.K., Lauder, S.D. & Rehrauer, W.M. Biochemistry of homologous recombination in *Escherichia coli*. *Microbiol. Mol. Biol. Rev.* **58**, 401–465 (1994).
12. Yu, D. *et al.* An efficient recombination system for chromosome engineering in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **97**, 5978–5983 (2000).
13. Butland, G. *et al.* Interaction network containing conserved and essential protein complexes in *Escherichia coli*. *Nature* **433**, 531–537 (2005).
14. Susman, M. General bacterial genetics. *Annu. Rev. Genet.* **4**, 135–176 (1970).
15. Bachmann, B.J. Pedigrees of some mutant strains of *Escherichia coli* K-12. *Bacteriol. Rev.* **36**, 525–557 (1972).
16. Fontecave, M., Ollagnier de Choudens, S. & Barras, B. Py, F. Mechanisms of iron-sulfur cluster assembly: the SUF machinery. *J. Biol. Inorg. Chem.* **10**, 713–721 (2005).
17. Tokumoto, U., Kitamura, S., Fukuyama, K. & Takahashi, Y. Interchangeability and distinct properties of bacterial Fe-S cluster assembly systems: functional replacement of the isc and suf operons in *Escherichia coli* with the nifSU-like operon from *Helicobacter pylori*. *J. Biochem.* **136**, 199–209 (2004).
18. Tokumoto, U. & Takahashi, Y. Genetic analysis of the *isc* operon in *Escherichia coli* involved in the biogenesis of cellular iron-sulfur proteins. *J. Biochem.* **130**, 63–71 (2001).
19. Outten, F.W., Djaman, O. & Storz, G. A suf operon requirement for Fe–S cluster assembly during iron starvation in *Escherichia coli*. *Mol. Microbiol.* **52**, 861–872 (2004).
20. Lu, J., Yang, J., Tan, G. & Ding, H. Complementary roles of SufA and IscA in the biogenesis of iron-sulfur clusters in *Escherichia coli*. *J. Biochem.* **409**, 535–543 (2008).
21. Mani, R., St. Onge, R.P., Hartman, J.L., Giaever, G. & Roth, F.P. Defining genetic interaction. *Proc. Natl. Acad. Sci. USA* **105**, 3461–3466 (2008).
22. Fox, M.S. Some features of genetic recombination in prokaryotes. *Annu. Rev. Genet.* **12**, 47–68 (1978).
23. St. Onge, R.P. *et al.* Systematic pathway analysis using high-resolution fitness profiling of combinatorial gene deletions. *Nat. Genet.* **39**, 199–206 (2007).
24. Von Mering, C. *et al.* STRING 7-recent developments in the integration and prediction of protein interactions. *Nucleic Acids Res.* **35**, D358–D362 (2007).
25. Lauhon, C.T. Requirement for IscS in biosynthesis of all thionucleosides in *Escherichia coli*. *J. Bacteriol.* **184**, 6820–6829 (2002).
26. Davierwala, A.P. *et al.* The synthetic genetic interaction spectrum of essential genes. *Nat. Genet.* **37**, 1147–1152 (2005).
27. Ding, H., Harrison, K. & Lu, J. Thioredoxin reducatse system mediates iron binding in iscA and iron delivery for the iron-sulfur cluster assembly in IscU. *J. Biol. Chem.* **280**, 30432–30437 (2005).
28. Fernandes, A.P. *et al.* A novel monothiol glutaredoxin (Grx4) from *Escherichia coli* can serve as a substrate for thioredoxin reductase. *J. Biol. Chem.* **280**, 24544–24552 (2005).
29. Mizoguchi, H., Mori, H. & Fujio, T. *Escherichia coli* minimum genome factory. *Biotechnol. Appl. Biochem.* **46**, 157–167 (2007).
30. Eisen, M.B., Spellman, P.T., Brown, P.O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA* **95**, 14863–14868 (1998).